Compressive sensing (CS) and sparse representations (SR) for computer vision and image processing

Rama Chellappa
University of Maryland

Compressive sensing: Is it another wave?

- Computer vision in the grip of many waves
  - MRFs (1980 – 1995), reemergence in graphics and vision
  - Regularization (From early eighties ..
  - Invariants (Mid eighties to mid nineties
  - Particle filters (From mid nineties; has reached the status of FFT in DSP!)
  - Graphical models (From early nineties)
  - Context (From early nineties…)
  - Computational sensing (Started in early nineties, died a bit later, reemerging as computational photography)
  - Manifolds

Why CVPR/PRIP folks should be interested in CS and SR

- Sensor and processing tightly coupled
  - Natural outgrowth of integrated sensing and processing, A2I efforts from DARPA/DSO since the early nineties
  - Already having impact on medical imaging, SAR imaging, face/iris recognition, new camera designs, etc
- Data deluge
  - From not enough data to too much data
  - Optimized data acquisition strategies and sparse representation will allow the full digestion of data
  - Data to action
- Hard sciences and tons of data can and do help each other!

Outline

- Compressive sensing (CS)
  - Basics
  - Restricted isometry property (RIP), Johnson Lindenstrauss (JL) lemma.
  - Recovery algorithms
- Examples
  - Compressive sensing-driven camera
  - Sparsity-induced iris recognition and cancelability using random projections.
  - Dictionary-based face recognition
  - Reconstruction from gradient fields using *l-1* optimization
  - Compressive video processing
  - CS and graphics
  - CS for SAR imaging
- Concluding remarks
  - Breakthroughs will come in integrated compressive sensing and processing
  - l-1 optimization and sparse representations suggest novel ways of sensing, reconstruction, processing and understanding of signals and images.

Collaborators

- Vishal Patel
- Jai Pillai
- Dikpal Reddy
- Aswin C. Sankaranarayanan (Rice)
- Pavan Turaga
- Rich Baraniuk (Rice) and
- Volkan Cevher (EPFL)

Motivation

- Shannon/Nyquist sampling theorem: Sampling rate must be at least twice the message signal bandwidth in order to achieve exact recovery.
- Compressed sensing is new method to capture and represent compressible signals at the rate well below Nyquist's rate.
  - Employs nonadaptive linear projections (random measurement matrix)
  - Preserves the signal structure (length of the sparse vectors is conserved)
  - Reconstructs the signal from the projections using optimization process (L1 norm)

An early CS paper for signal recovery

- S. Levy and P. K Fullagar, Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution, Geophysics, 46,

1235-1243(1981).

Classical Approach: Transform coding

PICTURE → ORTHO-GONALIZITION → SORTING → CODING

straightforward decoding

N samples (**ALL** measurements are taken)

Full set of projections is found

K largest coefficients selected
N-K coef. dumped

Only K coefficients are coded

K<<N

EXAUSTIVE SEARCH

Compressed sensing

PICTURE → CODING

Capture only significant components
Only M (K ≈ M) samples (measurements) are taken

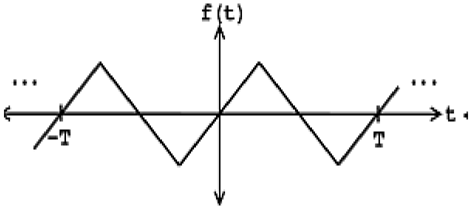
(a) Measurements must be carefully designed
(b) Original signal (picture) must be sparse

Signal Reconstruction

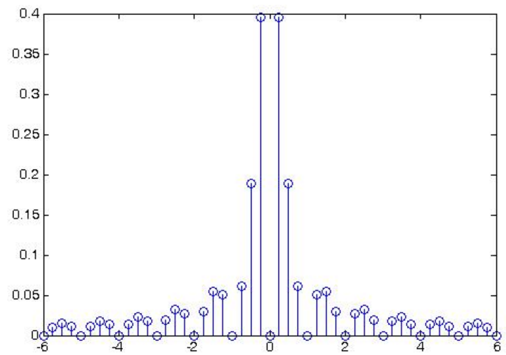
(1) Underdetermined system M<N
(2) Reconstructed signal must have N components
(3) **L1** norm is used to find sparse representation

