南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

模式计算与应用实验室
Pattern Computing & Applications Lab

Matrix Regression and Its Applications to Robust Face Recognition

Jian Yang
School of Computer Science and Engineering
Nanjing University of Science and Technology

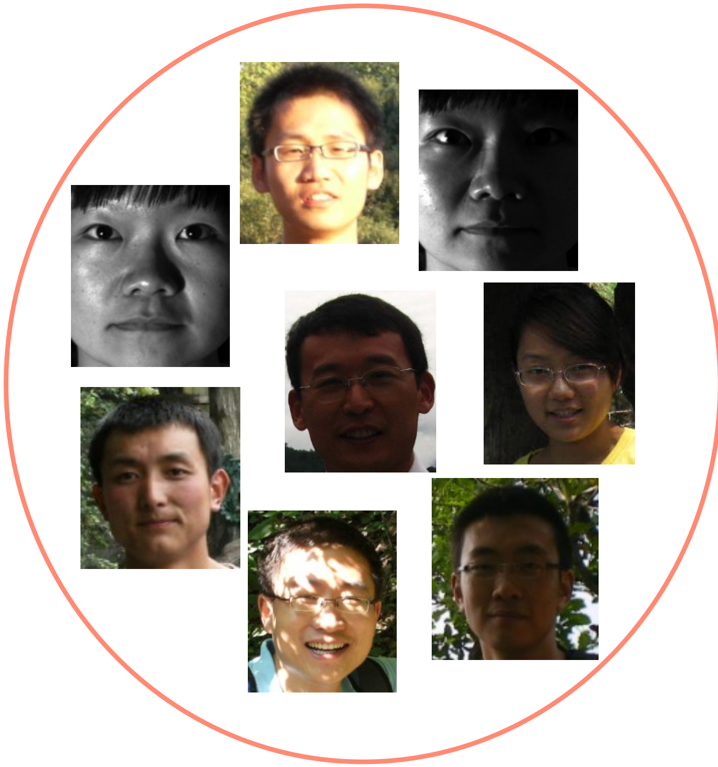


Outline

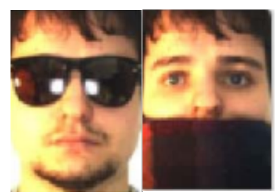
- Background and Overview
- Nuclear Norm based Matrix Regression
- Extended Versions

Background: Challenges in face recognition

illuminations



Illumination & occlusion



occlusions

Linear Regression

Linear regression model:

$$\min_{x} \|Y - \mathrm{D}x\|_2^2$$

- Closed-form solution

$$x = (\mathrm{D}^T \mathrm{D})^{-1} \mathrm{D}^T \mathrm{Y}$$

- LR based Classification

  Y is a testing sample

  D is the dictionary (representation basis), which is formed by a given set of samples with class labels

Two Representation Schemes

Class sample based representation (LRC, PAMI 2010)

- Given a test sample, using the samples of each class to represent it.
- Using the class representation error to design a classification rule.

Population sample based representation (CRC, ICCV 2011)

- Given a test sample, using the samples of all classes to represent it.
- Using the class representation error (or representation coefficients) to design a classification rule

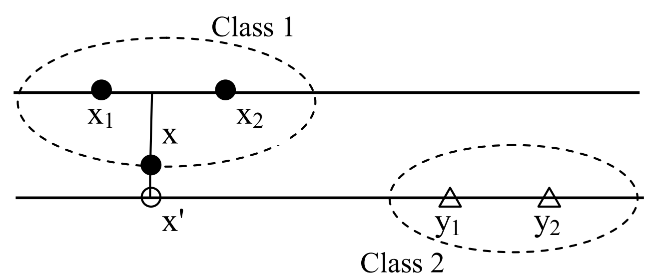
- I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, *IEEE Trans. PAMI*, 2010, 32(11): 2106- 2112.
- L. Zhang, M. Yang, and X. C. Feng. Sparse representation or collaborative representation which helps face recognition? In *ICCV*, 2011.

Regularized Linear Regression

- L2-norm Regularized Linear Regression (Ridge Regression)

$$\min_{x} \|Y - \mathrm{D}x\|_2^2 + \lambda\|x\|_2^2$$

- Avoid overfitting



- Close-form solution

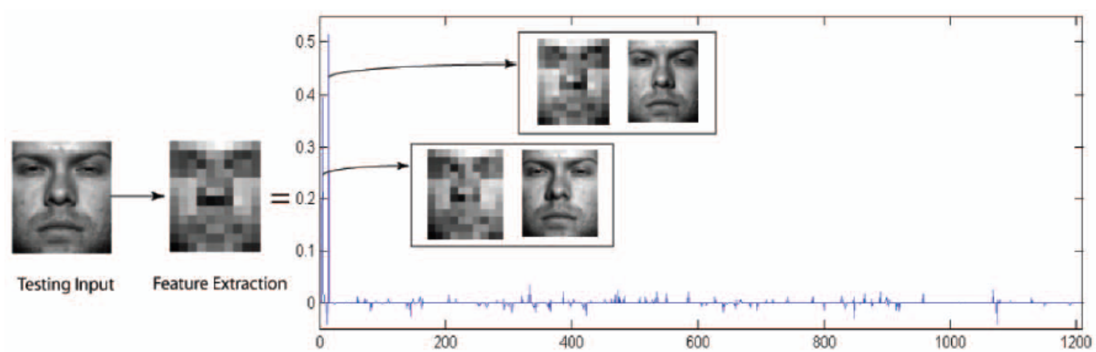
$$x = (\mathrm{D}^T \mathrm{D} + \lambda \mathbf{I})^{-1} \mathrm{D}^T \mathrm{Y}$$

Regularized Linear Regression

L1-norm Regularized Linear Regression
(Sparse Representation)

$$\min_{x} \|Y - Dx\|_2^2 + \lambda \|x\|_1$$

- No close-form solution (many algorithms for solving it)
- Sparse Representation Classifier (SRC)



The Role of Sparsity

Small dictionary Case

- Sparsity is **unnecessary** (L2 does as well as L1)

Big dictionary (over-complete) Case

- Sparsity is **necessary** for **Learning locality**

L. Zhang, M. Yang, and X. C. Feng. Sparse representation or collaborative representation which helps face recognition? In *ICCV*, 2011.
Q. Shi, A. Eriksson, A. Hengel, C. Shen. Is face recognition really a compressive sensing problem? In *CVPR* 2011.
R. Rigamonti, M. Brown and V. Lepetit. Are Sparse Representations Really Relevant for Image Classification? In *CVPR* 2011.
J. Yang et al., Beyond Sparsity: the Role of L1-optimizer in Pattern Classification, *Pattern Recognition*, 45 (2012), pp. 1104-1118.

Error Characterization

- L2-norm based error characterization

$$\min_x \|Y-Dx\|_2^2 + \lambda\|x\|_1$$

Optimal for noise with Gaussian distribution

- L1-norm based error characterization

$$\min_x \|Y-Dx\|_1 + \lambda\|x\|_1$$

Optimal for noise with Laplacian distribution

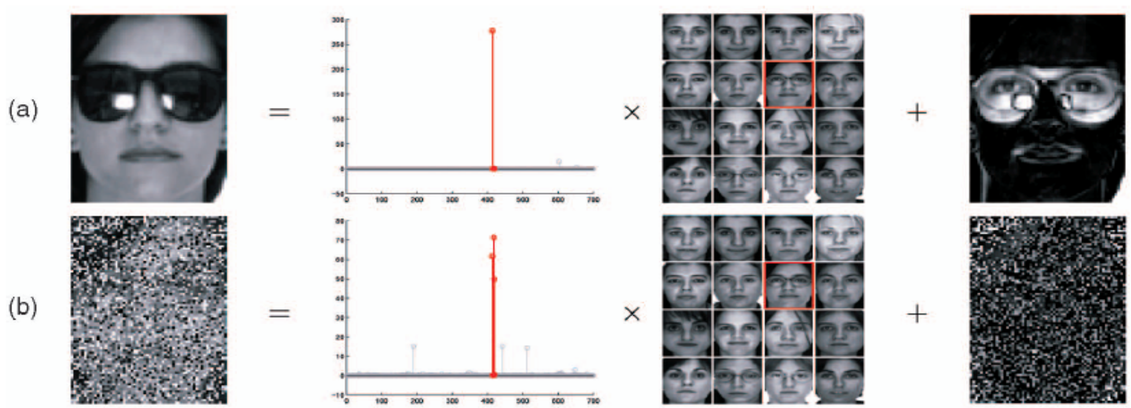
- M. Yang, L. Zhang, J. Yang and D. Zhang, Robust sparse coding for face recognition, In *CVPR*, 2011.

Sparse Representation Classifier

$$\min_{x} \|Y - Dx\|_1 + \lambda \|x\|_1$$

$D = [D, I]$ $\quad$ $w = [x, e]$

$$\min_{w} \|Y - D\, w\|_2^2 + \lambda \|w\|_1$$



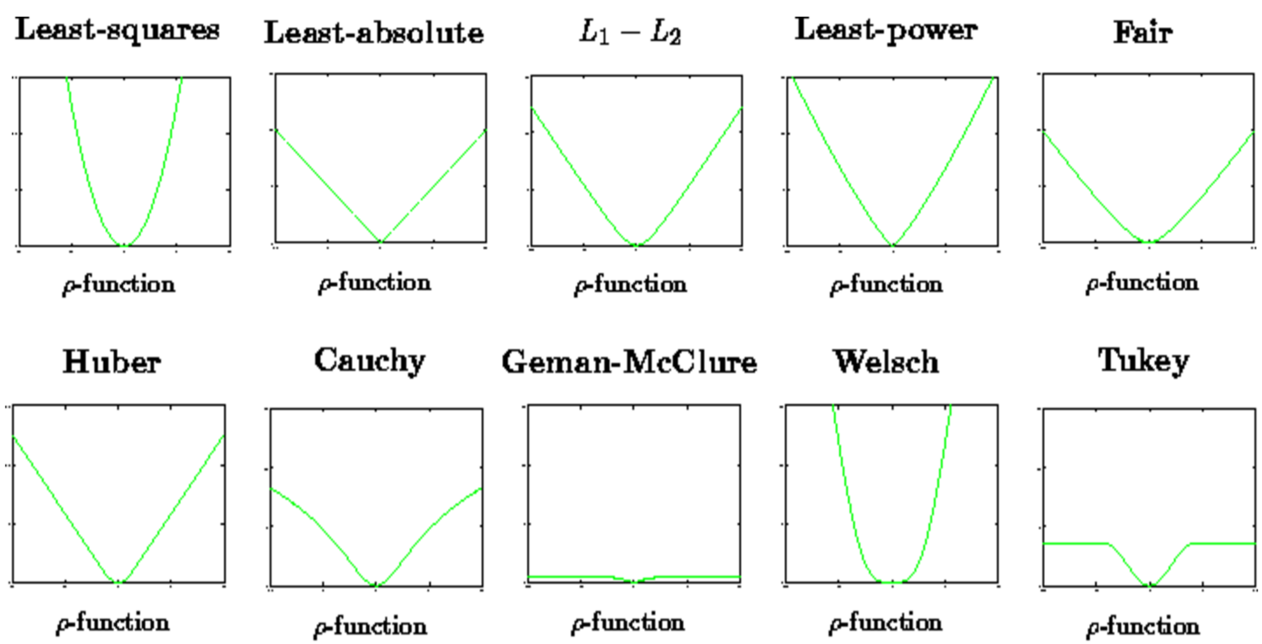
J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. PAMI*, 31(2):210–227, 2009.

