IMAGE RECOGNITION BASED ON DISCRIMINATIVE MODELS USING FEATURES GENERATED FROM SEPARABLE LATTICE HMMS

Yoshinari Tsuzuki, Kei Sawada, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda

Department of Computer Science and Engineering, Nagoya Institute of Technology, Nagoya Japan

ABSTRACT

This paper presents an image recognition technique based on discriminative models using features generated from separable lattice hidden Markov models (SL-HMMs). A major problem in image recognition is that the recognition performance is degraded by geometric variations such as that in position and size of the object to be recognized. SL-HMMs have been proposed to solve this problem. SL-HMMs are an extension of HMMs with size and locational invariances based on state transitions. An SL-HMM is a generative model and can represent generation processes of observations well. However, there is a possibility that the recognition performance of generative models is inferior to that of discriminative models because discriminative models are specialized to identification. In this paper, we propose image recognition based on log linear models (LLMs) using features extracted from SL-HMMs. The proposed method can extract features invariant to geometric variations by using SL-HMMs and built an accurate classifier based on discriminative models with the extracted features. Face recognition experiments showed that the proposed method obtained higher recognition rates than SL-HMMs and convolutional neural networks based methods.

Index Terms— Image recognition, hidden Markov model, separable lattice HMM, log linear model, derivative feature

1. INTRODUCTION

In image recognition, statistical models using big data have grown in popularity in the last decade, e.g., eigenfaces [1] and convolutional neural network (CNN) [2, 3]. However, such statistical models encounter a problem in terms of geometric variations, i.e., position, size and rotation of target objects. One of the major solutions to this problem is to use invariant features, e.g., a scale-invariant feature (SIFT)[4] and histograms of oriented gradients (HOG) [5]. Although these methods can avoid the influence of geometric variations by using only accumulated statistics of local features, there is a problem that most of such features ignore global shape information of target objects that seems to be effective in image recognition. Another solution is to pre-normalize geometric variations prior to applying statistical models. In general, the normalization is performed manually or by using an empirically developed normalization technique independently of the training and recognition. However, it takes a large cost and task-dependent normalization techniques need to be developed for each target dataset. Furthermore, the final objective of image recognition is not to accurately normalize images for human perception but to achieve better recognition performance. Therefore, it seems to be a good idea to integrate the normalization process into classifiers and optimize them simultaneously based on the unified criterion.

Hidden Markov models (HMMs) based techniques have been proposed as such kind of approaches for dealing with geometric variations [6, 7]. The geometric normalization is represented by discrete hidden variables, and the normalization process is performed in the calculation of probabilities. Although the extension of HMMs to multi-dimension generally leads to an exponential increase in the computational complexity, some efficient approximations of likelihood calculation and model structures have been proposed [8]–[14]. Separable lattice hidden Markov models (SL-HMMs) are feasible models that can perform an elastic matching in both horizontal and vertical directions and makes it possible to model invariances to the size and location of an object. Furthermore, some extensions to structures representing typical geometric variations have already been proposed, e.g., a structure for rotational variations [15], a structure with multiple horizontal and vertical Markov chains [16], and explicit state duration modeling [17].

Recently, discriminative models have intensively been studied, especially neural networks have shown great success in many applications. While generative models such as SL-HMMs focus on capturing the property of training data by assuming data generation processes, discriminative models focus on directly solving a discrimination problem to improve recognition performance. CNNs have successfully been used in image recognition, because of the robustness against geometric variations based on multiple convolutional and pooling layers. The most important advantage of CNNs is that the network structure has the feature extraction process robust to geometric variations, and that is simultaneously optimized with training of the classifier based on the discriminative criterion. However, CNNs still have a limitation in invariance to geometric transformations, i.e., it is difficult to represent global geometric transformations over an entire image because pooling is independently performed in each local window. Therefore, the structure of generative models assuming explicit image variations should be useful to construct discriminative models with higher invariance to geometric transformations.

