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

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

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

  Integrating two models
  - 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

      ⇒ Over-fitting problem when amount of data is insufficient
  - Bayesian criterion
    - Estimation of posterior distributions using prior information

      ⇒ High generalization ability

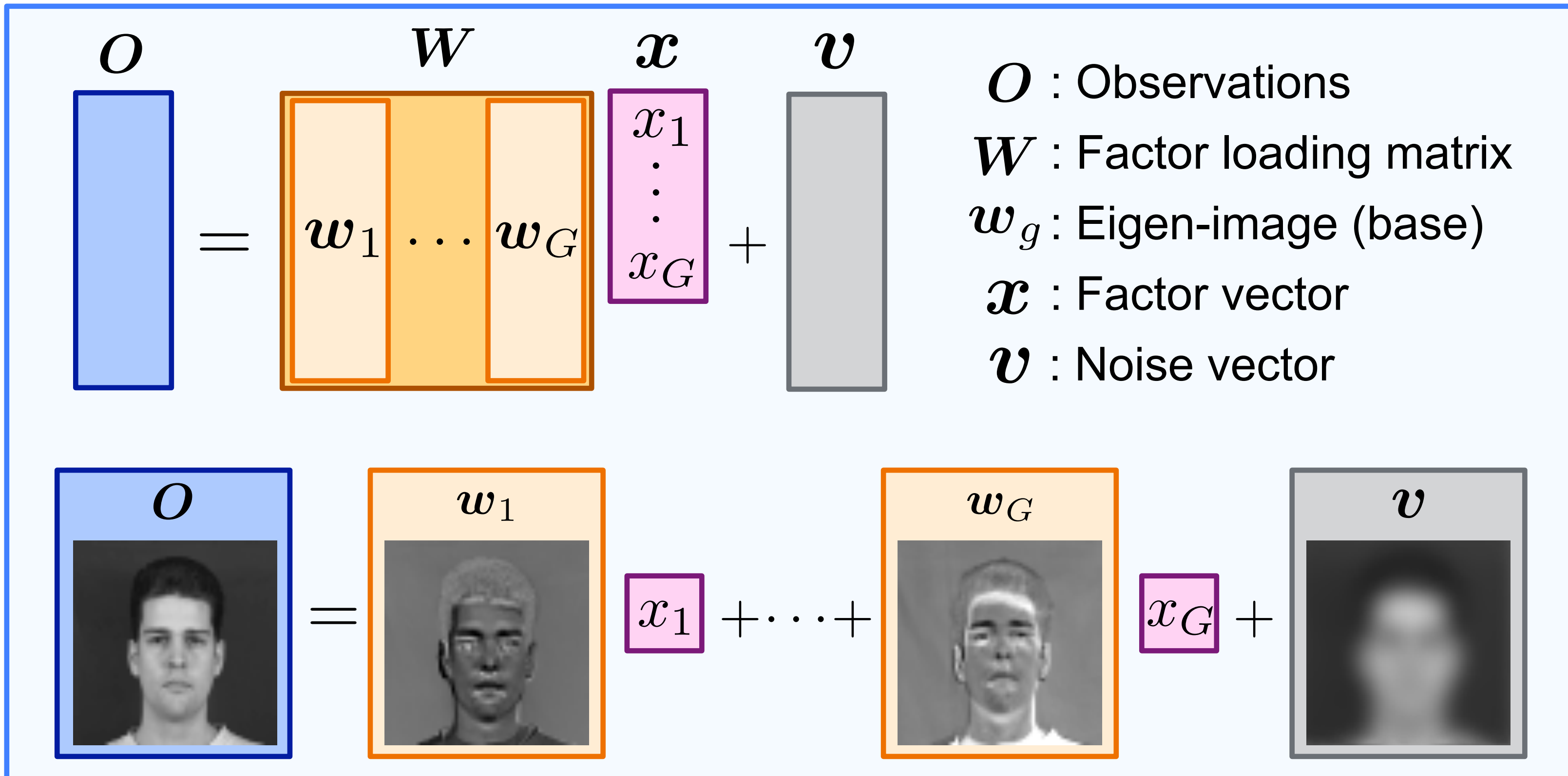
Apply Bayesian criterion to HMEMs

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)

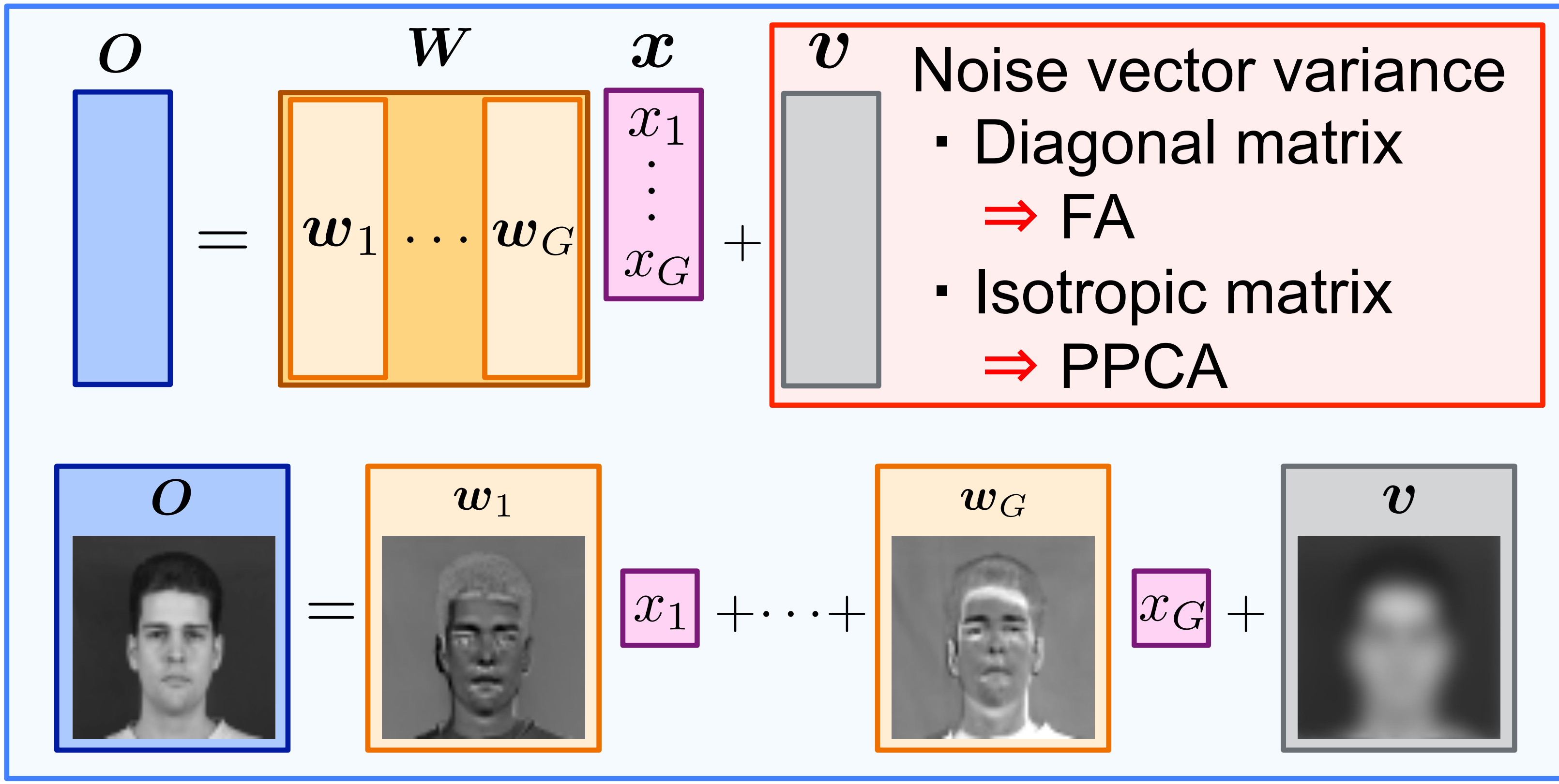


☺ Linear feature extraction based on statistical analysis

☹ 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)

