ICDM 2015
IEEE International Conference on Data Mining

LOGDET DIVERGENCE BASED SPARSE NON-NEGATIVE MATRIX FACTORIZATION FOR STABLE REPRESENTATION

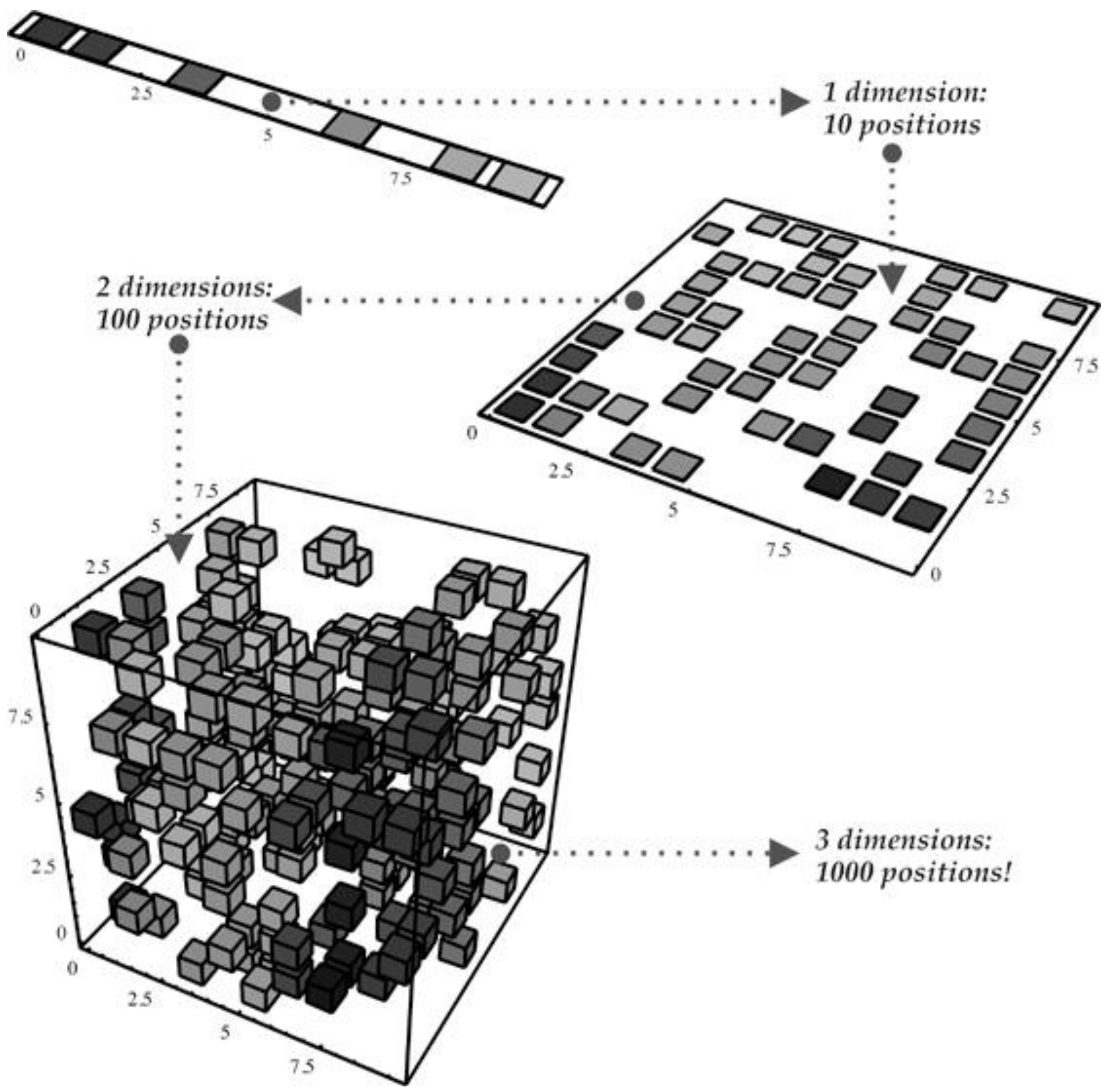
Qing Liao †, Naiyang Guan ‡, Qian Zhang †

