IMAGE RECOGNITION BASED ON SEPARABLE LATTICE TRAJECTORY 2-D HMMS

Akira Tamamori, Yoshihiko Nankaku, Keiichi Tokuda

Nagoya Institute of Technology, Nagoya, Japan

ABSTRACT

In this paper, a novel statistical model for image recognition based on separable lattice 2-D HMMs (SL2D-HMMs) is proposed. Although SL2D-HMMs can model invariance to size and location deformation, its modeling accuracy is still insufficient because of the following two assumptions: i) the statistics of each state are constant and ii) the state output probabilities are conditionally independent. In this paper, SL2D-HMMs are reformulated as a trajectory model that can capture dependencies between adjacent observations. The effectiveness of the proposed model was demonstrated in face recognition and image alignment experiments.

Index Terms— image recognition, hidden Markov models, separable lattice 2-D HMMs, trajectory HMMs.

1. INTRODUCTION

With the recent development of image recognition techniques, image recognition systems based on statistical approaches such as eigenface methods [1] and subspace methods [2] have become popular in many applications. In these image recognition systems, heuristic normalization techniques included in the pre-process part of the classification have been applied. Although high recognition performance can be obtained by using these heuristic techniques, it is still required for human to develop such techniques for each task by using task dependent information. Furthermore, in image recognition, the final objective is not to accurately normalize images for humans 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 are such statistical approaches and have been proposed recently to reduce the influence of geometric variations [3–7]. Geometric matching between input images and model parameters is represented by discrete hidden variables, and the normalization process is included in calculating probabilities. For another HMM based approach, separable lattice 2-D HMMs (SL2D-HMMs) were proposed [8] to reduce computational complexity while retaining outstanding properties that model multi-dimensional data. SL2D-HMMs can perform elastic matching both horizontally and vertically, which makes it possible to model not only invariance to the size and location of an object but also nonlinear warping in all dimensions. Nevertheless, due to the composite structure of hidden variables, SL2D-HMMs have the same constraints as 1-D HMMs [9] in that (i) the statistics of each state do not change dynamically and (ii) the output probability of an observation vector depends only on the current state, not on any other states nor observations. To capture the dependencies and improve the recognition performance, it was reported that using dynamic features (e.g., 1st and 2nd order delta coefficients) [10] can be effective [11, 12]. However, the static and dynamic features are assumed to be independent variables and these relationships are ignored even though the relationships between the static and dynamic features are essentially deterministic.

In previous work [13], trajectory HMMs were proposed and successfully applied to speech recognition and speech synthesis. The standard HMM is reformulated by imposing the explicit relationship between static and dynamic features, in order that the constraint of HMMs such as the conditional independence and the constant statistics in each state can be relaxed. In this paper, we propose a novel generative model that reformulate SL2D-HMMs as a trajectory model, referred to as separable lattice trajectory 2-D HMMs (SLT2D-HMMs). The proposed model can overcome the shortcomings of separable lattice HMMs and capture the dependencies of adjacent observations. Consequently, the modeling ability can be significantly improved.

The rest of the paper is organized as follows. In section 2, SL2D-HMMs are explained briefly. In section 3, the structure of the proposed model is defined. In Section 4, we derive the training algorithm for the proposed model. In Section 5, we describe face recognition experiments on the XM2VTS database [14] and finally conclude in Section 6.

2. SEPARABLE LATTICE 2-D HMMS

Separable lattice 2-D hidden Markov models [8] are defined for modeling two-dimensional data. The observations of two-dimensional data, e.g., pixel values of an image, are assumed to be given on a two-dimensional lattice:

$$\boldsymbol{O} = \{\boldsymbol{O}_{\boldsymbol{t}} | \boldsymbol{t} = (t^{(1)}, t^{(2)}) \in \boldsymbol{T}\}, \tag{1}$$

