Robust recovery of low-rank subspaces

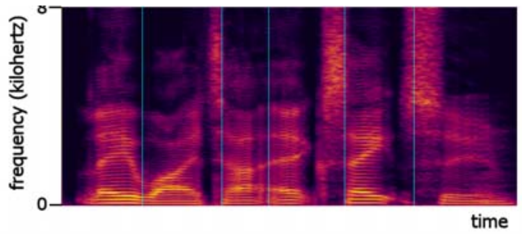
Yannis Panagakis

9/02/2014

Low-dimensional structures in high-dimensional data

Images

Videos



Music

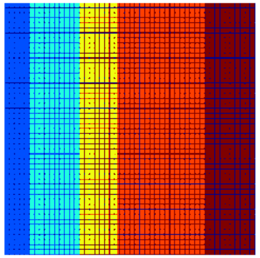
High-dimensional audio-visual data exhibit **low-dimensional (*low-rank, sparse, manifold, etc.*) structures** due to:

- **local** regularities,
- **global** symmetries,
- **repetitive** patterns,
- or **redundant** sampling.

Principal Component Analysis (PCA)

$$\mathbf{X} = \mathbf{A} + \mathbf{N}$$

- $\mathbf{X} \in \mathbb{R}^{m \times n}$: Observations matrix.
- $\mathbf{A} \in \mathbb{R}^{m \times n}$: Low-rank matrix, $r = \text{rank}(\mathbf{A}) \ll m$.
- $\mathbf{N} \in \mathbb{R}^{m \times n}$: Gaussian noise of small variance.

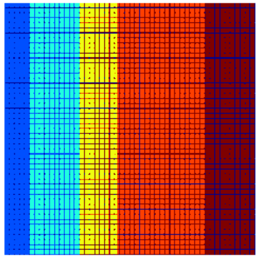


- Optimal estimate of the low- rank matrix under iid Gaussian noise.
- Efficient and scalable computation via SVD.
- Huge impact in image processing, vision, web search, etc.

Principal Component Analysis (PCA)

$$\mathbf{X} = \mathbf{A} + \mathbf{N}$$

- $\mathbf{X} \in \mathbb{R}^{m \times n}$: Observations matrix.
- $\mathbf{A} \in \mathbb{R}^{m \times n}$: Low-rank matrix, $r = \mathrm{rank}(\mathbf{A}) \ll m$.
- $\mathbf{N} \in \mathbb{R}^{m \times n}$: Gaussian noise of small variance.



- Optimal estimate of the low- rank matrix under iid Gaussian noise.
- Efficient and scalable computation via SVD.
- Huge impact in image processing, vision, web search, etc.

- **PCA breaks down under even a single corrupted observation.**

Real World Data

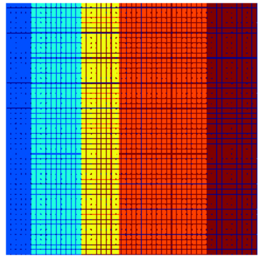
Real application data often contain:

- **missing observations,**
- **corruptions,**
- **unknown deformation,**
- **misalignment.**

- Classical methods (e.g., PCA, least square regression) **break down**.

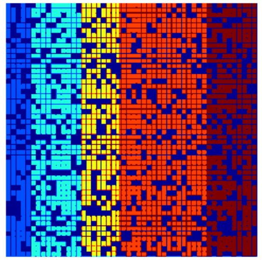
Low-Rank Models

- The data matrix should be **low-rank:**



$$\mathbf{A} \in \mathbb{R}^{m \times n}, r = \text{rank}(\mathbf{A}) \ll m$$

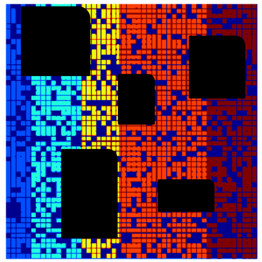
- but some of the observations are **grossly corrupted**:



$$\mathbf{A} + \mathbf{E},$$

$$|e_{ij}| \text{arbitrarily large, but most of them are zero.}$$

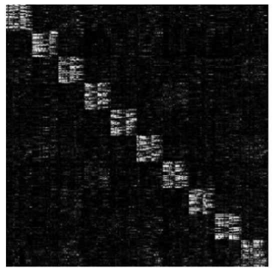
- and some of them can be **missing** too*:*



$$\mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{A} + \mathbf{E})$$

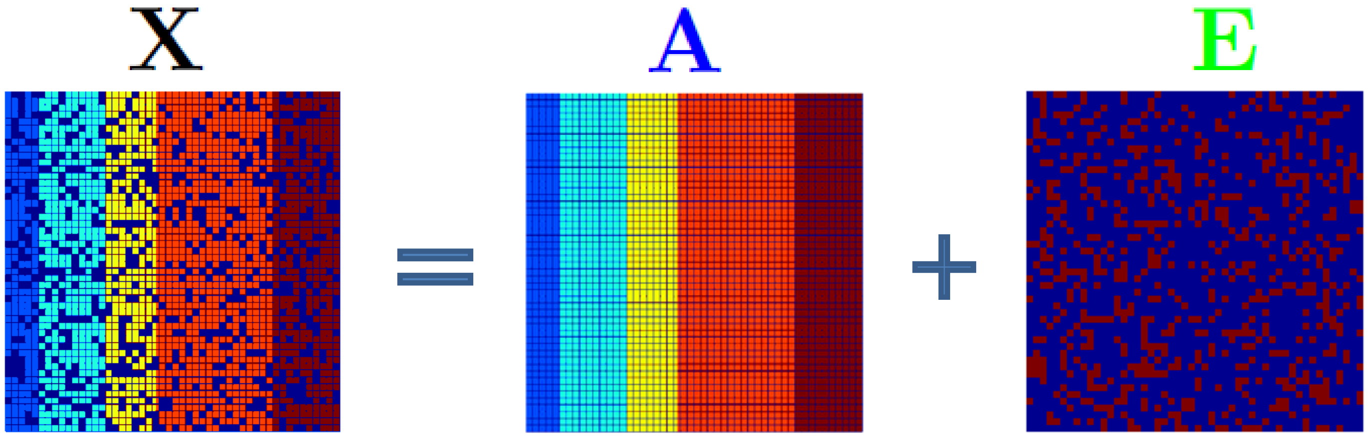
$$\Omega \subset [m] \times [n] \ \text{ is the set of observed entries.}$$

- or the data can be drawn from **a union of independent subspaces**:



$$\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

Robust PCA (?)



