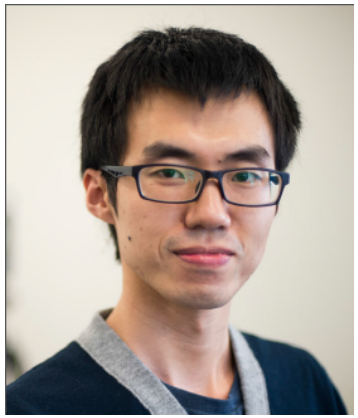
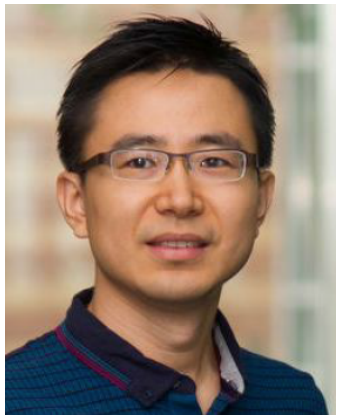
Smile lab
Synergetic Media Learning Lab

Northeastern University

Low-Rank and Sparse Modeling for Visual Analytics

Algorithms & More Models





Sheng Li & Yun Fu

Northeastern University, Boston, MA

June 26, 2016

Outline

- Optimization and Analysis
  - Convex Algorithms
  - Nonconvex Algorithms
- Scalable Algorithms
  - Randomized Algorithms
  - Distributed Algorithms
- More Models
  - Low-Rank Models for Domain Adaptation
  - Low-Rank Models for Multi-view Learning

Convex Algorithms

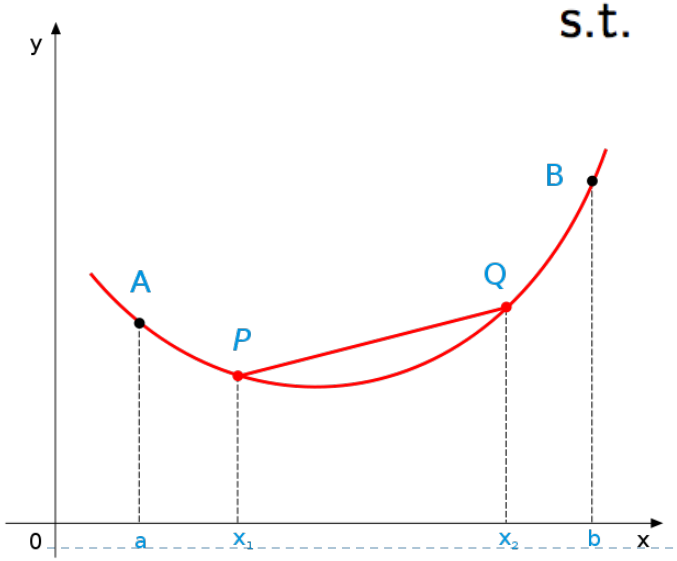
- Optimization
  - Finding the minimizer of a function subject to constraints

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \;\; i = \{1, \ldots, k\} \\ & h_j(x) = 0, \;\; j = \{1, \ldots, l\} \end{aligned}$$

Convex Algorithms

- Convex Optimization
  - An optimization problem is convex if its objective is a convex function, the inequality constraints are convex, and the equality constraints are affine

$$\begin{aligned} \underset{x}{\text{minimize}} \; & f_0(x) && \text{(Convex function)} \\ \text{s.t.} \; & f_i(x) \leq 0 && \text{(Convex sets)} \\ & h_j(x) = 0 && \text{(Affine)} \end{aligned}$$



Convex Algorithms

- Low-Rank Learning Models

Noncontinuous and nonconvex; NP-hard!

  - Matrix Completion

$$\begin{aligned} \text{minimize} \quad & \operatorname{rank}(A) \\ \text{subject to} \quad & A_{ij} = X_{ij}, \quad (i,j) \in \Omega \end{aligned}$$

  - Robust PCA (RPCA)

$$\begin{aligned} \text{minimize} \quad & \operatorname{rank}(A) + \lambda \|E\|_0 \\ \text{subject to} \quad & X = A + E \end{aligned}$$

  - Low-Rank Representation (LRR)

$$\begin{aligned} \text{minimize} \quad & \operatorname{rank}(Z) + \lambda \|E\|_0 \\ \text{subject to} \quad & X = XZ + E \end{aligned}$$

Convex Algorithms

- Low-Rank Learning Models—Convex Relaxations
  - Matrix Completion

Sum of singular values

$$\begin{aligned} \text{minimize} \quad & \|A\|_* \\ \text{subject to} \quad & A_{ij} = X_{ij}, \quad (i,j) \in \Omega \end{aligned}$$

  - Robust PCA (RPCA)

$$\begin{aligned} \text{minimize} \quad & \|A\|_* + \lambda \|E\|_l \\ \text{subject to} \quad & X = A + E \end{aligned}$$

  - Low-Rank Representation (LRR)

How to solve them?

$$\begin{aligned} \text{minimize} \quad & \|Z\|_* + \lambda \|E\|_l \\ \text{subject to} \quad & X = XZ + E \end{aligned}$$

Convex Algorithms

- Method of Multipliers
  - For a constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

  - Use the augmented Lagrangian function

$$L_\rho(x, y) = f(x) + y^T(Ax - b) + (\rho/2)\|Ax - b\|_2^2$$

Penalty term

  - Update steps

$$\begin{aligned} x^{k+1} &:= \operatorname*{argmin}_x L_\rho(x, y^k) \\ y^{k+1} &:= y^k + \rho(Ax^{k+1} - b) \end{aligned}$$

Convex Algorithms

- Alternating Direction Method of Multipliers (ADMM)
  - Form of optimization problem (two sets of variables)

$$\begin{aligned} \text{minimize} \quad & f(x) + g(z) \\ \text{subject to} \quad & A_1(x) + A_2(z) = c \end{aligned}$$

  - Augmented Lagrangian function

$$\begin{aligned} L_\rho(x, z, y) = \quad & f(x) + g(z) + y^\top (A_1(x) + A_2(z) - c) \\ & + (\rho/2) \|A_1(x) + A_2(z) - c\|^2 \end{aligned}$$

Zhouchen Lin, Minming Chen, Yi Ma. *The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices*. arXiv, 2010.

Convex Algorithms

- Alternating Direction Method of Multipliers (ADMM)
  - Form of optimization problem (two sets of variables)

$$\begin{aligned} \text{minimize} \quad & f(x) + g(z) \\ \text{subject to} \quad & A_1(x) + A_2(z) = c \end{aligned}$$

  - Update steps

$$\begin{aligned} x^{k+1} = \quad & \arg\min_x L_\rho(x, z^k, y^k) \\ & \arg\min_x f(x) + (\rho/2)\|A_1(x) + A_2(z^k) - c + y^k/\rho\|^2 \\ z^{k+1} = \quad & \arg\min_z L_\rho(x^{k+1}, z, y^k) \\ & \arg\min_z g(z) + (\rho/2)\|A_1(x^{k+1}) + A_2(z) - c + y^k/\rho\|^2 \\ y^{k+1} = \quad & y^k + \rho(A_1(x^{k+1}) + A_2(z^{k+1}) - c) \end{aligned}$$

Zhouchen Lin, Minming Chen, Yi Ma. *The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices*. arXiv, 2010.

Convex Algorithms

- Solving Low-Rank Representations (LRR) via ADMM
  - Relaxed objective function of LRR

$$\begin{aligned} \text{minimize} \quad & \|J\|_* + \lambda\|E\|_{2,1} \\ \text{subject to} \quad & X = AZ + E, \;\; Z = J \end{aligned}$$

  - Augmented Lagrangian function

$$\begin{aligned} \mathcal{L} = & \|J\|_* + \lambda\|E\|_{2,1} + \operatorname{tr}\left(Y_1^T(X - AZ - E)\right) \\ & + \operatorname{tr}\left(Y_2^T(Z - J)\right) + \frac{\mu}{2}\left(\|X - AZ - E\|_F^2 + \|Z - J\|_F^2\right) \end{aligned}$$

Penalty parameter

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, Yi Ma: *Robust Recovery of Subspace Structures by Low-Rank Representation.* IEEE Trans. PAMI, 2013

Convex Algorithms

- Solving Low-Rank Representations (LRR) via ADMM
  - Algorithm

Singular Value Thresholding (SVT)

Input: data matrix X, parameter λ.
Initialize: $Z = J = 0, E = 0, Y_1 = 0, Y_2 = 0, \mu = 10^{-6}$, $\mu_{max} = 10^6, \rho = 1.1$, and $\varepsilon = 10^{-8}$.
while not converged **do**
1. fix the others and update J by
$$J = \arg\min \tfrac{1}{\mu}\|J\|_* + \tfrac{1}{2}\|J - (Z + Y_2/\mu)\|_F^2.$$
2. fix the others and update Z by
$$Z = (\mathtt{I} + A^T A)^{-1}(A^T(X - E) + J + (A^T Y_1 - Y_2)/\mu).$$
3. fix the others and update E by
$$E = \arg\min \tfrac{\lambda}{\mu}\|E\|_{2,1} + \tfrac{1}{2}\|E - (X - AZ + Y_1/\mu)\|_F^2.$$
4. update the multipliers
$$Y_1 = Y_1 + \mu(X - AZ - E),$$
$$Y_2 = Y_2 + \mu(Z - J).$$
5. update the parameter μ by $\mu = \min(\rho\mu, \mu_{max})$.
6. check the convergence conditions:
$$\|X - AZ - E\|_\infty < \varepsilon \text{ and} \|Z - J\|_\infty < \varepsilon.$$
end while

Time Consuming

- SVD operation in SVT
- Number of iterations

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, Yi Ma: *Robust Recovery of Subspace Structures by Low-Rank Representation.* IEEE Trans. PAMI, 2013

Convex Algorithms

- Linearized ADMM with Adaptive Penalty (LADMAP)
  - Update steps in ADMM

Linearize the quadratic terms

$$
\begin{aligned}
x^{k+1} = \quad & \arg\min_x L_\rho(x, z^k, y^k) \\
& \arg\min_x f(x) + (\rho/2)\|A_1(x) + A_2(z^k) - c + y^k/\rho\|^2 \\
z^{k+1} = \quad & \arg\min_z L_\rho(x^{k+1}, z, y^k) \\
& \arg\min_z g(z) + (\rho/2)\|A_1(x^{k+1}) + A_2(z) - c + y^k/\rho\|^2
\end{aligned}
$$

  - Update steps in LADM

$$
\begin{aligned}
x^{k+1} = \quad & \arg\min_x f(x) + \langle A_1^*(y^k) + \rho^k A_1^*(A_1(x^k) + A_2(z^k) - c), x - x^k\rangle \\
& +(\rho^k\eta_1/2)\|x - x^k\|^2 \\
= \quad & \arg\min_x f(x) + (\rho^k\eta_1/2)\|x - x^k + A_1^*(y^k + \rho^k(A_1(x^k) + A_2(z^k) - c))/(\rho^k\eta_1)\|^2
\end{aligned}
$$

  - Adaptive Penalty

$$\beta_{k+1} = \min(\beta_{\max}, \rho\beta_k)$$

Zhouchen Lin, Risheng Liu, Zhixun Su: *Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation*. NIPS 2011

Convex Algorithms

- Empirical Study

Table 1: Comparison among APG, ADM, LADM, LADMAP and LADMAP(A) on the synthetic data. For each quadruple $(s, p, d, \tilde{r})$, the LRR problem, with $\mu = 0.1$, was solved for the same data using different algorithms. We present typical running time (in $\times 10^3$ seconds), iteration number, relative error (%) of output solution $(\hat{\mathbf{E}}, \hat{\mathbf{Z}})$ and the clustering accuracy (%) of tested algorithms, respectively.

Size $(s, p, d, \tilde{r})$	Method	Time	Iter.	$\frac{\|\hat{\mathbf{Z}}-\mathbf{Z}_0\|}{\|\mathbf{Z}_0\|}$	$\frac{\|\hat{\mathbf{E}}-\mathbf{E}_0\|}{\|\mathbf{E}_0\|}$	Acc.
(10, 20,200, 5)	APG	0.0332	110	2.2079	1.5096	81.5
	ADM	0.0529	176	0.5491	0.5093	**90.0**
	LADM	0.0603	194	**0.5480**	**0.5024**	**90.0**
	LADMAP	0.0145	**46**	**0.5480**	**0.5024**	**90.0**
	LADMAP(A)	**0.0010**	**46**	**0.5480**	**0.5024**	**90.0**
(15, 20,300, 5)	APG	0.0869	106	2.4824	1.0341	80.0
	ADM	0.1526	185	0.6519	0.4078	83.7
	LADM	0.2943	363	**0.6518**	**0.4076**	**86.7**
	LADMAP	0.0336	**41**	**0.6518**	**0.4076**	**86.7**
	LADMAP(A)	**0.0015**	**41**	**0.6518**	**0.4076**	**86.7**
(20, 25, 500, 5)	APG	1.8837	117	2.8905	2.4017	72.4
	ADM	3.7139	225	1.1191	1.0170	80.0
	LADM	8.1574	508	**0.6379**	**0.4268**	80.0
	LADMAP	0.7762	**40**	**0.6379**	**0.4268**	**84.6**
	LADMAP(A)	**0.0053**	**40**	**0.6379**	**0.4268**	**84.6**
(30, 30, 900, 5)	APG	6.1252	116	3.0667	0.9199	69.4
	ADM	11.7185	220	0.6865	0.4866	**76.0**
	LADM	N.A.	N.A.	N.A.	N.A.	N.A.
	LADMAP	2.3891	**44**	**0.6864**	**0.4294**	**80.1**
	LADMAP(A)	**0.0058**	**44**	**0.6864**	**0.4294**	**80.1**

Zhouchen Lin, Risheng Liu, Zhixun Su: *Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation*. NIPS 2011

Nonconvex Algorithms

- Motivation
  - Better approximation
  - Overcome the imbalanced penalization of different singular values

Nonconvex Algorithms

- Truncated Nuclear Norm Regularization (TNNR)
  - Matrix Completion Problem

$$\min_X \ \|X\|_* \quad s.t. \ \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \qquad \Rightarrow \qquad \min_X \ \|X\|_r \quad s.t. \ \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M)$$

Definition 3.1. *Given a matrix* $X \in \mathbb{R}^{m\times n}$, *the truncated nuclear norm* $\|X\|_r$ *is defined as the sum of* $\min(m,n)-r$ *minimum singular values, i.e.,* $\|X\|_r = \sum_{i=r+1}^{\min(m,n)} \sigma_i(X)$.

 Hu et al. *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE TPAMI, 2013.

Nonconvex Algorithms

- Truncated Nuclear Norm Regularization (TNNR)
  - Matrix Completion Problem

$$\min_X \ \|X\|_* \quad s.t. \ \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M)$$

Nonconvex

$$\min_X \ \|X\|_r \quad s.t. \ \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M)$$

Equivalent Form

$$\min_X \ \|X\|_* - \max_{AA^T=I, BB^T=I} \mathrm{Tr}(AXB^T) \quad s.t. \ \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M),$$

where $A \in \mathbb{R}^{r\times m}, B \in \mathbb{R}^{r\times n}$.

Convex Subproblem

$$\min_X \ \|X\|_* - \mathrm{Tr}(A_l X B_l^T) \quad s.t. \ \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M).$$

- A_l and B_l are computed from the SVD of X_l

 Hu et al. *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE TPAMI, 2013.

Nonconvex Algorithms

- TNNR-ADMM
  - Relaxed objective function

$$\min_{X,W} \|X\|_* - \mathrm{Tr}(A_l W B_l^T)$$
$$s.t. \quad X = W, \mathcal{P}_\Omega(W) = \mathcal{P}_\Omega(M)$$

  - Augmented Lagrangian function

$$L(X,Y,W,\beta) = \|X\|_* - \mathrm{Tr}(A_l W B_l^T) + \frac{\beta}{2}\|X-W\|_F^2 + \mathrm{Tr}(Y^T(X-W)),$$

Hu et al. *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE TPAMI, 2013.

Nonconvex Algorithms

- TNNR-ADMM
  - Relaxed objective function

$$\min_{X,W} \|X\|_* - \mathrm{Tr}(A_l W B_l^T)$$
$$s.t. \quad X = W, \mathcal{P}_\Omega(W) = \mathcal{P}_\Omega(M)$$

  - Update steps

$$X_{k+1} = \arg\min_X \|X\|_* + \frac{\beta}{2}\|X - \left(W_k - \frac{1}{\beta}Y_k\right)\|_F^2$$

$$W_{k+1} = \arg\min_{\mathcal{P}_\Omega(W)=\mathcal{P}_\Omega(M)} \frac{\beta}{2}\left\|W - \left(X_{k+1} + \frac{1}{\beta}\left(A_l^T B_l + Y_k\right)\right)\right\|_F^2$$

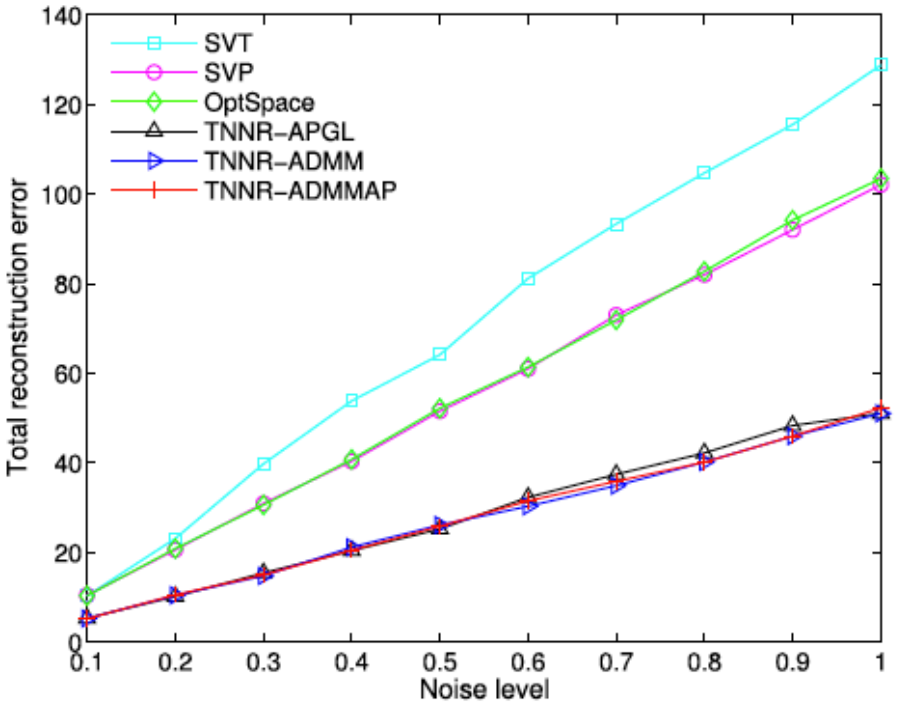
$$Y_{k+1} = Y_k + \beta(X_{k+1} - W_{k+1})$$

  - In addition, TNNR-APGL and TNNR-ADMMAP are introduced

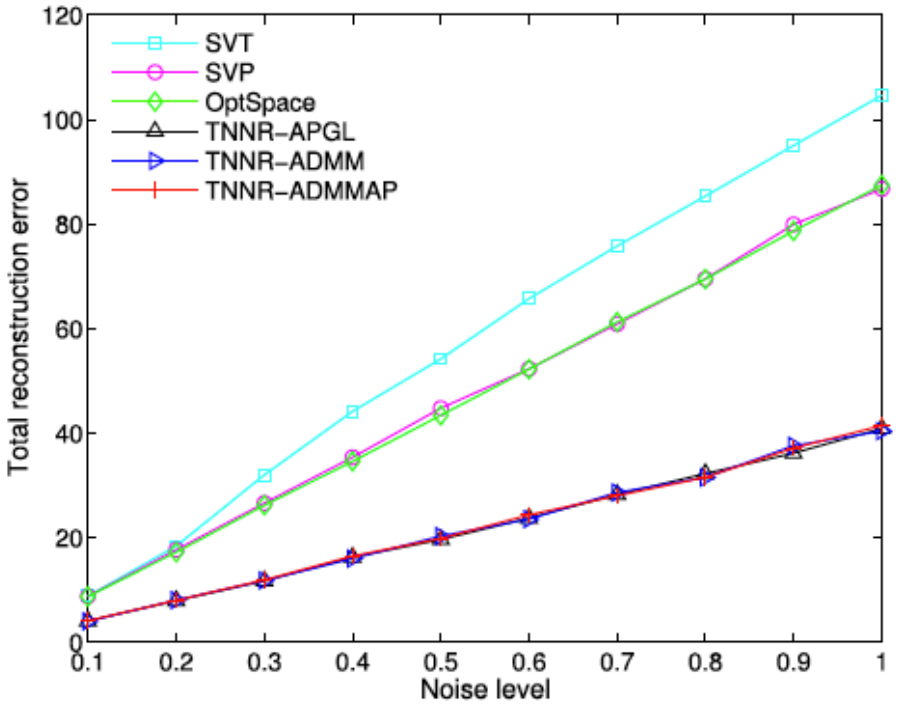
 Hu et al. *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE TPAMI, 2013.

Nonconvex Algorithms

- TNNR-ADMM



(a) 60% observed



(b) 70% observed

 Hu et al. *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE TPAMI, 2013.

Nonconvex Algorithms

- Iterative Shrinkage-Thresholding and Reweighted Algorithm (ISTRA)
  - Inspired by weighted L1 norm (Candes, Wakin and Boyd, 2008)
  - Weighted nuclear norm defined as a weighted sum of all singular values to enhance low-rank
  - Optimization problem

$$\min_X \ f(X) + \lambda w^T \sigma(X)$$

where $\sigma(X) = [\sigma_1(X) \cdots \sigma_q(X)]^T$, $\sigma_i(X)$ is the i-th largest singular value of X.