where $\boldsymbol{t}$ denotes the coordinates of the lattice in two dimensional space $\boldsymbol{T}$ and $t^{(m)} = 1, \dots, T^{(m)}$ is the coordinate of the m-th dimension. The observation $\boldsymbol{O}_{\boldsymbol{t}}$ is emitted from the state indicated by the hidden variable $\boldsymbol{S}_{\boldsymbol{t}} \in \boldsymbol{K}$. The hidden variables $\boldsymbol{S}_{\boldsymbol{t}} \in \boldsymbol{K}$ can take one of $K = K^{(1)}K^{(2)}$ states, which are assumed to be arranged on an two-dimensional state lattice $\boldsymbol{K} = \{(1,1),(1,2),\dots,(K^{(1)},K^{(2)})\}$. In other words, a set of hidden variables, $\{\boldsymbol{S_t} | \boldsymbol{t} \in \boldsymbol{T}\}$, 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 $\boldsymbol{O_t}$ is dependent only on the state $\boldsymbol{S_t}$ as in ordinary HMMs, dependencies among hidden variables determine the properties and the modeling ability of two-dimensional HMMs. In SL2D-HMMs, to reduce the number of possible state sequences, the hidden variables are constrained to be composed of two Markov chains:

$$\begin{aligned} \boldsymbol{S} &= \{\boldsymbol{S}^{(1)}, \boldsymbol{S}^{(2)}\}, \quad (2)\\ \boldsymbol{S}^{(m)} &= \{S_1^{(m)}, \dots, S_{t^{(m)}}^{(m)}, \dots, S_{T^{(m)}}^{(m)}\}, \quad (3) \end{aligned}$$

where $\boldsymbol{S}^{(m)}$ 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: $\boldsymbol{S_t} = (S_{t^{(1)}}^{(1)}, S_{t^{(2)}}^{(2)})$. 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.

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

$$P(\boldsymbol{O}, \boldsymbol{S} \mid \Lambda) = P(\boldsymbol{O} \mid \boldsymbol{S}, \Lambda) \prod_{m=1,2} P(\boldsymbol{S}^{(m)} \mid \Lambda), \quad (4)$$

where Λ is a set of model parameters.

3. SEPARABLE LATTICE TRAJECTORY 2-D HMMS

In the previous section, we described the structure of SL2D-HMMs, where the hidden variables are composed of two independent 1-D Markov chains. Therefore, similar to the 1-D HMMs, the following two limitations are imposed on SL2D-HMMs [9]:

1. The statistics of each state do not change dynamically.
2. The output probability of the observation is conditionally independent, given the horizontal and vertical states.

To overcome these shortcomings, augmenting the dimensionality of static feature vectors (e.g., pixel values) by appending their dynamic feature vectors (e.g., delta and delta-delta coefficients) [10] to capture dependencies between adjacent observations can enhance the performance of the HMM-based speech recognizers [15]. Generally, dynamic features are calculated as regression coefficients from their neighboring static features. Therefore, the relationship between static and dynamic features is deterministic. However, this relationship is ignored, and static and dynamic features are modeled as independent statistical variables in standard HMMs. In the next section, the proposed model is derived in order to avoid this problem.

3.1. Reformulation as a trajectory model

In this paper, the observation vector $\boldsymbol{O_t}$ is assumed to consist of the M-dimensional static feature vector

$$\boldsymbol{C_t} = [C_{\boldsymbol{t},1}, C_{\boldsymbol{t},2}, \dots, C_{\boldsymbol{t},M}]^\top, \quad (5)$$

and horizontal/vertical dynamic feature vectors [1]

$$\Delta^{(H)}\boldsymbol{C_t} = [\Delta^{(H)}C_{\boldsymbol{t},1}, \Delta^{(H)}C_{\boldsymbol{t},2}, \dots, \Delta^{(H)}C_{\boldsymbol{t},M}]^\top, \quad (6)$$

$$\Delta^{(V)}\boldsymbol{C_t} = [\Delta^{(V)}C_{\boldsymbol{t},1}, \Delta^{(V)}C_{\boldsymbol{t},2}, \dots, \Delta^{(V)}C_{\boldsymbol{t},M}]^\top, \quad (7)$$

that is

$$\boldsymbol{O_t} = \left[\Delta^{(S)}\boldsymbol{C_t}^\top, \Delta^{(H)}\boldsymbol{C_t}^\top, \Delta^{(V)}\boldsymbol{C_t}^\top\right]^\top, \quad (8)$$