Courtesy: Rich Baraniuk

Sparsity: Motivation

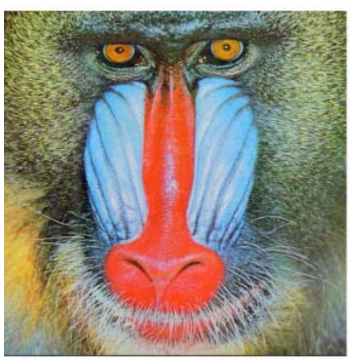


→

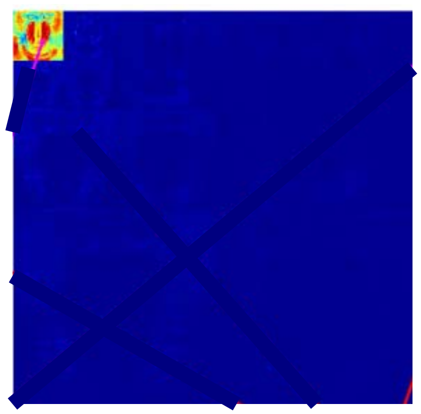


Periodic signal **Sparse in Fourier**

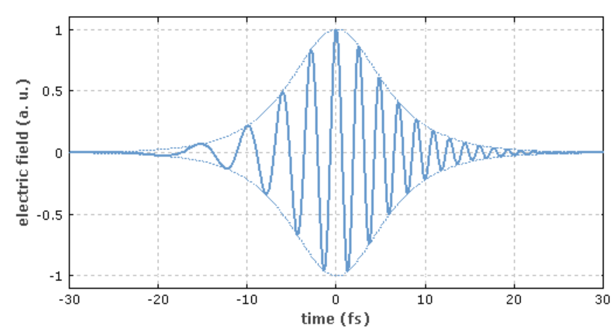
Background subtracted images



→



Images **Wavelet tree**

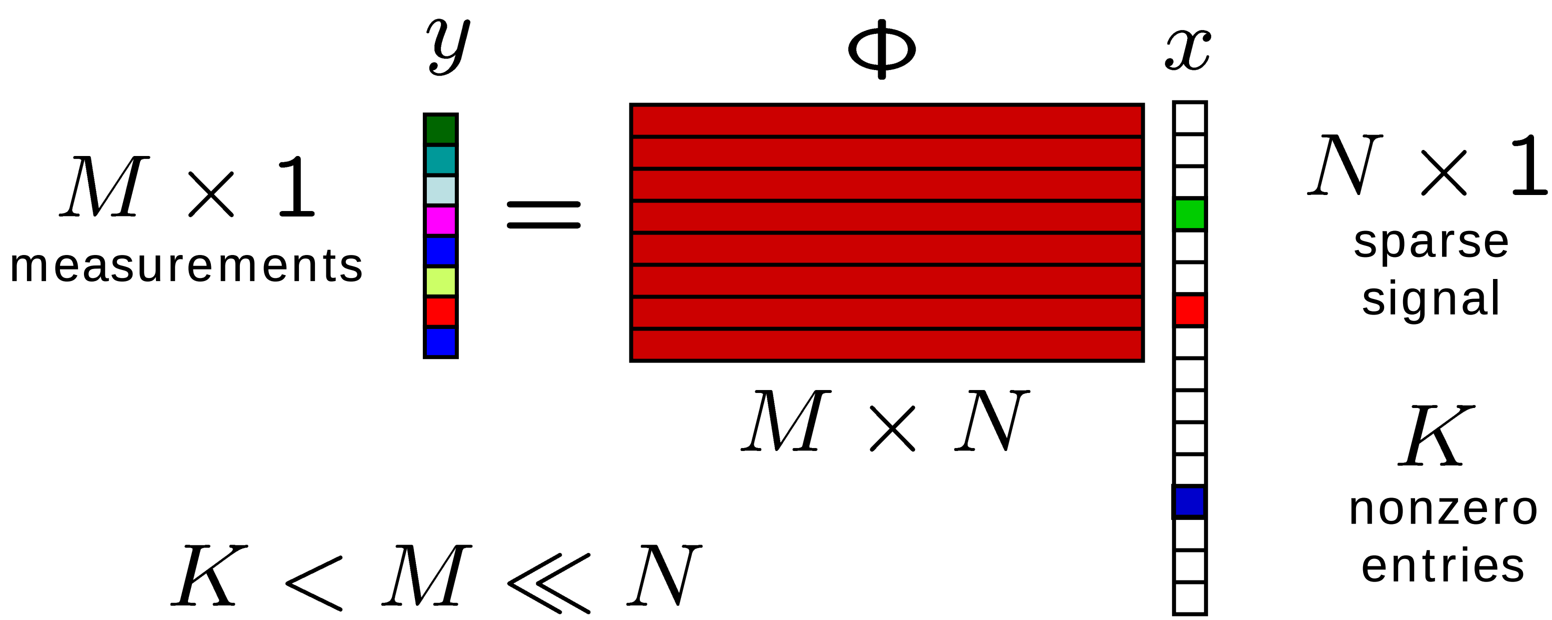


Chirp Signals → Sparse in Short Time Fourier Transform

Compressive sampling

- When data is sparse/compressible, can directly acquire a ***condensed representation*** with no/little information loss through linear ***dimensionality reduction***

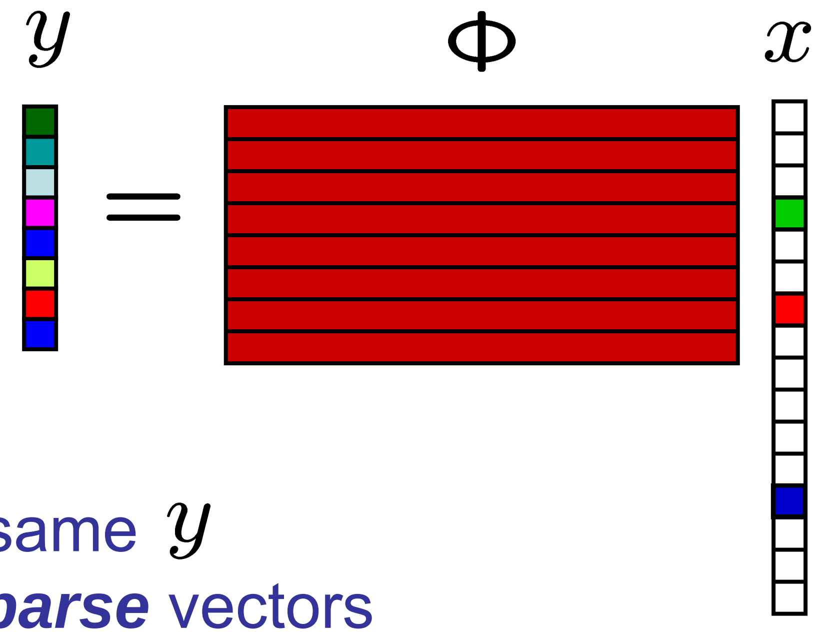
$$y = \Phi x$$



$$K < M \ll N$$

From Dr. V. Cevher

How does it work?



