《机器学习》课件

基于稀疏表示的人脸识别

John Wright, Allen Y. Yang,
Arvind Ganesh, S. Shankar Sastry, Yi Ma

北京航空航天大學
BEIHANG UNIVERSITY

内容的来源信息

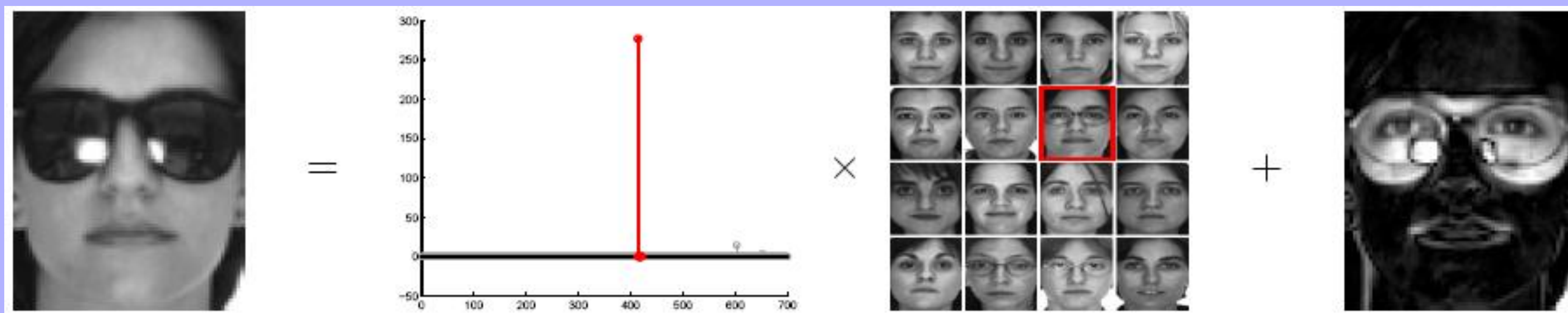
- PAMI
- 2008.3
- 相关的链接
  - http://perception.csl.uiuc.edu/recognition/Home.html
  - http://www.eecs.berkeley.edu/~yang/software/face_recognition/index.html

核心内容

- 正面人脸的识别问题，人脸图像包含光照和遮挡的变化。识别问题看作多个线性回归模型的分类问题，并证明稀疏信号表示的理论可以很好的解决这个问题。基于稀疏表示（通过L^1-Normalizaion计算），提出一种通用的基于图像的物体识别算法。
- 这个新的框架可以很好的处理人脸识别中的两个关键问题：特征提取和对遮挡的鲁棒性。
- 对于特征提取，证明：如果在识别过程中稀疏性能得到保证，那么采用何种特征就变得不重要了。
- 对于遮挡问题，稀疏表示理论可以帮助预测: (1)算法能处理多大的遮挡; (2)如何选择训练样本来最大化对遮挡的鲁棒性。

Overview

测试图像　稀疏表示　训练图像　重构误差



- 本文的主要思路
  - 测试图像可以由训练图像的线性组合来重构（允许一定重构误差），重构系数是稀疏的。
  - 通过这种稀疏表示，可以确定测试图像的类别：最大重构系数所对应的类别。
- 三个问题
  - 为什么能重构？
  - 为什么采用稀疏表示？
  - 如何保证重构系数的稀疏性？

为什么能重构？

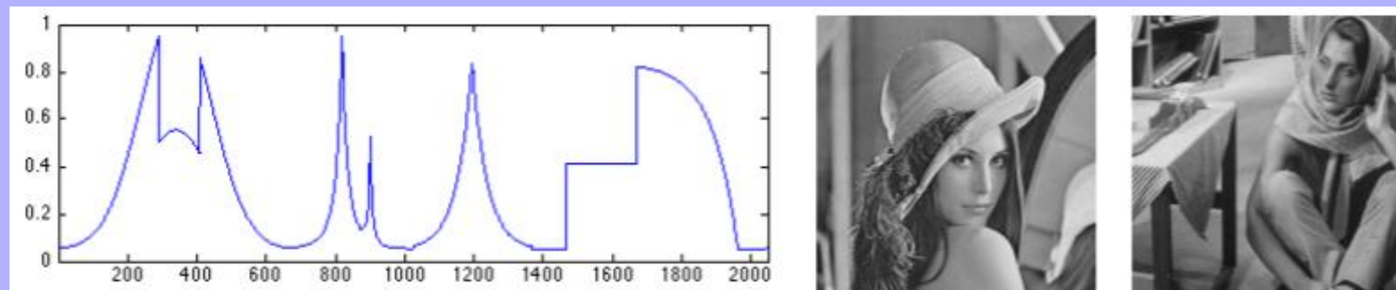
- 线性子空间假设
  - “for ease of presentation, assume that the training samples from a single class do lie on a subspace”
  - “If sufficient training samples are available, from each class, it will be possible to represent the test samples as a linear combination of just those training samples from the same class.”
  - 所作的唯一假设。

为什么采用稀疏表示？

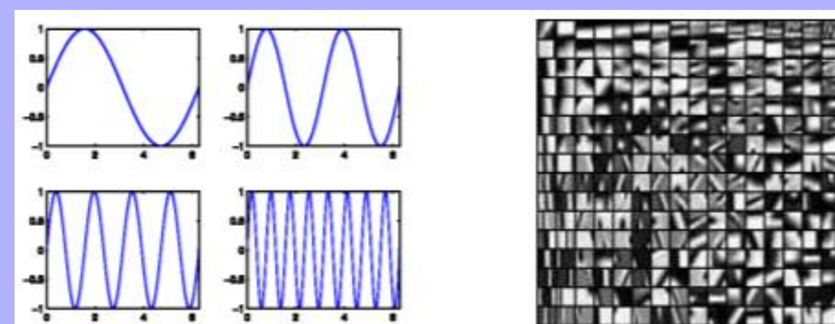
- Principle of Parsimony (节俭的)
  - also known as Occam's Razor: simple is better
  - SVM：只有支持向量被用于构建分类面。
  - human perception: 同一时刻，被激活的视觉细胞只占其总量的很小一部分。
- 如果线性子空间假设成立，那么重构系数的稀疏性是很显然的。

稀疏表示

- Let x in $\mathbb{R}^m$ be a signal



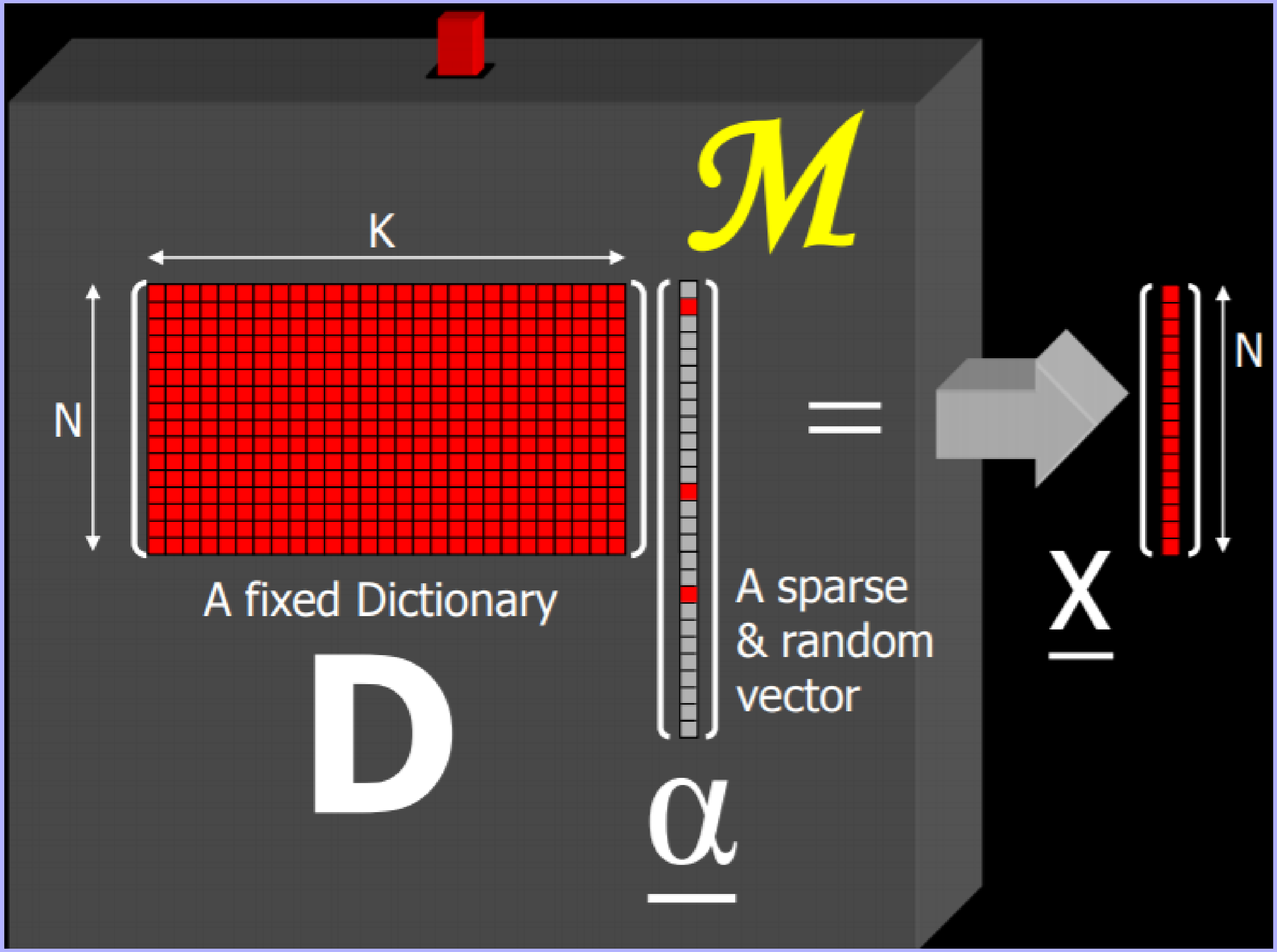
- Let $D = [d_1, \dots, d_p] \in \mathbb{R}^{m\times p}$ be a set of normalized “basis vectors”. We call it **dictionary**.