In this paper, we propose image recognition based on log linear models (LLMs) [18, 19] using features generated from SL-HMMs. The proposed method can extract features invariant to geometric variations by using SL-HMMs and built an accurate classifier based on discriminative models with the extracted features. Although there are many features that can extract from SL-HMMs, features based on log-likelihoods and derivatives with respect to parameters of SL-HMMs are used as the input features for LLMs in this paper. It is expected that the performance of the proposed method is improved by using defferent types of features and LLM-based systems using the features were evaluated comparing with the baseline SL-HMM-based system and CNN-based systems.

The rest of this paper is organized as follows. In section 2 and 3, SL-HMMs and LLMs are briefly explained. Section 4 describes the features generated from SL-HMMs. Section 5 presents face recognition experiments on the XM2VTS database [20] and we finally conclude the paper in section 6.

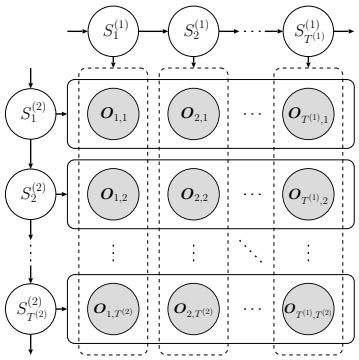

Fig. 1. Graphical model of SL-HMMs

2. SEPARABLE LATTICE HIDDEN MARKOV MODELS

Separable lattice hidden Markov models (SL-HMMs) [14] are defined for modeling multi-dimensional data. Observations are assumed to be given on a two-dimensional lattice as:

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

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

In SL-HMMs, the hidden variables are constrained to be composed of two Markov chains to reduce the number of possible state sequences as:

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

$$\boldsymbol{S}^{(m)} = \{S^{(m)}_{t^{(m)}} \mid 1 \le t^{(m)} \le T^{(m)}\}, \tag{3}$$

where $\boldsymbol{S}^{(m)}$ is the Markov chain along with the m-th coordinate and $S^{(m)}_{t^{(m)}} \in \{1, 2, \ldots, K^{(m)}\}$. The composite structure of hidden variables in SL-HMMs is defined as the product of hidden state sequences:$\boldsymbol{S_t} = (S^{(1)}_{t^{(1)}}, S^{(2)}_{t^{(2)}}) \in \boldsymbol{K}$. 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. Figure 1 shows a graphical model of SL-HMMs. The joint probability of observation vectors $\boldsymbol{O}$ and hidden variables $\boldsymbol{S}$ can be written as:

$$\begin{aligned} P(\boldsymbol{O}, \boldsymbol{S} \mid \boldsymbol{\Lambda}) &= P(\boldsymbol{O}, \boldsymbol{S}^{(1)}, \boldsymbol{S}^{(2)} \mid \boldsymbol{\Lambda}) \\ &= \prod_{m=1}^{2} \left[P(S^{(m)}_1 \mid \boldsymbol{\Lambda}) \prod_{t^{(m)}=2}^{T^{(m)}} P(S^{(m)}_{t^{(m)}} \mid S^{(m)}_{t^{(m)}-1}, \boldsymbol{\Lambda}) \right] \\ &\quad \times \prod_{\boldsymbol{t}} P(\boldsymbol{O_t} \mid \boldsymbol{S_t}, \boldsymbol{\Lambda}), \end{aligned} \tag{4}$$

where $\boldsymbol{\Lambda} = \left\{\boldsymbol{\pi}^{(m)}, \boldsymbol{a}^{(m)}, \boldsymbol{B_k}\right\}$ is a set of model parameters, $\boldsymbol{k}$ is the two-dimensional state index in the two-dimensional state lattice $\boldsymbol{K}$, $\boldsymbol{\pi}^{(m)}$ is the initial state probability, $\boldsymbol{a}^{(m)}$ is the state transition probability, $\boldsymbol{B_k} = \{\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}\}$ are model parameters of state output probability, and $\boldsymbol{\mu_k}$ and $\boldsymbol{\Sigma_k}$ are the mean vector and the convariance matrix of the Gaussian distribution on a two-dimensional state lattice.

