Lecture II:
Data Modeling and Applications

Yi Ma

Visual Computing Group
Microsoft Research Asia, Beijing

John Wright

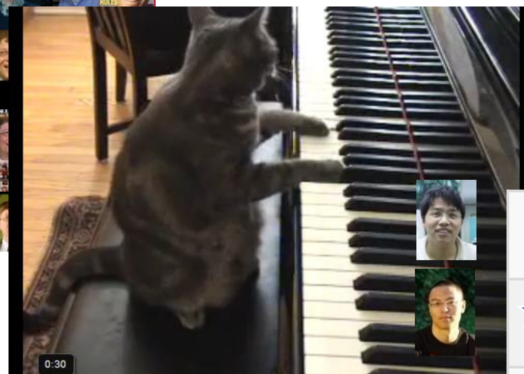
Electrical Engineering
Columbia University

CONTEXT – *Data increasingly massive, high-dimensional…*

Images

⇓➢ ***1M pixels***

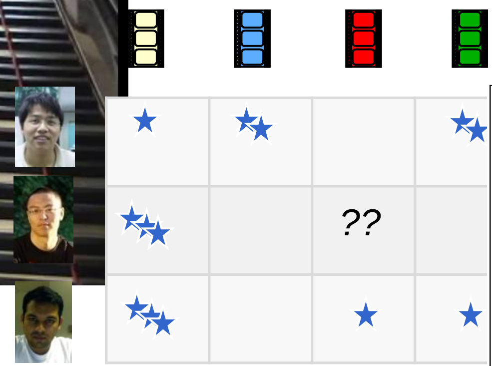
Compression
De-noising
Super-resolution
Recognition…



Videos

⇓➢ ***1B voxels***

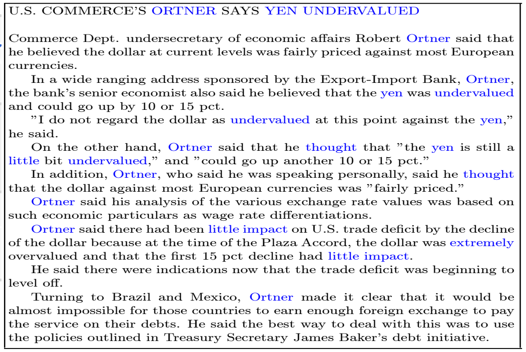
Streaming
Tracking
Stabilization…



User data

⇓➢ ***1B users***

Clustering
Classification
Collaborative filtering…



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 he thought that the dollar against most European currencies was "fairly priced."

Ortner said his analysis of the various exchange rate values was based on such economic particulars as wage rate differentiations.

Ortner said there had been little impact on U.S. trade deficit by the decline of the dollar because at the time of the Plaza Accord, the dollar was extremely overvalued and that the first 15 pct decline had little impact.

He said there were indications now that the trade deficit was beginning to level off.

Turning to Brazil and Mexico, Ortner made it clear that it would be almost impossible for those countries to earn enough foreign exchange to pay the service on their debts. He said the best way to deal with this was to use the policies outlined in Treasury Secretary James Baker's debt initiative.

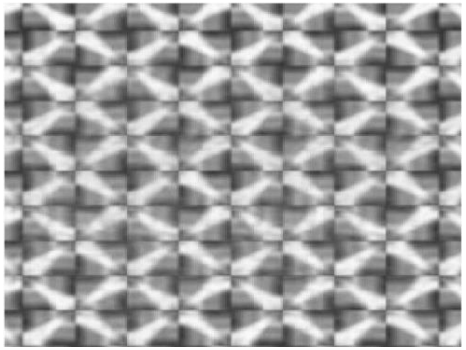
Web data

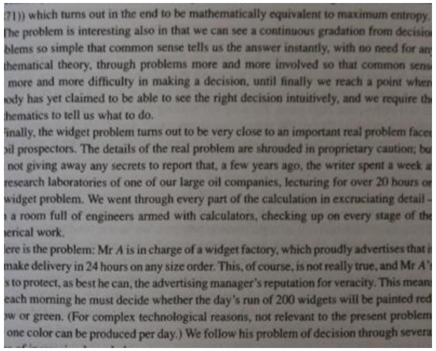
⇓➢ ***100B webpages***

Indexing
Ranking
Search…

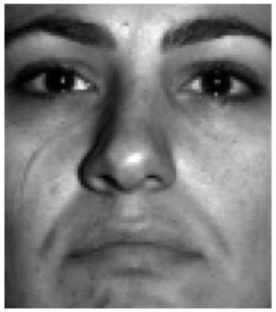
How to extract low-dim structures from such high-dim data?

CONTEXT – *Low dimensional structures in visual data*

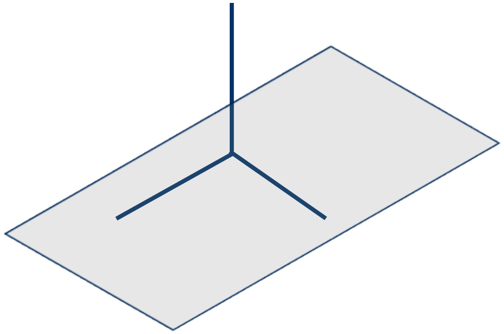




71)) which turns out in the end to be mathematically equivalent to maximum entropy.
he problem is interesting also in that we can see a continuous gradation from decisio
blems so simple that common sense tells us the answer instantly, with no need for an
thematical theory, through problems more and more involved so that common sens
more and more difficulty in making a decision, until finally we reach a point wher
ody has yet claimed to be able to see the right decision intuitively, and we require th
hematics to tell us what to do.
inally, the widget problem turns out to be very close to an important real problem face
il prospectors. The details of the real problem are shrouded in proprietary caution; bu
not giving away any secrets to report that, a few years ago, the writer spent a week a
research laboratories of one of our large oil companies, lecturing for over 20 hours o
widget problem. We went through every part of the calculation in excruciating detail –
a room full of engineers armed with calculators, checking up on every stage of the
erical work.
ere is the problem: Mr *A* is in charge of a widget factory, which proudly advertises that i
make delivery in 24 hours on any size order. This, of course, is not really true, and Mr *A*'
s to protect, as best he can, the advertising manager's reputation for veracity. This mean
each morning he must decide whether the day's run of 200 widgets will be painted red
w or green. (For complex technological reasons, not relevant to the present problem
one color can be produced per day.) We follow his problem of decision through severa

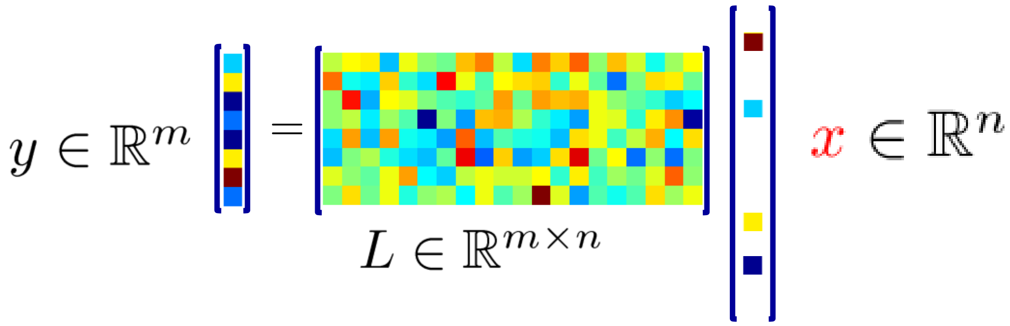


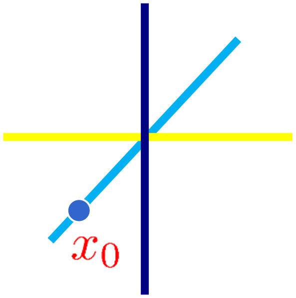
Visual data exhibit ***low-dimensional structures*** due to rich ***local*** regularities, ***global*** symmetries, ***repetitive*** patterns, or ***redundant*** sampling.



CONTEXT – *Recent related progress*

Sparse recovery: Given $y = Lx_0$, $L \in \mathbb{R}^{m \times n}$, $m \ll n$, recover x_0.





Impossible in general ($m \ll n$)

Well-posed if x_0 is structured *(sparse),* but still **NP-hard**

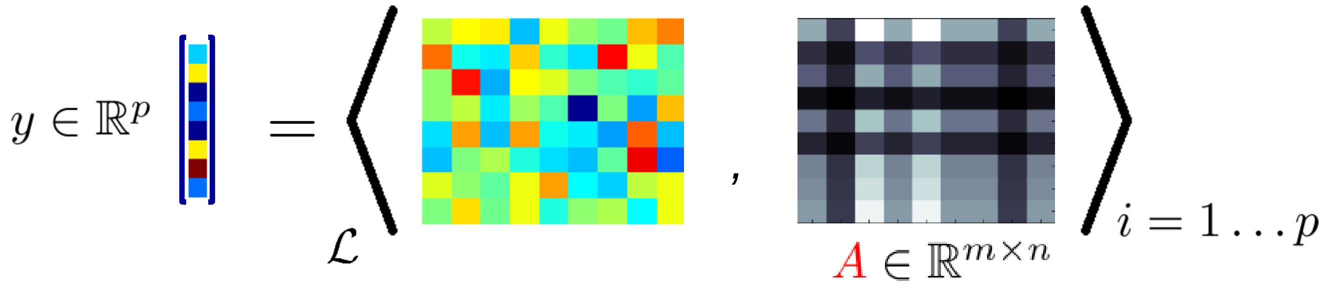
Tractable via convex optimization: $\min \|x\|_1 \text{ s.t. } y = Lx$

… if L is "nice" *(random, incoherent, RIP)*

Hugely active area: *Donoho+Huo '01, Elad+Bruckstein '03, Candès+Tao '04,'05, Tropp '04, '06, Donoho '04, Fuchs '05, Zhao+Yu '06, Meinshausen+Buhlmann '06, Wainwright '09, Donoho+Tanner '09 … and many others*

CONTEXT – *Recent related progress*

Low-rank sensing: Given $y = \mathcal{L}[A_0]$, $\mathcal{L} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$, recover A_0.



Impossible in general ($p \ll mn$)

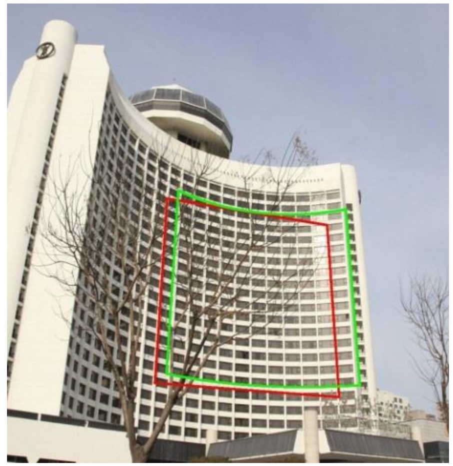
Well-posed if A_0 is structured *(low-rank),* but still **NP-hard**

Tractable via convex optimization: $\min \|A\|_* \text{ s.t. } y = \mathcal{L}(A)$

… if $\mathcal{L}$ is "nice" *(random, rank-RIP)*

Hugely active area: *Recht+Fazel+Parillo '07, Candès+Plan '10, Mohan+Fazel '10, Recht+Xu+Hassibi '11, Chandrasekaran+Recht+Parillo+Willsky '11, Negahban+Wainwright '11 …*

APPLICATIONS – *How to apply such models and tools?*

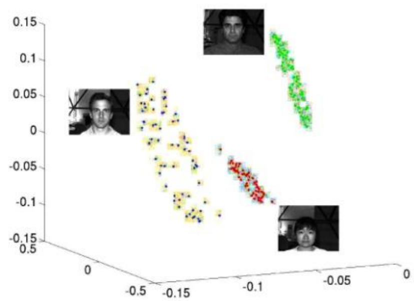


1. ***Justify models*** *of low-dimensional structures in practical data, such as face images, textures, objects etc…*

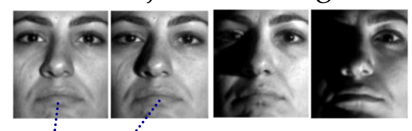
2. ***Customize models*** *to incorporate additional variations in a problem, such as incomplete or corrupted measurements, and/or deformation or misalignment.*

Ex: Face recognition with ℓ^1-regression:

Linear subspace model for images of same face under varying illumination:



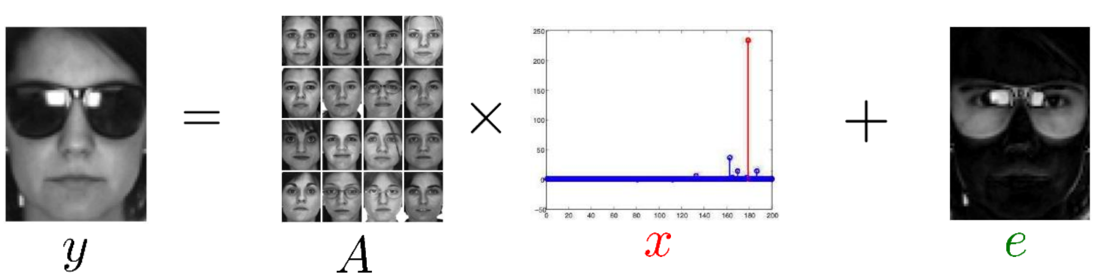
Subject i training



$$A_i = [\ \cdots\ ] \in \mathbb{R}^{m \times k}$$

If test image $\boldsymbol{y} \in \mathbb{R}^m$ is also of subject i, then $\boldsymbol{y} = A_i \boldsymbol{x_i}$ for some $\boldsymbol{x}_i \in \mathbb{R}^k$.

Given $y = Ax_0 + e_0$, $A \in \mathbb{R}^{m \times n}$, $m \ll n$, recover x_0 and e_0.



[Wright, Yang, Ganesh, Sastry, Ma, TPAMI'09]

Ex: Face recognition with ℓ^1-regression:

Underdetermined system of linear equations in unknowns $\boldsymbol{x}, \boldsymbol{e}$:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} = \left[\begin{array}{cc} \boldsymbol{A} & \boldsymbol{I} \end{array}\right] \left[\begin{array}{c} \boldsymbol{x} \\ \boldsymbol{e} \end{array}\right] \doteq \boldsymbol{B}\boldsymbol{w} \qquad \boldsymbol{B} \in \mathbb{R}^{m \times m+n} \quad \boldsymbol{w} \in \mathbb{R}^{m+n}$$

Solution is not unique … but

$\boldsymbol{x}$ should be *sparse*: ideally, only supported on images of the same subject

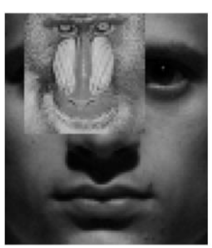
$\boldsymbol{e}$ expected to be *sparse*: occlusion only affects a subset of the pixels

Seek the *sparsest* solution:

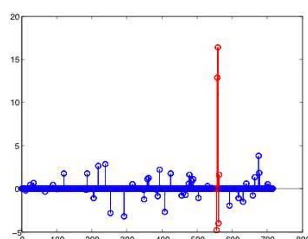
$$\min \|\boldsymbol{x}\|_0 + \|\boldsymbol{e}\|_0 \;\text{ subj }\; \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$$

$$\Downarrow \;\; \text{convex relaxation} \;\; \Downarrow$$

$$\min \|\boldsymbol{x}\|_1 + \|\boldsymbol{e}\|_1 \;\text{ subj }\; \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$$



$\boldsymbol{y}$



$\hat{\boldsymbol{x}}_1$

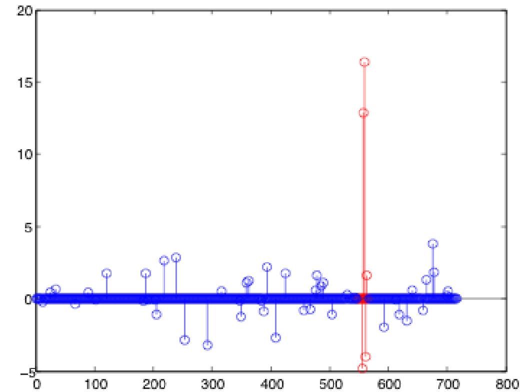
$\hat{\boldsymbol{e}}_1$

$\hat{\boldsymbol{y}} = \boldsymbol{y} - \hat{\boldsymbol{e}}_1$

[Wright, Yang, Ganesh, Sastry, Ma, TPAMI'09]