where $\Delta^{(S)}\boldsymbol{C_t} = \boldsymbol{C_t}$. Generally, these dynamic features are calculated as regression coefficients from their neighboring static features:

$$\Delta^{(d)}\boldsymbol{C_t} = \sum_{\tau=-L_-^{(d)}}^{L_+^{(d)}} w^{(d)}(\tau)\boldsymbol{C}_{\boldsymbol{t}^{(d,\tau)}}, \quad d = H, V, \quad (9)$$

where $\boldsymbol{t}^{(H,\tau)} = (t^{(1)}+\tau, t^{(2)})$ and $\boldsymbol{t}^{(V,\tau)} = (t^{(1)}, t^{(2)}+\tau)$. The observation vectors and static feature vectors on the 2-D lattice can be rewritten in $MT^{(1)}T^{(2)}$ size vector forms as

$$\boldsymbol{O} = \left[\boldsymbol{O}_{(1,1)}^\top, \boldsymbol{O}_{(1,2)}^\top, \dots, \boldsymbol{O}_{(T^{(1)},T^{(2)})}^\top\right]^\top, \quad (10)$$

$$\boldsymbol{C} = \left[\boldsymbol{C}_{(1,1)}^\top, \boldsymbol{C}_{(1,2)}^\top, \dots, \boldsymbol{C}_{(T^{(1)},T^{(2)})}^\top\right]^\top. \quad (11)$$

Then, a relationship between $\boldsymbol{O}$ and $\boldsymbol{C}$ can be arranged in a matrix form:

$$\boldsymbol{O} = \boldsymbol{W}\boldsymbol{C}, \quad (12)$$

where $\boldsymbol{W}$ is a $3MT^{(1)}T^{(2)} \times MT^{(1)}T^{(2)}$ window matrix to append dynamic features given by

$$\begin{aligned} \boldsymbol{W} &= [\boldsymbol{w}_{(1,1)}, \dots, \boldsymbol{w}_{(T^{(1)},T^{(2)})}]^\top \otimes \boldsymbol{I}_{M\times M}, \quad (13)\\ \boldsymbol{w_t} &= [\boldsymbol{w_t}^{(S)}, \boldsymbol{w_t}^{(H)}, \boldsymbol{w_t}^{(V)}]^\top. \quad (14) \end{aligned}$$

The output probability $P(\boldsymbol{O} \mid \boldsymbol{S}, \Lambda)$ is given by

$$P(\boldsymbol{O} \mid \boldsymbol{S}, \Lambda) = \mathcal{N}(\boldsymbol{O} \mid \boldsymbol{\mu_S}, \boldsymbol{\Sigma_S}) = \prod_{\boldsymbol{t}} \mathcal{N}(\boldsymbol{O_t} \mid \boldsymbol{\mu}_{\boldsymbol{S_t}}, \boldsymbol{\Sigma}_{\boldsymbol{S_t}}), \quad (15)$$

where $\mathcal{N}(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu_S}$ and $\boldsymbol{\Sigma_S}$ are the "image level" mean vector and covariance matrix given state sequences $\boldsymbol{S}$, respectively, and they are constructed by concatenating the "state level" mean vectors and covariance matrices in accordance with state sequences $\boldsymbol{S}$:

$$\begin{aligned} \boldsymbol{\mu_S} &= \left[\boldsymbol{\mu}_{\boldsymbol{S}_{(1,1)}}^\top, \boldsymbol{\mu}_{\boldsymbol{S}_{(1,2)}}^\top, \dots \boldsymbol{\mu}_{\boldsymbol{S}_{(T^{(1)},T^{(2)})}}^\top\right]^\top, \quad (16)\\ \boldsymbol{\Sigma_S} &= \operatorname{diag}\left[\boldsymbol{\Sigma}_{\boldsymbol{S}_{(1,1)}}, \boldsymbol{\Sigma}_{\boldsymbol{S}_{(1,2)}}, \dots \boldsymbol{\Sigma}_{\boldsymbol{S}_{(T^{(1)},T^{(2)})}}\right] \quad (17) \end{aligned}$$

[1] Using higher-order dynamic features is straightforward. Moreover, other dynamic features in other directions, e.g., diagonal dynamic features can be applied.