Robust regression

Robust regression model

$$\min_{x} \rho_{\theta}(\mathrm{Y}-\mathrm{D}x)+\lambda||x||_{q}^{q}$$

where $\rho_{\theta}(\mathrm{Y}-\mathrm{D}x)=\sum_{i} \rho(Y_{i}-D_{i}x)$

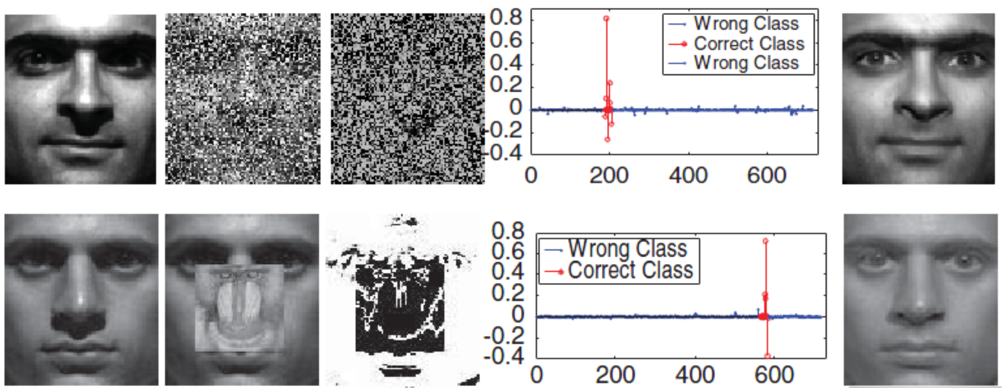


Commonly used M-estimators

Robust Sparse Representation Classifier

- Model $$\min_{x} \rho_{\theta}(\mathrm{Y}-\mathrm{D}x)+\lambda\|x\|_{1}$$

$$\min_{x} \|W^{1/2}(\mathrm{Y}-\mathrm{D}x)\|_{2}^{2}+\lambda\|x\|_{1}$$

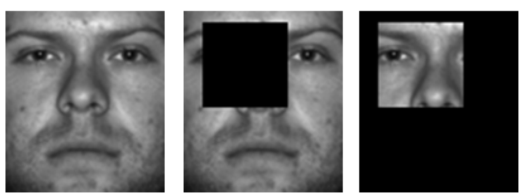


- M. Yang, L. Zhang, J. Yang and D. Zhang, Regularized robust coding for face recognition, *IEEE TIP*, VOL. 22, NO. 5, MAY 2013
- R. He, W.S. Zheng, and B.G. Hu, Maximum correntropy criterion for robust face recognition, *IEEE PAMI*, vol. 33, no. 8, pp. 1561-1576, 2011.
- R. He, W.-S. Zheng, T. Tan, and Z. Sun. Half-quadratic based Iterative Minimization for Robust Sparse Representation. IEEE Trans. PAMI, 2014, 36(2): 261 -275.

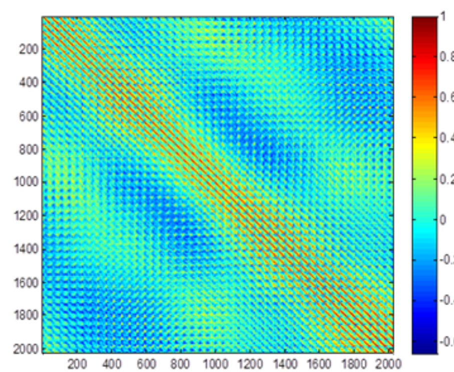
Problems of 1D Error Characterization

- Assume that $e_1, e_2, \ldots, e_m$ are independent and identically distributed (i.i.d.)
- Suitable for random-pixel noise

- Not suitable for the continuous occlusion



Original Image   Occluded Image   Error Image



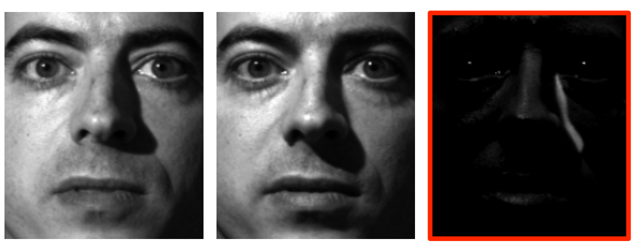
Correlation map of pixel-errors in the occluded part

Outline

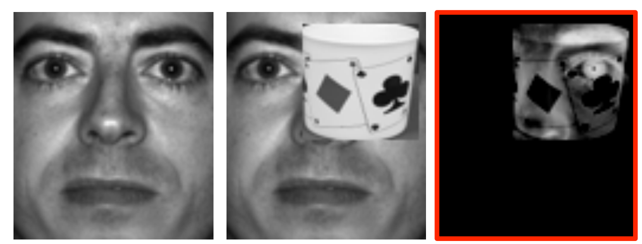
- Background and Overview
- Nuclear Norm based Matrix Regression
- Extended Versions

Motivation

- Make full use of the 2D structure of error images
- Structured noise (e.g. local illumination changes and block occlusion) caused error image is low rank



Error image



Error image

Model

- Given a set of n image matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n \in R^{p\times q}$ and an image matrix $\mathbf{B} \in R^{p\times q}$.
- **B** is linearly represented by $\mathbf{A}_1, \ldots, \mathbf{A}_n$

$$\mathbf{B} = \underbrace{x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + , \ldots, + x_n\mathbf{A}_n}_{\mathbf{A}(\mathbf{x})} + \mathbf{E}$$

$$\min_x \operatorname{rank}(\mathrm{A}(x) - \mathbf{B})$$

$$\min_x \|\mathrm{A}(x) - \mathbf{B}\|_*$$

$$\min_x \|\mathrm{A}(\mathbf{x}) - \mathbf{B}\|_* + \tfrac{1}{2}\lambda\|\mathbf{x}\|_2^2$$

Nuclear norm of a matrix M is the sum of its singular values: $\|M\|_* = \sum_i \sigma_i(M)$

Optimality of nuclear norm based error characterization

- **Optimality**:

The nuclear norm based error characterization is optimal if the structural noise matrix $\mathbf{E} = \mathbf{B} - A(\mathbf{x})$ follows the generalized matrix variate Elliptical Distribution:

$$f(\mathbf{E}) = C\exp\left(-\frac{1}{2}tr\left(\mathbf{E}^T\mathbf{E}\right)^{1/2}\right)$$

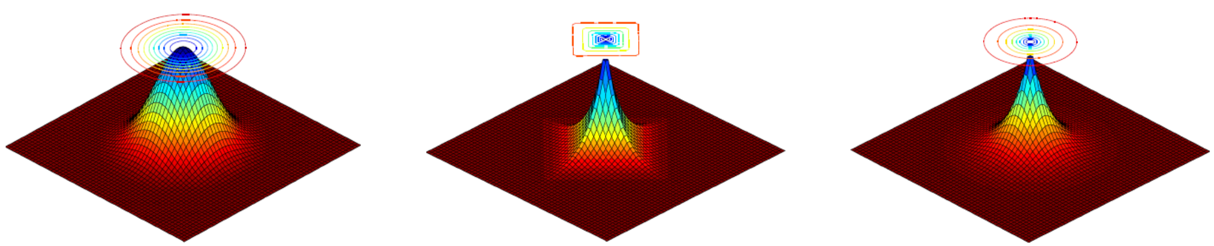


Fig. 1. Illustration of different multivariate distribution. (a) 2-D-independent Gaussian distribution $f(x,y) = C_g e^{-(x^2+y^2)}$. (b) 2-D-independent Laplace distribution $f(x,y) = C_l e^{-(|x|+|y|)}$. (c) 2-D-dependent Laplace distribution $f(x,y) = C_e e^{-(x^2+y^2)^{1/2}}$. C_g, C_l, and C_e are the positive proportionality constants.

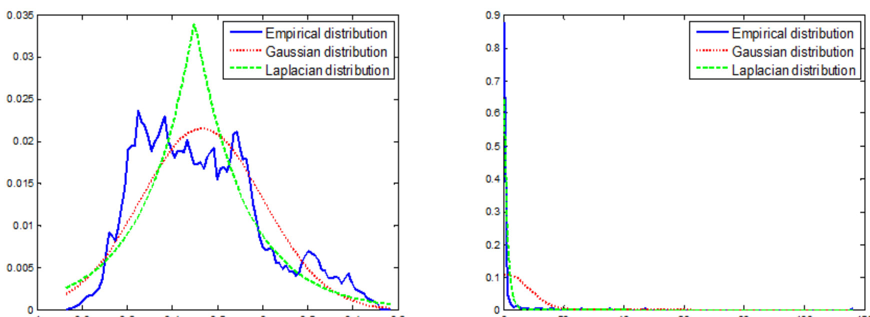
2D case of the distribution

Advantages of nuclear norm based error characterization

- **Analysis from singular value distribution point of view**

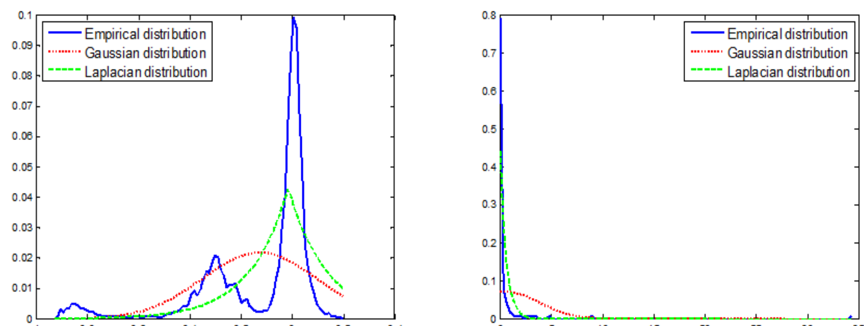
Singular value based "second-order" sparseness is more effective than **pixel based "first-order" sparseness** for describing the structured noise (occlusions, illuminations etc.)