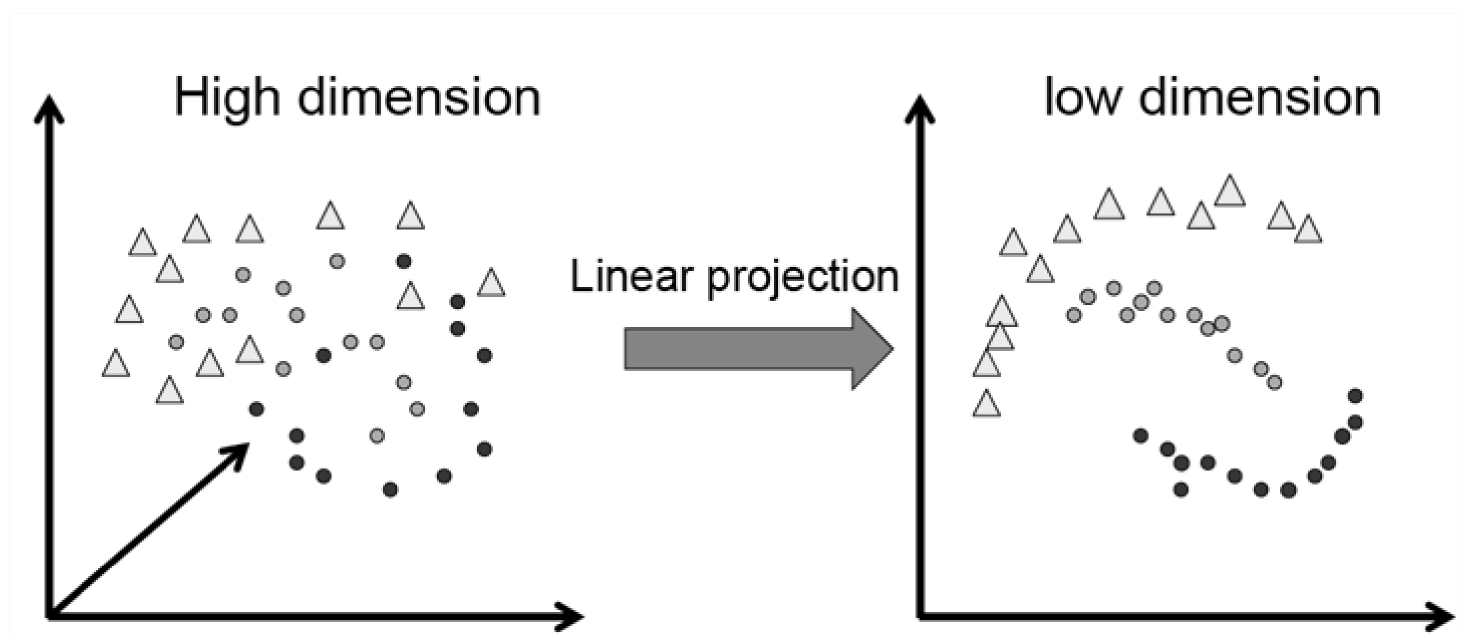
qnature@ust.hk, ny_guan@nudt.edu.cn, qzhang@ust.hk

† The Hong Kong University of Science and Technology
‡ National University of Defense Technology