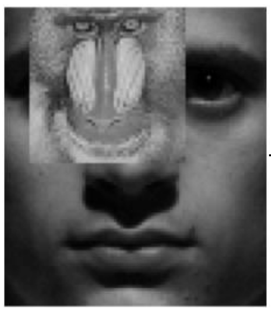
Ex: Face recognition with ℓ^1-regression:

$$\hat{\boldsymbol{x}}_1 \;=\; \arg\min \;\; \|\boldsymbol{x}\|_1 + \|\boldsymbol{e}\|_1.$$



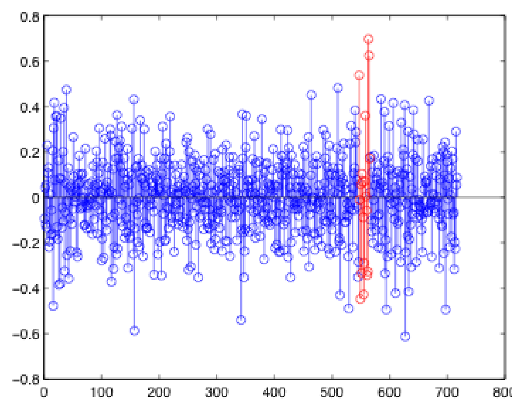
$\hat{\boldsymbol{e}}_1$ $\quad$ $\hat{\boldsymbol{y}}_0 = A\hat{\boldsymbol{x}}_1$

Input: $\boldsymbol{y} \in \mathbb{R}^D$



$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}$$

$$\hat{\boldsymbol{x}}_2 \;=\; \arg\min_{\boldsymbol{x}} \;\; \|\boldsymbol{y} - A\boldsymbol{x}\|_2.$$



$\hat{\boldsymbol{e}}_2$ $\quad$ $\hat{\boldsymbol{y}}_0 = A\hat{\boldsymbol{x}}_2$

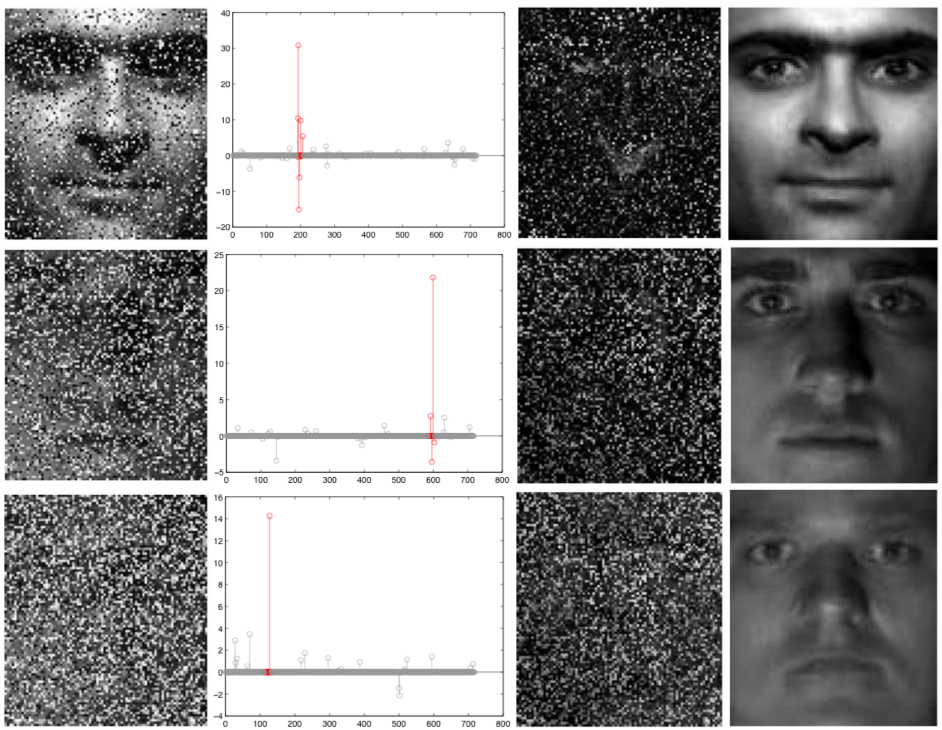
[Wright, Yang, Ganesh, Sastry, Ma, TPAMI’09]

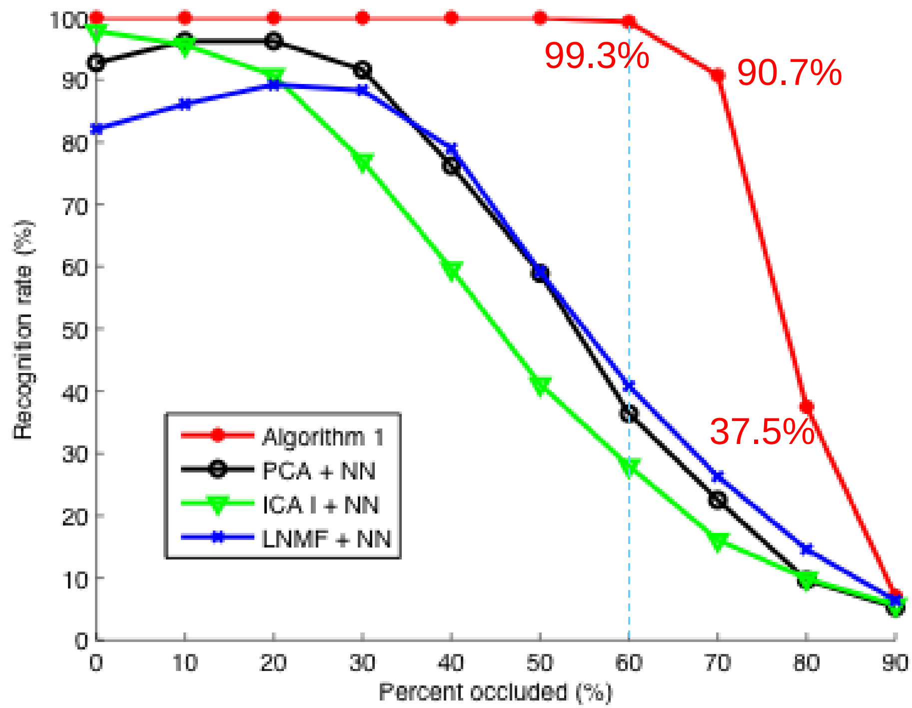
Ex: Face recognition with ℓ^1-regression:

Extended Yale B Database
(38 subjects)

Training: subsets 1 and 2 (717 images)
Testing: subset 3 (453 images)

$\boldsymbol{y}$ $\quad \hat{\boldsymbol{x}}_1 \quad \hat{\boldsymbol{e}}_1 \quad \hat{y}_0 = A\hat{\boldsymbol{x}}_1$





[Wright, Yang, Ganesh, Sastry, Ma, TPAMI'09]

Ex: Face recognition with ℓ^1-regression:

Theorem 1 (W. and Ma, '10). *For any $\delta > 0, \rho < 1$, $\exists\, \nu_0(\delta, \rho) > 0$, such that if $\nu < \nu_0$ and $\alpha < \alpha_0(\delta, \nu, \rho)$, then with error support J and signs $\boldsymbol{\sigma}$ chosen uniformly at random,*

$$\lim_{m \to \infty} \mathbb{P}_{A,J,\sigma}\left[\, \ell^1\text{-minimization recovers all } \alpha m\text{-sparse } \boldsymbol{x} \,\right] = 1.$$

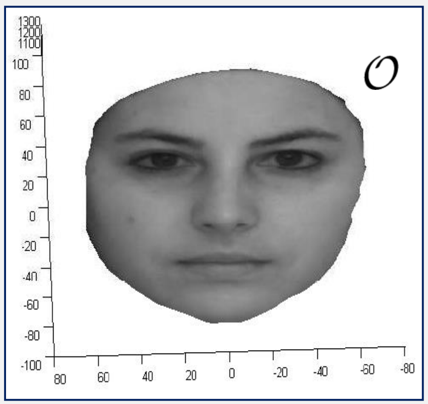
If $\boldsymbol{A}$ is "nice" and the model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$ fits, can make strong statements about the performance of ℓ^1 regression.

[Wright, Ma, Trans. Information Theory'09]

The space/model of face images of a person?

Conceptual Problem:

Given (information about) an object $\mathcal{O}$, can we provably ***detect*** *or* ***recognize*** *the object from a new image $\boldsymbol{y}$?*



$\boldsymbol{y}$

This talk:

Attempt this for ***fixed pose****, variations in* ***illumination,*** *(and possibly occlusion).*

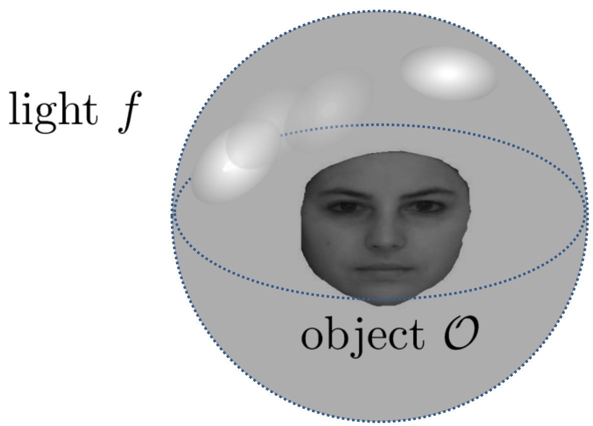
Geometry of illumination variations

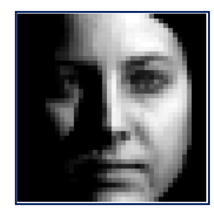
Distant illumination identified with a Riemann integrable, nonnegative function $f : \mathbb{S}^2 \to \mathbb{R}_+$.

Assume a **linear sensor response**. The **image** $\boldsymbol{y}[f]$ can often be written as

$$\boldsymbol{y}[f] = \int_{\boldsymbol{u} \in \mathbb{S}^2} f(\boldsymbol{u}) \bar{\boldsymbol{y}}[\boldsymbol{u}] \, d\boldsymbol{u}$$

What is the set of possible images $\boldsymbol{y}[f]$ of $\mathcal{O}$?





$\boldsymbol{y}[f] \in \mathbb{R}^m$

Geometry of illumination variations

Distant illumination identified with a Riemann integrable, nonnegative function $f : \mathbb{S}^2 \to \mathbb{R}_+$.

Assume a **linear sensor response**. The **image** $\boldsymbol{y}[f]$ can often be written as

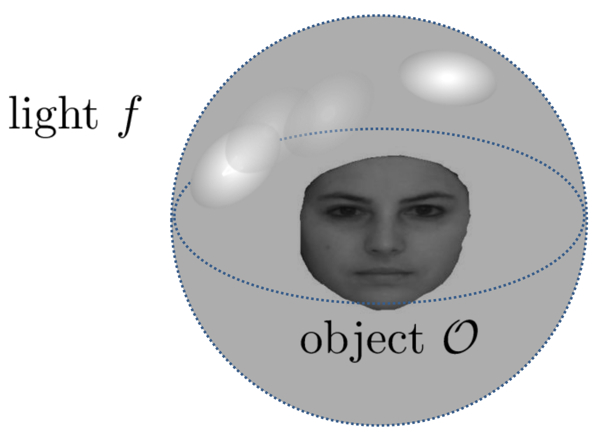
$$\boldsymbol{y}[f] = \int_{\boldsymbol{u} \in \mathbb{S}^2} f(\boldsymbol{u}) \bar{\boldsymbol{y}}[\boldsymbol{u}] \, d\boldsymbol{u}$$

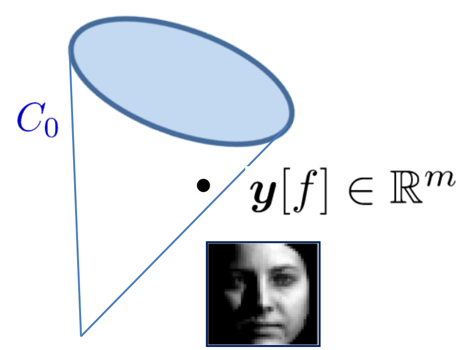
What is the set of possible images $\boldsymbol{y}[f]$ of $\mathcal{O}$?

Let $\mathcal{F}_0 \doteq \{ f : \mathbb{S}^2 \to \mathbb{R}_+ \mid f \text{ Riemann integrable} \}$, *and*

$$C_0 = \{ \boldsymbol{y}[f] \mid f \in \mathcal{F}_0 \} \subset \mathbb{R}^m.$$

The set C_0 *is a* ***convex cone.*** [Belhumeur + Kriegman '98]



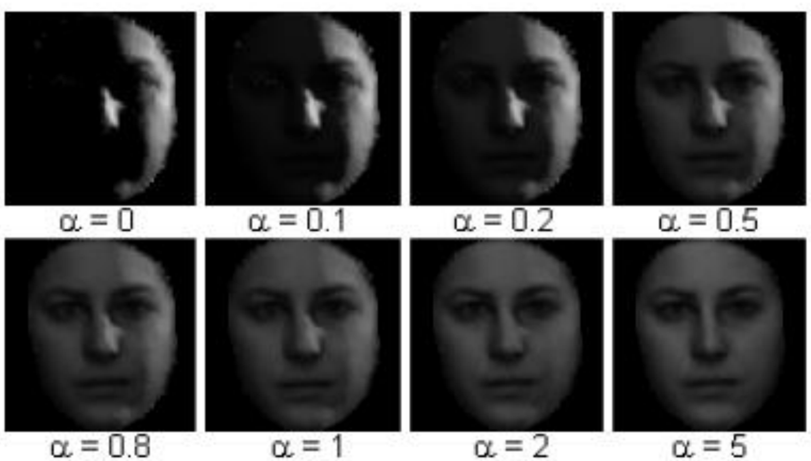


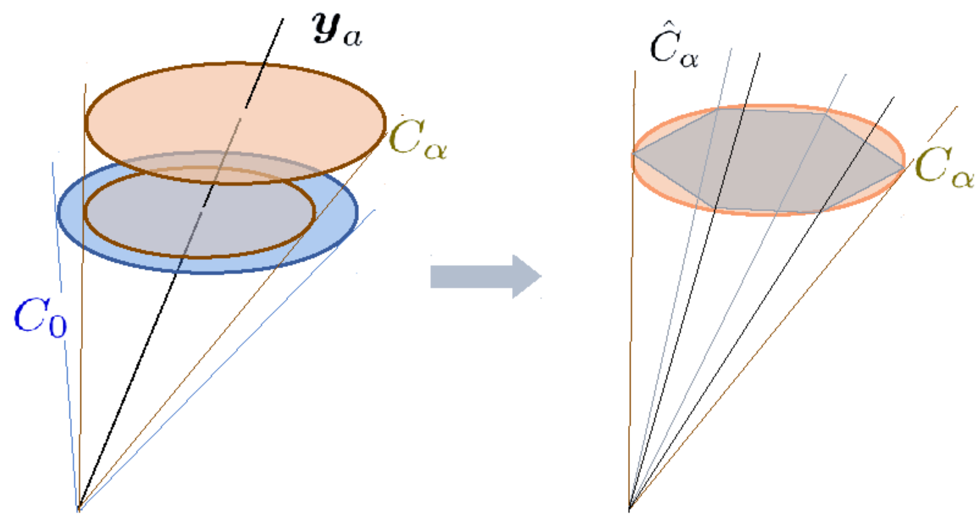
Geometry of illumination variations

Ambient cone model. *Illumination is the sum of ambient and directional (arbitrary) components:*

$$\mathcal{F}_\alpha \doteq \left\{ f_d + \alpha\omega \mid f_d : \mathbb{S}^2 \to \mathbb{R}_+, \text{Riemann integrable}, \ \|f_d\|_{L^1} \le 1 \right\}$$

and corresponding cone of possible images: $C_\alpha \doteq \operatorname{cone}\Big(\boldsymbol{y}[\mathcal{F}_\alpha] \Big)$



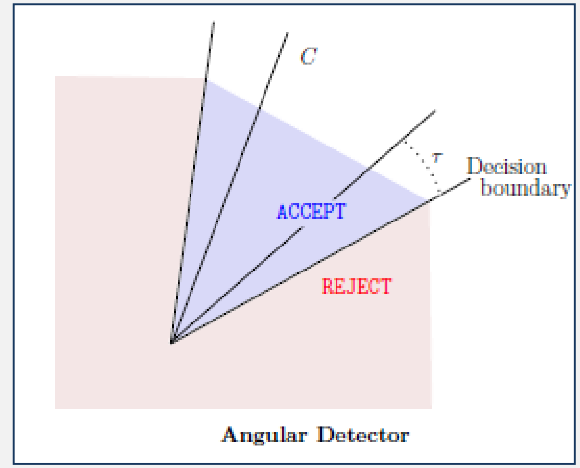


Approximating a convex cone

Angular detector: is $\boldsymbol{y}$ in or close to C_α ?