(a) (b) (c)



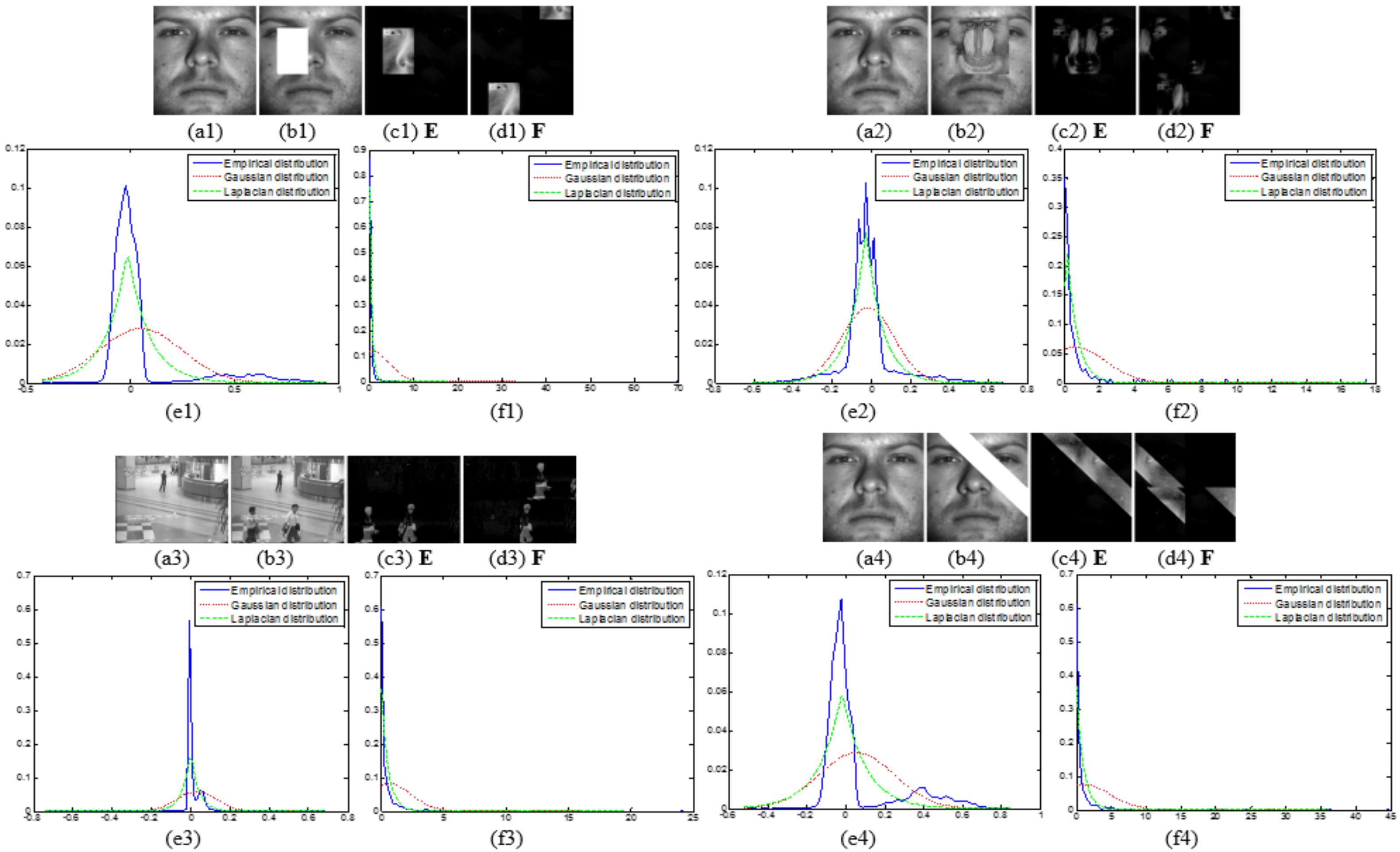
(d) (e)

(a) (b) (c)



(d) (e)

Advantages of nuclear norm based error characterization



Essence of Nuclear norm based error characterization

- **De-correlation + L1 norm of the decorrelated Matrix**

$$\|\mathbf{A}\|_* = \left\|\mathbf{U}^T\mathbf{A}\mathbf{V}\right\|_1 = \left\|\Sigma\right\|_1$$

- U is formed by eigenvectors of $\mathbf{A}\mathbf{A}^T$, which is exactly the PCA transformed matrix on image columns, so it can remove the correlation between rows.
- V is formed by eigenvectors of $\mathbf{A}^T\mathbf{A}$, which is exactly the PCA transformed matrix on image rows, so it can remove the correlation between columns.

J. Yang, C. Liu, "Horizontal and Vertical 2DPCA-based Discriminant analysis for Face Verification on a Large-scale Database", IEEE Trans. on Information Forensics and Security, 2007, 2(4), 781-792.

ADMM Algorithm for NMR

- The proposed model can be rewritten:

$$\min \|\mathbf{Y}\|_* + \tfrac{1}{2}\lambda \|\mathbf{x}\|_2^2 \quad \textit{subject to} \quad \mathrm{A}(\mathbf{x}) - \mathbf{B} = \mathbf{Y}$$

- The augmented Lagrangian function is defined by

$$L_\mu(\mathbf{Y}, \mathbf{x}, \mathbf{Z}) = \|\mathbf{Y}\|_* + \tfrac{1}{2}\lambda \|\mathbf{x}\|_2^2 + \mathrm{Tr}\left(\mathbf{Z}^T\left(\mathbf{A}(\mathbf{x}) - \mathbf{Y} - \mathbf{B}\right)\right) + \tfrac{\mu}{2}\|\mathbf{A}(\mathbf{x}) - \mathbf{Y} - \mathbf{B}\|_F^2$$

$$\begin{aligned}\mathbf{x}^{k+1} &= \arg\min_x L_\mu(\mathbf{Y}, \mathbf{x}, \mathbf{Z}) \\ &= \arg\min_x \left(\tfrac{\mu}{2}\|\mathbf{A}(\mathbf{x}) - (\mathbf{B} + \mathbf{Y} - \tfrac{1}{\mu}\mathbf{Z})\|_F^2 + \tfrac{1}{2}\lambda \|\mathbf{x}\|_2^2\right)\end{aligned}$$

Ridge regression

$$\begin{aligned}\mathbf{Y}^{k+1} &= \arg\min_{\mathbf{Y}} L_\mu(\mathbf{Y}, \mathbf{x}, \mathbf{Z}) \\ &= \arg\min_{\mathbf{Y}} \left(\|\mathbf{Y}\|_* + \tfrac{\mu}{2}\|\mathbf{A}(\mathbf{x}) - (\mathbf{B} + \mathbf{Y} - \tfrac{1}{\mu}\mathbf{Z})\|_F^2\right)\end{aligned}$$

Singular value shrinkage

ADMM Algorithm for NMR

Singular value shrinkage operator

- **Theorem**: For a given $\tau > 0$, let us define the singular value shrinkage operator

$$D_\tau(\mathbf{Q}) = \mathbf{U}_{p\times r}\mathrm{diag}\left(\{\max(0, \sigma_j - \tau)\}_{1\le j\le r}\right)\mathbf{V}_{q\times r}^T$$

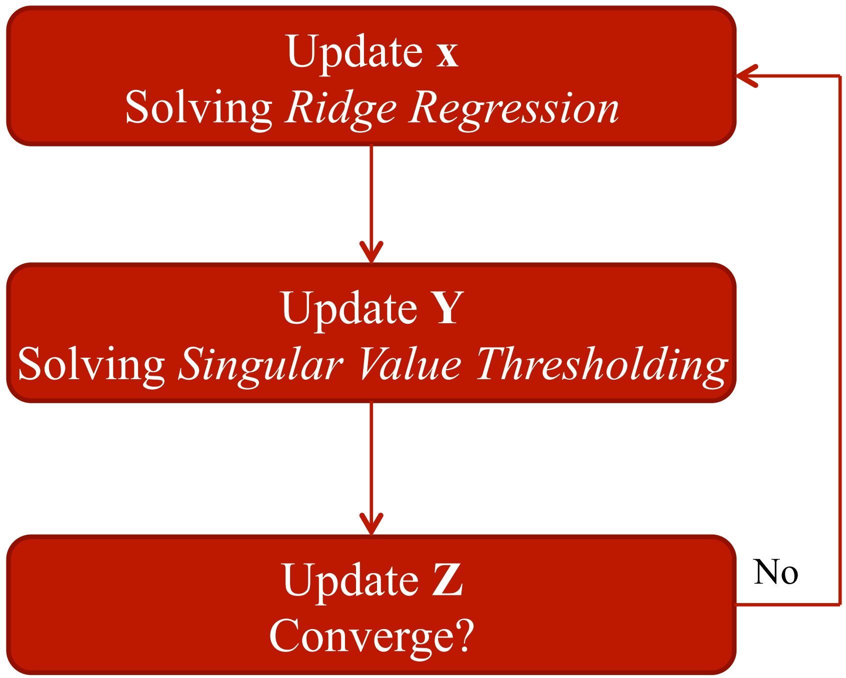
- Then

$$D_\tau(\mathbf{Q}) = \arg\min_{\mathbf{Y}}\left(\tau \| \mathbf{Y} \|_* + \tfrac{1}{2} \| \mathbf{Y} - \mathbf{Q} \|_F^2\right)$$

- J.F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM Journal on Optimization, 2010

ADMM Algorithm for NMR

- The pipeline of ADMM for NMR



ADMM Algorithm for NMR: Convergence

Convergence Theorem

- If $\mu > 0$, then the sequence $\{(\mathbf{Y}^k, \mathbf{x}^k, \mathbf{Z}^k)\}$ generated by ADMM converges to a saddle point $\{(\mathbf{Y}^k, \mathbf{x}^k, \mathbf{Z}^k)\}$of the Lagrangian function.

- **Convergence rate** ADMM Algorithm can achieve a convergence rate of $O(1/k)$.

Z. Lin, M. Chen, L. Wu, and Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. arXiv:1009.5055v2.
B. He and X. Yuan. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers, *Optimization Online*, Jan. 2012.

Fast ADMM Algorithm for NMR

- Approximate NMR : (Convex → strong convex)

$$\min \|\mathbf{Y}\|_* + \gamma \|\mathbf{Y}\|_F^2 + \tfrac{1}{2}\theta \|\mathbf{x}\|_2^2 \quad \textit{subject to} \quad \mathrm{A}(\mathbf{x}) - \mathbf{B} = \mathbf{Y}$$

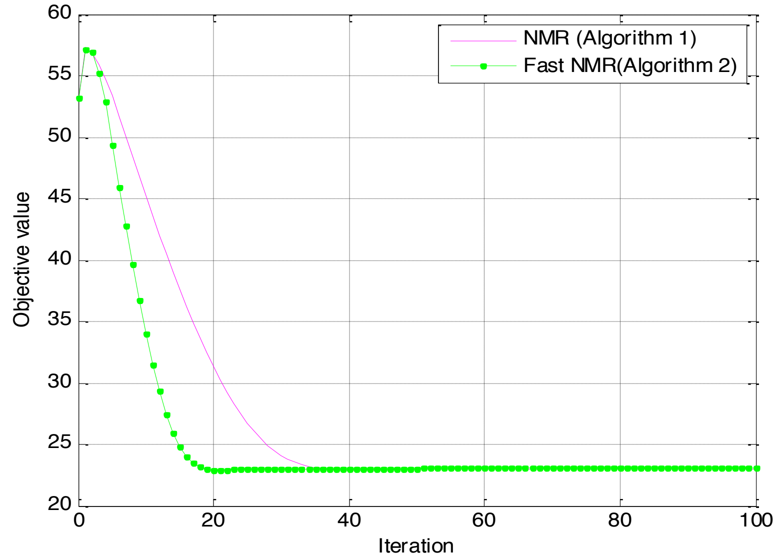
- The strong convex objective function terms ensure the optimal convergence rate
- Its solution is very close to that of NMR:

