IMAGE RECOGNITION BASED ON HIDDEN MARKOV EIGEN-IMAGE MODELS USING VARIATIONAL BAYESIAN METHOD

Kei Sawada, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda
Nagoya Institute of Technology

APSIPA ASC 10/30/2013
Background & introduction

Models
- Probabilistic eigen-image models (PEMs)
- Separable lattice hidden Markov models (SL-HMMs)
- Hidden Markov eigen-image models (HMEEMs)

Training criterion
- Maximum likelihood (ML) criterion
- Bayesian criterion
  - HEMEs using variational Bayesian (VB) method (proposed)

Experiments
- Face recognition experiments
- Conclusion & future work
BACKGROUND

- Image recognition
  - Assignment of a label to a given input image
    - Biometrics, OCR, video recognition, etc.
  - Increase in demand in various fields
    - Security, industrial inspection, entertainment, etc.

- Approach to geometric variation in image recognition
  - Heuristic normalization techniques
    - Development of task-dependent techniques require high cost
  - Local features (HOG, SIFT, etc.)
    - Global information can’t use
  - Classifiers
    - Subspace method : corresponding to only pattern variation
    - Characteristic of geometric variation is not considered

Focus on techniques for modeling geometric variations explicitly
INTRODUCTION

- Hidden Markov eigen-image models (HMEMs) [Nankaku, et al.; ’06]
  - Probabilistic PCA and factor analysis
    - Linear feature extraction
  - Separable lattice HMMs [Kurata, et al.; ’06]
    - Invariance size and location
  - Over-fitting problem because of complex model structures

- Training criterion of probabilistic models
  - Maximum likelihood (ML) criterion
    - ML criterion produces point estimation of model parameters
      \[\Rightarrow\] Over-fitting problem when amount of data is insufficient
  - Bayesian criterion
    - Estimation of posterior distributions using prior information
      \[\Rightarrow\] High generalization ability

Apply Bayesian criterion to HMEMs
OUTLINE

- Background & introduction
- Models
  - Probabilistic eigen-image models (PEMs)
  - Separable lattice hidden Markov models (SL-HMMs)
  - Hidden Markov eigen-image models (HMEMs)
- Training criterion
  - Maximum likelihood (ML) criterion
  - Bayesian criterion
    - HEMEs using variational Bayesian (VB) method (proposed)
- Experiments
  - Face recognition experiments
  - Conclusion & future work
PROBABILISTIC EIGEN-IMAGE MODELS (PEMs)

- Eigen-images are represented by probabilistic models
  - Probabilistic principal component analysis (PPCA)
  - Factor analysis (FA)

$$ O = Ww_g + uv $$

- $O$ : Observations
- $W$ : Factor loading matrix
- $w_g$ : Eigen-image (base)
- $x$ : Factor vector
- $v$ : Noise vector

 [~, ~, ~, ~]  

😊 Linear feature extraction based on statistical analysis
😊 Image normalization is required as pre-processing
**PROBABILISTIC EIGEN-IMAGE MODELS (PEMs)**

- Eigen-images are represented by probabilistic models
  - Probabilistic principal component analysis (PPCA)
  - Factor analysis (FA)

<table>
<thead>
<tr>
<th>$O$</th>
<th>$W$</th>
<th>$x$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1 \ldots w_G$</td>
<td>$x_1 \ldots x_G$</td>
<td>$\mathbf{v}$</td>
<td></td>
</tr>
</tbody>
</table>

$O$: observations  
$W$: factor loading matrix  
$x$: factor vector  
$\mathbf{v}$: noise vector  

**Noise vector variance**
- Diagonal matrix $\Rightarrow$ FA
- Isotropic matrix $\Rightarrow$ PPCA

😊 Linear feature extraction based on statistical analysis
😊 Image normalization is required as pre-processing
SEPARABLE LATTICE HMMS (SL-HMMS)

- SL-HMMS have horizontal and vertical Markov chains
  - State sequences of horizontal and vertical are independent

Images are divided into rectangular regions in the state

Each pixel is emitted from a corresponding output PDF

 довольны

Include size-and-location-normalization

sembles independence assumption of observations
SL-HMMs have horizontal and vertical Markov chains
- State sequences of horizontal and vertical are independent

Each pixel is emitted from a corresponding output PDF

Images are divided into rectangular regions in the state

Include size-and-location-normalization

Independence assumption of observations
SEPARABLE LATTICE HMMs (SL-HMMs)

- SL-HMMs have horizontal and vertical Markov chains
- State sequences of horizontal and vertical are independent

Each pixel is emitted from a corresponding output PDF

Include size-and-location-normalization
Independence assumption of observations
HIDDEN MARKOV EIGEN-IMAGE MODELS (HMEMs)

- Integration of PEMs and SL-HMMs
- Eigen-images and noise are generated from SL-HMMs

\[ O = \begin{bmatrix} w_1 & \cdots & w_G \end{bmatrix} x + \nu \]

- \( O \) : Observations
- \( W \) : Factor loading matrix
- \( w_g \) : Eigen-image (base)
- \( x \) : Factor vector
- \( \nu \) : Noise vector