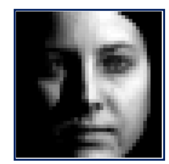
Accept any input $\boldsymbol{y}$ that is an image of $\mathcal{O}$ under some valid illumination.
Reject any input $\boldsymbol{y}$ that is not.

$$\mathfrak{D}_\tau[\boldsymbol{y}] = \begin{cases} \text{ACCEPT} & \angle(\boldsymbol{y}, C) \le \tau \\ \text{REJECT} & \angle(\boldsymbol{y}, C) > \tau \end{cases}$$

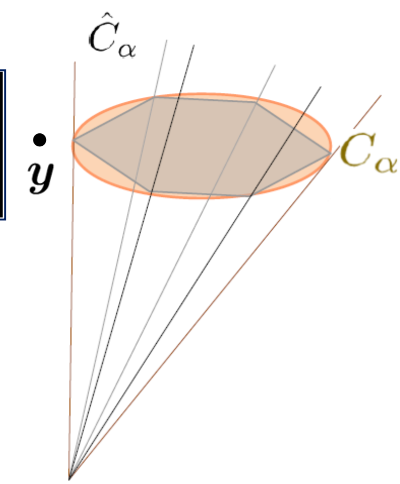


Work with a "Hausdorff distance" between cones:

$$\begin{aligned} \delta(C, C') &\doteq d_{\text{Hausdorff}}\Big(C \cap B(0,1), C' \cap B(0,1)\Big) \\ &= \max\left\{ \sup_{\boldsymbol{y} \in C \cap B(0,1)} d(\boldsymbol{y}, C'), \sup_{\boldsymbol{y}' \in C' \cap B(0,1)} d(\boldsymbol{y}', C) \right\} \end{aligned}$$



$\mathcal{O}$



If C' approximates C in Hausdorff sense, we lose little in working with C'.

Cone Approximation: Existing Results

Approximation of **general** convex bodies in high dimensions is a disaster:

Theorem 2. [Bronstein, Ivanov '76] *Let $K \subset \mathbb{R}^m$ be a convex body. There exists an ε-approximation to K in Hausdorff distance, with $O\Big((\mathrm{diam}(K)/\varepsilon)^{\frac{m-1}{2}}\Big)$ vertices. For the unit sphere, this is optimal up to a constant.*

In vision literature, results for **average case**, for **convex objects**:

Informal claim [Basri and Jacobs '03, Basri and Frolova '04]:
For an convex, Lambertian object $\mathcal{O}$, there exists a subspace Γ, with $\dim(\Gamma) = 9$, such that if $\boldsymbol{u} \sim \mathrm{uni}\left(\mathbb{S}^2\right)$ is a uniformly oriented random point source,

$$\frac{\mathbb{E}\left[\|\bar{\boldsymbol{y}}[\boldsymbol{u}] - \mathcal{P}_\Gamma \bar{\boldsymbol{y}}[\boldsymbol{u}]\|_2^2\right]}{\mathbb{E}\left[\|\bar{\boldsymbol{y}}[\boldsymbol{u}]\|_2^2\right]} \le .02$$

Is there any special structure we can use here?

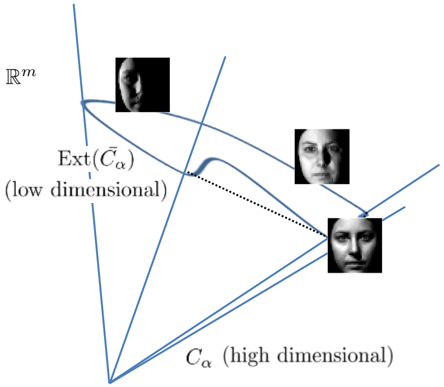
Extreme rays

Image formation: $\boldsymbol{y}[f] = \int_{\boldsymbol{u}\in\mathbb{S}^2} f(\boldsymbol{u})\bar{\boldsymbol{y}}[\boldsymbol{u}]\,d\boldsymbol{u}\,.$

$\bar{\boldsymbol{y}}[\boldsymbol{u}] \in \mathbb{R}^m$ are images under **directional illumination**:




Extreme rays lie on a low dimensional submanifold of $\mathbb{R}^m$:

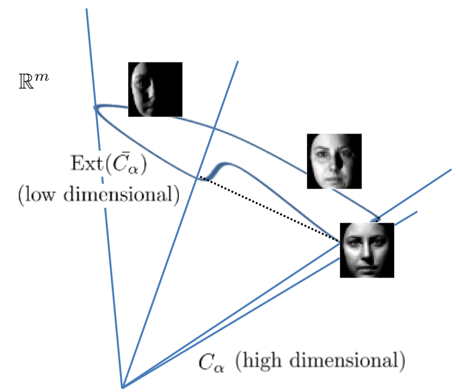


Extreme rays

Image formation: $\boldsymbol{y}[f] = \int_{\boldsymbol{u}\in\mathbb{S}^2} f(\boldsymbol{u})\bar{\boldsymbol{y}}[\boldsymbol{u}]\, d\boldsymbol{u}$.

$\bar{\boldsymbol{y}}[\boldsymbol{u}] \in \mathbb{R}^m$ are images under **directional illumination:**

Extreme rays lie on a low dimensional submanifold of $\mathbb{R}^m$ **:**



Lemma. *[Just approximate the extreme rays]. Let* $\boldsymbol{u}_1, \dots, \boldsymbol{u}_N \in \mathbb{S}^2$. *Then*

$$\delta\Big(C_\alpha,\ \mathrm{cone}\,\{\bar{\boldsymbol{y}}[\boldsymbol{u}_i] + \alpha \boldsymbol{y}_a\} \Big) \ \le\ \frac{2}{\eta_\star \alpha \|\boldsymbol{y}_a\|} \times \sup_{\boldsymbol{u}} \min_i \|\bar{\boldsymbol{y}}[\boldsymbol{u}] - \bar{\boldsymbol{y}}[\boldsymbol{u}_i]\|_2,$$

where $\eta_\star = \sup_{\|\boldsymbol{w}\|_2 \le 1} \min_i \left\langle \boldsymbol{w}, \frac{\bar{\boldsymbol{y}}[\boldsymbol{u}_i]}{\|\bar{\boldsymbol{y}}[\boldsymbol{u}_i]\|_2} \right\rangle \ \ge\ 1/\sqrt{m}.$

Cone Approximation: New Results

Theorem 3 (Zhang, Mu, Kuo, W. '12) *(sketch) Under our previously stated hypotheses, for all $\boldsymbol{u}$, $\boldsymbol{u}' \in \mathbb{S}^2$,*

$$\|\bar{\boldsymbol{y}}[\boldsymbol{u}] - \bar{\boldsymbol{y}}[\boldsymbol{u}']\|_2 \quad \leq \quad \frac{C_{sensor}}{1-\nu_\star} \times \left(\operatorname{area}(\partial\mathcal{O}) \, \|\boldsymbol{u} - \boldsymbol{u}'\|_2^2 + \chi_\star \operatorname{diam}(\mathcal{O}) \, \|\boldsymbol{u} - \boldsymbol{u}'\|_2 \right)^{1/2}$$

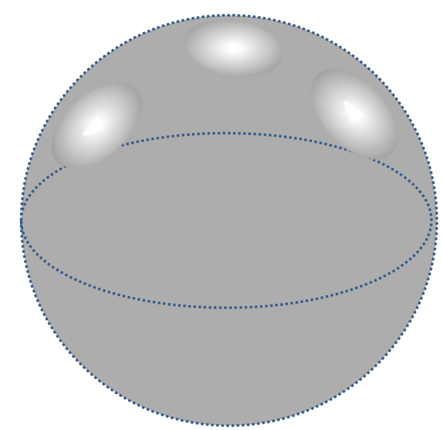
where C_{sensor} depends on the parameters of the imaging system.

If we define an **illumination covering number**

$$N(C,\gamma) \;=\; \min\left\{\#V \mid \delta(C, \operatorname{cone}(V)) \;\leq\; \gamma\right\},$$

This implies in general, $N(C_\alpha, \gamma) \;\leq\; \dfrac{h(\mathcal{O}, \text{sensor})}{(\alpha\gamma)^4}$,

and for convex objects $N(C_\alpha, \gamma) \;\leq\; \dfrac{h(\mathcal{O}, \text{sensor})}{(\alpha\gamma)^2}$.

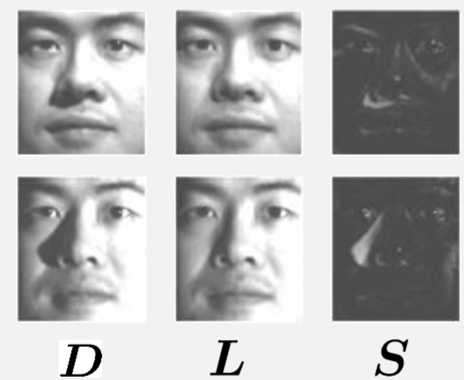


Proof Sketch

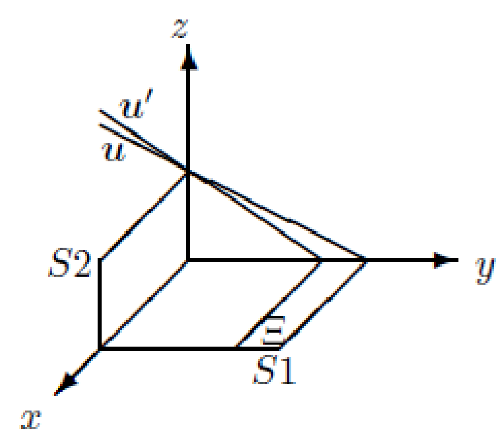
Control the complexity of cone approximation in terms of object convexity defect.

The conceptual idea:

Separate the direct illumination operator into a **low-rank** part (due to smooth variations) and a **sparse** part (due to cast shadows):



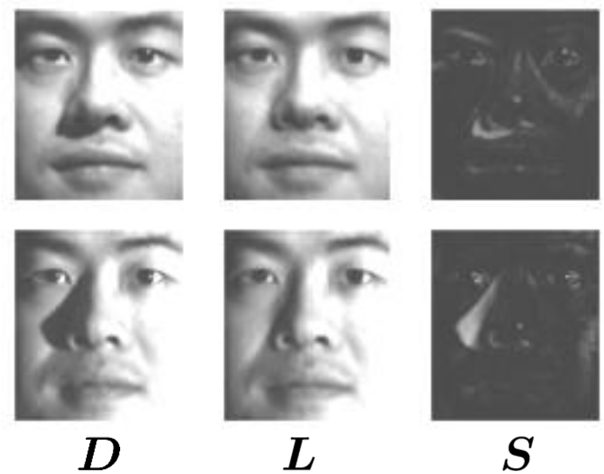
The low-rank term can be bounded by direct calculation; the sparse term needs a more involved accounting.



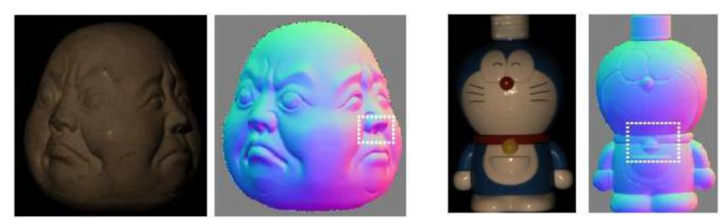
LR+S captures physical intuition

Calculations with the Lambertian sphere suggest that sets of images of **smooth, near-convex** objects should be approximately **low-rank**:
[Basri+Jacobs '03,Ramamoorthi '04].

Cast shadows are often **sparse**
[W., Yang, … Ma, '09] , [Candes, Li, Ma, W. '11]:



Observations used in photometric stereo:
[Wu et. Al. '11] :

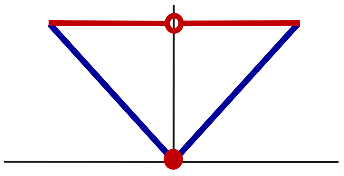


How can we exploit these structures to obtain compact representation/approximation of the cone?

Recover low-rank and sparse components

Convex relaxation: $\min \; \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t} \quad \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{D}.$

with $\begin{array}{l} \|\boldsymbol{L}\|_* = \sum_i \sigma_i(\boldsymbol{L}) \\ \|\boldsymbol{S}\|_1 = \sum_{ij} |\boldsymbol{S}_{ij}| \end{array}$



Provably effective for recovery (and **statistical estimation**):

Theorem 5 (Candès, Li, Ma, W. '11). *If* $\boldsymbol{L}_0 \in \mathbb{R}^{m \times n}$, $m \geq n$ *has rank*

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

and $\boldsymbol{S}_0$ *has Bernoulli support with error probability* $\rho \leq \rho_s^\star$, *then w.h.p.,*

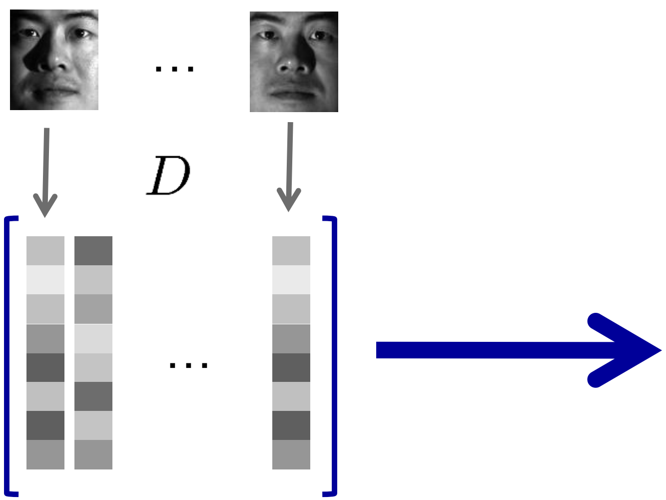
$$(\boldsymbol{L}_0, \boldsymbol{S}_0) \;=\; \arg\min \|\boldsymbol{L}\|_* + \frac{1}{\sqrt{m}} \|\boldsymbol{S}\|_1 \quad \text{subj} \quad \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{L}_0 + \boldsymbol{S}_0,$$

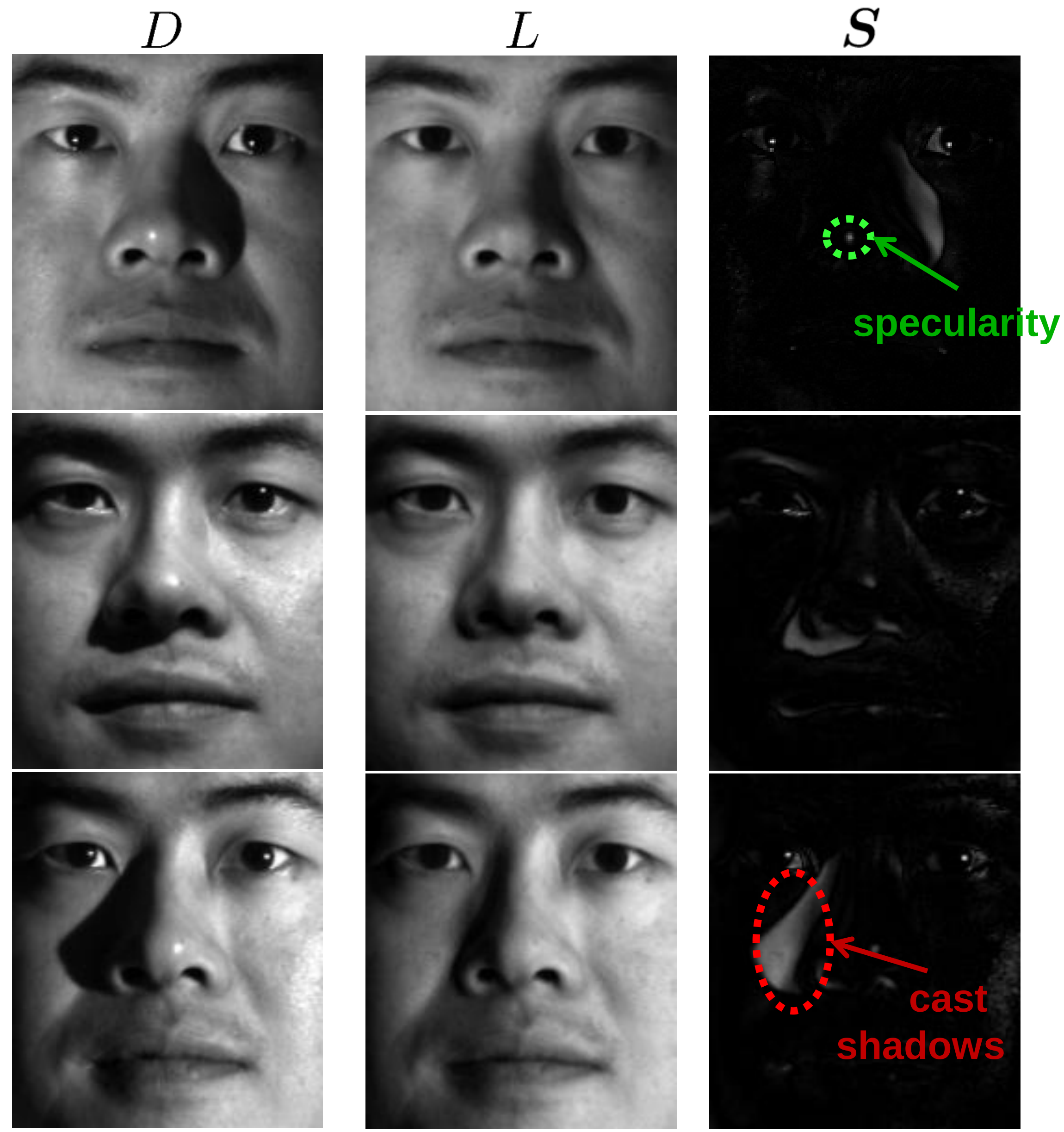
and the minimizer is unique.