3. LOG LINEAR MODELS

An SL-HMM is a generative model that represents the process of generating the observed data. On the other hand, discriminative models estimate the posterior probability for each target class directly. Discriminative models are specialized in classification problems. Therefore, discriminative models usually show better classification performance than generative models. Log linear models (LLMs) have been proposed as discriminative models [18, 19]. The posterior probability distribution is represented as:

$$P(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\lambda}) = \prod_{l=1}^{L} \frac{1}{W(\boldsymbol{X}_l)} \exp\left\{\boldsymbol{\lambda}^{(y)} \boldsymbol{X}_l\right\}, \tag{5}$$

$$W(\boldsymbol{X}_l) = \sum_{y'=1}^{C} \exp\left\{\boldsymbol{\lambda}^{(y')} \boldsymbol{X}_l\right\}, \tag{6}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_L)$ is a target class sequence, $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_L)$ is an input feature vector sequence, L is the number of training vectors, C is the number of classes, and $\boldsymbol{\lambda}^{(y)}$ is a net of model parameters for class y. In LLMs, dependence of the target class variables on input feature $\boldsymbol{X}$ is directly modeled by using model parameters $\boldsymbol{\lambda}$. In addition, LLMs can deal with various features and calculate the posterior probability from them. Since the input features affect the recognition performance, they are important in LLMs-based image recognition.

4. IMAGE RECOGNITION BASED ON LLMS WITH FEATURES GENERATED FROM SL-HMMS

LLMs can select features that are effective for recognition. Therefore, it is important to prepare features that may be effective for classification, i.e., features that may be highly dependent on target classes, in order to achieve high recognition performance. Previous studies have described techniques to enumerate candidate features by using the human knowledge and experience [21]. However, manually enumerating effective features incurs high costs. Furthermore, since a fixed-length vector is used as an input feature for LLMs, geometric variations such as that in position and size of an object affects the recognition performance in image recognition using LLMs.

Without limiting the recognition target, the feature generation using generative models has been proposed as for automatically generating features on the basis of training data [22]. Because generative models estimate the generation process of observation data, the features based on generative models seem effective in recognition. Additionally, generative models can use prior knowledge such as prior distribution. By generating features using the SL-HMMs incorporating the normalization process, the features may have fixed-lengths in consideration of the geometric variations. The purpose of this study is to improve in recognition performance by using features generated from SL-HMMs.

4.1. Features based on SL-HMMs

In this paper, log-likelihood and derivative features based on SL-HMMs are used as inputs to LLMs.

4.1.1. Log-likelihood features

One feature based on generative models is a log-likelihood feature, which expresses plausibility of the model for the observed data. Log-likelihood features for the image data are defined as:

$$\boldsymbol{X}^{(l)} = \begin{bmatrix} \ln P(\boldsymbol{O}^{(l)} \mid \boldsymbol{\Lambda}^{(1)}) \\ \vdots \\ \ln P(\boldsymbol{O}^{(l)} \mid \boldsymbol{\Lambda}^{(C)}) \end{bmatrix}, \tag{7}$$

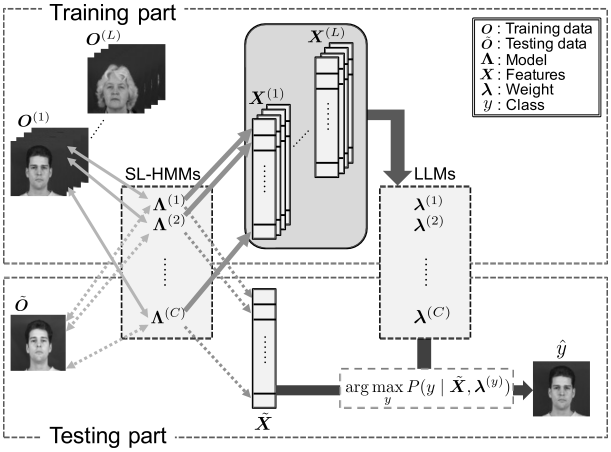

Fig. 2. Overview of proposed method

where $\boldsymbol{O}^{(l)}$ is the l-th image data. Log-likelihood is used for recognizing the image on the basis of SL-HMMs. Therefore, recognition of LLMs by using log-likelihood comprises recognition of the framework of SL-HMMs.