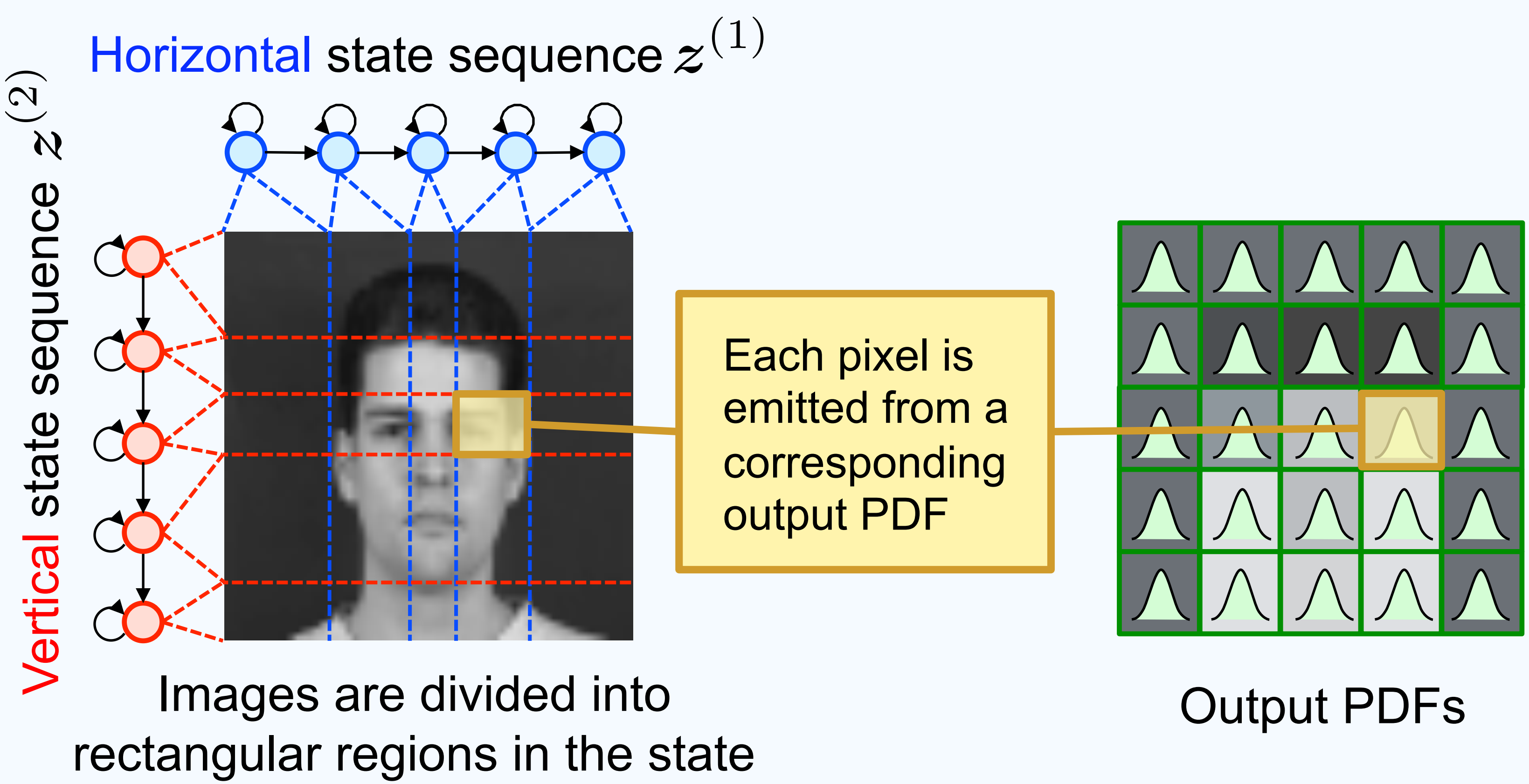


☺ Linear feature extraction based on statistical analysis

☹ 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

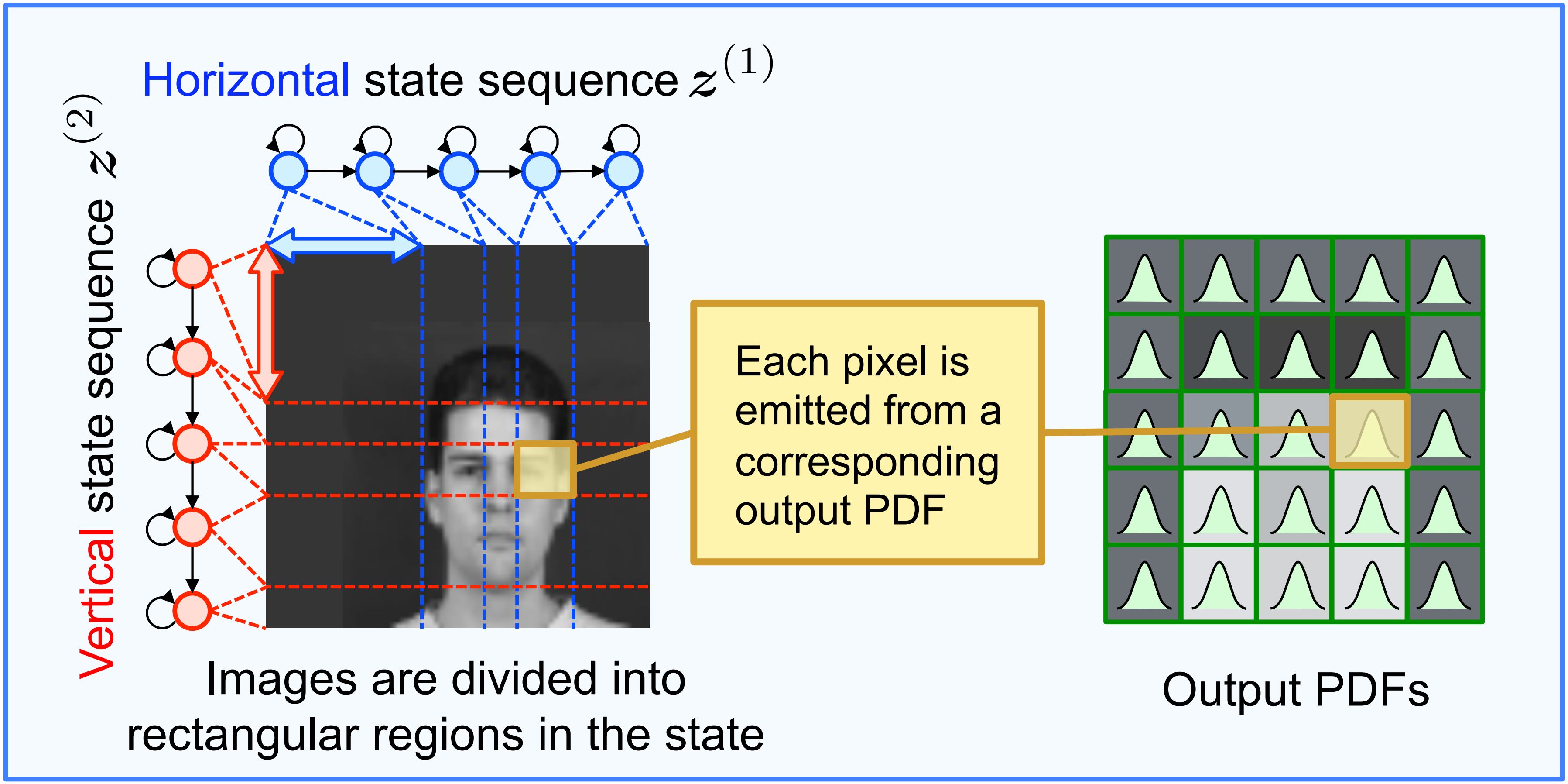


☺ Include size-and-location-normalization

☹ 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

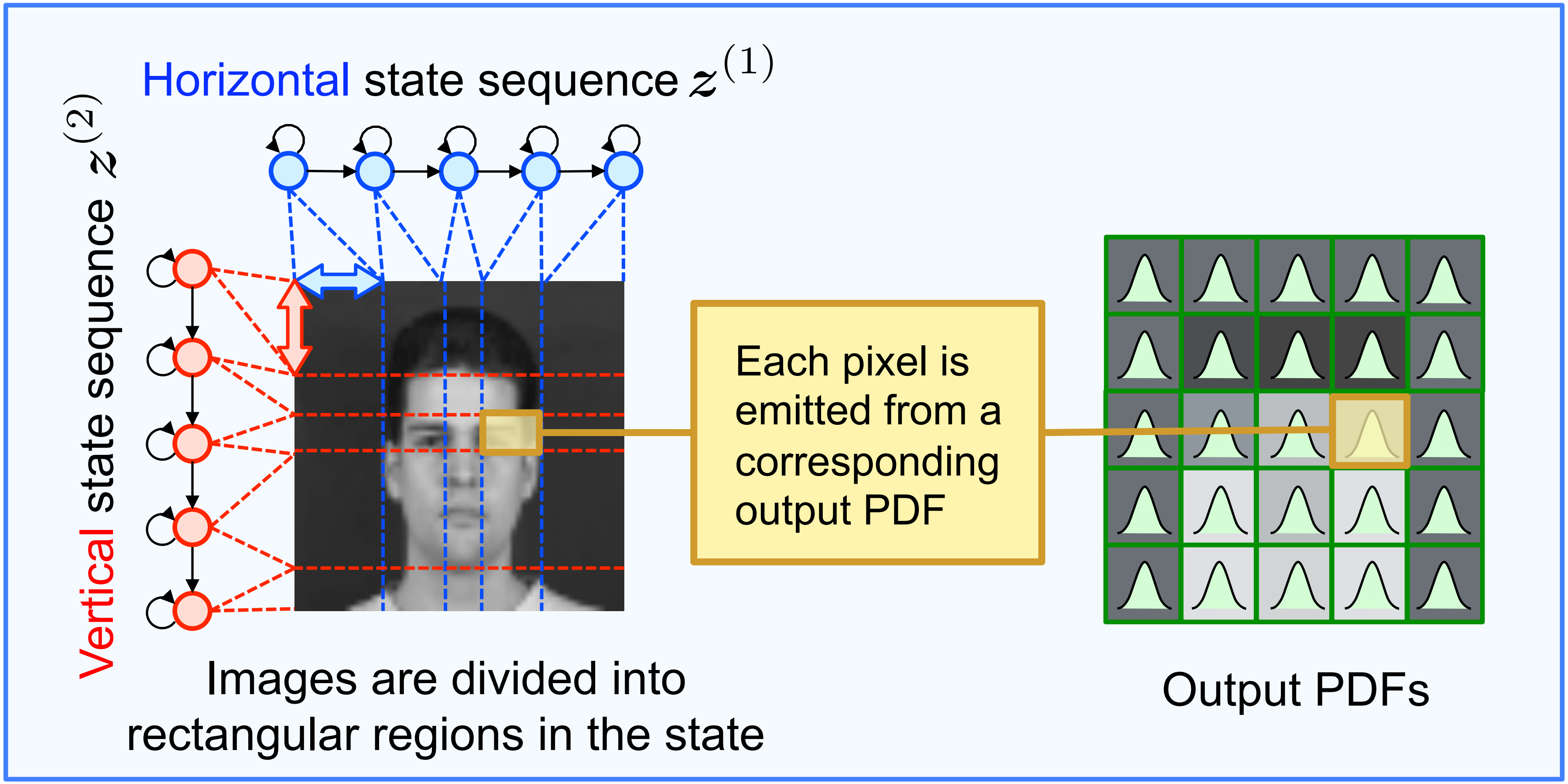


☺ Include size-and-location-normalization

☹ 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

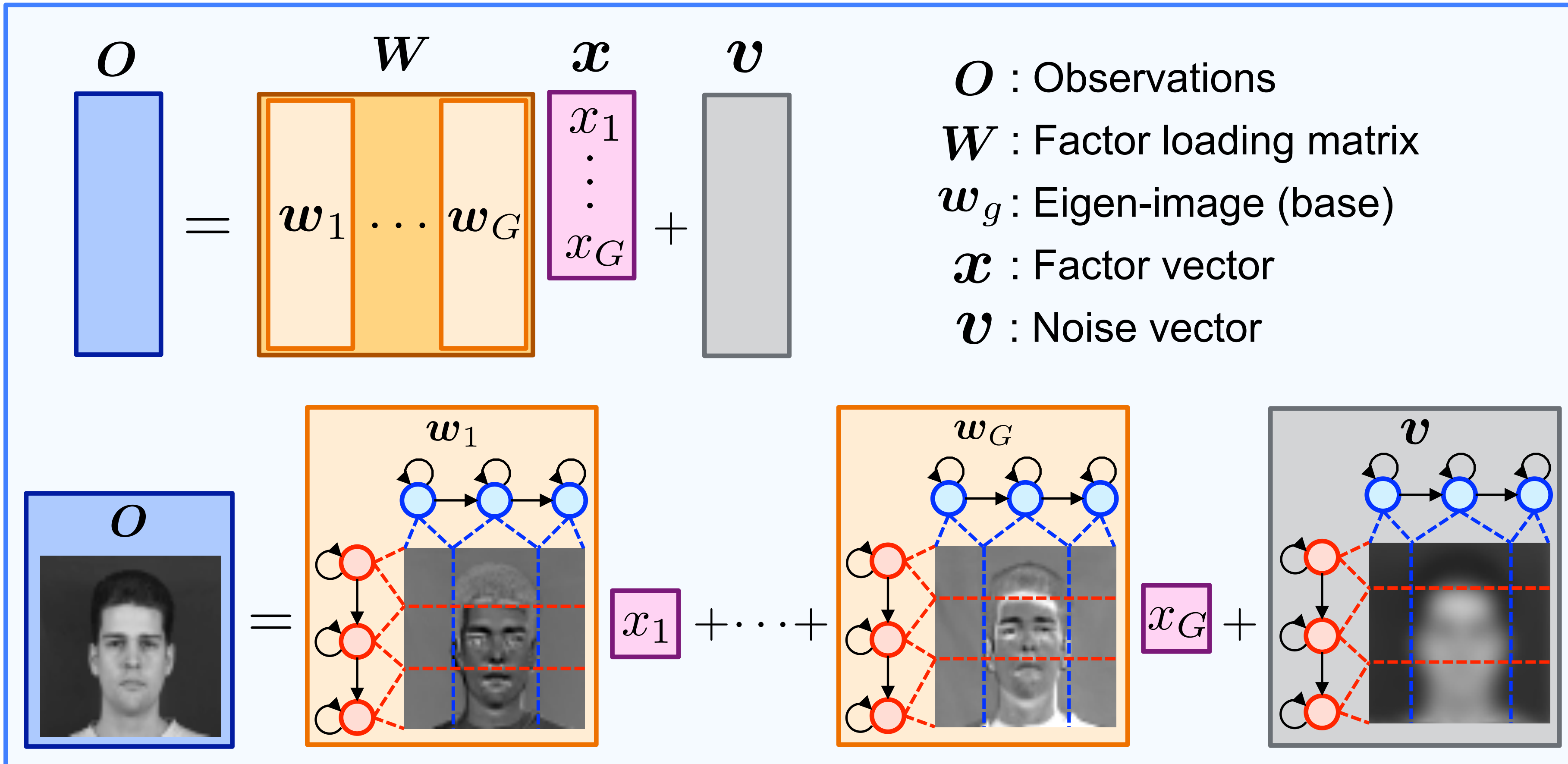


☺ Include size-and-location-normalization

☹ 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



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

OUTLINE

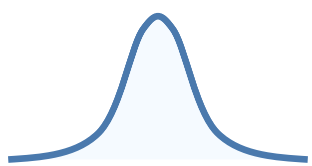
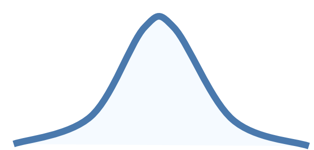
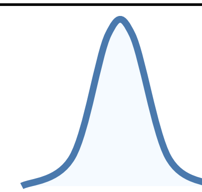
- Background & introduction
- Models
  - Probabilistic eigen-image models (PEMs)
  - Separable lattice hidden Markov models (SL-HMMs)
  - Hidden Markov eigen-image models (HMEMs)
- Training criterion
  - Maximum likelihood (ML) criterion
  - Bayesian criterion
    - HEMEs using variational Bayesian (VB)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

Maximum likelihood (ML) criterion

- Optimal model parameters $\mathbf{\Lambda}_{\mathrm{ML}}$ are estimated by maximizing the likelihood

Training data $\boldsymbol{O}$ → $\mathbf{\Lambda}_{\mathrm{ML}} = \arg\max_{\mathbf{\Lambda}} P(\boldsymbol{O} \mid \mathbf{\Lambda})$

Test data $\boldsymbol{Y}$ → Predictive dist. $P(\boldsymbol{Y} \mid \mathbf{\Lambda}_{\mathrm{ML}})$

Amount of training data	Parent population	Large	Small
Training data			
Estimated distribution			

☹ Estimation accuracy is decreased by over-fitting problem

