Sparse and Low-Rank Representation for Biometrics – Lecture II: Low-Rank Representation and Applications

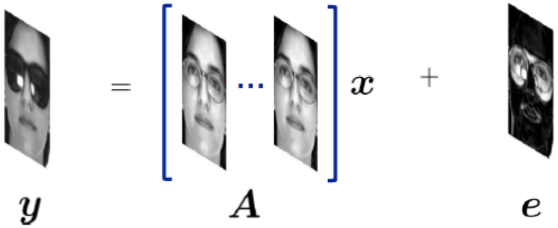
Allen Y. Yang
Department of EECS, UC Berkeley
yang@eecs.berkeley.edu

with John Wright and Yi Ma

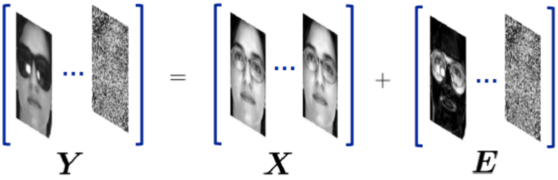
ICB 2013 Tutorial

From Sparsity to Low Rank

- Previously: Sparse representation-based classification



- This lecture: Recovering low-rank matrices (many correlated vectors)



Formulation: Robust PCA

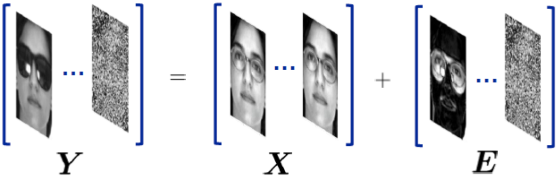


Figure : Given $Y = X + E$ with X low rank and E sparse, recover X and E.

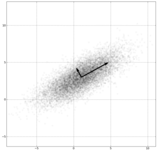
Existing approaches to **Robust PCA** in the literature:

- Multivariate trimming [Gnanadeskian & Kettering '72]
- Random sampling [Fischler & Bolles '81]
- Alternating minimization [Ke & Kanade '03]
- Influence functions [de la Torre & Black '03]

Can we find an efficient and provably correct algorithm?

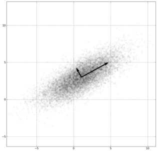
Related Solutions: Matrix Recovery

- **Classical singular-value decomposition (SVD)** [Hotelling '35, Karhunen & Loeve '72]
  Given $Y = X + Z$, where Z represents Gaussian noise, recover X
  SVD is a stable, efficient algorithm. Theoretically optimal → huge impact in practice.



Related Solutions: Matrix Recovery

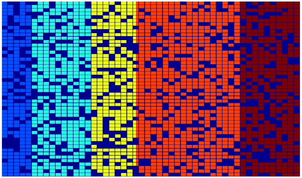
- **Classical singular-value decomposition (SVD)** [Hotelling '35, Karhunen & Loeve '72]
  Given $Y = X + Z$, where Z represents Gaussian noise, recover X
  SVD is a stable, efficient algorithm. Theoretically optimal $\rightarrow$ huge impact in practice.



- **Matrix completion:** low rank with missing data [Candès & Recht '08, Candès & Tao '09, Keshevan et al. '09, Gross '09, Ravikumar & Wainwright '10]

$$\text{From } Y = \mathcal{P}_\Omega[X], \text{ recover } X.$$

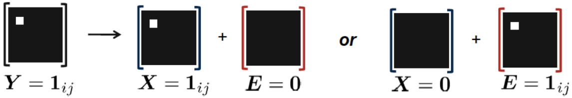
  The problem is solvable if X is low rank and the support Ω is large enough.



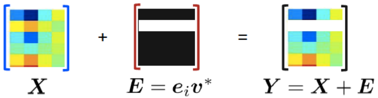
Robust PCA is a hard problem

(RPCA): $Y = X + E$, whereby unknowns X is low-rank and E is sparse.

- Sparse matrices can be also low-rank:



- Certain sparse error patterns E make exactly recovering X impossible:



Exclude these ambiguities from the possible solutions.

Incoherence Conditions

Theorem [Candès & Recht '08]

X can be recovered if it is **incoherent** with the standard basis on which E is sparse.