Nonconvex Algorithms

- Iterative Shrinkage-Thresholding and Reweighted Algorithm (ISTRA)
  - Update X

$$X^{k+1} = \arg\min_X P_{t^k}(X, X^k) + \lambda (w^k)^T \sigma(X)$$

where $P_{t^k}(X, X^k) = f(X^k) + \langle X - X^k, \nabla f(X^k) \rangle + \frac{t^k}{2} \|X - X^k\|^2$ is a first-order approximation of $f(X)$ at X^k regularized by a quadratic term

$$X^{k+1} = \arg\min_X \frac{t^k}{2\lambda} \left\| X - \left(X^k - \frac{1}{t^k} \nabla f(X^k) \right) \right\|_F^2 + (w^k)^T \sigma(X)$$

  - Update W

$$w_i^k = \frac{r}{(\sigma_i(X^k) + \epsilon)^{1-r}}$$

where $i = 1 \cdots q$, $0 < r < 1$, and $\epsilon > 0$ is a smoothing parameter.

 Zhong et al. A Nonconvex Relaxation Approach for Rank Minimization Problems. *AAAI 2015*

Nonconvex Algorithms

- Iterative Shrinkage-Thresholding and Reweighted Algorithm (ISTRA)

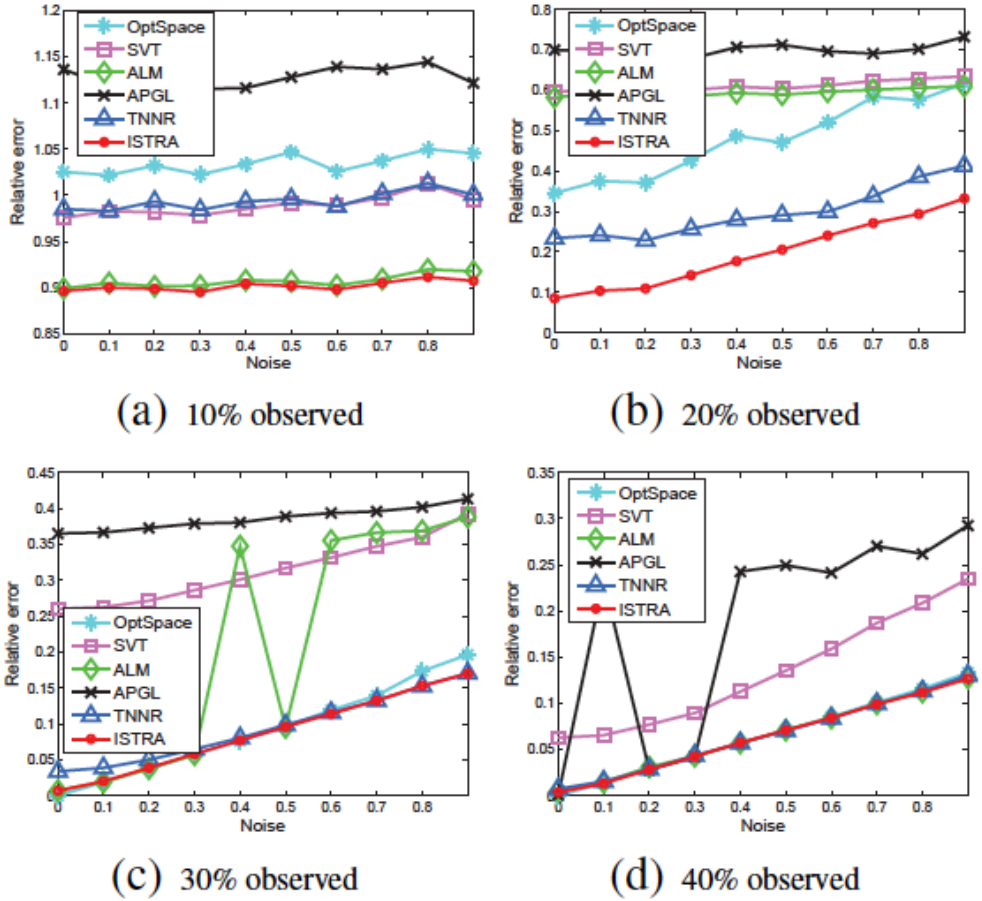


Figure 2: Relative error versus noise with different observations

 Zhong et al. A Nonconvex Relaxation Approach for Rank Minimization Problems. *AAAI 2015*

Nonconvex Algorithms

- Iteratively Reweighted Nuclear Norm (IRNN)
  - A general nonconvex and nonsmooth problem

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} F(\mathbf{X}) = \sum_{i=1}^{m} g_\lambda(\sigma_i(\mathbf{X})) + f(\mathbf{X})$$

where $\sigma_i(\mathbf{X})$ denotes the i-th singular value of $\mathbf{X} \in \mathbb{R}^{m\times n}$

  - Assumptions
    - g is continuous, concave and monotonically
    - f is a smooth function (the gradient is Lipschitz continuous)
  - Idea
    - Compute the proximal operator of the weighted nuclear norm, which has convergence guarantee

 Lu et al. Generalized Nonconvex Nonsmooth Low-Rank Minimization. *CVPR 2014*

Nonconvex Algorithms

- Iteratively Reweighted Nuclear Norm (IRNN)
  - g could be any nonconvex penalty function in Table 1.

Table 1: Popular nonconvex surrogate functions of $||\theta||_0$ and their supergradients.

Penalty	Formula $g_\lambda(\theta), \theta \geq 0, \lambda > 0$	Supergradient $\partial g_\lambda(\theta)$
L_p [11]	$\lambda\theta^p$	$\begin{cases} \infty, & \text{if } \theta = 0, \\ \lambda p\theta^{p-1}, & \text{if } \theta > 0. \end{cases}$
SCAD [10]	$\begin{cases} \lambda\theta, & \text{if } \theta \leq \lambda, \\ \frac{-\theta^2+2\gamma\lambda\theta-\lambda^2}{2(\gamma-1)}, & \text{if } \lambda < \theta \leq \gamma\lambda, \\ \frac{\lambda^2(\gamma+1)}{2}, & \text{if } \theta > \gamma\lambda. \end{cases}$	$\begin{cases} \lambda, & \text{if } \theta \leq \lambda, \\ \frac{\gamma\lambda-\theta}{\gamma-1}, & \text{if } \lambda < \theta \leq \gamma\lambda, \\ 0, & \text{if } \theta > \gamma\lambda. \end{cases}$
Logarithm [12]	$\frac{\lambda}{\log(\gamma+1)}\log(\gamma\theta+1)$	$\frac{\gamma\lambda}{(\gamma\theta+1)\log(\gamma+1)}$
MCP [23]	$\begin{cases} \lambda\theta - \frac{\theta^2}{2\gamma}, & \text{if } \theta < \gamma\lambda, \\ \frac{1}{2}\gamma\lambda^2, & \text{if } \theta \geq \gamma\lambda. \end{cases}$	$\begin{cases} \lambda - \frac{\theta}{\gamma}, & \text{if } \theta < \gamma\lambda, \\ 0, & \text{if } \theta \geq \gamma\lambda. \end{cases}$
Capped L_1 [24]	$\begin{cases} \lambda\theta, & \text{if } \theta < \gamma, \\ \lambda\gamma, & \text{if } \theta \geq \gamma. \end{cases}$	$\begin{cases} \lambda, & \text{if } \theta < \gamma, \\ [0,\lambda], & \text{if } \theta = \gamma, \\ 0, & \text{if } \theta > \gamma. \end{cases}$
ETP [13]	$\frac{\lambda}{1-\exp(-\gamma)}(1-\exp(-\gamma\theta))$	$\frac{\lambda\gamma}{1-\exp(-\gamma)}\exp(-\gamma\theta)$
Geman [15]	$\frac{\lambda\theta}{\theta+\gamma}$	$\frac{\lambda\gamma}{(\theta+\gamma)^2}$
Laplace [21]	$\lambda(1-\exp(-\frac{\theta}{\gamma}))$	$\frac{\lambda}{\gamma}\exp(-\frac{\theta}{\gamma})$

 Lu et al. Generalized Nonconvex Nonsmooth Low-Rank Minimization. *CVPR 2014*

Nonconvex Algorithms

- Iteratively Reweighted Nuclear Norm (IRNN)
  - Update rule

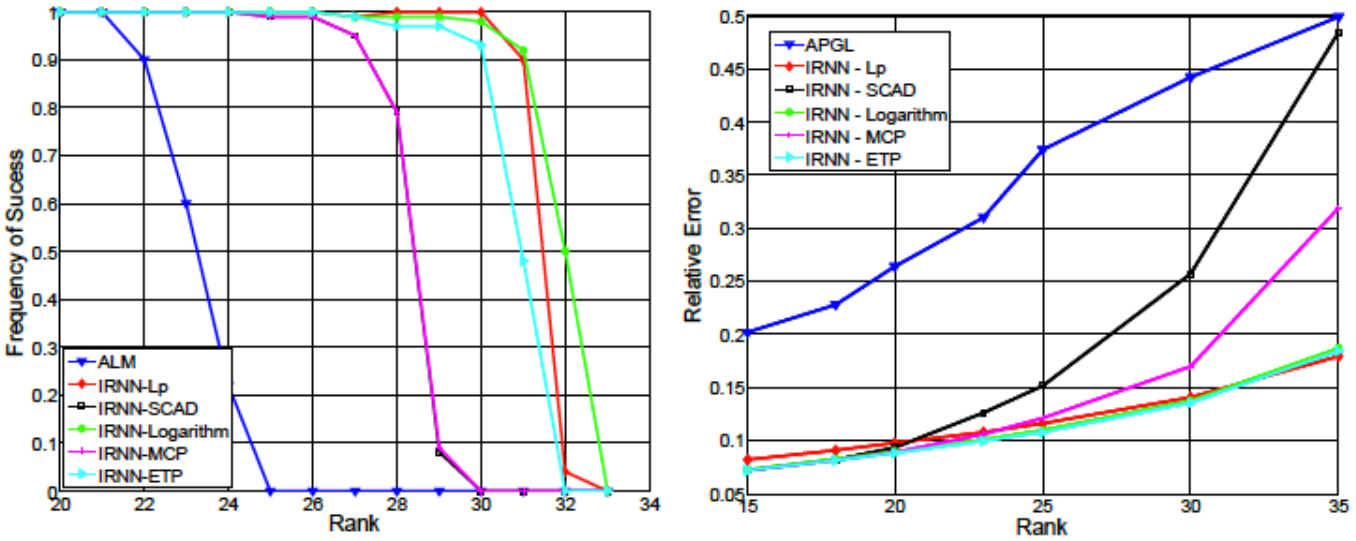
$$\begin{aligned}\mathbf{X}^{k+1} &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} w_i^k \sigma_i + f(\mathbf{X}^k) \\ &\quad + \langle \nabla f(\mathbf{X}^k), \mathbf{X} - \mathbf{X}^k \rangle + \frac{\mu}{2} ||\mathbf{X} - \mathbf{X}^k||_F^2 \\ &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} w_i^k \sigma_i + \frac{\mu}{2} \left\| \mathbf{X} - \left(\mathbf{X}^k - \frac{1}{\mu} \nabla f(\mathbf{X}^k) \right) \right\|_F^2\end{aligned}$$

It is nonconvex, but has closed form solution

 Lu et al. Generalized Nonconvex Nonsmooth Low-Rank Minimization. *CVPR 2014*

Nonconvex Algorithms

- Iteratively Reweighted Nuclear Norm (IRNN)



(a) random data without noise (b) random data with noise

Figure 3: Comparison of matrix recovery on (a) random data without noise, and (b) random data with noise.

 Lu et al. Generalized Nonconvex Nonsmooth Low-Rank Minimization. *CVPR 2014*

Nonconvex Algorithms

- Log-determinant Approximation
  - Model

$$\min_{Z} \quad rank(Z) \qquad \min_{Z} \sum_{i=1}^{n} h(\sigma_i(Z))$$
$$s.t \quad \mathcal{A}(Z) = b \qquad s.t. \quad \mathcal{A}(Z) = b$$

where $h()$ is a nonsmooth and nonconvex function

  - Idea: use log-determinant function to approximate rank function

$$F(Z) = logdet(I + Z^T Z) = \sum_{i=1}^{n} log(1 + \sigma_i^2(Z))$$

  - For LRR model

$$\min_{Z,E} logdet(I + Z^T Z) + \lambda \|E\|_l$$
$$s.t. \quad X = XZ + E$$

 Peng et al. Subspace clustering using log-determinant rank approximation. *KDD 2015*

Nonconvex Algorithms

- Log-determinant Approximation
  - Relaxed objective function

$$\min_{E,J,Z} \ logdet(I + J^T J) + \lambda ||E||_l \qquad s.t.\ X = XZ + E, Z = J$$

  - Update rules

$$\begin{aligned}
Z^{t+1} &= \arg\min_Z Tr[(Y_1^t)^T(J^t - Z)] + \frac{\mu^t}{2}\|J^t - Z\|_F^2 \\
&\quad + Tr[(Y_2^t)^T(X - XZ - E^t)] + \frac{\mu^t}{2}\|X - XZ - E^t\|_F^2, \\
J^{t+1} &= \arg\min_J logdet(I + J^T J) + \\
&\quad \frac{\mu^t}{2}\|J - (Z^{t+1} - \frac{Y_1^t}{\mu^t})\|_F^2, \\
E^{t+1} &= \arg\min_E \lambda||E||_l + Tr[(Y_2^t)^T(X - XZ^{t+1} - E)] \\
&\quad + \frac{\mu^t}{2}||X - XZ^{t+1} - E||_F^2.
\end{aligned}$$

 Peng et al. Subspace clustering using log-determinant rank approximation. *KDD 2015*

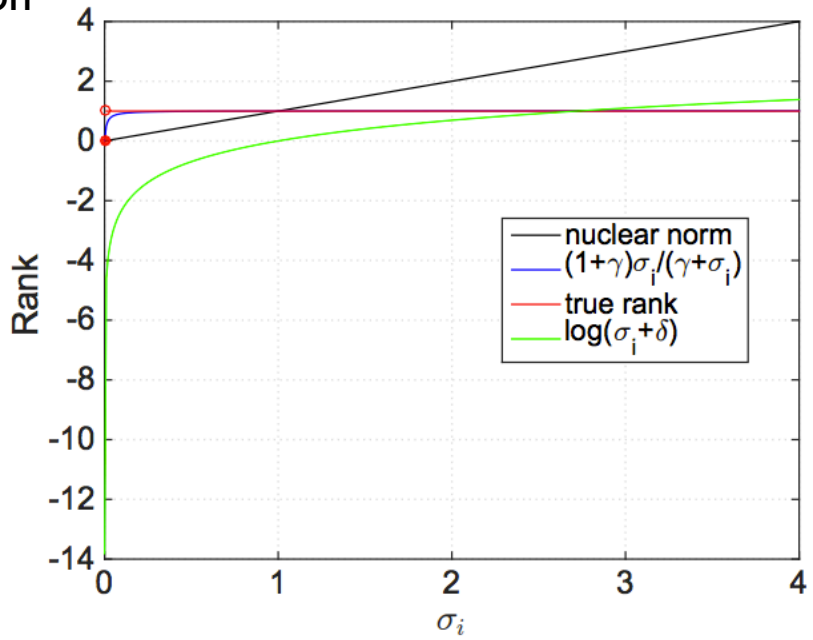
Nonconvex Algorithms

- Robust PCA via Nonconvex Rank Approximation
  - Model

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad s.t. \quad X = L + S \qquad \Rightarrow \qquad \min_{L,S} \|L\|_\gamma + \lambda \|S\|_l \quad s.t. \quad X = L + S$$

where $||\ ||_r$ denotes a rank approximation

$$\|L\|_\gamma = \sum_i \frac{(1+\gamma)\sigma_i(L)}{\gamma + \sigma_i(L)}, \quad \gamma > 0.$$



 Kang et al. *Robust PCA via Nonconvex Rank Approximation*. ICDM 2015

Nonconvex Algorithms

- Robust PCA via Nonconvex Rank Approximation
  - Augmented Lagrangian function

$$\mathcal{L}(L,S,Y,\mu) = \|L\|_\gamma + \lambda\|S\|_l + \langle Y, L+S-X\rangle + \frac{\mu}{2}\|L+S-X\|_F^2$$

  - Update L

$$L^{t+1} = \arg\min_L \|L\|_\gamma + \frac{\mu^t}{2}\left\|L - (X - S^t - \frac{Y^t}{\mu^t})\right\|_F^2$$

  - Update S

$$S^{t+1} = \arg\min_S \lambda\|S\|_l + \frac{\mu^t}{2}\left\|S - (X - L^{t+1} - \frac{Y^t}{\mu^t})\right\|_F^2$$

 Kang et al. *Robust PCA via Nonconvex Rank Approximation*. ICDM 2015

Convex and Nonconvex Algorithms

- Summary
  - Both of them mainly try to approximate rank function and L0 norm
  - Nonconvex algorithms usually obtain better performance than nuclear norm based convex algorithms
  - Challenge: Scalability when dealing with large matrices

Scalable Algorithms: Randomized

- Fast Randomized SVT
  - Rank Minimization Problem

$$X^* = \arg\min_X f(X) + \text{rank}(X) \quad \Rightarrow \quad X^* = \arg\min_X f(X) + \tau\|X\|_*$$

  - Unavoidable subproblem

$$S_\tau(Y) = \min_X \tfrac{1}{2}\|Y - X\|_F^2 + \tau\|X\|_*$$

  - Closed-form solution: Singular Value Thresholding (SVT)

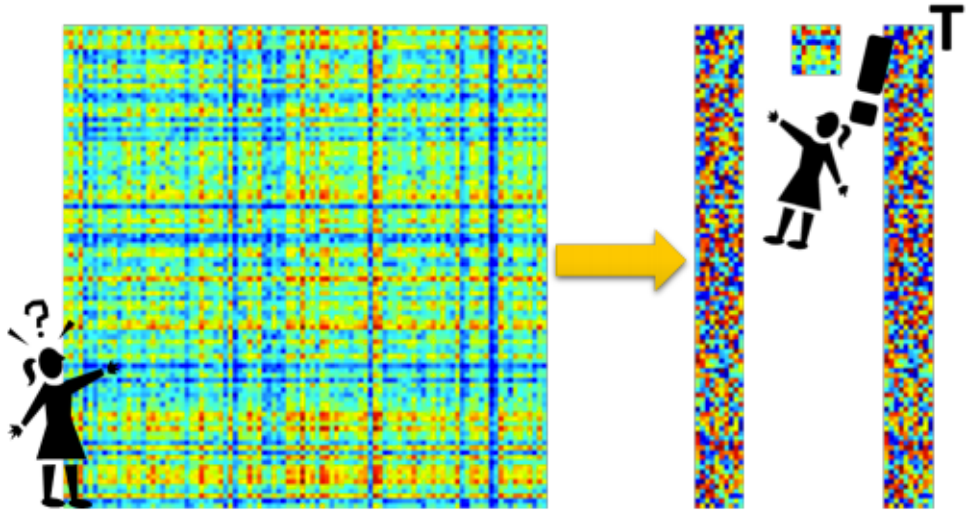
$$\mathbf{X}^* = \mathbb{S}_\tau(\mathbf{A}) = \mathbf{U}_\mathbf{A}\mathcal{S}_\tau(\mathbf{\Sigma}_\mathbf{A})\mathbf{V}_\mathbf{A}^\top, \qquad (3)$$

where $\mathcal{S}_\tau(x) = \text{sgn}(x) \cdot \max(|x| - \tau, 0)$ *is the soft shrinkage operator [10], and* $\mathbf{U}_\mathbf{A}\Sigma_\mathbf{A}\mathbf{V}_\mathbf{A}^\top$ *is the SVD of* $\mathbf{A}$.