Problem: Given $\mathbf{X} = \mathbf{A} + \mathbf{E}$ recover $\mathbf{A}$ and $\mathbf{E}$.

- Various approaches in the literature:
  - Multivariate trimming [Gnanadesikan and Kettering '72]
  - Power Factorization [Wieber'70s]
  - Random sampling [Fischler and Bolles '81]
  - Alternating minimization [Shum & Ikeuchi'96, Ke and Kanade '03]
  - Influence functions [de la Torre and Black '03]

- Key question: **can guarantee correctness with an efficient algorithm?**

Robust PCA

$$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda \left\|\mathbf{E}\right\|_0 \ \text{ s.t. } \ \mathbf{X} = \mathbf{A} + \mathbf{E}$$

- Seek the lowest rank matrix that agrees with the data up to some sparse error.

Robust PCA

$$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \;\; \text{s.t.} \;\; \mathbf{X} = \mathbf{A} + \mathbf{E}$$

- Seek the lowest rank matrix that agrees with the data up to some sparse error.
- NP-hard!

Robust PCA

$$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{A} + \mathbf{E}$$

- Seek the lowest rank matrix that agrees with the data up to some sparse error.
- NP-hard!
- Convex relaxation:

$$\operatorname{rank}(\mathbf{A}) = \#\{\sigma_i(\mathbf{A}) \neq 0\} \rightarrow \|\mathbf{A}\|_* = \textstyle\sum_i \sigma_i(\mathbf{A})$$

$$\|\mathbf{E}\|_0 = \#\{e_{ij} \neq 0\} \rightarrow \|\mathbf{E}\|_1 = \textstyle\sum_{i,j} |e_{ij}|$$

Robust PCA

$$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \;\; \text{s.t.} \;\; \mathbf{X} = \mathbf{A} + \mathbf{E}$$

- Seek the lowest rank matrix that agrees with the data up to some sparse error.
- NP-hard!
- Convex relaxation:

$$\operatorname{rank}(\mathbf{A}) = \#\{\sigma_i(\mathbf{A}) \neq 0\} \rightarrow \|\mathbf{A}\|_* = \textstyle\sum_i \sigma_i(\mathbf{A})$$

$$\|\mathbf{E}\|_0 = \#\{e_{ij} \neq 0\} \rightarrow \|\mathbf{E}\|_1 = \textstyle\sum_{i,j} |e_{ij}|$$

$$\boxed{\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \;\; \text{s.t.} \;\; \mathbf{X} = \mathbf{A} + \mathbf{E}}$$

Robust PCA

$$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \ \text{ s.t. } \ \mathbf{X} = \mathbf{A} + \mathbf{E}$$

- Seek the lowest rank matrix that agrees with the data up to some sparse error.
- NP-hard!
- Convex relaxation:

$$\operatorname{rank}(\mathbf{A}) = \#\{\sigma_i(\mathbf{A}) \neq 0\} \rightarrow \|\mathbf{A}\|_* = \textstyle\sum_i \sigma_i(\mathbf{A})$$

$$\|\mathbf{E}\|_0 = \#\{e_{ij} \neq 0\} \rightarrow \|\mathbf{E}\|_1 = \textstyle\sum_{i,j} |e_{ij}|$$

$$\boxed{\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \ \text{ s.t. } \ \mathbf{X} = \mathbf{A} + \mathbf{E}}$$

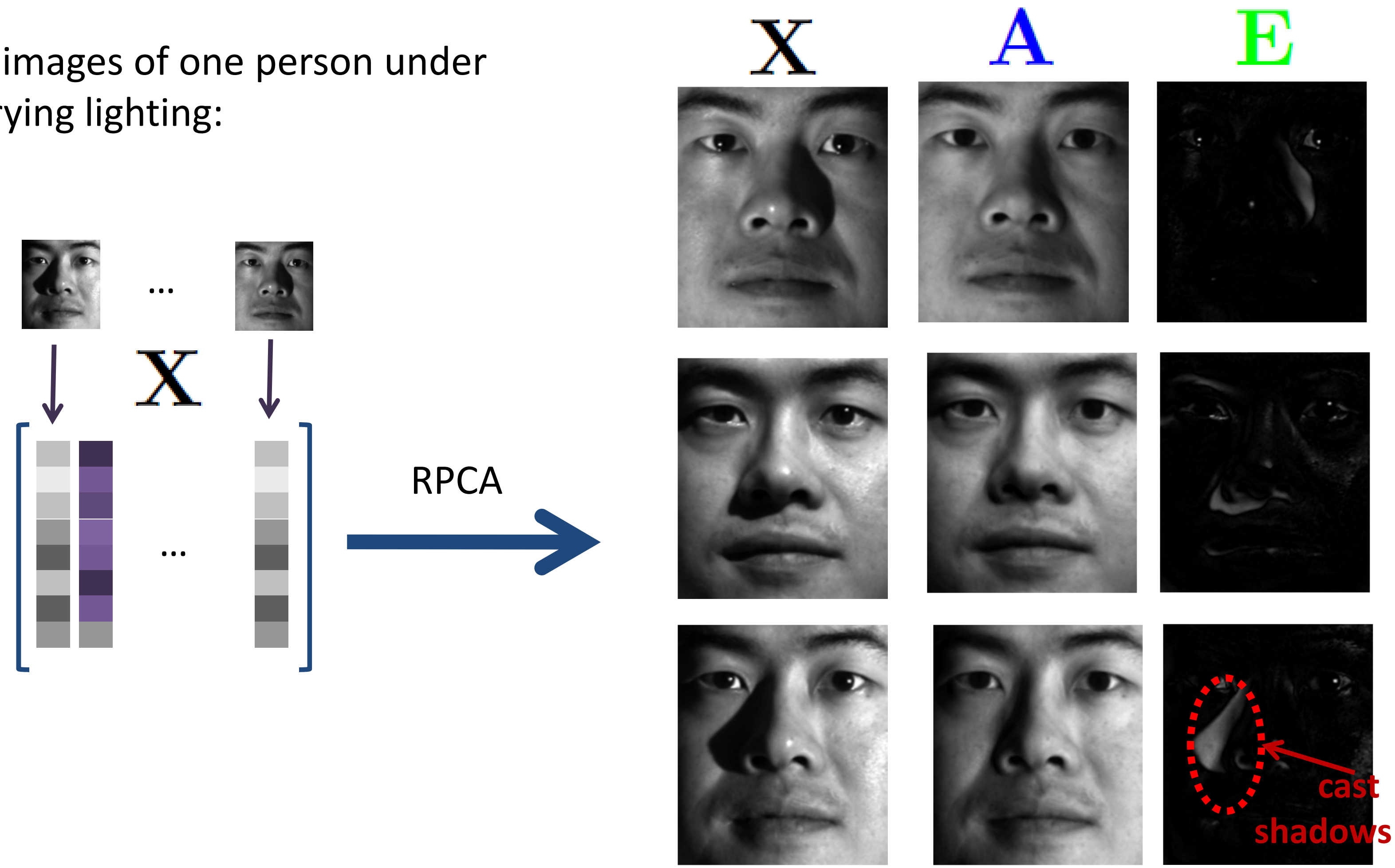
- Solvable in polynomial time!
- Convex optimization recovers almost any matrix of rank $\mathcal{O}(\frac{m}{\log^2 n})$ from errors corrupting $\mathcal{O}(mn)$ of the observations!