Linear feature extraction and include size-and-location-normalization
OUTLINE

- Background & introduction
- Models
  - Probabilistic eigen-image models (PEMs)
  - Separable lattice hidden Markov models (SL-HMMs)
  - Hidden Markov eigen-image models (HMEEMs)
- Training criterion
  - Maximum likelihood (ML) criterion
  - Bayesian criterion
    - HEMEs using variational Bayesian (VB) method (proposed)
- Experiments
  - Face recognition experiments
  - Conclusion & future work
Optimal model parameters $\Lambda_{ML}$ are estimated by maximizing the likelihood

$$\Lambda_{ML} = \arg \max_{\Lambda} P(O | \Lambda)$$

Training data $O$ ➔ $\Lambda_{ML}$ ➔ Predictive dist. $P(Y | \Lambda_{ML})$

Test data $Y$

<table>
<thead>
<tr>
<th>Amount of training data</th>
<th>Parent population</th>
<th>Large</th>
<th>Small</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data</td>
<td></td>
<td>⬠⬤⬤⬤⬤⬤⬤⬤⬤⬤</td>
<td>⬠⬤⬤⬤⬤圜圜圜圜圜圜</td>
</tr>
<tr>
<td>Estimated distribution</td>
<td></td>
<td>⬠⬤⬤⬤⬤⬤⬤⬤⬤⬤</td>
<td>⬠⬤⬤⬤⬤圜圜圜圜圜圜</td>
</tr>
</tbody>
</table>

Estimation accuracy is decreased by over-fitting problem
**BAYESIAN CRITERION**

- Estimation of posterior distribution

Prior information $\rightarrow$ Prior dist. $P(\Lambda)$

Training data $O$ $\rightarrow$ Posterior dist. $P(\Lambda | O) = \frac{P(O | \Lambda) P(\Lambda)}{P(O)}$

Test data $Y$ $\rightarrow$ Predictive dist. $P(Y | O) = \int P(Y | \Lambda) P(\Lambda | O) d\Lambda$
BAYESIAN CRITERION

- Estimation of posterior distribution

Prior information

Large amount of face images

Prior dist. $P(\Lambda)$
Bayesian Criterion

- Estimation of posterior distribution

Prior information $\rightarrow$ Prior dist. $P(\Lambda)$

Training data $O$
Face images of one subject

Posterior dist. $P(\Lambda | O) = \frac{P(O | \Lambda)P(\Lambda)}{P(O)}$
Bayesian Criterion

- Estimation of posterior distribution

Prior information → Prior dist. $P(\Lambda)$ → Posterior dist. $P(\Lambda | O) = \frac{P(O | \Lambda)P(\Lambda)}{P(O)}$

Training data $O$ → Posterior dist. $P(\Lambda | O)$ → Predictive dist. $P(Y | O) = \int P(Y | \Lambda)P(\Lambda | O)d\Lambda$

- Using prior dist. and marginalization of model parameters
- Complicated integral and expectation computations
**Bayesian Criterion**

- Estimation of posterior distribution

Prior information \rightarrow \text{Prior dist. } P(\Lambda) \rightarrow \text{Posterior dist. } P(\Lambda \mid O) = \frac{P(O \mid \Lambda)P(\Lambda)}{P(O)}

Training data \(O\) \rightarrow \text{Predictive dist. } P(Y \mid O) = \int P(Y \mid \Lambda)P(\Lambda \mid O)d\Lambda

Test data \(Y\) \rightarrow \text{Using prior dist. and marginalization of model parameters}

- Complicated integral and expectation computations

\[ P(O) = \sum_{z} \int \int P(O, x, z \mid \Lambda)P(\Lambda)dxd\Lambda \]

\(x, z\) : Hidden variable
Bayesian Criterion

- Estimation of posterior distribution

Prior information $\rightarrow$ Prior dist. $P(\Lambda)$

Training data $O$ $\rightarrow$ Posterior dist. $P(\Lambda | O) = \frac{P(O | \Lambda)P(\Lambda)}{P(O)}$

Test data $Y$ $\rightarrow$ Predictive dist. $P(Y | O) = \int P(Y | \Lambda)P(\Lambda | O)d\Lambda$

- Using prior dist. and marginalization of model parameters
- Complicated integral and expectation computations
  - MCMC method [Gilks, et al.; '96]
  - MAP method [Gauvain, et al.; '94]
  - VB method [Attias; '99]
VARIATIONAL BAYESIAN (VB) METHOD (1/2)

- Estimation of approximated posterior dist.
- Define a lower bound $\mathcal{F}(Q)$ of log marginal likelihood

\[
\ln P(O) \geq \sum_{z^{(1)}} \sum_{z^{(2)}} \int \int Q(x, z^{(1)}, z^{(2)}, \Lambda) \ln \frac{P(O, z^{(1)}, z^{(2)}, \Lambda)}{Q(x, z^{(1)}, z^{(2)}, \Lambda)} \, dx \, d\Lambda
\]

\[
= \mathcal{F}(Q)
\]

$x$: Factor vector  
$z^{(1)}$: Horizontal state sequence  
$Q(x, z^{(1)}, z^{(2)}, \Lambda)$: Arbitrary dist.  
$z^{(2)}$: Vertical state sequence