Bayesian criterion

- Estimation of posterior distribution

Prior information → Prior dist. $P(\boldsymbol{\Lambda})$

Training data $\boldsymbol{O}$ → Posterior dist. $P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) = \dfrac{P(\boldsymbol{O} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda})}{P(\boldsymbol{O})}$

Test data $\boldsymbol{Y}$ → Predictive dist. $P(\boldsymbol{Y} \mid \boldsymbol{O}) = \int P(\boldsymbol{Y} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) \mathrm{d}\boldsymbol{\Lambda}$

Bayesian criterion

- Estimation of posterior distribution

Prior information

Prior dist. $P(\mathbf{\Lambda})$

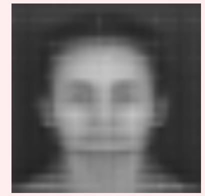
Large amount of face images

Bayesian criterion

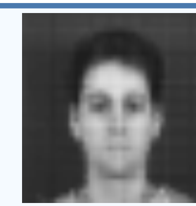
- Estimation of posterior distribution

Prior information

Prior dist. $P(\mathbf{\Lambda})$



Posterior dist. $P(\mathbf{\Lambda} \mid \mathbf{O}) = \dfrac{P(\mathbf{O} \mid \mathbf{\Lambda}) P(\mathbf{\Lambda})}{P(\mathbf{O})}$



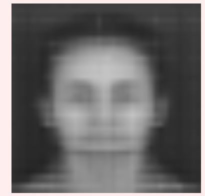
Training data $\boldsymbol{O}$

Face images of one subject

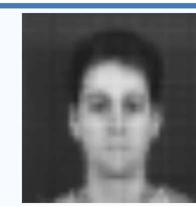
Bayesian criterion

- Estimation of posterior distribution

Prior information → Prior dist. $P(\boldsymbol{\Lambda})$



Training data $\boldsymbol{O}$ → Posterior dist. $P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) = \dfrac{P(\boldsymbol{O} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda})}{P(\boldsymbol{O})}$



Test data $\boldsymbol{Y}$ → Predictive dist. $P(\boldsymbol{Y} \mid \boldsymbol{O}) = \int P(\boldsymbol{Y} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) \mathrm{d}\boldsymbol{\Lambda}$