 Oh et al. *Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization*. CVPR 2015

Scalable Algorithms: Randomized

- Fast Randomized SVT



Proposition 1. *Let* $\mathbf{A} = \mathbf{QB} \in \mathbb{R}^{m \times n}$, *where* $\mathbf{Q} \in \mathbb{R}^{m \times n}$ *has orthonormal columns. Then,*

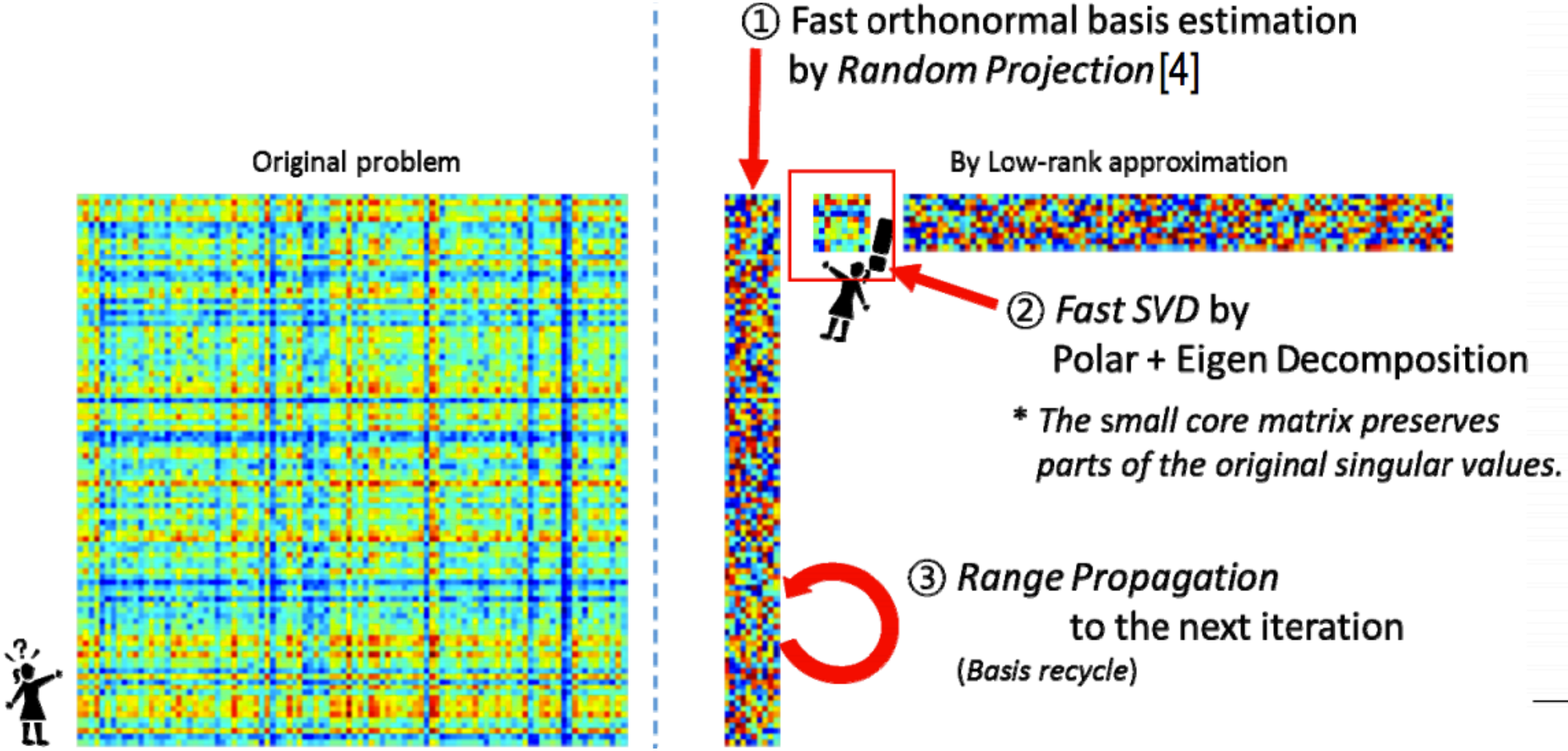
$$\mathbb{S}_\tau(\mathbf{A}) = \mathbf{Q}\,\mathbb{S}_\tau(\mathbf{B}), \tag{4}$$

where $\mathbb{S}_\tau(\cdot)$ *is the SVT operator.*

 Oh et al. *Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization*. CVPR 2015

Scalable Algorithms: Randomized

- Fast Randomized SVT



 Oh et al. *Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization*. CVPR 2015

Scalable Algorithms: Randomized

- Fast Randomized SVT

Theorem 1 (Average error bound of the approximate SVT). *Let $\mathbb{S}_\tau(\cdot)$ be the SVT operator. Then, the average error satisfies the following inequality.*

$$\mathbb{E}\|\mathbb{S}_\tau(\mathbf{A}) - \mathbb{S}_\tau(\hat{\mathbf{A}}_k)\|_F^2 \leq (1 + k/(p-1)) \cdot \left(\sum_{j>k} \sigma_j^2(\mathbf{A})\right) - G(\mathbf{A}),$$

where $G(\mathbf{A}) = \sum_{j>k} \min(\sigma_j(\mathbf{A}), \tau)^2 \geq 0.$

 Oh et al. *Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization*. CVPR 2015

Scalable Algorithms: Randomized

- Fast Randomized SVT

Algorithms	#NM	SIT	TT	SPG	ERR
$\text{iALM}_{\text{econSVD10}}$	23	947.9	21668	–	1.8e-7
$\text{iALM}_{\text{econSVD14}}$	23	46.5	1069	20×	1.8e-7
$\text{iALM}_{\text{LTSVD}}$ [5]	41	1.7	70	312×	4.6e-7
$\text{iALM}_{\text{BLWS}}$ [19]	24	2.2	53	405×	4.8e-7
$\text{iALM}_{\text{FSVT}}$ [2]	23	110.0	2527	9×	1.8e-7
$\text{iALM}_{\text{RSVD}}$ [11]	23	3.6	81	268×	1.8e-7
$\text{iALM}_{\text{FRSVT}}$ (ours)	23	1.5	33	665×	1.8e-7
$\text{iALM}_{\text{FRSVT-RP}}$ (ours)	23	1.3	30	716×	1.8e-7

- #NM: The number of NNM,
- TT: Total elapsed time (sec),
- SIT: Elapsed time (sec) of a single iteration,
- SPG: Speed-up gain against the baseline,
- ERR: $\|\mathbf{A}_{GT} - \hat{\mathbf{A}}\|_F / \|\mathbf{A}_{GT}\|_F$.

Table 1. Quantitative Comparisons on RPCA. 4000×4000 matrices are used. The results of other sizes and convergence graphs are shown in the supplementary material. In iALM procedure, #NM is the number of iteration to solve the NNM subproblem, which corresponds to the total number of iterations.

 Oh et al. *Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization*. CVPR 2015

Scalable Algorithms: Distributed

- Divide-Factor-Combine LRR (DFC-LRR)

$$\begin{aligned} \text{minimize} \quad & \|Z\|_* + \lambda \|E\|_{2,1} \\ \text{subject to} \quad & X = XZ + E \end{aligned}$$

  - D Step: Divide input matrix into submatrices
    - DFC-LRR randomly partitions the columns of X into t l-column submatrices, $\{C_1, \ldots, C_t\}$.
  - F Step: Factor submatrices in parallel
    - The i-th subproblem

$$\begin{aligned} \text{minimize} \quad & \|Z_i\|_* + \lambda \|E_i\|_{2,1} \\ \text{subject to} \quad & C_i = XZ_i + E_i \end{aligned}$$

Ameet Talwalkar, Lester W. Mackey, Yadong Mu, Shih-Fu Chang, Michael I. Jordan: *Distributed Low-Rank Subspace Segmentation*. ICCV 2013.

Scalable Algorithms: Distributed

- Divide-Factor-Combine LRR (DFC-LRR)

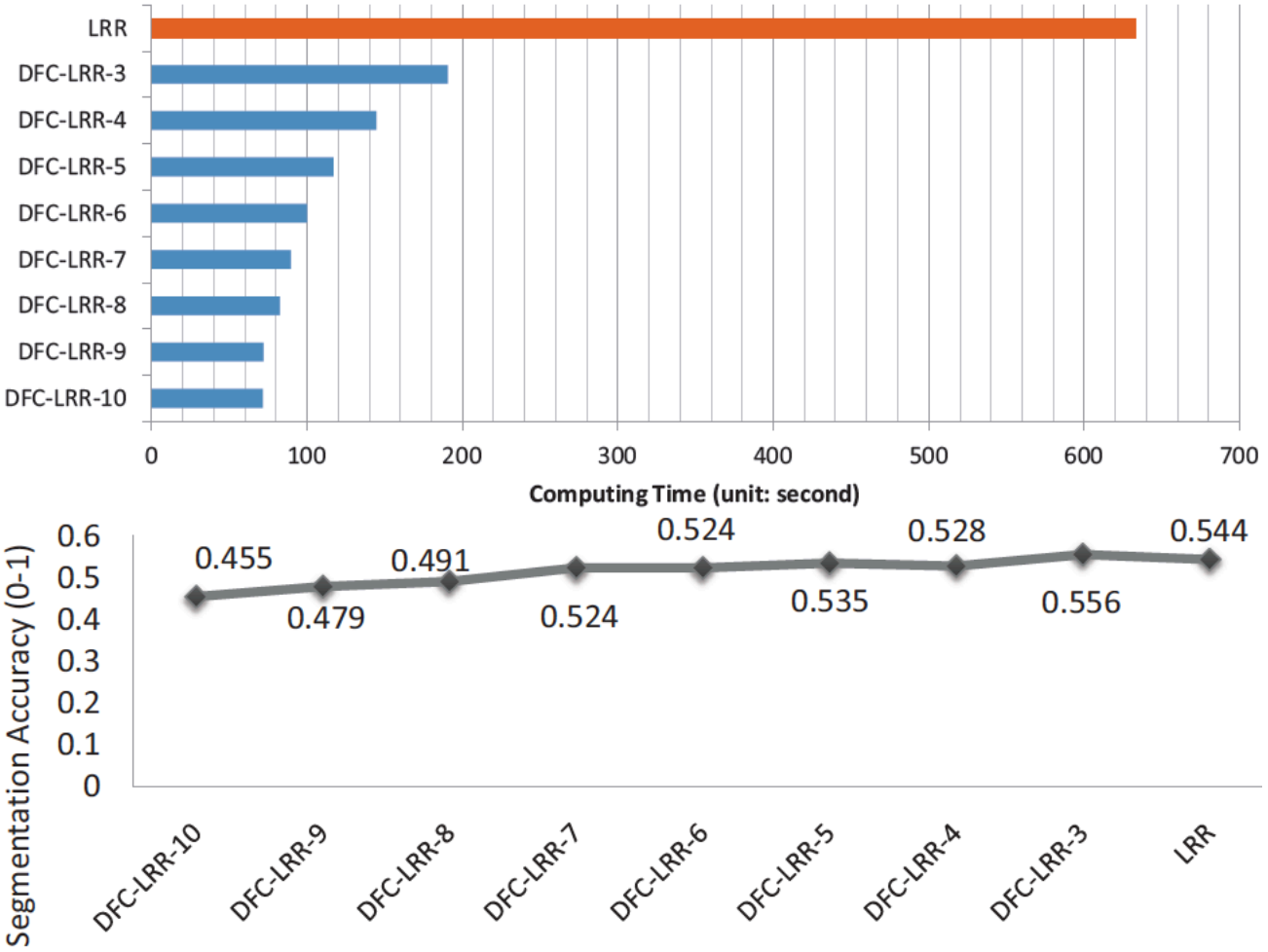
$$\begin{aligned} \text{minimize} \quad & \|Z\|_* + \lambda \|E\|_{2,1} \\ \text{subject to} \quad & X = XZ + E \end{aligned}$$

  - C Step: Combine submatrix estimates
    - DFC-LRR use *column projection* technique to generates a final approximation Z^p, by projecting $[Z^1, \ldots, Z^t]$ onto the column space of Z^1.
  - Complexity
    - Traditional nuclear-norm based methods usually take $\Omega(mnk_M)$ per iteration, due to the truncated SVD operation
    - DFC-LRR reduces it to $\mathrm{O}(mlk_{C_i})$

Ameet Talwalkar, Lester W. Mackey, Yadong Mu, Shih-Fu Chang, Michael I. Jordan: *Distributed Low-Rank Subspace Segmentation*. ICCV 2013.

Scalable Algorithms: Distributed

- Divide-Factor-Combine LRR (DFC-LRR)



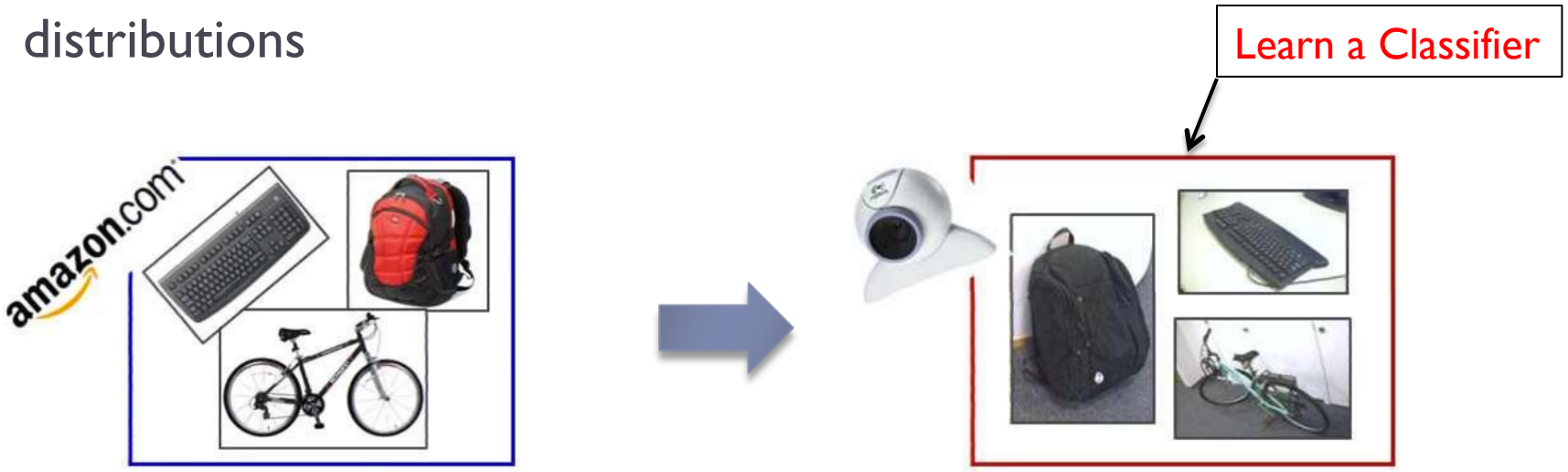
Ameet Talwalkar, Lester W. Mackey, Yadong Mu, Shih-Fu Chang, Michael I. Jordan: *Distributed Low-Rank Subspace Segmentation*. ICCV 2013.

Outline

- Optimization and Analysis
  - Convex Algorithms
  - Nonconvex Algorithms
- Scalable Algorithms
  - Randomized Algorithms
  - Distributed Algorithms
- More Models
  - Low-Rank Models for Domain Adaptation
  - Low-Rank Models for Multi-view Learning

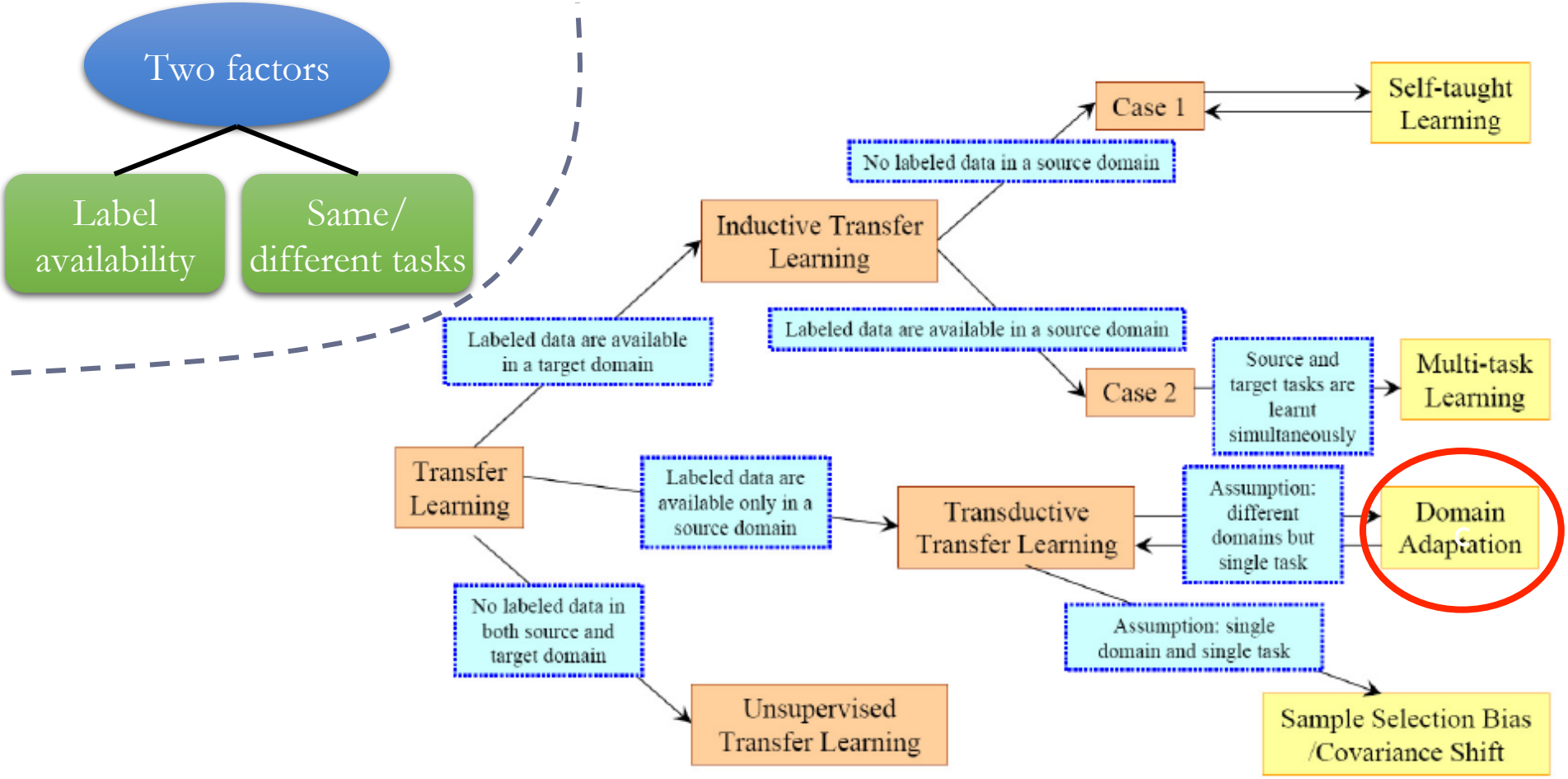
Low-Rank Models for Domain Adaptation

- Domain Adaptation vs. Traditional Learning Methods
  - Traditional machine learning methods: training and test data are from the same distribution
  - Domain Adaptation: training and test data are from different distributions



Low-Rank Models for Domain Adaptation

- Domain Adaptation: Related Concepts



 Sinno Jialin Pan, Qiang Yang: *A Survey on Transfer Learning*. IEEE Trans. KDE, 2010

Low-Rank Models for Domain Adaptation

- Models
  - Robust Domain Adaptation with Low-Rank Reconstruction (RDALR) [Jhuo et al., CVPR'12]
  - Low-Rank Transfer Subspace Learning (LTSL) [Shao et al., ICDM'12, IJCV'14]
  - Missing Modality Transfer Learning via Latent Low-Rank Constraints [Ding et al., AAAI'14, IEEE TIP'15]
  - Deep Low-Rank Coding for Domain Adaptation [Ding et al., IJCAI'15]

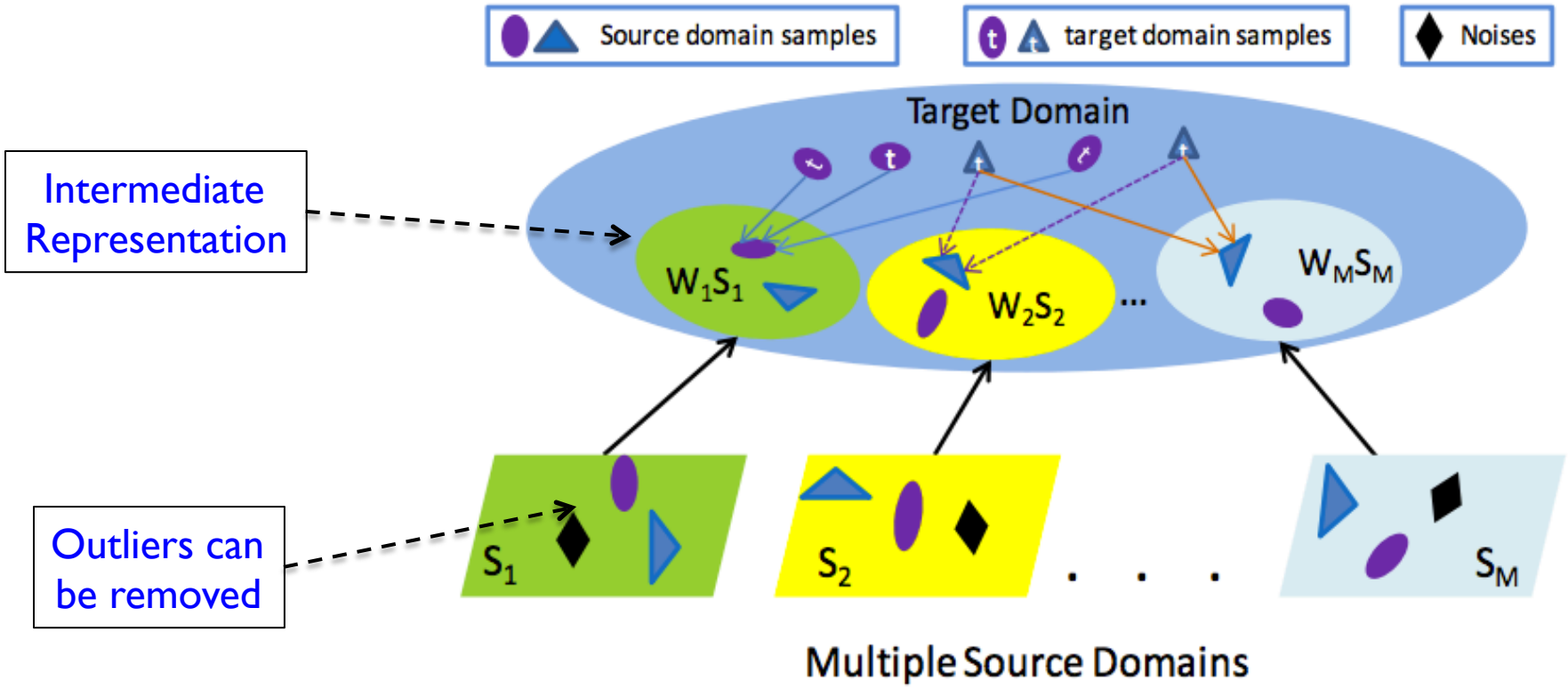
Low-Rank Models for Domain Adaptation

- Robust Domain Adaptation with Low-Rank Reconstruction (RDALR)
  - Goal: To effectively adapt the sample distribution of source domain to match that of the target domain.
  - Transform the samples in source domain into an intermediate representation.
  - Each transformed source sample can be linearly reconstructed by the samples of the target domain.
  - The reconstruction coefficient matrix w.r.t. all source samples is enforced to be low-rank.

I-Hong Jhuo, Dong Liu, D.T. Lee, Shih-Fu Chang. *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR 2012

Low-Rank Models for Domain Adaptation

- Robust Domain Adaptation with Low-Rank Reconstruction (RDALR)



I-Hong Jhuo, Dong Liu, D.T. Lee, Shih-Fu Chang. *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR 2012

Low-Rank Models for Domain Adaptation

- RDALR: Problem Definition
  - Notations
    - W: Transformation matrix
    - $S = [s_1, s_2, \ldots, s_n]$: Sample set in source domain
    - $T = [t_1, t_2, \ldots, t_n]$: Sample set in target domain
    - Z: Reconstruction coefficient matrix
  - Goal: Find a transformation matrix W to transform the source domain into an intermediate representation matrix.
  - Reconstruction: each transformed source sample will be linearly reconstructed by the target samples.

$$WS = TZ$$

Why Z should be low-rank?
- Capture structure information in source domain
- Remove outliers

I-Hong Jhuo, Dong Liu, D.T. Lee, Shih-Fu Chang. *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR 2012

Low-Rank Models for Domain Adaptation

- RDALR
  - **Single Source** Domain Adaptation

$$\min_{W,Z,E} \quad \operatorname{rank}(Z) + \alpha\|E\|_{2,1}, \qquad \min_{W,Z,E} \quad \|Z\|_* + \alpha\|E\|_{2,1},$$
$$s.t. \quad WS = TZ + E, \qquad s.t. \quad WS = TZ + E,$$
$$WW^\top = I, \qquad WW^\top = I,$$

  - Optimization: ADMM.
  - Training: Once obtain the optimal solution $(\hat{W}, \hat{Z}, \hat{E})$, the source data can be transformed into the target domain as

$$\hat{W}S - \hat{E} = [\hat{W}s_1 - \hat{e}_1, \ldots, \hat{W}s_n - \hat{e}_n]$$

    Finally, the transformed source samples are mixed with targeted samples for training classifiers.

