AN EXTENSION OF SEPARABLE LATTICE 2-D HMMS FOR ROTATIONAL DATA VARIATIONS

Akira Tamamori, Yoshihiko Nankaku, Keiichi Tokuda

Nagoya Institute of Technology, Nagoya, Japan

ABSTRACT

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 normalization is required. SL2D-HMMs can perform an elastic matching in both horizontal and vertical directions; this makes it possible to model invariances to size and location. To deal with rotational variations, we introduce additional HMM states which represent the shifts of the state alignments of the observation lines in a particular direction. Face recognition experiments show that the proposed method improves the performance significantly for rotational variation data.

Index Terms— Face recognition, Hidden Markov model, Separable lattice 2-D HMM, Deterministic Annealing EM Algorithm

1. INTRODUCTION

For many years, many researchers of pattern recognition have developed the field of image recognition as the main theme of the 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. 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 such kind of 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 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 2-D HMMS (SL2D-HMMs) have been proposed [1]. 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 invariances 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 our 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 [2] to the our proposed model. Variational method approximate the posterior distribution over the hidden variables by a tractable distribution. Furthermore, to improve the training algorithm, we applied the deterministic annealing EM (DAEM) algorithm [3] to the variational EM algorithm for the proposed model.

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. In Section 4, we derive the training algorithm for the proposed model. In Section 5, we describe face recognition experiments on the XM2VTS database and finally conclude in Section 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:

\[ O = \{ O_t | t = (t^{(1)}, t^{(2)}) \in T \} \quad (1) \]

where \( t \) denotes the coordinates of the lattice in two-dimensional space \( T \) and \( t^{(m)} = 1, \ldots, T^{(m)} \) is the coordinate of the \( m \)-th dimension. The observation \( O_t \) 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 \( K = \{1, \ldots, K\} \). 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 \( O_t \) is dependent only on the state \( S_t \) as in ordinary HMMs, dependencies 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:

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

\[
S^{(m)} = \{S_1^{(m)}, \ldots, S_T^{(m)}\} \quad (3)
\]

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

\[
S_t = (S_t^{(1)}, S_t^{(2)}) \quad (4)
\]

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 \( \prod_m K^{(m)} \) to \( \prod_m \{K^{(m)}\}^{T^{(m)}} \).

The joint probability of observation vectors \( O \) and hidden variables \( S \) can be written as

\[
P(O, S | \Lambda) = P(O | S, \Lambda) \cdot P(S | \Lambda) \cdot P(d | \Lambda)
\]

\[
= \prod_t P(O_t | S_t, \Lambda) \times \prod_{m=1,2} P(S_t^{(m)} | \Lambda) \prod_{t=m}^{T} P(S_t^{(m)} | S_{t-1}^{(m)} | \Lambda) \quad (5)
\]

where \( \Lambda \) is a set of model parameters.

In image modeling, SL2D-HMMs can perform 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, SL2D-HMMs assume that a target object does not rotate in an image, because SL2D-HMMs have only one state sequence in each direction. Therefore, when images have such rotational variations, the mapping between pixels and the states fails, and this leads to degradation of the recognition performance.

3. MODEL STRUCTURE FOR 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. Figure 1 shows the graphical representation for the proposed model and the likelihood function of the proposed model is defined as follows:

\[
P(O, S, d | \Lambda) = P(O | S, d, \Lambda) \cdot P(S | \Lambda) \cdot P(d | \Lambda)
\]

\[
= \prod_t P(O_t | S_t, \Lambda) \cdot \prod_{m=1,2} P(S_t^{(m)} | \Lambda) \cdot \prod_{m=1}^{T} P(d^{(m)} | \Lambda) \quad (6)
\]

where \( d \) represents to shift state sequences and consists of two Markov chains for each dimension:

\[
d = \{d^{(1)}, d^{(2)}\} \quad (7)
\]

\[
d^{(m)} = \{d_1^{(m)}, d_2^{(m)}, \ldots, d_{T^{(m)}}^{(m)}\} \quad (8)
\]

\[
d_1^{(m)} \in \{D_{\min}^{(m)}, D_{\min}^{(m)} + 1, \ldots, D_{\max}^{(m)}\}, \quad n \neq m \quad (9)
\]

where \( D_{\min}^{(m)} \) and \( D_{\max}^{(m)} \) represent the minimum and maximum shift of the \( m \)-th coordinate respectively, and \( \overline{S_t} \) is the shifted state defined as

\[
\overline{S_t} = (\overline{S_{t^{(1)}}^{(1)}}, \overline{S_{t^{(2)}}^{(2)}}) = \left( S_{t^{(1)}}^{(1)+d_1^{(1)}}, S_{t^{(2)}}^{(2)+d_2^{(2)}} \right) \quad (10)
\]

where the following boundary conditions are assumed:

\[
\overline{S_t^{(m)}} = \begin{cases} 
1 & \left( t^{(m)} \leq T^{(m)} \right) \\
K^{(m)} & \left( t^{(m)} > T^{(m)} \right)
\end{cases} \quad (11)
\]

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 expectations of the complete data log-
likelihood so called Q-function:

$$Q(\Lambda, \Lambda') = \sum_S \sum_d P(S, d | O, \Lambda) \ln P(O, S, d | \Lambda)$$

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 [2] 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 $Q(S, d)$ over the hidden variables defines a lower bound on the log-likelihood

$$\ln P(O | \Lambda) = \ln \sum_S \sum_d Q(S, d) \frac{P(O, S, d | \Lambda)}{Q(S, d)}$$

$$\geq \sum_S \sum_d Q(S, d) \ln \frac{P(O, S, d | \Lambda)}{Q(S, d)}$$

$$= \mathcal{F}(Q, \Lambda)$$

where Jensen’s inequality is 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 $P(S, d | O, \Lambda)$. 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:

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

(M-step) : $\Lambda^{(k+1)} = \arg \max_{\Lambda} \mathcal{F}(Q^{(k+1)}, \Lambda)$

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 $Q(S, d)$ instead of the true posterior distribution $P(S, d | O, \Lambda)$. 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

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

where $\sum_{S^{(m)}} Q(S^{(m)}) = 1$ and $\sum_{d^{(m)}} Q(d^{(m)}) = 1$.