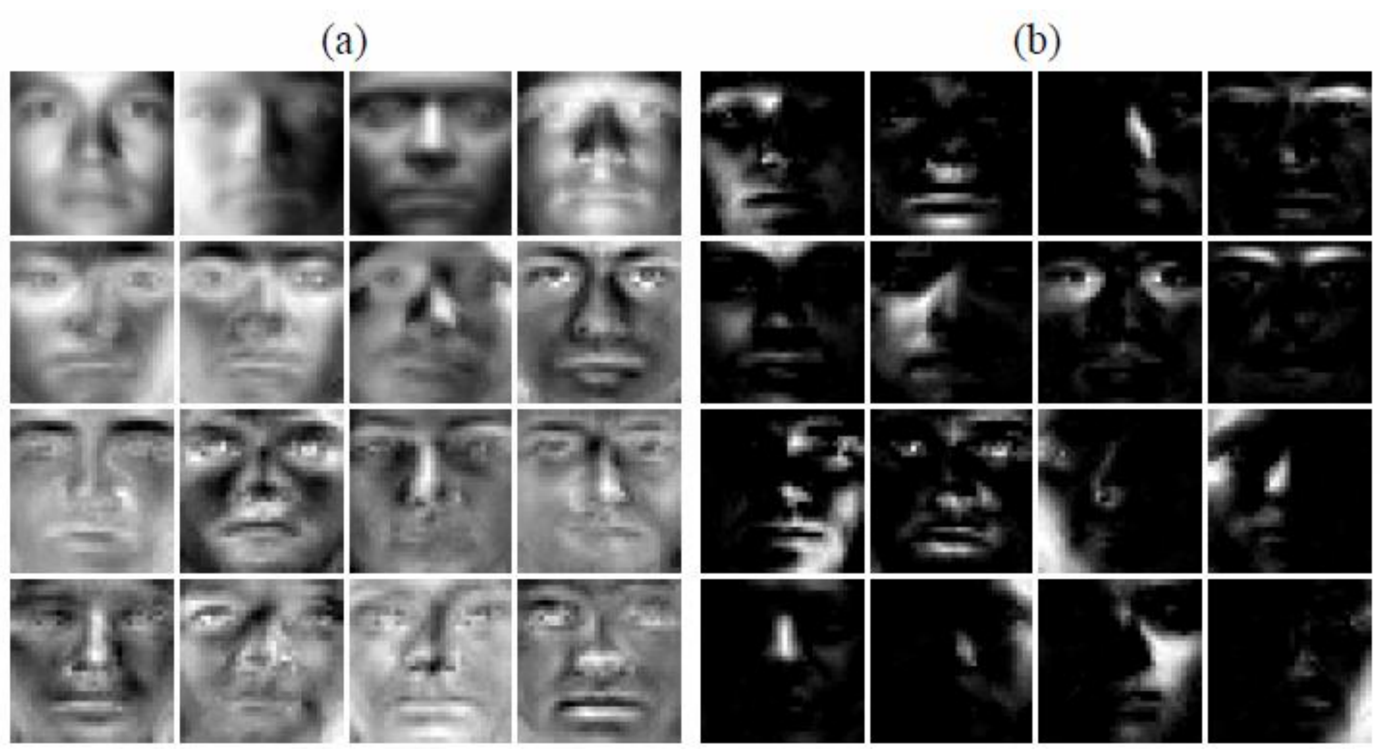
November 17, 2015

CURSE OF DIMENSIONALITY





MATRIX FACTORIZATION



- Non-negative matrix factorization has been widely applied in data representation
  - Non-negativity constraints
  - Learn parts-based representation
- Non-negative matrix factorization (NMF) vs. principle component analysis (PCA)
- The basis learned by:
  a) PCA, and
  b) Our method on the YaleB dataset

MOTIVING EXAMPLE

(a)
$$\overset{X}{\begin{bmatrix} 110 & 113 & 107 & 112 & 172 \\ 106 & 70 & 94 & 92 & 96 \\ 147 & 81 & 133 & 114 & 116 \\ 171 & 100 & 135 & 158 & 108 \\ 242 & 153 & 205 & 216 & 193 \end{bmatrix}} \approx \overset{W_1}{\begin{bmatrix} 1 & 11 & 3 & 8 \\ 2 & 4 & 6 & 4 \\ 3 & 2 & 9 & 6 \\ 4 & 6 & 12 & 1 \\ 5 & 9 & 15 & 6 \end{bmatrix}} \times \overset{H_1}{\begin{bmatrix} 1 & 6 & 8 & 3 & 2 \\ 3 & 6 & 2 & 6 & 7 \\ 12 & 3 & 7 & 9 & 4 \\ 5 & 4 & 7 & 2 & 10 \end{bmatrix}}$$

(b)
$$\overset{X}{\begin{bmatrix} 110 & 113 & 107 & 112 & 172 \\ 106 & 70 & 94 & 92 & 96 \\ 147 & 81 & 133 & 114 & 116 \\ 171 & 100 & 135 & 158 & 108 \\ 242 & 153 & 205 & 216 & 193 \end{bmatrix}} \approx \overset{W_2}{\begin{bmatrix} 7.91 & 5.18 & 0.79 & 2.56 \\ 3.18 & 5.75 & 2.10 & 0.87 \\ 5.10 & 5.73 & 5.14 & 0.53 \\ 5.46 & 10.53 & 1.72 & 0.06 \\ 8.31 & 13.06 & 4.44 & 1.10 \end{bmatrix}} \times \overset{H_2}{\begin{bmatrix} 4.89 & 5.19 & 4.57 & 6.10 & 6.29 \\ 11.92 & 6.44 & 8.53 & 11.31 & 5.74 \\ 10.27 & 1.96 & 11.00 & 3.46 & 6.37 \\ 0.45 & 14.41 & 6.88 & 0.93 & 33.62 \end{bmatrix}}$$

