Low-rank modeling for data representation

Chong Peng

College of Science and Technology, Qingdao University

May 16, 2018

Introduction

Images

- ➢ *1M pixels*

Videos

- ➢ *1B voxels*


U.S. COMMERCE'S ORTNER SAYS YEN UNDERVALUED

Commerce Dept. undersecretary of economic affairs Robert Ortner said that he believed the dollar at current levels was fairly priced against most European currencies.

In a wide ranging address sponsored by the Export-Import Bank, Ortner, the bank's senior economist also said he believed that the yen was undervalued and could go up by 10 or 15 pct.

"I do not regard the dollar as undervalued at this point against the yen," he said.

On the other hand, Ortner said that he thought that "the yen is still a little bit undervalued," and "could go up another 10 or 15 pct."

In addition, Ortner, who said he was speaking personally, said

that the dollar against most European currencies was "fairly pric

Ortner said his analysis of the various exchange rate values w

such economic particulars as wage rate differentiations.

Ortner said there had been little impact on U.S. trade deficit by

of the dollar because at the time of the Plaza Accord, the dollar wa

overvalued and that the first 15 pct decline had little impact.

He said there were indications now that the trade deficit was

level off.

Turning to Brazil and Mexico, Ortner made it clear that

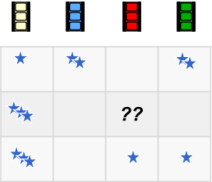
almost impossible for those countries to earn enough foreign exch

the service on their debts. He said the best way to deal with thi

the policies outlined in Treasury Secretary James Baker's debt in

Web data

- ➢ *100B webpages*



User data

- ➢ *1B users*

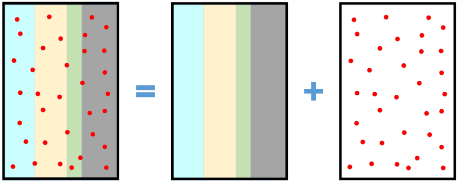
Robust Principal Component Analysis

- PCA

$$\min_{rank(A)\leq r} \|X - A\|_F^2. \tag{1}$$

- Robust PCA

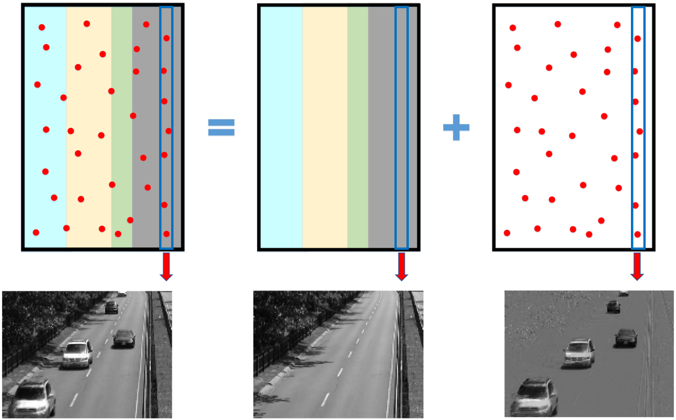
$$\min_{A,S} \|A\|_* + \lambda\|S\|_1, \quad s.t. \quad X = A + S. \tag{2}$$



Candes et al., Robust Principal Component Analysis? J. ACM, 2011.

Robust Principal Component Analysis

In a surveillance video, the background forms a low-rank part while the moving objects form a sparse part.



Robust Principal Component Analysis

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1, \quad s.t. \quad X = L + S. \tag{3}$$

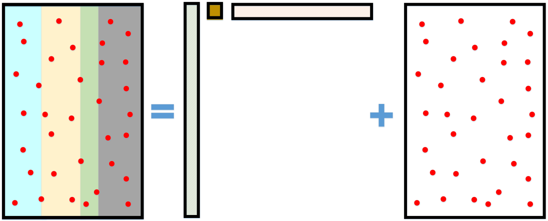
The nuclear norm is not accurate for matrices with large singular values!

$$\|L\|_{ld} = \log\det(I + (Z^T Z)^{1/2}) = \sum_i \log(1 + \sigma_i(L)) \tag{4}$$

$$\min_{L,S} \|L\|_{ld} + \lambda \|S\|_1, \quad s.t. \quad X = L + S. \tag{5}$$

Fast factorization-based approach:

$$\min_{C,S,U,V} \|C\|_{Id} + \lambda \|S\|_1, \quad s.t. \quad X = UCV^T + S, U^T U = I_r, V^T V = I_r. \tag{6}$$



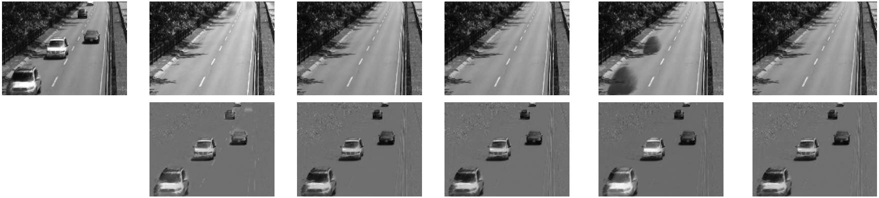
Nonconvex: factorization, nonconvex rank approximation

Peng et al., A fast factorization-based approach to robust PCA, ICDM 2016.

Robust Principal Component Analysis

- Background-foreground separation: In a surveillance video, the background usually forms a low-rank part while the moving foreground forms a sparse part.
- Shadow removal from face images: In a set of face images from the same person, the face usually forms a low-rank part wile the shadow forms a sparse part.
- Anomaly detection: In a set of handwritten digits, the majority number forms a low-rank part while the anomaly forms a sparse part.
- Denoising of hyperspectral images: In hyperspectral images, the ground truth image forms a low-rank part while the noise forms a sparse part.

Foreground-background Separation



(a) Original (b) IALM (c) AltProj (d) F-FFP (e) AltProj (k=5) (f) U-FFP

Figure 1: Foreground-background separation in the Highway video. The top left is the original frame and the rest are extracted background (top) and foreground (bottom).

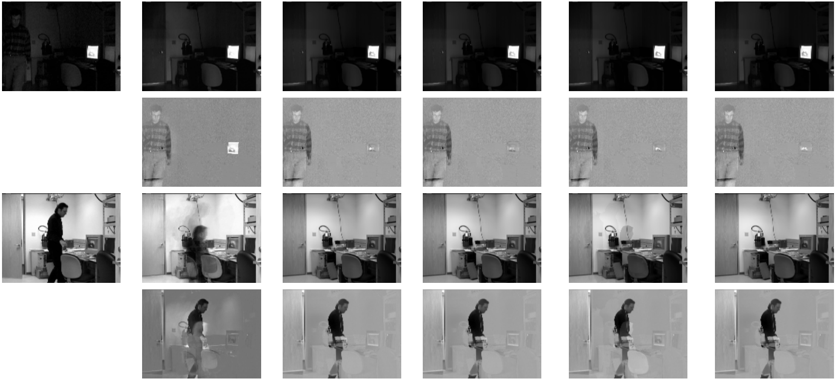
Foreground-background Separation

Table 1: Results with r Known for Datasets with Single Background

Data	Method	Rank(L)	$\|S\|_0/(dn)$	$\frac{\|X-L-S\|_F}{\|X\|_F}$	# of Iter.	# of SVDs	Time
Highway	AltProj	1	0.9331	2.96e-4	37	38	49.65
	IALM	539	0.8175	6.02e-4	12	13	269.10
	F-FFP	1	0.8854	5.74e-4	24	24	14.83
Escalator Airport	AltProj	1	0.9152	2.29e-4	40	41	110.75
	IALM	957	0.7744	7.76e-4	11	12	1,040.91
	F-FFP	1	0.8877	5.45e-4	23	23	30.78
PETS2006	AltProj	1	0.8590	5.20e-4	35	36	44.64
	IALM	293	0.8649	5.63e-4	12	13	144.26
	F-FFP	1	0.8675	5.61e-4	24	24	14.33
Shopping Mall	AltProj	1	0.9853	3.91e-5	45	46	45.35
	IALM	328	0.8158	9.37e-4	11	12	123.99
	F-FFP	1	0.9122	7.70e-4	23	23	11.65

For IALM and AltProj, (partial) SVDs are for $d \times n$ matrices. For F-FFP, SVDs are for $n \times k$ matrices, which are computationally far less expensive than those required by IALM and AltProj.

Foreground-background Separation



(a) Original (b) IALM (c) AltProj (d) F-FFP (e) AltProj (k=5) (f) U-FFP

Figure 2: Foreground-background separation in the Light Switch-2 video. The top and bottom two panels correspond to two frames, respectively. For each frame, the top left is the original image while the rest are the extracted background (top) and foreground (bottom), respectively.