☺ Using prior dist. and marginalization of model parameters

☹ Complicated integral and expectation computations

Bayesian criterion

- Estimation of posterior distribution

Prior information → Prior dist. $P(\boldsymbol{\Lambda})$

Training data $\boldsymbol{O}$ → Posterior dist. $P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) = \dfrac{P(\boldsymbol{O} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda})}{\boxed{P(\boldsymbol{O})}}$

Test data $\boldsymbol{Y}$ → Predictive dist. $P(\boldsymbol{Y} \mid \boldsymbol{O}) = \displaystyle\int P(\boldsymbol{Y} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) \mathrm{d}\boldsymbol{\Lambda}$

Using prior dist. and marginalization of model parameters

Complicated integral and expectation computations

$$P(\boldsymbol{O}) = \sum_{\boldsymbol{z}} \iint P(\boldsymbol{O}, \boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{\Lambda}$$

$\boldsymbol{x}, \boldsymbol{z}$: Hidden variable

BAYESIAN CRITERION

- Estimation of posterior distribution

Prior information → Prior dist. $P(\boldsymbol{\Lambda})$

Training data $\boldsymbol{O}$ → Posterior dist. $P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) = \dfrac{P(\boldsymbol{O} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda})}{P(\boldsymbol{O})}$

Test data $\boldsymbol{Y}$ → Predictive dist. $P(\boldsymbol{Y} \mid \boldsymbol{O}) = \int P(\boldsymbol{Y} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) \mathrm{d}\boldsymbol{\Lambda}$

☺ Using prior dist. and marginalization of model parameters

☹ Complicated integral and expectation computations $\Rightarrow$
- 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