- Projection Φ
  not full rank…

  $M < N$

  … and so
  loses information in general
- Infinitely many x's map to the same y
- But we are only interested in ***sparse*** vectors
- **Design** Φ so that each of its *M*x*K* submatrices are full rank
  - preserve information in *K*-sparse signals
  - **Restricted Isometry Property** (RIP) of order 2*K*

From Dr. V. Cevher

Restricted Isometry Property (RIP)

A matrix Θ is said to satisfy the RIP of order K with constants $\delta_K \in (0,1)$ if

$$(1-\delta_K) \parallel v \parallel_2^2 \leq \parallel \Theta v \parallel_2^2 \leq (1+\delta_K) \parallel v \parallel_2^2$$

for any v such that $\parallel v \parallel_0 \leq K$. [Candes and Tao, 2005].

- RIP is a sufficient condition to find the sparsest solution.
- When RIP holds, Θ approximately preserves the Euclidian length of K-sparse signals.
- All subsets of K -columns taken from Θ are nearly orthogonal.
- RIP has been established for some matrices such as random Gaussian, Hadamard, and Fourier, however, in practice there is no computationally feasible way to check this property for a given matrix, as it is combinatorial in nature.

Direct construction of measurement matrix (Φ)

- Verifying RIP requires $\binom{N}{K}$ all possible combinations of K

  non-zero entries of length N
- However RIP could be achieved by simply selecting Φ randomly
  - φ_{ij} are independent and identically distributed random variables from Gaussian probability density with zero mean and 1/N variance.
  - So measurements

$$y = \Phi x$$

  are randomly weighted linear combinations of elements of signal.

For compressible signals

$$\mathrm{x}_{\substack{\text{target signal,}\\ \text{represenation of an image}\\ \text{in time or space domain}\\ \text{x is in N dimensional domain}}} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad \Psi_{\text{orthonormal basis}} = \begin{bmatrix} \psi_{11} & \cdots & \psi_{1N} \\ \vdots & & \vdots \\ \psi_{N1} & \cdots & \psi_{NN} \end{bmatrix} \quad \psi_{\mathrm{i}_{\text{othonormal vector}}} = \begin{bmatrix} \psi_{i1} \\ \vdots \\ \psi_{iN} \end{bmatrix}$$

Classical Approach: Find the signal projections on a given basis : $\mathrm{x} = \sum_{i=1}^{N} s_i \psi_i$

$$s_{i_{\text{represenation of the image in } \Psi \text{ domain}}} = \psi_i^T \mathrm{x}$$

x is K-sparce if only K basis vecotors ψ_i have $s_i \neq 0$ (thoeretical)

x is compressible if it has just a few large s_i coeficients and many small (practical)

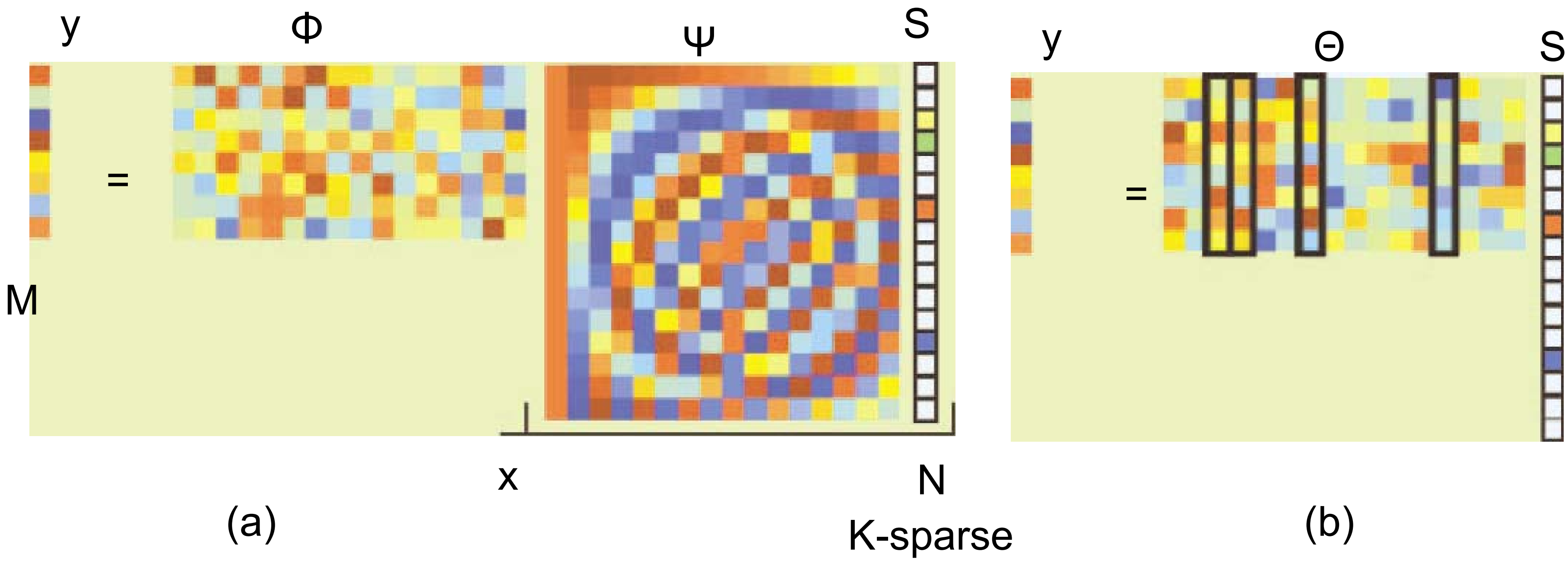
Compressable signals are well represented by K-sparse representations.