Theorem 3. Let $(\mathbf{Y}_\gamma^\star, \mathbf{x}_\gamma^\star)$ be the solution to (22) and $(\mathbf{Y}^\star, \mathbf{x}^\star)$ be the solution to problem (6), then

$$\min_{\gamma \to 0} \left\|\mathbf{Y}_\gamma^\star - \mathbf{Y}^\star\right\|_F^2 + \left\|\mathbf{x}_\gamma^\star - \mathbf{x}^\star\right\|_2^2 = 0.$$

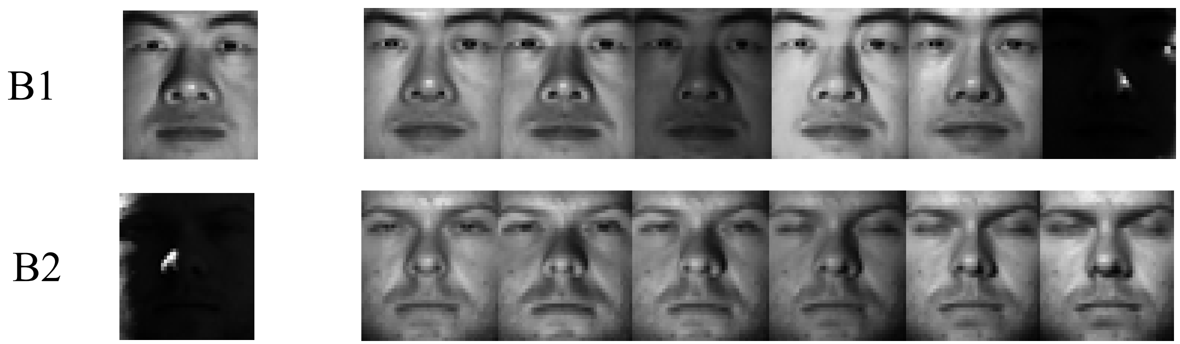
Fast ADMM: Convergence rate

- **Convergence rate** Algorithm 2 can achieve a convergence rate of $O(1/k^2)$.



- T. Goldstein, B. O'Donoghue, and S. Setzer. R. Baraniuk, Fast alternating direction optimization methods. Fast alternating direction optimization methods. SIAM Journal on Imaging Sciences, 2014, 7(3), 1588-1623.

Example: robust to illumination



Testing and Training images of two classes of faces

$\mathbf{E}_{11}^{NMR}$ $\mathbf{E}_{12}^{NMR}$ $\hat{\mathbf{B}}_{11}^{NMR}$ $\hat{\mathbf{B}}_{12}^{NMR}$

The residual images (left two) and reconstructed images (right two) of B_1 using NMR

Example: robust to illumination

$\mathbf{E}_{21}^{Ridge}$ $\mathbf{E}_{22}^{Ridge}$ $\hat{\mathbf{B}}_{21}^{Ridge}$ $\hat{\mathbf{B}}_{22}^{Ridge}$

The residual images (left two) and reconstructed images (right two) of B_2 using Ridge regression

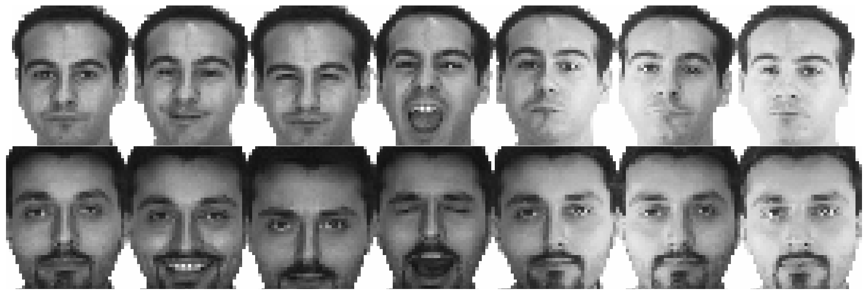
$\mathbf{E}_{21}^{NMR}$ $\mathbf{E}_{22}^{NMR}$ $\hat{\mathbf{B}}_{21}^{NMR}$ $\hat{\mathbf{B}}_{22}^{NMR}$

The residual images (left two) and reconstructed images (right two) of B_2 using NMR

$$\| \mathbf{E}_{22}^{Ridge} \|_2 = 4.54 > \| \mathbf{E}_{21}^{Ridge} \|_2 = 4.46$$

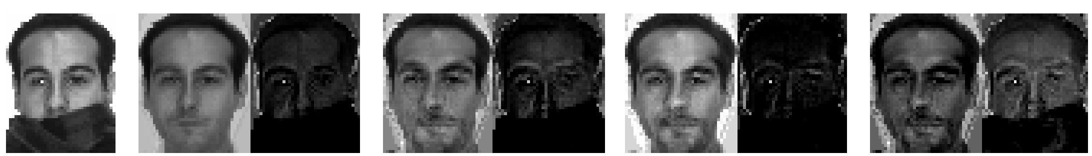
$$\| \mathbf{E}_{22}^{NMR} \|_* = 11.02 < \| \mathbf{E}_{21}^{NMR} \|_* = 11.34$$

Example: robust to occlusion



Two class of sample images from the AR database

OI-3 NMR Ridge-R SR Robust-R



OI-4 NMR Ridge-R SR Robust-R

NMR based Classifier

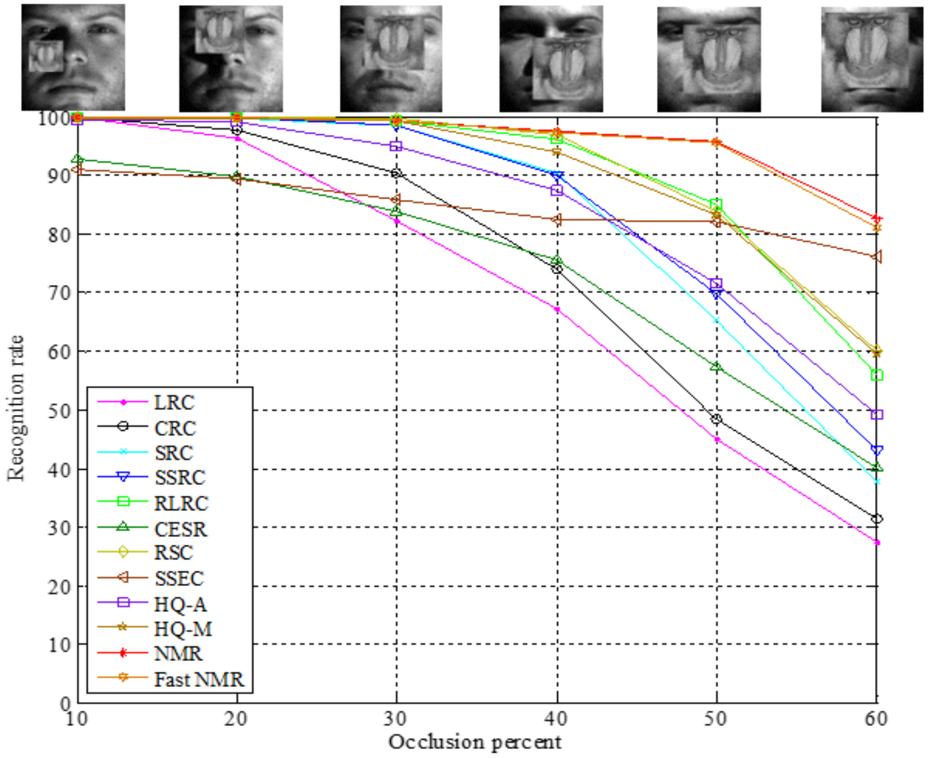
- Similar to the strategy of SRC, we use the training samples of all classes to form the set of regressors.

$$\mathbf{x}^* = \arg\min_x \| A(\mathbf{x}) - \mathbf{B} \|_* + \tfrac{1}{2}\lambda \| \mathbf{x} \|_2^2$$

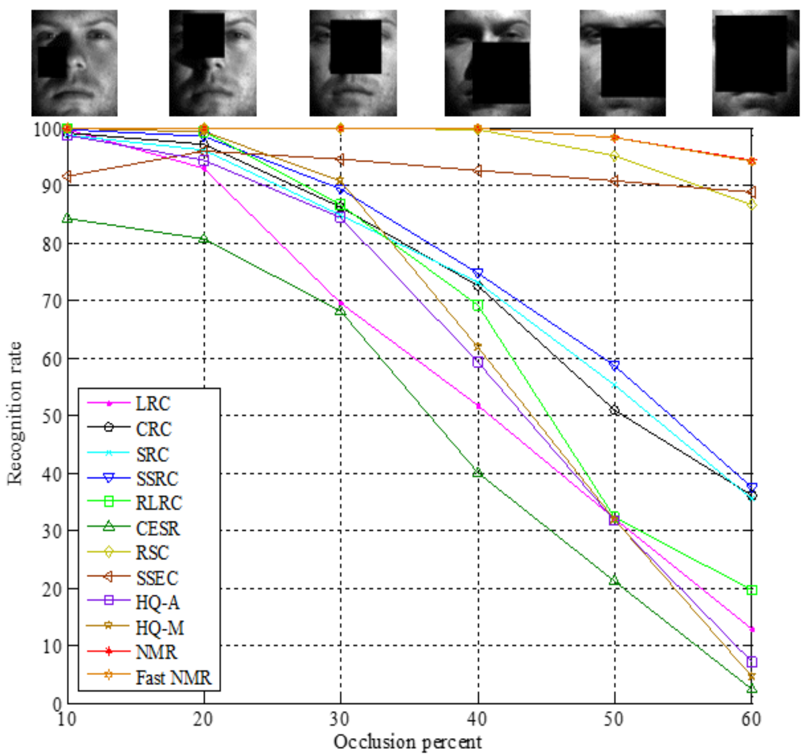
- The decision rule is defined as: if $e_l(\mathbf{B}) = \min_i e_i(\mathbf{B})$, then **B** is assigned to Class *l*.

$$e_i(\mathbf{B}) = \| \hat{\mathbf{B}} - \hat{\mathbf{B}}_i \|_* = \left\| A(\mathbf{x}^*) - A\left(\delta_i(\mathbf{x}^*)\right) \right\|_*$$