I-Hong Jhuo, Dong Liu, D.T. Lee, Shih-Fu Chang. *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR 2012

Low-Rank Models for Domain Adaptation

- RDALR
  - **Multiple Source** Domain Adaptation

$$\min_{Z_i,E_i,W_i} \sum_{i=1}^{M} (\|Z_i\|_* + \alpha\|E_i\|_{2,1}) + \beta\|Q\|_*$$

$$s.t. \quad W_i S_i = T Z_i + E_i,$$

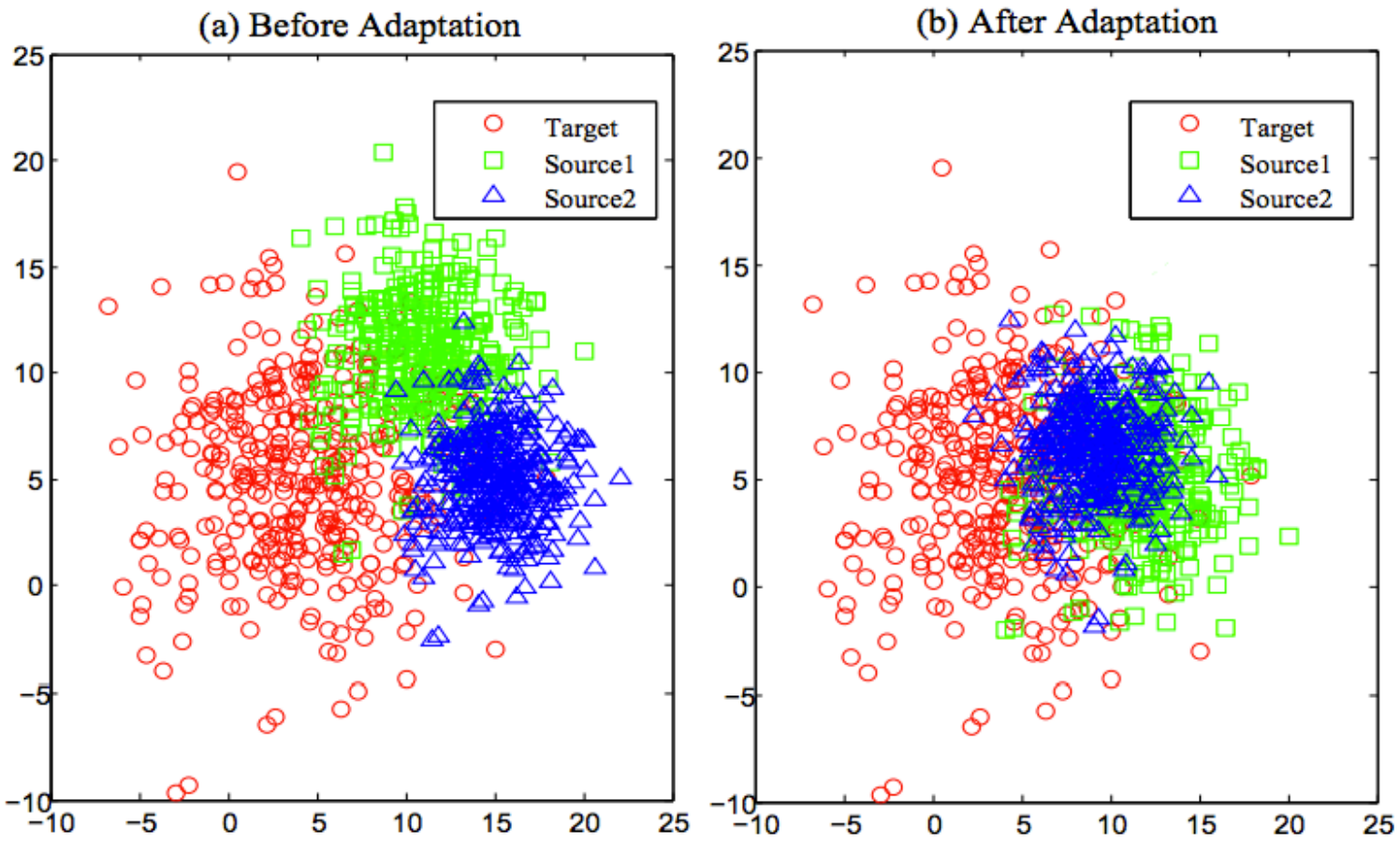
$$W_i W_i^\top = I, i = 1, ..., M,$$

  - For each source domain, low-rank and sparse constrains are still used as in the single source domain case
  - $Q=[W_1S_1 \mid W_2S_2 \mid \ldots \mid W_MS_M]$, where W_iS_i denotes the *i-th* transformed source domain.
  - Q is used to discover the low-rank structure across domains

I-Hong Jhuo, Dong Liu, D.T. Lee, Shih-Fu Chang. *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR 2012

Low-Rank Models for Domain Adaptation

- RDALR: A Toy Example



I-Hong Jhuo, Dong Liu, D.T. Lee, Shih-Fu Chang. *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR 2012

Low-Rank Models for Domain Adaptation

- Low-Rank Transfer Subspace Learning (LTSL)
  - Goal: Align the source and target data in a learned subspace by minimizing the reconstruction error

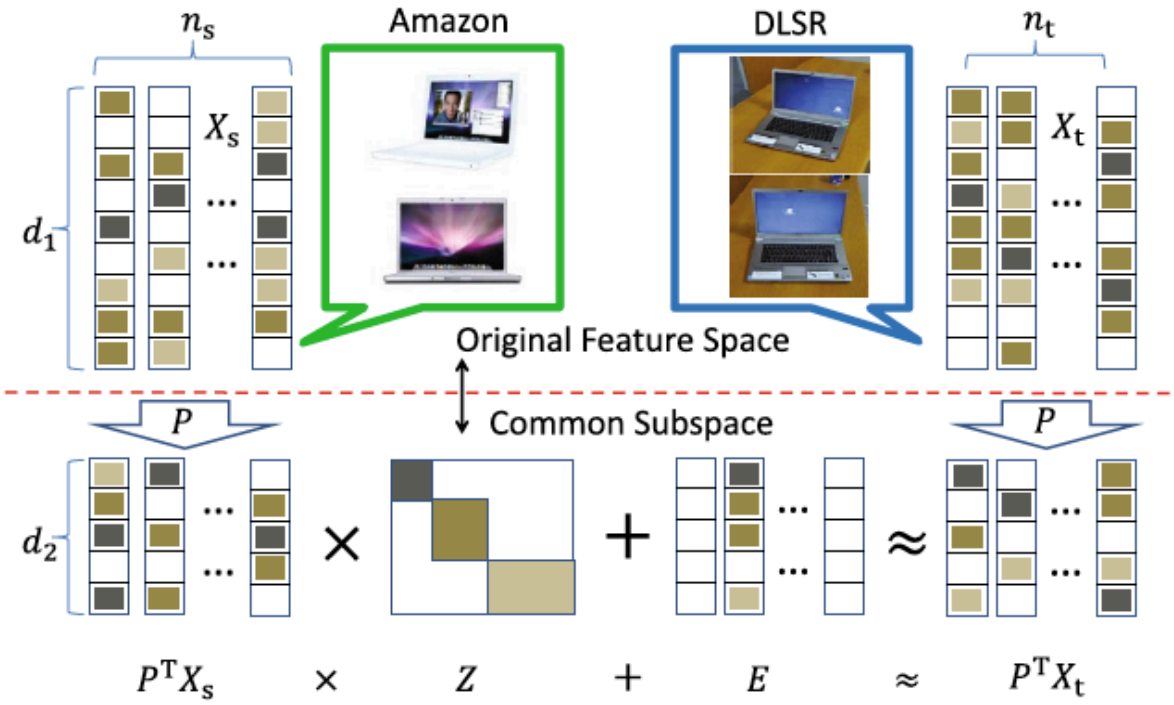
    (1) How to reconstruct?

    (2) How to avoid negative transfer?
  - Incorporate *low-rank* constraint to enforce locality aware reconstruction
  - Low-rank and sparse constraints are used to compensate for the noisy data
  - LTSL generalizes traditional subspace learning techniques to transfer learning scenario

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Framework



d_1 – dimensionality of original feature space
d_2 – dimensionality of common subspace
n_s, n_t – number of samples in source and target domains

X_s – data in the source domain
X_t – data in the target domain
P – projection matrix
Z – low-rank coefficient matrix
E – sparse matrix

- LTSL iteratively seeks appropriate subspace P that satisfies the above constraint

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Problem Definition
  - Given test data (target domain) and training data (source domain), the **goal** is to find a discriminative subspace P where each test datum can be linearly represented by the data from subspace $\mathcal{S}_i$ in source domain:

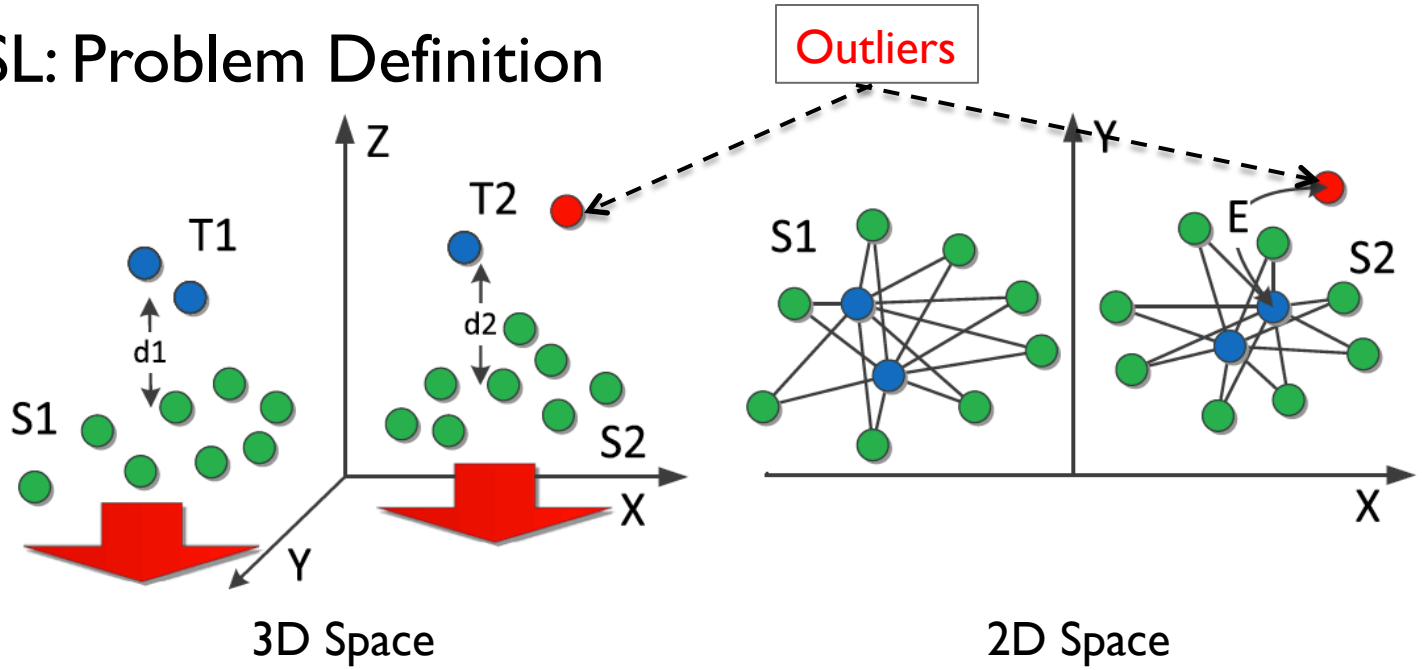
$$\min_{P,Z} F(P, X_s) + \lambda_1 \text{rank}(Z)$$
$$\text{s.t., } P^T X_t = P^T X_s Z,$$

  - $F(P, X_S)$: generalized subspace learning loss function
  - Z: coefficient matrix
  - Subspace P: a projection matrix
  - Subspace $\mathcal{S}_i$: a group of data that lie in P

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Problem Definition



- $\mathcal{S}_1$, $\mathcal{S}_2$: two independent subspaces in source domain
- $\mathcal{T}_1$, $\mathcal{T}_2$: two independent subspaces in target domain
- d_1, d_2: (large) distances between two domains

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Model and Theoretical Analysis
  - Objective function $\min_{P,Z,E} F(P, X_s) + \lambda_1 \|Z\|_* + \lambda_2 \|E\|_{2,1}$

$$\text{s.t.}, P^T X_t = P^T X_s Z + E.$$

  - Reconstruction Performance of subspace P

Theorem 2 *Given P, Z learned from problem in Eq. (5), energy function $F(P, \mathcal{S}_i)$ and $F(P, \mathcal{T}_i)$[1], where $\mathcal{S}_i$ is a subspace in the source domain and $\mathcal{T}_i$ is a subspace in the target domain that can be linearly represented by $\mathcal{S}_i$ using a coefficient vector Z_i. Suppose $|\mathcal{S}_i| = |\mathcal{T}_i|$, i.e., Z_i is a square matrix. Then $F(P, \mathcal{T}_i)$ is bounded by:*

$$F(P, \mathcal{T}_i) \geq F(P, \mathcal{S}_i) \|\widehat{Z}_i^{-1}\|_F^{-2} + \xi, \tag{8}$$

where $\widehat{Z}_i = Z_i + \gamma \mathbf{I}$ is an invertible matrix, $\gamma \mathbf{I}$ is a small perturbation term to make sure $\widehat{Z}_i$ is non-singular, and $\xi = F(P, \mathcal{T}_i) - F(P, \widehat{\mathcal{T}}_i)$. $\widehat{\mathcal{T}}_i$ represents the data in the target domain perturbed by $\widehat{Z}_i$.

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Model and Theoretical Analysis
  - Objective function

$\mathrm{Tr}(P^T U_1 P)$

$$\min_{P,Z,E} F(P, X_s) + \lambda_1 \|Z\|_* + \lambda_2 \|E\|_{2,1}$$
$$\text{s.t., } P^T X_t = P^T X_s Z + E.\ P^T U_2 P = \mathbf{I}$$

  - U_1 and U_2 are selected based on subspace learning methods

Table 1 Examples of several objective functions for LTSL

Method	U_1	U_2
PCA	$-\Sigma$	$\mathbf{I}_{d_1}$
LDA	S_w	S_b
LPP[a]	$X_s \mathcal{L} X_s^T$	$X_s D X_s^T$
NPE	$X_s(\mathbf{I}_{n_s} - W)^T(\mathbf{I}_{n_s} - W)X_s^T$	$X_s X_s^T$
MFA[b]	$X_s(D - W)X_s^T$	$X_s(D_p - W_p)X_s^T$
DLA[c]	$X_s L X_s^T$	$\mathbf{I}_{d_1}$

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Optimization
  - Objective function

$$\min_{P,Z,E} F(P, X_s) + \lambda_1 \|Z\|_* + \lambda_2 \|E\|_{2,1}$$
$$\text{s.t., } P^T X_t = P^T X_s Z + E.$$

  - Relaxed formulation

$$\min_{Z,E,J,P} F(P, X_s) + \lambda_1 \|J\|_* + \lambda_2 \|E\|_{2,1}$$
$$\text{s.t., } P^T X_t = P^T X_s Z + E,$$
$$\mathbf{1}_{n_s}^T Z = \mathbf{1}_{n_t}^T, P^T U_2 P = \mathbf{I}_{d_2}, Z = J,$$

  - Augmented Lagrange function

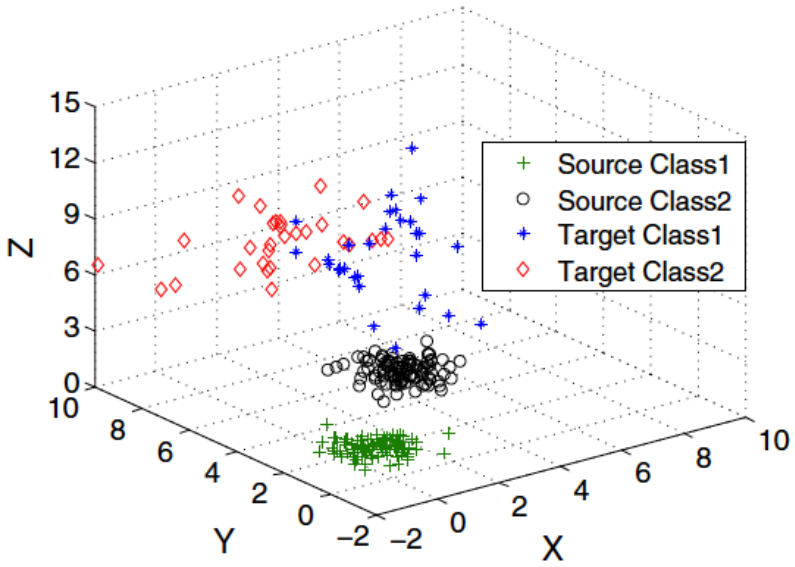
$$\begin{aligned}\mathcal{L} = & F(P, X_s) + \lambda_1 \|J\|_* + \lambda_2 \|E\|_{2,1} \\ & + < Y_1, P^T X_t - P^T X_s Z - E > + < Y_2, Z - J > \\ & + < Y_3, \mathbf{1}_{n_s}^T Z - \mathbf{1}_{n_t}^T > + < Y_4, P^T U_2 P - \mathbf{I}_{d_2} > \\ & + \frac{\mu}{2}(\|P^T X_t - P^T X_s Z - E\|_F^2 + \|Z - J\|_F^2 \\ & + \|\mathbf{1}_{n_s}^T Z - \mathbf{1}_{n_t}^T\|_F^2 + \|P^T U_2 P - \mathbf{I}_{d_2}\|_F^2), \end{aligned} \quad (11)$$

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

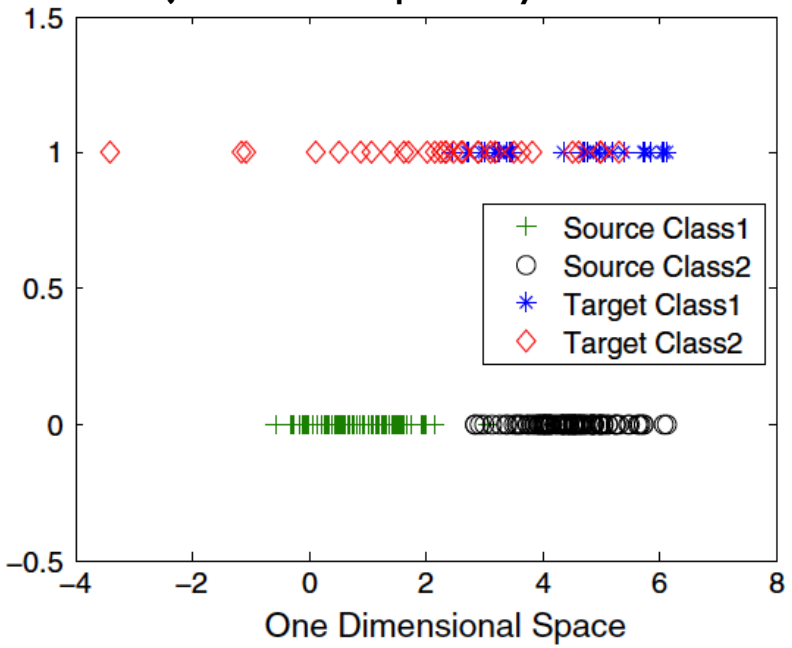
Low-Rank Models for Domain Adaptation

- LTSL: Toy Example 1

Original 2D Space



Projected 1D Space by LTSL-LDA



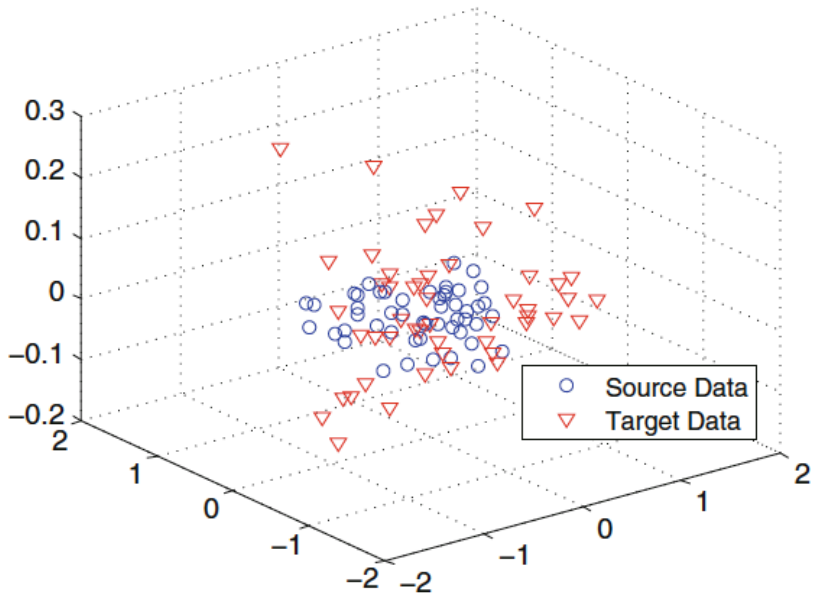
- In the 2D space, the decision boundary for target domain is unclear
- In the 1D space, the source and target data are well aligned