However, Eq. (15) becomes an invalid probabilistic distribution over static feature vectors because the integral of Eq. (15) over $\boldsymbol{C}$ is not equal to 1. To yield a valid probability distribution over $\boldsymbol{C}$, Eq. (15) can be re-normalized as the probability distribution of $\boldsymbol{C}$ and can be written as

$$P(\boldsymbol{C} \mid \boldsymbol{S}, \Lambda) = \frac{1}{Z_{\boldsymbol{S}}} \mathcal{N}(\boldsymbol{W}\boldsymbol{C} \mid \boldsymbol{\mu}_{\boldsymbol{S}}, \boldsymbol{\Sigma}_{\boldsymbol{S}}) = \mathcal{N}(\boldsymbol{C} \mid \overline{\boldsymbol{C}}_{\boldsymbol{S}}, \boldsymbol{P}_{\boldsymbol{S}}), \tag{18}$$

where $Z_{\boldsymbol{S}}$ is a normalization term, and $\overline{\boldsymbol{C}}_{\boldsymbol{S}}$ and $\boldsymbol{P}_{\boldsymbol{S}}$ are the mean vector and the covariance matrix, respectively, defined as

$$\boldsymbol{P}_{\boldsymbol{S}} = (\boldsymbol{W}^\top \boldsymbol{\Sigma}_{\boldsymbol{S}}^{-1} \boldsymbol{W})^{-1}, \quad \overline{\boldsymbol{C}}_{\boldsymbol{S}} = \boldsymbol{P}_{\boldsymbol{S}} \boldsymbol{W}^\top \boldsymbol{\Sigma}_{\boldsymbol{S}}^{-1} \boldsymbol{\mu}_{\boldsymbol{S}}. \tag{19}$$

Using the above distribution, the joint distribution of static feature vectors $\boldsymbol{C}$ and hidden variables $\boldsymbol{S}$ can be written as:

$$P(\boldsymbol{C}, \boldsymbol{S} \mid \Lambda) \quad = \quad P(\boldsymbol{C} \mid \boldsymbol{S}, \Lambda) \prod_{m=1,2} P(\boldsymbol{S}^{(m)} \mid \Lambda), \tag{20}$$

where Λ is a set of model parameters of the proposed model. In this paper, the proposed model is referred to as separable lattice trajectory 2-D HMMs (SLT2D-HMMs). Note that covariance matrix $\boldsymbol{P}_{\boldsymbol{S}}$ is generally full even when using the completely same number of model parameters as in SL2D-HMMs. Therefore, the inter-pixel correlation can be modeled by the covariance matrix $\boldsymbol{P}_{\boldsymbol{S}}$. Moreover, SLT2D-HMMs can be viewed as an HGMRF [16] and its graphical representation can be specified by the window matrix $\boldsymbol{W}$.

4. TRAINING ALGORITHM

The parameters of the proposed model can be estimated via the expectation maximization (EM) algorithm [17]. This algorithm maximizes the expectation of the complete data log-likelihood so called $\mathcal{Q}$-function:

$$\mathcal{Q}(\Lambda, \Lambda') \quad = \quad \sum_{\boldsymbol{S}} \gamma_{\boldsymbol{S}} \ln P(\boldsymbol{C}, \boldsymbol{S} \mid \Lambda'), \tag{21}$$

where $\gamma_{\boldsymbol{S}}$ is the posterior probability of $\boldsymbol{S}$ given $\boldsymbol{C}$ and Λ. All unique mean vectors $\boldsymbol{m}$ and inverse covariance matrices $\boldsymbol{\phi}$ in the model set Λ can be optimized by using the following partial derivatives:

$$\frac{\partial \mathcal{Q}(Q, \Lambda)}{\partial \boldsymbol{m}} \quad = \quad \sum_{\boldsymbol{S}} \gamma_{\boldsymbol{S}} \boldsymbol{F}_{\boldsymbol{S}}^\top \boldsymbol{\Sigma}_{\boldsymbol{S}}^{-1} \boldsymbol{W}(\boldsymbol{C} - \overline{\boldsymbol{C}}_{\boldsymbol{S}}), \tag{22}$$