Foreground-background Separation

Table 2: Results with r Known for Datasets with Multiple Backgrounds

Data	Method	Rank(L)	$\|S\|_0/(dn)$	$\frac{\|X-L-S\|_F}{\|X\|_F}$	# of Iter.	# of SVDs	Time
Lobby	AltProj	2	0.9243	1.88e-4	39	41	47.32
	IALM	223	0.8346	6.19e-4	12	13	152.54
	F-FFP	2	0.8524	6.42e-4	24	24	15.20
Light Switch-2	AltProj	2	0.9050	2.24e-4	47	49	87.35
	IALM	591	0.7921	7.93e-4	12	13	613.98
	F-FFP	2	0.8323	7.54e-4	24	24	24.12
Camera Parameter	AltProj	2	0.8806	5.34e-4	47	49	84.99
	IALM	607	0.7750	6.86e-4	12	13	433.47
	F-FFP	2	0.8684	6.16e-4	24	24	22.25
Time Of Day	AltProj	2	0.8646	4.72e-4	44	46	61.63
	IALM	351	0.6990	6.12e-4	13	14	265.87
	F-FFP	2	0.8441	6.81e-4	25	25	18.49

For IALM and AltProj, (partial) SVDs are for $d \times n$ matrices. For F-FFP, SVDs are for $n \times k$ matrices, which are computationally far less expensive than those required by IALM and AltProj.

Shadow Removal from Face Images

Table 3: Recovery Results of Face Data with $k = 1$

Data	Method	Rank(Z)	$\|S\|_0/(dn)$	$\frac{\|X-Z-S\|_F}{\|X\|_F}$	# of Iter.	# of SVDs	Time
Subject 1	AltProj	1	0.9553	8.18e-4	50	51	4.62
	IALM	32	0.7745	6.28e-4	25	26	2.43
	F-FFP	1	0.9655	8.86e-4	36	36	1.37
Subject 2	AltProj	1	0.9755	2.34e-4	49	50	5.00
	IALM	31	0.7656	6.47e-4	25	26	2.66
	F-FFP	1	0.9492	9.48e-4	36	36	1.37

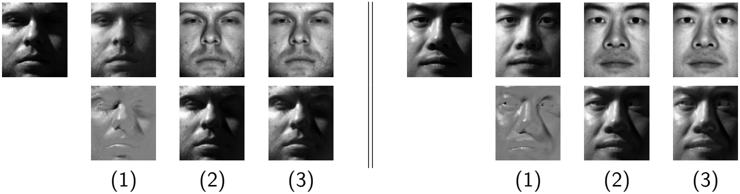


Figure 3: Shadow removal results from EYaleB data. For each of the two parts, the top left is the original image and the rest are recovered clean image (top) and shadow (bottom) by (1) IALM, (2) AltProj, and (3) F-FFP, respectively.

Shadow Removal from Face Images

Table 4: Recovery Results of Face Data with $k = 5$

Data	Method	Rank(Z)	$\|S\|_0/(dn)$	$\frac{\|X-Z-S\|_F}{\|X\|_F}$	# of Iter.	# of SVDs	Time
Subject 1	AltProj	5	0.9309	3.93e-4	51	55	6.08
	U-FFP	5	0.9632	9.01e-4	36	36+36	1.44
Subject 2	AltProj	5	0.8903	6.40e-4	54	58	7.92
	U-FFP	1	0.9645	5.85e-4	37	37+37	1.53

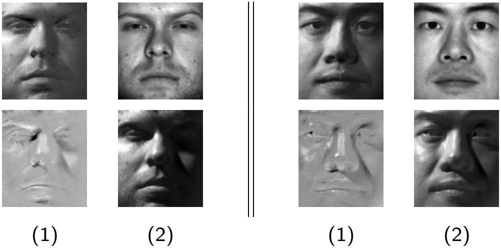


Figure 4: Shadow removal results from EYaleB data. The top panel are the recovered clean image and the bottom panel are the shadows by (1) AltProj (k=5) and (2) U-FFP, respectively.

Anomaly Detection

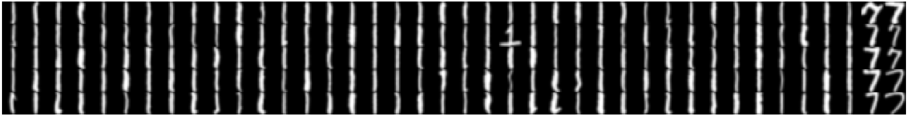


Figure 5: Selected ‘1’s and ‘7’s from USPS dataset.

Anomaly Detection

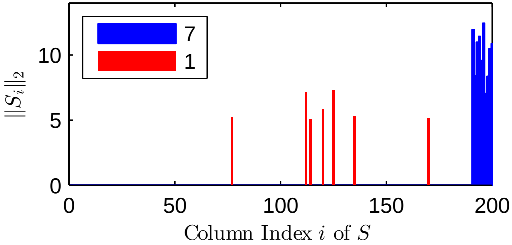
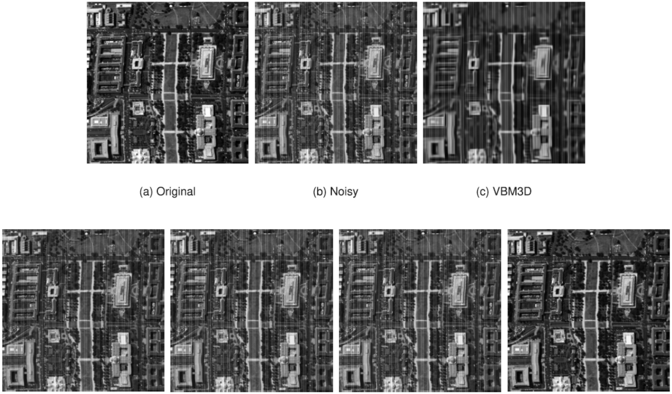


Figure 6: ℓ_2-norms of each row of S.

Figure 7: Written '1's and outliers identified by F-FFP.

Denoising of HSI


(a) Original (b) Noisy (c) VBM3D

(d) LRMR (e) NAILRMA (f) WSN-LRMA (g) NonLRMA

Figure 8: Restoration results on synthetic data: Washington DC Mall. (a) Original image. (b) Noisy image. The resorted image obtained by (c) VBM3D, (d) LRMR, (e) NAILRMA, (f) WSN-LRMA, and (g) U-FFP.

Denoising of HSI

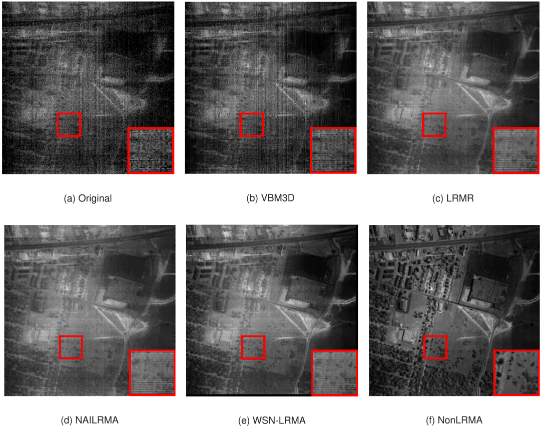


Figure 9: Restoration results on HYDICE urban data set: severe noise band. (a) Original image located at the 108th band. Resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA.

Scabality

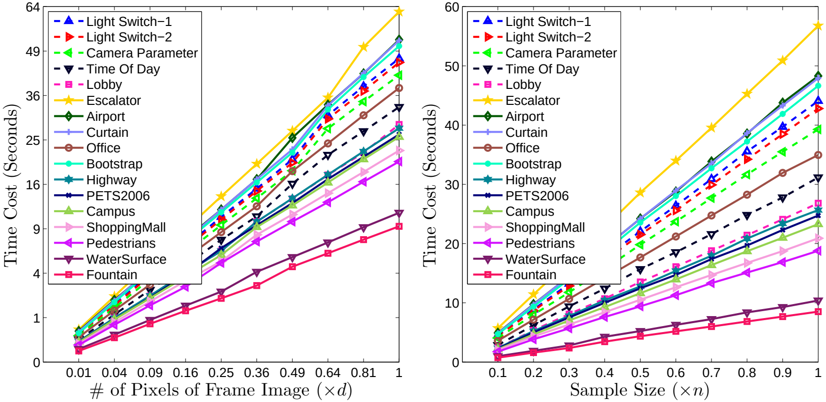
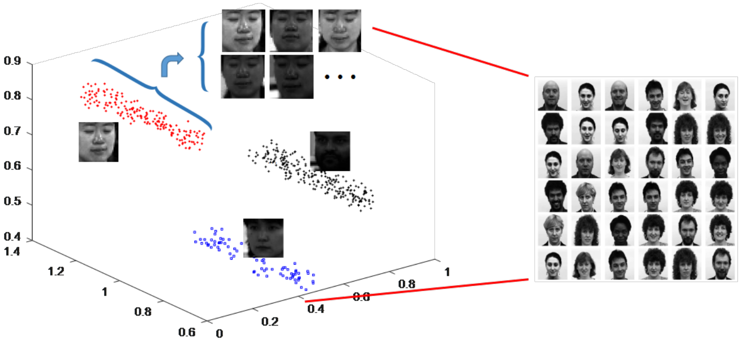


Figure 10: Time cost of F-FFP changes with respect to dimension and sample size of the data.