4.1.2. Derivative features

Some features based on generative models are derivative features, which are defined by derivatives of the log-likelihood function with respect to the model parameters [22]. The derivative features $\boldsymbol{X}_D^{(l)}$ are thus represented as:

$$\boldsymbol{X}_D^{(l)} = \left[\boldsymbol{X}_{D,1,1}^{(l)\top} \boldsymbol{X}_{D,1,2}^{(l)\top} \cdots \boldsymbol{X}_{D,K^{(1)},K^{(2)}}^{(l)\top} \right]^\top, \quad (8)$$

$$\boldsymbol{X}_{D,\boldsymbol{k}}^{(l)} = \begin{bmatrix} \dfrac{\partial}{\partial \boldsymbol{B}_{\boldsymbol{k}}^{(1)}} \ln P(\boldsymbol{O}^{(l)} \mid \boldsymbol{\Lambda}^{(1)}) \\ \vdots \\ \dfrac{\partial}{\partial \boldsymbol{B}_{\boldsymbol{k}}^{(C)}} \ln P(\boldsymbol{O}^{(l)} \mid \boldsymbol{\Lambda}^{(C)}) \end{bmatrix}. \quad (9)$$

Derivative features of each model parameter are represented as:

$$\frac{\partial}{\partial \boldsymbol{\mu}_{\boldsymbol{k}}^{(y)}} \ln P(\boldsymbol{O}^{(l)} \mid \boldsymbol{\Lambda}^{(y)}) = \sum_{\boldsymbol{t}} \langle S_{\boldsymbol{k},\boldsymbol{t}} \rangle_{Q(\boldsymbol{S})} {\boldsymbol{\Sigma}_{\boldsymbol{k}}^{(y)}}^{-1} (\boldsymbol{O}_{\boldsymbol{t}}^{(l)} - \boldsymbol{\mu}_{\boldsymbol{k}}^{(y)}), \quad (10)$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_{\boldsymbol{k}}^{(y)}} \ln P(\boldsymbol{O}^{(l)} \mid \boldsymbol{\Lambda}^{(y)}) = \sum_{\boldsymbol{t}} \langle S_{\boldsymbol{k},\boldsymbol{t}} \rangle_{Q(\boldsymbol{S})} \frac{1}{2} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{(y)} - \sum_{\boldsymbol{t}} \langle S_{\boldsymbol{k},\boldsymbol{t}} \rangle_{Q(\boldsymbol{S})} \frac{1}{2} (\boldsymbol{O}_{\boldsymbol{t}}^{(l)} - \boldsymbol{\mu}_{\boldsymbol{k}}^{(y)})(\boldsymbol{O}_{\boldsymbol{t}}^{(l)} - \boldsymbol{\mu}_{\boldsymbol{k}}^{(y)})^\top, \quad (11)$$

where $\langle S_{\boldsymbol{k},\boldsymbol{t}} \rangle_{Q(\boldsymbol{S})}$ is posterior distribution of state $\boldsymbol{k}$ at coordinate $\boldsymbol{t}$ and $Q(\boldsymbol{S})$ is approximate posterior probability of $P(\boldsymbol{S} \mid \boldsymbol{O}, \boldsymbol{\Lambda})$. The derivative features of state $\boldsymbol{k}$ are derived using the statistics related to the model parameters of state $\boldsymbol{k}$.

4.2. Image recognition using features generated from SL-HMMs

Figure 2 shows an overview of proposed method. First, an SL-HMM of each class is trained from training data in the training part of the proposed method. Second, features, such as log-likelihood and derivative features described in Section 4.1, are generated by using the trained SL-HMMs $\boldsymbol{\Lambda}^{(y)}$ and training data $\boldsymbol{O}^{(l)}$. Then, an LLM is trained from the generated features. In the testing part of the proposed method, features corresponding to the testing data $\tilde{\boldsymbol{O}}$ are generated by the same procedures as the feature generation in the training part. Recognition is performed by calculating the posterior probabilities of all classes from the trained LLMs and the generated features. Features generated from SL-HMMs can consider geometric variations such as that in the position and size of an object. Furthermore, LLMs can directly model posterior probabilities. Therefore, the proposed method can perform accurate recognition.