- Rank-deficiency problem: the rank of the learned basis is not equivalent to the predefined reduced dimensionality

a) NMF obtains perfect factorization without any loss, but the rank-deficient basis may cause redundant representation

b) LDS-NMF obtains an approximate factorization with a small loss 8.28, but it can yield full column-rank basis

- We incorporate the Logdet divergence regularization into NMF to reduce the risk of the rank-deficiency problem
- We develop a multiplicative update rule (MUR) to optimize LDS-NMF and proven its convergence

LOGDET DIVERGENCE BASED SPARSE NMF

Given any two positive definite matrices with the same dimensionality, the Logdet divergence is defined:

$$D_{ld}(A, A_0) = \varphi(A) - \varphi(A_0) - \left\langle \nabla_{\varphi}(A), A - A_0 \right\rangle$$

where $\varphi(A) = -logdet(A)$, and $\langle B, C \rangle = trace(B^T C)$.

- Scale invariance: $D_{ld}(\alpha A, \alpha A_0) = D_{ld}(A, A_0)$, for any positive α
- Translation invariance: $D_{ld}(SAS^T, SA_0S^T) = D_{ld}(A, A_0)$, for any invertible matrix S
- Rang space preservation: $D_{ld}(A, A_0)$ is finite if and only if $range(A) = range(A_0)$

LEMMA 1

Given two positives m and r which satisfy $r \leq m$, if the matrix $W' \in R^{m \times r}$ minimizes the following objective:

$$W' = argmin_W D_{ld}(W^T W, I_r)$$

then $range(W') = r$.

LOGDET DIVERGENCE BASED SPARSE NMF

- LDS-NMF final objective function:

$$\min_{W,H\geq 0} D(X|WH) + \frac{\lambda}{2} D_{ld}(W^T W, I_r) + \gamma \sum_{j=1}^{n} \left\| H_{.j} \right\|_1$$

- Optimization algorithm:

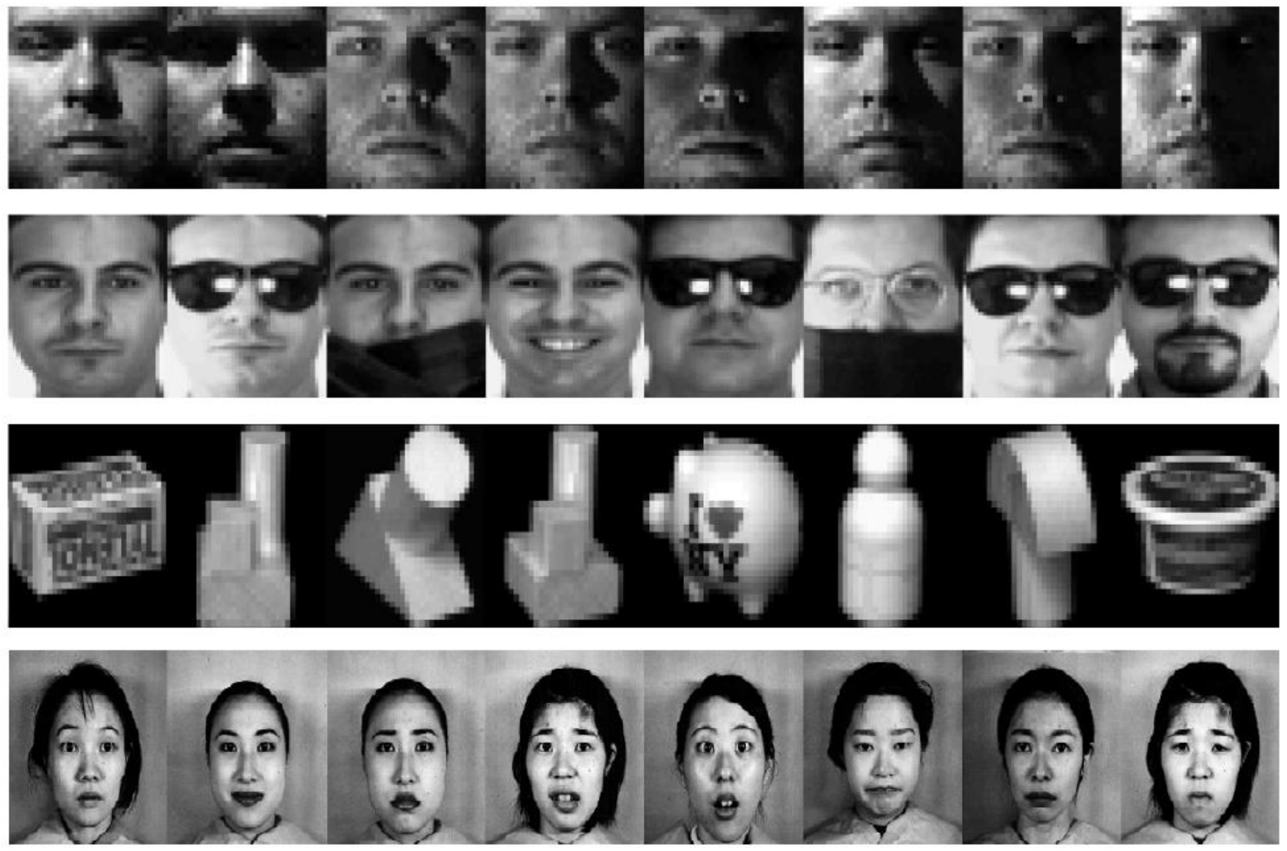
$$W \leftarrow W \otimes \frac{XDH^T + \lambda W \lfloor (W^T W)^{-1} \rfloor_+}{WHDH^T + \lambda W + \lambda W \lfloor (W^T W)^{-1} \rfloor_-}$$