[Chadrasekaran,Sanghavi,Willsky,Parillo'11], [Candes, Li, Ma, W. '11], [Hsu,Kakade,Zhang '12], [Agarwal, Wainwright '12], …

Recover low-rank and sparse components

58 images of one person under varying lighting:





Cone-preserving dimensionality reduction

A low-rank and sparse decomposition that provably **preserves verification performance**?

Would need to solve

$$\text{minimize } \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \delta\Big(\text{cone}(\boldsymbol{L}+\boldsymbol{S}), \text{cone}(\boldsymbol{A})\Big) \;\leq\; \gamma$$

… but constraint is very complicated. Relax to:

Theorem 6 (Zhang, Mu, Kuo, W. '13) *Let* $(\boldsymbol{L}_\star, \boldsymbol{S}_\star)$ *solve*

$$\text{minimize } \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \begin{bmatrix} \boldsymbol{I} & \boldsymbol{L}+\boldsymbol{S}-\boldsymbol{A} \\ \boldsymbol{L}^*+\boldsymbol{S}^*-\boldsymbol{A}^* & \gamma' \boldsymbol{A}^*\boldsymbol{A} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0}, \; \boldsymbol{\mu} \geq \mathbf{0},$$

with $\gamma' = \frac{\gamma}{1+\gamma}$. *Then* $\delta\Big(\text{cone}(\boldsymbol{L}_\star+\boldsymbol{S}_\star), \text{cone}(\boldsymbol{A})\Big) \;\leq\; \gamma.$

Provable Low-Dim Structures of Visual Data

Guaranteed Illumination Models for Nonconvex Objects, Zhang, Mu, Kuo, W., Arxiv '13

Compressive Principal Component Pursuit, W., Ganesh, Min, Ma, I&I '13

Towards a Practical Automatic Face Recognition, Wagner, W., Ganesh, Zhou, Mobahi, Ma, PAMI '12

Robust Principal Component Analysis? Candes, Li, Ma, W. JACM '11

Dense Error Correction via L1 Minimization. Wright and Ma, Information Theory '10

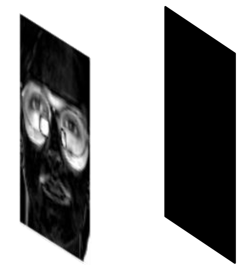
Robust Face Recognition via Sparse Representation, W., Yang, Ganesh,, Sastry, Ma, PAMI '09

Real Face Images from the Internet: any low-dim structures?

*48 images collected from internet

Robust Alignment of Multiple (Face) Images

D – corrupted & misaligned observation | A – aligned low-rank signals | E – sparse errors

 $\circ\ \tau =$  $+$ 

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:

$$\min \ \|A\|_* + \lambda \|E\|_1 \quad \text{subj} \quad \underline{A + E = D \circ \tau_k + J \Delta\tau}$$

$$(\text{or} \quad Q(A+E) = QD \circ \tau_k, \ QJ = 0)$$

RASL: *Faces Detected*

Input: faces detected by a face detector (D)

Average

RASL: *Faces Aligned*

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

Average

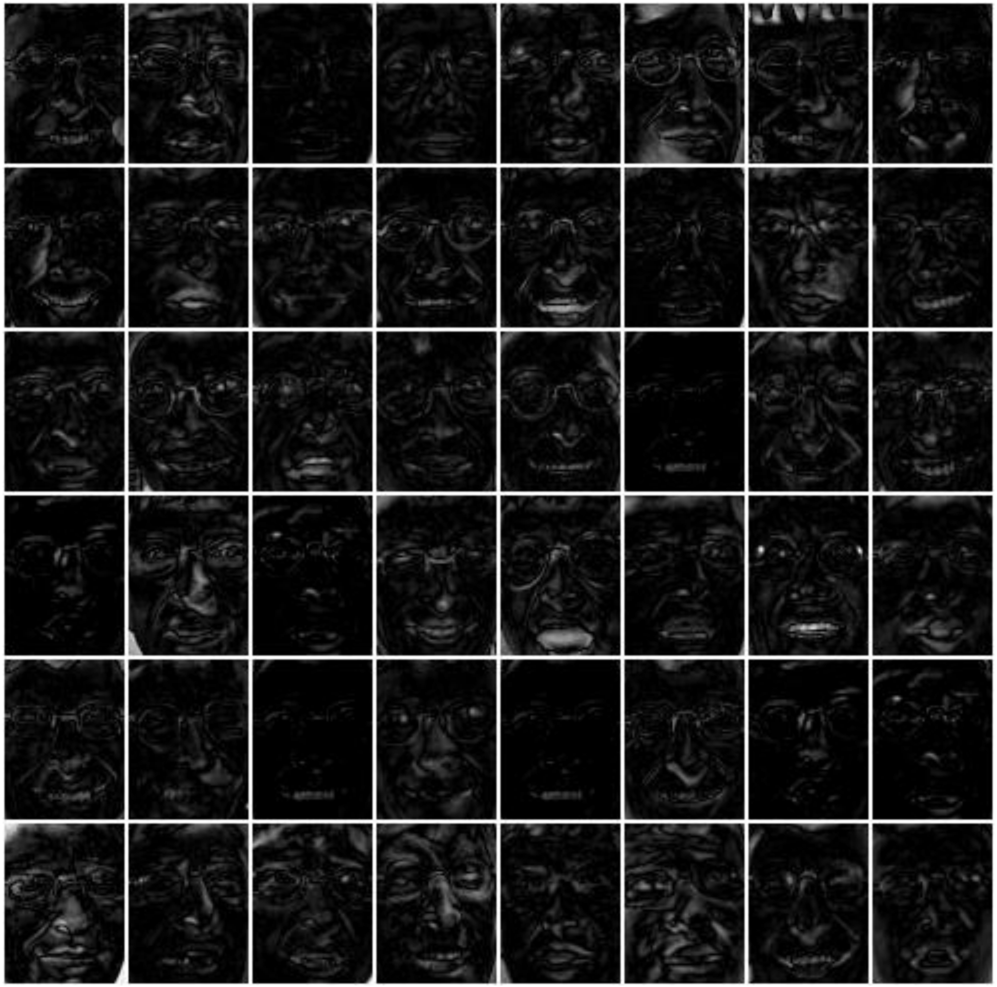
RASL: *Faces Cleaned as the Low-Rank Component*

Output: clean low-rank faces (A)

Average

RASL: *Sparse Errors of the Face Images*

Output: sparse error images (E)



RASL: *Video Stabilization and Enhancement*

Original video (D)

Aligned video ($D \circ \tau$)

Low-rank part (A)

Sparse part (E)

RASL: *Aligning Handwritten Digits*

D

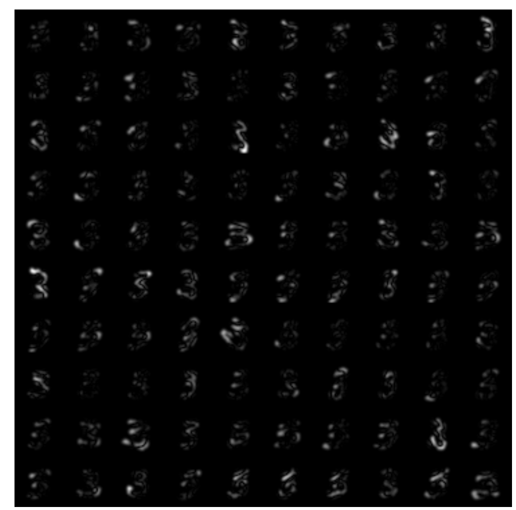
Learned-Miller PAMI'06

Vedaldi CVPR'08

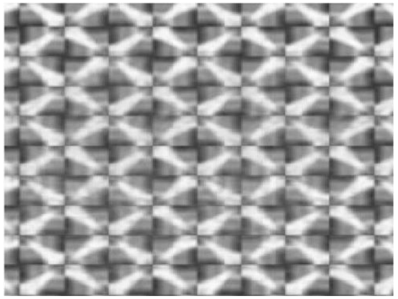
$D \circ \tau$

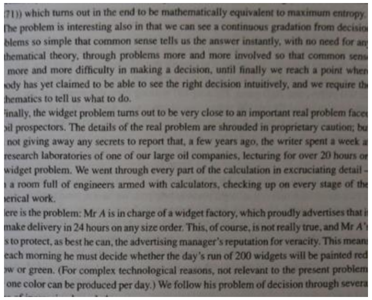
A

E



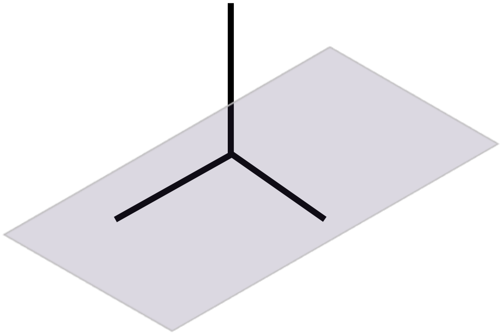
CONTEXT – *Low dimensional structures in visual data*





71)) which turns out in the end to be mathematically equivalent to maximum entropy.
The problem is interesting also in that we can see a continuous gradation from decisio
blems so simple that common sense tells us the answer instantly, with no need for an
thematical theory, through problems more and more involved so that common sens
more and more difficulty in making a decision, until finally we reach a point whe
ody has yet claimed to be able to see the right decision intuitively, and we require th
hematics to tell us what to do.
inally, the widget problem turns out to be very close to an important real problem face
il prospectors. The details of the real problem are shrouded in proprietary caution; bu
not giving away any secrets to report that, a few years ago, the writer spent a week a
research laboratories of one of our large oil companies, lecturing for over 20 hours o
widget problem. We went through every part of the calculation in excruciating detail –
a room full of engineers armed with calculators, checking up on every stage of the
erical work.
ere is the problem: Mr *A* is in charge of a widget factory, which proudly advertises that i
make delivery in 24 hours on any size order. This, of course, is not really true, and Mr *A*'
s to protect, as best he can, the advertising manager's reputation for veracity. This mean
each morning he must decide whether the day's run of 200 widgets will be painted red
w or green. (For complex technological reasons, not relevant to the present problem
one color can be produced per day.) We follow his problem of decision through severa

Individual images exhibit ***low-dimensional structures*** due to rich ***local*** regularities, ***global*** symmetries, ***repetitive*** patterns, or ***redundant*** sampling.

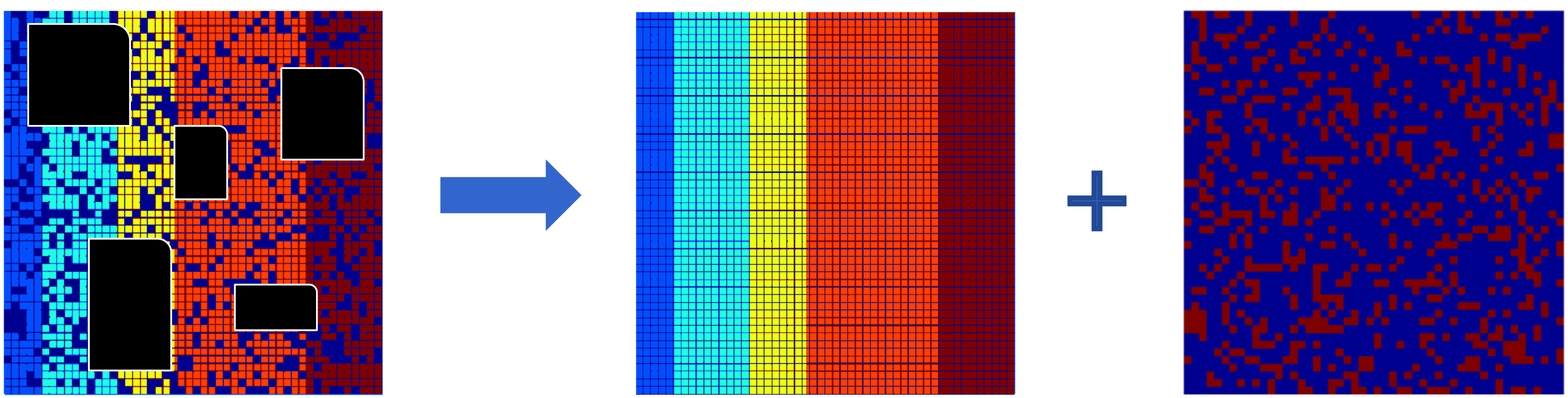


Robust Recovery of Low-Dimensional Models

Recover low-dimensional structures from a fraction of missing measurements with structured support and/or a fraction of random corruptions.

$$D = \mathcal{P}_\Omega[\ {\color{red}A_0} + {\color{green}E_0}\ ], \qquad \Omega \sim \mathrm{uni}\binom{[m]\times[n]}{mn}$$