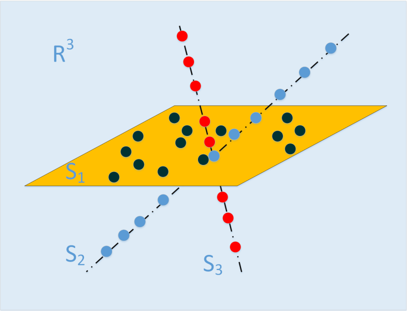
Multiple Subspaces

PCA/RPCA recover a single subspace.
But data may have multiple subspaces...



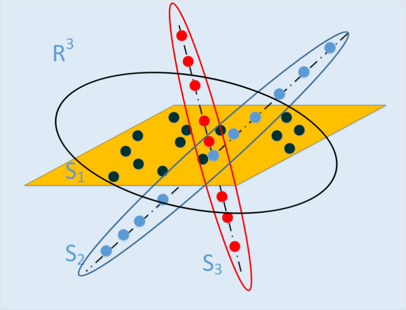
Low-dimensional Subspaces

Rather than uniformally distributed in the high-dimensional space, high-dimensional data often come from a union of low-dimensional subspaces, i.e., high-dimensional data often have low-dimensional structures.



Low-dimensional Subspaces

Can we exploit low-dimensional structures?



Subspace Clustering

- Iterative Methods: K-subspace, q-flat
- Algebraic Methods: matrix factorization based, generalized PCA, robust algebraic segmentation
- Statistical Methods: mixture of probabilistic PCA, agglomerative lossy compression, random sample consensus
- Spectral Clustering-Based Methods: factorization-based affinity, GPCA-based affinity, local-subspace based affinity, locally linear manifold clustering ...

Vidal, Subspace Clustering, IEEE Signal Processing Magzine, 2011.

Sparse subspace clustering

- Sparse Representation:

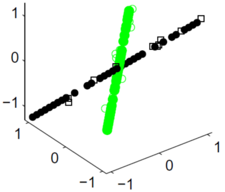
$$\min_{z} \|z\|_0 \quad s.t. \quad x = Az. \tag{7}$$

- Sparse subspace clustering: self-expressiveness of the data

$$\min_{z} \|z\|_0 \quad s.t. \quad x_i = Xz_i, z_{ii} = 0. \tag{8}$$

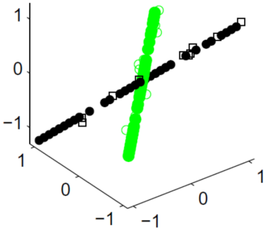
or

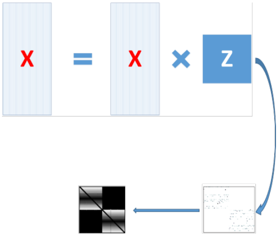
$$\min_{Z} \|Z\|_0 \quad s.t. \quad X = XZ, \mathrm{diag}(Z) = 0. \tag{9}$$



Sparse subspace clustering

$$\min_{Z} \|Z\|_1 \quad s.t. \quad X = XZ, \mathrm{diag}(Z) = 0. \tag{10}$$

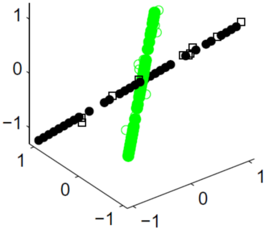


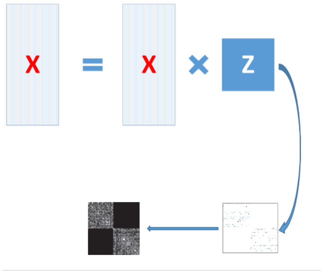


Elhamifar and Vidal, Sparse Subspace Clustering, CVPR 2009.

Low-rank Representation

$$
\begin{aligned}
&\text{Clean data:} && \min_{Z} \|Z\|_* \quad s.t. \quad X = XZ. \\
&\text{Noisy data:} && \min_{Z} \|Z\|_* + \lambda \|E\|_{2,1} \quad s.t. \quad X = XZ + E.
\end{aligned}
\qquad (11)
$$





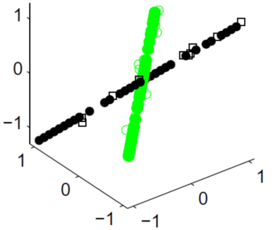
Liu et al., Robust Subspace Segmentation by Low-Rank Representation, ICML 2010.

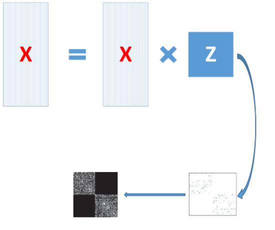
Liu et al., Robust Recovery of Subspace Structures by Low-Rank Representation, IEEE T. PAMI, 2013.

The nuclear norm is not accurate for rank approximation.

$$\min_{Z} \log\det(I + Z^T Z) + \lambda \|S\|_1 + \gamma \|E\|_F^2 \quad s.t. \quad X = XZ + S + E. \quad (12)$$

Here $$\log\det(I + Z^T Z) = \sum_i \log(1 + \sigma_i^2(Z)) \quad (13)$$





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

Thresholding Ridge Regression

SSC and LRR eliminates the noise effects from the data, where some prior knowledge is required.

$$\min_{z_i} \|x_i - X_{\hat{i}} z_i\|_2^2 + \lambda \|z_i\|_2^2 \tag{14}$$

where

$$X_{\hat{i}} = [x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n]. \tag{15}$$

OR

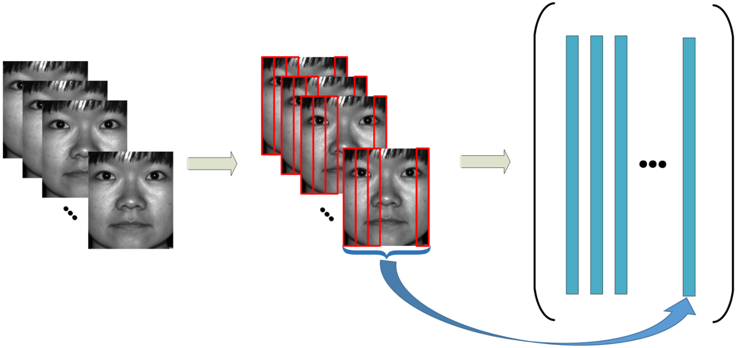
$$\min_{Z} \|X - XZ\|_F^2 + \lambda \|Z\|_F^2, \quad s.t. \quad \mathrm{diag}(Z) = 0. \tag{16}$$

TRR eliminates the noise effect by thresholding small values in Z.

Peng et al., Robust subspace clustering via thresholding ridge regression, AAAI 2015.

Variance Regularized Ridge Regression

Existing methods usually convert 2D data to vectorial (1-dimensional) data!



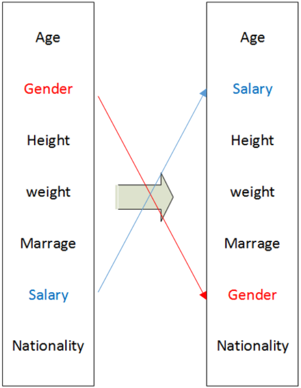
2-Dimensional Data

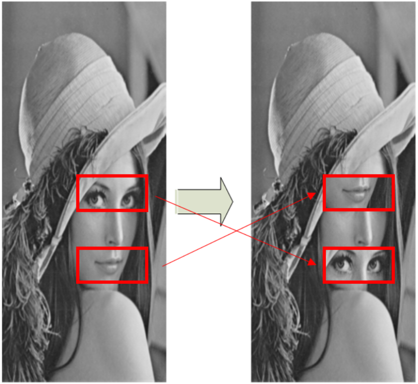
2-dimensional (2D) data means that each example is a 2D matrix.
Examples:

- A gray image is a matrix;
- Each frame of a video sequence;
- A user-item in recommender system is changing over time;
- A community in a social network is changing over time;
- There is often a need to partition a scene into multiple segments;
- Satellite images bring daily weather reports and provide farmers with information for precision agriculture;
- Real estate sales use geographic information systems.

Spatial Information

Spatial information identifies the geographical location of the features and reveals the inherent structures of the data.





Spatial Information

Why not tensor method?