- D is “adapted” to x vectors--that is, the that $x \approx D\alpha$. We ca

$$\underbrace{\begin{pmatrix} \mathbf{x} \end{pmatrix}}_{\mathbf{x}\in\mathbb{R}^m} \approx \underbrace{\begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{pmatrix}}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix}\alpha[1]\\ \alpha[2]\\ \vdots \\ \alpha[p]\end{pmatrix}}_{\alpha\in\mathbb{R}^p,\ \text{sparse}}$$

稀疏表示



Every column in D(**dictionary**) is a prototype signal(**Atom**).

The vector α contains very few non-zeros.

问题的形式化

- 第i类样本 $A_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in \Re^{m \times n_i}$
  - n_i:第i类样本数目，m:样本维数
- 整个训练集 $A = [A_1, A_2, \ldots A_k] = [v_{1,1}, v_{1,2}, \ldots v_{k,n_k}]$
- 根据线性子空间假设，如果第i类样本足够多，那么来自此类的测试样本y可以写成

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \ldots + \alpha_{i,n_i} v_{i,n_i}$$

- 扩展到整个训练集上

$$y = A x_0 \in \Re^m$$

$$x_0 = [0, \ldots 0, \alpha_{i,1}, \alpha_{i,2}, \ldots \alpha_{i,n_i}, 0, \ldots 0]^T \in \Re^n$$

问题的求解

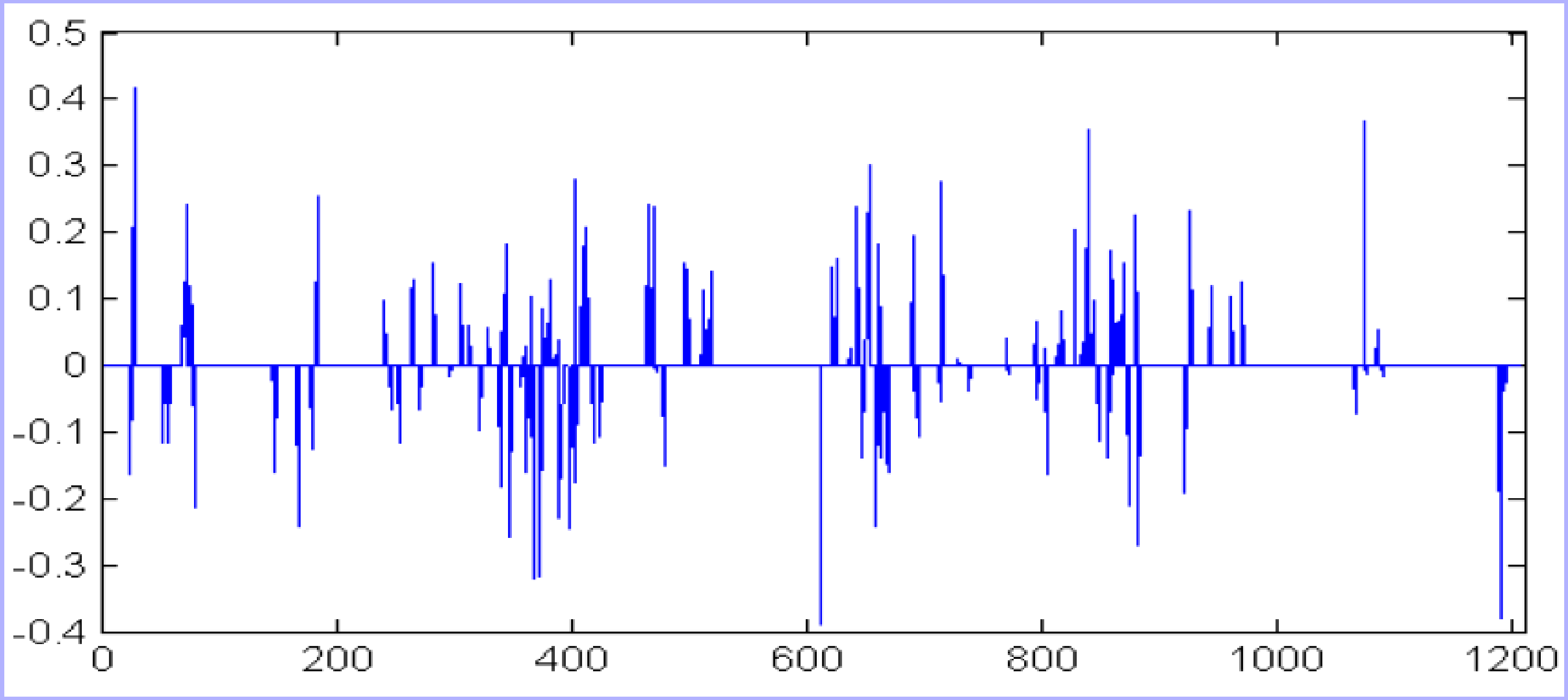
- 解线性方程组 $y_m = A_{m\times n} x_{n\times 1}$
  - 如果m>n, 方程组是overdetermined的, x有唯一解。
  - 如果m<n, 方程组是underdetermined的，x有多个解，一般采用minimum L^2-norm solution。

$$(L^2):\ \hat{x}_2 = \arg\min \|x\|_2 \ \text{ subject to } \ Ax = y$$

可以通过A的伪逆矩阵A^+来求解：$x = A^+ y$

L^2-Minimization的结果

- 没有稀疏性



L^0-Minimization

$$(L^0): \quad \hat{x}_0 = \arg\min \|x\|_0 \ \text{ subject to } \ Ax = y$$

- $\|x\|_0$：X中非零元素的个数。
- 当x中的非零元素数小于$m/2$时，x就是最稀疏的唯一解$\hat{x}_0 = x$（参见文献[33]）。
- 但是，这是个NP-Hard问题，目前没有方法能比对x穷举快很多，哪怕只是近似求解。

L^1-Minimization

$$(L^1):\quad \hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad Ax = y$$

- 如果x_0足够稀疏，那么L^0-Minimization的解和L^1- Minimization的解相同 (参见文献[9,10,11]) 。
- L^1-Minimization的问题可以通过线性规划算法在多项式时间内求解(参见文献[34])。

噪声问题

- 容忍一定范围内的重构误差

$$y = Ax_0 + z, \quad \|z\| < \varepsilon$$

$$(L_s^1): \quad \hat{x}_1 = \arg\min \|x\|_1 \ \text{ subject to } \ \|Ax - y\| \le \varepsilon$$

- 这是一个convex optimization问题，可以用second-order cone programming来求解。
- 参考文献[38]中证明(L_s^1)问题的解是稀疏的。

基于稀疏表示的分类方法

Algorithm 1 : Sparse Representation-based Classification (SRC)

1: **Input:** a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $\boldsymbol{y} \in \mathbb{R}^m$, (and an optional error tolerance $\varepsilon > 0$.)

2: Normalize the columns of A to have unit ℓ^2-norm.

3: Solve the ℓ^1-minimization problem:

$$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{y}. \qquad (13)$$

(Or alternatively, solve

$$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \|A\boldsymbol{x} - \boldsymbol{y}\|_2 \le \varepsilon.)$$

4: Compute the residuals $\boxed{r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\,\delta_i(\hat{\boldsymbol{x}}_1)\|_2}$ for $i = 1, \ldots, k$.

5: **Output:** $\text{identity}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$.

基于稀疏表示的分类方法

1: **Input:** a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $\boldsymbol{y} \in \mathbb{R}^m$, (and an optional error tolerance $\varepsilon > 0$.)

2: Normalize the columns of A to have unit ℓ^2-norm.