Compressed sensing: Measure only significant components (M ≪ N)

$$\mathrm{y}_{\substack{\text{measurements}\\ \text{y is in M dimensional domain}}} = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} \quad \Phi_{\text{measurement matrix}} = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1N} \\ \vdots & & \vdots \\ \phi_{M1} & \cdots & \phi_{MN} \end{bmatrix}$$

$$\mathrm{y} = \Phi \mathrm{x} = \Phi \Psi s \Rightarrow \text{ Goal } \mathrm{y} = \Theta s$$

Courtesy: Rich Baraniuk



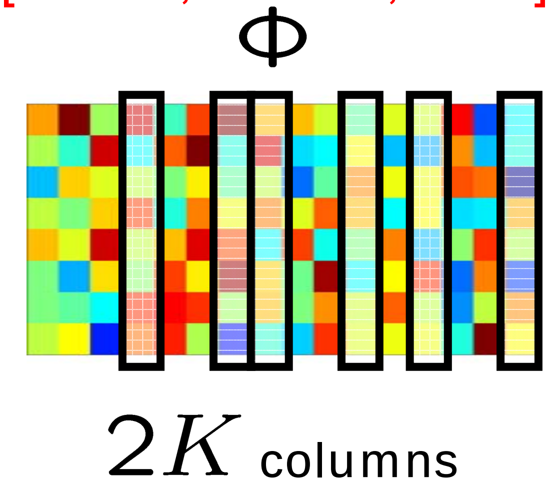
(a) Compressive sensing measurement process with a random Gaussian measurement matrix and discrete cosine transform (DCT) matrix . The vector of coefficients s **is sparse with** *K* = 4**.**

Φ (phi, measurement matrix) *Ψ (psi, orthonormal basis)*
Θ (theta, Compressed Sensing reconstruction matrix)

(b) Measurement process with $\Theta = \Phi\Psi$ **There are four columns that correspond to nonzero** *si* **coefficients; the measurement vector** y **is a linear combination of these columns.**

From IEEE Signal processing magazine July 2007, R. G. Baraniuk, Compressive sensing

Insight from the 80's [Kashin, Gluskin, 1984]



- Draw Φ at **random**
  - iid Gaussian
  - iid Bernoulli ± 1

- Then Φ has the RIP with high probability as long as

$$M = O(K \log(N/K)) \ll N$$

  - *M*x2*K* submatrices are full rank

From Dr. V. Cevher

Compressive sensing solution

- Measurement matrix must be stable and produce reconstruction of original signal, x, (length N) from M measurements
- 1) Ψ othonormal basis must be selected
- 2) Stable Φ (measurement matrix) must be designed (CS measures PROJECTIONS of signal onto a basis) and
- 3) the signal x has to be reconstructed from the underdetermined system using optimization (norm L1) linear combinations, since the projections(linear combinations are measured) we are extracting the sparse components using optimization algorithm (L1) with restrictions

Courtesy: Rich Baraniuk

Designing a signal reconstruction algorithm

- *Signal reconstruction algorithm aims to find signal's sparse coefficient vector.*
  - *<u>L2 norm</u> (energy) minimized. Pseudo inverse is closed form to find the solution but it DOES NOT find sparse solution.*
  - *<u>L0 norm</u> counts the number of non-zero elements of s. This optimization can recover K sparse signal exactly with high probability using only M=K+1 measurements but solving it is unstable and NP-complete requiring exhaustive enumeration of all (N K) possible locations of the non-zero entries in s.*
  - *<u>L1 norm</u> (adding absolute values of all elements) can* ***exactly recover K sparse signals*** *and closely approximate compressible signals with high probability using only* $M \geq cK\log(N/K)$ iid Gaussian measurements.
    - *Convex optimization reducing to* ***linear*** *programming known as Basis pursuit with the computation complexity about* $O(N^3)$

CS signal recovery

- Recovery: given $y = \Phi x$
  (ill-posed inverse problem) find x
  (sparse)

- ℓ_2 fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$
- ℓ_0 correct, slow $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$
- ℓ_1 **correct, efficient**
  mild oversampling $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_1$
  [Candes, Romberg, Tao; Donoho] ***linear program***

number of measurements required

$$M = O(K \log(N/K)) \ll N$$

From Dr. Cevher

ℓ_0 / ℓ_1 equivalence

- D. L. Donoho, “High-Dimensional centrally symmetric polytopes with neighborliness proportional to dimension,” *Discrete and Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.
- Let $p = M/N$ be a measure of undersampling factor, and $q = K/M$ be a measure of sparsity. A plot of the pairing of the variables p and q describes a two-dimensional phase space’
- p and q ranged through 40 equi-spaced points in the interval [0*;* 1] and N = 800*: a*t each point on the grid, we recorded the mean number of coordinates at which original and reconstruction differed by more than 0.001*;* averaged over 20 independent realizations (see the paper by Donoho.