- **On X:** Incoherence condition on singular vectors
  1. Singular vectors of X not too spiky: $\begin{cases} \max_i \|U_i\|^2 \leq \mu r/m \\ \max_i \|V_i\|^2 \leq \mu r/n \end{cases}$.
  2. Not too cross-correlated: $\|UV^*\|_\infty \leq \sqrt{\mu r/mn}$.

Incoherence Conditions

Theorem [Candès & Recht '08]

X can be recovered if it is **incoherent** with the standard basis on which E is sparse.

- **On X:** Incoherence condition on singular vectors
  1. Singular vectors of X not too spiky: $\begin{cases} \max_i \|U_i\|^2 \leq \mu r/m \\ \max_i \|V_i\|^2 \leq \mu r/n \end{cases}$.
  2. Not too cross-correlated: $\|UV^*\|_\infty \leq \sqrt{\mu r/mn}$.
- **On E:** Uniform model on error support, but signs and magnitudes are arbitrary:

$$\mathrm{supp}(E) \sim \mathrm{uni}([m] \times [n], \rho).$$

Convex Optimization

- Exact solution is nonconvex and NP-hard

$$\min \operatorname{rank}(X) + \gamma \|E\|_0 \quad \text{subj. to} \quad Y = X + E.$$

Neither $\operatorname{rank}(X)$ nor $\|E\|_0$ is a smooth convex function.

Convex Optimization

- Exact solution is nonconvex and NP-hard

$$\min \operatorname{rank}(X) + \gamma \|E\|_0 \quad \text{subj. to} \quad Y = X + E.$$

  Neither $\operatorname{rank}(X)$ nor $\|E\|_0$ is a smooth convex function.

- **Convex relaxation:** [Fazel et al. '01, Recht et al. '08]
  1. Rank: $\operatorname{rank}(X)$ equivalent to ℓ_0-norm of its singular values.
     Nuclear norm $\|X\|_* \doteq \sum_i \sigma_i(X)$ equivalent to ℓ_1-norm of its singular values.

$$\operatorname{rank}(X) \Rightarrow \|X\|_*$$

  2. Sparse error:

$$\|E\|_0 \Rightarrow \|E\|_1 = \sum_{ij} |E_{ij}|.$$

Main Result

Theorem (Principal Component Pursuit) [Candès et al. '09]

If $X_0 \in \mathbb{R}^{m\times n}$, assuming $m \geq n$, has rank

$$r \leq \rho_r \frac{n}{\mu \log^2(m)}$$

and E_0 has Bernoulli support with error probability $\rho \leq \rho_s$, then with very high probability

$$(X_0, E_0) = \arg\min \|X\|_* + \frac{1}{\sqrt{m}}\|E\|_1 \quad \text{subj. to} \quad X + E = X_0 + E_0,$$

and the minimizer is unique.

Convex optimization recovers matrices of rank $O\left(\frac{n}{\log^2 m}\right)$ from errors corrupting $O(mn)$ entries.

Big Picture: Parallelism of Sparsity and Low-Rank

	Sparse Vector	*Low-Rank Matrix*
Degeneracy of	one signal	correlated signals
Measure	ℓ^0 norm $\|x\|_0$	$\text{rank}(X)$
Convex Surrogate	ℓ^1 norm $\|x\|_1$	Nuclear norm $\|X\|_*$
Compressed Sensing	$y = Ax$	$Y = A(X)$
Error Correction	$y = Ax + e$	$Y = A(X) + E$
Domain Transform	$y \circ \tau = Ax + e$	$Y \circ \tau = A(X) + E$

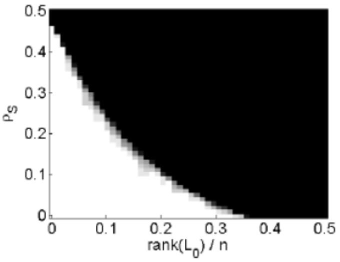
Two Important Variations

- **Matrix completion:** $Y = \mathcal{P}_\Omega[A_0 + E_0]$
  With conditions similar to RPCA, but the observation Y is only a random subset of entries of size
  $$|\Omega| = mn/10.$$
  Then with very high probability, solving the convex program
  $$\min \|X\|_* + \frac{1}{\sqrt{m}}\|E\|_1 \quad \text{subj. to} \quad \mathcal{P}_\Omega[X + E] = Y,$$
  uniquely recovers (X_0, E_0).