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

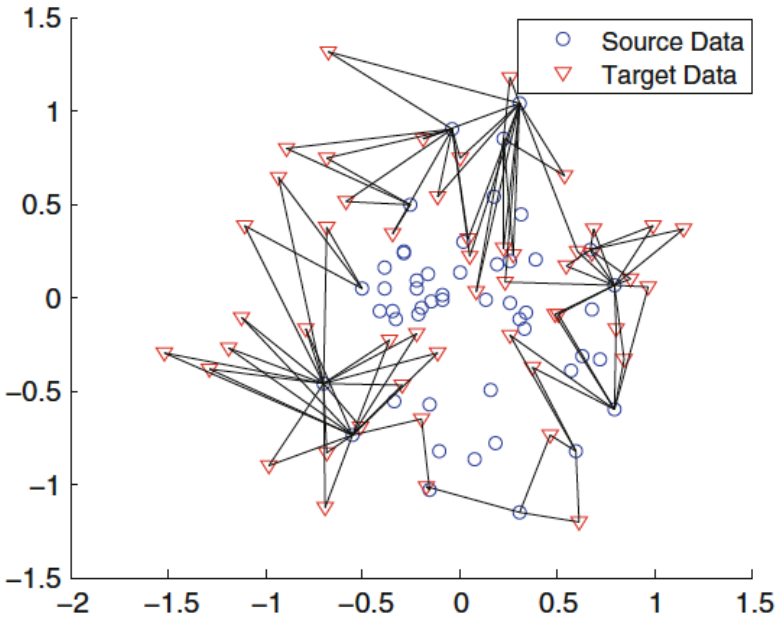
Low-Rank Models for Domain Adaptation

- LTSL: Toy Example 2

Original 3D Space



Projected 2D Space by LTSL-PCA



- Target data in 3D space cannot find ideal neighborhood from source
- In 2D space, all target data choose neighbor source points (boundary of source domain)

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

- LTSL: Benchmark Datasets (Source: YaleB; Target: CMU PIE)

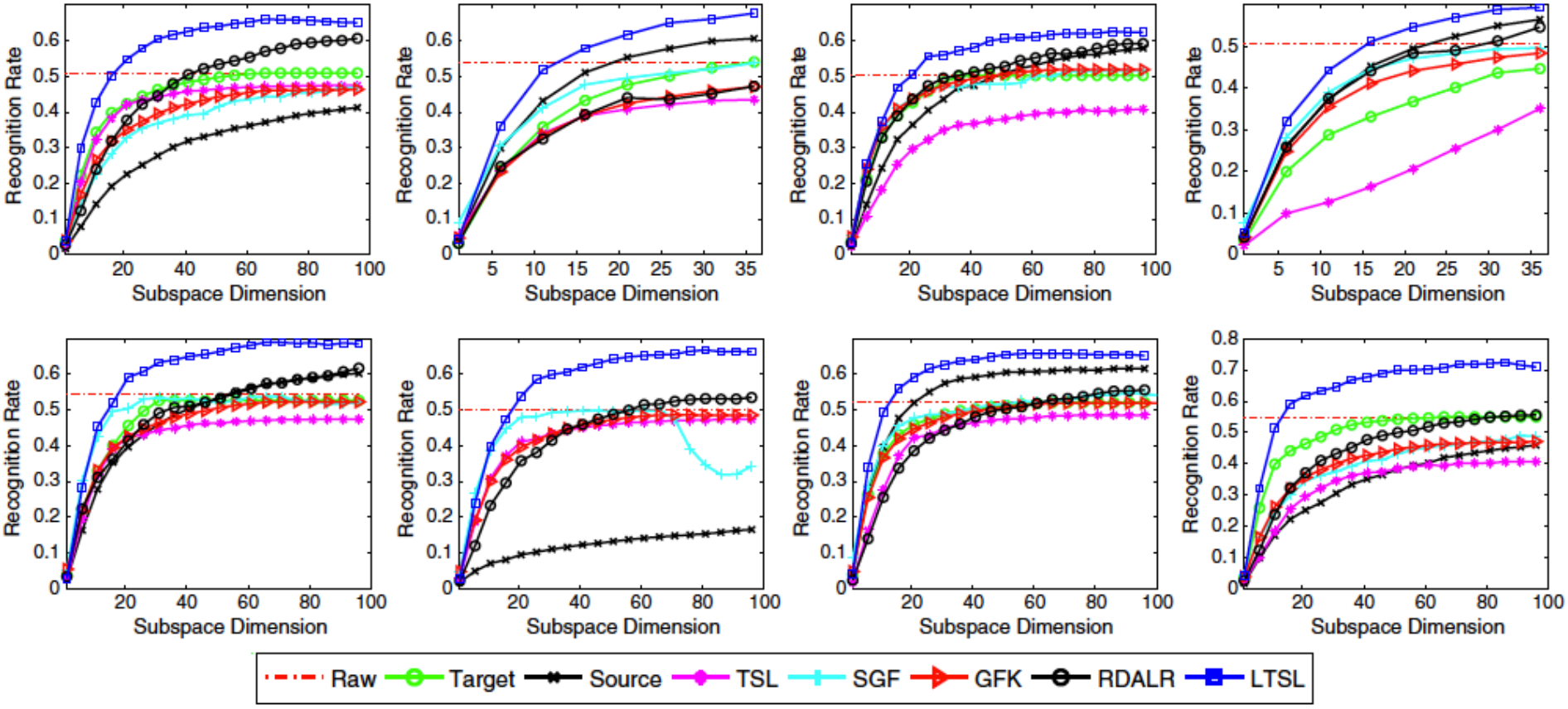
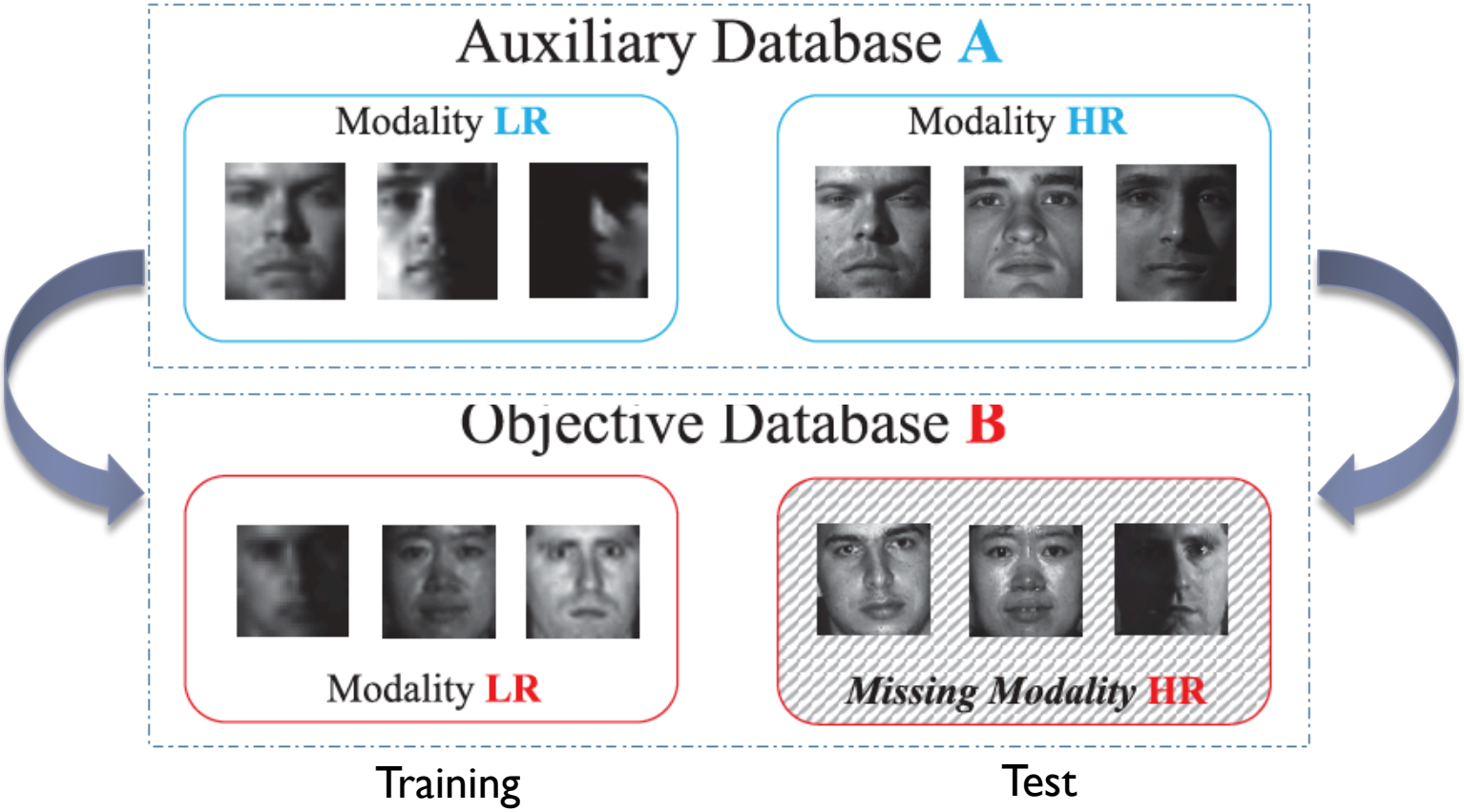


Fig. 8 Recognition results over different dimensions of problem Y2P. Subspace learning methods from *left* to *right*, from *top* to *bottom* are: PCA, LDA, ULPP, SLPP, UNPE, SNPE, MFA, DLA

Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu: *Low-Rank Transfer Subspace Learning*. ICDM 2012
Ming Shao, Dmitry Kit, Yun Fu: *Generalized Transfer Subspace Learning Through Low-Rank Constraint*. IJCV, 2014.

Low-Rank Models for Domain Adaptation

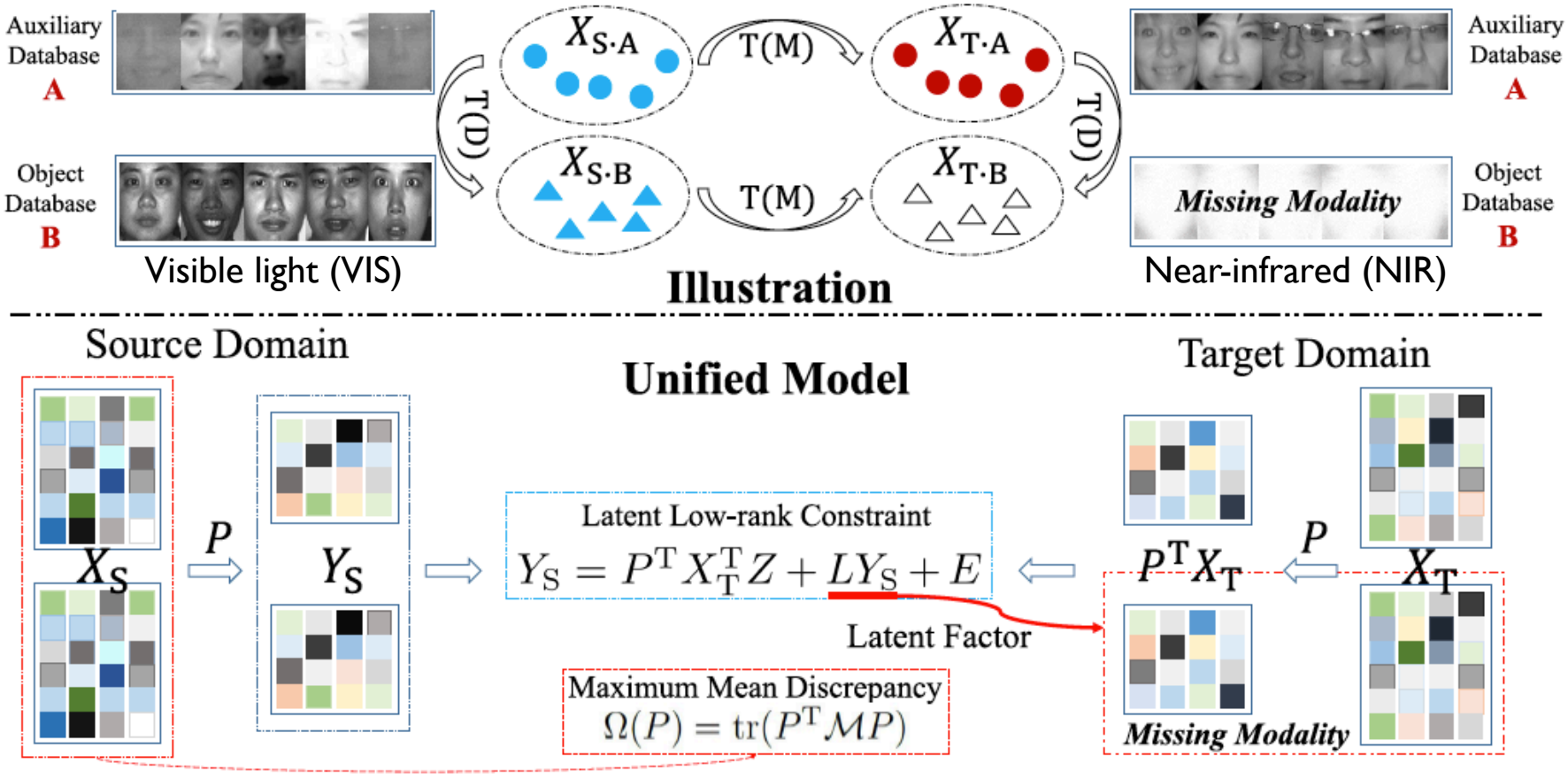
- Missing Modality Transfer Learning via Latent Low-Rank Constraints
  - Missing Modality Problem



Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

- Missing Modality Transfer Learning via Latent Low-Rank Constraints
  - Framework



Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

- Missing Modality Transfer Learning via Latent Low-Rank Constraints
  - Formulation

Auxiliary data:
$$\min_{Z_A} \ \text{rank}(Z_A), \quad \text{s.t.} \quad Y_{S\cdot A} = P_A^T X_{T\cdot A} Z_A$$

Target data:
$$\min_{Z_B} \ \text{rank}(Z_B), \quad \text{s.t.} \quad Y_{S\cdot B} = P_B^T X_{T\cdot B} Z_B$$

Missing

Unified objective:
$$\min_{Z_c} \ \text{rank}(Z_c), \quad \text{s.t.} \quad Y_S = P^T D_T Z_c,$$
$$\text{where } Y_S = [Y_{S\cdot A}, Y_{S\cdot B}],\ D_T = [X_{T\cdot A}, X_{T\cdot B}], \text{ and}$$
$$Z_c = \begin{bmatrix} Z_A & 0 \\ 0 & Z_B \end{bmatrix}.$$

Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

- Missing Modality Transfer Learning via Latent Low-Rank Constraints
  - Formulation

Objective function:

$$\min_{Z_c} \ \operatorname{rank}(Z_c), \quad \text{s.t.} \quad Y_S = P^T D_T Z_c,$$

Missing

$$\begin{aligned}
Y_S &= P^T D_T Z_c^* = [P^T X_{T\cdot A}, P^T X_{T\cdot B}] Z_c^* \\
&= [P^T X_{T\cdot A}, P^T X_{T\cdot B}] \begin{bmatrix} V_{T\cdot A} \\ V_{T\cdot B} \end{bmatrix} V_S^T \\
&= P^T X_{T\cdot A}(V_{T\cdot A} V_S^T) + P^T X_{T\cdot B} V_{T\cdot B} V_S^T \\
&= P^T X_{T\cdot A}(V_{T\cdot A} V_S^T) + U \Sigma V_{T\cdot B}^T V_{T\cdot B} V_S^T \\
&= P^T X_T Z + (U \Sigma V_{T\cdot B}^T V_{T\cdot B} \Sigma^{-1} U^T) Y_S \\
&= P^T X_T Z + L Y_S,
\end{aligned}$$

Latent factor

$$\min_{Z,\ L} \ \|Z\|_* + \|L\|_*,$$

$$\text{s.t.} \quad Y_S = P^T X_T Z + L Y_S$$

Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

- Missing Modality Transfer Learning via Latent Low-Rank Constraints
  - Objective Function

$$\min_{P,Z,L,E} \quad \|Z\|_* + \|L\|_* + \lambda\|E\|_1 + \alpha\|P\|_{2,1} + \beta\Omega(P)$$
$$\text{s.t.} \quad Y_S = P^T X_T Z + L Y_S + E, \quad P^T P = I_p, \qquad ($$

  - $\|P\|_{2,1}$: Group structure sparsity term to select important features
  - $\Omega(P)$: Push the means of auxiliary and target data sets to be closer

$$\Omega(P) = \|\frac{1}{n_a}\sum_{i=1}^{n_a} P^T x_i - \frac{1}{n_b}\sum_{j=n_a+1}^{n} P^T x_j\|_F^2$$
$$= \|P^T\mu_A - P^T\mu_B\|_F^2 = \mathrm{tr}(P^T\mathcal{M}P)$$

Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

- Missing Modality Transfer Learning via Latent Low-Rank Constraints

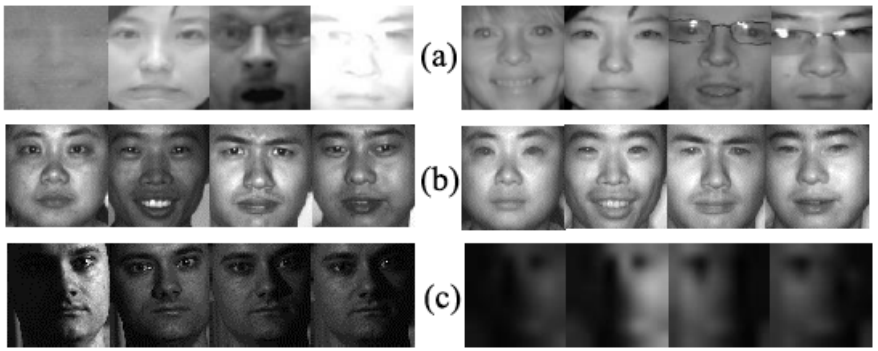
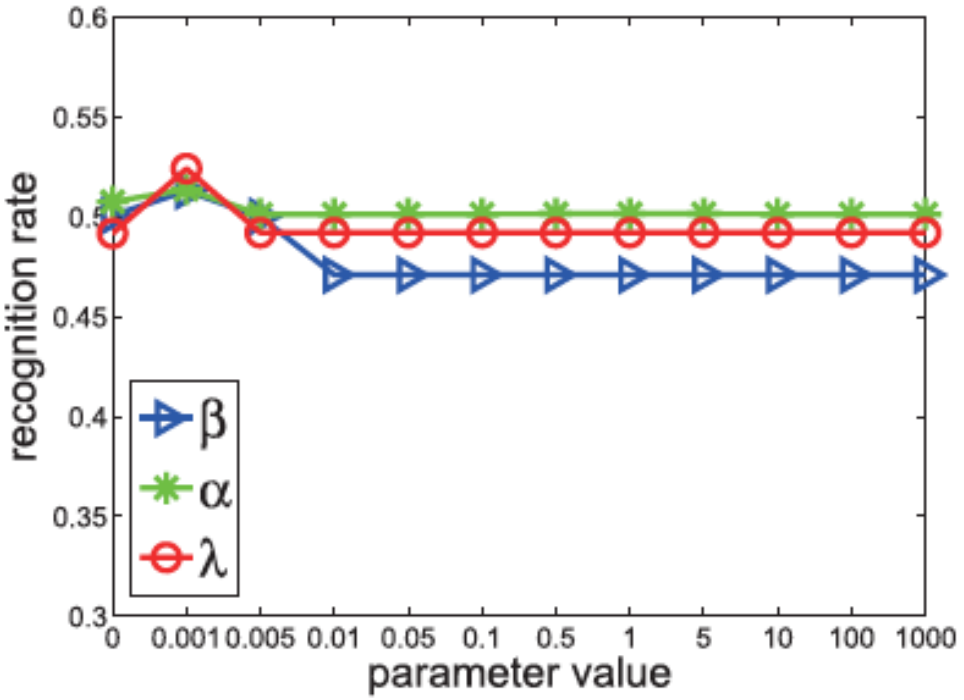


Fig. 3. Samples from (a) Oulu NIR-VIS database (Left: VIS image; Right: NIR image.), (b) BUAA NIR-VIS database (Left: VIS image; Right: NIR image.), (c) CMU-PIE database (Left: HR image; Right: LR image.)