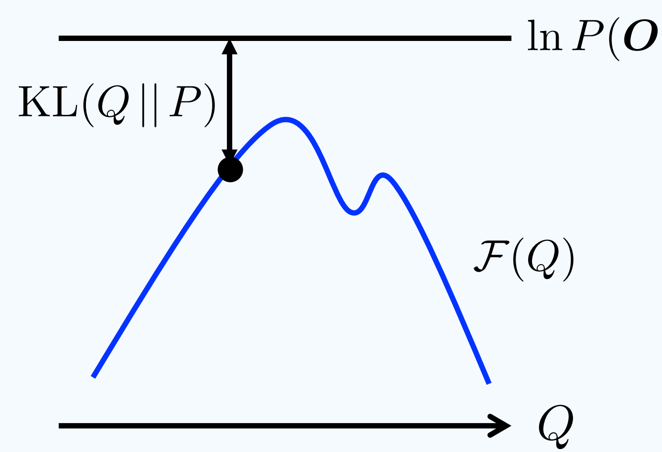
$$
\begin{aligned}
\ln P(\boldsymbol{O}) &\geq \sum_{\boldsymbol{z}^{(1)}} \sum_{\boldsymbol{z}^{(2)}} \iint Q(\boldsymbol{x}, \boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \boldsymbol{\Lambda}) \ln \frac{P(\boldsymbol{O}, \boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \boldsymbol{\Lambda})}{Q(\boldsymbol{x}, \boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \boldsymbol{\Lambda})} \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{\Lambda} \\
&= \mathcal{F}(Q)
\end{aligned}
$$

$\boldsymbol{x}$: Factor vector $\quad$ $\boldsymbol{z}^{(1)}$: Horizontal state sequence

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

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

$$
\mathrm{KL}(Q \,||\, P) = \ln P(\boldsymbol{O}) - \mathcal{F}(Q)
$$



- Maximizing lower bound $\mathcal{F}(Q)$
  $\Leftrightarrow$ Minimizing KLD
- Arbitrary dist. Q represents approximated posterior dist.

Variational Bayesian (VB) method (2/2)

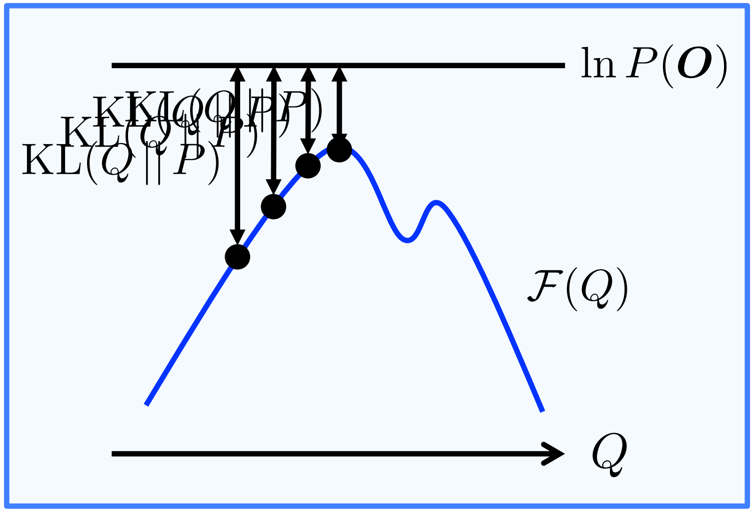
- Assume the independency of random variables

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

$Q(\boldsymbol{x}, \boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \boldsymbol{\Lambda})$: Arbitrary dist. $Q(\cdot)$: VB posterior dist.

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

VB E-step	$\bar{Q}(\boldsymbol{z}^{(1)}) = \arg\max_{Q(\boldsymbol{z}^{(1)})} \mathcal{F}$ $\bar{Q}(\boldsymbol{z}^{(2)}) = \arg\max_{Q(\boldsymbol{z}^{(2)})} \mathcal{F}$ $\bar{Q}(\boldsymbol{x}) = \arg\max_{Q(\boldsymbol{x})} \mathcal{F}$
VB M-step	$\bar{Q}(\boldsymbol{\Lambda}) = \arg\max_{Q(\boldsymbol{\Lambda})} \mathcal{F}$



Apply VB method to HMEMs

Outline

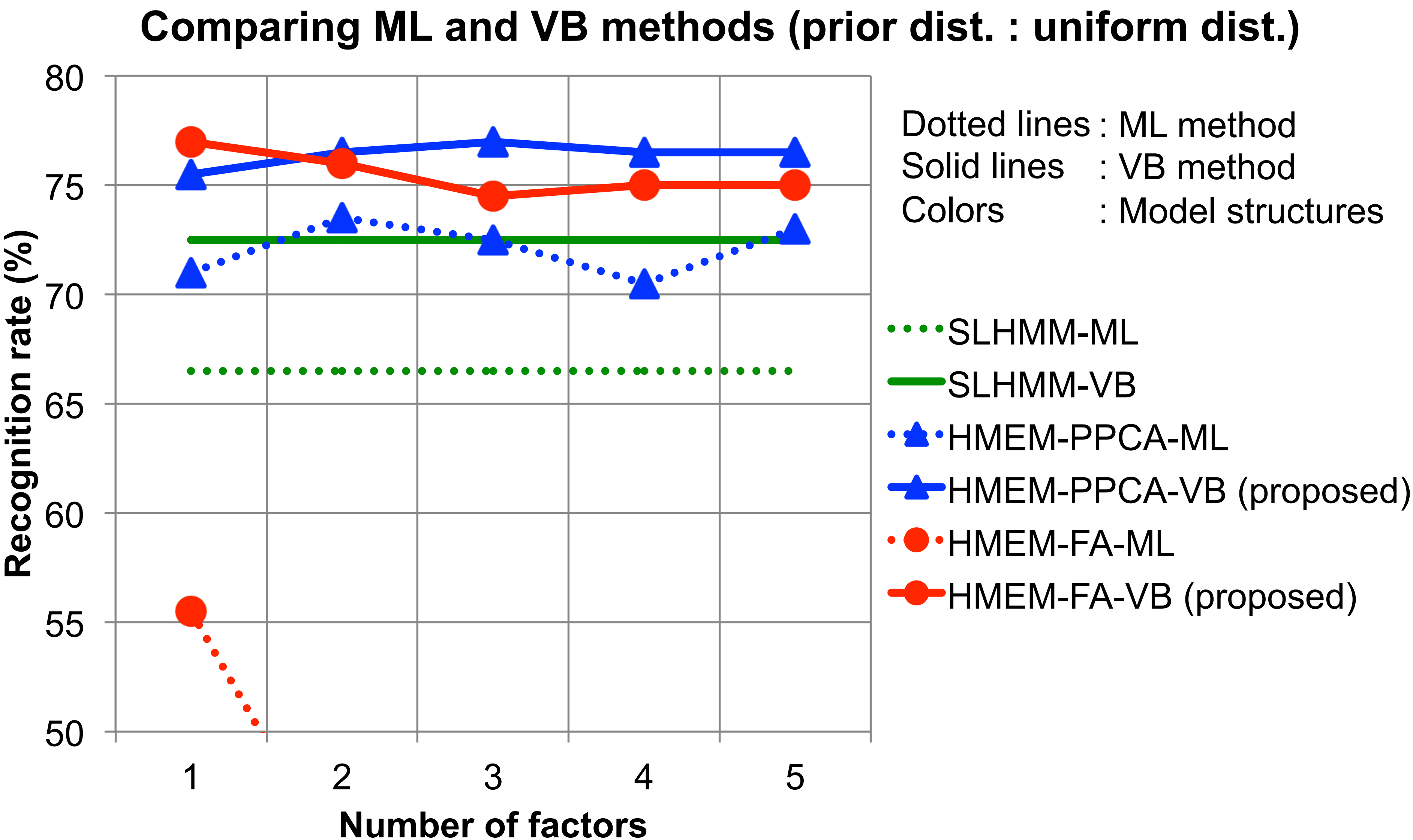
- Background & introduction
- Models
  - Probabilistic eigen-image models (PEMs)
  - Separable lattice hidden Markov models (SL-HMMs)
  - Hidden Markov eigen-image models (HMEMs)
- Training criterion
  - Maximum likelihood (ML) criterion
  - Bayesian criterion
    - HEMEs using variational Bayesian (VB)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