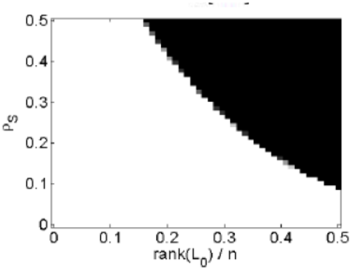
Two Important Variations

- **Matrix completion:** $Y = \mathcal{P}_\Omega[A_0 + E_0]$
  With conditions similar to RPCA, but the observation Y is only a random subset of entries of size
  $$|\Omega| = mn/10.$$
  Then with very high probability, solving the convex program
  $$\min \|X\|_* + \frac{1}{\sqrt{m}}\|E\|_1 \quad \text{subj. to} \quad \mathcal{P}_\Omega[X + E] = Y,$$
  uniquely recovers (X_0, E_0).
- **RPCA with Noise:** $Y = A_0 + E_0 + Z$
  With conditions similar to RPCA, but assuming $\|Z\|_F \leq \eta$. Then with very high probability, solving the convex program
  $$\min \|X\|_* + \frac{1}{\sqrt{m}}\|E\|_1 \quad \text{subj. to} \quad \|Y - X - E\|_F \leq \eta,$$
  satisfies the following bound for some constant $C > 0$:
  $$\|(X^*, E^*) - (X_0, E_0)\|_F \leq C\eta.$$

Simulations



Robust PCA, Random Signs



Matrix Completion

- White regions are instances with perfect recovery.
- Correct recovery when X is low-rank and E is sparse.

Other Useful Regularizers in Sparse and Low-Rank Representation

There are many types of low-dimensional structures:

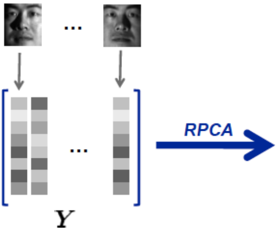
- [Zhou et al. '09]: Spatially contiguous sparse errors via MRF
- [Bach '10]: Structured relaxations from sub modular functions
- [Negahban et al. '10]: Geometric analysis of recovery
- [Becker et al. '10]: Algorithmic templates
- [Xu et al. '11]: Column sparse errors $L_{2,1}$ norm
- [Recht et al. '11]: Compressive sensing of various structures
- [Candès & Recht '11]: Compressive sensing of decomposable structures
- [McCoy & Tropp '11]: Decomposition of sparse and low-rank structures
- [Wright et al. '12]: Superposition of decomposable structures
- [Ohlsson et al. '13]: Quadratic basis pursuit

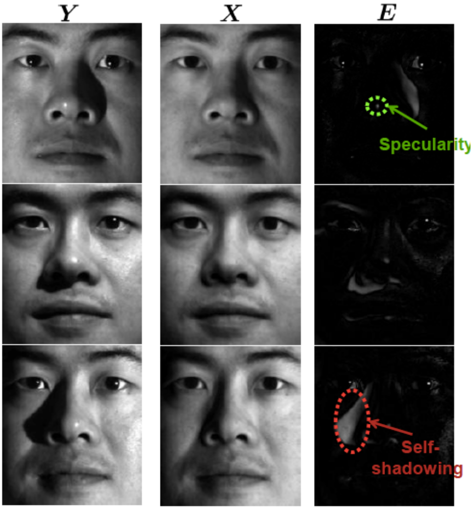
Take-Home Message

Let the data tell you the right structure: **geometry, statistics, learning algorithms, computation**.