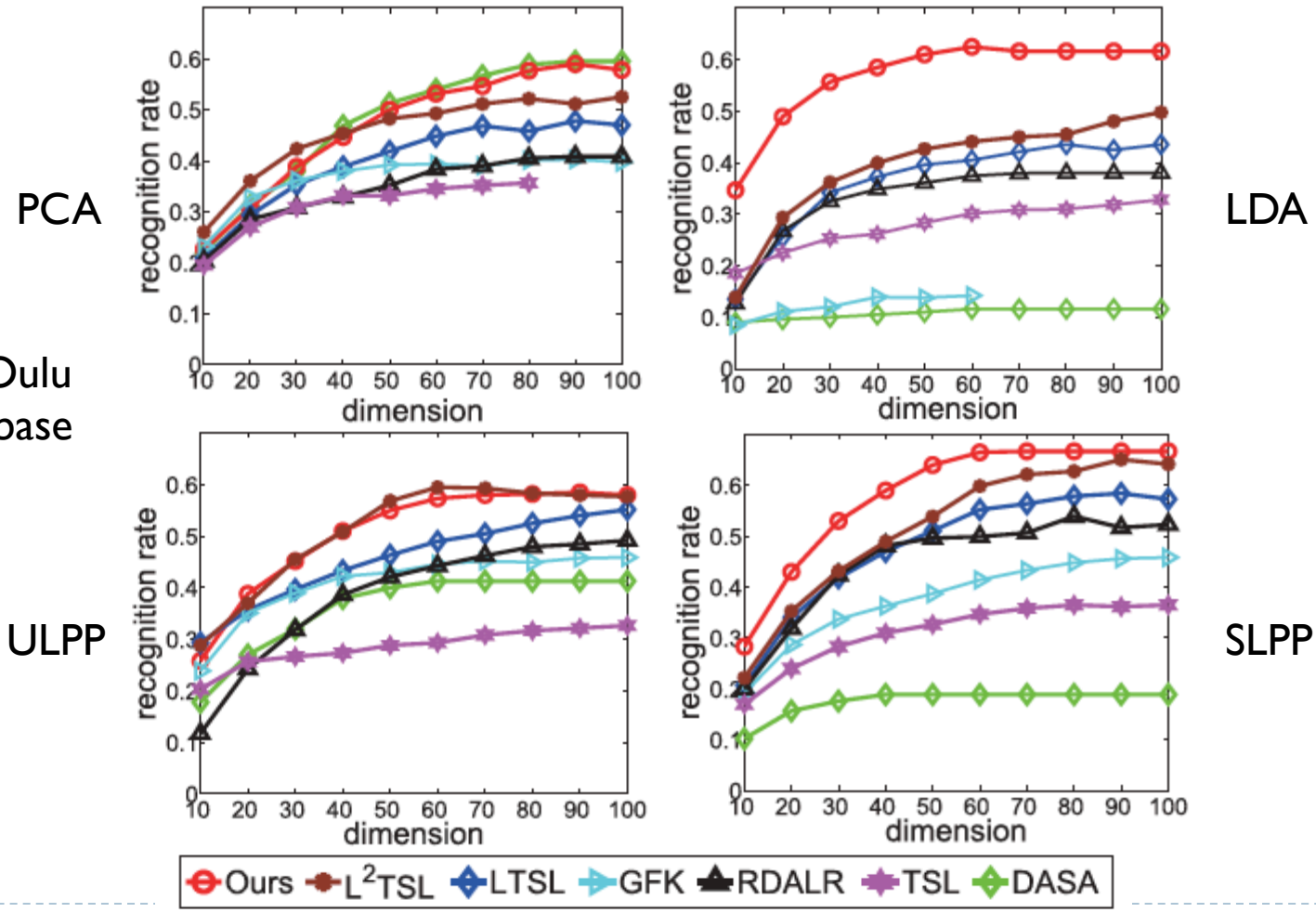


Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

- Missing Modality Transfer Learning via Latent Low-Rank Constraints

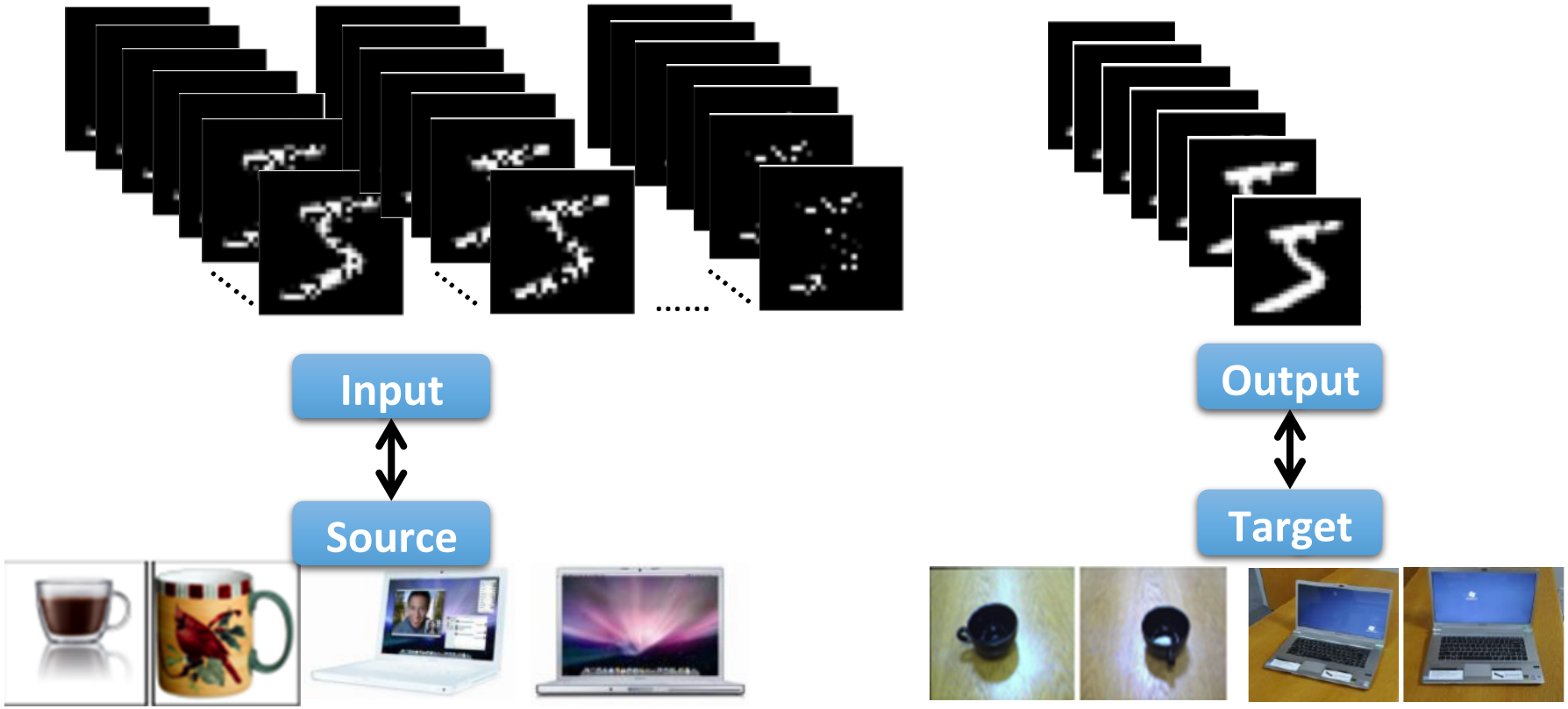
BUAA & Oulu Face Database



Zhengming Ding, Ming Shao, Yun Fu: *Missing Modality Transfer Learning via Low-Rank Constraint*. IEEE Trans. Image Processing, 2015.

Low-Rank Models for Domain Adaptation

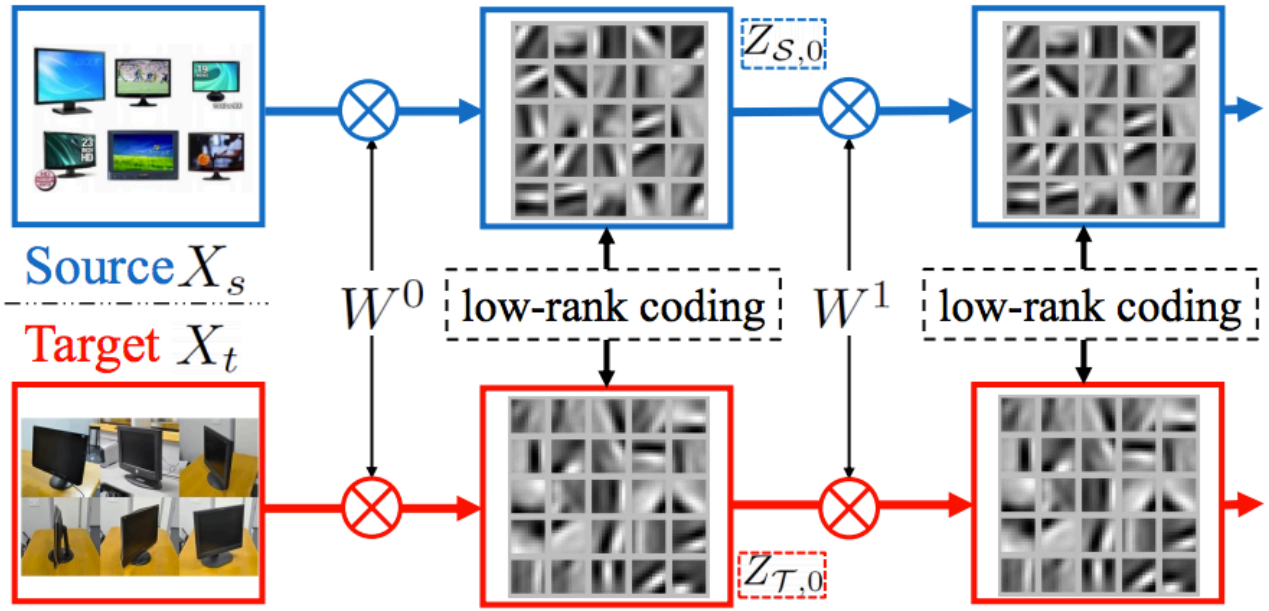
- Deep Low-Rank Coding for Domain Adaptation
  - From Denoising Auto-Encoder (DAE) to Domain Adaptation



69 Zhengming Ding, Ming Shao, Yun Fu: *Deep Low-Rank Coding Transfer Learning*. IJCAI 2015.

Low-Rank Models for Domain Adaptation

- Deep Low-Rank Coding for Domain Adaptation

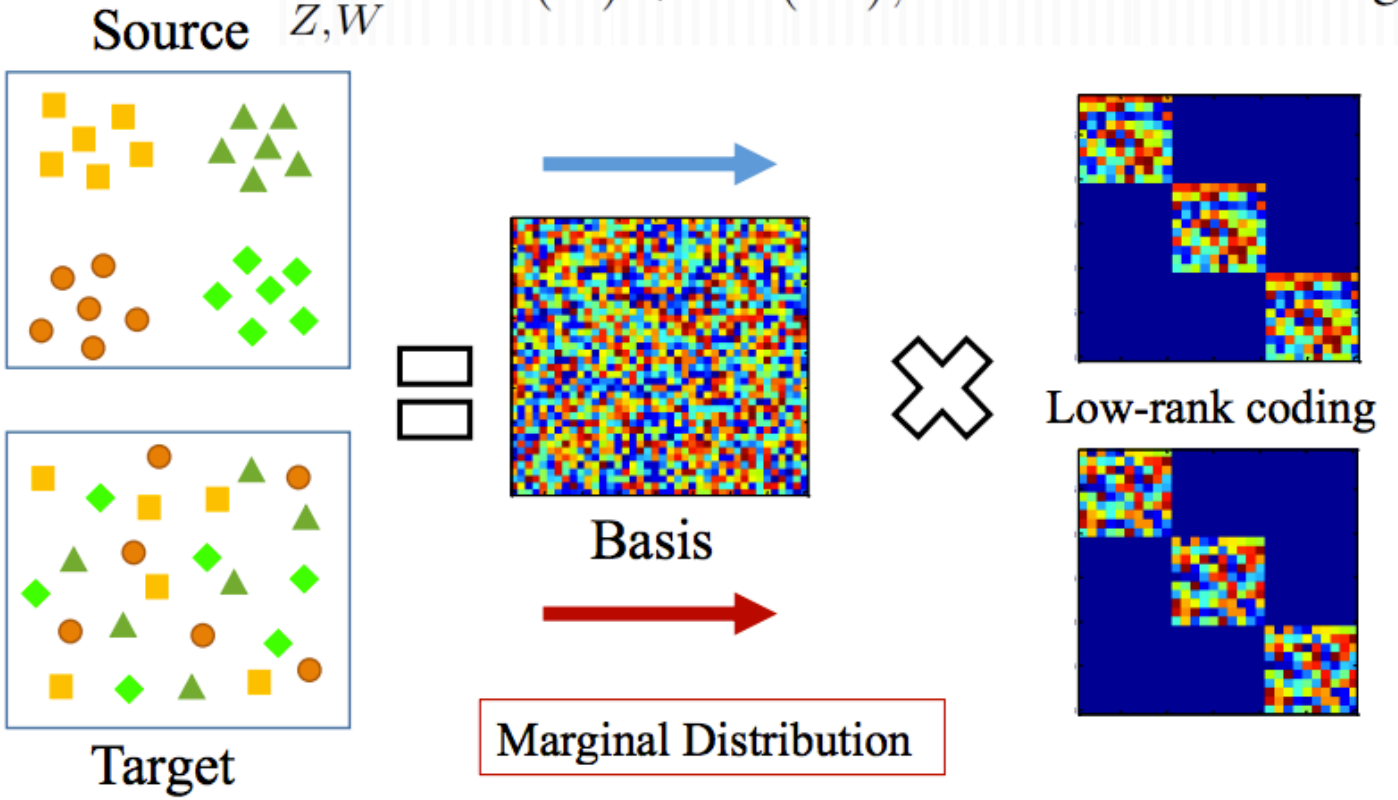


 Zhengming Ding, Ming Shao, Yun Fu: *Deep Low-Rank Coding Transfer Learning*. IJCAI 2015.

Low-Rank Models for Domain Adaptation

- Deep Low-Rank Coding for Domain Adaptation

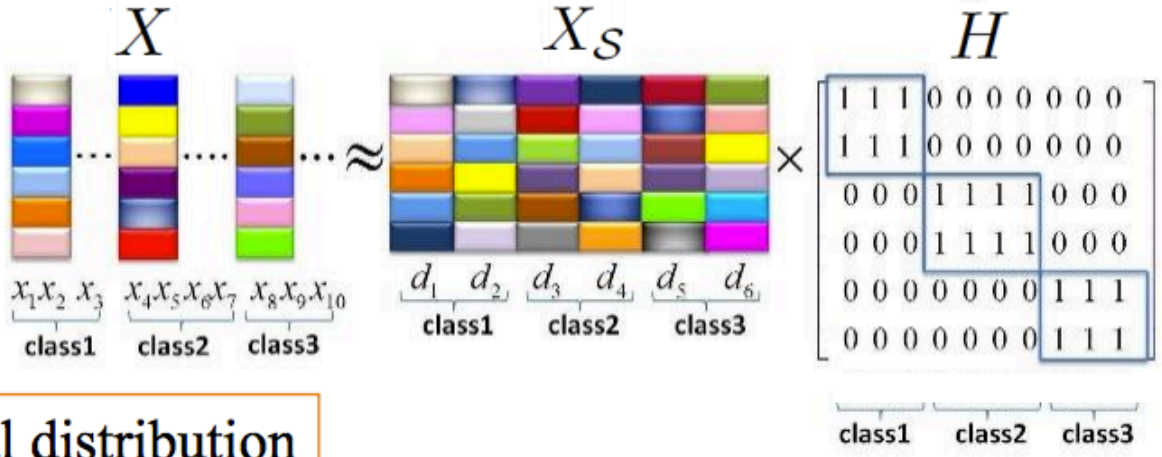
$$\min_{Z,W} \operatorname{rank}(Z) + \lambda\Omega(W), \ \text{s.t.} \ WX = WX_SZ$$



 Zhengming Ding, Ming Shao, Yun Fu: *Deep Low-Rank Coding Transfer Learning*. IJCAI 2015.

Low-Rank Models for Domain Adaptation

- Deep Low-Rank Coding for Domain Adaptation



Conditional distribution

$$\min_{W,Z} \|Z\|_* + \lambda\Omega(W) + \alpha\|Z_l - H\|_{\mathrm{F}}^2,$$
$$\text{s.t. } WX = WX_SZ,$$

- $\Omega(W) = \mathrm{tr}\left[(\bar{X} - W\tilde{X})^{\mathrm{T}}(\bar{X} - W\tilde{X})\right]$ (Marginalized denoising auto-encoder)

 Zhengming Ding, Ming Shao, Yun Fu: *Deep Low-Rank Coding Transfer Learning*. IJCAI 2015.

Low-Rank Models for Domain Adaptation

- Deep Low-Rank Coding for Domain Adaptation

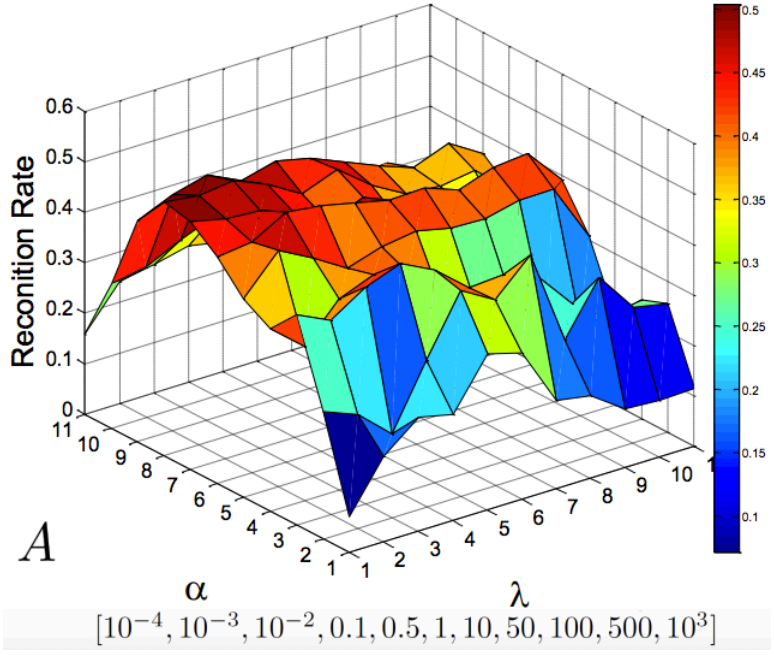

MNIST

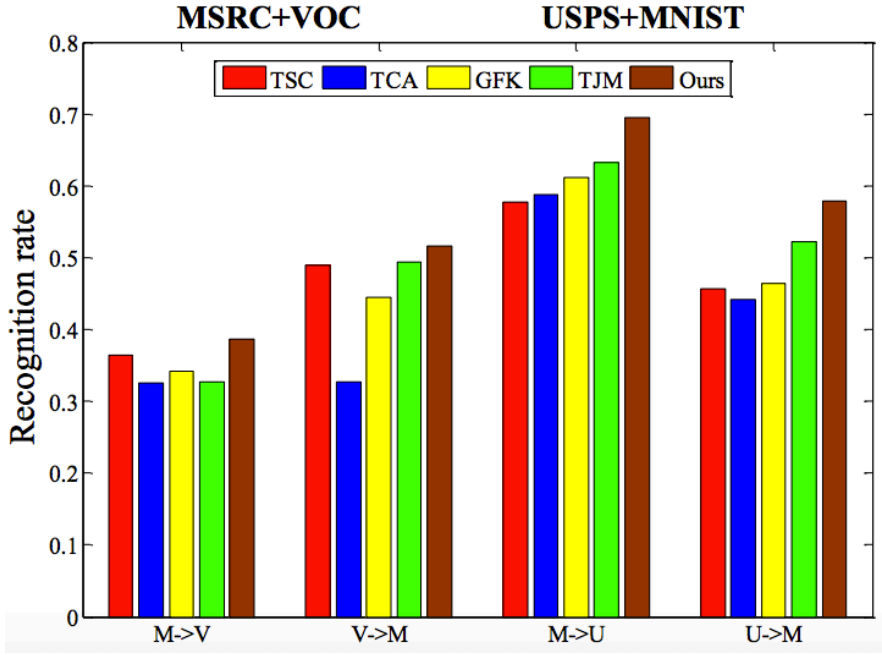

USPS


VOC


MSRC


Office+Caltech

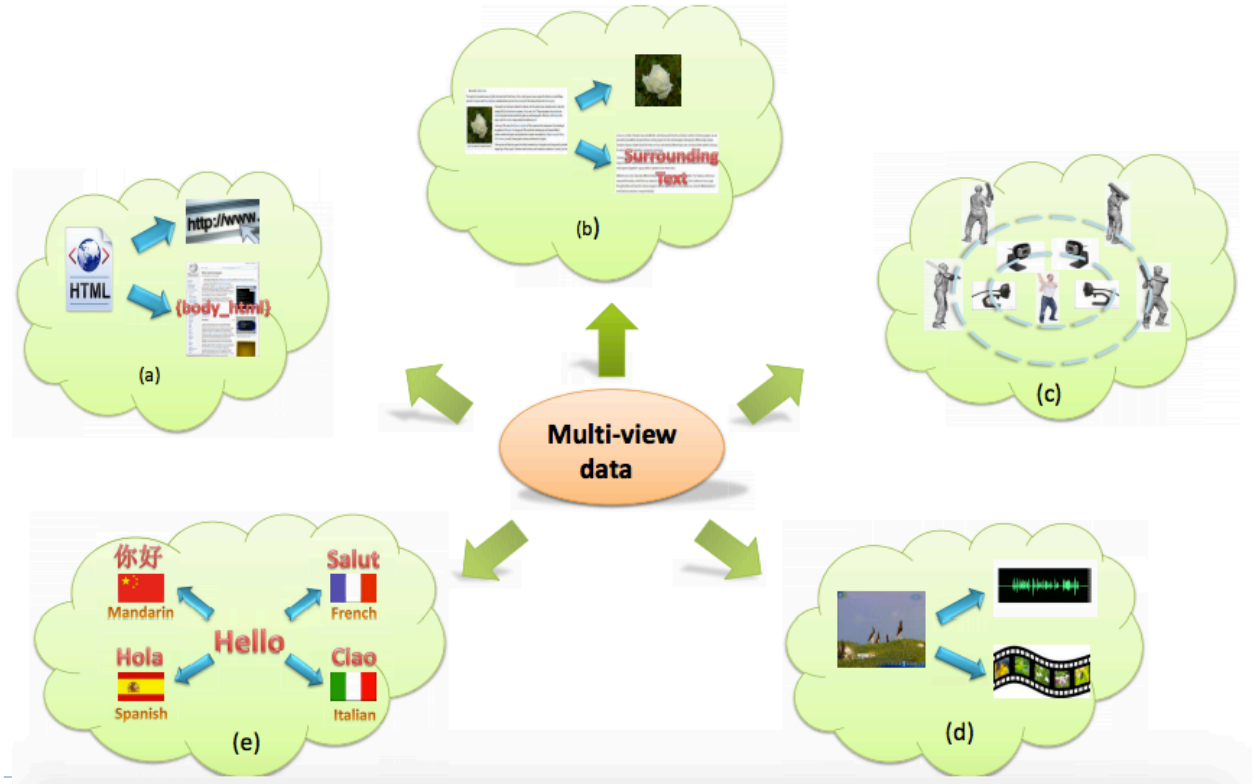




 Zhengming Ding, Ming Shao, Yun Fu: *Deep Low-Rank Coding Transfer Learning*. IJCAI 2015.

Low-Rank Models for Multi-View Learning

- Multi-View Learning
  - Learning from data where observations are represented by multiple independent sets of features

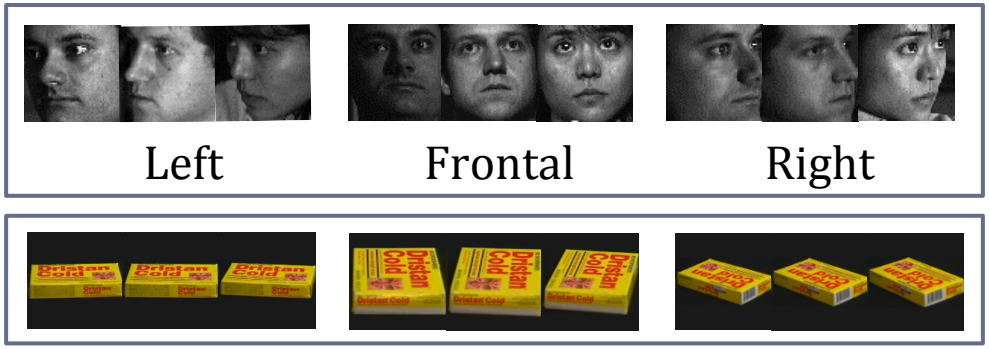


 Chang Xu, Dacheng Tao, Chao Xu: *A Survey on Multi-view Learning*. arXiv 2013.