[Candes, Li, Ma, and Wright, Journal of the ACM, May 2011.]

Applications of RPCA

Repairing multiple correlated images

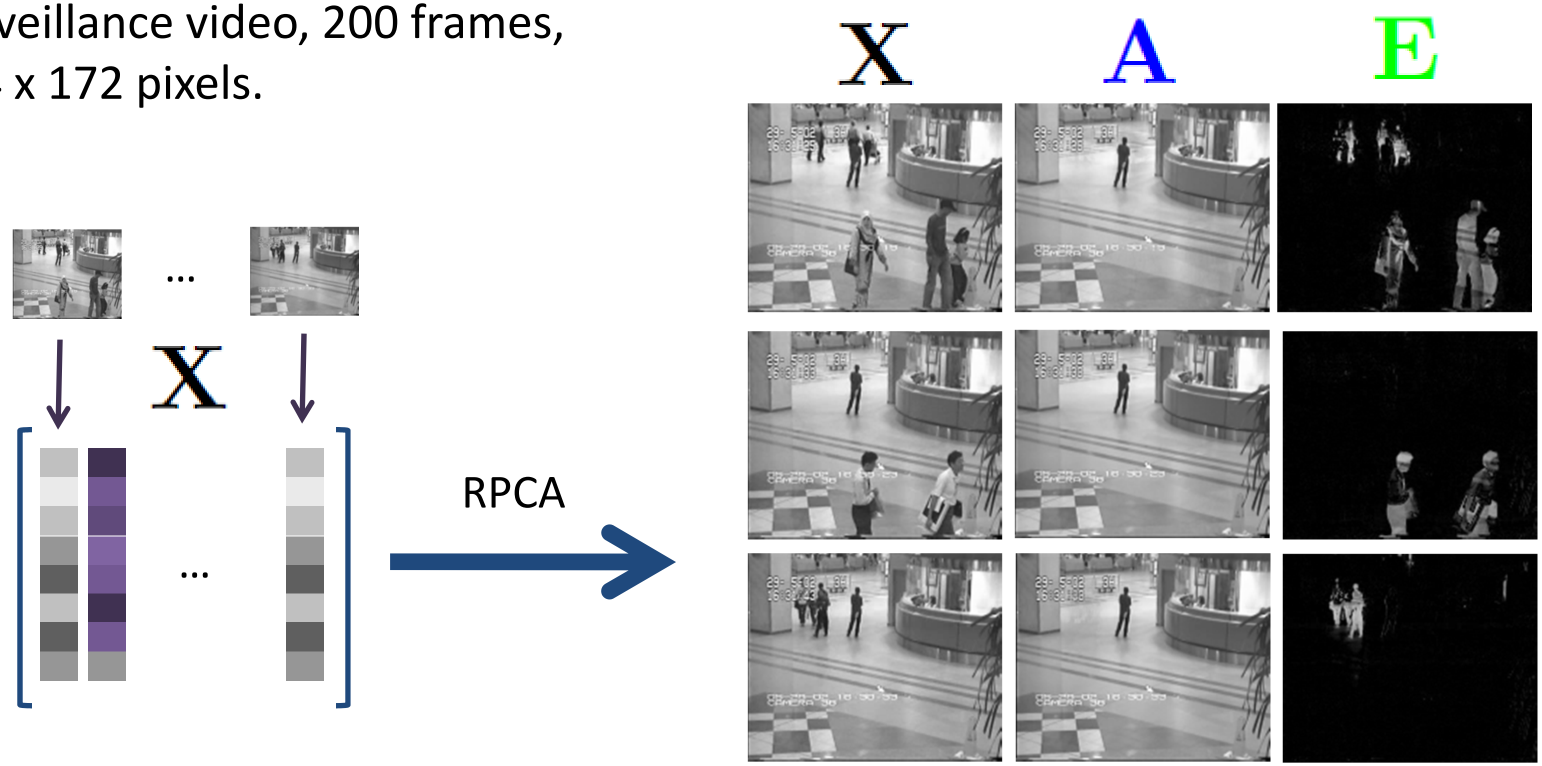
- 58 images of one person under varying lighting:



[Candes, Li, Ma, and Wright, Journal of the ACM, May 2011.]