CS recovery algorithms

- Convex optimization
  - noise-free signals
    - Linear programming (Basis pursuit) [Chen et al, 2001, Donoho 2006]
    - Bregman iteration, [Osher, et al, 2008]
  - noisy signals
    - Basis Pursuit De-Noising (BPDN)
    - Second-Order Cone Programming (SOCP)
- Iterative greedy algorithms
  - Matching Pursuit (MP) [Mallat and Zhang, 1993]
  - Orthogonal Matching Pursuit (OMP) [Pati et al, 1993, Mallat et al, 1994 and Tropp and Gilbert, 2007]

software @
dsp.rice.edu/cs

From Dr. Cevher

BPDN

- Basis Pursuit Formulation (A Linear program)

$$\min \|x\|_1 \quad \text{subject to } y = \Phi x$$

- Noisy measurements

$$y = \Phi x + n$$

- Basis Pursuit De-Noising

$$\min \|x\|_1 \; s.t \|y - \Phi x\|_2 \leq \varepsilon$$

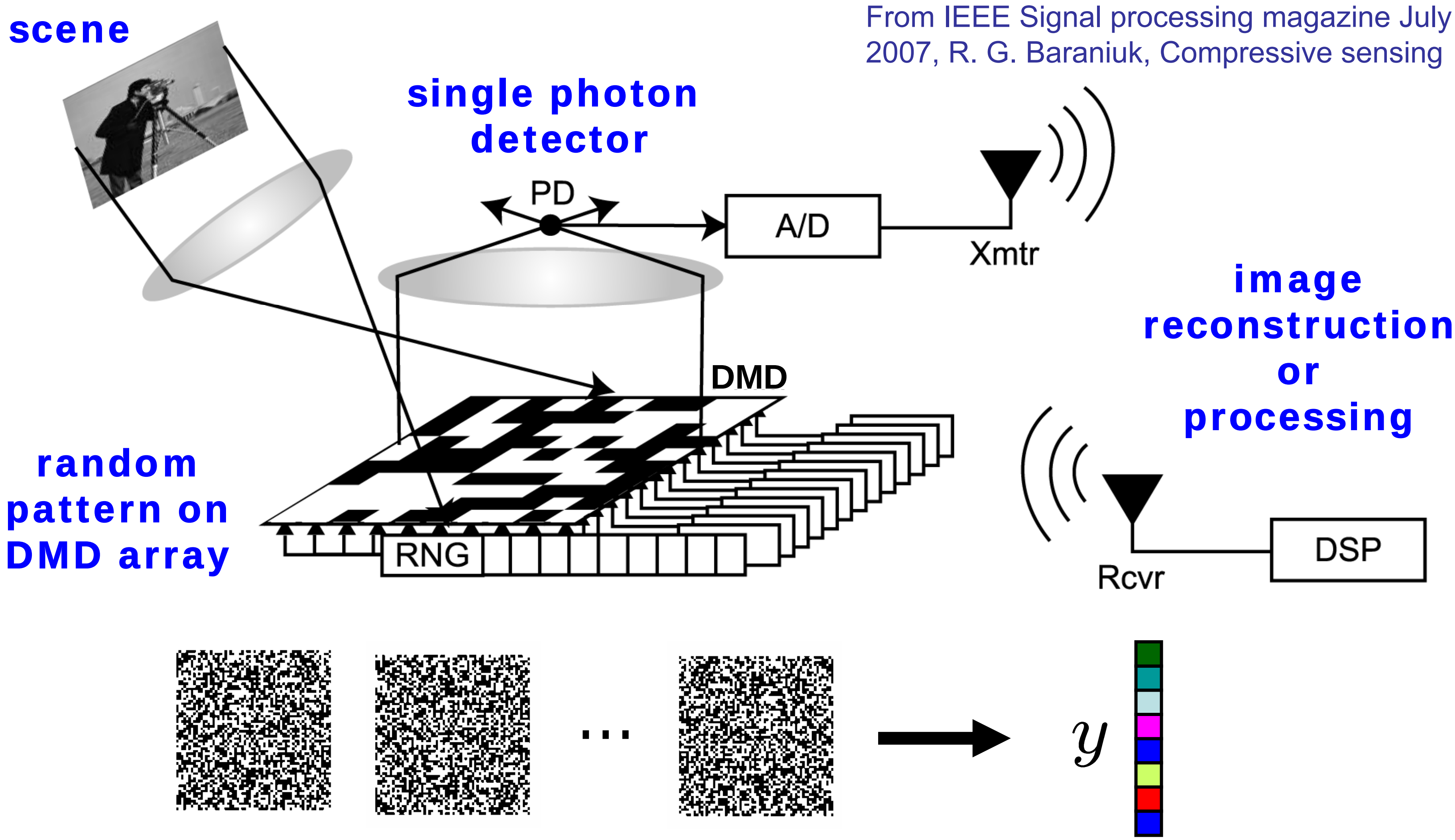
- Convex, **quadratic program**

From Dr. Cevher

Recent developments in computer vision/graphics

- Single pixel camera [Duarte, et al, 2008]
- Compressive sensing for coded aperture imaging [Veeraraghavan and Reddy, PAMI 2010]
- Sparse representations for face recognition [Wright et al, PAMI 2009, CVPR 2009]
- Robust and secure iris recognition (BTAS 2009, ICASSP2010,PAMI 2011]
- Shape from gradients (CVPR 2009)
- Image reconstruction from gradients
- Compressive Sensing of Reflectance Fields
  - Peers et al, ACM Transactions on Graphics, 2009)
- Compressive sensing for background subtraction [Cevher et al, ECCV 2008]
- Compressive SAR imaging, ICIP 2009, IEEE Jl. SASP, 04/10