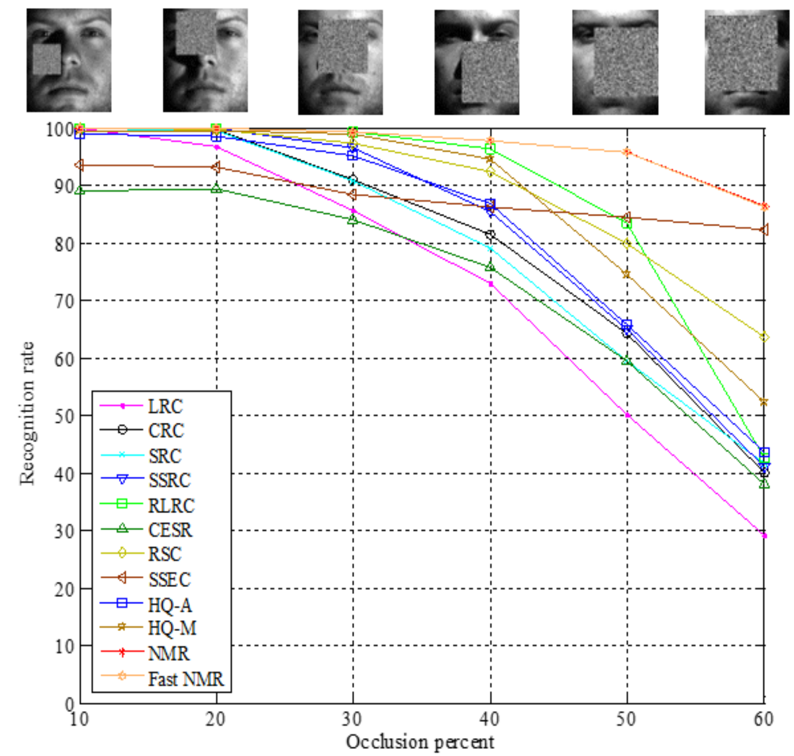
Experiment on the Extended Yale B



Experiment on the Extended Yale B

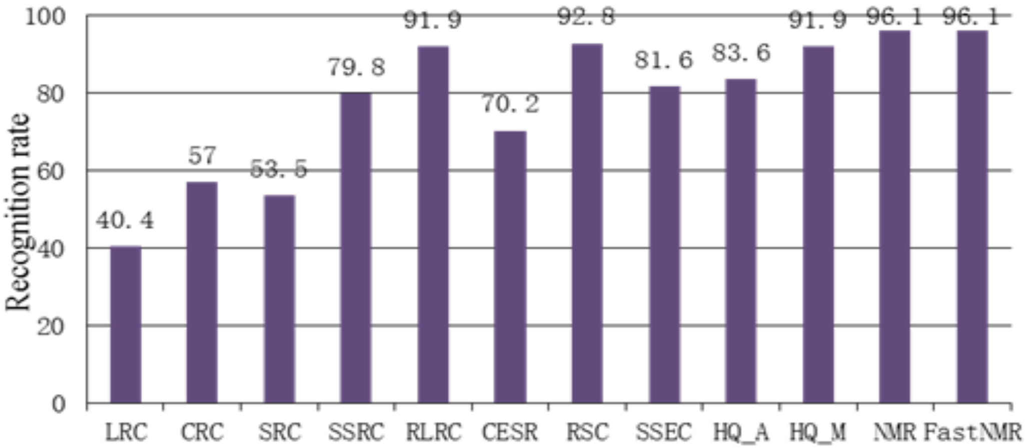


(a)

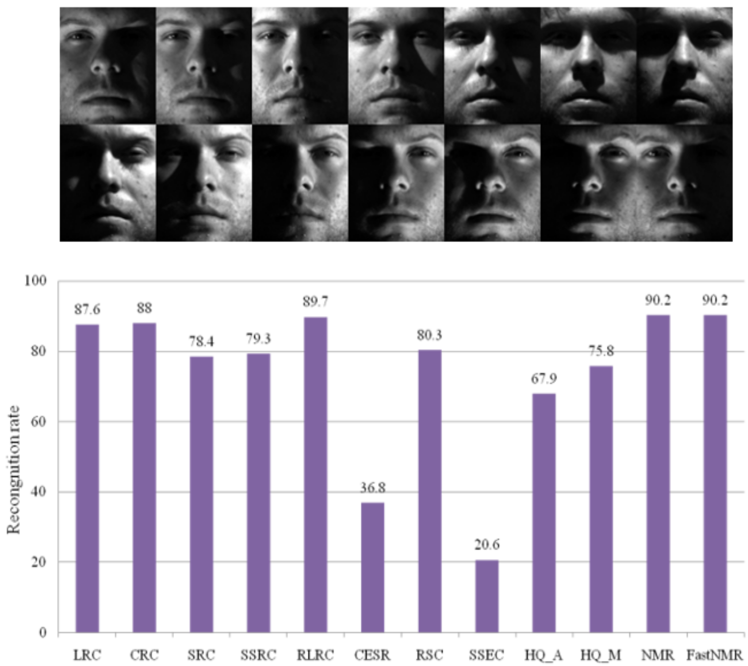


(b)

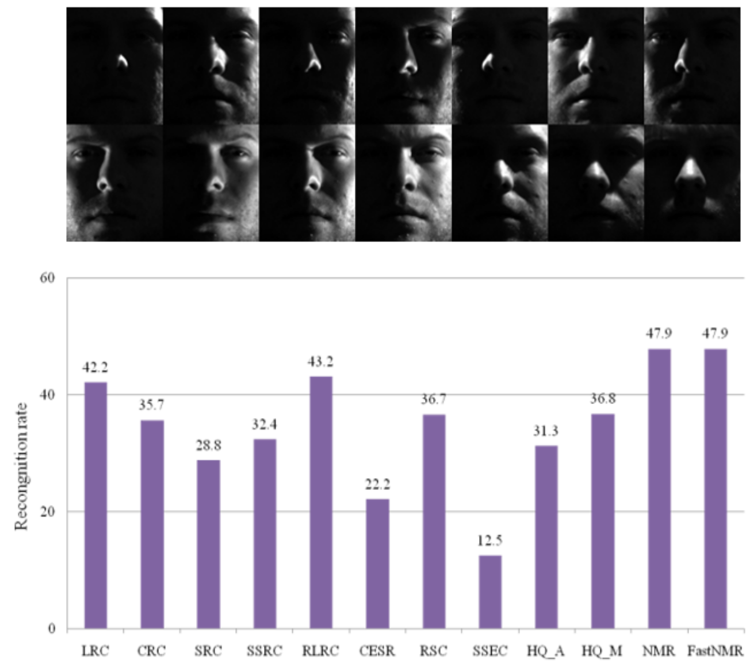
Experiment on the Extended Yale B



Experiment on the Extended Yale B



(a)



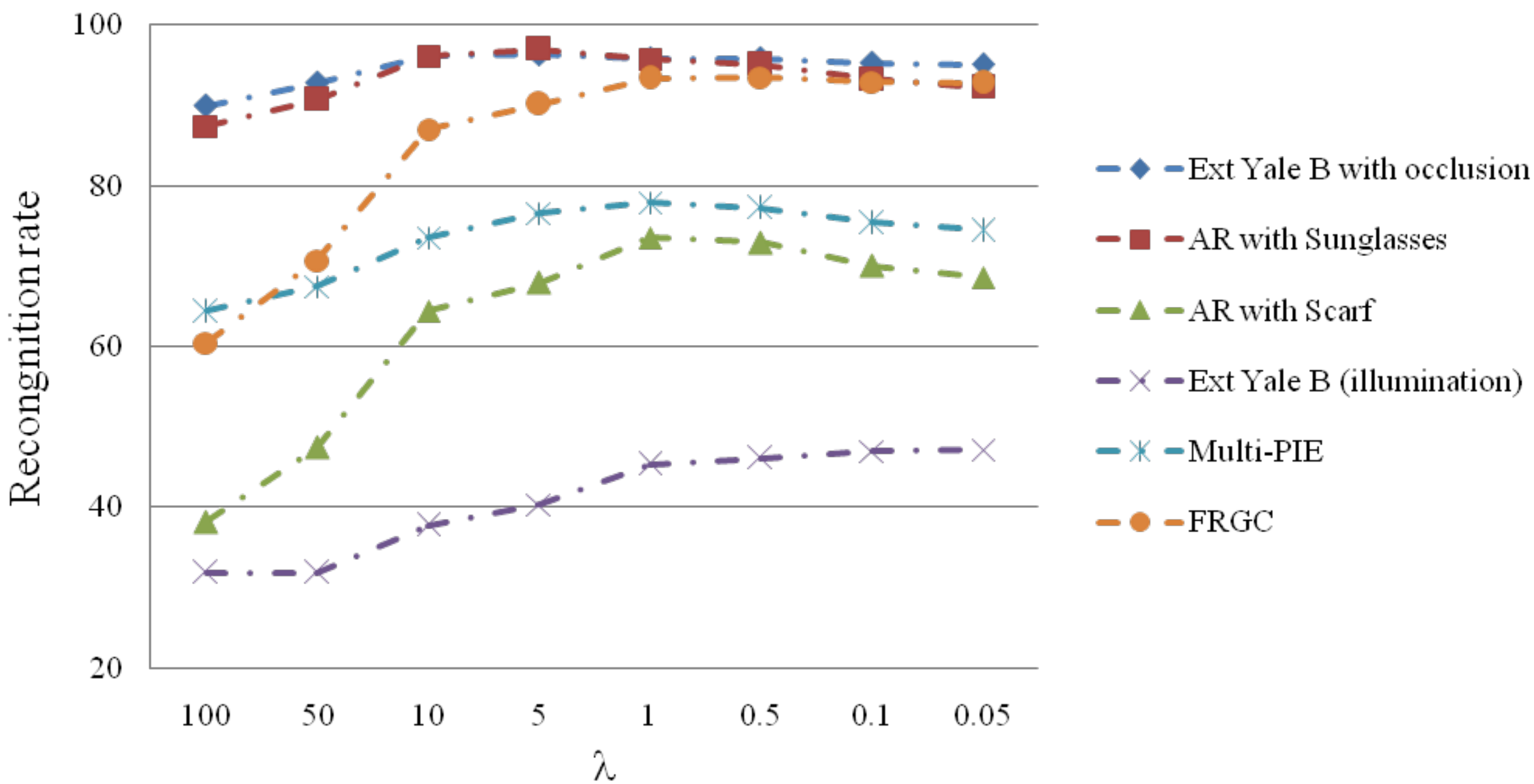
(b)

Experiment on the AR

Table 1 Recognition rates (%) of LRC, CRC, SRC, SSRC, RLRC, CESR, RSC, SSEC, HQ_A, HQ_M, NMR and Fast NMR on the AR database

	LRC	CRC	SRC	SSRC	RLRC	CESR	RSC	SSEC	HQ_A	HQ_M	NMR	Fast NMR
Sunglasses	92.8	93.5	94.4	95.4	94.6	95.0	89.2	79.0	94.7	95.0	*96.9*	*96.9*
Scarf	30.7	63.6	57.6	66.7	53.3	33.5	66.8	49.1	48.7	50.1	*73.5*	73.3