Background modelling from video

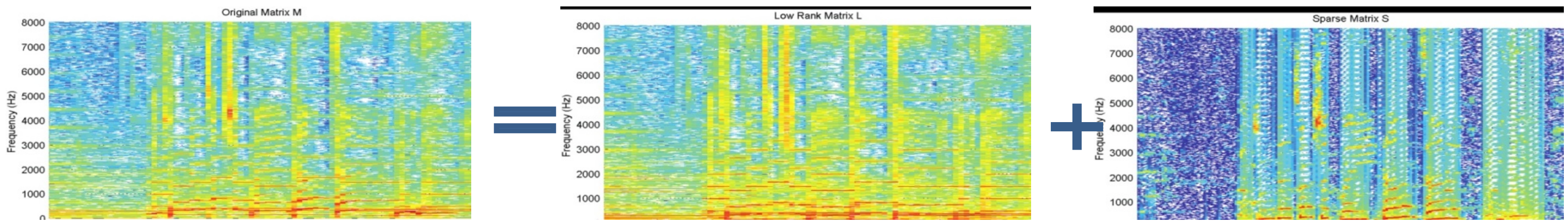
- Surveillance video, 200 frames, 144 x 172 pixels.

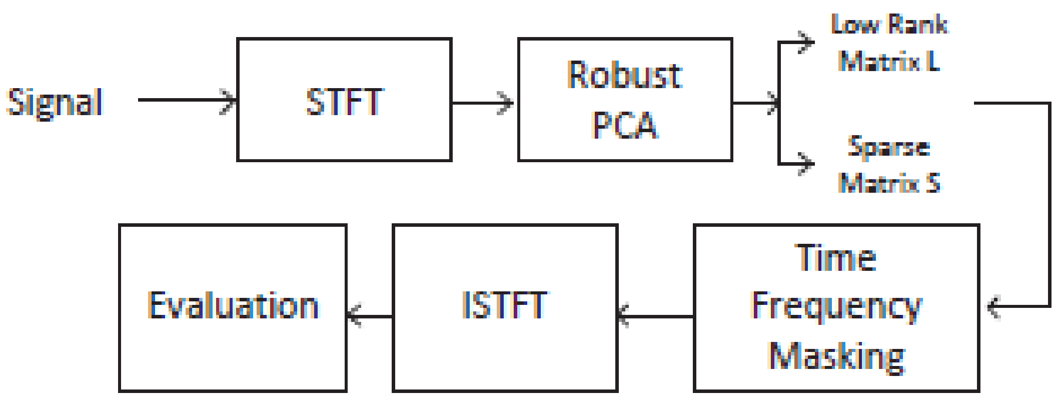


[Candes, Li, Ma, and Wright, Journal of the ACM, May 2011.]

Singing voice and music separation

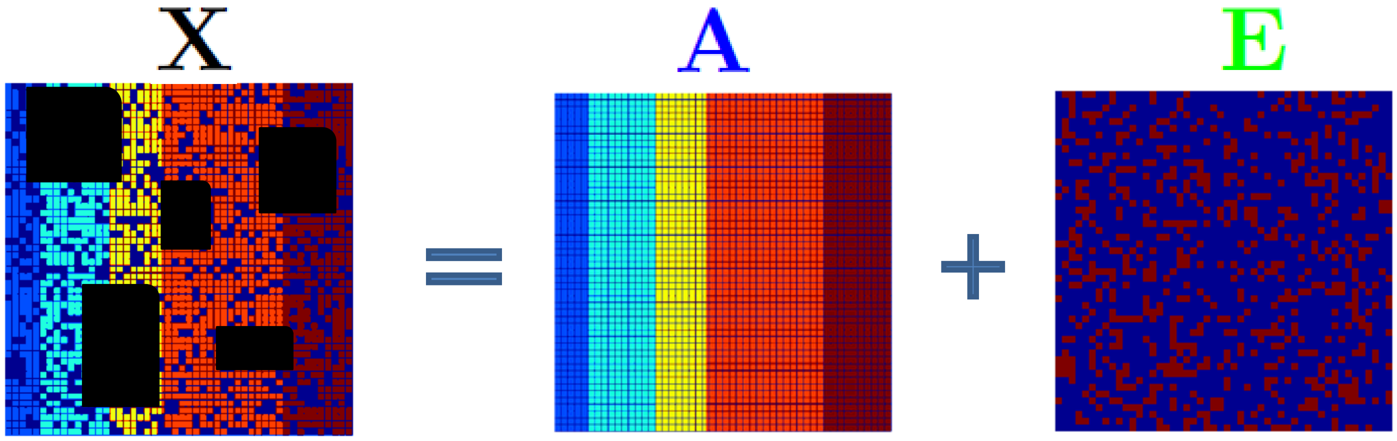
Song (STFT) Low-rank (music) Sparse (voices)





[Huang, Chen, Smaragdis, Hasegawa-Johnson, ICASSP 2012.]

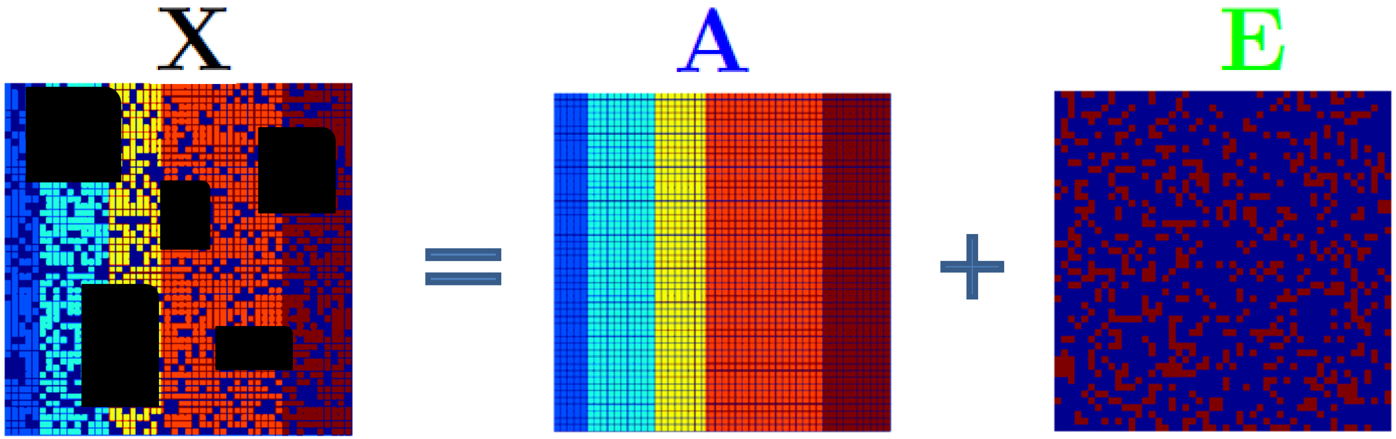
Robust Matrix Completion



Problem: Given $\mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{A} + \mathbf{E})$ recover $\mathbf{A}$ and $\mathbf{E}$.