$$H \leftarrow H \otimes \frac{W^T XD}{W^T WHD + \gamma}$$

$$D_{ii} \leftarrow \frac{1}{\sqrt{\sum_{j=1}^{m} (X - WH)_{ji}^2}}$$

EXPERIMENTS

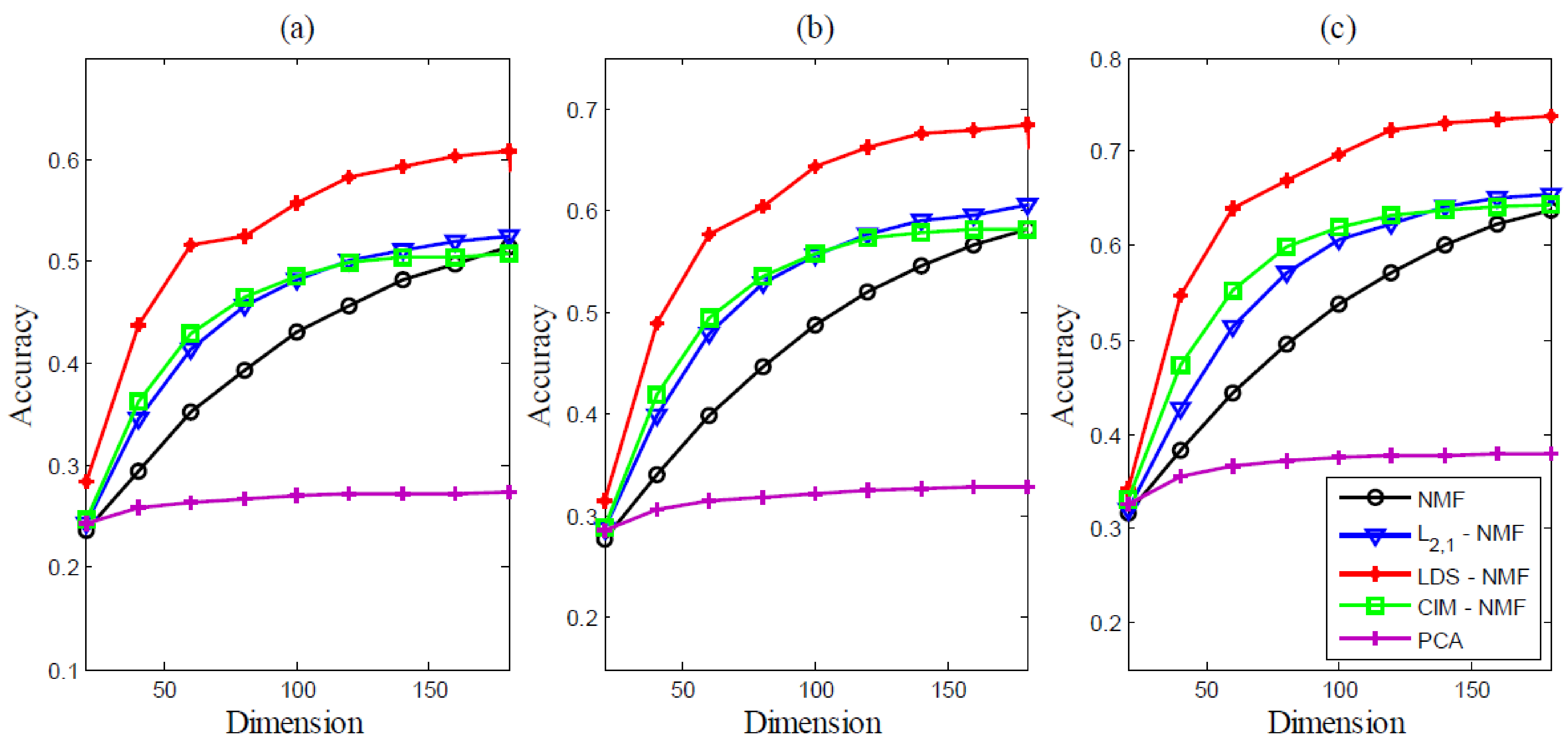
(a)
(b)
(c)
(d)