D — Low-rank Texture A — Sparse Errors E



MAIN THEORY – *Corrupted, Incomplete Matrix*

Theorem 2 (Matrix Completion and Recovery). *If* $A_0, E_0 \in \mathbb{R}^{m\times n}, m \geq n$, *with*

$$\operatorname{rank}(A_0) \;\leq\; C\,\frac{n}{\mu\log^2(m)}, \quad \textit{and} \quad \|E_0\|_0 \;\leq\; \rho^\star mn,$$

and we observe only a random subset of size

$$|\Omega| = mn/10$$

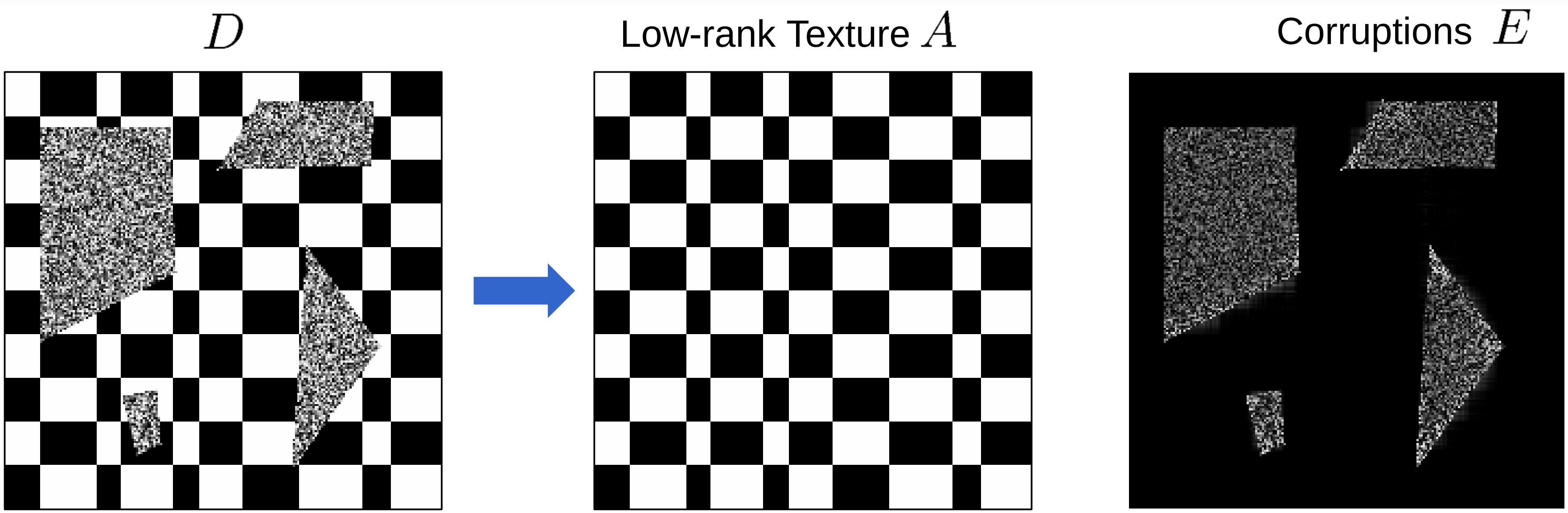
entries, then with very high probability, solving the convex program

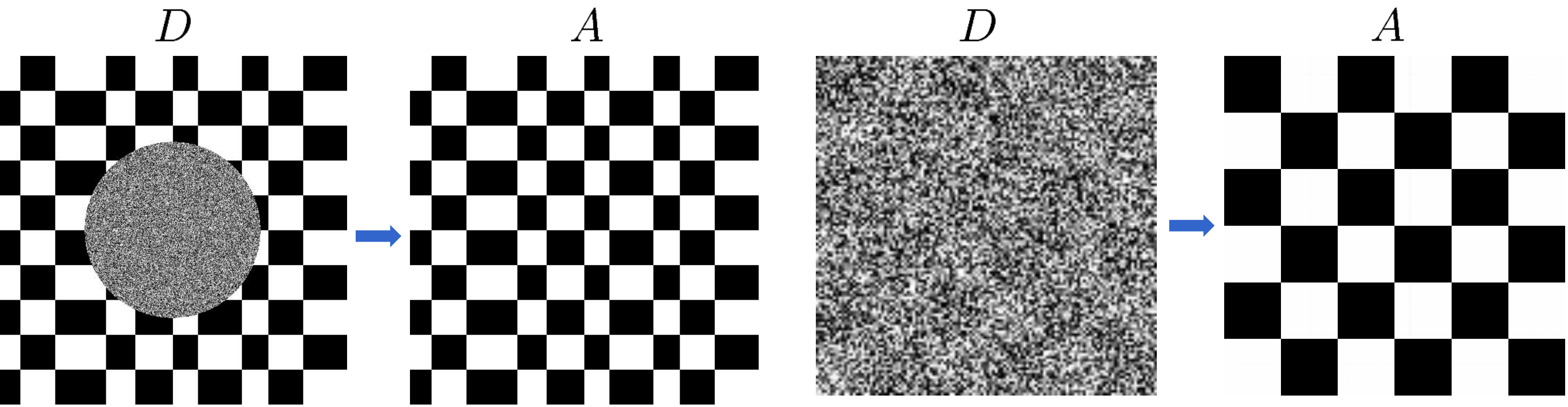
$$\min \|A\|_* + \frac{1}{\sqrt{m}}\|E\|_1 \quad \text{subj} \quad P_\Omega[A+E] = D,$$

uniquely recovers (A_0, E_0).

"Convex optimization recovers almost any matrix of rank $O\left(\frac{m}{\log^2 n}\right)$ from any small fraction of entries, even with $O\left(mn\right)$ entries corrupted!"

Repairing Images: Highly Robust Repairing of Low-rank Textures!





What about repairing distorted low-rank textures?

Low-rank Method | Photoshop

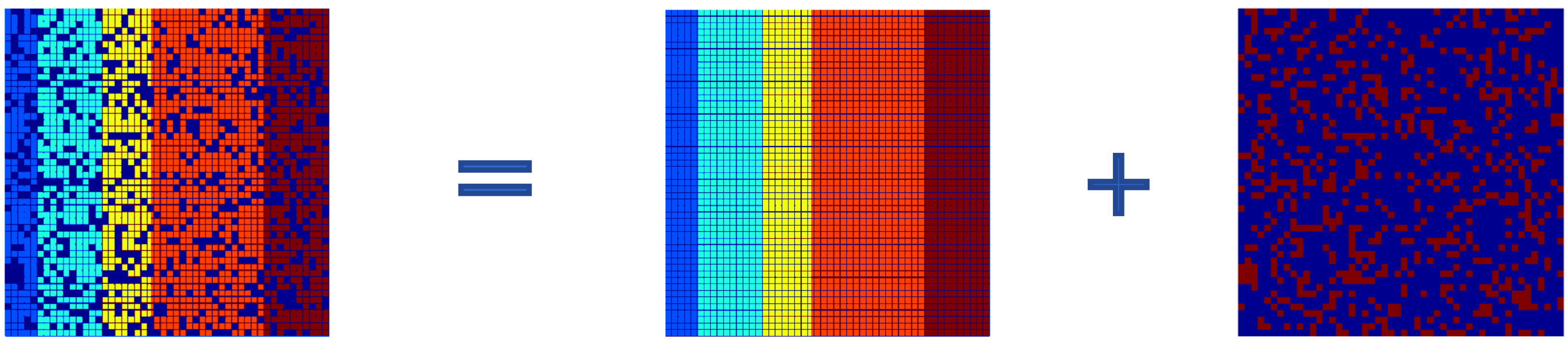
Input

Output

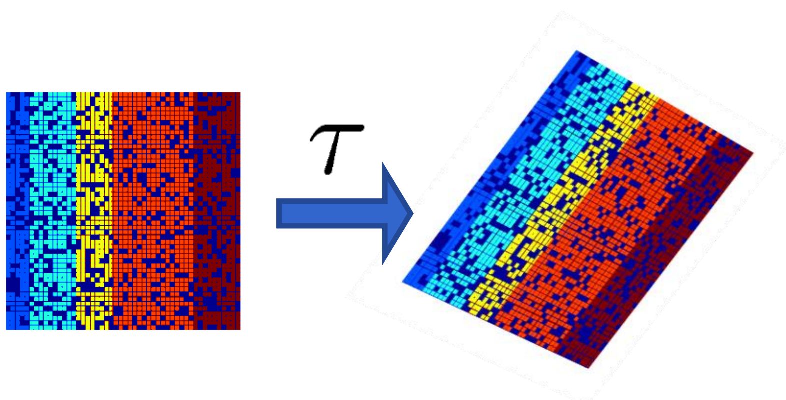
Sensing or Imaging of Low-rank and Sparse Structures

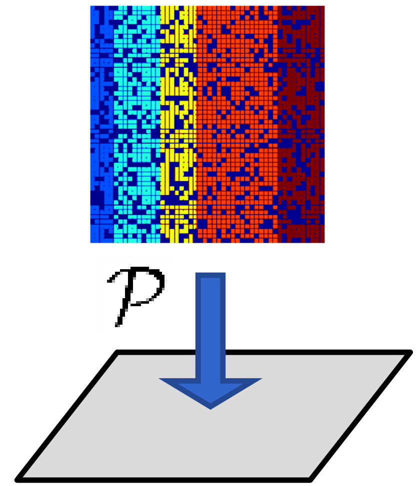
Fundamental Problem: ***How to recover low-rank and sparse structures from corrupted data***

Low-rank Structures Sparse Structures



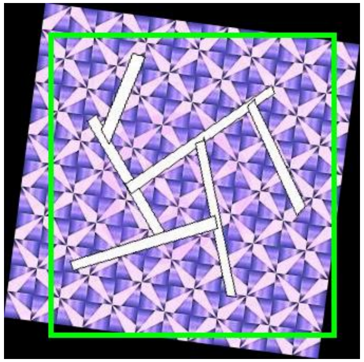
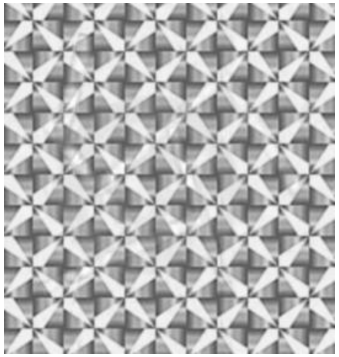
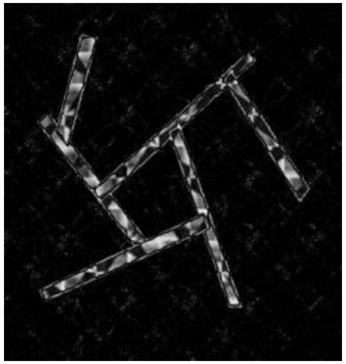
subject to either nonlinear deformation τ or linear compressive sampling $\mathcal{P}$?





Reconstructing 3D Geometry and Structures

D – deformed observation | A – low-rank structures | E – sparse errors

 $\circ\, \tau \;=$  $+$ 

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

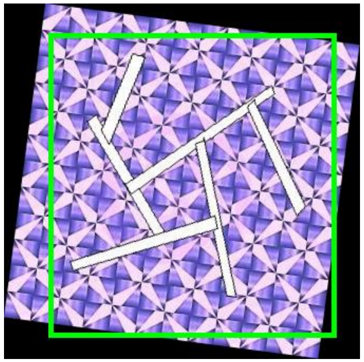
Low-rank component
(regular patterns…)

Sparse component
(occlusion, corruption, foreground…)

Parametric deformations
(affine, projective, radial distortion, 3D shape…)

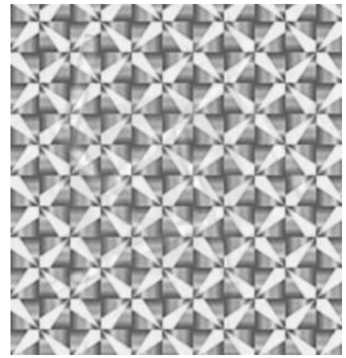
Transform Invariant Low-rank Textures (TILT)

D – deformed observation



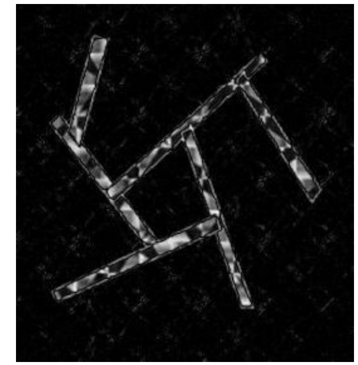
$\circ\, \tau \;=$

A – low-rank structures



$+$

E – sparse errors



Objective: *Transformed Principal Component Pursuit:*

$$\min \ \|A\|_* + \lambda \|E\|_1 \quad \text{subj} \quad A + E = D \circ \tau$$

Solution: *Iteratively solving the linearized convex program:*

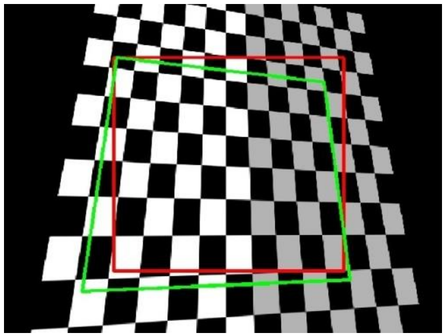
$$\min \ \|A\|_* + \lambda \|E\|_1 \quad \text{subj} \quad A + E = D \circ \tau_k + J \cdot \Delta\tau$$

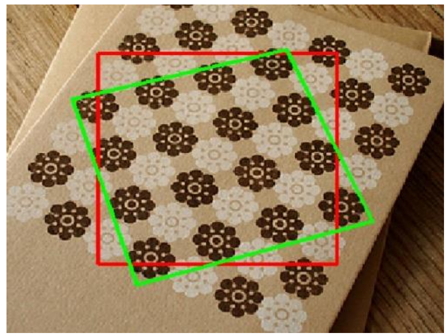
Or reduced version: $\text{subj} \quad \mathcal{P}_Q[A + E] = \mathcal{P}_Q[D \circ \tau_k], \ \mathcal{P}_Q[J] = 0$

Zhang, Liang, Ganesh, Ma, ACCV'10, IJCV'12

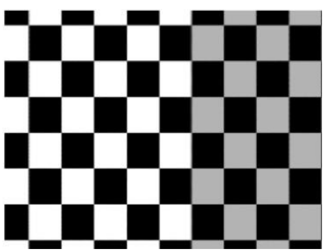
TILT: *Shape from texture*

Input (red window D)





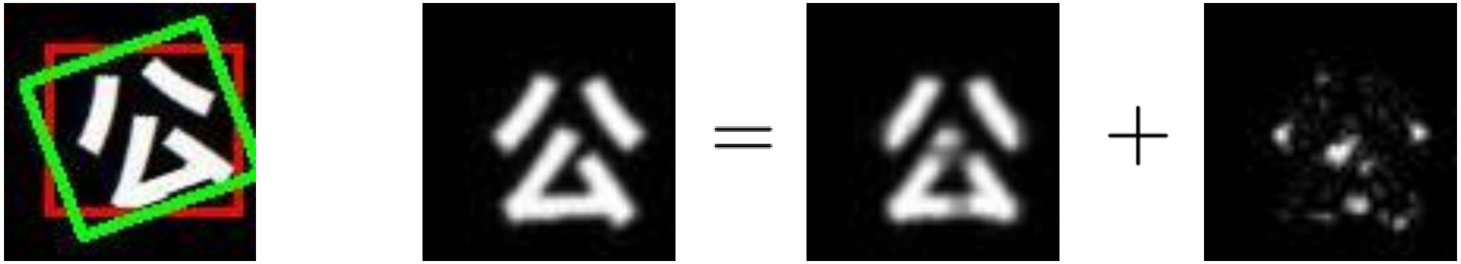
Output (rectified green window A)



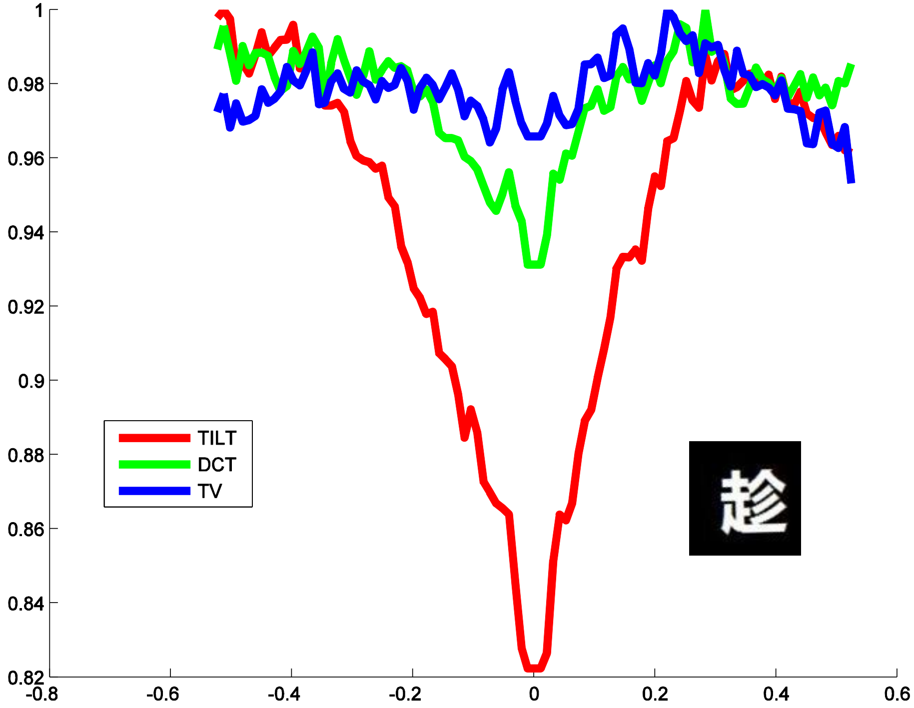
Zhang, Liang, Ganesh, Ma, ACCV'10, IJCV'12