Removing varying illumination from face images

58 images of one person under varying lighting:





Background subtraction from video

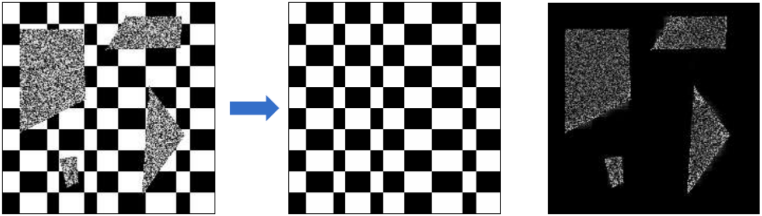
Static camera surveillance video

200 frames, 144 x 172 pixels,

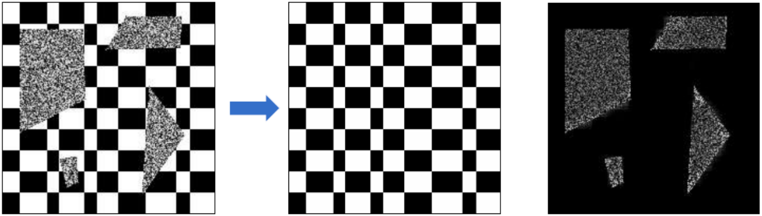
Significant foreground motion

Video Y = Low-rank appx. X + Sparse error E

Repairing low-rank textures



Repairing low-rank textures



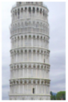
Minimize Image Rank Under Transformation

- Most symmetric image patterns (if treated as matrices) are low-rank

(e) Output ($r = 14$)

(f) Output ($r = 8$)

(g) Output ($r = 19$)



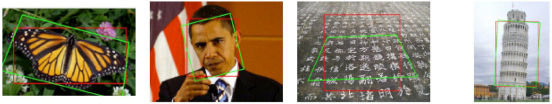
(h) Output ($r = 6$)

Minimize Image Rank Under Transformation

- Most symmetric image patterns (if treated as matrices) are low-rank

(e) Output ($r = 14$) (f) Output ($r = 8$) (g) Output ($r = 19$) (h) Output ($r = 6$)

- Camera projection and pose variation distort/destroy the low-rank representation



(a) Input ($r = 35$) (b) Input ($r = 15$) (c) Input ($r = 53$) (d) Input ($r = 13$)

Recover camera projection and pose

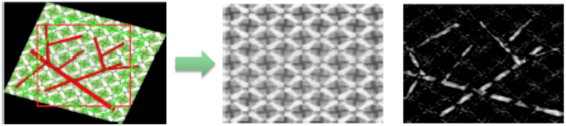
Minimizing the rank of texture images may recover the hidden information about the orientation of the pattens in 3-D space.

Transform Invariant Low-rank Texture (TILT)

- Objective function [Zhang et al. '10]

$$\min_{A,E,\tau} \|A\|_* + \lambda \|E\|_1 \quad \text{subj. to} \quad I \circ \tau = A + E,$$

where A is low-rank and E is sparse, τ parametrizes an image transformation.

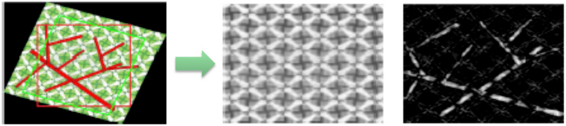


Transform Invariant Low-rank Texture (TILT)

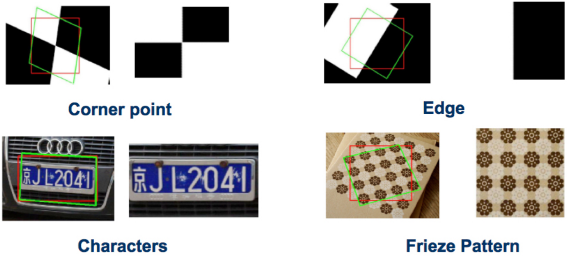
- Objective function [Zhang et al. '10]

$$\min_{A,E,\tau} \|A\|_* + \lambda\|E\|_1 \quad \text{subj. to} \quad I \circ \tau = A + E,$$