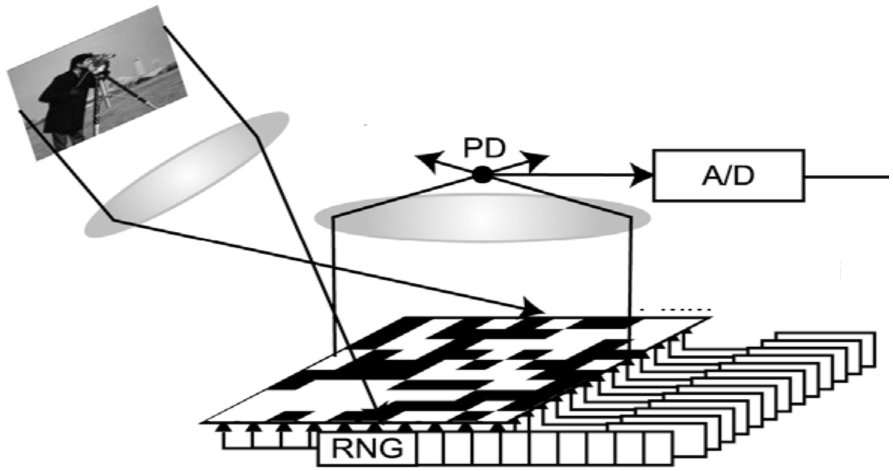
"Single-pixel" CS camera



- Flip mirror array M times to acquire M measurements
- Sparsity-based (linear programming) recovery

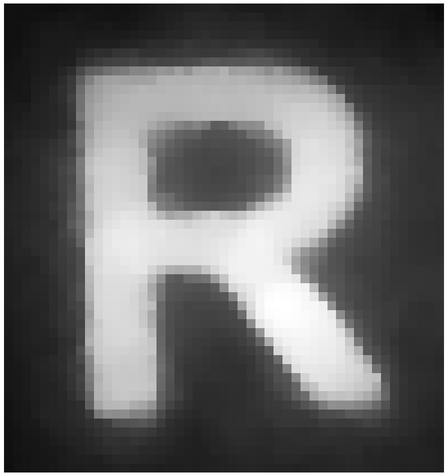
First image acquisition

Takhar et al, 2006
Duarte et al, 2008



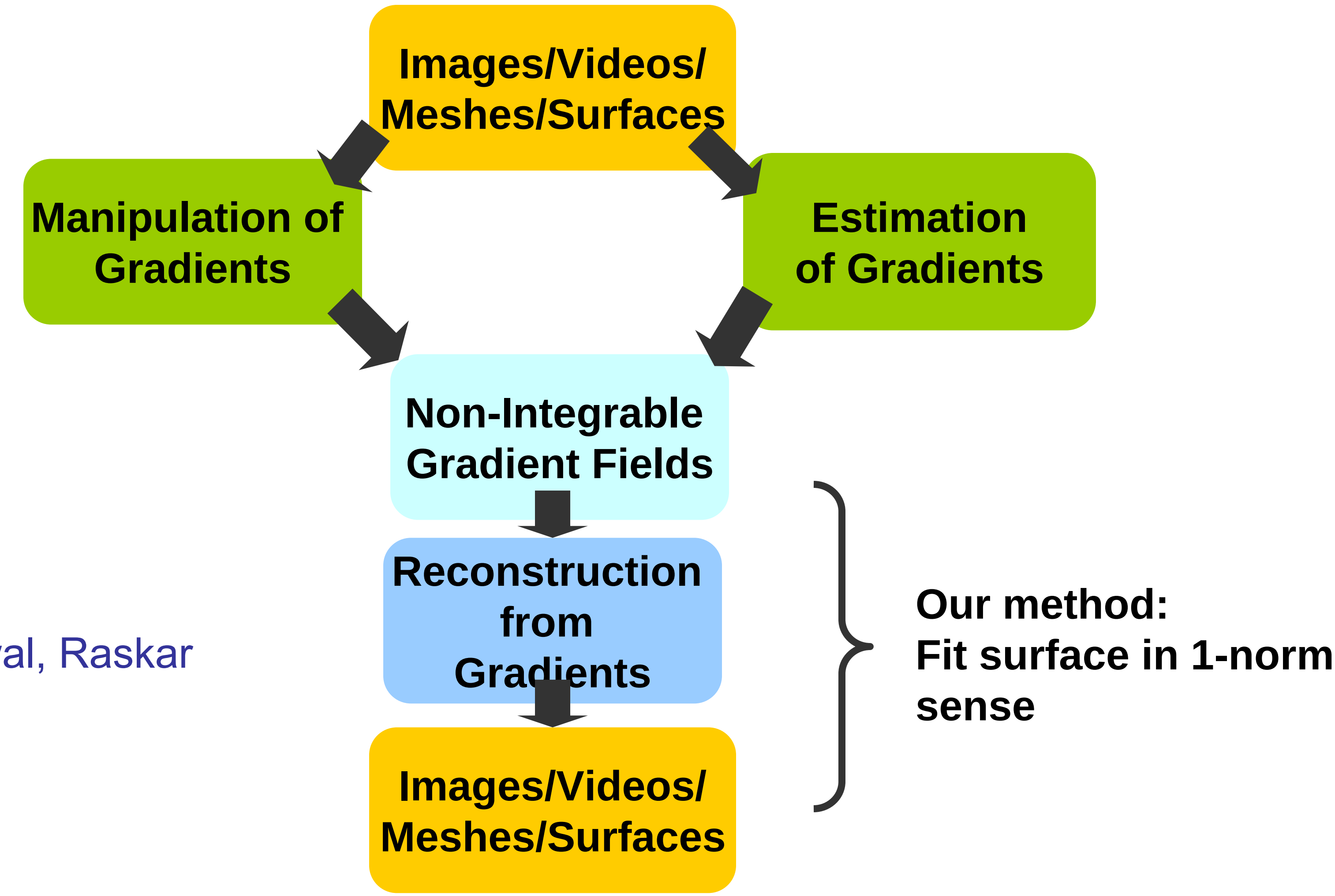
target
65536 pixels

11000 measurements
(16%)