5. EXPERIMENTS

5.1. Experimental conditions

To verify the effectiveness of the proposed method, face recognition experiments on the XM2VTS database [20] were conducted. Eight images of 100 subjects were prepared for experiments; six or four images were used for training and two images were used for testing. Face images of 64×64 grayscale pixels were extracted from the original images. The example images are shown in Figs. 3 and 4. We prepared two datasets for experiments. Dataset 1 did not include many size and location variations, while dataset 2 did. SL-HMMs with 24×24, 32×32, 40×40, 48×48, and 56×56 states were used. SL-HMMs were estimated by the maximum likelihood (ML) estimation and maximum a posteriori (MAP) estimation [23]. As the training algorithm, EM algorithm [24] and deterministic annealing EM (DAEM) algorithm [25, 26] were applied. The hyper-parameters of the prior distribution were determined by using statistics on a universal background model (UBM) [26, 27], which was trained using all training data. LLMs used features obtained from each SL-HMM. As features for LLMs, log-likelihood features (**L**), derivative features with respect to the means of the SL-HMMs (**M**), and derivative features with respect to the variance of the SL-HMMs (**V**) were used. Therefore, the results for **LLM-{L, LMV}** were compared.

Additionally, two convolutional neural network (CNN)-based approaches [2, 3] (**CNN** and **CaffeNet**) were compared with the proposed method. In **CNN**, CNNs were trained by using the Caffe [28] based on datasets 1 and 2. In **CaffeNet**, a pre-trained CNN (CaffeNet) [3, 28], which was trained by using the dataset from the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [29], was used to extract image features. The details of the CNN approaches are as follows:

- **CNN**: The architecture of the CNNs was $I(64,1)-C(128,10,1,55)-P(3,2,27)-C(256,5,1,23)-P(3,2,11)-F(800)-F(600)-F(400)-O(100)$, where $I(i,d)$ indicates an input layer with d dimensional $i \times i$ sized image, $C(f,w,s,o)$ indicates a convolutional layer with f filters of a $w \times w$ sized window with a stride of s and $o \times o$ sized output, $P(w,s,o)$ indicates a pooling layer, $F(n)$ indicates a fully-connected layer with n units, and $O(c)$ indicates an output layer with c classes. The ReLU function and dropout with probability of 0.5 were used in the convolutional and fully-connected layers.
- **CaffeNet**: The image-feature vectors were composed of 4096 dimensions extracting the pre-trained CaffeNet of the 7^{th} fully-connected layer. The one-nearest neighbor was then used as the classifier.

Fig. 3. Examples of images in dataset 1

Fig. 4. Examples of images in dataset 2

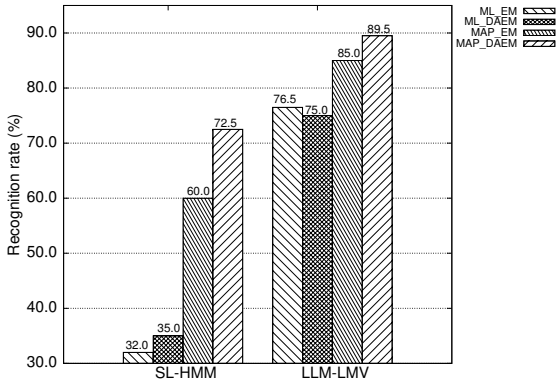

Fig. 5. Accuracy of estimation methods in SL-HMMs

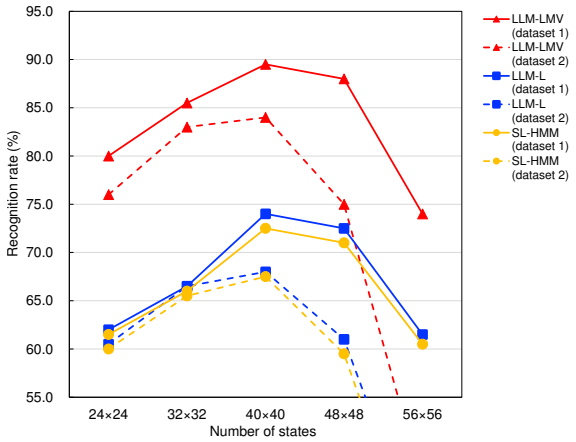

Fig. 6. Relationship between features and recognition rate

Table 1. Comparison with CNNs

	Dataset 1	Dataset 2
SL-HMM	85.0	81.0
LLM-LMV	98.0	91.5
CNN	82.5	63.0
CaffeNet	85.5	73.0