3: Solve the ℓ^1-minimization problem:

$$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{y}. \tag{13}$$

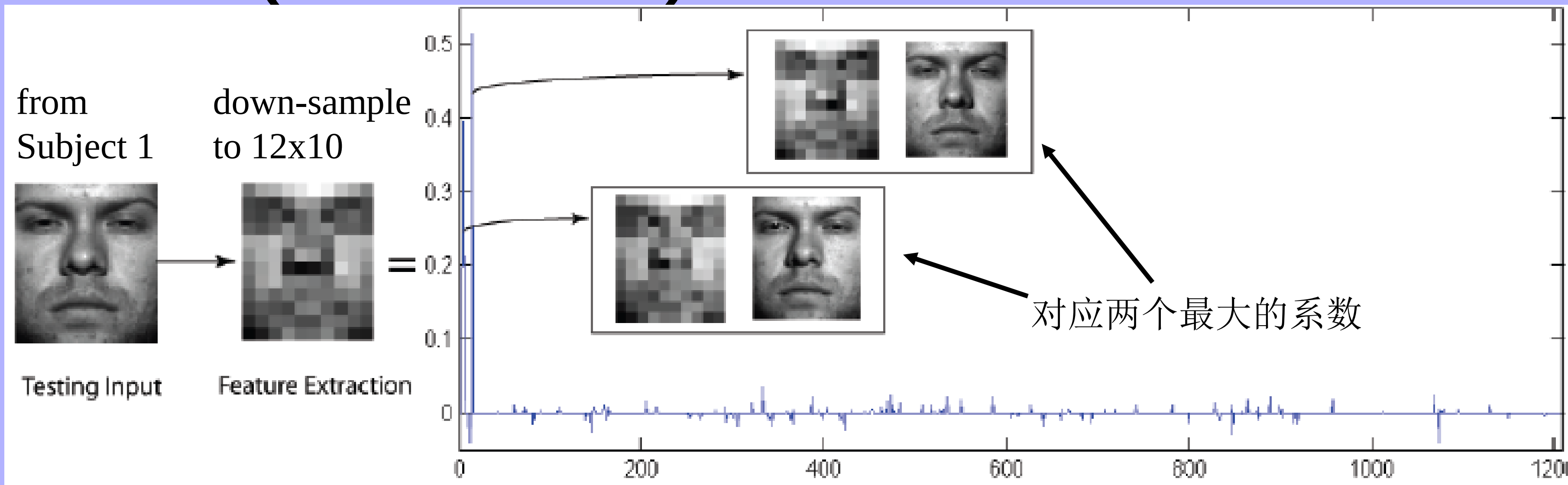
(Or alternatively, solve

$$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \|A\boldsymbol{x} - \boldsymbol{y}\|_2 \le \varepsilon.)$$

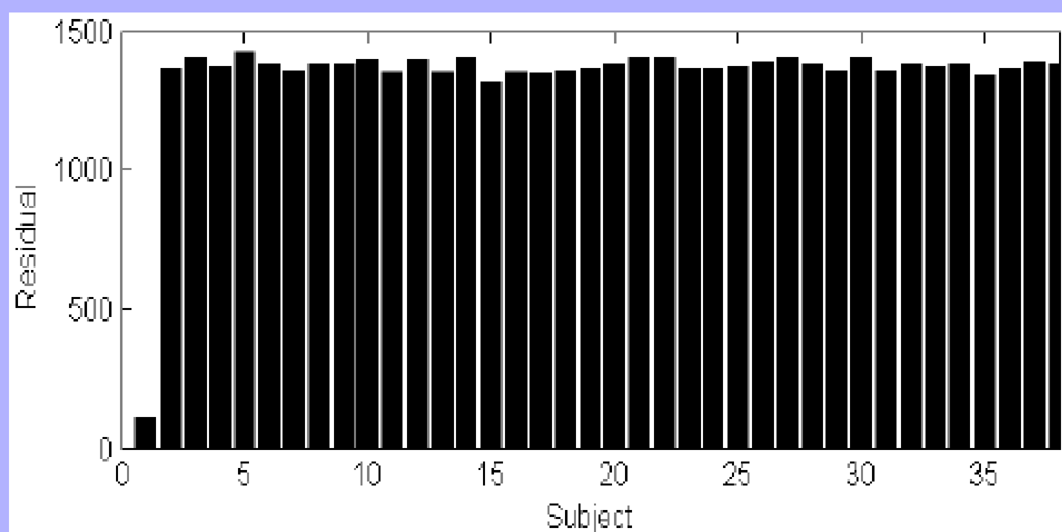
4: Compute the residuals $r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\,\delta_i(\hat{\boldsymbol{x}}_1)\|_2$ for $i = 1, \ldots, k$.

5: **Output:** identity$(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$.

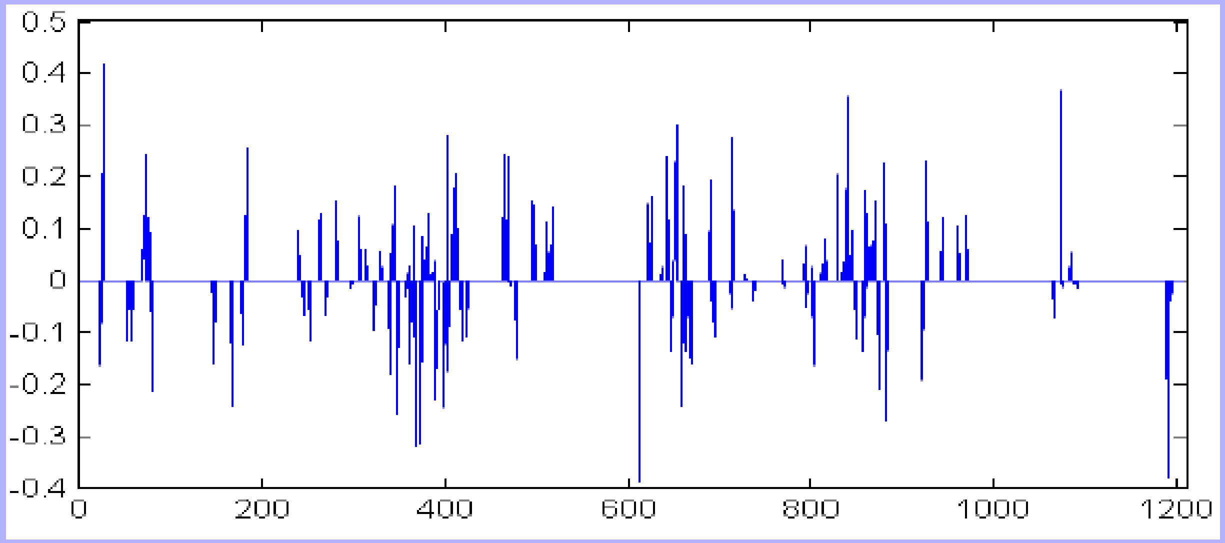
例子 (L^1-Norm)



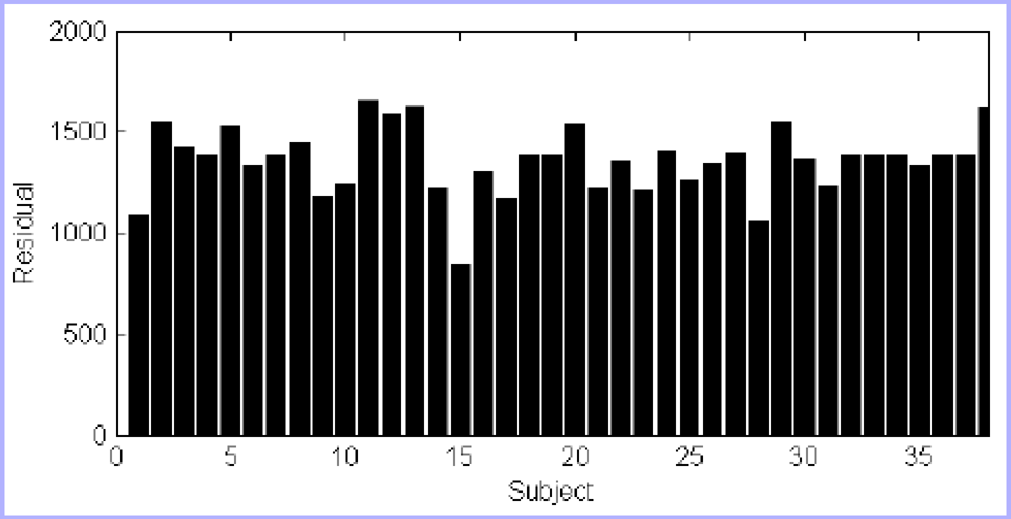
Residuals



例子 (L^2-Norm)



0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
200
400
600
800
1000
1200

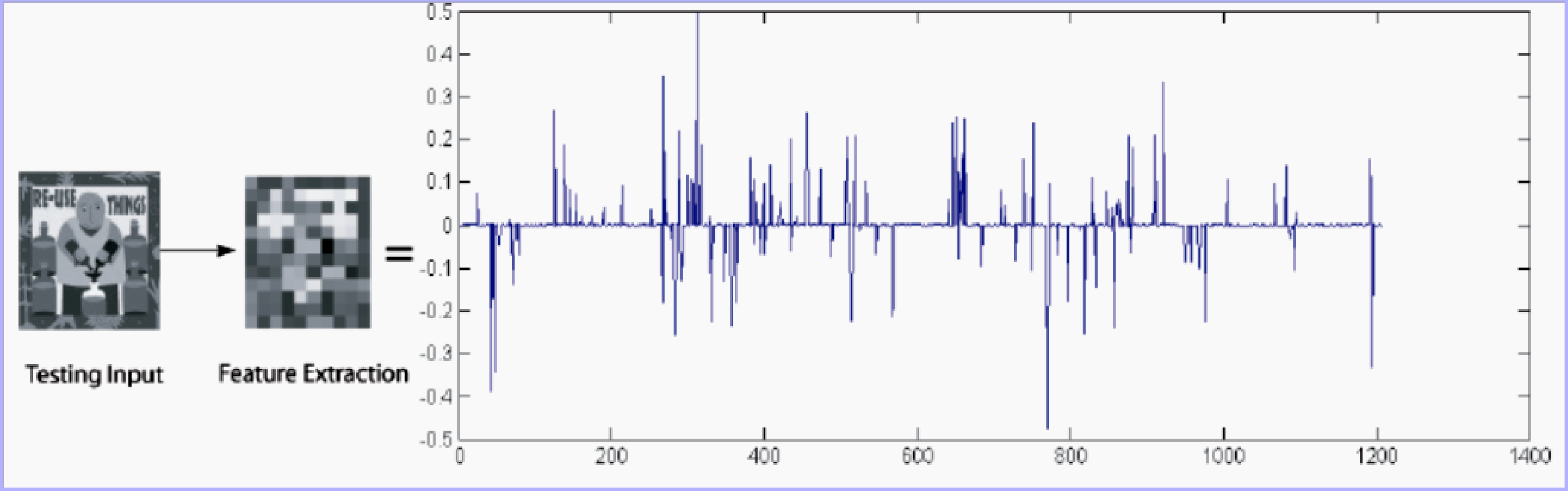


2000
1500
1000
500
0
Residual
0
5
10
15
20
25
30
35
Subject

Invalid Test Image

- 根据系数的稀疏程度可以判断测试图像是否属于训练集中的某一类别。

稀疏性度量 $$\mathrm{SCI}(x) = \frac{k \cdot \max_i \|\delta_i(x)\|_1 / \|x\|_1 - 1}{k - 1}$$