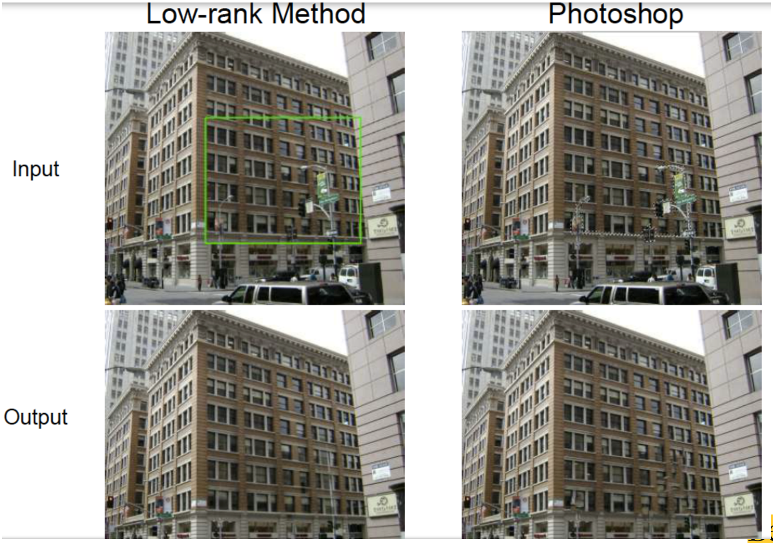
where A is low-rank and E is sparse, τ parametrizes an image transformation.



- More Examples



Repair Distorted Low-rank Textures



Robust Alignment via Low-rank and Sparse Decomposition (RASL)

D – corrupted & misaligned observation

A – aligned low-rank signals

E – sparse errors

Problem: Given $D \circ \tau = A_0 + E_0$, recover τ, A_0 and E_0.

Parametric deformations (rigid, affine, projective…)

Low-rank component

Sparse component

Solution: Robust Alignment via Low-rank and Sparse (**RASL**) Decomposition

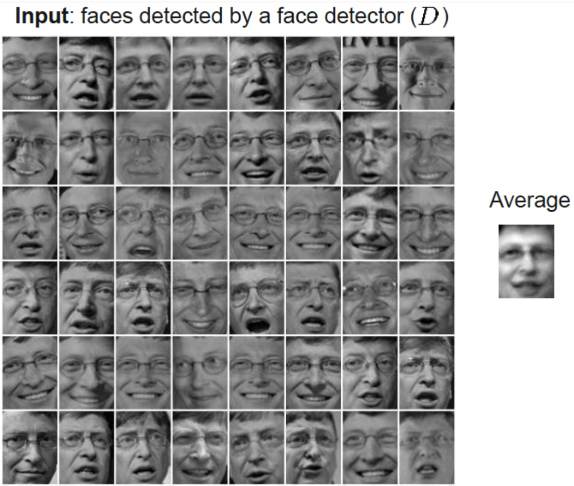
Iteratively solving the linearized convex program:

$$\min \|A\|_* + \lambda \|E\|_1 \quad \text{subj} \quad A + E = D \circ \tau_k + J\Delta\tau$$
$$(\text{or} \quad Q(A+E) = QD \circ \tau_k, \; QJ = 0)$$

Reference: Peng, Ganesh, Wright, Ma, '10.