- Recover low-dimensional structures from a fraction of missing measurements with structured support.

Robust Matrix Completion



Problem: Given $\mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{A} + \mathbf{E})$ recover $\mathbf{A}$ and $\mathbf{E}$.

- Recover low-dimensional structures from a fraction of missing measurements with structured support.
- Convex formulation:

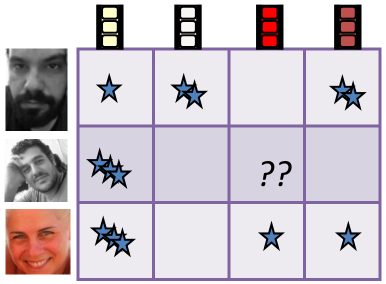
$$\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{A} + \mathbf{E})$$

[Candès, Recht. Foundations of Computational Mathematics, 2009.]

Applications of RMC

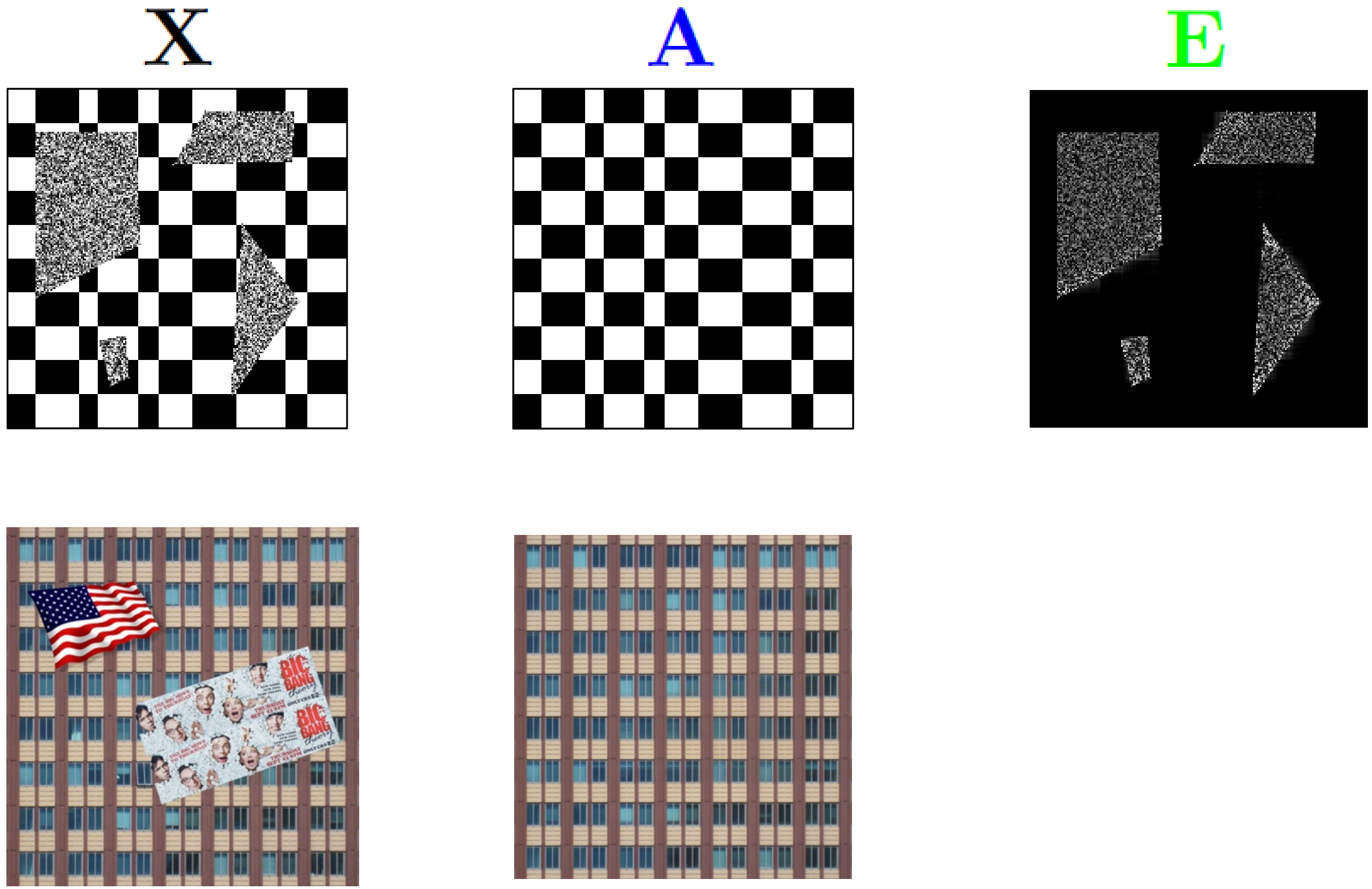
Recommender systems

- Users (rows of the data matrix) are given the opportunity to rate movies (columns of the data matrix) but users typically rate only very few movies so that there are very few scattered observed entries of this data matrix.
- One would like to complete this matrix so that the vendor (e.g., Netflix) might recommend titles that any particular user is likely to be willing to order.
- Assumption: the data matrix of all user-ratings may be approximately low-rank because it is commonly believed that only a few factors contribute to an individual's tastes or preferences.