5.2. Accuracy of estimation methods of generative models

Impacts of the estimation accuracy of the SL-HMMs in the proposed method were verified in this section. Figure 5 shows recognition rates of four SL-HMM-based systems and four LLM-based systems on dataset 1. The training data was four images per subject and 40×40-state SL-HMMs were used. It can be seen from Figure 5 that the recognition rate of **LLM-LMV** was improved as improving the recognition rate of **SL-HMM** that was used for feature generation in **LLM-LMV**. These results indicate that the estimation accuracy of the generative models used for feature generation has a strong impact on the performance of the proposed method.

5.3. Comparison with features

The effectiveness of the features generated from SL-HMMs in the proposed method was evaluated by comparing the several feature sets. SL-HMMs were used for feature generation by the MAP method and with the DAEM algorithm. Four images per subject were used as the training data. Figure 6 shows the recognition rate of **SL-HMM**, **LLM-L**, and **LLM-LMV** on two datasets. When comparing **SL-HMM** and **LLM-L**, the experiment results show that **LLM-L** achieved higher recognition rate than **SL-HMM**. This is because the LLM can use the likelihood of the all models and take into account the relation of them for calculating the posterior probabilities. Moreover, **LLM-LMV** significantly improved the recognition rate from **LLM-L** on both datasets. These results clearly show that derivative features of SL-HMMs are effective for image recognition even when data includes large geometric variations.

5.4. Comparison with CNNs

In this section, the proposed method was evaluated by comparing with two CNN-based systems. The training data consisted of six images per subject and in total 600 images for 100 subjects. The model structure of SL-HMMs was 40×40 states. Table 1 shows the experimental results of **SL-HMM**, **LLM-LMV**, **CNN**, and **CaffeNet**.

Although CNNs can obtain geometric invariants by repeating convolutional and pooling layers, **CNN** and **CaffeNet** showed the large degradation of the recognition performance when dataset 2, which consists of images including large geometric variations, was used. On the other hand, SL-HMMs can take accout of geometric variations by state transitions, and the degradation of the recognition rate of **SL-HMM** was smaller than ones of **CNN** and **CaffeNet**. By using SL-HMMs for feature generation, **LLM-LMV** obtained the highest recognition rate in both datasets. These results indicate that the proposed method is more robust to geometric variations than **CNN** and **CaffeNet**. However, the number of training images in the experiments seems to be small to train CNNs. Therefore, the proposed method should be compared with CNN-based systems on large database for detailed evaluation.

6. CONCLUSION

This paper proposed image recognition based on log linear models using features generated from separable lattice hidden Markov models (SL-HMMs). The proposed method can obtain features taking account of geometric variations by using SL-HMMs as feature generator. The results obtained in this paper suggest that features generated from SL-HMMs are effective for classification and robust to geometric variations. Moreover, it is clearly shown that the recognition performance is significantly improved by using derivative features. In future work, we will extend the proposed method to the classification based on neural networks, and future work also includes the detailed comparison of CNNs on large datasets.

7. ACKNOWLEDGEMENTS

This work was supported by Grant-in-aid for JSPS Fellows Grant Number 15J08391 and the Hori Sciences and Arts Foundation.

8. REFERENCES

[1] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," Conference on Computer Vision and Pattern Recognition, pp. 586–591, 1991.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffiner,"Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] A. Krizhevsky, I. Sutskever, and G.E. Hinton,"ImageNet classification with deep convolutional neural networks," Conference on Neural Information Processing Systems, pp. 1097–1105, 2012.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[5] N. Dalal, "Histograms of oriented gradients for human detection," Conference on Computer Vision and Pattern Recognition, pp. 886–893, 2005.

[6] F. S. Samaria, "Face recognition using hidden Markov models," Ph. D. dissertation, University of Cambridge, 1994.