$$\frac{\partial \mathcal{Q}(Q, \Lambda)}{\partial \boldsymbol{\phi}} \quad = \quad \sum_{\boldsymbol{S}} \frac{\gamma_{\boldsymbol{S}}}{2} \boldsymbol{F}_{\boldsymbol{S}}^\top \mathrm{diag}^{-1} \Big[\boldsymbol{W} \boldsymbol{G}_{\boldsymbol{S}} \boldsymbol{W}^\top + 2\boldsymbol{\mu}_{\boldsymbol{S}} (\boldsymbol{C} - \overline{\boldsymbol{C}}_{\boldsymbol{S}})^\top \boldsymbol{W}^\top \Big], \tag{23}$$

where $\boldsymbol{G}_{\boldsymbol{S}} = \boldsymbol{P}_{\boldsymbol{S}} + \overline{\boldsymbol{C}}_{\boldsymbol{S}} \overline{\boldsymbol{C}}_{\boldsymbol{S}}^\top - \boldsymbol{C}\boldsymbol{C}^\top$, and $\boldsymbol{F}_{\boldsymbol{S}}$ is a matrix whose elements are 0 or 1 determined in accordance with the state sequence $\boldsymbol{S}$ so that the following relationships are satisfied:

$$\boldsymbol{\mu}_{\boldsymbol{S}} = \boldsymbol{F}_{\boldsymbol{S}} \boldsymbol{m}, \qquad \boldsymbol{\Sigma}_{\boldsymbol{S}}^{-1} = \mathrm{diag}[\boldsymbol{F}_{\boldsymbol{S}} \boldsymbol{\phi}]. \tag{24}$$

In the present paper, a Viterbi approximation is applied because it is computationally intractable to evaluate the posterior probability over all possible state sequences $\boldsymbol{S}$ in Eqs. (21), (22), and (23). The training procedure of SLT2D-HMMs can be summarized as:

1. Initialize the model parameters and the state sequences of SLT2D-HMMs by using the parameters and Viterbi state sequences of SL2D-HMMs, respectively.
2. Update $\boldsymbol{m}$ and $\boldsymbol{\phi}$ with Eqs. (22) and (23), respectively.
3. Search sub-optimal state sequences by adding a small variation, e.g., ± 1 on the state boundaries.
4. If the Viterbi-approximated $\mathcal{Q}$-function has not converged, return to step 2. otherwise, stop iteration.

5. EXPERIMENTS

5.1. Experimental conditions

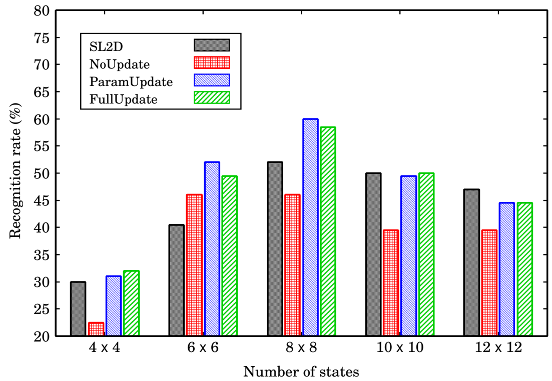
To demonstrate the effectiveness of the proposed model, experiments on modeling faces from the XM2VTS database [14] were conducted. The face images were extracted from the original images (720×576 pixels and transformed into gray-scale) and then sub-sampled to 16×16 and 32×32 pixels. Two datasets were prepared with this process:

- "dataset 1": size-location normalized data (the original size and location in the database are used).
- "dataset 2": 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 sizes $500 \times 500 \sim 600 \times 600$ with a fixed aspect ratio.