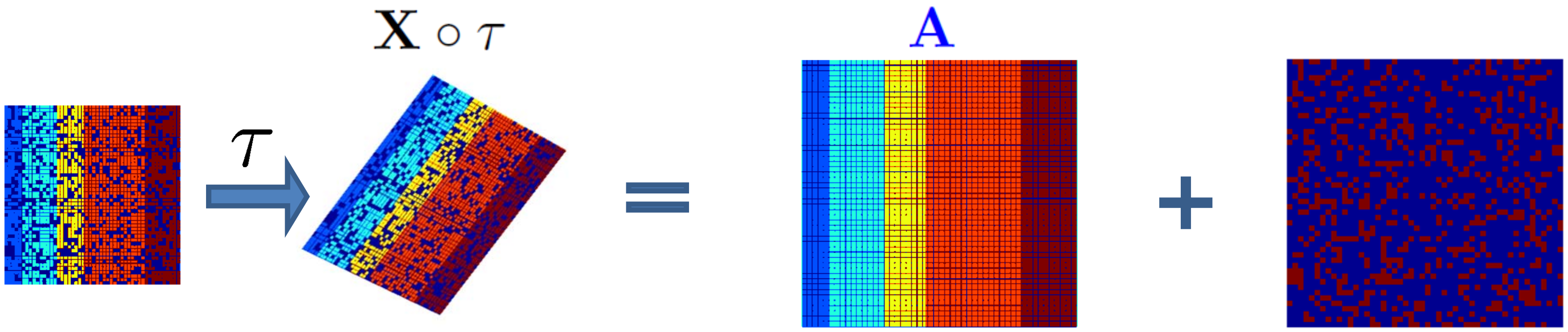
[Candès, Recht. Foundations of Computational Mathematics, 2009.]

Repairing of Low-rank Textures



[Liang, Ren, Zhang, Ma, ECCV 2012.]

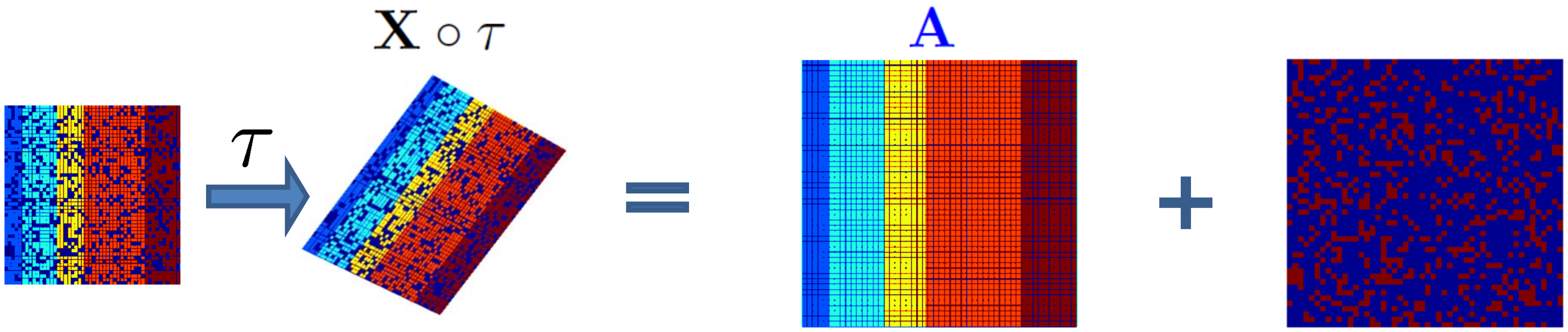
Robust subspace recovery under deformations



Problem: Given $\mathbf{X} \circ \tau = \mathbf{A} + \mathbf{E}$ recover, τ, $\mathbf{A}$ and $\mathbf{E}$.

- Recover low-dimensional structures from deformed measurements.

Robust subspace recovery under deformations



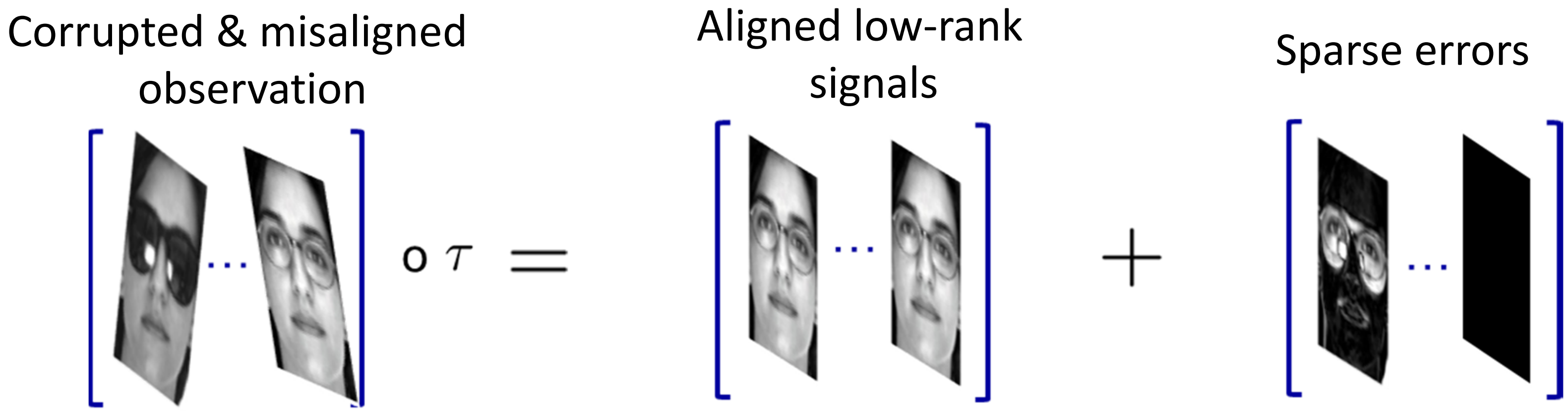
Problem: Given $\mathbf{X} \circ \tau = \mathbf{A} + \mathbf{E}$ recover, τ, $\mathbf{A}$ and $\mathbf{E}$.