人脸识别中的两个问题

- 特征提取
- 对遮挡和噪声的鲁棒性

特征提取

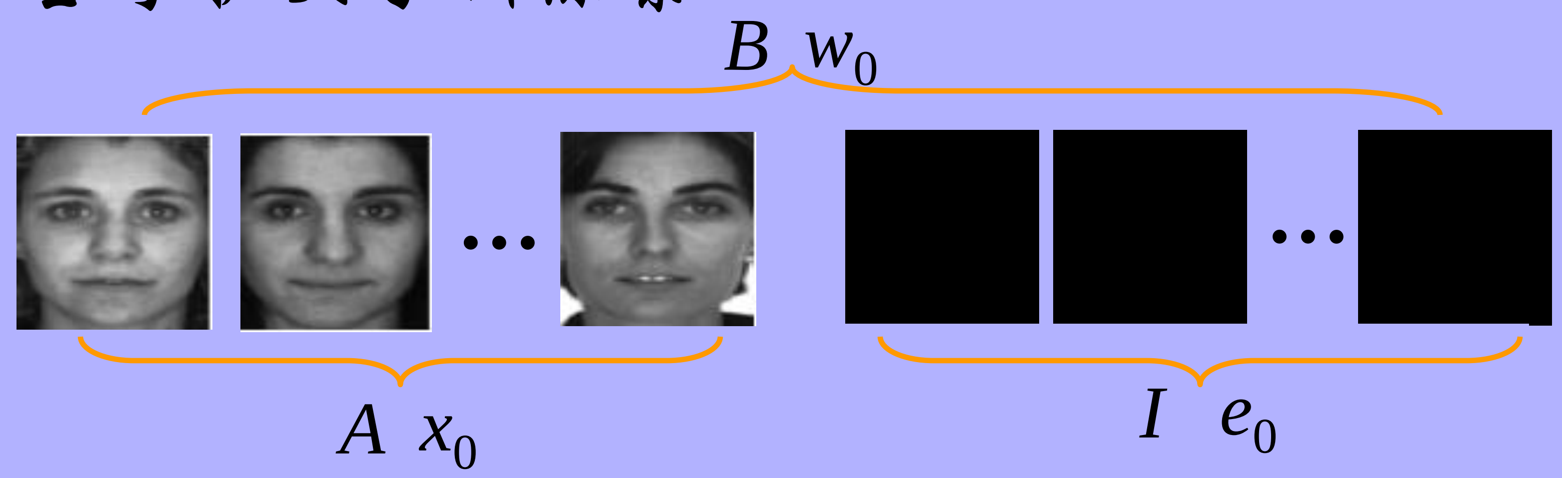
- 提取何种特征才能得到正确的稀疏系数?
  - 如果x包含t个非零值(t<<n)，那么d维特征就可以保证通过L^1-Minimiazaion得到正确的稀疏系数，其中 $d \geq 2t\log(n/d)$。
  - 因此只要维数大于d，任何特征的分类正确率都基本一致，甚至包括最简单的随机特征和下采样特征。
  - 这种现象称作“blessing of dimensionality”。

对遮挡和噪声的鲁棒性

- 很多情况下，人脸图像包含遮挡和噪声（occlusion and corruption）。
- 此时，问题变成：

$$y = Ax_0 + e_0 \Rightarrow y = [A, I]\begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = Bw_0$$

- 相当于扩大了训练集



Extended L^1-Minimization

$$(L_e^1): \quad \hat{w}_1 = \arg\min \|w\|_1 \ \text{ subject to } Bw = y$$

- 遮挡小于图像面积的33%时，仍然可以得到正确的稀疏系数，因此也就可以正确的重构图像(参见文中证明)。
- Residual需要重新定义

$$r_i(y) = \|y_r - A\delta_i(\hat{x}_1)\|_2 = \|y - \hat{e}_1 - A\delta_i(\hat{x}_1)\|_2$$

SRC vs. NN、NS

- SRC
  - Sparse Representation-based Classification
- NN、NS
  - NN：Nearest Neighbor
    - 测试样本的类别为距离其最近的训练样本的类别
  - NS：Nearest Subspace
    - 测试样本的类别为距离其最近的子空间的类别