4.2. Variational DAEM algorithm

The EM algorithm has the problem that the solution converges to a local optimum and the convergence point depends on the initial model parameters. In the variational EM algorithm for the proposed model, the decoupled posterior distributions are updated individually based not only on the initial model parameters but also on the other distributions, both of which are unreliable at an early stage of training. To avoid this problem, we apply the DAEM algorithm to the algorithm derived in the previous section and show that the expectations with respect to the decoupled posterior distributions for the DAEM can also be calculated by the forward-backward procedure.

In the DAEM algorithm, the problem of maximizing the log-likelihood is reformulated as minimizing the thermodynamic free energy defined as

$$\mathcal{F}_\beta = -\frac{1}{\beta} \ln \sum_S \sum_d P(O, S, d | \Lambda)$$

where $1/\beta$ called the “temperature” and maximizing $\mathcal{F}_\beta(\Lambda)$ with a fixed temperature can be interpreted as the approach to thermodynamic equilibrium. In the algorithm, the temperature is gradually decreased and the function is deterministically optimized at each temperature. The procedure of the DAEM algorithm can be summarized as follows:

1. Give an initial model and set $\beta = \beta_{min}$.

2. Iterate EM-steps with $\beta$ fixed until $\mathcal{F}_\beta$ converged:

   (E-step) : $\Lambda^{(k+1)} = \arg \max_{\Lambda} \mathcal{F}_\beta(Q, \Lambda^{(k)})$

   (M-step) : $Q^{(k+1)} = \arg \max_{Q} \mathcal{F}_\beta(Q^{(k+1)}, \Lambda)$

3. Increase $\beta$.

4. If $\beta > 1$, stop the procedure. Otherwise go to step 2.

where $1/\beta_{min}$ is an initial temperature and should be chosen as a high enough value that the EM-steps can achieve a single global maximum of $\mathcal{F}_\beta$. At the initial temperature, the entropy of $Q_{\beta}(S, d)$ is intended to be maximized rather than the Q-function (the first term of equation (15)); therefore $Q_{\beta}(S, d)$ takes a form nearly uniform distribution. While the temperature is decreasing, the form of $Q_{\beta}(S, d)$ changes from uniform to the original posterior and at the final temperature $1/\beta = 1$, the negative free energy $\mathcal{F}_\beta$ becomes equal to the lower bound $\mathcal{F}$, accordingly the DAEM algorithm agrees with the original EM algorithm.

5. EXPERIMENTS

In order to demonstrate the modeling ability of the proposed model, face recognition experiments on the XM2VTS database [4] were conducted. We prepared eight images of 100 subjects; seven images are used for training and one image for testing. Face images of grayscale $64 \times 64$ pixels were extracted from the original images. For rotational variations, all images in the testing data are transformed with rotational angles randomly generated within $-10 \sim 10$ degree from Gaussian distribution with 0.0 mean and 5.0 standard deviation. The number of states of SL2D-HMMs is $24 \times 24$ and the number of shift states is varied among $24 \times 2k (3 \leq k \leq 8)$.

Figure 2 shows recognition rates of the proposed model trained by the variational EM algorithm. In the figure, “SL2D” indicates the conventional SL2D-HMMs, and “Normalized data” represents SL2D-HMMs using the normalized data (without rotational transformation) for both training and testing. Therefore, the result of “Normalized data” can be regarded as the upper limit of the proposed model. From the figure, it can be seen that the proposed model improved the recognition performance significantly compared with “SL2D.” The highest recognition rate of 80% was obtained at $12 \times 12$. 

2208
shift states, and this result is better than “Normalized data.” This means that the proposed model can normalize rotational variations accurately. Figure 3 shows the recognition rates of the variational DAEM algorithm. The temperature parameter $\beta$ were updated according to $(i/I)^{\alpha}$, $I = 40$, $\alpha = 8$ where $i$ is the update step. This update scheduling is selected from other different scheduleings so that the selected one leads to the best recognition performance. Comparing figure 2 and 3, the recognition performance was improved by applying the variational DAEM algorithm. This result indicates that the local optimum problem was improved.

Figure 4 and 5 show the mean vectors of the proposed model and visualized state alignments obtained by the Viterbi algorithm, respectively. The alignments are represented by the mean values of the states corresponding to the observations of the testing data. Therefore, when the visualized alignment is similar to the testing data, it means that the model appropriately normalized the variations of testing data. From the results, we can observe that “SL2D” could not deal with the rotational variations due to the constraint of the model structure. On the other hand, the proposed models obtained appropriate state alignments and this would lead to the improvement of the recognition performance.

6. CONCLUSION

This paper proposed an extension of the separable lattice HMMs for dealing with rotational variations. In face recognition experiments on the XM2VTS database, the proposed model achieved better results than the conventional models. Furthermore, the DAEM algorithm improved the performance of the proposed models. From the results of the state alignments, it can be seen that the proposed model can normalize rotational variations appropriately. Experiments using training data with rotational variations and extensions to more flexible models will be future work.

7. REFERENCES