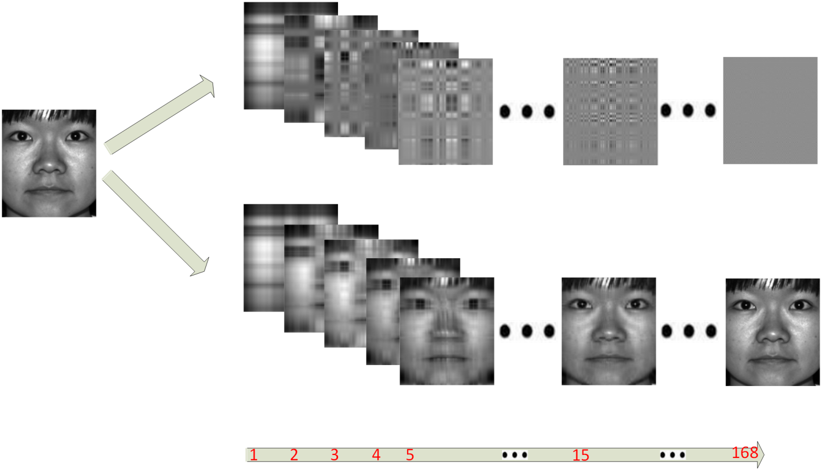
- For candecomp/parafac (CP) decomposition based methods, it is generally NP-hard to compute the CP rank;
- Tucker decomposition is not unique;
- The application of a core tensor and a high-order tensor product would incur information loss of spatial details;
- Tensor computation and methods usually involve flattening and folding operations, which, more or less, have issues similar to those of vectorization operation and thus might not fully exploit the true structures of the data.

Lu et al., Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization, CVPR 2016.

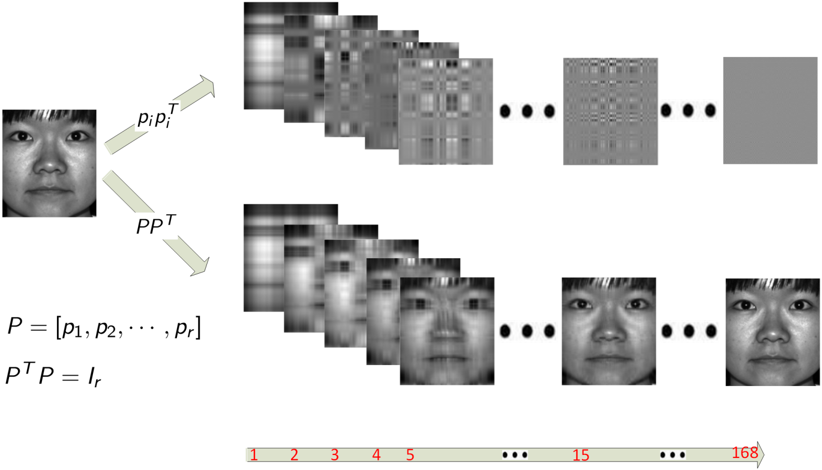
Kolda et al., Tensor decompositions and applications, SIAM Review, 2009.

Letexier et al., Noise removal from hyperspectral images by multidimensional filtering, IEEE T-GRS, 2008

2-Dimensional Features



2-Dimensional Projections



$P = [p_1, p_2, \cdots, p_r]$

$P^T P = I_r$

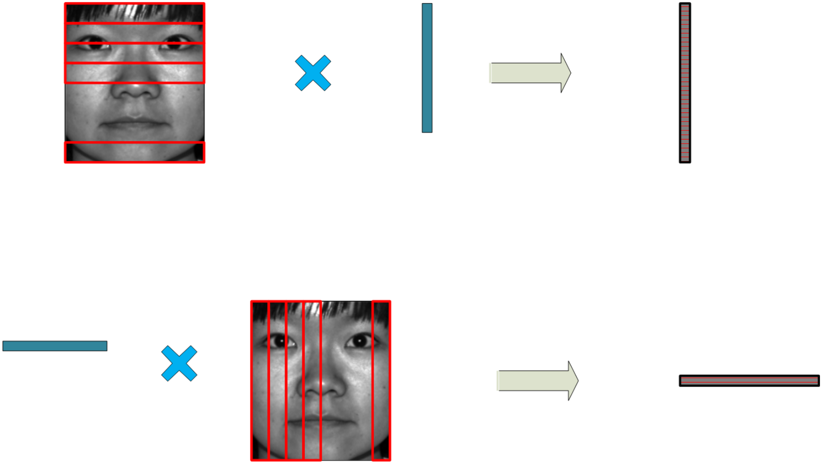
Variance Regularized Ridge Regression (VR3)

Construct the coefficients for low-dimensional representation with projected data:

$$\min_{Z, P^T P = I_r} \sum_{i=1}^{n} \left\| X_i P - \sum_{j=1}^{n} z_{ji} X_j P \right\|_F^2 + \tau \|Z\|_F^2 + \gamma_1 \mathrm{Tr}(P^T G_P P), \qquad (17)$$

where G_P is inverse of 2D covariance matrix of X_i's.

Variance Regularized Ridge Regression (VR3)



Variance Regularized Ridge Regression (VR3)

Preserve spatial information by seeking projection directions from both horizontal and vertical directions:

$$\min_{Z, P^T P = I_r, Q^T Q = I_r} \sum_{i=1}^{n} \left\| X_i P - \sum_{j=1}^{n} z_{ji} X_j P \right\|_F^2 + \gamma_1 \mathrm{Tr}(P^T G_P P) + \tau \|Z\|_F^2 \\ \underbrace{+ \sum_{i=1}^{n} \left\| X_i^T Q - \sum_{j=1}^{n} z_{ji} X_j^T Q \right\|_F^2 + \gamma_2 \mathrm{Tr}(Q^T G_Q Q)}_{\text{counter part to } P}, \tag{18}$$

where G_Q is inverse of 2D covariance matrix of X_j^T's.

Optimization-Q

The subproblem of optimizing Q is

$$\min_{Q^TQ=I_r} \sum_{i=1}^{n} \left\| X_i^T Q - \sum_{j=1}^{n} z_{ji} X_j^T Q \right\|_F^2 + \gamma_2 \mathrm{Tr}(Q^T G_Q Q). \tag{19}$$

Theorem 1

Define $F_1 = \sum_{i=1}^n X_i X_i^T$, $F_2 = \sum_{i=1}^n \sum_{j=1}^n z_{ji} X_i X_j^T$, and $F_3 = \sum_{i=1}^n \sum_{j=1}^n z_{(i)} z_{(j)}^T X_i X_j^T$. The problem of (19) is a constrained quadratic optimization, and admits a closed-form solution,

$$\mathbf{eig}_r(F_1 - 2F_2 + F_3 + \gamma_2 G_Q), \tag{20}$$

where $F_1 - 2F_2 + F_3 + \gamma G_Q$ is positive definite and $\mathbf{eig}_r(F)$ returns eigenvectors of F associated with its r smallest eigenvalues.

Optimization-P

The subproblem of optimizing P is

$$\min_{P^T P = I_r} \sum_{i=1}^{n} \left\| X_i P - \sum_{j=1}^{n} z_{ji} X_j P \right\|_F^2 + \gamma_1 \mathrm{Tr}(P^T G_P P). \tag{21}$$

Theorem 2

Define $H_1 = \sum_{i=1}^n X_i^T X_i$, $H_2 = \sum_{i=1}^n \sum_{j=1}^n z_{ji} X_i^T X_j$, *and* $H_3 = \sum_{i=1}^n \sum_{j=1}^n z_{(i)} z_{(j)}^T X_i^T X_j$. *The problem of* (21) *is a constrained quadratic optimization and admits a closed-form solution,*

$$\mathbf{eig}_r(H_1 - 2H_2 + H_3 + \gamma_1 G_P), \tag{22}$$

where $H_1 - 2H_2 + H_3 + \gamma_1 G_P$ *is positive definite.*

Optimization-Z

The subproblem of optimizing Z is

$$\min_{Z} \sum_{i=1}^{n} \left\| X_i P - \sum_{j=1}^{n} z_{ji} X_j P \right\|_F^2 + \sum_{i=1}^{n} \left\| X_i^T Q - \sum_{j=1}^{n} z_{ji} X_j^T Q \right\|_F^2 + \tau \|Z\|_F^2. \tag{23}$$

We define a matrix $\mathcal{K}$ with

$$\mathcal{K}_{ij} = \mathrm{Tr}(P^T (X_i^T X_j) P) + \mathrm{Tr}(Q^T (X_i X_j^T) Q),$$

then it is seen that (23) admits a closed-from solution:

$$Z = (\mathcal{K} + \tau I_n)^{-1} \mathcal{K}. \tag{24}$$

Convergence Analysis of VR3

Theorem 3

Denoting the objective function of (18) *by* $\mathcal{J}(Q, P, Z)$, *under the updating rules of* (20), (22) *and* (24), *the value sequence of the objective function* $\left\{\mathcal{J}(Q^k, P^k, Z^k)\right\}_{k=1}^{\infty}$ *is non-increasing and converges, where* k *denotes the iteration number.*

Nonlinear VR3

To further account for nonlinear relationships of the data, we try to ensure the smoothness between linear and nonlinear spaces on manifold:

$$\min_{Z, P^T P = I_r, Q^T Q = I_r} \mathcal{J}(Q, P, Z) \underbrace{+ \eta_1 \mathrm{Tr}(Z L_P Z^T) + \eta_2 \mathrm{Tr}(Z L_Q Z^T)}_{\text{Learning on manifold}}, \tag{25}$$

where L_P and L_Q are constructed using projected data $X_i P P^T$ and $Q Q^T X_i$, respectively.