RASL Example: Face Alignment

Detected Faces

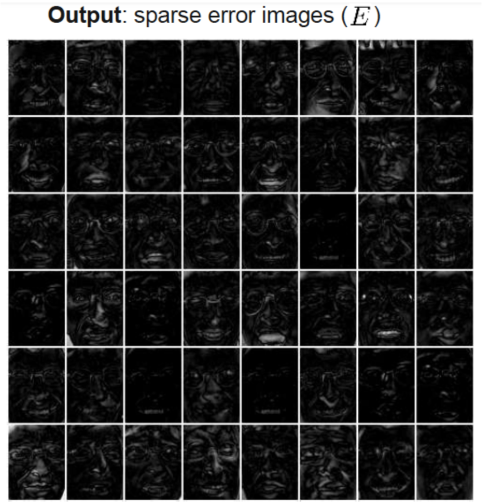


After RASL Alignment

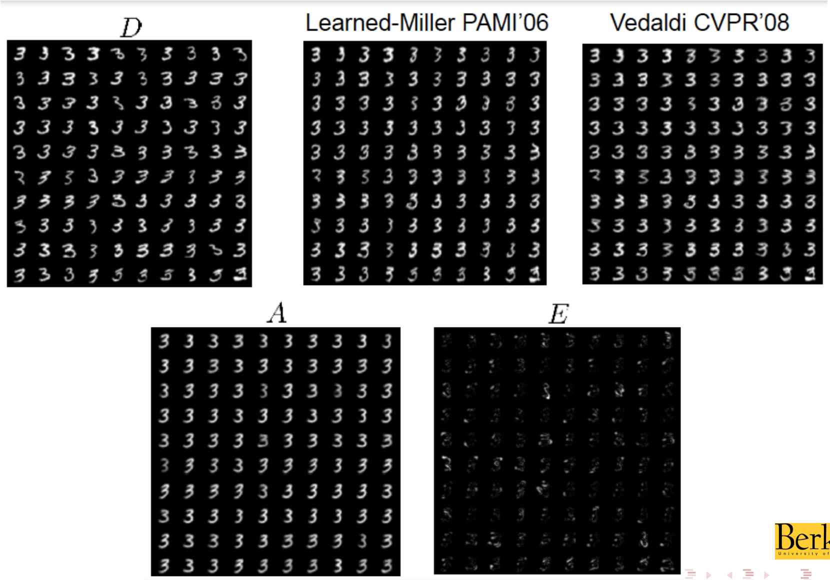
Output: aligned faces ($D \circ \tau$)

Average

Sparse error of the face image

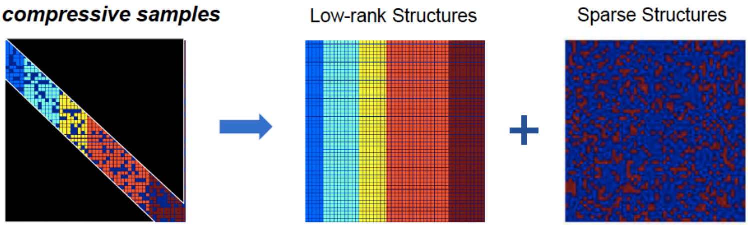


Aligning hand-written digits using RASL



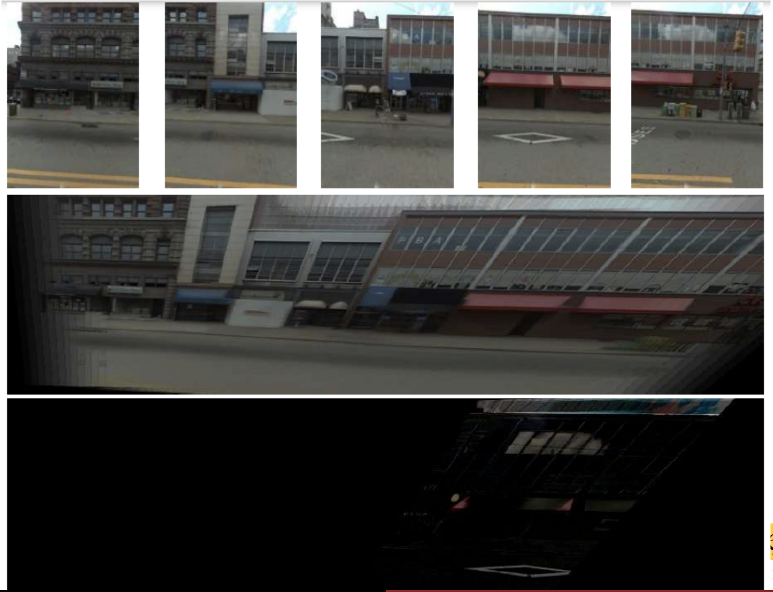
Repair Street Panorama

Recover low-dimensional structures from diminishing fraction of corrupted measurements.