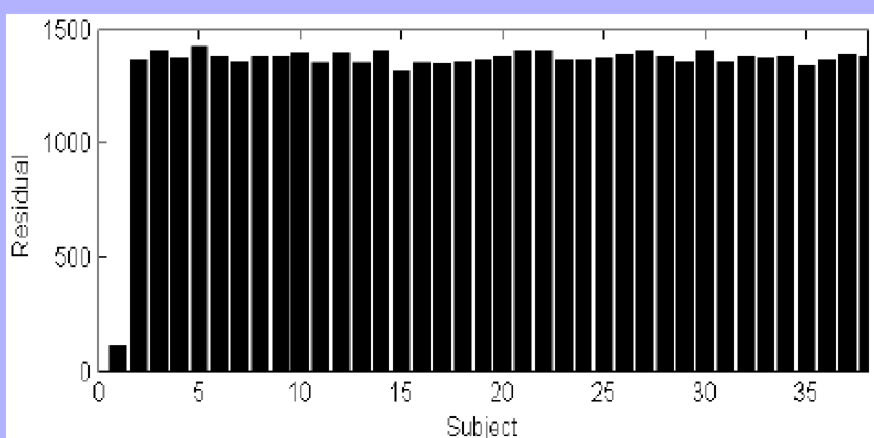
SRC vs. NN、NS

- SRC vs. NN
  - 当训练样本足够稠密，以至于测试样本可以用一个训练样本表示时，SRC等同于NN。
  - 实际上，训练集不可能包含所有变化。

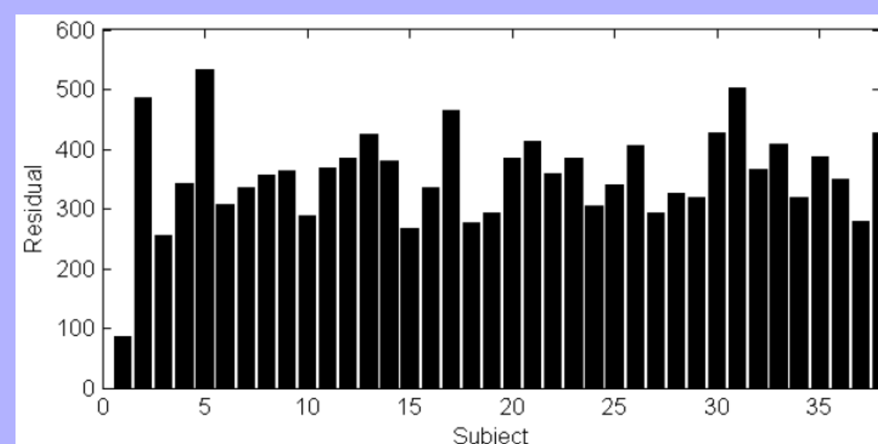
SRC vs. NN、NS

- SRC vs. NS
  - 当测试样本可以用训练集中某类样本的线性组合表示时，SRC等同与NS。
  - 但是，SRC比NS更具有区分性。

SRC

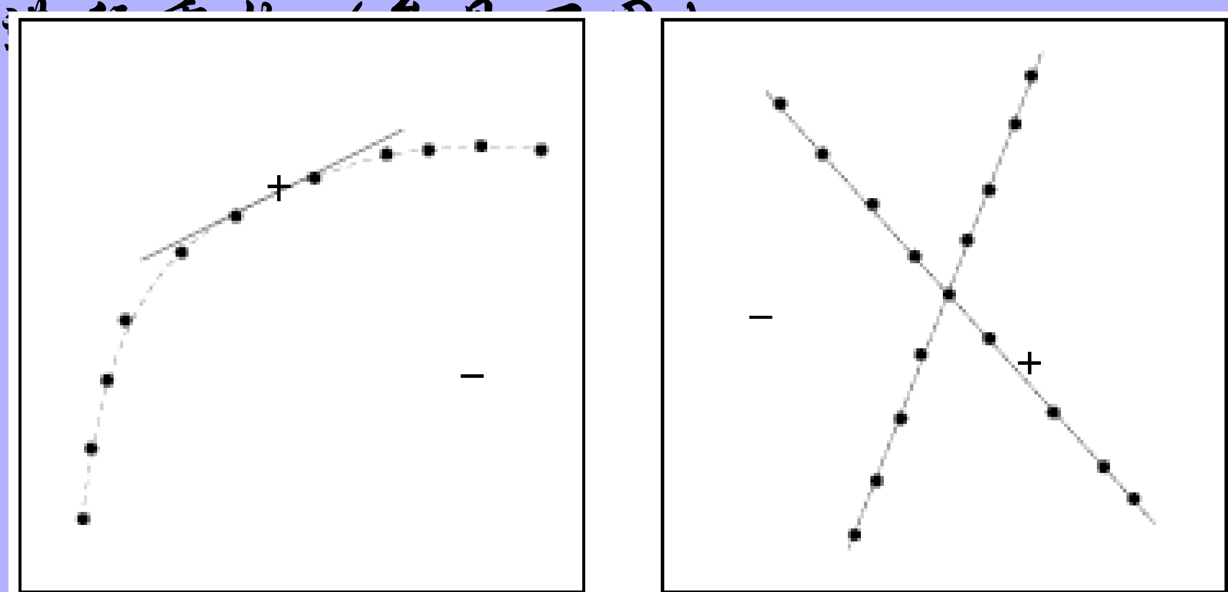


NS



SRC vs. NN、NS

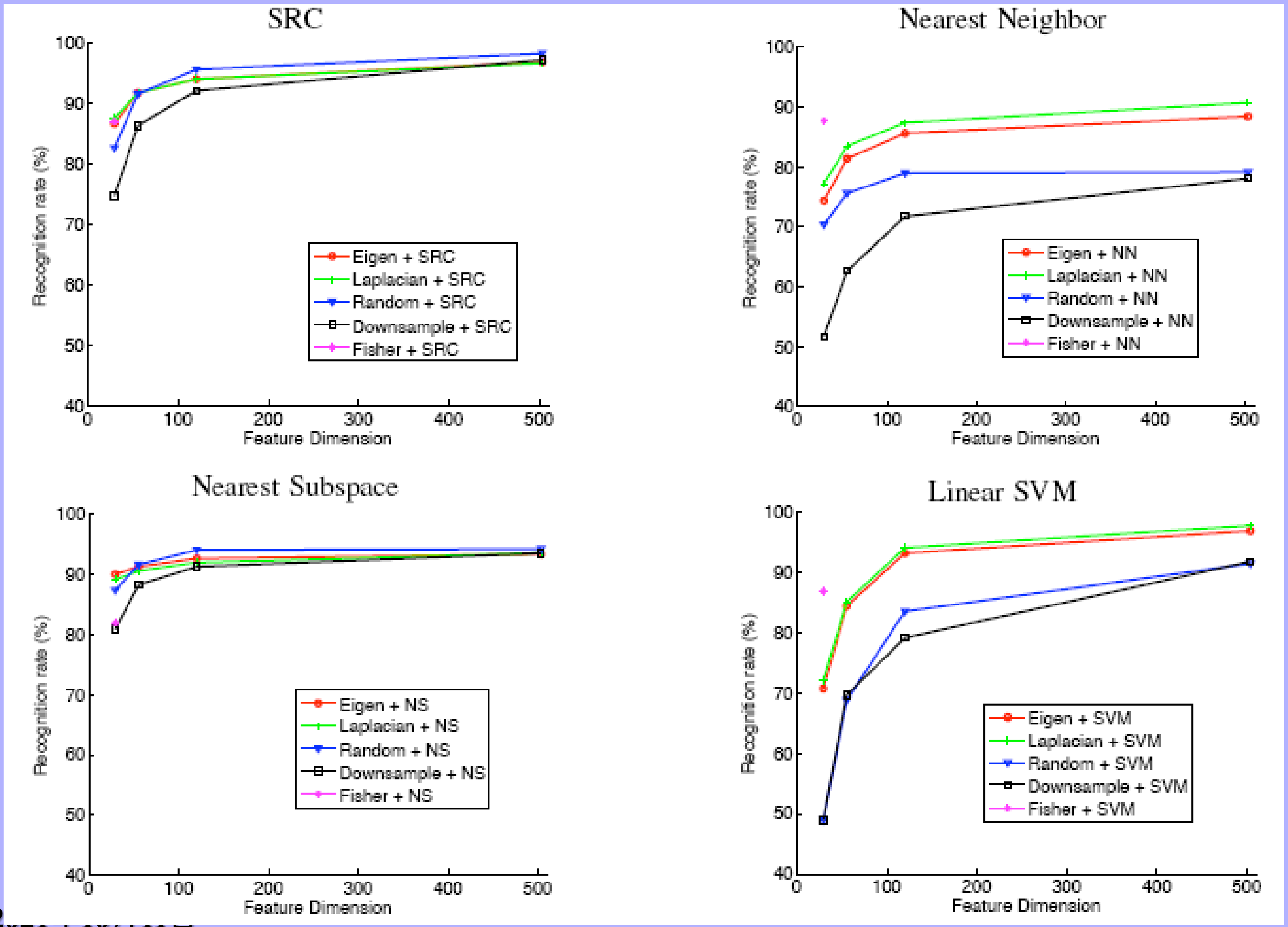
- SRC vs. NS
  - 用某一类别的所有样本来重构测试样本可能会产生overfitting，而SRC可以只用其中的部分样本来



实验

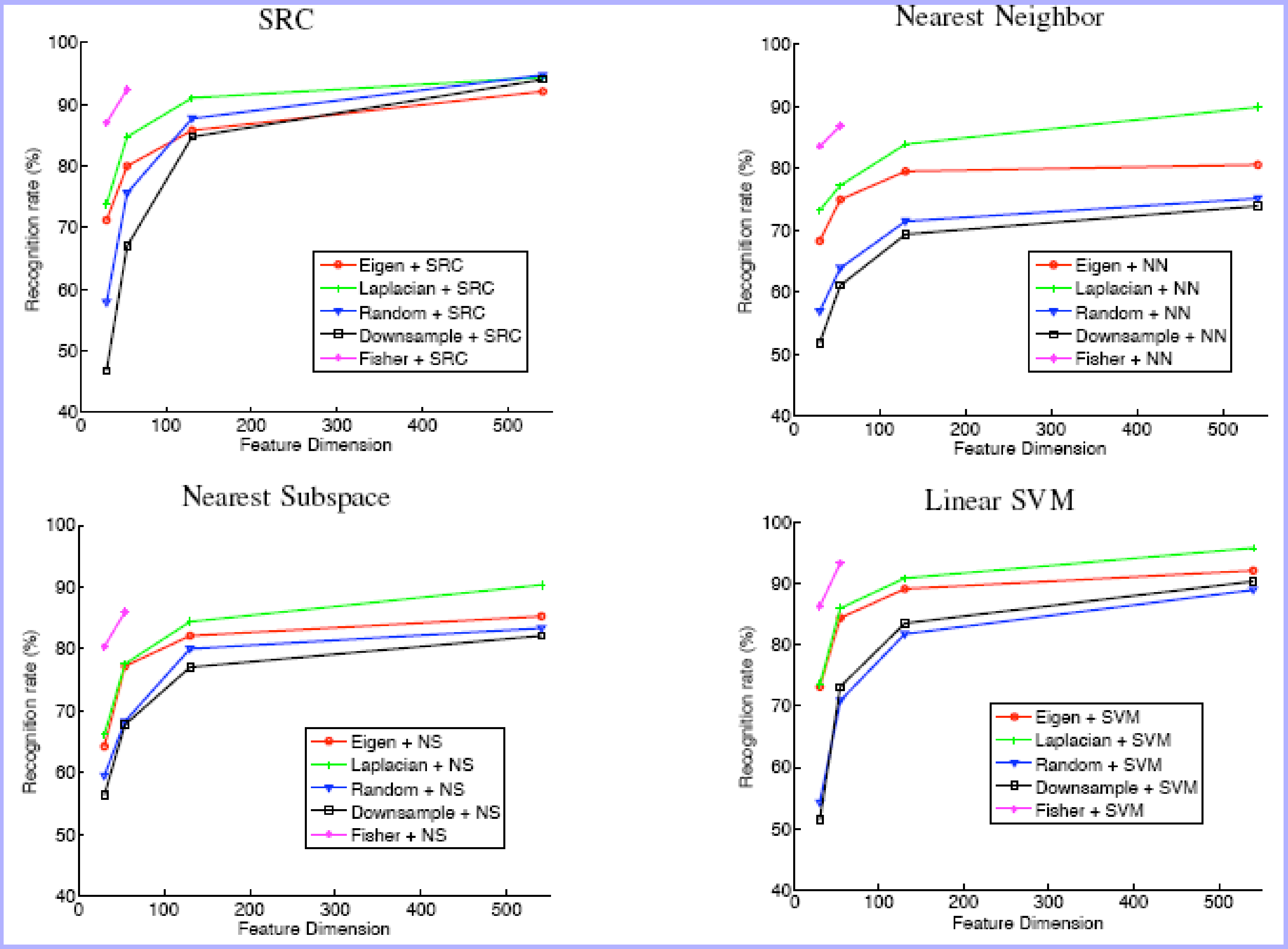
- 不同的特征
  - Eigen、Laplacian、Random、Downsample、Fisher
- 遮挡和噪声
- 有拒识的情况
- 训练集的选取