Examples of images in the four datasets: (a) YaleB, (b) AR, (c) COIL-20, and (d) JAFFE

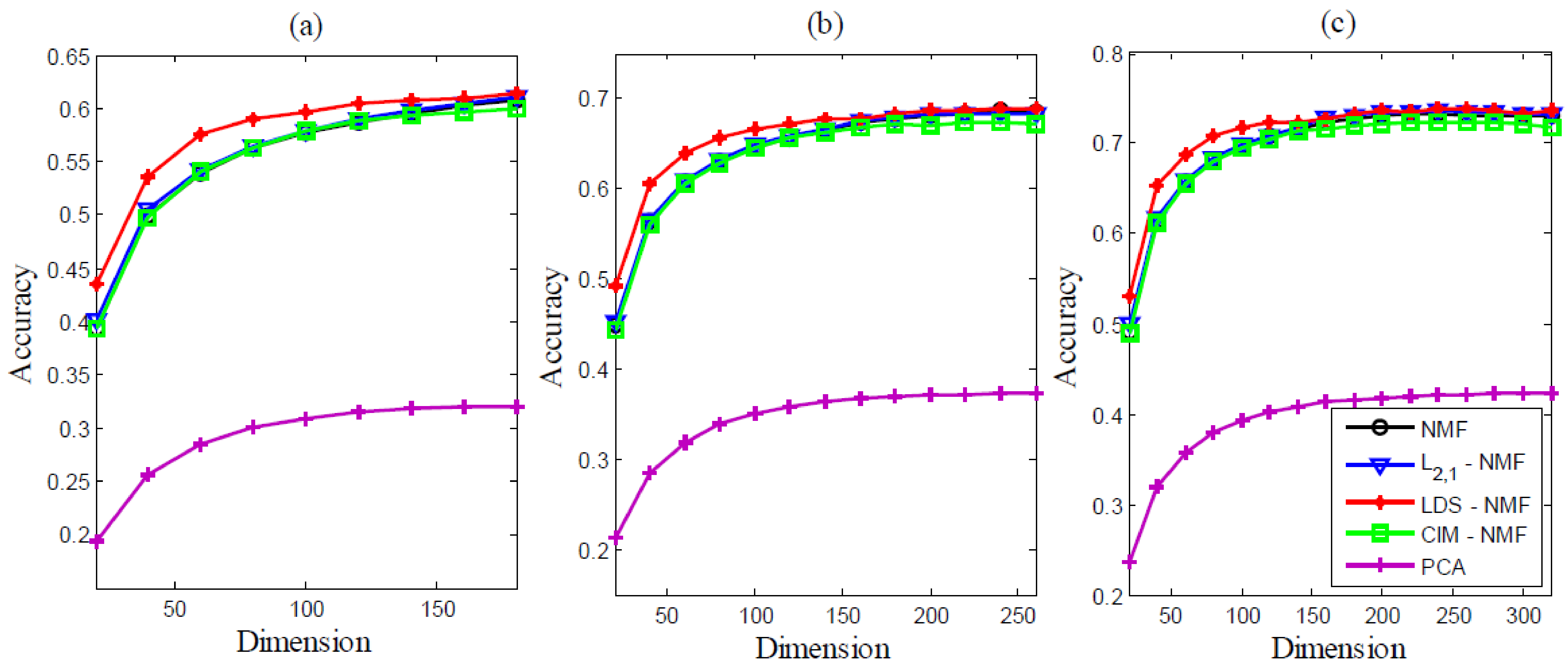
- Comparison methods
  - NMF
  - PCA
  - $L_{2,1}$-NMF
  - CIM-NMF
- Face recognition
  - Nearest neighbor (NN), SVM
  - YaleB, AR
  - Size of training set: 5 / 7 / 9
- Image clustering
  - Accuracy, NMI
  - COIL-20, JAFFE