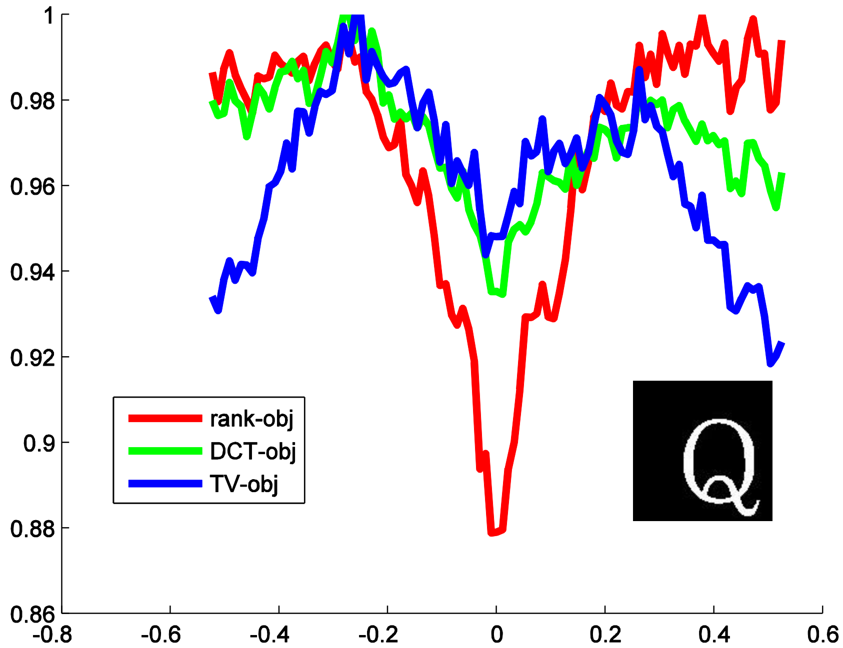
Structured Texture Completion and Repairing

Recognition: *Character/Text Rectification*



$D \qquad D \circ \tau \qquad A \qquad E$

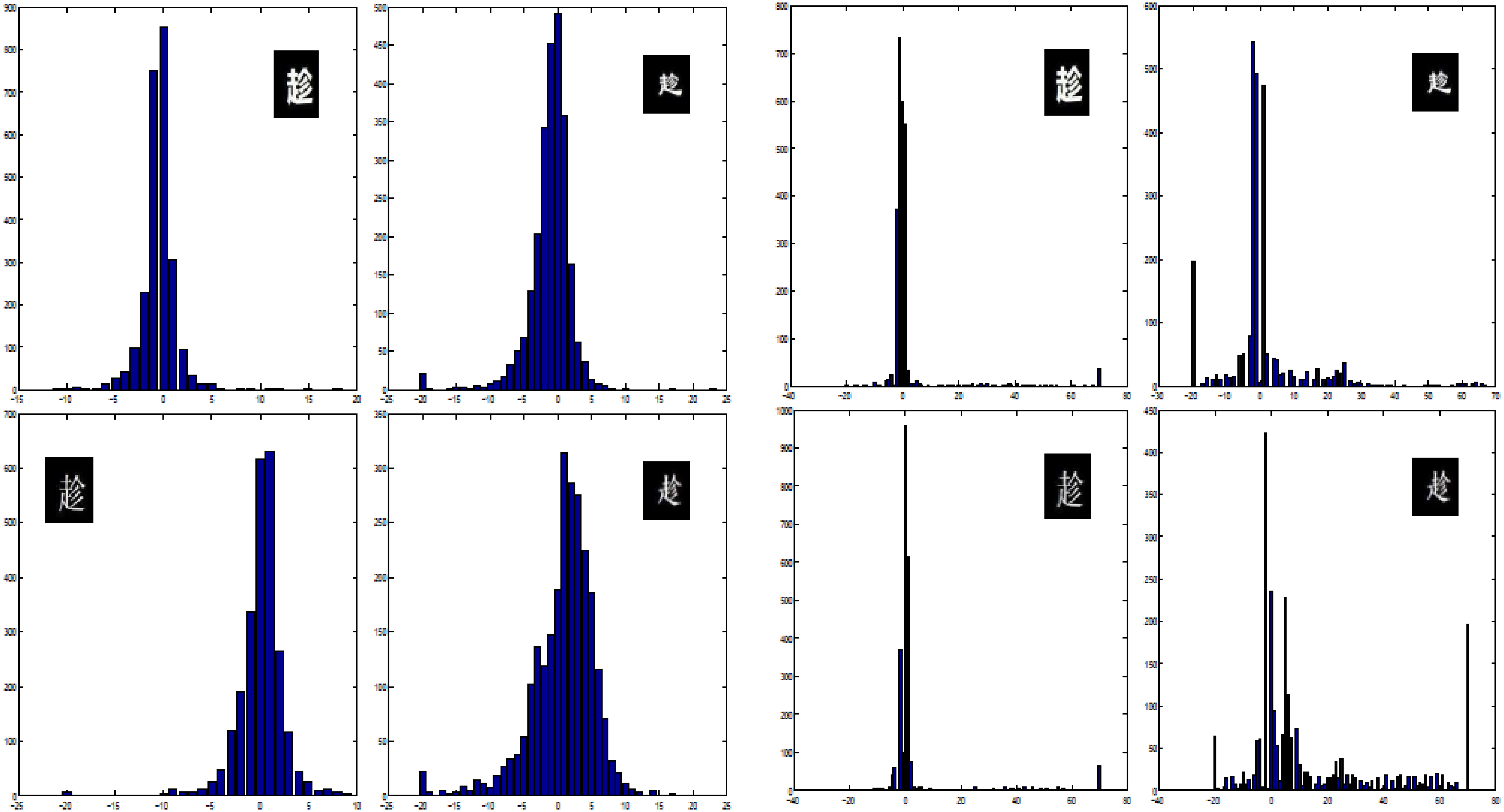




Xin Zhang, Zhouchen Lin, and Ma, ICDAR 2013

Recognition: *Character/Text Rectification*

TILT **versus** **Hough Transform**

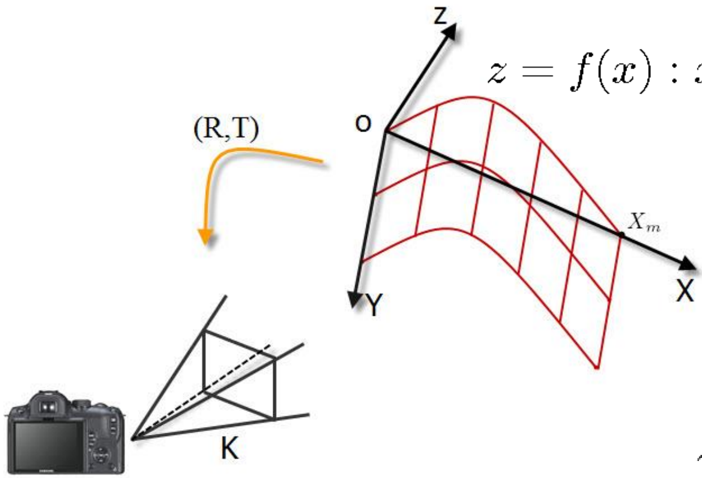


Object Recognition: *Regularity of Texts at All Scales!*

Input (red window D)

Output (rectified green window A)

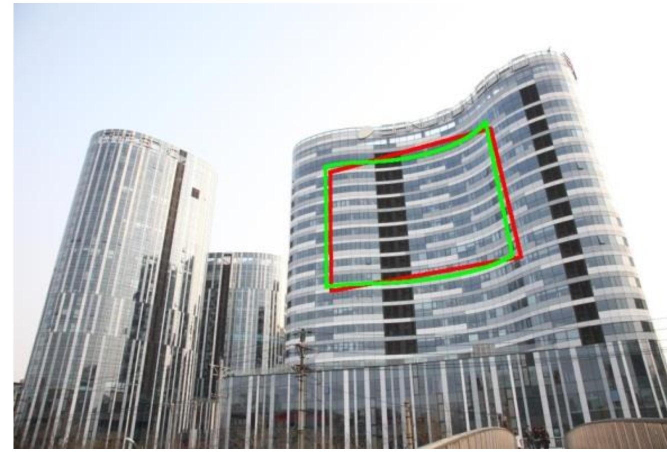
TILT: *Shape and geometry from textures*



$$z = f(x) : x(x - x_m) \sum_{i=1}^{d} a_i x^i$$

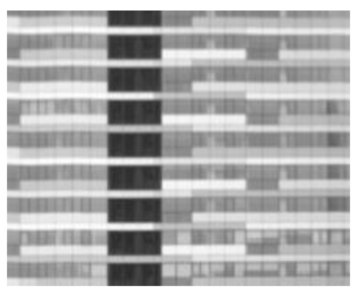
$$D \circ \tau = A + E$$

$$\tau = (K, R, T, \{a_i\})$$

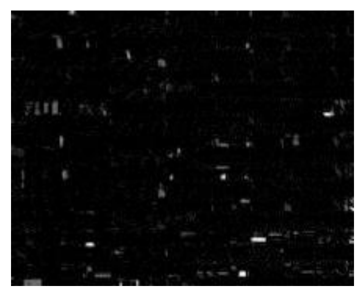


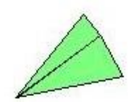
D

A



E

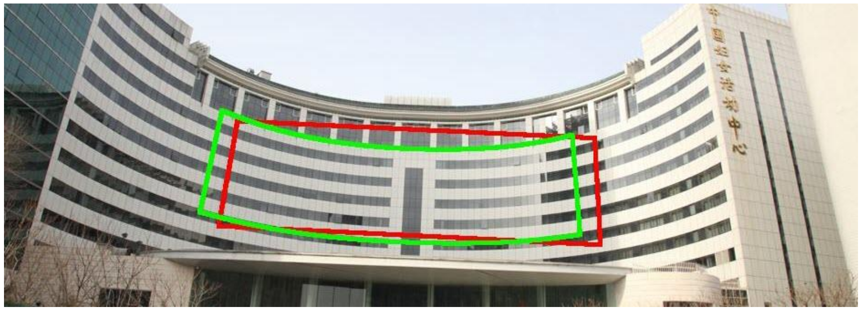




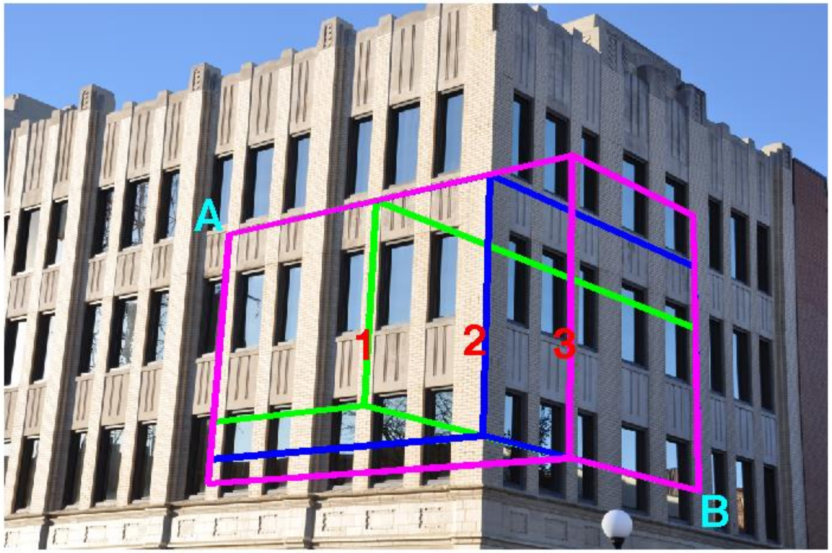
TILT: *Shape and geometry from textures*

360º panorama

TILT: *Virtual reality*

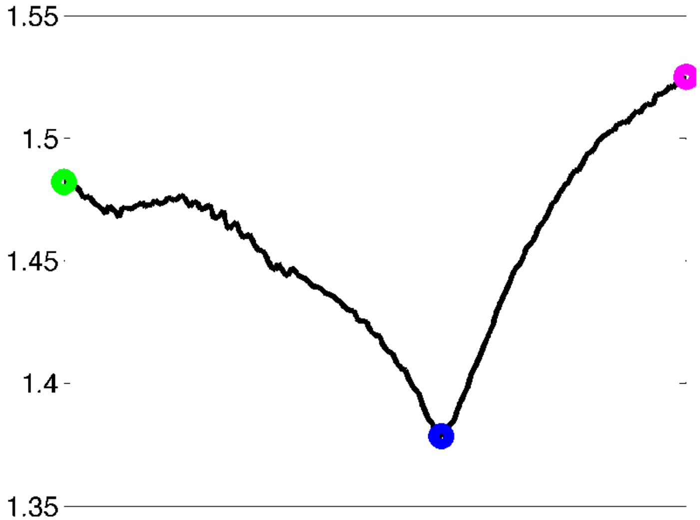


TILT: *Holistic 3D Reconstruction of Urban Scenes*



$$\min \|A\|_* + \|E\|_1 \quad \text{s.t.}$$

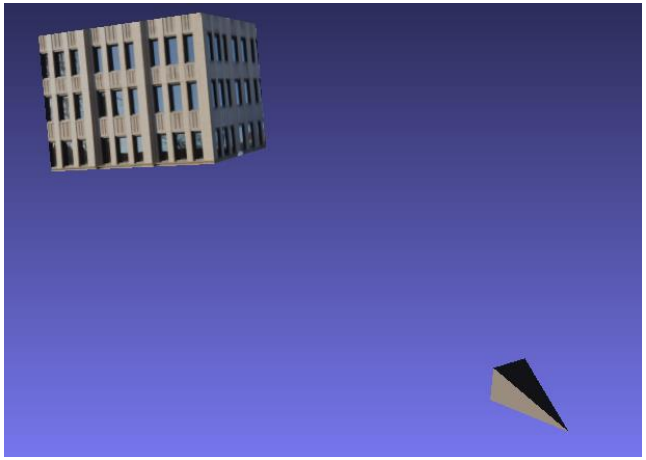
$$A + E = [D_1 \circ \tau_1, D_2 \circ \tau_2]$$