• Fully connected graphs:
$[W_P]_{ij} = \mathrm{Tr}((X_i P)^T (X_j P))$ and $[W_Q]_{ij} = \mathrm{Tr}((X_i^T Q)^T (X_i^T Q))$.

Optimiation-Q

The subproblem for Q-minimization is

$$\min_{Q^T Q = I_r} \sum_{i=1}^{n} \left\| X_i^T Q - \sum_{j=1}^{n} z_{ji} X_j^T Q \right\|_F^2 + \gamma_2 \mathrm{Tr}(Q^T G_Q Q) + \eta_2 \mathrm{Tr}(Z L_Q Z^T). \tag{26}$$

Theorem 4

Define $F_4 = \sum_{i=1}^n \sum_{j=1}^n \|z_i - z_j\|_2^2 X_i X_j^T$*. Given that* Z *is bounded, the problem in (26) is a constrained quadratic optimization, and admits a closed-form solution,*

$$\mathbf{eig}_r \left(F_1 - 2F_2 + F_3 + \gamma_2 G_Q + \frac{\eta_2}{2} F_4 \right), \tag{27}$$

where F_1*,* F_2*, and* F_3 *are defined in Theorem 1.*

Optimiation-P

The subproblem of optimizing P is

$$\min_{P^TP=I_r} \sum_{i=1}^{n} \left\| X_i P - \sum_{j=1}^{n} z_{ji} X_j P \right\|_F^2 + \gamma_1 \mathrm{Tr}(P^T G_P P) + \eta_1 \mathrm{Tr}(Z L_P Z^T). \tag{28}$$

Theorem 5

Define $H_4 = \sum_{i=1}^{n} \sum_{j=1}^{n} \|z_i - z_j\|_2^2 X_i^T X_j$. *Given that* Z *is bounded, the problem of* (28) *is a constrained quadratic optimization, and admits a closed-form solution,*

$$\mathbf{eig}_r \left(H_1 - 2H_2 + H_3 + \gamma_1 G_P + \frac{\eta_1}{2} H_4 \right), \tag{29}$$

where H_1, H_2, *and* H_3 *are defined in Theorem 2.*

Optimiation-Z

The subprblem of optimizing Z is

$$\min_{Z} \sum_{i=1}^{n} J(Q,P,Z) + \eta_1 \mathrm{Tr}(ZL_P Z^T) + \eta_2 \mathrm{Tr}(ZL_Q Z^T). \tag{30}$$

$$\textit{Solution} \qquad Z = \mathrm{lyap}\Big(\mathcal{K} + \frac{\tau}{2} I_n, \eta_1 L_P + \eta_2 L_Q + \frac{\tau}{2} I_n, \mathcal{K}\Big). \tag{31}$$

Theorem 6

If $$\tau \geq \Bigg\{ -\eta_1 r\Big(\min_i \{\lambda_b(\sum_{j=1}^{n} X_i^T X_j)\} - \lambda_1(\sum_{j=1}^{n} X_j^T X_j)\Big) - \eta_2 r\Big(\min_i \{\lambda_a(\sum_{j=1}^{n} X_i X_j^T)\} - \lambda_1(\sum_{j=1}^{n} X_j X_j^T)\Big)\Bigg\},$$

then (31) *is bounded and is the optimal solution to* (30)*. Here,* $\lambda_i(\cdot)$ *is the* i*th largest eigenvalue.*

Convergence Analysis of NVR3

Theorem 7

Let

$$\mathcal{L}(Q, P, Z) = \mathcal{J}(Q, P, Z) + \eta_1 Tr(ZL_P Z^T) + \eta_2 Tr(ZL_Q Z^T)$$

denote the objective function of (25). *If the condition of Theorem 6 is satisfied, then under the updating rules of* (27), (29) *and* (31), $\left\{\mathcal{L}(Q^k, P^k, Z^k)\right\}_{k=1}^{\infty}$ *is non-increasing and converges, where* k *denotes the iteration number.*

Experimental Results

Table 5: Clustering Performance on Extended Yale B data set

No. of Subjects	2 Subjects		3 Subjects		5 Subjects		8 Subjects		10 Subjects	
Accuracy (%)	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
LSA	67.20	52.34	47.71	50.00	41.98	43.13	40.81	41.41	39.58	42.50
SCC	83.38	92.18	61.84	60.94	41.10	40.62	33.89	35.35	26.98	24.22
LRR	90.48	94.53	80.48	85.42	65.84	65.00	58.81	56.25	61.15	58.91
LRR-H	97.46	99.22	95.79	97.40	93.10	94.37	85.66	89.94	77.08	76.41
LRSC	94.68	95.31	91.53	92.19	87.76	88.75	76.28	71.97	69.64	71.25
SSC	98.14	**100.0**	96.90	98.96	95.69	97.50	94.15	95.51	89.06	94.37
LatLRR	97.46	99.22	95.79	97.40	93.10	94.37	85.66	89.94	77.08	76.41
BDLRR	96.09	–	89.98	–	87.03	–	72.30	–	69.16	–
BDSSC	96.10	–	82.30	–	72.50	–	66.80	–	60.47	–
S^3C	98.57	**100.0**	96.91	**99.48**	95.92	97.81	95.16	95.90	93.91	94.84
NSN	98.29	99.22	96.37	96.88	94.19	95.31	91.54	92.38	90.18	90.94
TRR	97.87	99.22	97.07	98.44	96.17	97.50	95.69	96.48	95.10	95.78
VR3	**99.12**	**100.0**	99.23	**99.48**	98.96	**99.38**	98.73	98.63	98.85	98.75
NVR3	99.07	**100.0**	**99.26**	**99.48**	**99.25**	**99.38**	**99.16**	**99.22**	**99.38**	**99.38**

Table 6: Clustering Performance on Alphadigits data set

No. of Subjects	2 Subjects		3 Subjects		5 Subjects		8 Subjects		10 Subjects	
Accuracy (%)	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
LSA	89.30	96.15	77.31	77.78	66.19	66.15	59.24	59.94	57.35	58.72
SSC	94.30	97.44	86.42	91.46	76.74	74.88	70.00	69.99	67.86	67.18
LRR	92.24	96.16	85.79	88.89	76.66	76.41	69.50	69.56	66.33	67.44
LRSC	84.19	91.03	74.35	74.36	62.23	62.05	52.02	51.92	49.23	48.97
KSSC (P)	94.58	97.44	87.15	92.31	77.36	76.92	68.94	67.95	66.15	65.64
KSSC (G)	94.07	97.44	86.36	91.45	76.16	73.85	68.81	68.91	67.52	66.67
LS3C	93.77	96.15	87.33	90.17	74.76	73.85	65.94	66.03	62.99	63.51
S^3C	93.34	94.87	86.34	88.03	72.89	71.28	64.17	65.06	62.82	61.79
TRR	95.60	97.44	90.71	93.59	81.02	**83.59**	72.06	72.44	68.38	69.49
VR3	96.10	98.72	91.14	94.02	81.19	83.08	73.13	74.04	73.85	76.67
NVR3	**96.21**	**98.72**	**91.84**	**94.87**	**81.57**	**83.59**	**73.27**	**74.04**	**76.41**	**76.92**

Parameter Sensitivity

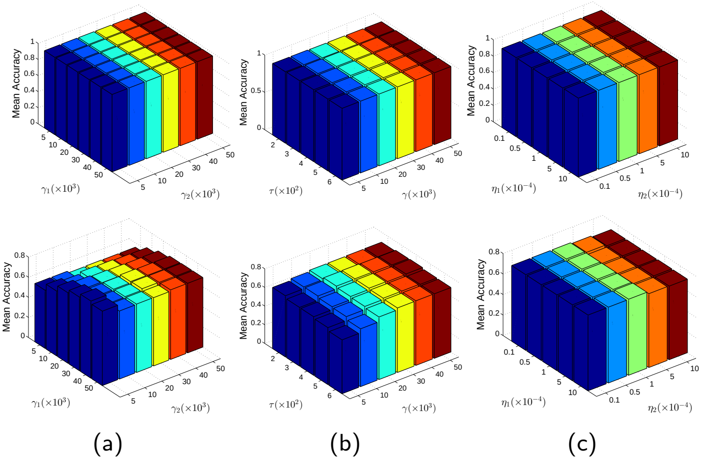


Figure 11: Performance of VR3 and NVR3. $K = 2$ on the top while $K = 10$ on the bottom. Performance of NVR3 with respect to different combinations of: (a) γ_1 and γ_2; (b) τ and γ; (c) η_1 and η_2.