FACE RECOGNITION EXPERIMENTS

- Experimental conditions

Database	XM2VTS
Image size	64 × 64 pixel, Gray-scale
Training data	6 images per subject × 100 subjects
Test data	2 images per subject × 100 subjects
Model structure	SL-HMM, HMEM-PPCA, HMEM-FA
Number of states	32 × 32 states
Estimate method	ML method (ML criterion), VB method (Baysian criterion)
Prior distribution	Uniform distribution (flat), Universal background model (UBM)

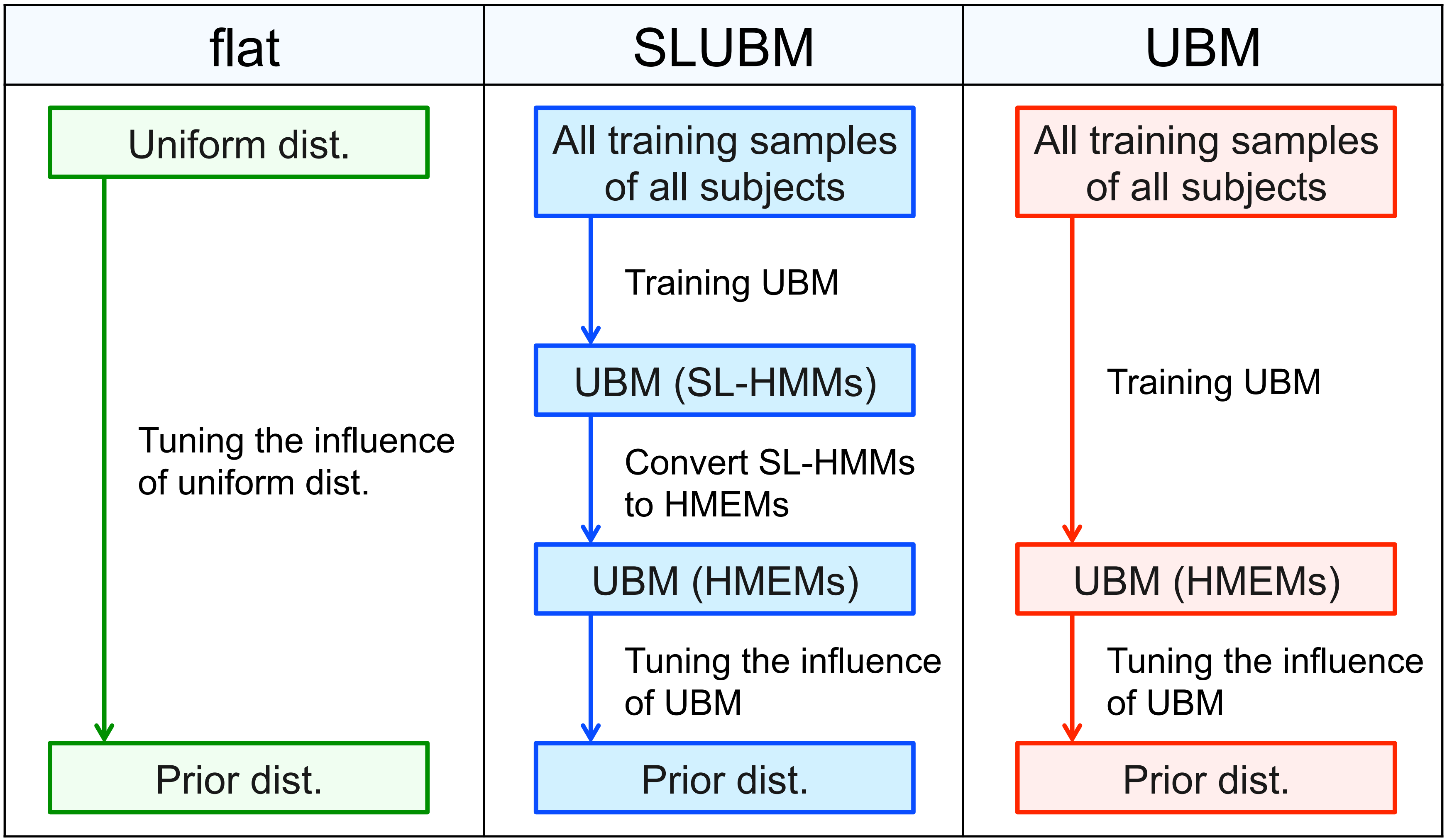
RECOGNITION RATES (COMPARING ML AND VB)