[7] A. V. Nefian and M. H. Hayes, "A Hidden Markov Model for face recognition," International Conference on Acoustics, Speech and Signal Processing, vol. 5, pp. 2721–2724, 1998.

[8] S. S. Kuo and O. E. Agazzi, "Keyword spotting in poorly printed documentions using pseudo 2-D hidden Markov models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 8, pp. 842–848, 1994.

[9] A. V. Nefian and M. H. Hayes III, "Maximum likelihood training of the embedded HMM for face detection and recognition," International Conference on Image Processing, vol. 1, pp. 33–36, 2000.

[10] X. Ma, D. Schonfeld, and A. Khokhar, "Image segmentation and classification based on a 2D distributed hidden Markov model," Society of Photo-optical Instrumentation Engineers, vol. 6822, 2008.

[11] J. Li, A. Najmi, and R. M. Gra, "Image classification by a two dimensional hidden Markov model," IEEE Transactions on Signal Processing, vol. 48, no. 2, pp. 517–533, 2000.

[12] H. Othman and T. Aboiilnasr, "A simplified second-order HMM with application to face recognition," International Symposium on Circuits and Systems, vol. 2, pp. 161–164, 2001.

[13] 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, no. 4, pp. 606–616, 2008.

[14] D. Kurata, Y. Nankaku, K. Tokuda, T. Kitamura, and Z. Gharamani, "Face Recognition based on Separable Lattice HMMs," International Conference on Acoustics, Speech and Signal Processing, vol. 5, pp. 737–740, 2006.

[15] A. Tamamori, Y. Nankaku, and K. Tokuda. "An extension of separable lattice 2-D HMMs for rotational data variations," IEICE transactions on information and systems, vol. 95, no. 8, pp. 2074–2083, 2012.

[16] K. Kumaki, Y. Nankaku, and K. Tokuda, "Face recognition based on extended separable lattice 2-D HMMs," International Conference on Acoustics, Speech and Signal Processing, pp. 2209–2212, 2012.

[17] Y. Takahashi, A. Tamamori, Y. Nankaku, and K. Tokuda, "Face recognition based on separable lattice 2-D HMM with state duration modeling," International Conference on Acoustics, Speech and Signal Processing, pp. 2162–2165, 2010.

[18] S. Wiesler, M. Nubaum-Thom, G. Heigold, R. Schluter, and H. Ney, "Investigations on features for log-linear acoustic models in continuous speech recognition," Automatic Speech Recognition and Understanding, pp. 52–57, 2009.

[19] S. Wiesler, A. Richard, Y. Kubo, R. Schluter, and H. Ney, "Feature Selection for Log-Linear Acoustic Models," International Conference on Acoustics, Speech and Signal Processing, pp. 5324–5327, 2011.

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

[21] Lee, S. Y., Young Kug Ham, and R-H. Park, "Recognition of human front faces using knowledge-based feature extraction and neurofuzzy algorithm," Pattern recognition 29.11, pp. 1863–1876, 1996.

[22] C. Longworth and M. J. F. Gales, "Combining Derivative and Parametric Kernels for Speaker Verification," IEEE Transactions on Audio, Speech and Language Processing, pp. 748–757, 2009.

[23] J. L. Gauvain and C. H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains,"IEEE Transactions on Speech and Audio Processing, 2:291–298, 1994.

[24] Z. Ghahramani and M. I. Jordan,"Factorial Hidden Markov Models," Machine Learning, vol. 29, pp. 245–273, 1997.

[25] N. Ueda and R. Nakano,"Deterministic Annealing Variant of the EM Algorithm," Neural Networks, vol. 11, no. 2, pp. 271–282, 1998.

[26] K. Sawada, A. Tamamori, K. Hashimoto, Y. Nankaku, and K. Tokuda, "Bayesian approach to image recognition based on separable lattice hidden Markov models," IEICE Transactions on Information and Systems, vol.E99-D, no.12, pp3119–3131, 2016.

[27] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," Eurospeech, pp. 963–966, 1997.

[28] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R.B. Girshick, S. Guadarrama, and T. Darrell,"Caffe: Convolutional architecture for fast feature embedding," ACM international conference on Multimedia, pp. 675–678, 2014.

[29] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei,"ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.