FACE RECOGNITION



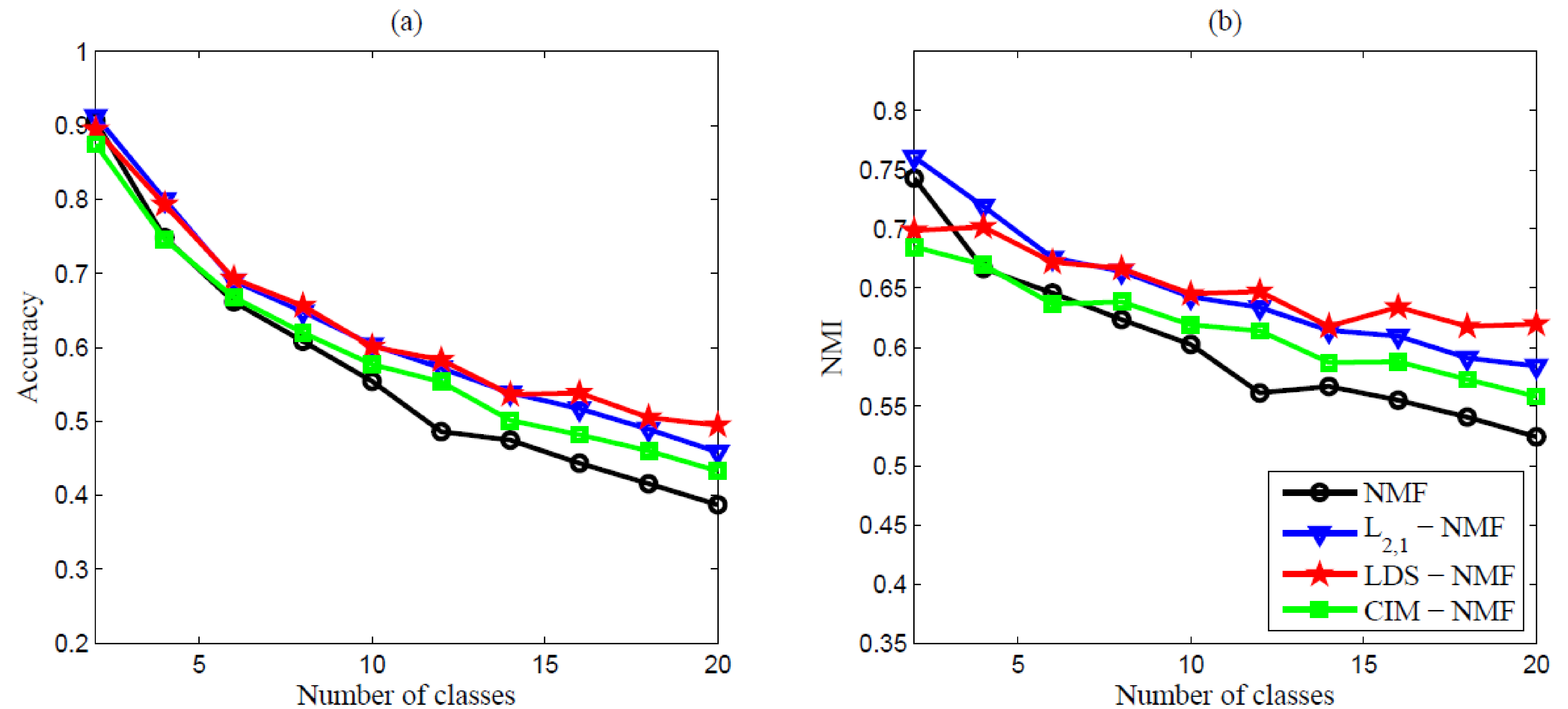
- AR dataset: 2,600 face images of 100 subjects
- Average accuracy vs. reduced dimensionality when (a) 5, (b) 7, and (c) 9 images of each subject are randomly selected for training on the AR dataset