Low-Rank Models for Multi-View Learning

- Multi-View Learning
  - Multi-View Visual Data

Multi-modal image
(different sensors)



Multi-pose image
(different viewpoints)

  - Challenge: large within-class divergence across multiple views

Low-Rank Models for Multi-View Learning

- Models
  - Robust Multi-View Spectral Clustering (RMSC) [Xia et al., AAAI'14]
  - Low-Rank Common Subspace (LRCS) [Ding et al., ICDM'14]
  - Robust Multi-View Subspace Learning (RMSL) [Ding et al., AAAI'16]

Low-Rank Models for Multi-View Learning

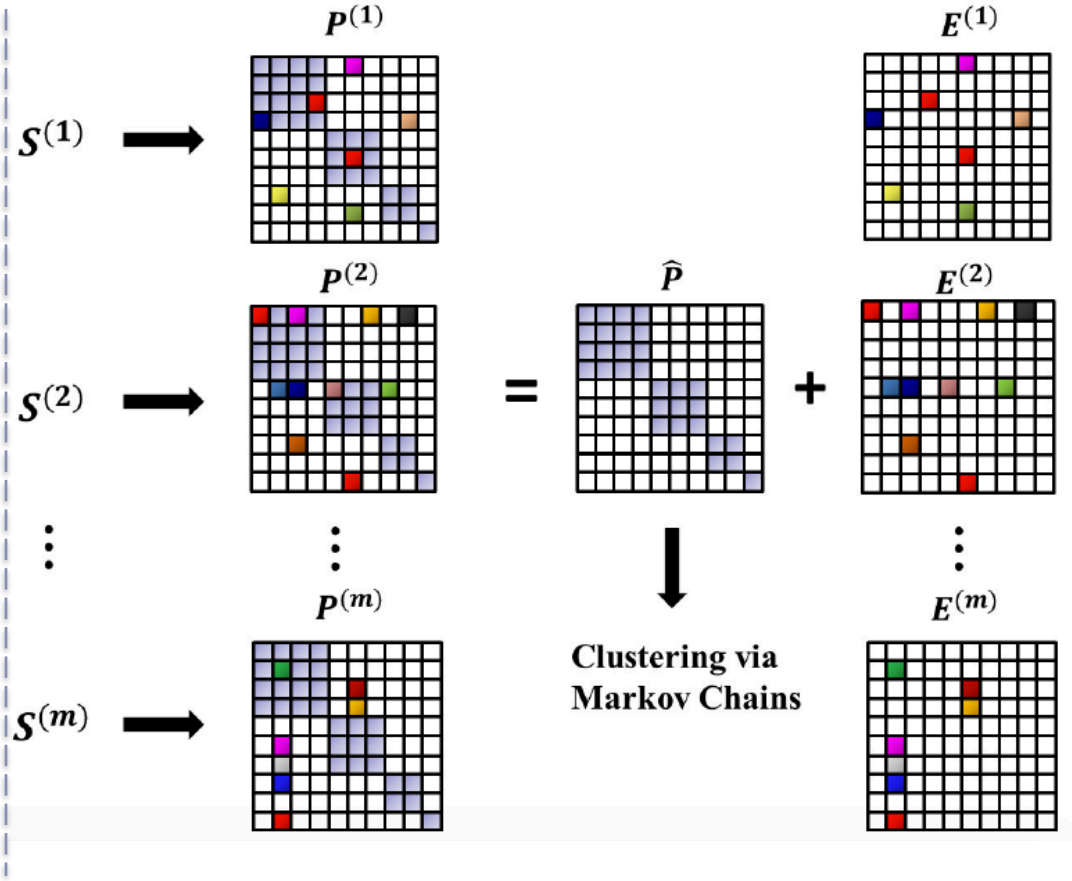
- Robust Multi-View Spectral Clustering (RMSC)
  - Spectral clustering has a natural connection to Markov chains methods, which construct an accurate transition probability matrix
  - Input data might be noisy. Thus, the transition matrix might be corrupted (**Solution**: Low-Rank + Sparse)
  - **RMSC** is a Markov chain method that explicitly handles the possible noise in the transition probability matrices associated with different views

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin: *Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition*. AAAI 2014.

Low-Rank Models for Multi-View Learning

- Robust Multi-View Spectral Clustering (RMSC)

- Construct similarity matrix $S^{(i)}$ for each view
- Calculate transition matrix $P^{(i)}$
- Recover a low-rank latent transition probability matrix $\widehat{P}$ from $P^{(i)}$
- $\widehat{P}$ is used for single-view standard spectral clustering



Rongkai Xia, Yan Pan, Lei Du, and Jian Yin: *Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition*. AAAI 2014.

Low-Rank Models for Multi-View Learning

- Robust Multi-View Spectral Clustering (RMSC)
  - Model

$$\min_{\hat{P},E^{(i)}} rank(\hat{P}) + \lambda \sum_{i=1}^{m} \|E^{(i)}\|_0$$

$$s.t.\ \ i = 1, 2, ..., m,\ P^{(i)} = \hat{P} + E^{(i)},\ \hat{P} \geq 0, \hat{P}\mathbf{1} = \mathbf{1}$$

$$\min_{\hat{P},E^{(i)}} \|\hat{P}\|_* + \lambda \sum_{i=1}^{m} \|E^{(i)}\|_1$$

$$s.t.\ \ i = 1, 2, ..., m,\ P^{(i)} = \hat{P} + E^{(i)},\ \hat{P} \geq 0, \hat{P}\mathbf{1} = \mathbf{1}.$$

  - ALM algorithm is used to solve the problem

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin: *Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition*. AAAI 2014.

Low-Rank Models for Multi-View Learning

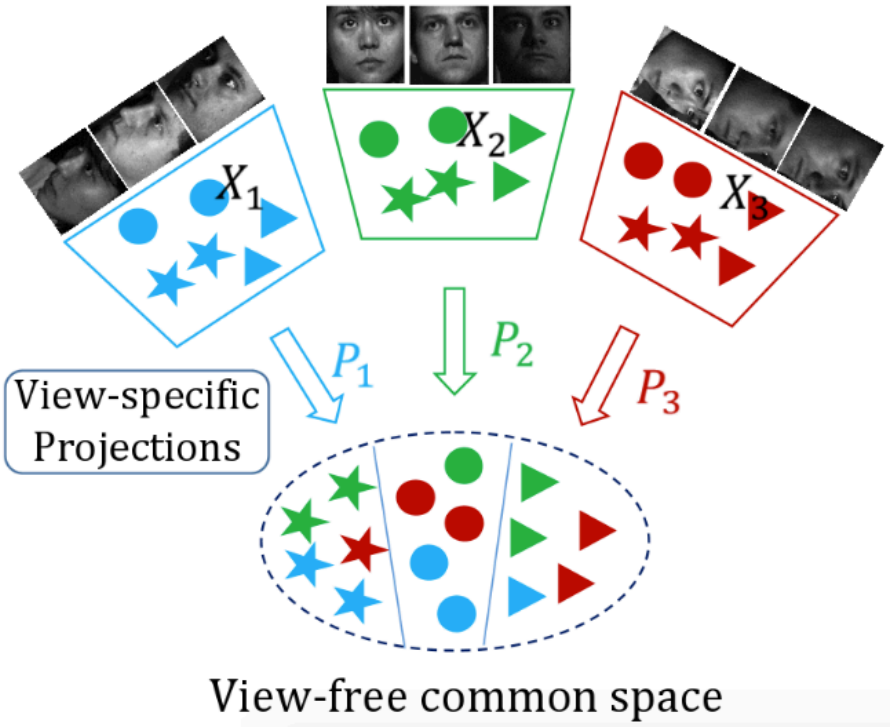
- Robust Multi-View Spectral Clustering (RMSC)

dataset	method	F-score	Precision	Recall	Entropy	NMI
BBC	Best Single View	0.798(0.058)	0.804(0.071)	0.792(0.045)	0.600(0.084)	0.735(0.033)
	Feature Concat	0.828(0.074)	0.822(0.101)	0.838(0.047)	0.498(0.131)	0.784(0.048)
	Kernel Addition	0.828(0.077)	0.818(0.110)	0.842(0.045)	0.503(0.146)	0.783(0.052)
	Co-Reg	0.829(0.049)	0.836(0.057)	0.822(0.042)	0.516(0.076)	0.771(0.031)
	MMC	0.829(0.075)	0.825(0.097)	0.834(0.051)	0.498(0.133)	0.783(0.051)
	Ours	**0.871(0.053)**	**0.879(0.078)**	**0.864(0.046)**	**0.431(0.109)**	**0.808(0.059)**
BBCSport	Best Single View	0.768(0.004)	0.781(0.015)	0.756(0.023)	0.616(0.028)	0.715(0.006)
	Feature Concat	0.657(0.015)	0.667(0.019)	0.649(0.030)	0.862(0.041)	0.604(0.016)
	Kernel Addition	0.657(0.020)	0.649(0.007)	0.667(0.044)	0.886(0.034)	0.600(0.022)
	Co-Reg	0.766(0.002)	0.786(0.008)	0.748(0.012)	0.606(0.015)	0.718(0.003)
	MMC	0.657(0.019)	0.658(0.018)	0.658(0.039)	0.877(0.039)	0.601(0.018)
	Ours	**0.869(0.035)**	**0.871(0.022)**	**0.869(0.049)**	**0.405(0.023)**	**0.818(0.017)**
WebKB	Best Single View	0.889(0.003)	0.824(0.005)	0.956(0.000)	0.406(0.001)	0.532(0.002)
	Feature Concat	0.947(0.001)	0.947(0.003)	0.947(0.001)	0.214(0.001)	0.718(0.002)
	Kernel Addition	0.947(0.006)	0.947(0.004)	0.947(0.010)	0.214(0.003)	0.718(0.001)
	Co-Reg	0.933(0.002)	0.958(0.003)	0.910(0.001)	0.209(0.002)	0.700(0.008)
	MMC	0.947(0.005)	0.947(0.002)	0.947(0.001)	0.214(0.001)	0.718(0.003)
	Ours	**0.960(0.001)**	**0.965(0.002)**	**0.965(0.001)**	**0.164(0.001)**	**0.779(0.004)**

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin: *Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition*. AAAI 2014.

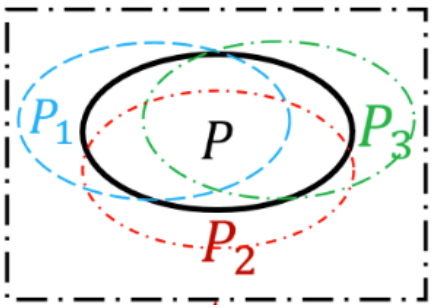
Low-Rank Models for Multi-View Learning

- Low-Rank Common Subspace (LRCS)
  - Previous work: learn multiple view-specific projections, and project test data (with known view info) to a specific projection



What if the view info of test data is unknown?

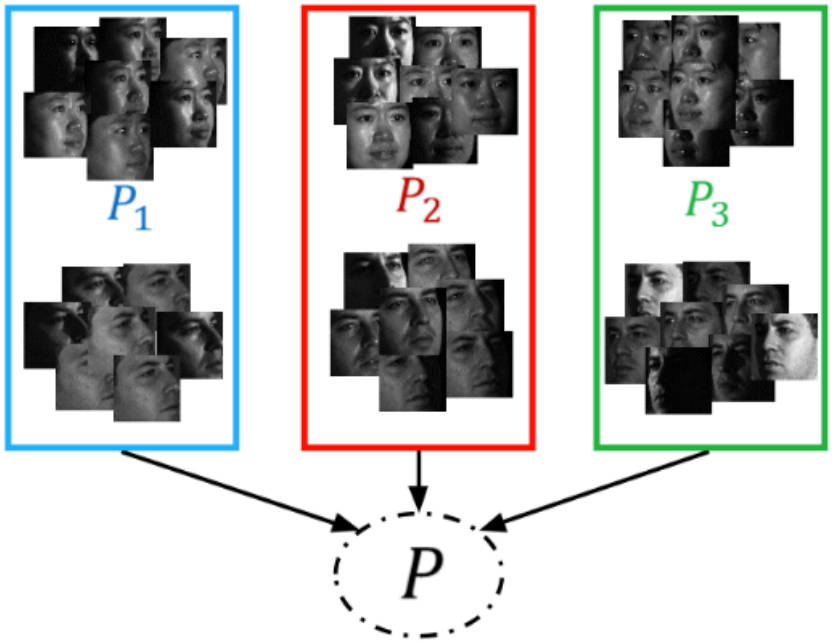
Learn a common projection!



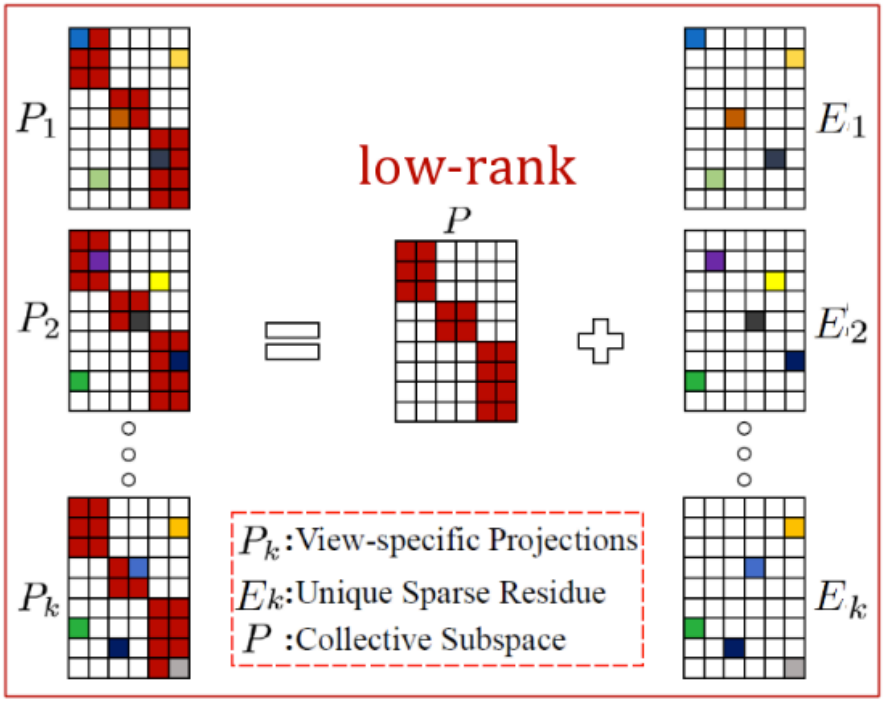
 Zhengming Ding and Yun Fu: *Low-Rank Common Subspace for Multi-View Learning*. ICDM 2014.

Low-Rank Models for Multi-View Learning

- Low-Rank Common Subspace (LRCS)



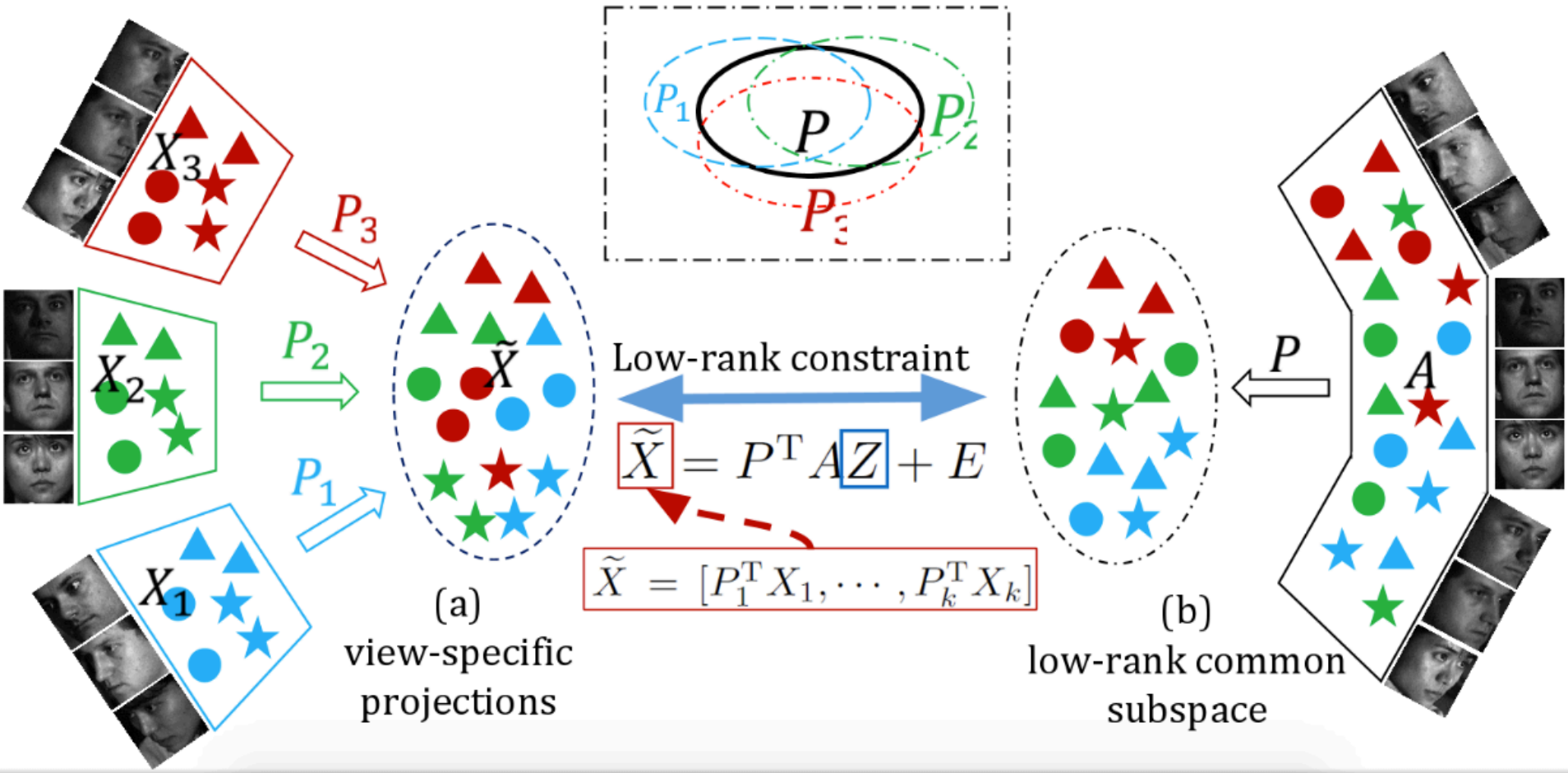
$$\min_{P,E_i,P_i} \boxed{\text{rank}(P)} + \lambda_0 \sum_{i=1}^{k} \|E_i\|_1$$
$$\text{s.t.} \quad P_i = P + E_i, \quad i = 1, \cdots, k,$$



 Zhengming Ding and Yun Fu: *Low-Rank Common Subspace for Multi-View Learning*. ICDM 2014.

Low-Rank Models for Multi-View Learning

- Low-Rank Common Subspace (LRCS)



 Zhengming Ding and Yun Fu: *Low-Rank Common Subspace for Multi-View Learning*. ICDM 2014.

Low-Rank Models for Multi-View Learning

- Low-Rank Common Subspace (LRCS)
  - Model

$$\min_{P,Z,E,E_i,P_i} \|Z\|_* + \|P\|_* + \lambda_0 \sum_{i=1}^{k} \|E_i\|_1 + \lambda_1 \|E\|_{2,1}$$

$$\text{s.t. } \widetilde{X} = P^{\mathrm{T}} A Z + E, \quad P^{\mathrm{T}} P = \mathrm{I},$$

$$P_i = P + E_i, \quad i = 1, \cdots, k.$$

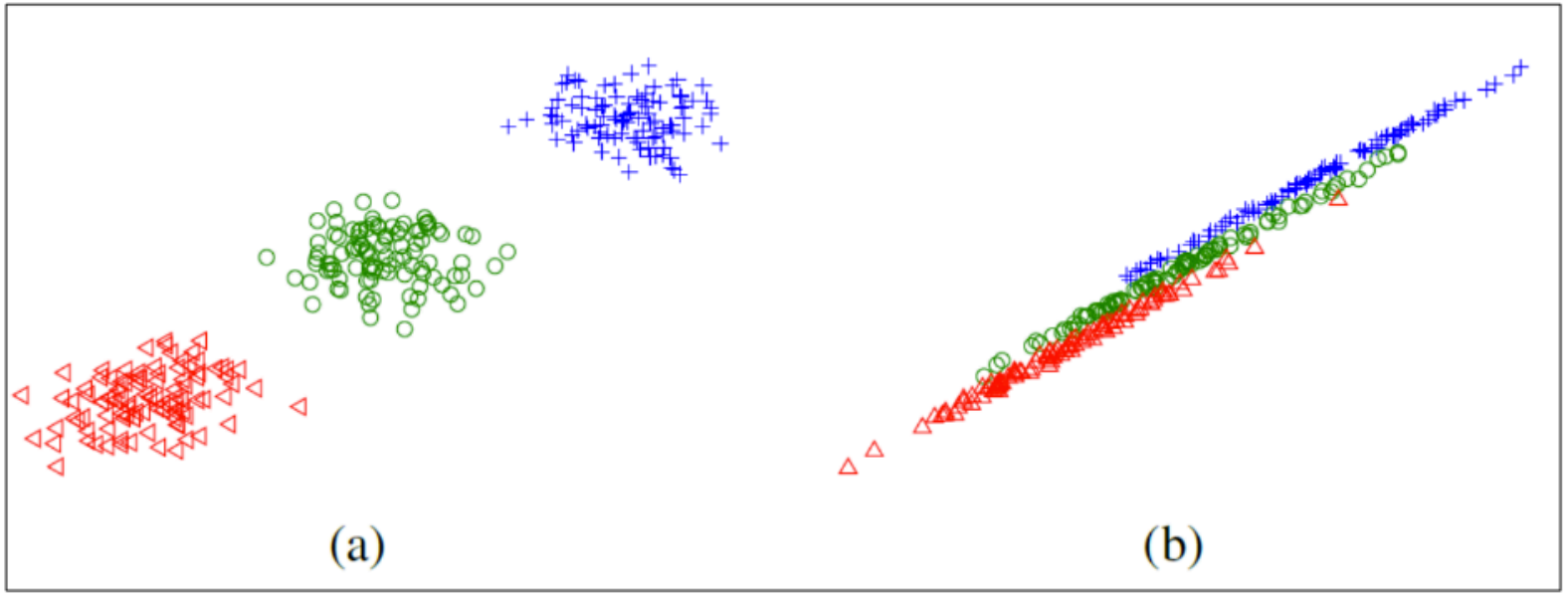
$$\widetilde{X} = [P_1^{\mathrm{T}} X_1, \cdots, P_k^{\mathrm{T}} X_k]$$

  - Low-rank constraints on two directions
  - Consider multi-view structure
  - Two scenarios (representation & transfer)

 Zhengming Ding and Yun Fu: *Low-Rank Common Subspace for Multi-View Learning*. ICDM 2014.