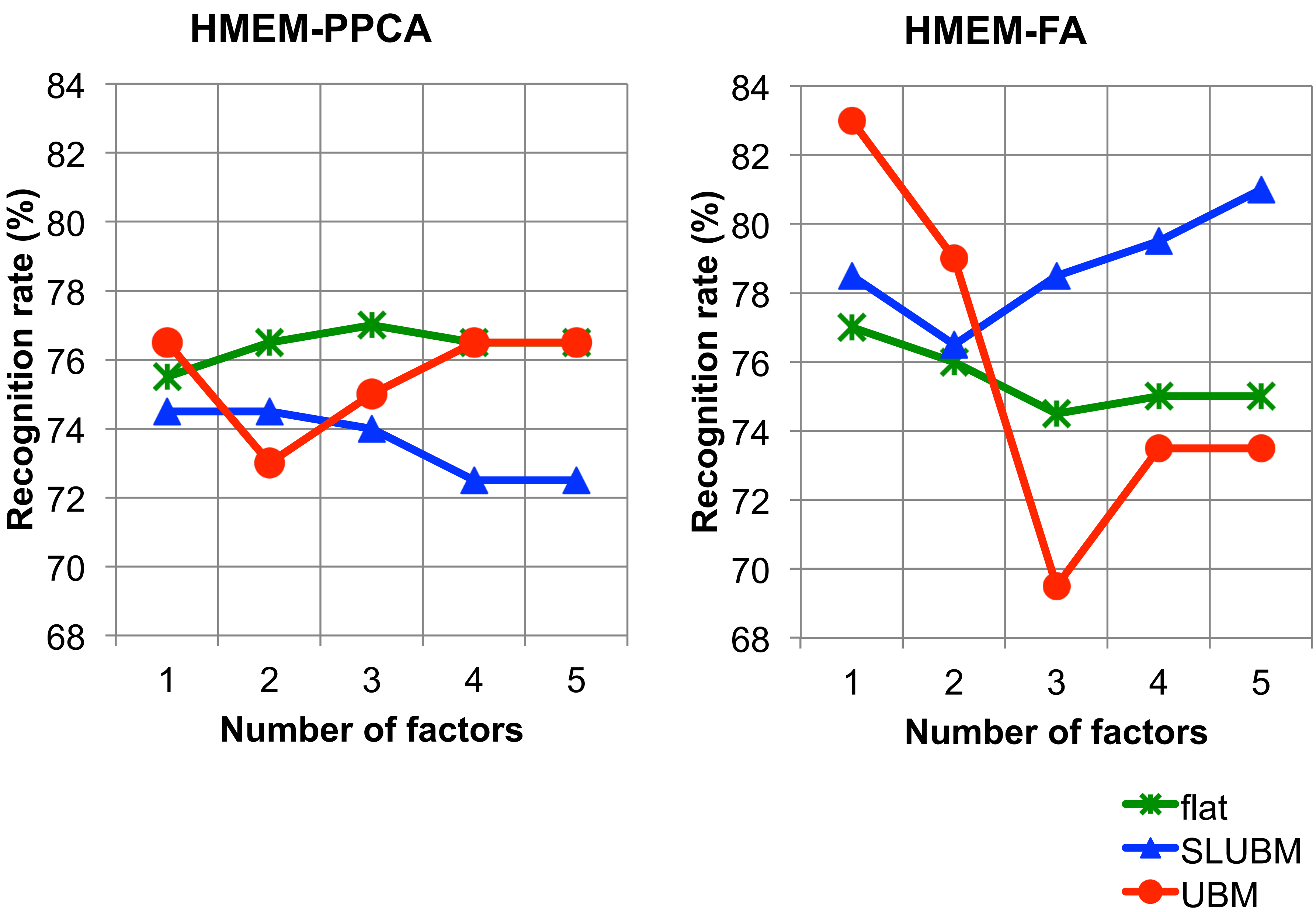
VB method achieved higher recognition rates than ML method

PRIOR DISTRIBUTION

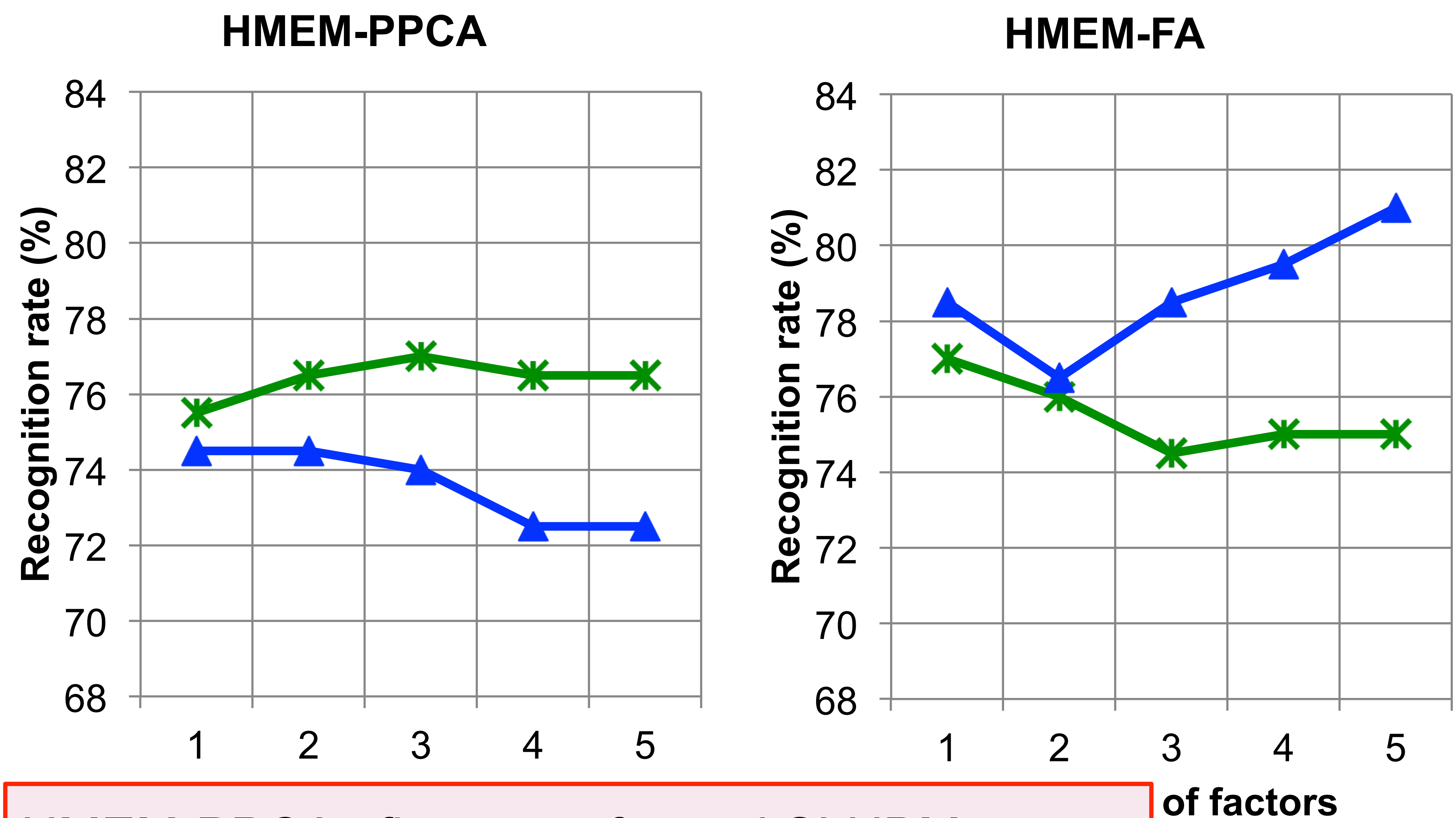
- Prior dist. affects the estimation of posterior dist.
  - Uniform distribution (flat)
  - Universal background model (UBM)



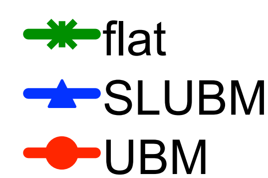
Recognition rates (Comparing prior dist.)