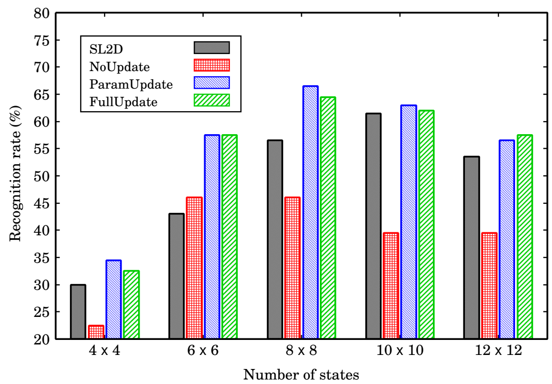
The images were modeled with 4×4, 6×6, 8×8, 10×10, and 12×12 states and single Gaussian distributions. The transition probabilities for each state sequence were assumed to be a left-to-right and top-to-bottom no skip topology. The observation vectors $\boldsymbol{O}$ were constructed by appending the 1st order horizontal and vertical dynamic feature vectors to the static features $\boldsymbol{C}$. The window matrix $\boldsymbol{W}$ was designed to satisfy Eq. (12), where $L_+^{(H)} = L_-^{(H)} = L_+^{(V)} = L_-^{(V)} = 1$, $w^{(H)}(-1) = w^{(V)}(-1) = -0.5$, $w^{(H)}(0) = w^{(V)}(0) = 0.0$, and $w^{(H)}(1) = w^{(V)}(1) = 0.5$. The model parameters of SLT2D-HMMs were estimated in accordance with the training procedure as summarized in Section 4 by using the Rprop [18] method which is an first order gradient-based optimization method.

5.2. Face recognition experiments

Face recognition experiments on the XM2VTS database were conducted. We prepared eight images of 100 subjects; six images were used for training and two images for testing. In this experiment, the size of face images was 16×16. Figure 1 (a) and (b) show the recognition rate of SL2D-HMMs and SLT2D-HMMs. "SL2D" means the recognition



(a) No variations

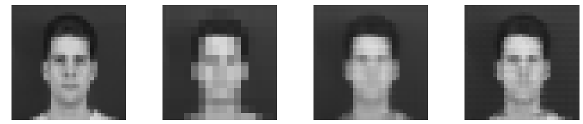


(b) Size and location variations

Fig. 1. Recognition rates

rates of SL2D-HMMs. "NoUpdate" means the results of SLT2D-HMMs with the same model parameters as SL2D-HMMs. "ParamUpdate" and "FullUpdate" mean the results of SLT2D-HMMs with the model parameters of SLT2D-HMMs updated but with state sequences fixed, and with both the model parameters and the state sequences updated, respectively.

First, the recognition rates in figure 1 (b) were higher than those in figure 1 (a). This indicates that both SL2D-HMMs and SLT2D-HMMs could normalize the variations on "dataset 2" successfully. It also can be seen that "NoUpdate" remained lower than did "SL2D", though the same model parameters were used between them. This is obviously because the parameters were not optimized for the likelihood function of the SLT2D-HMMs. After the model parameters were optimized, "ParamUpdate" and "FullUpdate" achieved better results than did "SL2D" and "NoUpdate." However, when comparing "ParamUpdate" and "FullUpdate," significant improvement of the performance could not be obtained. The reason for this results can be explained as follows: In previous work [19], for 1-D HMMs, changing state boundaries by adding small variations is effective for searching sub-optimal state sequences. However, for SLT2D-HMMs, it must be considered that the observations depend on horizontal and vertical state sequences and that the combinations of both state sequences affect the likelihoods. The re-estimation algorithm for state sequences adopted in this paper is strongly approximated in this viewpoint because it finds only one sub-optimum state boundary from all the state boundaries of both directions and the effect of the combinations is ignored. This explains why a significant improvement of the recognition performance was not obtained.



(a) (b) (c) (d)

Fig. 2. Visualization of state alignment. (a) is the test data, (b) is the state alignment of SL2D-HMMs, and (c) and (d) are the mean trajectory of SLT2D-HMMs w/ and w/o parameters updated, respectively.

5.3. Image alignment experiments

To demonstrate the advantageous property of SLT2D-HMMs for image recognition, an image alignment experiment was conducted, where the size of the face images was 32×32. Figure 2 presents the test image and its state alignments of SL2D-HMMs and SLT2D-HMMs. From figure 2 (b), it can be seen that a rectangular state alignment was obtained by using the SL2D-HMMs because of the constraint that the statistics within a state do not change dynamically. In comparison, from figure 2 (c), it can be seen that the mean trajectory $\overline{\boldsymbol{C}}$ of the proposed model seemed smoother than did the state alignment of the SL2D-HMMs. This indicates that the constraint of the SL2D-HMMs of constant statistics was mitigated. However, the detailed parts of the test data (e.g., eyes and nose) became blurred, since the model parameters were not optimized for SLT2D-HMMs. After the model parameters were optimized, from figures 2 (c) and (d), it can be observed that the details of these parts became clearer. This also explains the improvement of the recognition performance.