Parameter Selection



Comparison Running Time

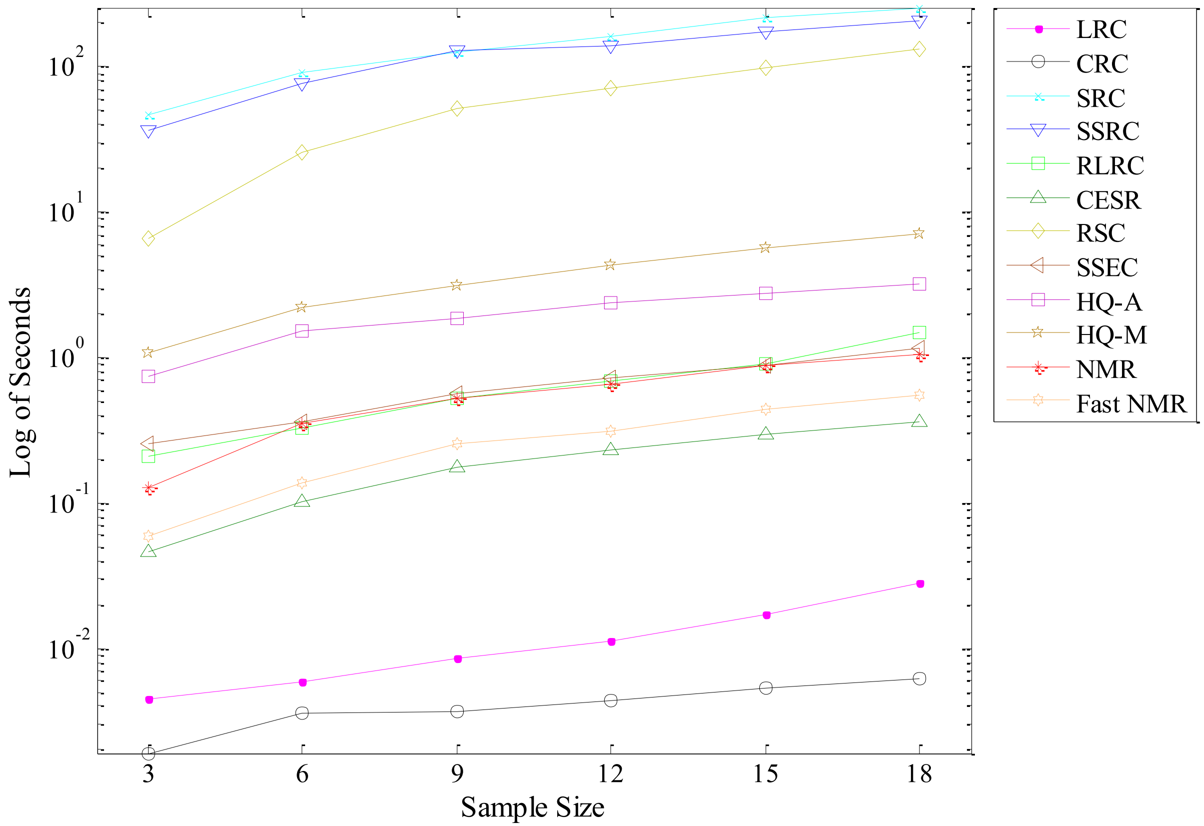


Illustration of the average running time (second, in logs) of recognizing one testing sample for each method on the Extended Yale B database

Outline

- Background and Overview
- Nuclear Norm based Matrix Regression
- Extended Versions

  (1) Schatten *p*-norm based matrix regression

  (2) Structured nuclear norm based matrix regression

(1) Schatten *p*-norm based matrix regression

- Schatten p-norm:

$$\|\mathbf{E}\|_{S_p} = \left(\sum_{i=1}^{\min\{l,m\}} \sigma_i^{\,p}\right)^{1/p}$$

- It is nuclear norm when p=1, and is Frobenius norm when p=2.
- Schatten p-norm based matrix regression:

$$\min_{\mathbf{x}} \|\mathbf{B} - \mathrm{A}(\mathbf{x})\|_{S_p}^{p} + \tfrac{1}{2}\lambda \|\mathbf{x}\|_q^q$$

Optimality of Schatten *p*-norm based error characterization

- **Optimality**:

The Schatten *p*-norm based error characterization is optimal if the structural noise matrix $\mathbf{E} = \mathbf{B} - \mathrm{A}(\mathbf{x})$ follows the generalized matrix variate Power Exponential Distribution:

$$f(\mathbf{E}) = C \exp\left(-\tfrac{1}{2} tr\left(\mathbf{E}^T \mathbf{E}\right)^{p/2}\right)$$

2D Power Exponential Distribution

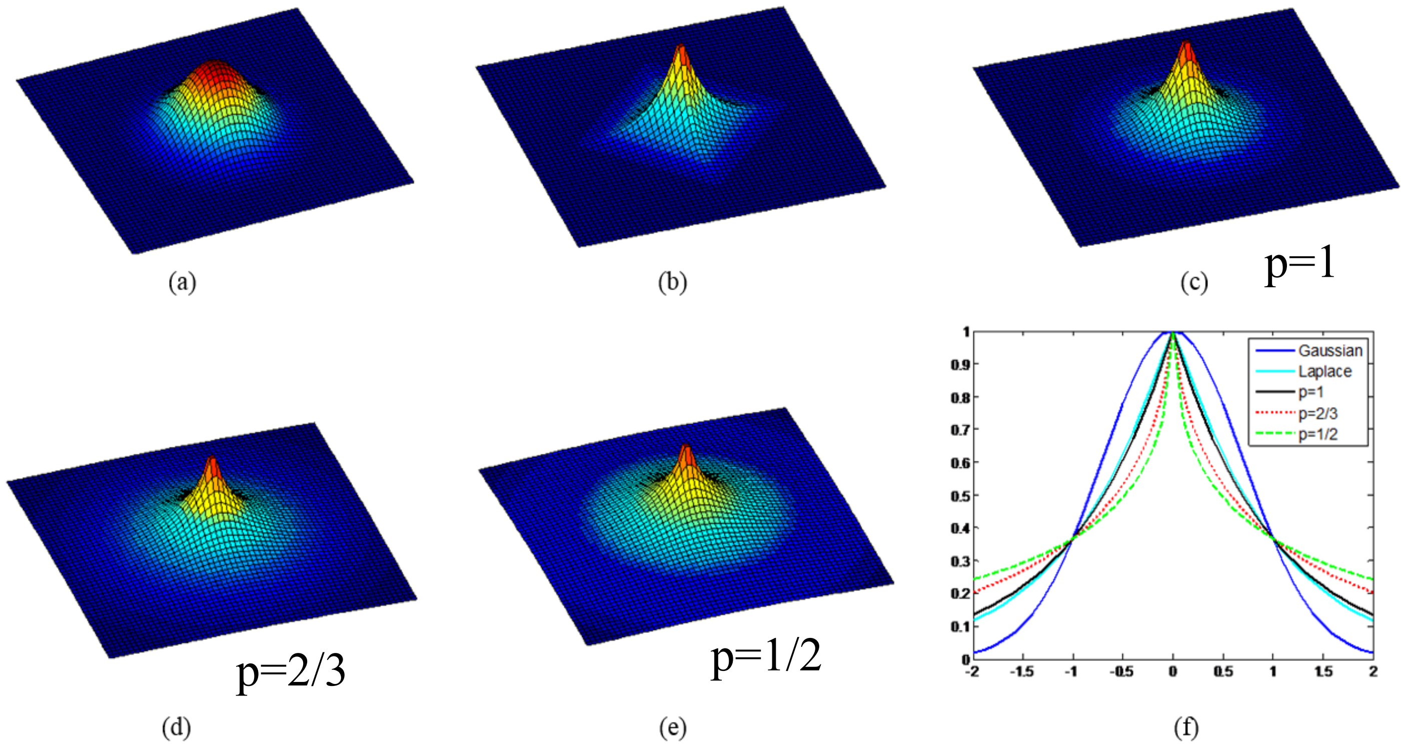


Fig. 2. (a) 2D-independent Gaussian distribution: $g(x,y)=C_1 e^{-(x^2+y^2)}$, (b) 2D-independent Laplace distribution: $g(x,y)=C_2 e^{-(|x|+|y|)}$, (c) 2D- power exponential distribution: $g(x,y)=C_3 e^{-\sqrt{x^2+y^2}}$ ($p=1$), (d) 2D-power exponential distributions: $g(x,y)=C_4 e^{-\sqrt[3]{x^2+y^2}}$ ($p=2/3$), (e) 2D-power exponential distributions: $g(x,y)=C_5 e^{-\sqrt[4]{x^2+y^2}}$ ($p=1/2$), (f) the longitudinal cutting figures for (a-e), where $C_i\,(i=1,2,3,4,5)$ are the positive proportionality constants.

Optimality of Schatten p-norm based error characterization: an equivalence

- The structural noise matrix $\mathbf{E} = \mathbf{B} - \mathrm{A}(\mathbf{x})$ follows the extended matrix variate Power Exponential Distribution:

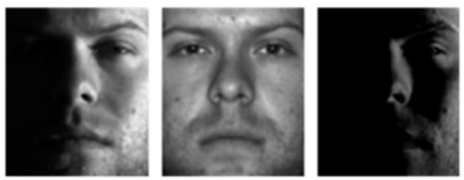
$$f(\mathbf{E}) = C\exp\left(-\tfrac{1}{2}tr\left(\mathbf{E}^T\mathbf{E}\right)^{p/2}\right)$$

$$\mathbf{W}_p = \left(\mathbf{E}^T\mathbf{E}\right)^{p/2-1} \qquad \Updownarrow \qquad \mathbf{Y} = \mathbf{E}\mathbf{W}_p^{1/2}$$

$$f(\mathbf{Y}) = C\exp\left(-\tfrac{1}{2}tr\left(\mathbf{Y}^T\mathbf{Y}\right)\right)$$

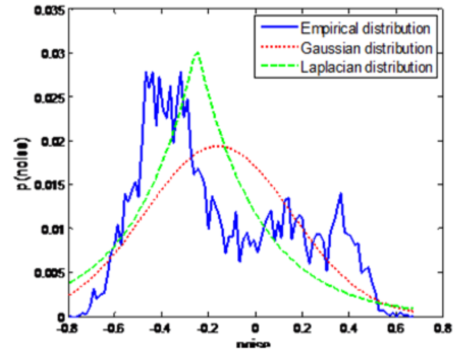
Equivalent to: Y follows matrix variate independent Gaussian distribution

Understanding Optimality from the linear transformation $\mathbf{Y} = \mathbf{E}\mathbf{W}_p^{1/2}$

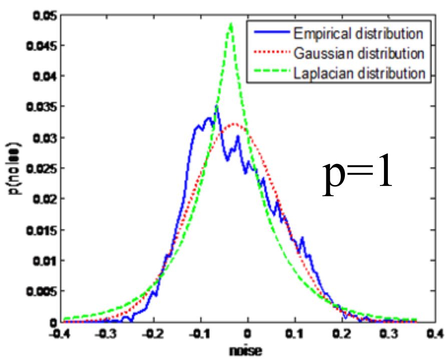


(a) (b) (c)

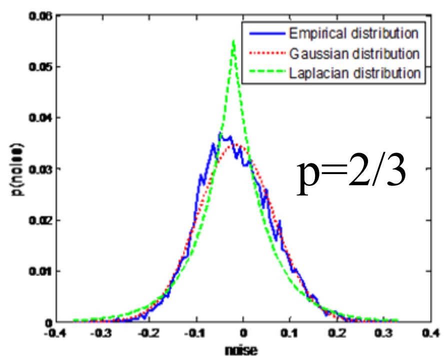
(a) Original image, (b) recovered image, (c) noise image **E**



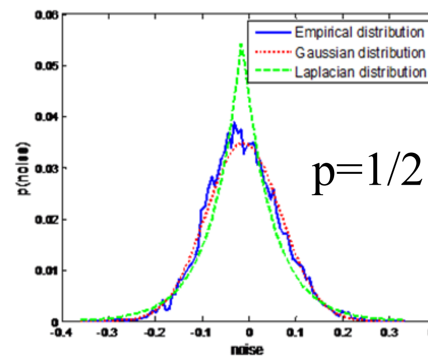
(d) The empirical distribution and the fitted distributions of the noise image **E**