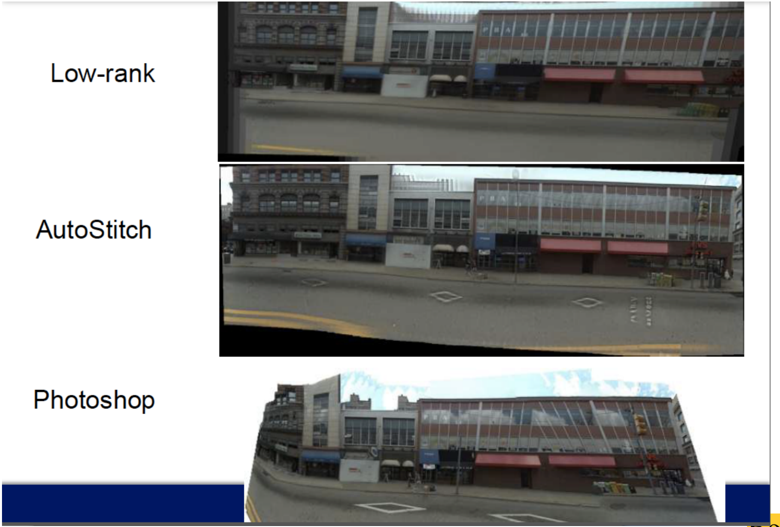


Reference: Zhou, Min, & Ma '12.

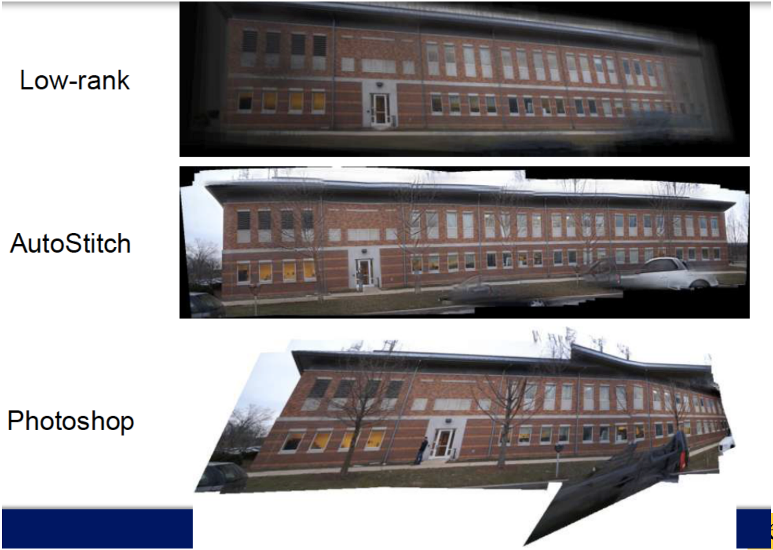
Example



Compare to AutoStitch and PhotoShop



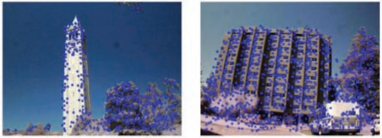
Compare to AutoStitch and PhotoShop



Informative feature selection via Sparse PCA

Goal

- Select strong/informative object features for low-end mobile cameras, surveillance apps.



- Geometric methods are often ill-posed

SfM plus *camera adjacency graph*

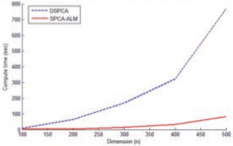
- **Challenges:** 1. objects lack invariant features; 2. excessive outliers; 3. pairwise matching is costly.

Approach

- A batch statistical solution: Sparse PCA

$$\max_{\|x\|_2 \le 1} x^T \Sigma x - \rho \|x\|_0$$

- An efficient convex algorithm as a SDP problem.



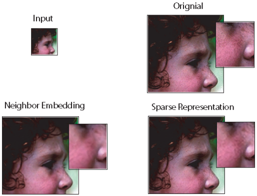
- Outperforms in both accuracy and speed

Reference: Naikal, AY, Sastry. *Informative feature selection for object recognition via Sparse PCA*. ICCV 2011.

More Applications in Sparse and Low-Rank Representation:

- **Target Tracking** [Mei & Ling 2009, Liu et al. 2010, Li et al. 2011]

- **Superresolution** [Yang et al. 2009]



- **Sparse dictionary learning** [Aharon et al. 2006, Mairal et al. 2008, Duarte-Carvajalino & Sapiro 2009]

Take-Home Messages

1. (Transformed) **low-rank and sparse** structures are central to visual data modeling, processing, and analyzing.
2. Such structures can now be extracted **robustly and efficiently** from raw image pixels.
3. Low-rank and sparse representation capable of **capturing local or global information** from high-resolution images, surpassing human perception.
4. The new algorithms have exhibited tremendous impact to **board applications** in image processing, pattern recognition, and biometrics.