6. CONCLUSION

In this paper, SLT2D-HMMs were proposed. SLT2D-HMMs can be defined by reformulating SL2D-HMMs and imposing explicit relationships between static and dynamic features. Experiments on image recognition and alignment were conducted on the XM2VTS database. The proposed model achieved better results than did the SL2D-HMMs. Implementing more precise search algorithms such as the delayed decision Viterbi algorithm [13] will be future work. The comparison of recognition performance between the proposed method and other classifiers such as support vector machines [20] will also be future work.

7. ACKNOWLEDGEMENTS

This work was partially supported by the Artificial Intelligence Research Promotion Foundation.

8. REFERENCES

[1] M. Turk and A. Pentland, "Face recognition using eigenfaces," *IEEE Computer Society Conference*, pp. 586–591, 1991.

[2] E. Oja, *Subspace Method for Pattern Recognition*, Research Studies Press, 1983.

[3] E. Levin adn R. Pieraccini, "Dynamic Planar Warping for Optical Character Recognition," 1992, vol. 3, pp. 149–152.

[4] S. Kuo and O.E. Agazzi, "Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 16, pp. 842–848, 1994.

[5] H. Othman and T. Aboulnasr, "A Simplifed Second-Order HMM with Application to Face Recognition," *International Symposium on Circuits and Systems*, vol. 2, pp. 161–164, 2001.

[6] A. Nefian and M.H. Hayes III, "Maximum Likelihood Training of Embedded HMM for Face Detection and Recognition," *IEEE International Conference on Image Processing (ICIP)*, vol. 25, no. 10, pp. 1229–1238, 2003.

[7] J.-T. Chien and C.-P. Liao, "Maximum Confidence Hidden Markov Modeling for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 606–616, 2008.

[8] D. Kurata, Y. Nankaku, K. Tokuda, T. Kitamura, and Z. Ghahramani, "Face recognitinon based sepalable lattice HMMs," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 737–740, 2006.

[9] S. Nakagawa, "A survey on automatic speech recognition," *IEICE Transactions Information and Systems* , vol. E85-D (3), pp. 465–486, 2002.

[10] S. Furui, "Speaker independent isolated word recognition using dynamic features of speech spetrum," *IEEE Transactions Acoustics, Speech, and Signal Processing*, vol. 34, pp. 52–59, 1986.

[11] K. Kumaki, Y. Nankaku, and K. Tokuda, "Face Recognition based on Extended Separable Lattice 2-D HMMs," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 25, pp. 2209–2212, 2012.

[12] K. Kumaki, "Face Recognition based on Extended Separable Lattice HMMs," Bachelor thesis, Nagoya Insititute of Technology, 2010.

[13] H. Zen, K. Tokuda, and T. Kitamura, "Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences," *Computer and Speech and Language*, vol. 21, pp. 153–173, 2007.

[14] K. Messer, J. Mates, J. Kitter, J. Luettin, and G. Maitre, "XM2VTS: The Extended M2VTS Database," *Audio and Video-Based Biometric Person Authentication*, pp. 72–77, 1999.

[15] C.-H. Lee and E. Giachin, "Speaker independent isolated word recognition using dynamic features of speech spetrum," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 161–164, 1991.

[16] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, Chapamn & Hall/CRC, 2005.

[17] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from imcomplete data via the EM algorithm," *Journal of Royal Statistics Sosiety*, vol. 39, pp. 1–38, 1997.

[18] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The Rprop algorithm," 1993, pp. 586–591.

[19] K. Tokuda, T. Kobayashi, and S. Imai, "Speech parameter generation from HMM using dynamic features," *IEICE technical reports*, vol. 103, no. 210, pp. 23–28, 2003.

[20] N. Cristianini and J.S. Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.