(e) The empirical distribution and the fitted distributions of $\mathbf{E}\mathbf{W}_1^{1/2}$



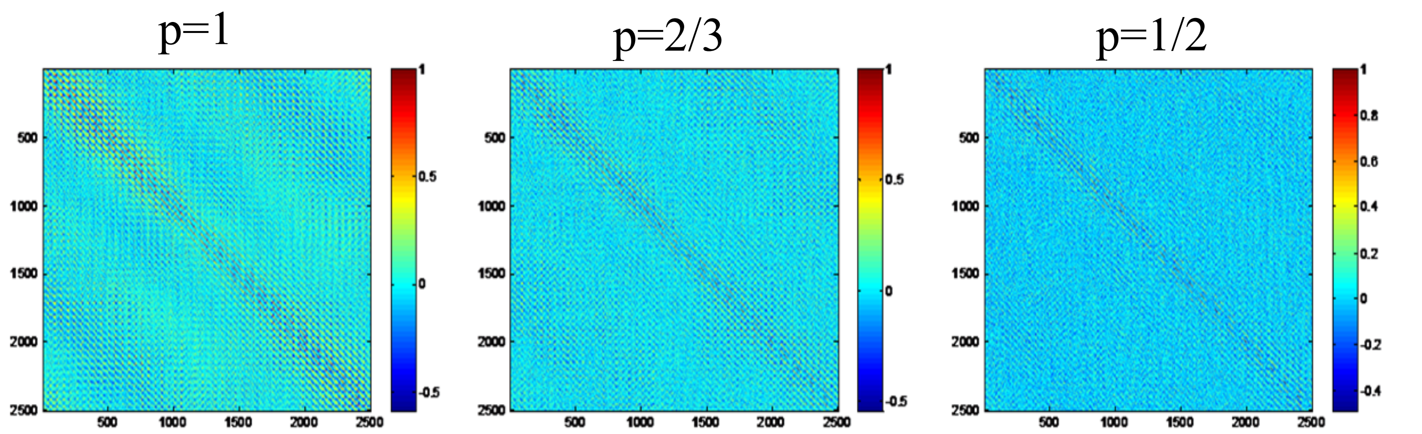
(f) The empirical distribution and the fitted distributions of $\mathbf{E}\mathbf{W}_{2/3}^{1/2}$



(g) The empirical distribution and the fitted distributions of $\mathbf{E}\mathbf{W}_{1/2}^{1/2}$

Fig. 4. An example that shows the empirical distributions and the fitted distributions of the noise image **E** and $\mathbf{E}\mathbf{W}_p^{1/2}$

Understanding Optimality from the linear transformation $\mathbf{Y} = \mathbf{E}\mathbf{W}_p^{1/2}$



(a)The correlation map of pixels in $\mathbf{E}\mathbf{W}_1^{1/2}$ (b) The correlation map of pixels in $\mathbf{E}\mathbf{W}_{2/3}^{1/2}$ (c) The correlation map of pixels in $\mathbf{E}\mathbf{W}_{1/2}^{1/2}$

Fig. 3. An example that shows the pixels in $\mathbf{Y} = \mathbf{E}\mathbf{W}_p^{1/2}$ are nearly uncorrelated

Singular Value Function Thresholding

- We focus on the case where 0<p<1, and Schatten *p*-norm is nonconvex
- Fortunately, when p=1/2 or 2/3, there is a close-form solution:

We introduce a single value function thresholding operator based on Schatten $1/2$ -norm:

$$D_\eta^{1/2}(\mathbf{G}) := \mathbf{U}\Omega_\eta^{1/2}(\mathbf{G})\mathbf{V}^T,\ \Omega_\eta^{1/2}(\mathbf{G}) = diag(\varpi(\sigma_i)\cdot\varepsilon),$$

where $\varpi(\sigma_i) = \frac{2}{3}|\sigma_i|\left(1+\cos\left(\frac{2\pi}{3}-\frac{2\varphi(\sigma_i)}{3}\right)\right)$, $\varphi(\sigma_i) = \arccos(\frac{\eta}{4}\left(\frac{\sigma_i}{3}\right)^{-3/2})$,

$$\varepsilon = \begin{cases} 1, \sigma_i > \omega(\eta) \\ 0, 0 \le \sigma_i \le \omega(\eta) \end{cases},\ \omega(\eta) = \frac{\sqrt[3]{54}}{4}(2\eta)^{2/3}.$$

Theorem 1. For each $\mathbf{E} \in R^{l\times m}$ and $\eta \ge 0$, the singular value function shrinkage operator in (15) obeys

$$D_\eta^{1/2}(\mathbf{G}) = \arg\min_{\mathbf{E}}\left(\eta\|\mathbf{E}\|_{S_{1/2}}^{1/2} + \tfrac{1}{2}\|\mathbf{E}-\mathbf{G}\|_F^2\right).$$

Reconstruction results

Fig.5. Fourteen training images of two persons from AR face database

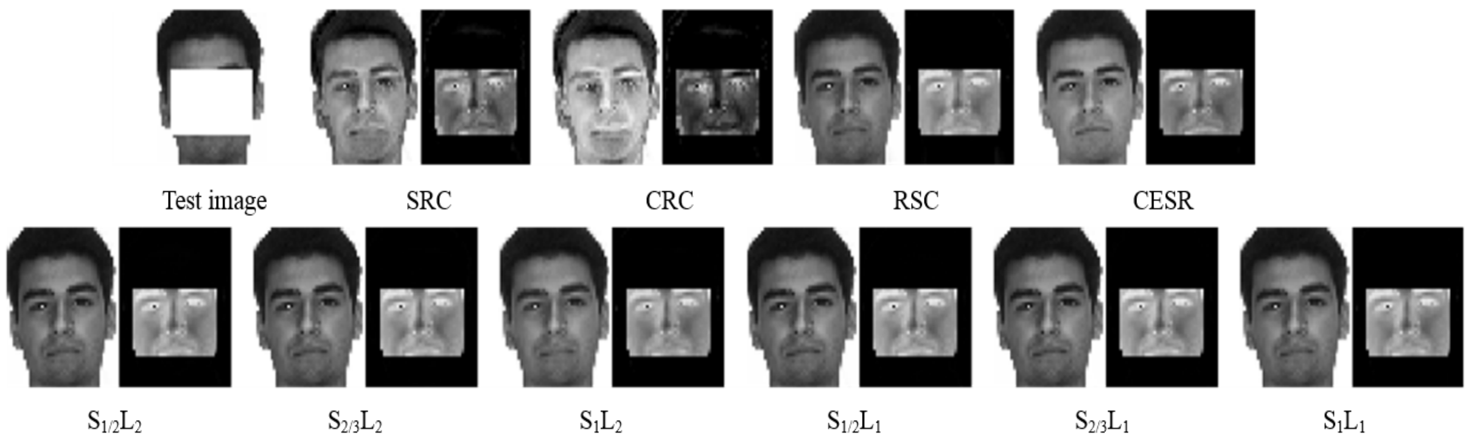


Fig.6. Recovered clean image and occluded part via five methods for the test image B with white block image

TABLE I

THE COMPARISON OF THE AVERAGE ERROR RATES AND STANDARD DEVIATIONS (%) FOR FACE RECONSTRUCTION VIA FIVE METHODS

SRC	CRC	RSC	CESR	S_1L_2	$S_{2/3}L_2$	$S_{1/2}L_2$	S_1L_1	$S_{2/3}L_1$	$S_{1/2}L_1$
20.45 ±1.69	41.1 ±0.87	4.02 ±0.73	8.54 ±3.11	6.68 ±0.85	0.68 ±0.71	1.64 ±0.73	3.98 ±0.88	**0.69** ±0.60	1.05 ±0.59

Robustness to occlusions

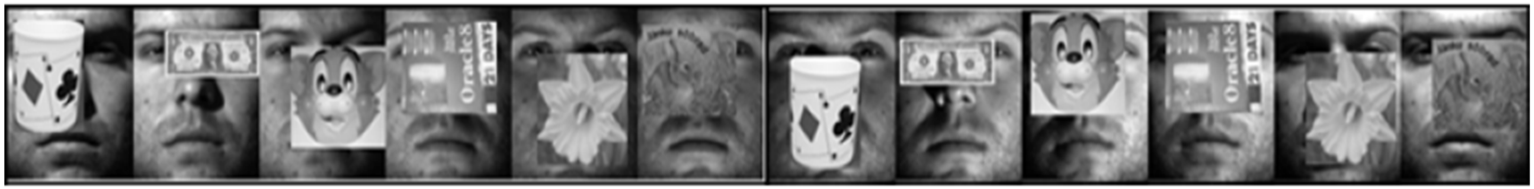


Fig. 8. Sample images of one person from the Extended Yale B database with different occlusions

TABLE IV

RECOGNITION RATES (%) OF LRC, CRC, SRC, CESR, RSC, RSC, SSEC, S_PL_Q UNDER DIFFERENT LEVELS OF OCCLUSION

LRC	CRC	SRC	CESR	RSC	SSEC	S_1L_2	$S_{2/3}L_2$	$S_{1/2}L_2$	S_1L_1	$S_{2/3}L_1$	$S_{1/2}L_1$
40.4	57	53.5	70.2	92.0	81.6	96.5	99.3	**100.0**	97.4	99.6	99.8

Fig. 9. Sample images of one person from the Extended Yale B database with different occlusion percent (10%-60%)

TABLE V

RECOGNITION RATES (%) OF LRC, CRC, SRC, CESR, RSC, RSC, SSEC, S_PL_Q UNDER DIFFERENT LEVELS OF OCCLUSION

Occlusion Percent	LRC	SRC	CRC	RSC	CESR	SSEC	S_1L_2	$S_{2/3}L_2$	$S_{1/2}L_2$	S_1L_1	$S_{2/3}L_1$	$S_{1/2}L_1$
10%	**100**	**100**	**100**	**100**	92.7	91.0	**100**	**100**	**100**	99.8	**100**	**100**
20%	96.3	99.8	97.8	**100**	89.8	89.5	99.8	99.8	**100**	99.8	**100**	**100**
30%	82.2	98.5	90.4	**99.8**	83.9	85.8	99.1	**99.8**	**99.8**	99.6	**99.8**	**99.8**
40%	67.3	90.3	73.9	98.5	75.5	82.5	97.2	**99.8**	**99.8**	98.2	**99.8**	**99.8**
50%	45.0	65.3	48.5	87.6	57.4	82.2	95.5	**99.8**	**99.8**	96.1	**99.8**	**99.8**
60%	27.4	37.5	31.3	60.1	40.1	76.1	82.4	97.4	99.1	84.4	97.6	**99.6**

Robustness to illumination

(a)Subset 4 (b)Subset 5

Fig. 10. Sample images with different illumination conditions from the Extended Yale B database

TABLE VI

RECOGNITION RATES (%) OF EACH CLASSIFIER UNDER DIFFERENT ILLUMINATION CONDITIONS ON THE EXTENDED YALE B DATABASE

Cases	LRC	SRC	CRC	RSC	CESR	SSEC	ESRC	SSRC	S_1L_2	$S_{2/3}L_2$	$S_{1/2}L_2$	S_1L_1	$S_{2/3}L_1$	$S_{1/2}L_1$
Subset 4	87.6	78.4	88	80.5	36.8	20.6	78.6	78.4	90.1	92.1	**93.8**	91.7	92.1	92.2
Subset 5	42.2	28.8	35.7	36.7	22.2	12.5	40.7	41.0	47.9	72.5	73.8.	49.7	73.6	**75.2**

Outline

- Background and Overview
- Nuclear Norm based Matrix Regression
- Extended Versions

(1) Schatten p-norm based matrix regression

(2) Structured nuclear norm based matrix regression

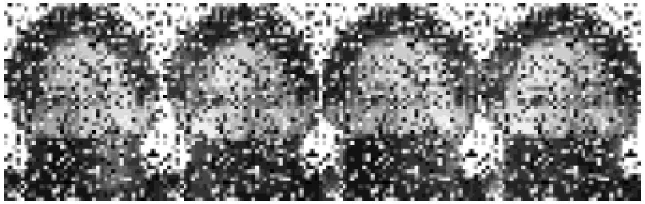
Motivations

- How to deal with mixed noise?