Extracted 2D Features and Recovered Data

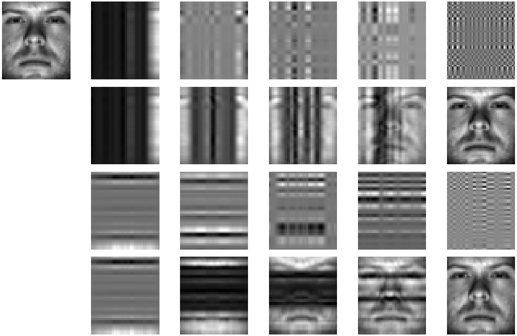


Figure 12: The top left is the original image. For the rest, the first (resp. third) row are the ith column (row) component image $Xp_ip_i^T$ (resp. $q_iq_i^TX$), and the second (resp. fourth) row are the reconstructed images $\sum_{j=1}^{i} Xp_jp_j^T$ (resp. $\sum_{j=1}^{i} q_jq_j^TX$) using the first i column (resp. row) component images, which from left to right represents $i = 1, 3, 8, 15$, and 30, respectively.

K-means

$$\text{K-means:} \qquad \min_{U,V} \|Y - UV^T\|_F^2, \quad s.t. \quad v_{ij} \geq 0, VV^T = I_k. \tag{32}$$

- K-means is the most widely used clustering method and expanding it for 2D data is attractive.
- Is it possible to expand K-means for 2D data? How?

Double-Sided Two-Dimensional K-Means

$$\min_{\mathbf{U},V,P,Q} \sum_{i=1}^{n} \left\{ \left\| X_i PP^T - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 + \left\| QQ^T X_i - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 \right\}$$
$$+ \lambda \left\{ \sum_{i=1}^{n} \left\| X_i - X_i PP^T \right\|_F^2 + \sum_{i=1}^{n} \left\| X_i - QQ^T X_i \right\|_F^2 \right\} \tag{33}$$
$$+ \gamma \left\{ \mathrm{Tr}(V^T L_P V) + \mathrm{Tr}(V^T L_Q V) \right\},$$
$$s.t. \quad \mathbf{U} = \{U_1, \cdots, U_k\}, P^T P = I_r, Q^T Q = I_r, VV^T = I_k, v_{ij} \geq 0.$$

Double-Sided Two-Dimensional K-Means

$$
\begin{aligned}
&\min_{\mathbf{U},V,P,Q} \sum_{i=1}^{n} \left\{ \left\| X_i P P^T - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 + \left\| Q Q^T X_i - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 \right\} \\
&\quad + \lambda \left\{ \sum_{i=1}^{n} \left\| X_i - X_i P P^T \right\|_F^2 + \sum_{i=1}^{n} \left\| X_i - Q Q^T X_i \right\|_F^2 \right\} \\
&\quad + \gamma \left\{ \mathrm{Tr}(V^T L_P V) + \mathrm{Tr}(V^T L_Q V) \right\}, \\
&s.t. \quad \mathbf{U} = \{U_1, \cdots, U_k\}, P^T P = I_{r_1}, Q^T Q = I_{r_2}, r_1 + r_2 = 2r, \\
&\qquad\ V V^T = I_k, v_{ij} \geq 0.
\end{aligned}
\tag{34}
$$

Closely connected with some existing methods, such as K-means, spectral clustering, and 2DPCA.

Optimization

Augmented Lagrange Multiplier (ALM)-based Optimization.

$$
\begin{aligned}
\min_{\mathbf{U},V,P,Q} & \sum_{i=1}^{n} \left\{ \left\| X_i PP^T - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 + \left\| QQ^T X_i - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 \right\} \\
& + \lambda \left\{ \sum_{i=1}^{n} \left\| X_i - X_i PP^T \right\|_F^2 + \sum_{i=1}^{n} \left\| X_i - QQ^T X_i \right\|_F^2 \right\} \\
& + \gamma \left\{ \mathrm{Tr}(V^T L_P V) + \mathrm{Tr}(V^T L_Q V) \right\} + \frac{\rho}{2} \left\| Z - V + \Theta/\rho \right\|_F^2, \\
s.t. \quad & \mathbf{U} = \{U_1, \cdots, U_k\}, P^T P = I_{r_1}, Q^T Q = I_{r_2}, r_1 + r_2 = 2r, \\
& VV^T = I_k, z_{ij} \geq 0.
\end{aligned}
\tag{35}
$$

Optimization-Jointly for $\{P, Q, r_1, r_2\}$

$\mathbf{H} := \begin{bmatrix} \mathbf{H}_{11} & 0 \\ 0 & \mathbf{H}_{22} \end{bmatrix}$, where

$$\mathbf{H}_{11} := (1-\lambda_1)G_P - \sum_{i=1}^{n}(X_i^T\Theta_i + \Theta_i^T X_i),$$

$$\mathbf{H}_{22} := (1-\lambda_1)G_Q - \sum_{i=1}^{n}(X_i\Theta_i^T + \Theta_i X_i^T),$$

and $\Theta_i := \sum_{j=1}^{k} U_j v_{ji}$.

$$\textit{Solution} \quad \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} = F = \mathbf{perm}(\mathbf{eig}_{2r}(\mathbf{H}+\xi I)) = \mathbf{perm}(\mathbf{eig}_{2r}(\mathbf{H})), \quad (36)$$

where $\mathbf{perm}(\cdot)$ is a permutation of the matrix columns, which ensures the block diagonal structure of F.

Optimization-V

Reduced to

$$\min_{VV^T=I_k} \|V-E\|_F^2, \tag{37}$$

where

$$E := \left(\vec{\mathbf{v}}(\mathbf{U})\right)^T \left(\vec{\mathbf{v}}(\mathbf{X}^P)+\vec{\mathbf{v}}(\mathbf{X}^Q)\right) - \frac{\lambda_2}{2} Z\left(L_P + L_Q\right) + \frac{\rho}{2}\left(Z+\Theta/\rho\right). \tag{38}$$

$$\mathit{Solution} \qquad V = AB^T, \tag{39}$$

where A and B are matrices containing the left and right singular vectors of E, respectively.

Optimization-Z

$$\min_{z_{ij} \geq 0} \left\| Z - \left(V - \frac{\Theta}{\rho} - \frac{\lambda}{\rho} V(L_P^T + L_Q^T) \right) \right\|_F^2 . \tag{40}$$

$$\textit{Solution} \qquad Z = \left(V - \frac{\Theta}{\rho} - \frac{\lambda_2}{\rho} V(L_P^T + L_Q^T) \right)_+ , \tag{41}$$

where $(\cdot)_+$ is element-wisely defined by $(\theta)_+ = \frac{1}{2}(|\theta| + \theta)$.

Optimization-$\mathbf{U}$, Θ, and ρ

$\mathbf{U}$-minimization subproblem is:

$$\min_{\mathbf{U}} \sum_{i=1}^{n} \left\{ \left\| X_i P P^T - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 + \left\| Q Q^T X_i - \sum_{j=1}^{k} U_j v_{ij} \right\|_F^2 \right\} \tag{42}$$

$$\textit{Solution} \qquad \mathbf{U} = \left\{ \frac{1}{2} \sum_{i=1}^{n} (X_i P P^T + Q Q^T X_i) v_{ij} \right\}_{j=1}^{k}. \tag{43}$$

$$\Theta = \Theta + \rho(Z - V), \quad \rho = \rho\kappa. \tag{44}$$

Table 7: Clustering Performance on Yale

N	Accuracy (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	51.36±02.20	71.36±22.78	70.00±10.97	71.36±19.05	73.64±24.97	73.18±223.61	53.64±04.18	86.36±13.03	**92.27±07.44**
3	40.00±08.43	61.82±15.20	62.73±11.52	52.12±13.08	63.33±21.40	65.45±18.75	52.12±04.24	72.12±12.10	**84.55±06.62**
4	30.91±05.79	46.14±11.54	58.18±12.69	44.09±06.96	49.55±11.33	48.41±10.28	40.45±07.93	60.45±10.45	**78.41±12.96**
5	28.18±06.43	50.00±08.45	52.73±12.03	50.73±16.47	53.27±09.35	54.36±19.59	39.82±05.65	58.73±11.37	**70.91±14.97**
6	23.64±05.40	50.61±04.91	43.94±05.39	37.58±07.92	53.18±03.53	51.97±04.04	33.64±03.90	51.36±04.07	**65.15±09.56**
7	22.73±06.13	47.27±06.57	47.53±07.75	39.61±11.19	54.55±04.37	51.95±05.09	34.81±06.12	52.73±04.02	**67.01±06.40**
8	21.36±04.54	48.30±05.55	46.25±08.02	37.73±10.67	53.98±04.92	52.61±06.16	32.95±05.59	49.89±04.68	**63.86±07.42**
9	19.29±03.31	45.45±07.68	45.15±06.95	32.53±06.28	54.44±02.67	48.28±05.92	30.00±03.20	50.30±06.85	**58.59±04.97**
10	17.64±03.30	43.36±04.29	41.82±06.23	28.91±05.71	50.00±04.56	46.45±04.91	28.73±02.36	46.73±03.81	**57.64±05.47**
12	16.97±01.72	40.83±04.23	43.03±04.31	30.30±03.50	49.92±05.74	46.44±03.03	27.80±01.43	45.08±06.04	**52.05±04.07**
14	15.39±00.31	41.75±02.82	37.92±04.10	25.52±03.02	45.65±02.87	46.04±02.09	25.06±00.98	44.61±03.33	**51.62±02.96**
15	14.55	39.39	43.03	21.82	44.24	43.03	21.82	44.85	**52.12**
Average	25.17	48.88	49.36	39.36	53.81	52.35	35.07	55.27	**66.13**