1300 measurements
(2%)

Gradient domain processing: vision and graphics



Agrawal, Raskar

Few applications

Shape from Shading, Photometric Stereo

Height Field

High Dynamic Range (HDR) Compression

Edge suppression under significant illumination variations

Gradient fields and integrability

Image or surface: $S(x, y)$

Gradients: $\nabla S = \{\frac{\partial S}{\partial x}, \frac{\partial S}{\partial y}\}$

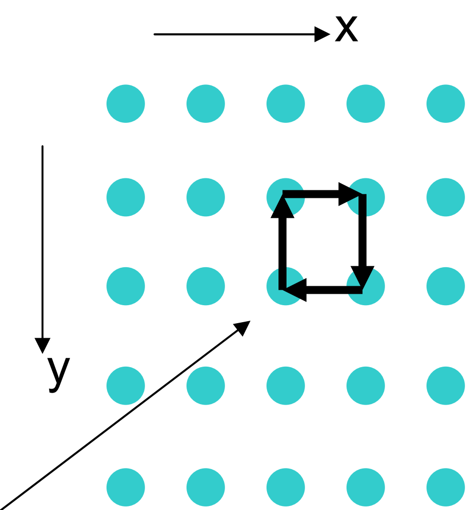
Divergence: $Div(\nabla S) = \nabla \bullet \nabla S$

Curl: $Curl(\nabla S) = \nabla \times \nabla S$

Integrability: Conservative vector field

For a scalar field $S(x, y)$

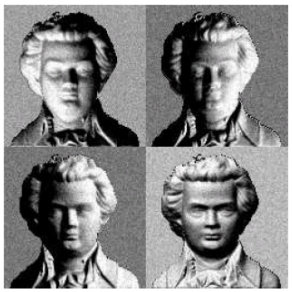
$$\boxed{\begin{gathered}\nabla \times \nabla S = 0 \\ Curl(\nabla S) = S_{yx} - S_{xy} = 0\end{gathered}}$$



Slides on this topic are from Reddy, Agrawal and Chellappa CVPR 2009

Non-integrable gradient fields

- Estimation of gradients
  - E.g. Shape from Shading, Photometric Stereo
  - Noise and outliers in estimation



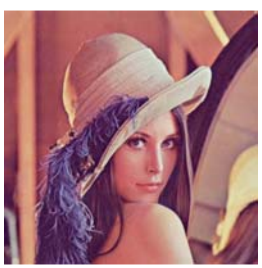
Input Images

Surface Normals/Gradients

Not-integrable

- Manipulation of integrable gradients
  - Synthesize new gradient field

Not-integrable



Image

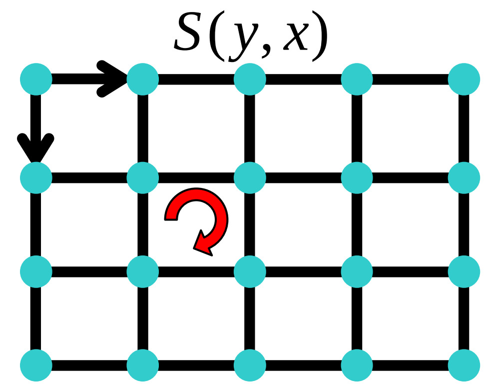
S_x

S_y

Gradient Manipulations

New Gradients

Integrability problem



Desired surface $\mathbf{s}$. Corrupted gradient field $\mathbf{g} = [\mathbf{g}_x^T \ \mathbf{g}_y^T]^T$

$$\mathbf{g} = \mathbf{D}\mathbf{s} + \mathbf{e}$$

Curl of the gradient field given by

$$\mathbf{d} = \mathbf{C}\mathbf{g} = \mathbf{C}\mathbf{e}$$

$\mathbf{d} = \mathbf{C}\mathbf{e}$ Under-determined system of equations

- Similar to error correction.
- Estimate gradient error given the curl values

Gradient operator

$$\mathbf{D} = \left(\begin{array}{cccccc} -1 & & 1 & & & \\ & -1 & & 1 & & \\ & & & & -1 & 1 \\ \hdashline -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & -1 & 1 & & \end{array}\right)$$

Poisson equation

$$\text{(P2)} \qquad \hat{\mathbf{e}} = \arg\min \|\mathbf{e}\|_2 \ \text{ s.t. } \mathbf{d} = \mathbf{C}\mathbf{e}$$

Curl operator

$$\mathbf{C} = \left(\begin{array}{cccccccccc} -1 & 1 & & & & 1 & & -1 & & \\ & -1 & 1 & & & & 1 & & -1 & \\ & & & & & & & & & \\ & & & -1 & 1 & & & & 1 & -1 \end{array}\right)$$

Integrability problem

Sparse recovery:

Sparse errors (arbitrarily large) can be recovered using

(P0) $$\hat{\mathbf{e}} = \arg\min \|\mathbf{e}\|_0 \text{ s.t. } \mathbf{d} = \mathbf{C}\mathbf{e}$$

(P0) NP-hard. So solve a convex formulation

(P1) $$\hat{\mathbf{e}} = \arg\min \|\mathbf{e}\|_1 \text{ s.t. } \mathbf{d} = \mathbf{C}\mathbf{e}$$

- (P1) not only **corrects outliers, corrects noise** as well as (P2)
- Has **local error confinement property**
- Motivated by recent results on equivalence of (P0) and (P1) in compressed sensing literature