(a)

(b)

(c)



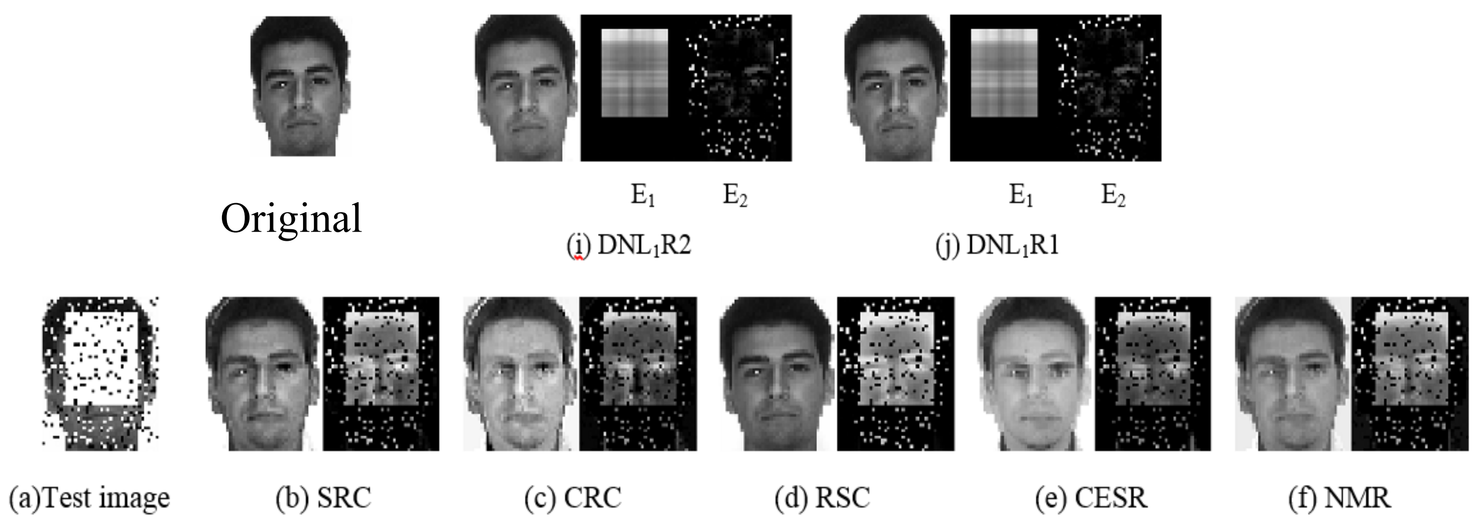
(d)

Nuclear-L_1 Norm Joint Regression

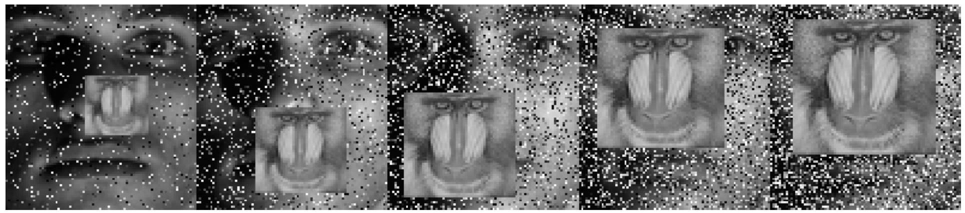
- Joint nuclear norm and L_1 norm into one model

$$\min_{\mathbf{E}_1,\mathbf{x}} \|\mathbf{E}_2\|_* + \alpha_2 \|\mathbf{E}_1\|_1 + \beta_2 \|\mathbf{x}\|_q^q, \quad \text{s.t.} \ \ \mathbf{B} - \mathbf{A}(\mathbf{x}) = \mathbf{E}_1 + \mathbf{E}_2$$

- The model degenerates to Robust PCA when x=0
- The model is effective for separate mixed noise



Recognition results



Extended Yale B: test images with different level of mixed noise

Comparison with other methods

Cases	LRC	CRC	SRC	CESR	RSC	SSEC	NMR	DNL_1R2	DNL_1R1
10%	99.6	100	99.8	94.7	100	93.4	99.8	100	**100**
20%	88.8	97.1	93.6	91.7	99.3	81.8	94.3	99.8	**100**
30%	63.8	75.0	67.5	76.8	89.9	48.0	79.4	98.2	**98.9**
40%	41.9	50.7	46.1	53.1	58.1	17.1	51.8	85.5	**87.7**
50%	17.8	24.1	28.1	22.1	23.7	7.5	27.6	50.2	**52.4**

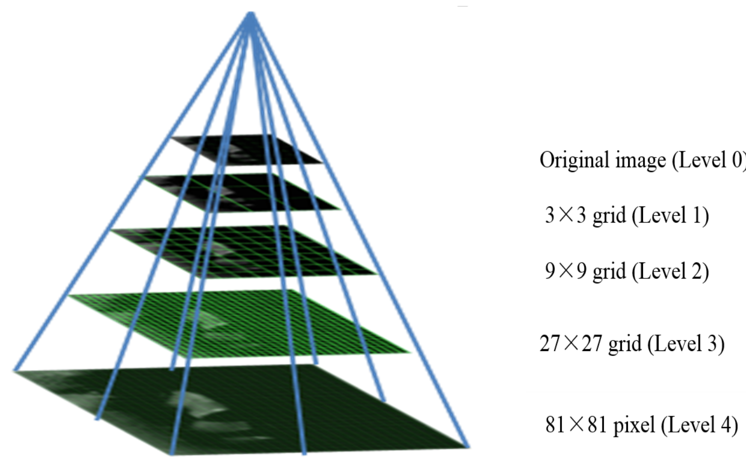
Go one step further: Structured nuclear norm based matrix regression

- Borrow the idea of structured sparsity to further extend the Nuclear-L_1 Norm
- Define the structured nuclear norm

$$\Omega(\mathbf{E}) = \sum_{i=0}^{d}\sum_{j=1}^{n_i} w_j^i \left\|\mathbf{E}_{H_j^i}\right\|_* = \left\|\mathbf{E}_{H_j^i}\right\|_{w_j^i,*}$$

- Based on the norm, we can build a new model:

$$\min_{\mathbf{x},\mathbf{E}} \left\|\mathbf{E}_{H_j^i}\right\|_{w_j^i,*} + \gamma\|\mathbf{x}\|_1, \quad \text{s.t.} \quad \mathrm{M}(\mathbf{x}) - \mathbf{L} = \mathbf{E},$$



Spatial Pyramid structure

- The model provides a general framework for regression based representation

Classification results

Fig. 9. The AR face database (a) Test image with glasses; (b) Test images with scarf

Table 4. Recognition rates (%) of LRC, CRC, SRC, CESR, RSC, SSEC, SSRC, NMR, NL_1R and our methods on the AR database

Cases	LRC	CRC	SRC	CESR	RSC	SSRC	SSEC	NMR	NL_1R	TSNA1	TSNA2	TSNA3
glasses	92.8	93.5	94.4	95.0	89.2	95.4	79.0	96.9	96.7	96.2	95.3	**97.8**
scarf	30.7	63.6	57.6	33.5	66.8	66.7	49.1	73.5	73.3	77.1	76.3	**77.1**

Structured sparsity

Conclusions and Future Efforts

Conclusions:

- We present nuclear norm based matrix regression and its variants
- Matrix regression is a simple but effective tool for robust face representation

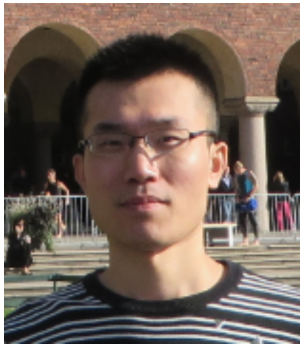
Future efforts:

- The model uses image based matching rule, which be further improved
- It is a shallow model, it is interesting to extend it to be a deep model via some techniques such as auto-encoder, etc.

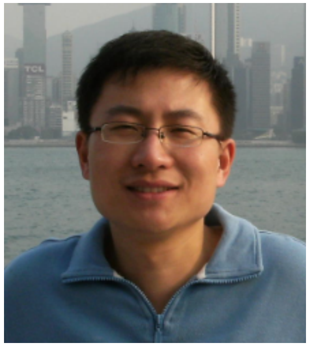
Related papers

- Jian Yang, Jianjun Qian, Lei Luo, Fanlong Zhang, Yicheng Gao, Nuclear Norm based Matrix Regression with Applications to Face Recognition with Occlusion and Illumination Changes , *arXiv:1405.1207*
- Jian Yang, Lei Luo, Jianjun Qian, Ying Tai, Fanlong Zhang andYicheng Gao, Nuclear Norm based Matrix Regression with Applications to Face Recognition with Occlusion and Illumination Changes, *IEEE Trans. PAMI*, in the second round of revision
- Jinhui Chen, Jian Yang, Lei Luo, Jianjun Qian, Wei Xu, Matrix Variate Distribution Induced Sparse Representation for Robust Image Classification, *IEEE Trans. Neural Networks and Learning Systems,* DOI:10.1109/TNNLS.2014.2377477.
- Lei Luo, Jian Yang, Jianjun Qian, Ying Tai, Nuclear-L1 Norm Joint Regression for Face Reconstruction and Recognition with Mixed Noise, *Pattern Recognition,* 2015,48(12), 3811–3824.
- Fanlong Zhang, Jian Yang, Ying Tai, Jinhui Tang, Double Nuclear Norm Based Matrix Decomposition For Occluded Image Recovery and Background Modeling, *IEEE Transactions on Image Processing,2015,* 24(6): 1956-1966
- Lei Luo, Jian Yang, et al. Tree-Structured Nuclear Norm Approximation with Applications to Robust Face Recognition, submitted

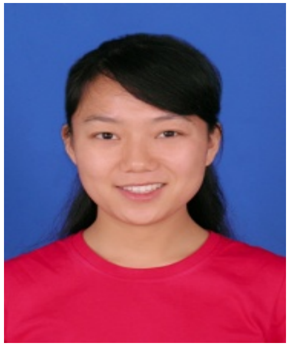
Contributors



Lei Luo



Jianjun Qian



Jinghui Chen

Fanlong Zhang

Ying Tai

Yicheng Gao

Thank you very much!