FACE RECOGNITION



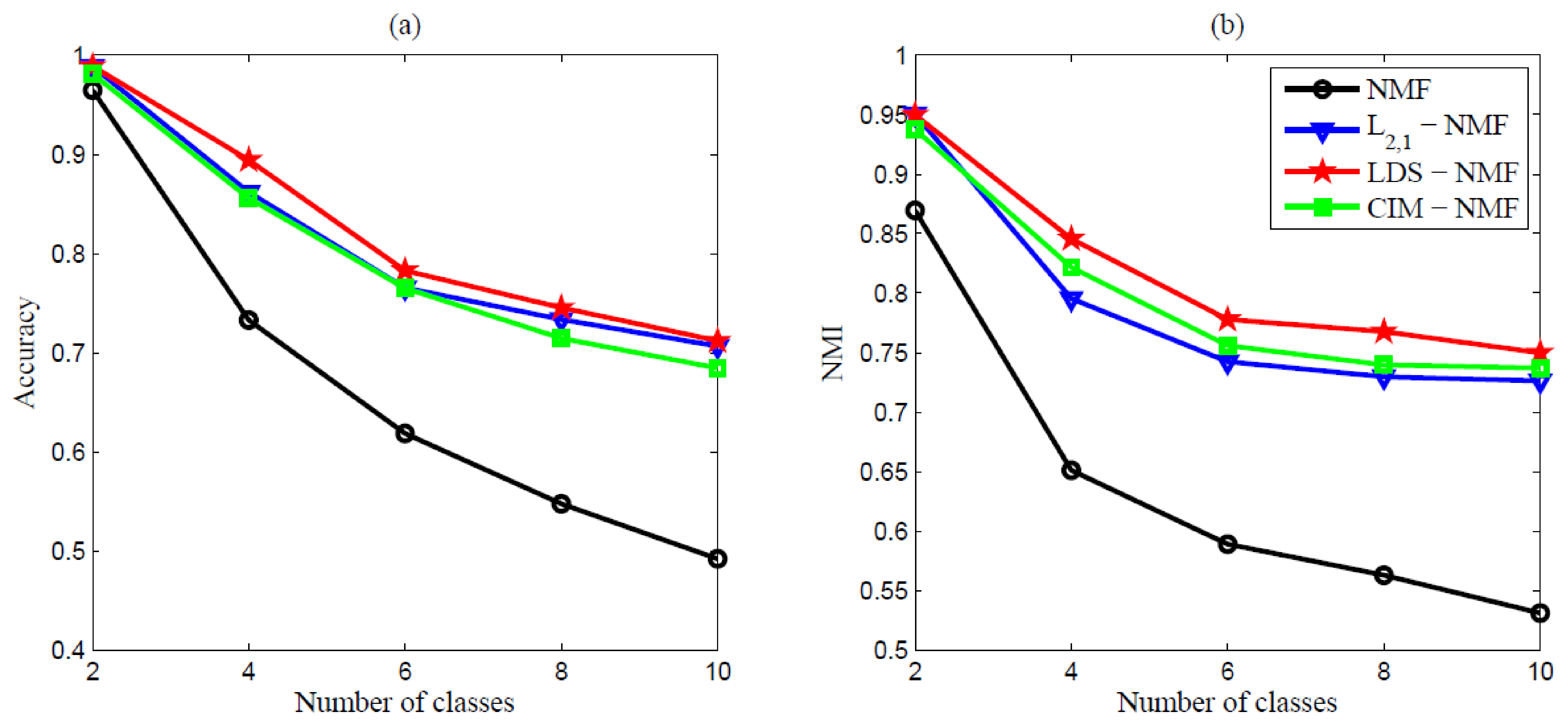
- YaleB dataset: 2,424 face images of 38 subjects
- Average accuracy vs. reduced dimensionality when (a) 5, (b) 7, and (c) 9 images of each subject are randomly selected for training on the YaleB dataset

IMAGE CLUSTERING



- COIL20 dataset: 1,440 images with the uniform black background for 20 objects
- (a) Average accuracy and (b) averaged NMI of LDS-NMF, $L_{2,1}$-NMF, CIM-NMF, and NMF on the COIL20 dataset

IMAGE CLUSTERING



- JAFFE dataset: 213 face images of 10 Japanese females
- (a) Average accuracy and (b) averaged NMI of LDS-NMF, $L_{2,1}$-NMF, CIM-NMF, and NMF on the JAFFE dataset

C O N C L U S I O N

- We proposed a Logdet divergence based sparse NMF method to solve the rank-deficiency problem of the learned lower dimensional basis
- We developed a multiplicative update rule (MUR) to optimize LDS-NMF, and the convergence of which has been proved

THANKS!

Presented by Qing Liao

qnature@ust.hk

The Hong Kong University of Science and Technology