N	Normalized Mutual Information (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	01.41±02.27	35.14±42.14	17.22±15.69	30.43±29.47	43.23±46.31	39.63±42.56	01.15±01.90	52.87±29.87	**67.87±27.05**
3	10.49±16.72	33.94±22.01	33.19±16.80	25.24±20.95	43.87±31.44	41.45±28.65	18.91±10.05	46.94±18.56	**62.74±12.59**
4	09.91±11.06	23.75±16.61	37.56±14.27	21.25±11.23	33.70±15.98	28.47±14.65	14.95±10.54	39.72±12.17	**63.39±16.14**
5	14.66±12.12	35.68±14.72	36.96±15.16	33.17±22.40	44.13±12.21	39.87±26.15	23.49±09.30	44.78±15.32	**61.59±17.07**
6	13.48±09.35	40.74±07.93	33.76±07.42	21.77±12.82	44.67±05.95	42.47±06.28	20.54±07.17	38.37±06.02	**57.72±09.79**
7	16.59±08.73	39.44±07.78	40.62±07.07	29.02±16.15	48.63±06.08	46.31±05.75	27.41±08.44	43.89±06.09	**63.12±07.54**
8	16.60±06.88	43.51±05.72	43.17±09.20	33.19±14.25	49.65±06.68	47.97±08.00	26.32±06.87	44.85±05.35	**61.17±06.62**
9	15.94±05.53	42.94±08.06	41.48±06.11	25.19±08.62	51.76±03.55	46.80±06.43	26.60±05.05	45.97±06.09	**57.43±05.23**
10	15.70±04.55	43.73±03.17	41.53±05.41	23.94±06.23	49.02±05.39	46.39±05.34	27.50±03.24	44.59±04.10	**58.00±05.78**
12	16.82±02.40	42.83±04.26	44.42±04.20	31.53±05.45	51.76±02.82	50.24±02.37	30.29±01.43	46.56±04.94	**55.99±03.03**
14	16.82±00.13	46.20±02.59	44.88±03.26	28.78±04.08	50.86±02.13	51.36±01.95	30.63±00.76	47.77±02.04	**55.53±02.72**
15	16.40	43.84	46.95	26.86	49.64	52.56	29.25	48.20	**52.59**
Average	13.74	39.31	38.48	27.53	46.74	44.46	23.09	45.38	**59.76**

Experimental Results

Table 8: Clustering Performance on PIX

N	Accuracy (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	94.50±10.39	94.50±10.39	99.50±01.58	94.50±10.39	99.50±01.58	99.50±01.58	73.00±11.11	96.50±07.84	**100.0±00.00**
3	85.33±14.42	96.00±05.84	96.67±03.85	95.00±06.89	96.00±05.84	97.67±06.30	60.67±09.27	97.33±03.06	**99.67±01.05**
4	87.00±22.26	96.25±04.60	77.00±13.43	90.50±13.58	97.25±03.81	99.25±01.69	59.25±12.47	96.50±04.44	**99.75±00.79**
5	72.80±24.37	87.20±11.00	75.40±11.85	71.80±13.62	90.80±09.34	95.40±08.17	55.00±10.38	90.80±07.50	**99.40±00.97**
6	63.33±24.47	84.83±12.38	74.17±12.65	72.83±12.74	90.17±09.51	90.00±11.92	47.33±05.89	89.00±08.72	**98.17±04.19**
7	68.14±23.00	84.14±05.89	71.86±05.68	65.57±08.31	90.27±07.64	**94.86±05.64**	51.29±06.48	87.14±07.85	94.43±05.53
8	61.25±08.27	84.12±05.65	78.13±04.18	58.00±06.07	87.25±0714	**95.25±02.27**	51.38±03.79	82.37±05.38	88.50±05.92
9	72.11±09.83	83.22±07.45	88.89±04.89	57.11±04.03	92.89±00.57	**96.44±00.70**	45.33±05.02	87.00±06.83	94.33±05.53
10	73.00	80.00	70.00	62.00	80.00	87.00	11.00	81.00	**95.00**
Average	75.27	87.81	81.29	74.15	91.57	95.04	50.47	89.74	**96.58**

N	Normalized Mutual Information (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	83.81±28.77	83.81±28.77	97.58±07.64	83.81±28.77	97.58±07.64	97.58±07.64	25.16±21.49	88.28±22.45	**100.0±00.00**
3	77.98±11.64	89.87±11.64	90.84±08.94	88.11±14.96	89.87±11.64	94.95±12.80	34.04±13.10	92.32±08.18	**98.98±03.22**
4	86.63±07.13	93.39±07.13	74.99±12.76	86.82±16.72	94.67±05.78	98.42±03.44	42.33±16.08	93.45±07.08	**98.75±02.01**
5	75.40±07.04	87.42±08.35	77.38±10.89	67.15±14.69	90.13±07.54	93.60±10.71	47.52±10.15	88.04±07.35	**98.04±03.78**
6	70.44±05.09	86.71±08.47	73.30±08.81	69.75±11.39	90.64±05.68	91.12±08.83	45.19±07.04	87.05±07.23	**93.71±05.88**
7	75.50±04.18	86.74±03.63	77.56±06.43	64.08±09.35	91.54±04.57	**93.92±04.67**	49.54±08.12	87.06±07.05	92.46±02.09
8	72.33±01.28	85.70±03.87	81.73±04.35	60.00±05.12	90.56±03.07	93.29±02.72	49.33±03.90	83.54±04.15	**93.50±03.19**
9	80.16±01.76	86.31±04.83	87.91±06.31	60.67±03.91	92.78±00.82	**94.76±00.94**	45.26±03.24	87.89±04.59	94.31±03.27
10	81.42	88.09	69.29	62.45	87.01	92.52	11.73	86.02	**95.85**
Average	78.19	87.56	81.18	71.43	91.64	94.46	38.90	88.18	**96.83**

Experimental Results

Table 9: Clustering Performance on 40% Corrupted ORL

N	Accuracy (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	60.00±14.14	78.50±15.82	55.00±00.00	———	65.50±09.85	77.50±20.03	61.00±06.58	76.50±16.67	**93.50±09.73**
4	31.00±03.16	49.75±09.68	37.50±00.00	———	49.50±06.65	59.00±10.08	38.75±04.45	53.50±10.01	**68.75±12.09**
6	23.83±01.37	36.33±07.89	31.67±00.00	———	43.67±07.28	48.17±12.36	31.67±02.48	45.50±10.45	**68.67±17.12**
8	18.50±01.84	30.38±04.72	31.25±00.00	———	41.25±05.56	42.75±06.42	29.75±04.16	34.25±04.87	**59.12±06,92**
10	15.90±00.88	29.27±03.58	27.00±00.00	———	34.50±04.22	39.40±05.13	11.00±00.00	33.10±03.60	**53.20±07.39**
12	13.67±01.43	27.42±02.40	22.50±00.00	———	33.50±04.02	34.50±04.93	23.33±01.80	30.75±02.65	**47.17±04.16**
14	12.14±00.95	25.50±02.78	21.43±00.00	———	32.64±03.63	36.21±04.73	23.29±01.31	28.50±02.98	**44.00±02.84**
16	11.37±00.77	23.81±02.80	21.88±00.00	———	29.25±02.55	33.88±03.06	21.69±00.84	25.94±01.57	**43.19±04.73**
18	10.61±00.81	23.06±02.03	19.44±00.00	———	30.72±03.74	32.06±02.80	21.39±01.29	25.83±01.58	**44.00±04.44**
20	10.20±00.89	22.10±02.00	20.00±00.00	———	27.70±03.10	31.95±03.07	19.85±01.36	26.35±01.43	**39.45±03.64**
Average	20.72	34.61	28.77	———	35.57	43.54	25.84	38.02	**55.70**