不同特征-Yale B



SRC
Nearest Neighbor
Nearest Subspace
Linear SVM
Recognition rate (%)
Feature Dimension
Eigen + SRC
Laplacian + SRC
Random + SRC
Downsample + SRC
Fisher + SRC
Eigen + NN
Laplacian + NN
Random + NN
Downsample + NN
Fisher + NN
Eigen + NS
Laplacian + NS
Random + NS
Downsample + NS
Fisher + NS
Eigen + SVM
Laplacian + SVM
Random + SVM
Downsample + SVM
Fisher + SVM

不同特征-AR


SRC
Nearest Neighbor
Nearest Subspace
Linear SVM
Recognition rate (%)
Feature Dimension
Eigen + SRC
Laplacian + SRC
Random + SRC
Downsample + SRC
Fisher + SRC
Eigen + NN
Laplacian + NN
Random + NN
Downsample + NN
Fisher + NN
Eigen + NS
Laplacian + NS
Random + NS
Downsample + NS
Fisher + NS
Eigen + SVM
Laplacian + SVM
Random + SVM
Downsample + SVM
Fisher + SVM

结论

- SRC和SVM的性能优于NN和NS。
- SVM、NN和NS的性能依赖与选择何种特征，并且性能不随着特征数增加而收敛。
- 理论上给出的可以恢复稀疏系数的最小特征数分别是：YaleB (128维)、AR (88维)。与实验结果基本一致。

不同人脸区域的特征

- 1207个训练样本，因此线性方程组是overdetermined (d>n)
- 证明了SRC在处理高维数据时的稳定性。

Features	Nose	Right Eye	Mouth & Chin
Dimension (d)	4,270	5,040	12,936
SRC	**87.3%**	**93.7%**	**98.3%**
NN	49.2%	68.8%	72.7%
NS	83.7%	78.6%	94.4%
SVM	70.8%	85.8%	95.3%

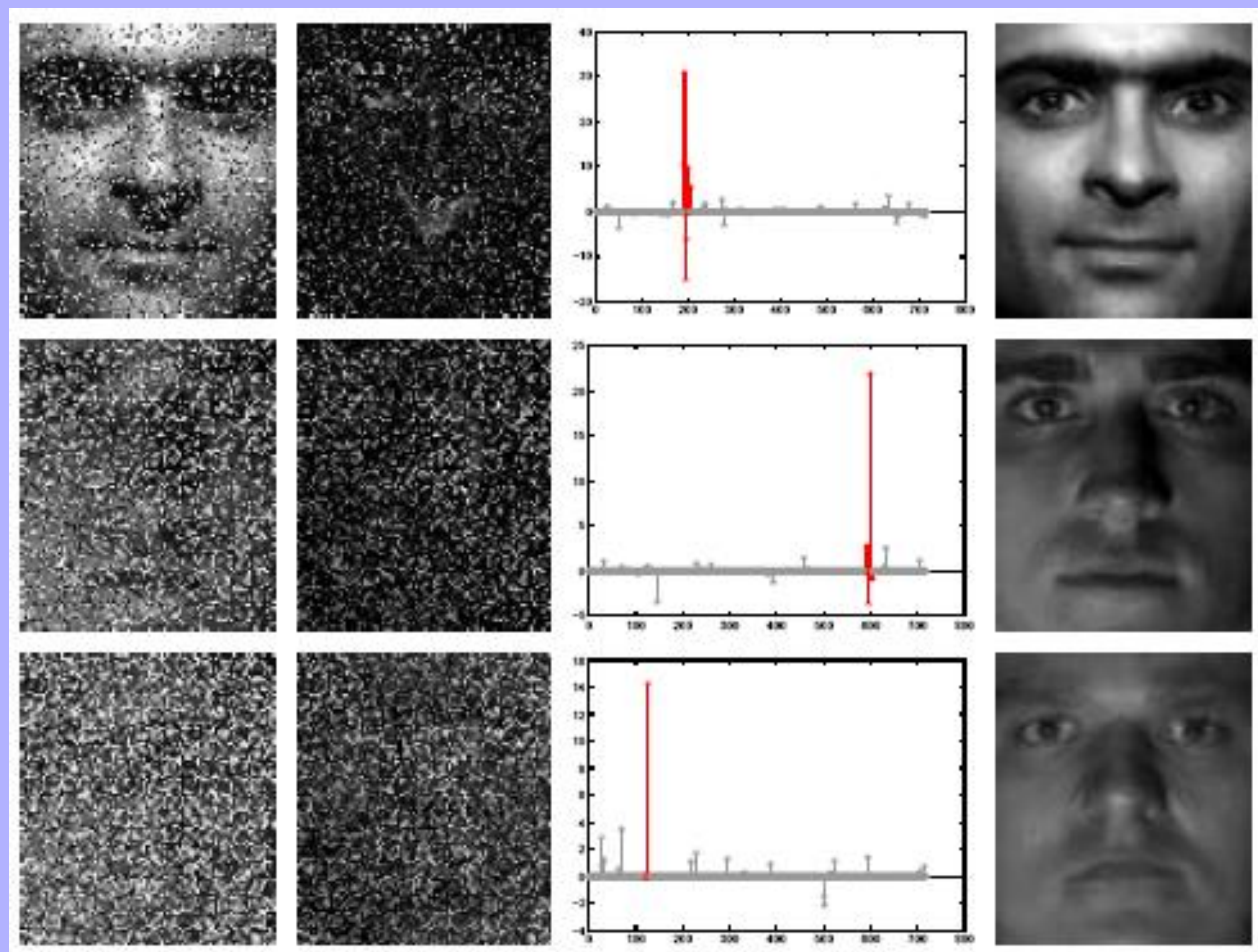
有噪声的情况

测试图像　重构误差　重构系数　重构图像

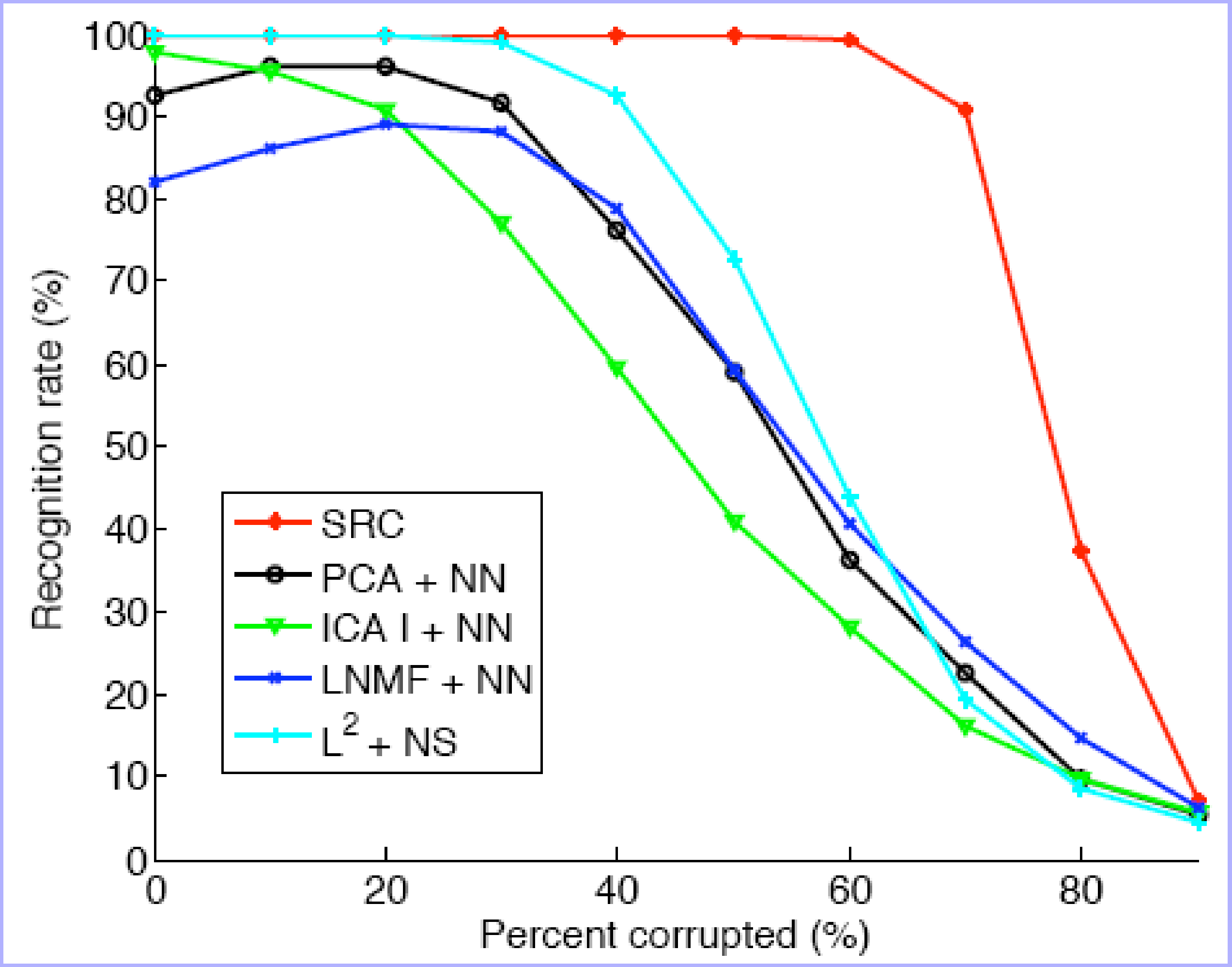
30% corrupted

50% corrupted

70% corrupted



有噪声的情况



Recognition rate (%)
Percent corrupted (%)
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
SRC
PCA + NN
ICA I + NN
LNMF + NN
L2 + NS