- KLD between arbitrary dist. $Q$ and true posterior dist. $P$

\[
\text{KL}(Q \| P) = \ln P(O) - \mathcal{F}(Q)
\]

- Maximizing lower bound $\mathcal{F}(Q)$  
  $\Leftrightarrow$ Minimizing KLD

- Arbitrary dist. $Q$ represents approximated posterior dist.
**VARIATIONAL BAYESIAN (VB) METHOD (2/2)**

- Assume the independency of random variables

\[
P(x, z^{(1)}, z^{(2)}, \Lambda | O) \approx Q(x, z^{(1)}, z^{(2)}, \Lambda)
= Q(x)Q(z^{(1)})Q(z^{(2)})Q(\Lambda)
\]

\[Q(x, z^{(1)}, z^{(2)}, \Lambda) : \text{Arbitrary dist.} \quad Q(\cdot) : \text{VB posterior dist.}\]

- Updates of VB posterior dist. increase the value of lower bound \(\mathcal{F}(Q)\) at each iteration until convergence

<table>
<thead>
<tr>
<th>VB E-step</th>
<th>(\bar{Q}(z^{(1)}) = \arg \max_{Q(z^{(1)})} \mathcal{F})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\bar{Q}(z^{(2)}) = \arg \max_{Q(z^{(2)})} \mathcal{F})</td>
</tr>
<tr>
<td></td>
<td>(\bar{Q}(x) = \arg \max_{Q(x)} \mathcal{F})</td>
</tr>
<tr>
<td>VB M-step</td>
<td>(\bar{Q}(\Lambda) = \arg \max_{Q(\Lambda)} \mathcal{F})</td>
</tr>
</tbody>
</table>

Apply VB method to HMEMs
OUTLINE

- Background & introduction
- Models
  - Probabilistic eigen-image models (PEMs)
  - Separable lattice hidden Markov models (SL-HMMs)
  - Hidden Markov eigen-image models (HMEMs)
- Training criterion
  - Maximum likelihood (ML) criterion
  - Bayesian criterion
    - HEMEs using variational Bayesian (VB) method (proposed)
- Experiments
  - Face recognition experiments
  - Conclusion & future work
## Face Recognition Experiments

### Experimental conditions

<table>
<thead>
<tr>
<th>Database</th>
<th>XM2VTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image size</td>
<td>64 × 64 pixel, Gray-scale</td>
</tr>
<tr>
<td>Training data</td>
<td>6 images per subject × 100 subjects</td>
</tr>
<tr>
<td>Test data</td>
<td>2 images per subject × 100 subjects</td>
</tr>
<tr>
<td>Model structure</td>
<td>SL-HMM, HMEM-PPCA, HMEM-FA</td>
</tr>
<tr>
<td>Number of states</td>
<td>32 × 32 states</td>
</tr>
<tr>
<td>Estimate method</td>
<td>ML method (ML criterion), VB method (Bayesian criterion)</td>
</tr>
<tr>
<td>Prior distribution</td>
<td>Uniform distribution (flat), Universal background model (UBM)</td>
</tr>
</tbody>
</table>
VB method achieved higher recognition rates than ML method
Prior distribution affects the estimation of posterior distribution.
- Uniform distribution (flat)
- Universal background model (UBM)

<table>
<thead>
<tr>
<th>flat</th>
<th>SL-UBM</th>
<th>UBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform dist.</td>
<td>All training samples of all subjects</td>
<td>All training samples of all subjects</td>
</tr>
<tr>
<td></td>
<td>Training UBM</td>
<td>Training UBM</td>
</tr>
<tr>
<td></td>
<td>UBM (SL-HMMs)</td>
<td>UBM (HMEMs)</td>
</tr>
<tr>
<td></td>
<td>Convert SL-HMMs to HMEMs</td>
<td>Tuning the influence of UBM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prior dist.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tuning the influence of UBM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prior dist.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UBM (HMEMs)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tuning the influence of UBM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prior dist.</td>
</tr>
</tbody>
</table>
RECOGNITION RATES (COMPARING PRIOR DIST.)

HMEM-PPCA

HMEM-FA

Recognition rate (%)

Number of factors

Number of factors

Recognition rate (%)

flat
SLUBM
UBM
Recognition rates (Comparing prior dist.)

HMEM-PPCA: flat outperformed SLUBM
HMEM-FA: SLUBM outperformed flat
⇒ SLUBM is effective for FA (diagonal) structure
No significant difference between flat and UBM
⇒ Prior dist. had tuned under the condition
⇒ that the number of factor was one
Significant high recognition rate
⇒ High recognition rate can be expected
⇒ if prior dist. can be set adequately
CONCLUSION

- Focus on technique for modeling geometric variations
- Apply Bayesian criterion to HMEMs
  - Derive VB method for HMEMs
  - Face recognition experiments
    - HMEMs based on VB method outperformed ML method
    - Recognition rate is improved by using an appropriate prior distribution

- Future work
  - Investigation of appropriate parameter sharing structures of HMEMs
  - Experiments on various image recognition tasks