- Recover low-dimensional structures from deformed measurements.
- Optimization problem:

$$\min_{\mathbf{A},\mathbf{E},\tau} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \ \text{ s.t. } \ \mathbf{X} \circ \tau = \mathbf{A} + \mathbf{E}$$

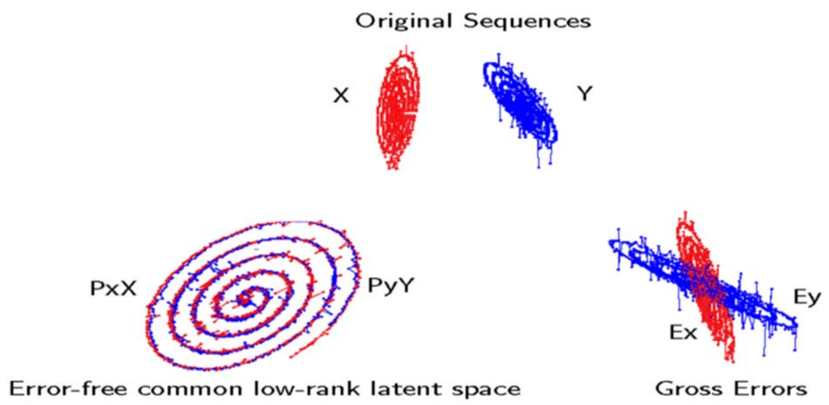
[Peng, Ganesh, Wright, Ma, CVPR 2010, TPAMI 2011.]

Robust alignment of multiple images



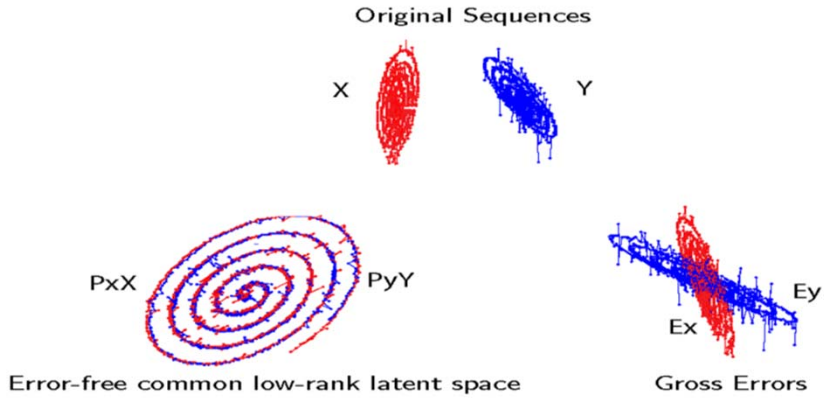
[Peng, Ganesh, Wright, Ma, CVPR 2010, TPAMI 2011.]

Robust temporal alignment



- **Problem**: Given two grossly corrupted temporally deformed sequences recover a low-rank subspace where the sequences are aligned in time.

Robust temporal alignment



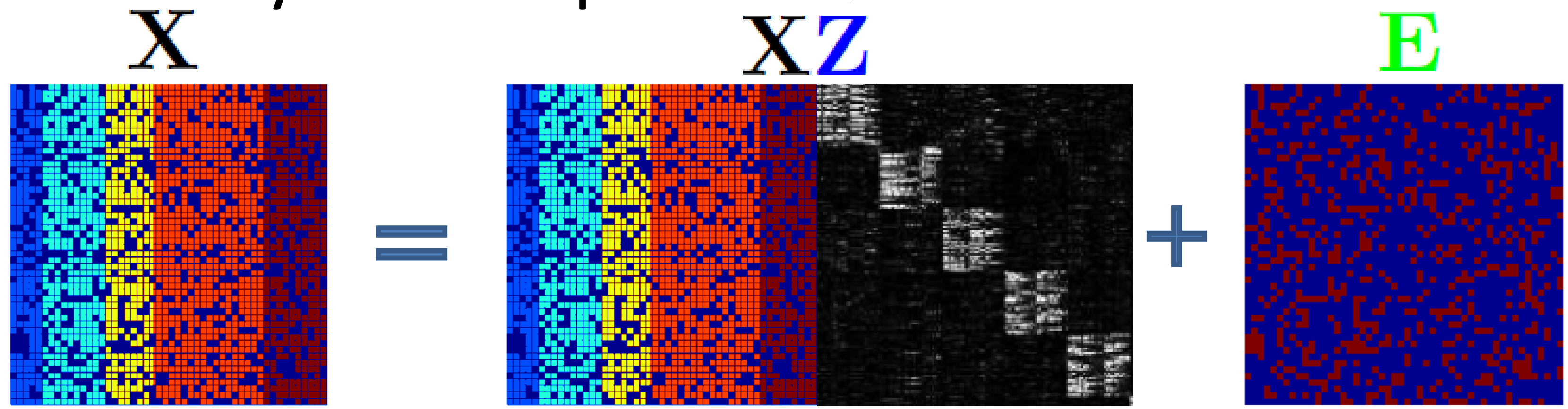
- **Problem**: Given two grossly corrupted temporally deformed sequences recover a low-rank subspace where the sequences are aligned in time.
- Optimization problem:

$$\text{argmin}_{\mathbf{P}_x,\mathbf{P}_y,\mathbf{E}_x,\mathbf{E}_y,\mathbf{\Delta}_x,\mathbf{\Delta}_y} \|\mathbf{P}_x\|_* + \|\mathbf{P}_y\|_* + \lambda_x \|\mathbf{E}_x\|_1 + \lambda_y \|\mathbf{E}_y\|_1 + \frac{\mu}{2}\|\mathbf{P}_x\mathbf{X}\mathbf{\Delta}_x - \mathbf{P}_y\mathbf{Y}\mathbf{\Delta}_y\|_F^2$$

$$\mathbf{X} = \mathbf{P}_x\mathbf{X} + \mathbf{E}_x, \mathbf{Y} = \mathbf{P}_y\mathbf{Y} + \mathbf{E}_y, \ \mathbf{\Delta}_x \in \{0,1\}^{T_x \times T}, \mathbf{\Delta}_y \in \{0,1\}^{T_y \times T}.$$

[Panagakis, Nicolaou, Zafeiriou, Pantic, CVPR 2013.]