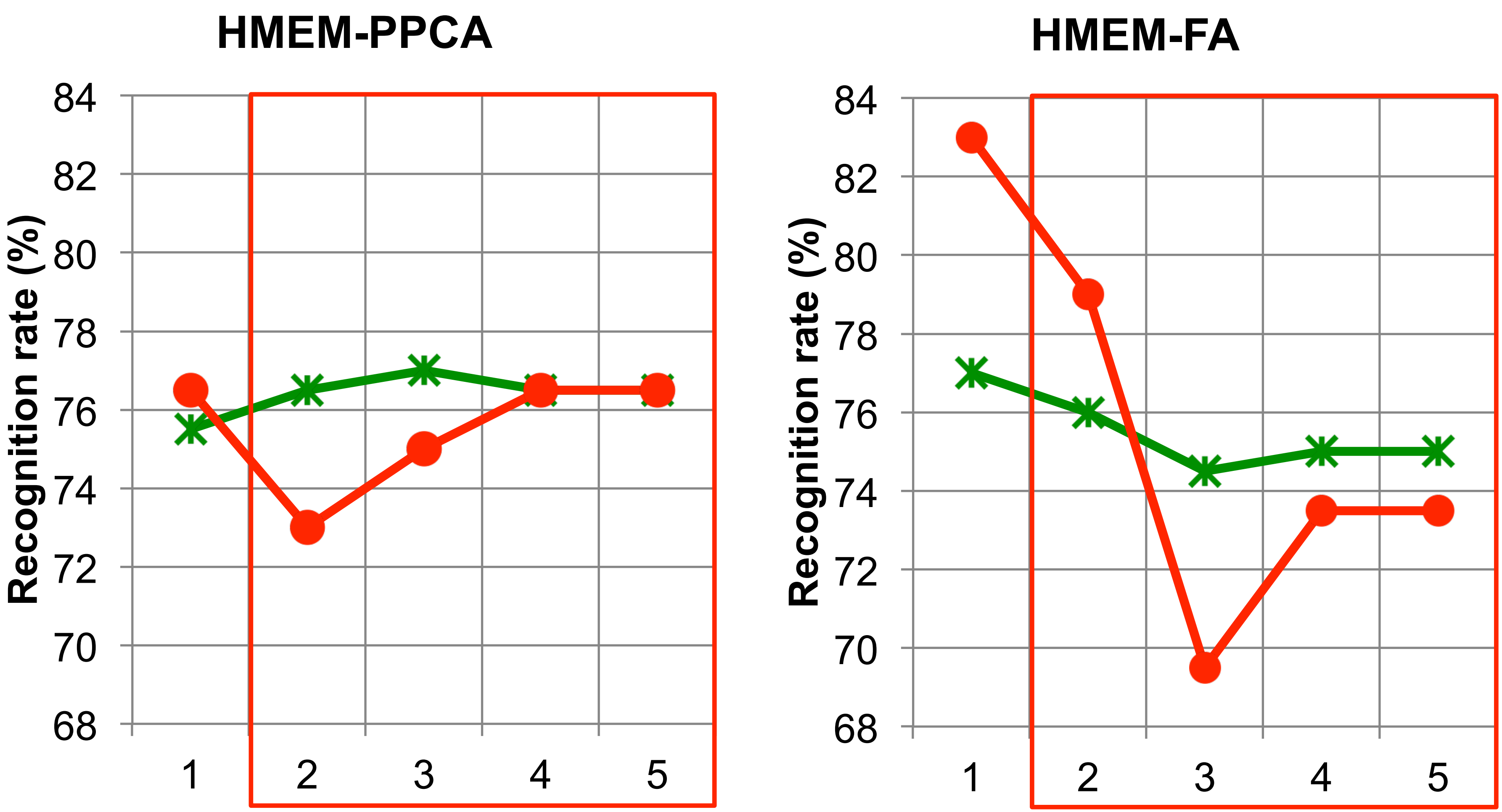
RECOGNITION RATES (COMPARING PRIOR DIST.)



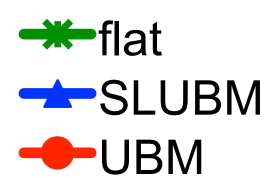
HMEM-PPCA : flat outperformed SLUBM
HMEM-FA : SLUBM outperformed flat
⇒ SLUBM is effective for FA (diagonal) structure



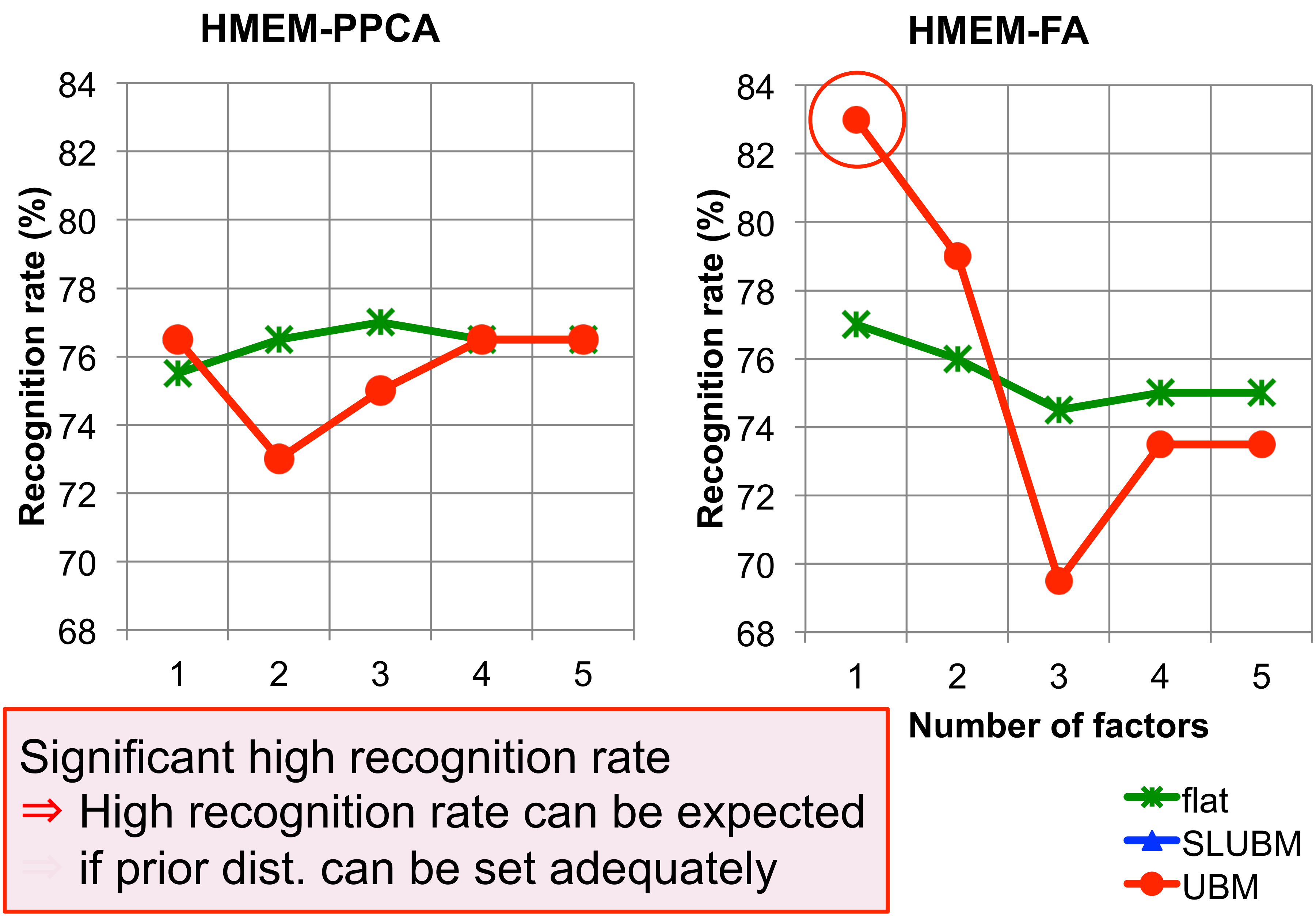
Recognition rates (Comparing prior dist.)



No significant difference between flat and UBM
⇒ Prior dist. had tuned under the condition
that the number of factor was one



Recognition rates (Comparing prior dist.)



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