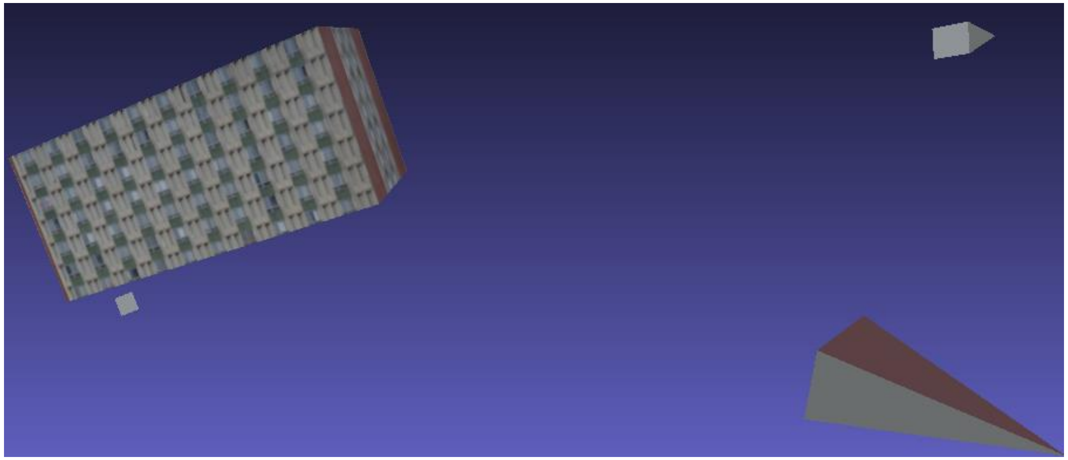
Virtual Reality in Urban Scenes

TILT: *Holistic 3D Reconstruction of Urban Scenes*

From one input image

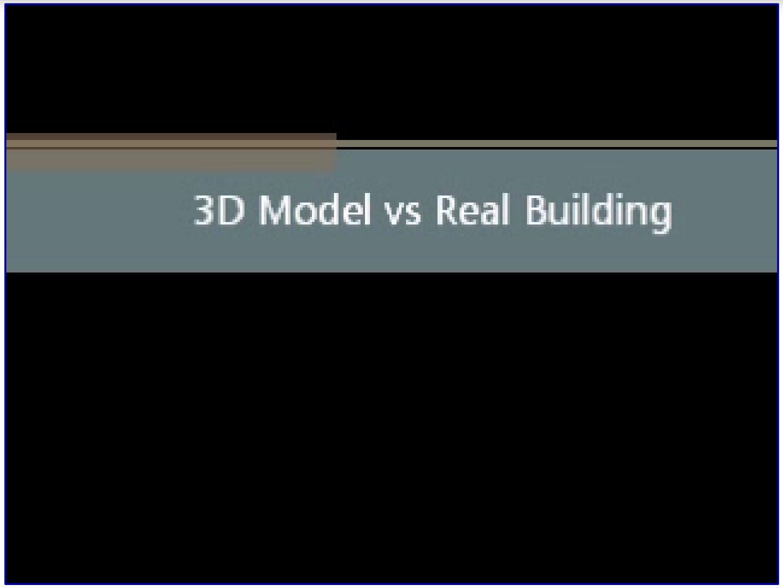


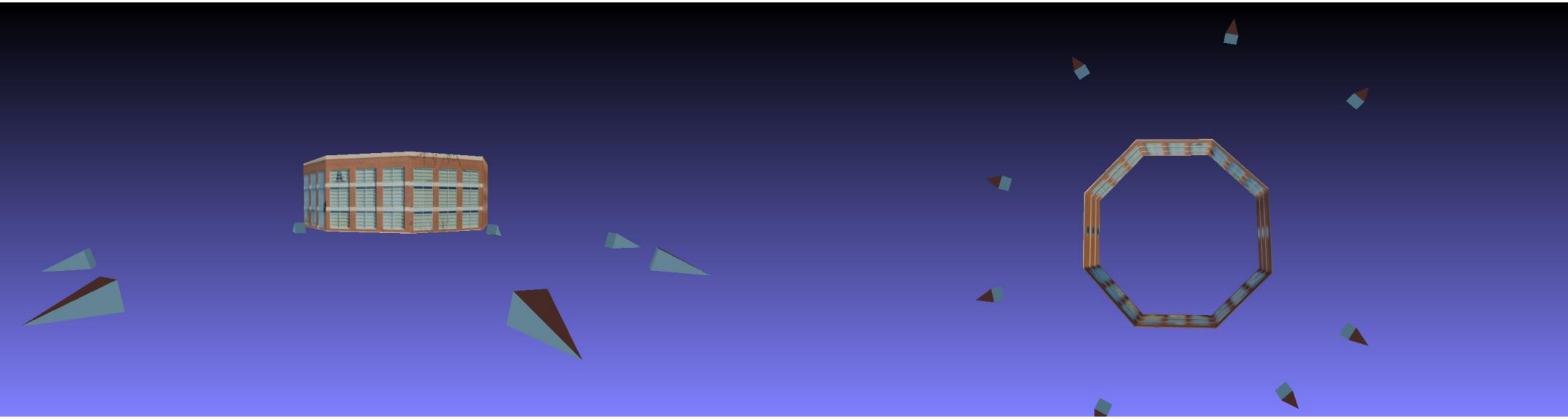
From four input images



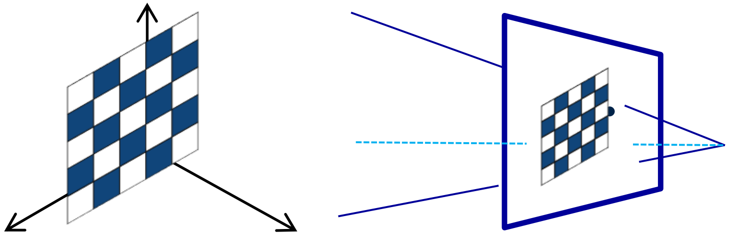
TILT: *Holistic 3D Reconstruction of Urban Scenes*

From eight input images





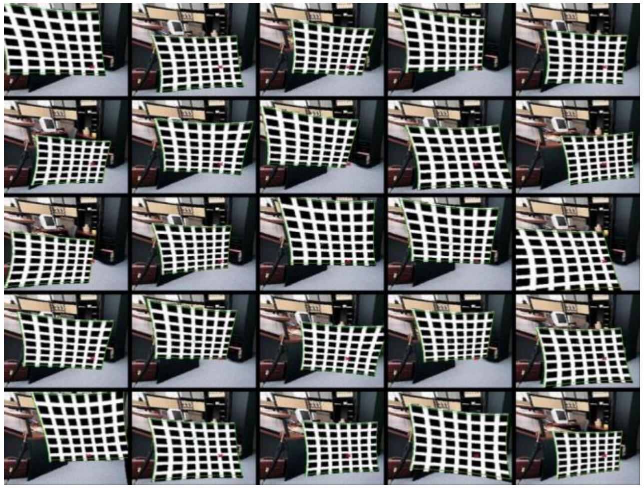
TILT: *Camera Calibration with Radial Distortion*

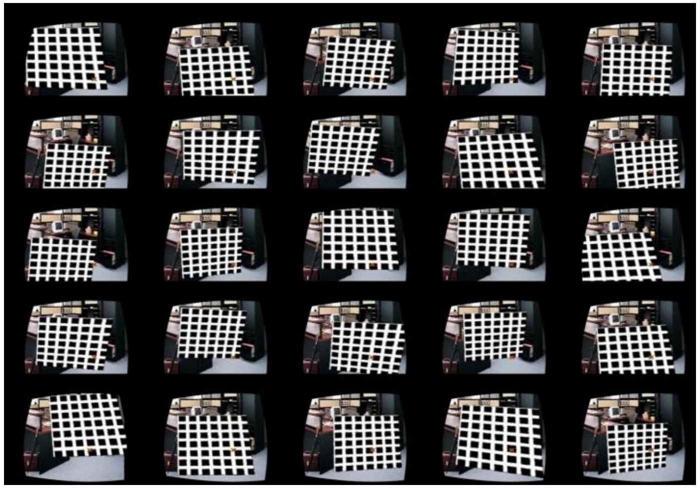


$$r = \sqrt{x_0^2 + y_0^2}, f(r) = 1 + kc(1)r^2 + kc(2)r^4 + kc(5)r^6$$

$$\binom{x}{y} = \begin{pmatrix} f(r)x_0 + 2kc(3)x_0y_0 + kc(4)(r^2 + 2x_0^2) \\ f(r)y_0 + 2kc(4)x_0y_0 + kc(3)(r^2 + 2y_0^2) \end{pmatrix}$$

$$K = \begin{bmatrix} f_x & \theta & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$





Zhang, Matsushita, and Ma, in CVPR 2011

TILT: *Camera Calibration with Radial Distortion*

$$\min \sum_{i=1}^{N} \|A_i\|_* + \lambda \|E_i\|_1 \quad \text{subj} \quad A_i + E_i = D \circ (\tau_0, \tau_i)$$

$$\tau_0 = (K, K_c), \quad \tau_i = (R_i, T_i).$$

Previous approach

Low-rank method

Zhang, Matsushita, and Ma, in CVPR 2011

Take-home Messages for Visual Data Processing:

1. (Transformed) **low-rank and sparse** structures are central to visual data modeling, processing, and analyzing;

2. Such structures can now be extracted **correctly, robustly, and efficiently**, from raw image pixels (or high-dim features);

3. These new algorithms **unleash tremendous local or global information** from multiple or single images, emulating or surpassing human perception;

4. These algorithms start to exert significant impact on **image/video processing, 3D reconstruction, and object recognition**.

… …

But try not to abuse or misuse them…

Other Data/Applications: *Rectifying 3D Object Orientation*

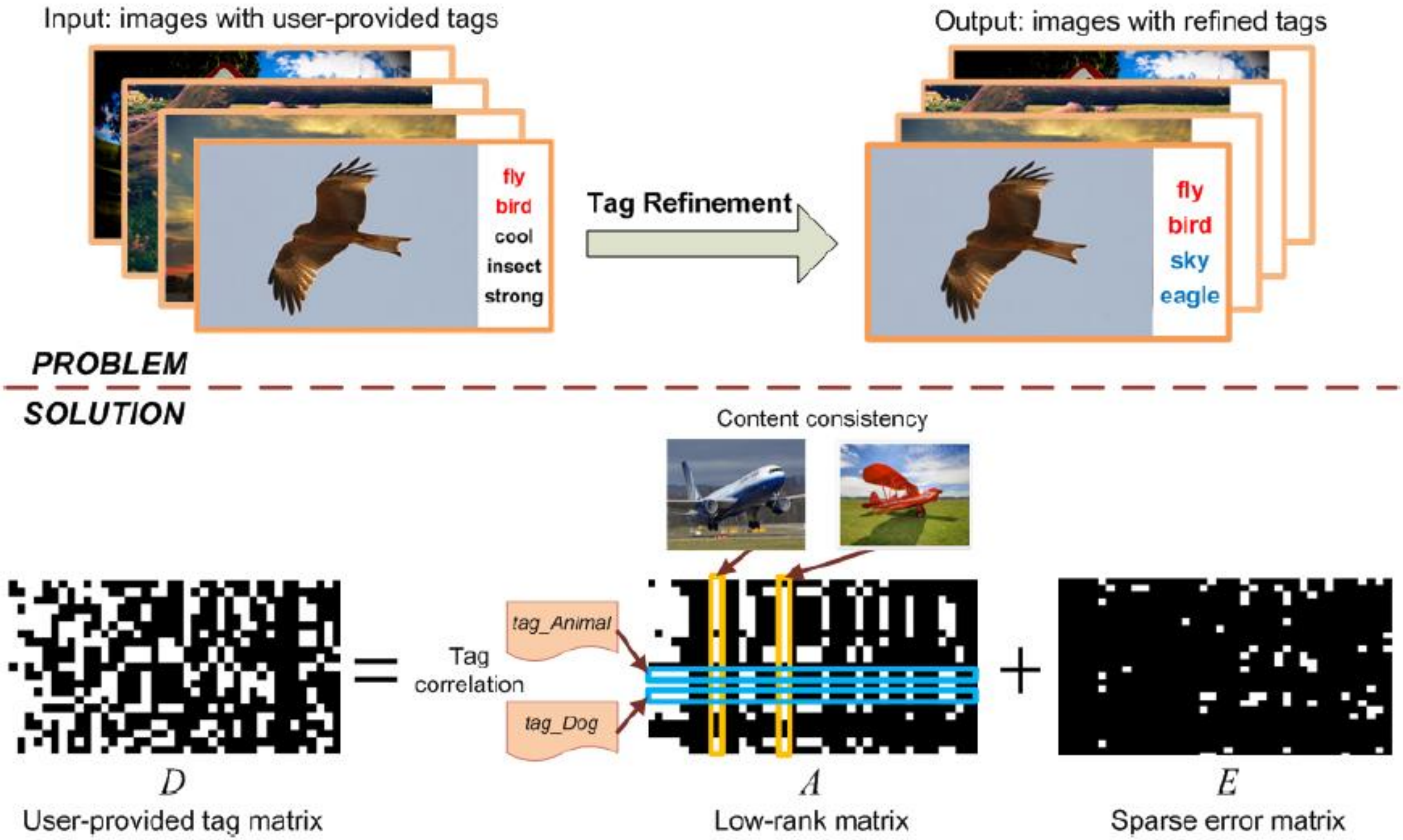
TILT for 3D: Unsupervised upright orientation of man-made 3D objects

$$\min \sum_{i=1}^{3} \|A_i\|_* + \lambda \|E_i\|_1$$

$$\text{st } D \circ \tau = [A_1, A_2, A_3] + [E_1, E_2, E_3].$$

Fig. 10. More models which have been successfully tested through our algorithm.

Other Data/Applications: Web Image/Tag Refinement



Zhu, Yan, and Ma, ACM MM 2010.

Other Data/Applications: Web Document Corpus Analysis

Latent Semantic Indexing: the classical solution (PCA)

Documents

Words

CHRYSLER SETS STOCK SPLIT, HIGHER DIVIDEND

Chrysler Corp said its board declared a three-for-two stock split in the form of a 50 pct stock dividend and raised the quarterly dividend by seven pct.

The company said the dividend was raised to 37.5 cts a share from 35 cts on a pre-split basis, equal to a 25 ct dividend on a post-split basis.

Chrysler said the stock dividend is payable April 13 to holders of record March 23 while the cash dividend is payable April 15 to holders of record March 23. It said cash will be paid in lieu of fractional shares.

With the split, Chrysler said 13.2 mln shares remain to be purchased in its stock repurchase program that began in late 1984. That program now has a target of 56.3 mln shares with the latest stock split.

Chrysler said in a statement the actions "reflect not only our outstanding performance over the past few years but also our optimism about the company's future."

D

d_{ij} word frequency (or TF/IDF)

$$D = A + Z$$
$$= U_1 \Sigma_1 V_1^T + \underline{U_2 \Sigma_2 V_2^T}$$

Dense, difficult to interpret

a better model/solution?

$$D = A + \underline{E}$$

Low-rank "background" topic model

Informative, discriminative "keywords"

Other Data/Applications: Sparse Keywords Extracted

Reuters-21578 dataset: 1,000 longest documents; 3,000 most frequent words

CHRYSLER SETS STOCK SPLIT, HIGHER DIVIDEND

Chrysler Corp said its board declared a three-for-two stock split in the form of a 50 pct stock dividend and raised the quarterly dividend by seven pct.

The company said the dividend was raised to 37.5 cts a share from 35 cts on a pre-split basis, equal to a 25 ct dividend on a post-split basis.

Chrysler said the stock dividend is payable April 13 to holders of record March 23 while the cash dividend is payable April 15 to holders of record March 23. It said cash will be paid in lieu of fractional shares.

With the split, Chrysler said 13.2 mln shares remain to be purchased in its stock repurchase program that began in late 1984. That program now has a target of 56.3 mln shares with the latest stock split.

Chrysler said in a statement the actions "reflect not only our outstanding performance over the past few years but also our optimism about the company's future."

Other Data/Applications: Protein-Gene Correlation

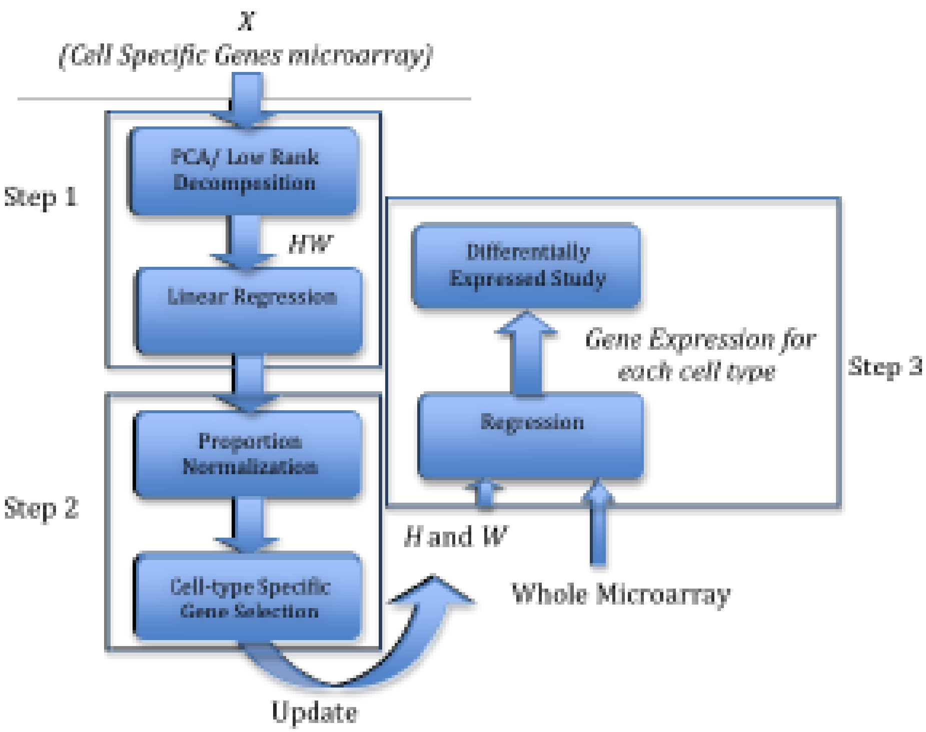


Fig. 1. The diagram of the workflow of the method presented in this paper.