有遮挡的情况

测试图像　重构误差　重构系数　重构图像

30% occluded

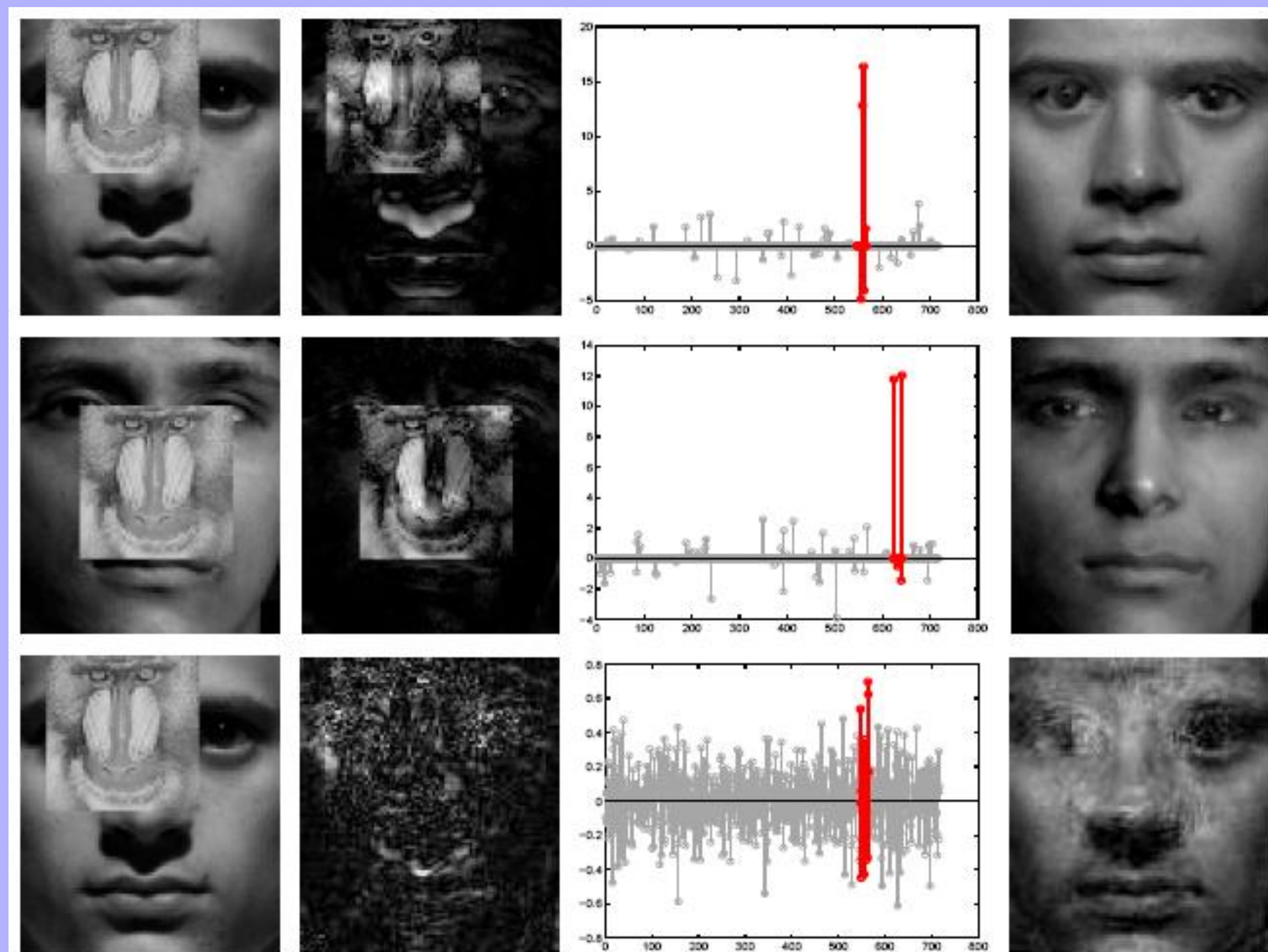
L^1-Minimization

30% occluded

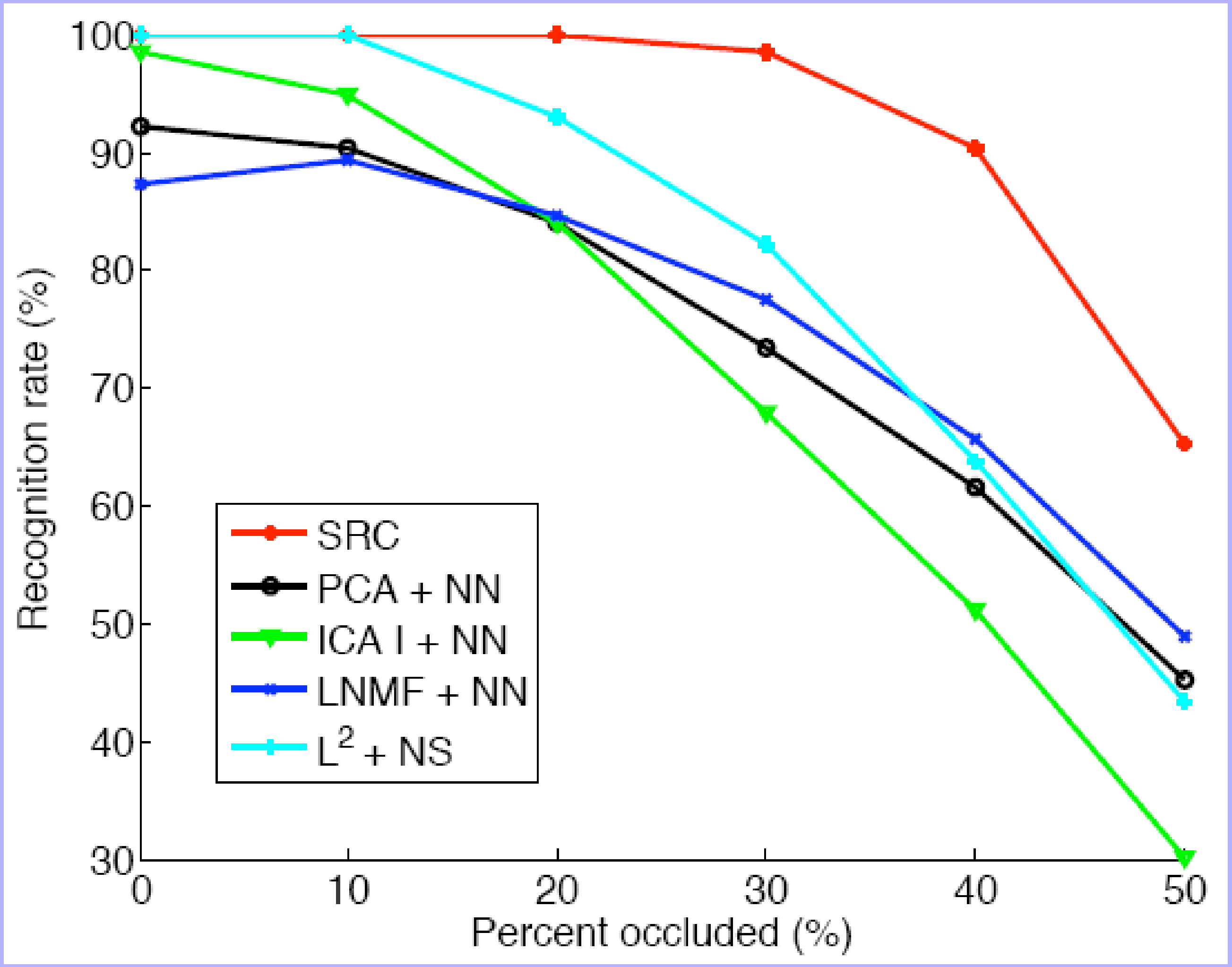
L^1-Minimization

30% occluded

L^2-Minimization



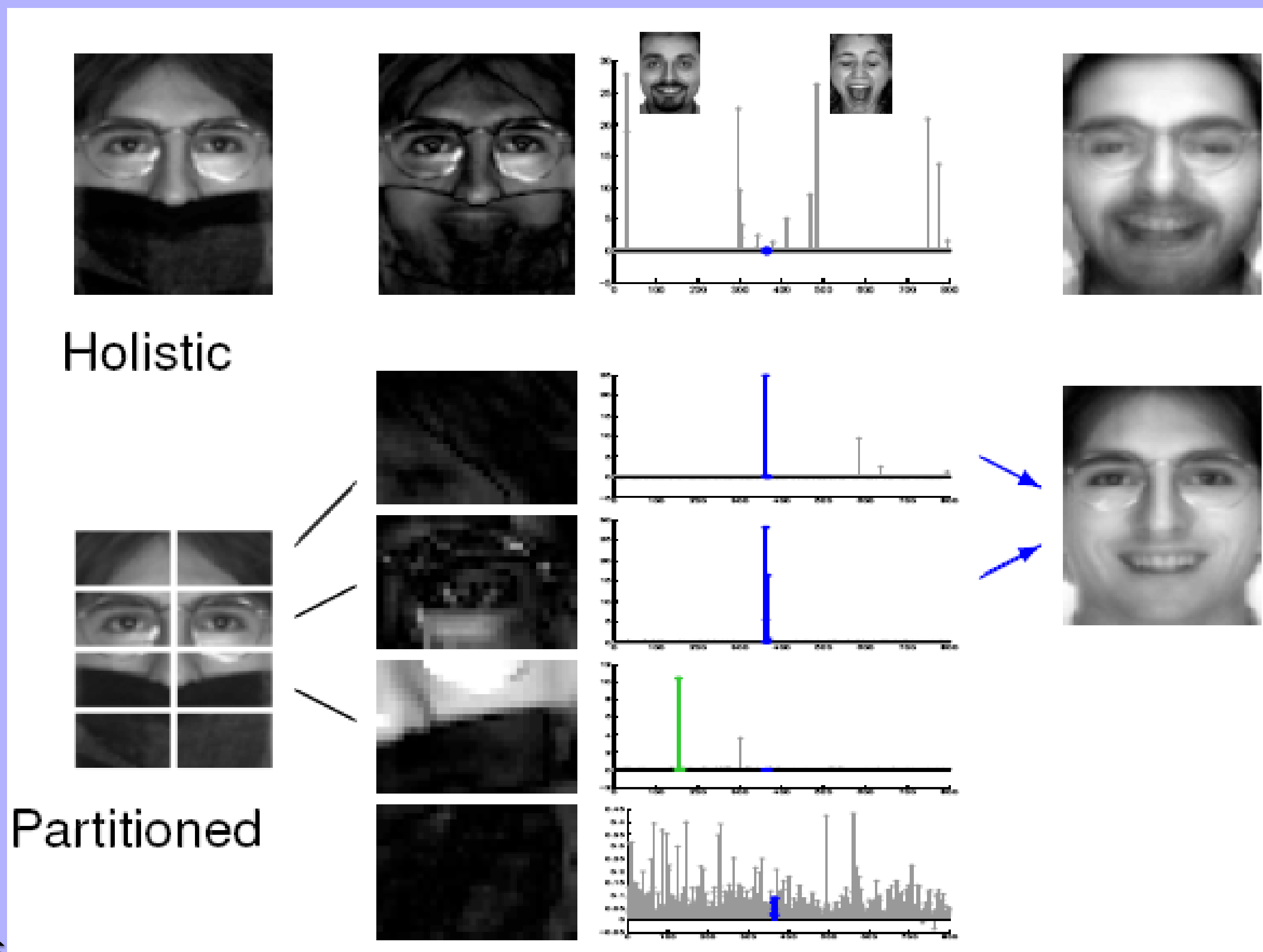
有遮挡的情况



2021年10月11日

有遮挡的情况

测试图像　　重构误差　　重构系数　　重构图像

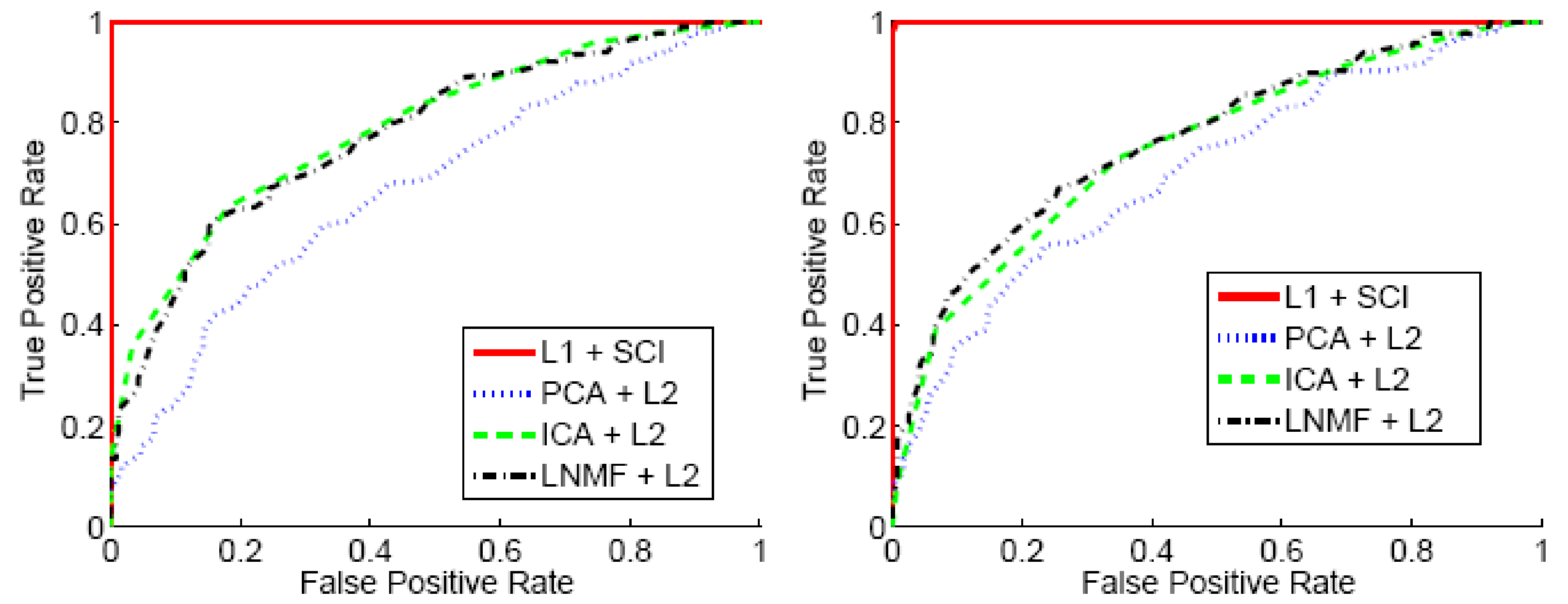


有遮挡的情况