Low-Rank Models for Multi-View Learning

- Low-Rank Common Subspace (LRCS)
  - Toy Example



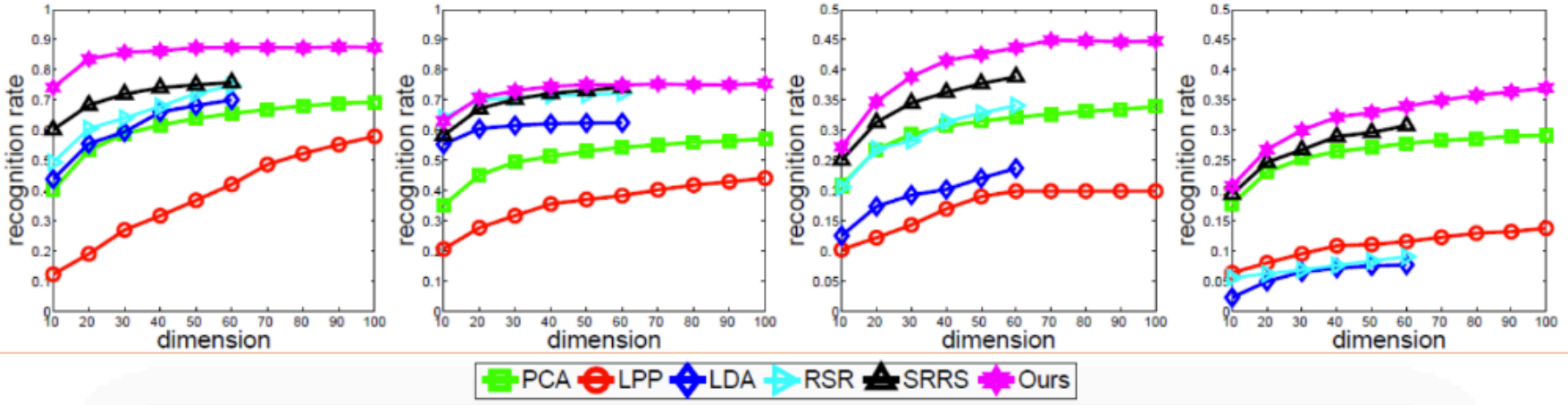
(a) The synthesized data of three views (different colors means different views); (b) The projected data with the learned common projection. Notice that the three views of projected data are aligned well. The figure is better viewed with color.

 Zhengming Ding and Yun Fu: *Low-Rank Common Subspace for Multi-View Learning*. ICDM 2014.

Low-Rank Models for Multi-View Learning

- LRCS: Results on CMU PIE

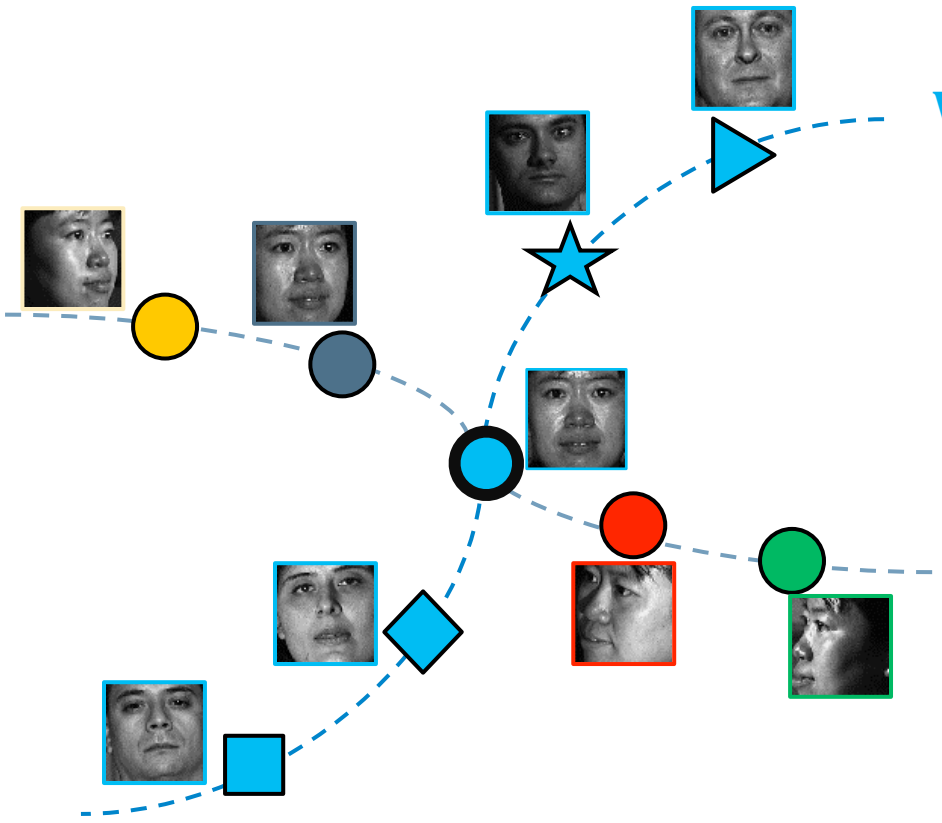
Methods	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
PCA[22]	69.03±0.08	69.21±0.08	68.52±0.12	52.65±0.04	34.94±0.08	29.09±0.01
LDA[24]	70.46±0.05	71.32±0.02	63.51±0.75	56.53±0.02	24.07±0.25	7.06±0.01
LPP[23]	57.25±0.06	58.83±0.07	59.25±0.56	43.56±0.08	19.67±0.05	13.11±0.01
RSR[38]	77.51±0.01	74.74±0.17	71.10±0.04	67.57±0.01	29.72±0.01	9.44±0.02
TFRR[35]	77.92±0.03	76.24±0.12	75.29±0.07	69.74±0.05	33.91±0.12	28.36±0.04
SRRS[18]	78.27±0.04	78.74±0.23	77.45±0.02	71.44±0.03	38.86±0.02	30.16±0.02
Ours	**87.78±0.02**	**86.67±0.01**	**87.38±0.99**	**74.84±0.04**	**44.48±0.03**	**36.17±0.01**



 Zhengming Ding and Yun Fu: *Low-Rank Common Subspace for Multi-View Learning*. ICDM 2014.

Low-Rank Models for Multi-View Learning

- Robust Multi-View Subspace Learning (RMSL)



View structure

intertwined

Class structure

$$\min_{Z,E} ||Z||_* + \lambda||E||_{2,1}$$
$$\text{s.t.,}\ X = XZ + E$$

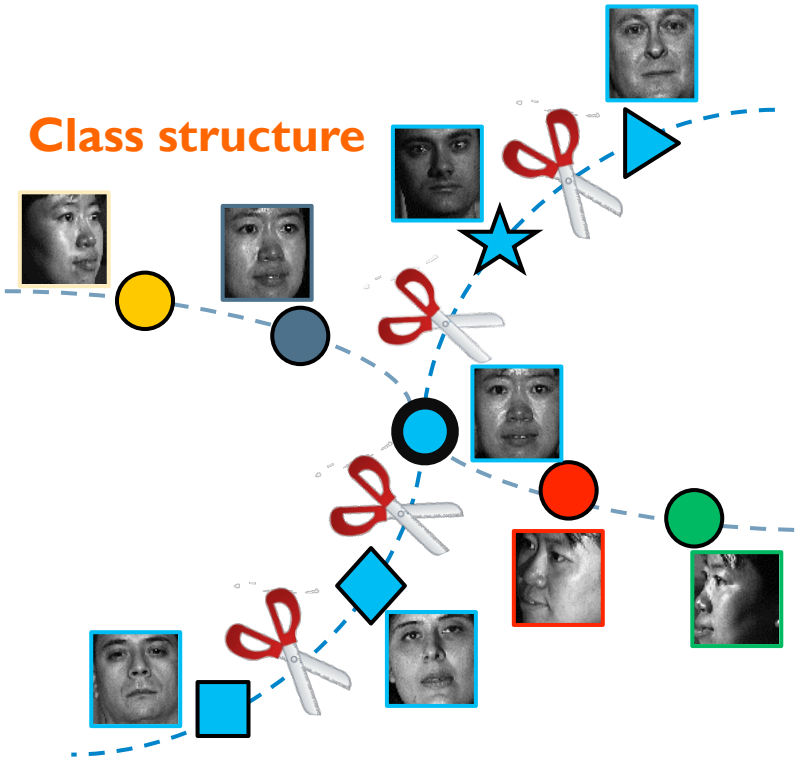
Remark 1: It is hard for low-rank representation (LRR) to uncover the global class structure for multi-view data.

Remark 2: View structure would hinder the classification task in multi-view learning.

Zhengming Ding and Yun Fu: *Robust Multi-View Subspace Learning through Dual Low-Rank Decompositions*. AAAI 2016.

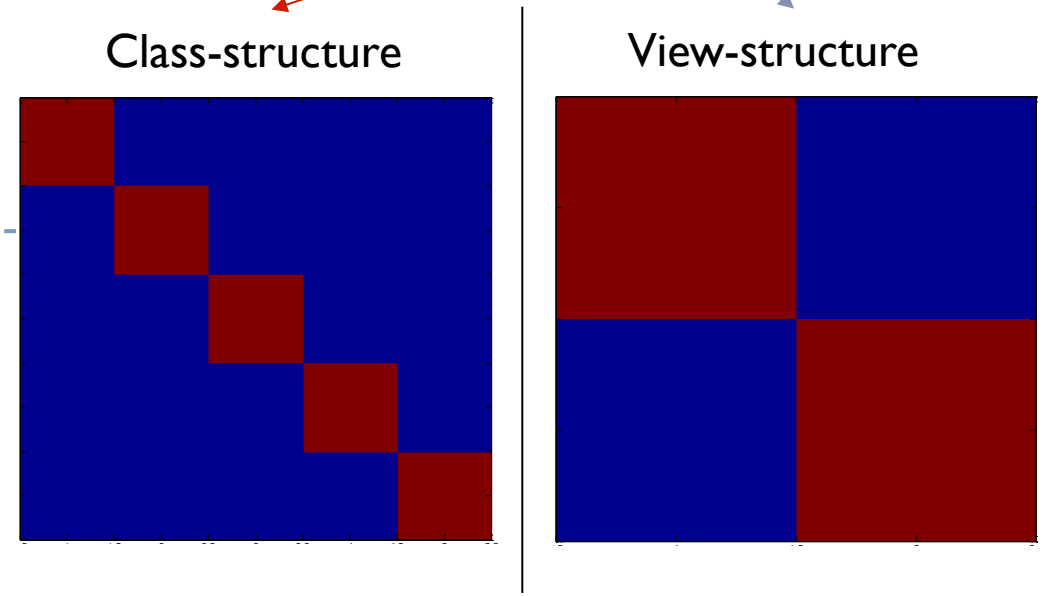
Low-Rank Models for Multi-View Learning

- Robust Multi-View Subspace Learning (RMSL)



Goal: decompose two structures in dual low-rank constraints

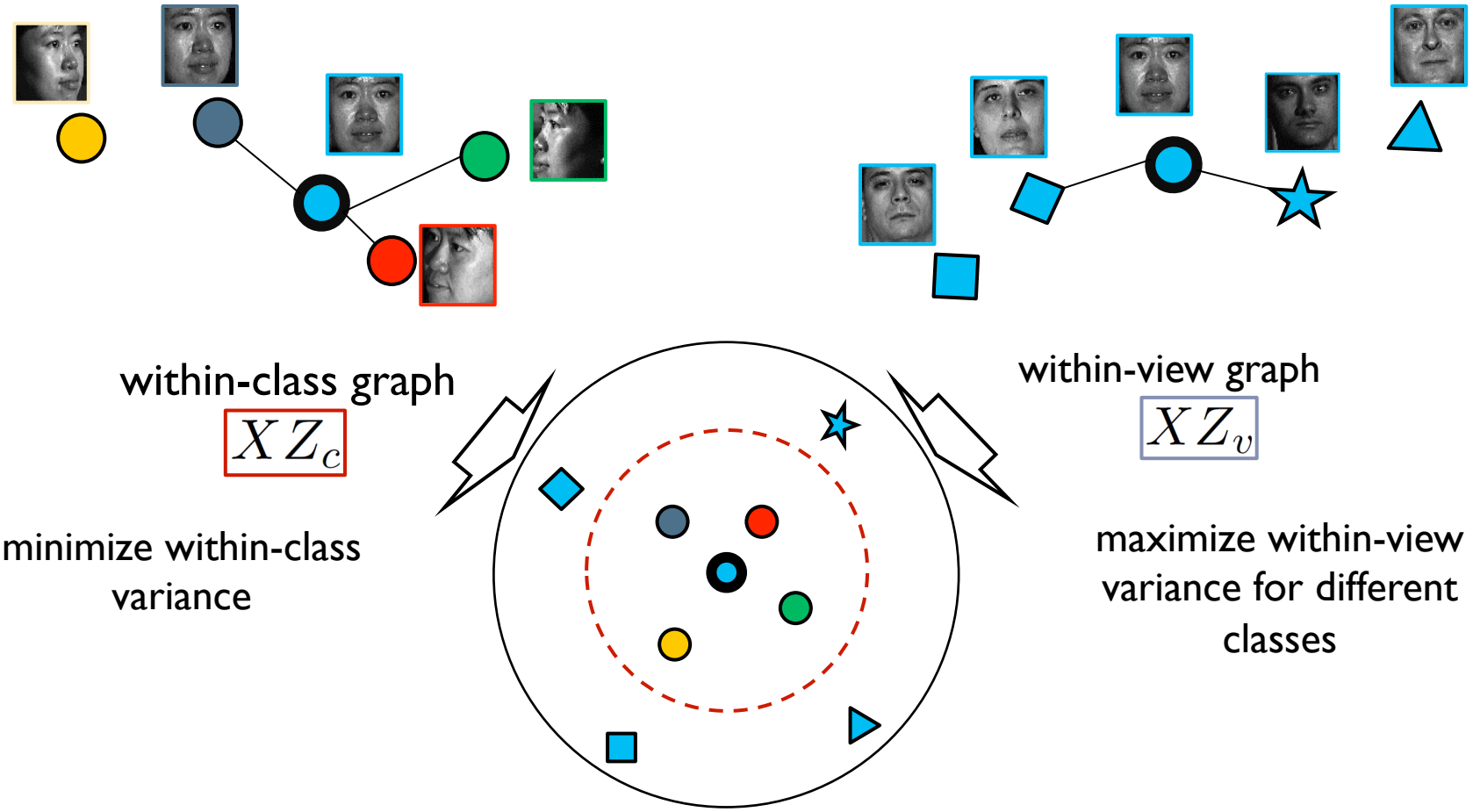
$$Z = Z_C + Z_V$$



Zhengming Ding and Yun Fu: *Robust Multi-View Subspace Learning through Dual Low-Rank Decompositions*. AAAI 2016.

Low-Rank Models for Multi-View Learning

- Robust Multi-View Subspace Learning (RMSL)



Zhengming Ding and Yun Fu: *Robust Multi-View Subspace Learning through Dual Low-Rank Decompositions*. AAAI 2016.

Low-Rank Models for Multi-View Learning

- Robust Multi-View Subspace Learning (RMSL)
  - Model

$$\min_{P,Z_c,Z_v,E} \|Z_c\|_* + \|Z_v\|_* + \lambda\|E\|_1 + \alpha\mathcal{G}(P,Z_c,Z_v)$$
$$\text{s.t. } P^{\mathrm{T}}X = P^{\mathrm{T}}X(Z_c+Z_v)+E,\ P^{\mathrm{T}}P=\mathrm{I_p},$$

  - $\mathcal{G}(P,Z_c,Z_v)$ is a graph regularizer

Nearest neighbor graphs

$$W^c_{i,j} = \begin{cases} 1, & \text{if } x_i \in \mathcal{N}_{k_1}(x_j),\ \text{and } l_i = l_j \\ 0, & \text{otherwise} \end{cases}$$

$$W^v_{i,j} = \begin{cases} 1, & \text{if } x_i \in \mathbf{N}_{k_2}(x_j),\ \text{but } l_i \neq l_j \\ 0, & \text{otherwise} \end{cases}$$

$$\mathcal{G}_c = \textstyle\sum_{i,j}(Y_{c,i}-Y_{c,j})^2 W^c_{i,j}$$
$$\mathcal{G}_v = \textstyle\sum_{i,j}(Y_{v,i}-Y_{v,j})^2 W^v_{i,j}$$

$$Y_c = P^{\mathrm{T}}XZ_c$$
$$Y_v = P^{\mathrm{T}}XZ_v$$

$$\mathcal{G}(P,Z_c,Z_v) = \frac{\mathcal{G}_c}{\mathcal{G}_v} = \frac{\mathrm{tr}(P^{\mathrm{T}}XZ_cL_c(P^{\mathrm{T}}XZ_c)^{\mathrm{T}})}{\mathrm{tr}(P^{\mathrm{T}}XZ_vL_v(P^{\mathrm{T}}XZ_v)^{\mathrm{T}})}$$

  - ADMM is used for optimization

Zhengming Ding and Yun Fu: *Robust Multi-View Subspace Learning through Dual Low-Rank Decompositions*. AAAI 2016.

Low-Rank Models for Multi-View Learning

- Robust Multi-View Subspace Learning (RMSL): Results

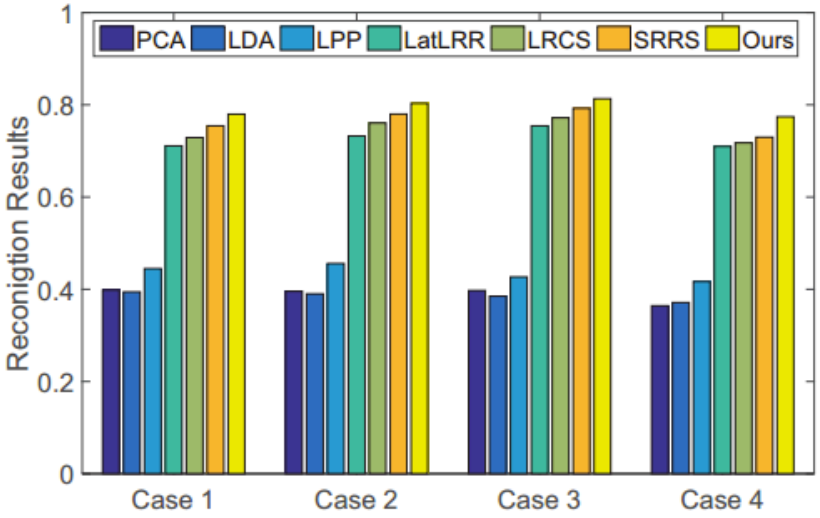
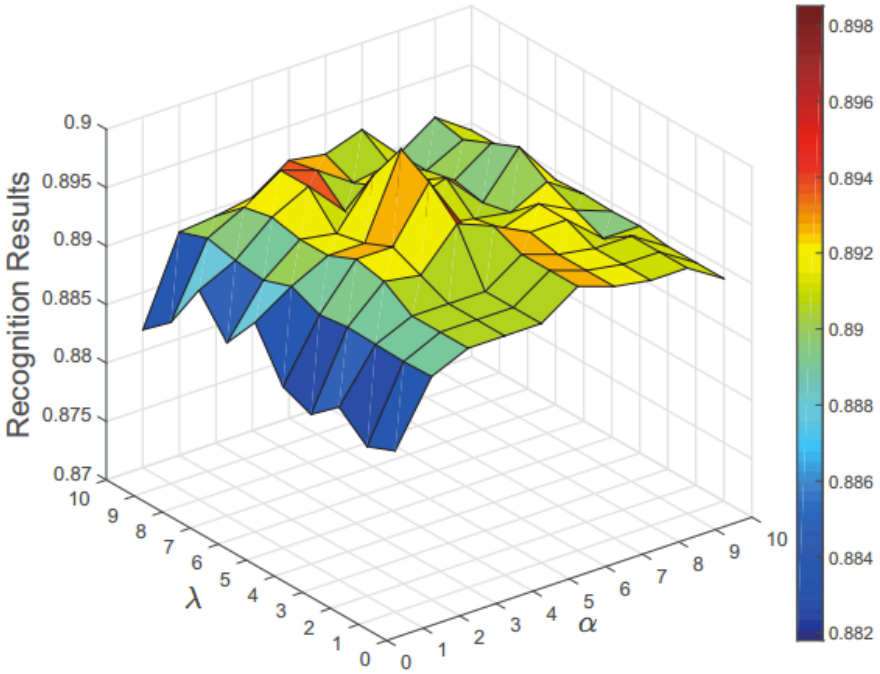


Figure 2: Recognition results of seven algorithms on 4 Cases of COIL-100 object dataset, where Case 1: View 1 & 3; Case 2: View 1 & 4; Case 3: View 2 & 3; Case 4: View 2 & 4.



Zhengming Ding and Yun Fu: *Robust Multi-View Subspace Learning through Dual Low-Rank Decompositions*. AAAI 2016.