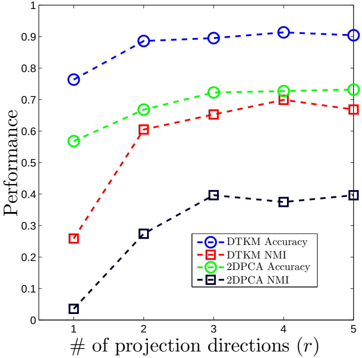
N	Normalized Mutual Information (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	14.62±30.00	43.76±35.61	00.79±00.00	———	12.74±12.09	43.50±44.67	07.33±09.71	34.46±33.07	**76.60±30.23**
4	09.63±03.12	26.82±12.41	08.93±00.00	———	24.67±06.12	38.37±12.68	11.68±04.82	32.00±12.80	**54.17±14.51**
6	10.45±02.09	20.64±09.85	17.02±00.00	———	30.53±09.59	35.82±13.95	15.68±03.16	31.24±10.86	**61.41±20.70**
8	10.38±01.55	22.45±06.22	21.29±00.00	———	36.09±06.28	36.88±07.35	20.69±03.54	26.83±06.68	**56.09±06.98**
10	09.72±00.86	25.37±04.16	23.72±00.00	———	33.73±04.38	38.11±05.97	11.73±00.00	29.77±04.75	**54.80±05.89**
12	10.16±00.78	26.16±03.84	24.40±00.00	———	35.80±03.90	36.13±06.41	24.13±01.96	30.12±02.11	**51.80±04.46**
14	10.00±00.42	26.10±02.79	26.36±00.00	———	38.26±03.47	41.14±05.66	27.07±02.20	30.89±03.02	**50.78±02.45**
16	10.08±00.36	27.18±02.80	29.41±00.00	———	35.31±02.43	41.68±03.10	27.75±00.68	31.74±02.58	**51.45±04.41**
18	10.35±00.65	27.78±01.76	28.80±00.00	———	38.96±03.73	42.12±02.52	29.92±01.17	33.37±01.92	**49.46±04.50**
20	10.17±00.43	28.41±02.30	30.99±00.00	———	37.79±03.30	43.96±03.97	30.09±01.33	36.87±01.33	**49.75±03.64**
Average	10.56	27.47	21.17	———	32.39	39.77	20.61	31.74	**55.63**

Experimental Results

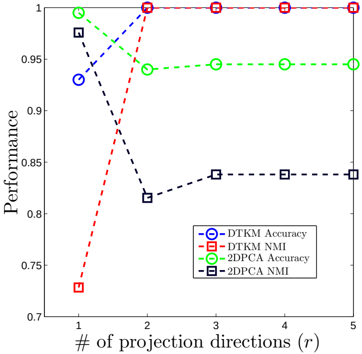
Table 10: Clustering Performance on 60% Corrupted ORL

N	Accuracy (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	55.00±00.00	72.50±13.18	55.00±00.00	————	55.00±00.00	67.00±12.74	59.00±06.58	66.50±11.56	**87.50±11.84**
4	30.50±01.05	42.25±05.83	37.50±00.00	————	27.50±00.00	44.00±08.51	38.00±03.07	44.75±06.06	**61.50±05.55**
6	23.50±01.46	33.00±04.50	31.67±00.00	————	18.33±00.00	37.00±09.26	31.50±03.80	35.50±04.91	**51.67±16.10**
8	18.75±01.95	29.25±01.79	31.25±00.00	————	13.75±00.00	32.25±03.27	28.50±02.75	29.13±02.95	**42.50±05.30**
10	16.70±01.42	26.20±01.69	27.00±00.00	————	11.00±00.00	28.30±03.56	11.00±00.00	26.30±02.36	**37.50±03.72**
12	14.58±01.06	23.92±02.39	22.50±00.00	————	10.00±00.00	27.17±02.49	24.33±01.88	24.33±02.63	**34.58±01.58**
14	12.93±01.09	23.36±01.69	21.43±00.00	————	08.57±00.00	25.21±02.36	22.14±00.89	24.14±02.23	**34.86±04.61**
16	12.06±00.84	22.00±01.21	21.88±00.00	————	07.50±00.00	24.69±02.29	21.25±01.41	21.94±01.65	**30.69±02.53**
18	11.44±00.65	20.94±01.17	19.44±00.00	————	06.67±06.00	24.22±01.05	21.06±01.21	21.33±01.18	**29.28±03.78**
20	10.60±00.57	21.10±01.76	20.00±00.00	————	06.00±00.00	23.65±02.59	20.75±01.32	20.85±01.93	**28.25±03.63**
Average	20.61	31.45	28.77	————	16.43	33.35	27.75	31.48	**43.83**
N	Normalized Mutual Information (%)								
	HCA	K-Means	KKM	SC	RPCA	2DPCA	NMF	RMNMF	DTKM
2	05.19±00.00	26.61±23.96	00.81±00.05	————	05.19±00.00	17.01±16.16	03.86±04.01	14.33±15.97	**56.93±36.99**
4	07.97±00.86	15.79±08.02	08.93±00.00	————	08.20±00.00	15.80±08.36	07.87±02.82	16.33±08.77	**38.29±08.99**
6	09.32±00.94	16.06±04.27	17.02±00.00	————	09.56±00.00	23.44±11.20	14.05±03.90	20.78±05.58	**41.03±19.23**
8	09.57±00.66	20.66±02.71	21.29±00.00	————	10.60±00.00	23.65±04.99	18.06±02.48	18.89±02.97	**35.96±04.88**
10	09.84±00.74	21.07±02.14	23.72±00.00	————	11.73±00.00	23.88±04.12	11.73±00.00	20.44±03.48	**36.60±04.20**
12	09.86±00.47	23.22±03.14	24.40±00.00	————	12.36±00.00	28.30±02.50	24.64±02.97	22.14±04.06	**36.21±02.61**
14	09.77±00.36	25.32±02.41	26.36±00.00	————	11.96±00.00	29.14±02.51	25.36±01.62	24.18±04.22	**39.28±04.05**
16	09.66±00.15	26.09±01.70	29.41±00.00	————	11.81±00.00	30.63±02.90	26.75±02.27	25.45±01.67	**37.04±02.36**
18	09.99±00.32	27.10±01.42	28.80±00.00	————	11.86±00.00	31.93±01.43	29.55±01.23	27.93±01.71	**38.29±03.80**
20	09.87±00.27	28.84±02.03	30.99±00.00	————	12.13±00.00	33.40±03.26	31.41±01.48	28.52±03.42	**38.04±03.52**
Average	09.10	23.08	21.17	————	10.54	25.72	19.33	21.90	**39.77**

Effect of Number of Projection Directions



(a) Yale



(b) PIX

Figure 13: Performance variations in accuracy and NMI with respect to different r values on Yale and PIX.

Extracted Features

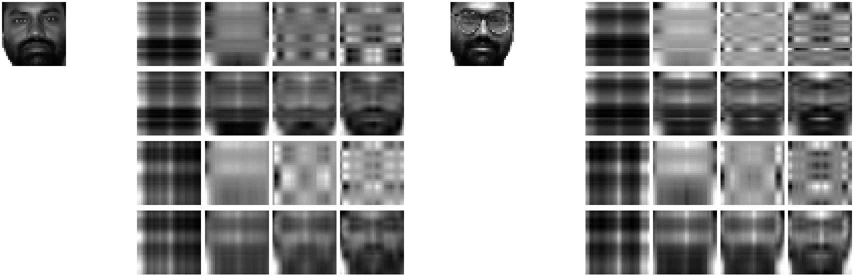
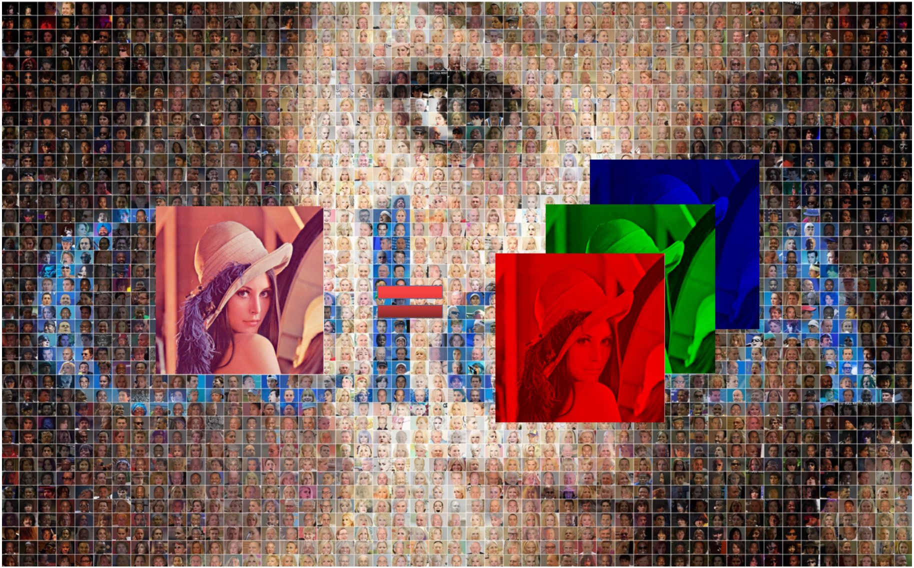


Figure 14: Extracted 2D features from two sample images of Yale data set. In each block of (a) and (b), the top left is the original image; in the right, from top to bottom are $Xp_ip_i^T$, $\sum_{j=1}^{i} Xp_ip_i^T$, $q_iq_i^TX$, and $\sum_{j=1}^{i} q_iq_i^TX$, i.e., the ith feature extracted by P, the recovered image by the top i projection directions of P, the ith feature extracted by Q, and the recovered image by the top i projection directions of Q, respectively. From left to right, i equals to 1, 2, 3, and 4, respectively.

Future Work




Questions

Thank You!