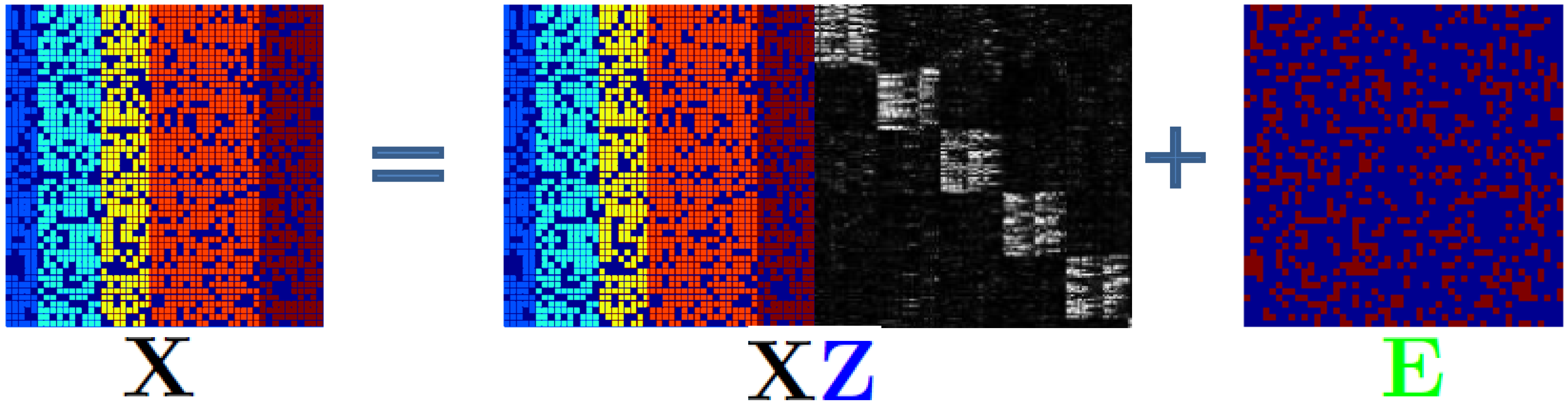
Recovery of multiple subspaces



Problem: Given $\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$ recover $\mathbf{Z}$ and $\mathbf{E}$.

- $\mathbf{Z}$ is a low-rank block diagonal matrix which reveals the subspace membership of each column of the data matrix.

Recovery of multiple subspaces via low-rank representation (LRR)



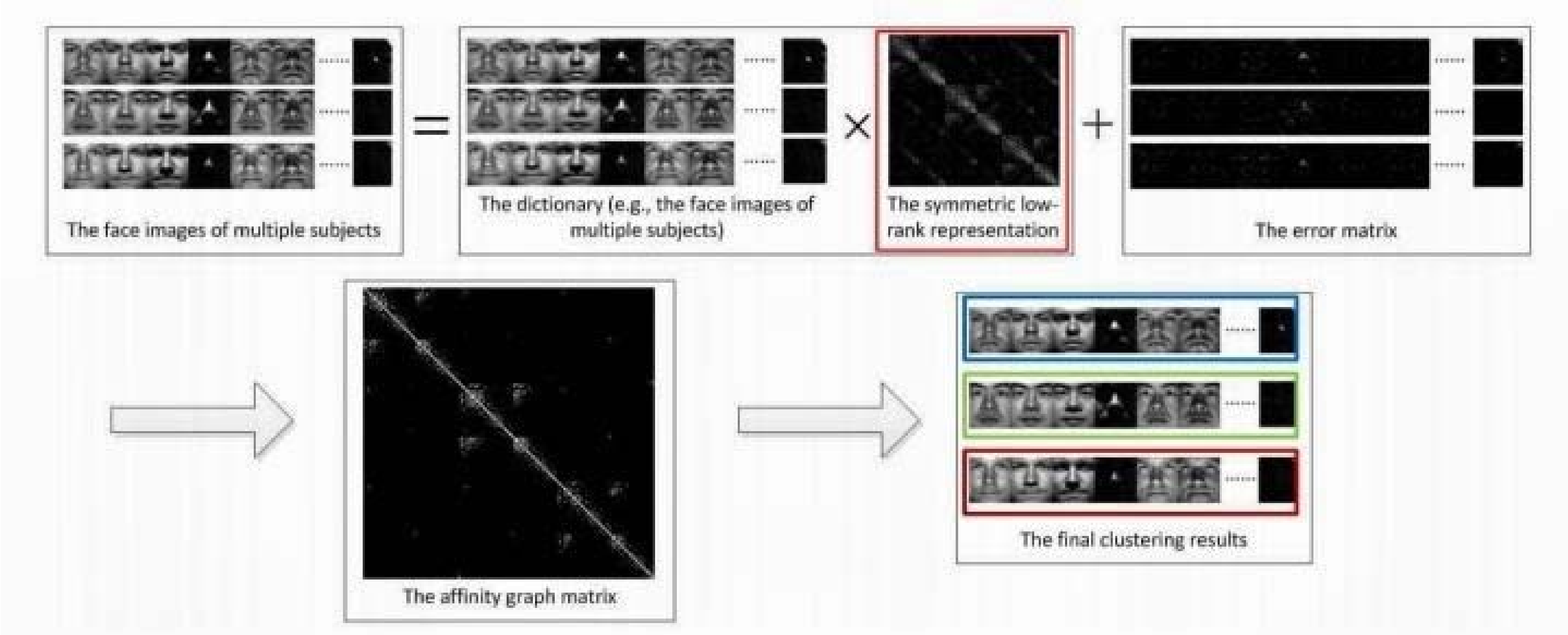
Problem: Given $\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$ recover $\mathbf{Z}$ and $\mathbf{E}$.

- $\mathbf{Z}$ is a low-rank block diagonal matrix which reveals the subspace membership of each column of the data matrix.
- Convex formulation:

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

[Liu, Lin, Yan, Sun, Yu, Ma, TPAMI 2013.]

Face Clustering via LRR



[Liu, Lin, Yan, Sun, Yu, Ma, TPAMI 2013.]

Algorithms?

- ADMM, ISTA, FISTA, FASTA, Semidefinite Programming etc.
- Easy to implement (50 - 100 lines of Matlab code)

Algorithms?

- ADMM, ISTA, FISTA, FASTA, Semidefinite Programming etc.
- Easy to implement (50 - 100 lines of Matlab code)
- We will NOT discuss about algorithms today.

Thank you!