Algorithms	Rec. rate sunglasses	Rec. rate scarves
SRC (partitioned)	**87.0%** **(97.5%)**	**59.5%** **(93.5%)**
PCA + NN	70.0%	12.0%
ICA I + NN	53.5%	15.0%
LNMF + NN	33.5%	24.0%
ℓ^2 + NS	64.5%	12.5%

有拒识的情况

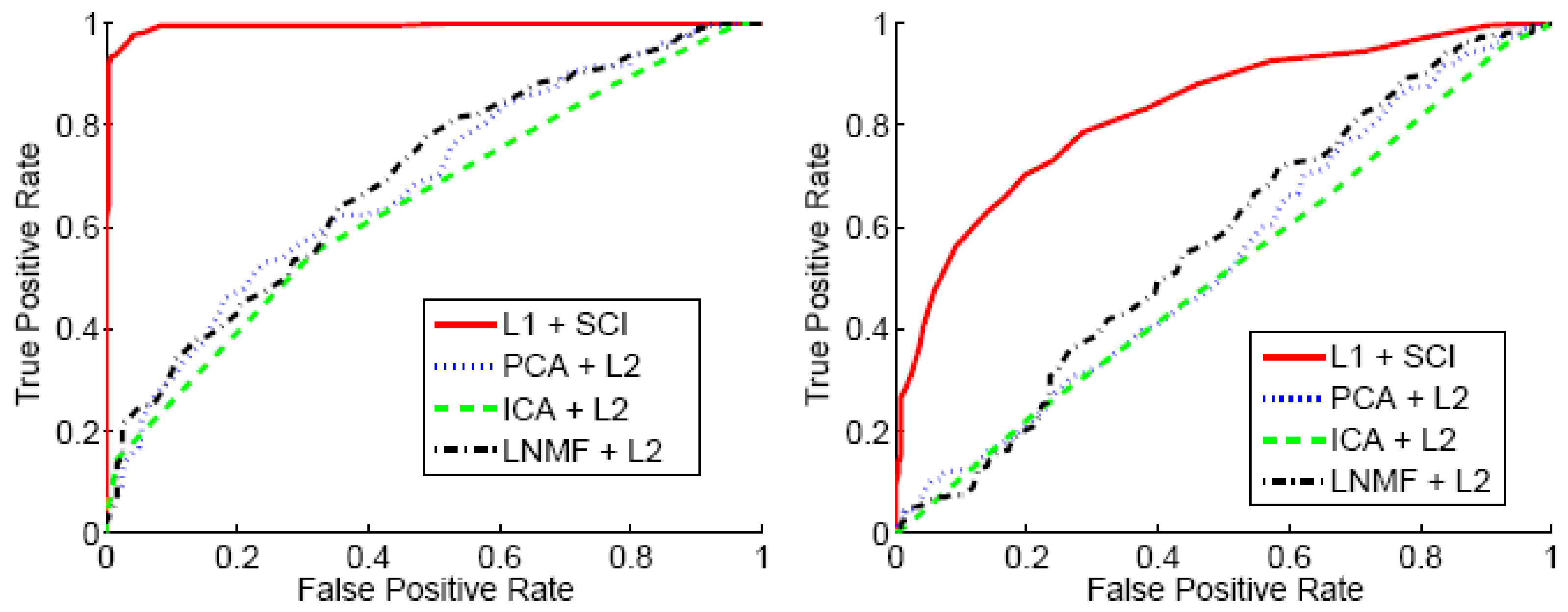
- 测试集中只有50%的人在训练集中出现



(a) No occlusion (b) 10% occlusion

有拒识的情况

- 测试集中只有50%的人在训练集中出现



(c) 30% occlusion (d) 50% occlusion