Microarray data

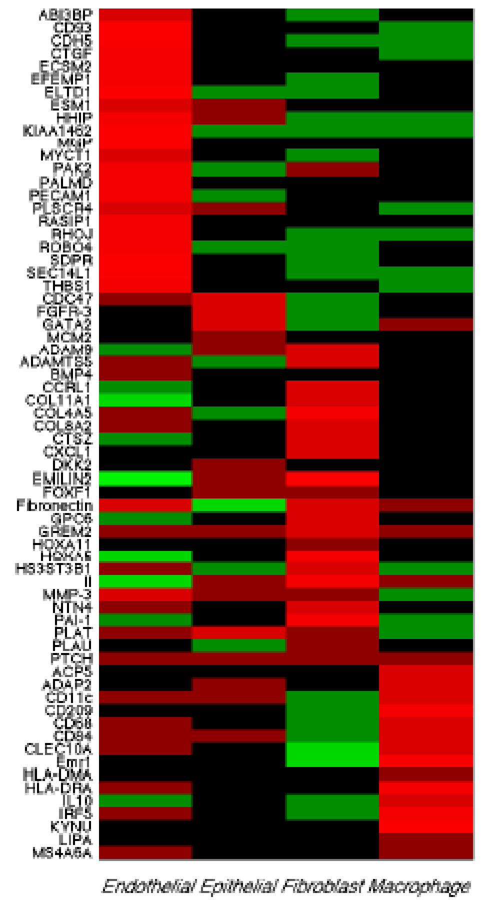
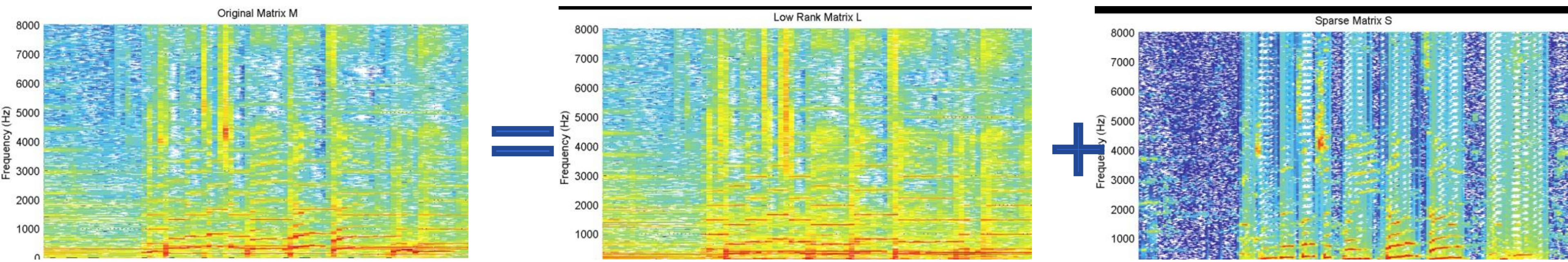


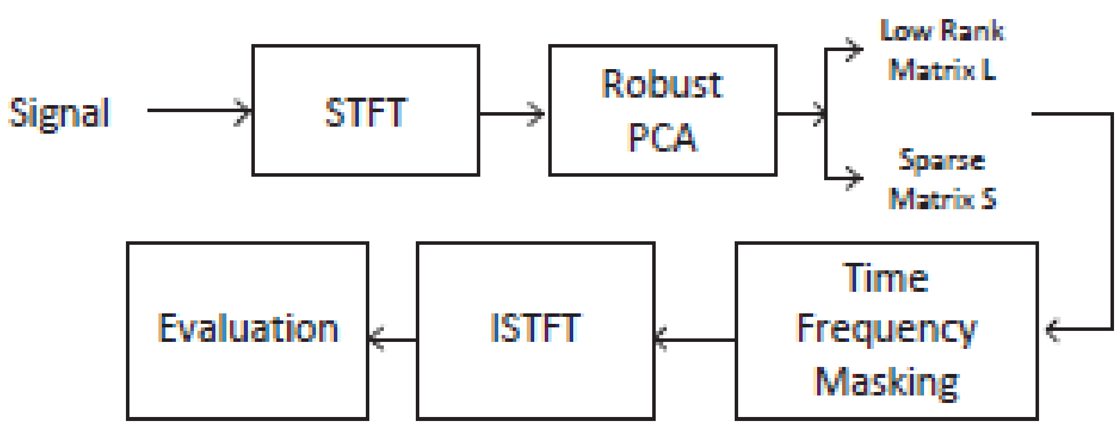
Fig. 6. HeatMap of estimated gene signatures for the sorted cell specific genes after adjustments based on fold changes. RPCA is used in the first step. It is clear that this matrix is close to a block diagonal structure.

Wang, Machiraju, and Huang, submitted to Bioinformatics 2012.

Other Data/Applications: Lyrics and Music Separation

Songs (STFT) | Low-rank (music) | Sparse (voices)





Other Data/Applications: Internet Traffic Anomalies

Network Traffic = Normal Traffic + Sparse Anomalies + Noise

$$D = L + RS + N$$

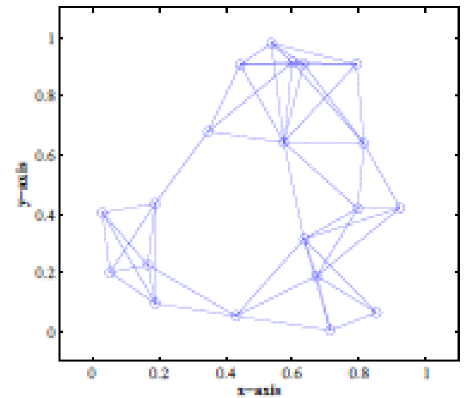
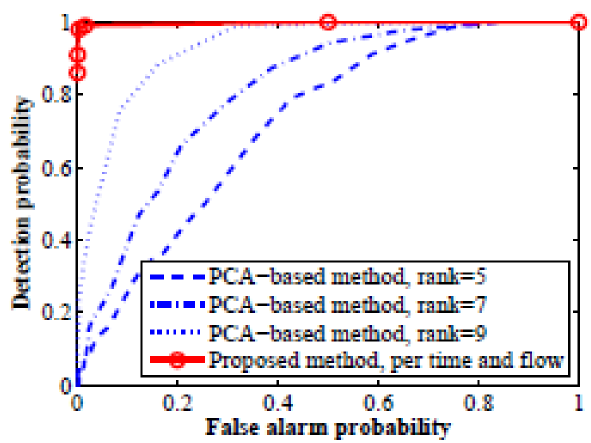
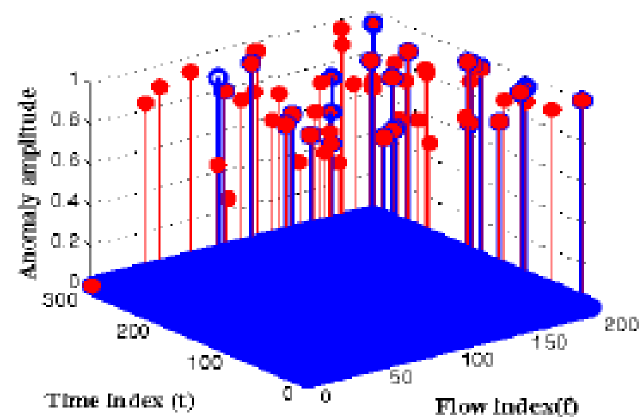


Fig. 2. Network topology graph.

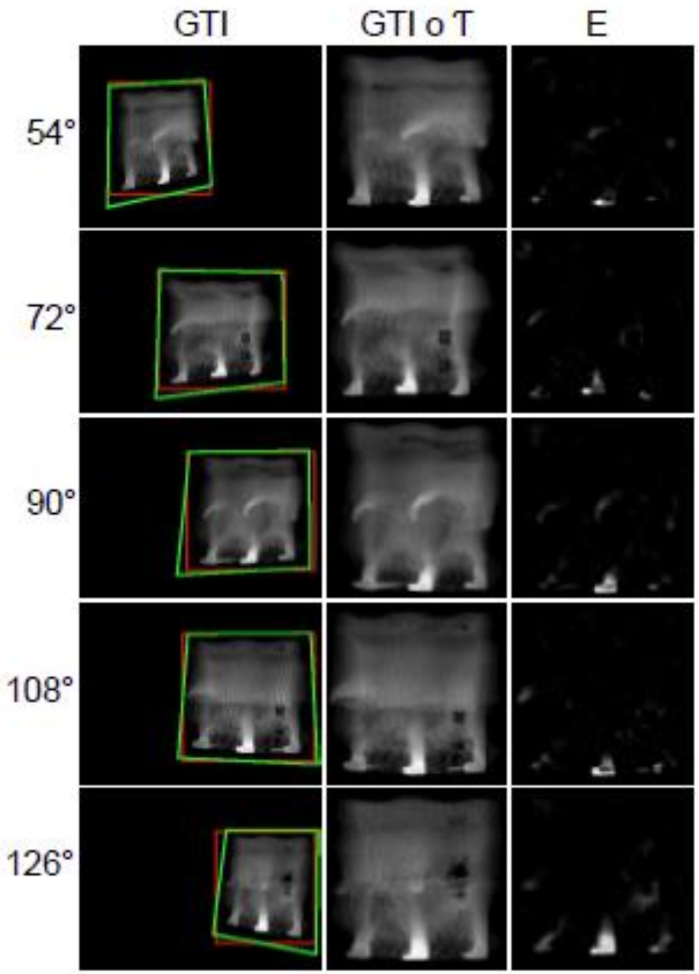




Other Data/Applications: View-Invariant Gait Recognition

Same gait from different views

Perspective distortion rectified



Other Data/Applications: Robust Filtering and System ID

GPS on a Car:

$$\begin{cases} \dot{x} &= Ax + Bu, \quad A \in \Re^{r\times r} \\ y &= Cx + z + {\color{red}e} \end{cases}$$

gross sparse errors
(due to buildings, trees…)

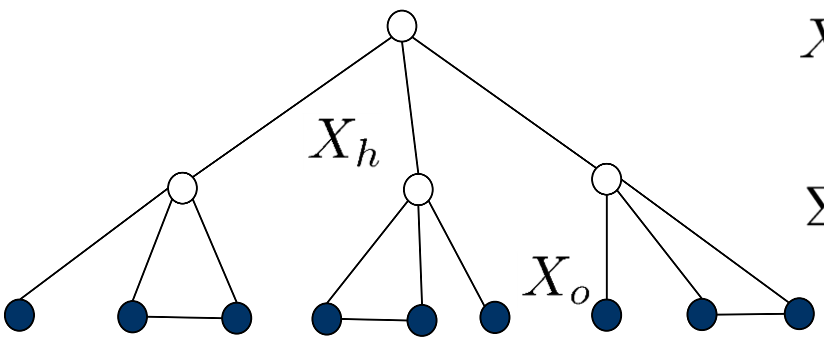
Robust Kalman Filter: $\hat{x}_{t+1} = Ax_t + K(y_t - C\hat{x}_t)$

Robust System ID:

$$\underbrace{\begin{bmatrix} y_n & y_{n-1} & y_{n-2} & \cdots & y_0 \\ y_{n-1} & y_{n-2} & \cdots & \ddots & y_{-1} \\ y_{n-2} & \cdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & y_{-n+2} \\ y_0 & y_{-1} & \cdots & y_{-n+2} & y_{-n+1} \end{bmatrix}}_{\text{Hankel matrix}} = \mathcal{O}_{n\times r} X_{r\times n} + S$$

Dynamical System Identification, Maryan Fazel, Stephen Boyd, 2000

Other Data/Applications: Learning Graphical Models



$$X = (X_o, X_h) \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{bmatrix} \Sigma_o & \Sigma_{oh} \\ \Sigma_{ho} & \Sigma_h \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} J_o & J_{oh} \\ J_{ho} & J_h \end{bmatrix}$$

$X_i, \; X_j$ cond. indep. given other variables $\quad \Leftrightarrow \quad \left(\Sigma^{-1}\right)_{ij} = 0$

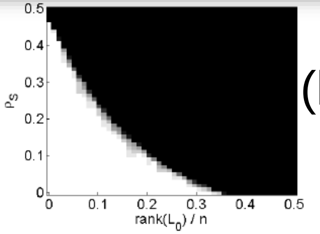
Separation Principle:

$$\begin{array}{ccccc} \Sigma_o^{-1} & = & J_o & - & J_{oh} J_h^{-1} J_{ho} \\ \text{observed} & = & \text{sparse} & + & \text{low-rank} \end{array}$$

- sparse pattern → conditional (in)dependence
- rank of second component → number of hidden variables

Chandrasekharan, Parrilo, and Wilsky, Annual of Statistics, 2012

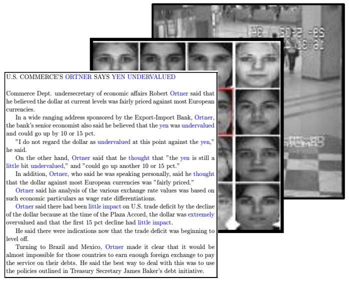
A Perfect Storm in the Cloud…



(a) Robust PCA, Random Signs

Mathematical Theory
(high-dimensional statistics, convex geometry measure concentration, combinatorics…)

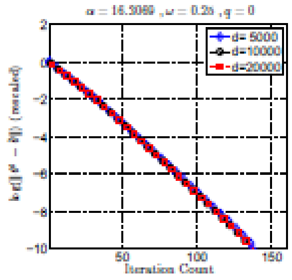
Massive Data
(images, videos, texts, audios, speeches, stocks, user rankings…)



Cloud Computing
(parallel, distributed, networked)

Applications & Services
(data processing, analysis, compression, knowledge discovery, search, recognition…)

Computational Methods
(convex optimization, first-order algorithms random sampling, approximate solutions…



REFERENCES + ACKNOWLEDGEMENT

Core References:

- *RASL: Robust Alignment by Sparse and Low-rank Decomposition*? Peng, Ganesh, Wright, Xu, and Ma, Trans. PAMI, 2012.
- *TILT: Transform Invariant Low-rank Textures,* Zhang, Liang, Ganesh, and Ma, IJCV 2012.
- *Compressive Principal Component Pursuit*, Wright, Ganesh, Min, and Ma, ISIT 2012.

More references, codes, and applications on the website:

http://perception.csl.illinois.edu/matrix-rank/home.html

Colleagues:

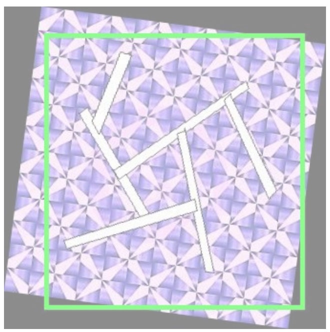
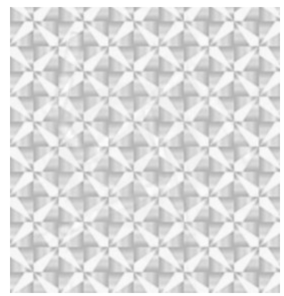
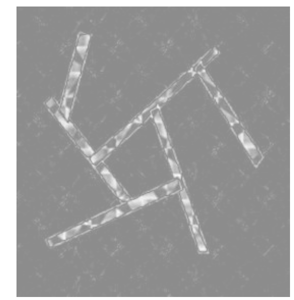
- Prof. Emmanuel Candes (Stanford)
- Prof. John Wright (Columbia)
- Prof. Zhouchen Lin (Peking University)
- Dr. Yasuyuki Matsushita (MSRA)
- Dr. Allen Yang (Berkeley)
- Dr. Arvind Ganesh (IBM Research, India)
- Prof. Shuicheng Yan (Na. Univ. Singapore)
- Prof. Jian Zhang (Sydney Tech. Univ.)
- Prof. Lei Zhang (HK Polytech Univ.)
- Prof. Liangshen Zhuang (USTC)

Students:

- Zhengdong Zhang (MSRA, Tsinghua University)
- Xiao Liang (MSRA, Tsinghua University)
- Xin Zhang (MSRA, Tsinghua University)
- Kerui Min (UIUC)
- Dr. Zhihan Zhou (UIUC)
- Dr. Hossein Mobahi (UIUC)
- Dr. Guangcan Liu (UIUC)
- Dr. Xiaodong Li (Stanford)

THANK YOU!

Questions, please?

 $\circ\,\tau =$  $+$ 

$$D \circ \tau = A + E \qquad \min \ \|A\|_* + \lambda \|E\|_1$$