CS and reconstruction from gradient fields

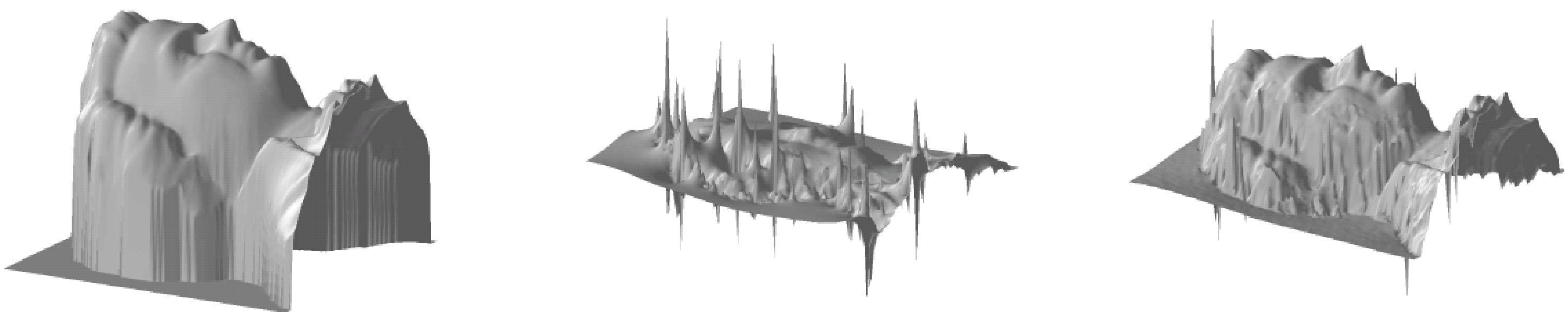
- Using CS we can analyze when recovery is guaranteed.
- For non contiguous errors C has RIP -1.
- How many outliers can $l1$ minimization fully correct?
- How should they be distributed?
- If large number of outliers then what outliers does $l1$ find and correct?

RIP -1: Berinde, et al, 2008

l_1 - minimization

$$\hat{e} = \arg\min \|e\|_1 \text{ s.t. } d = Ce$$

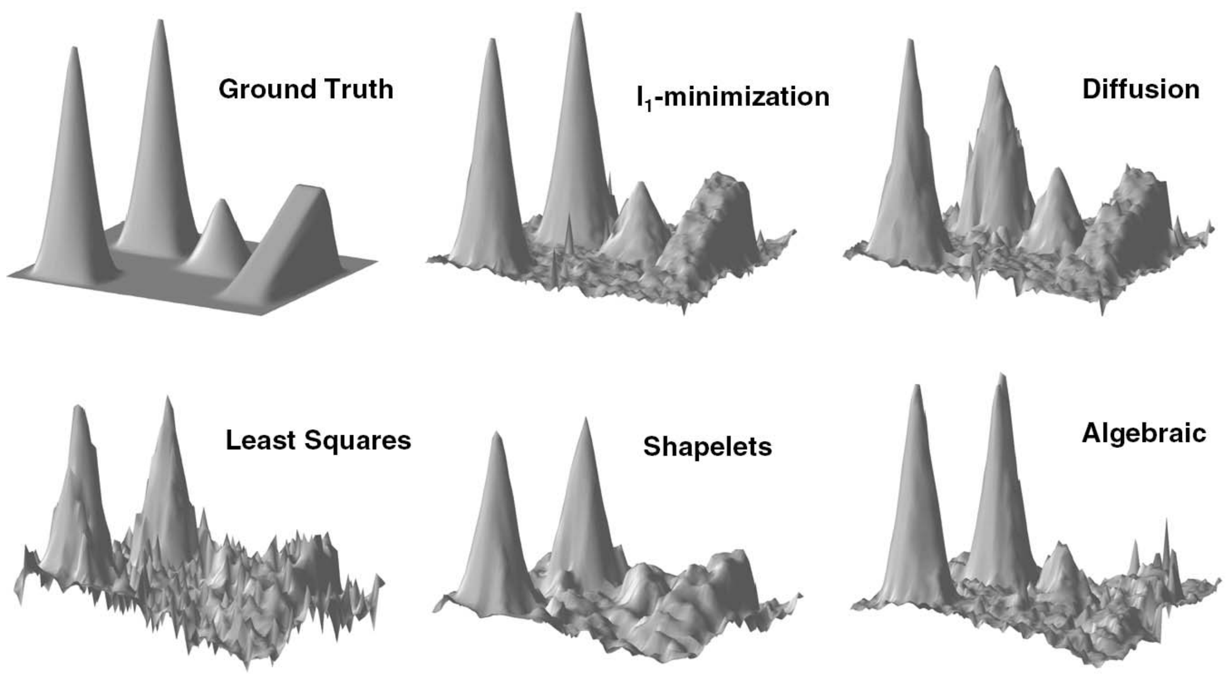
- Corrects outliers. Performs well in noise too.
- Good error confinement property locally

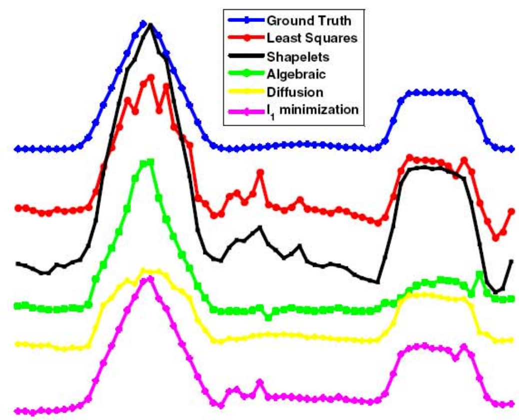


Ground truth Least squares 1-norm fit

l_1 minimization

Under noise and outliers

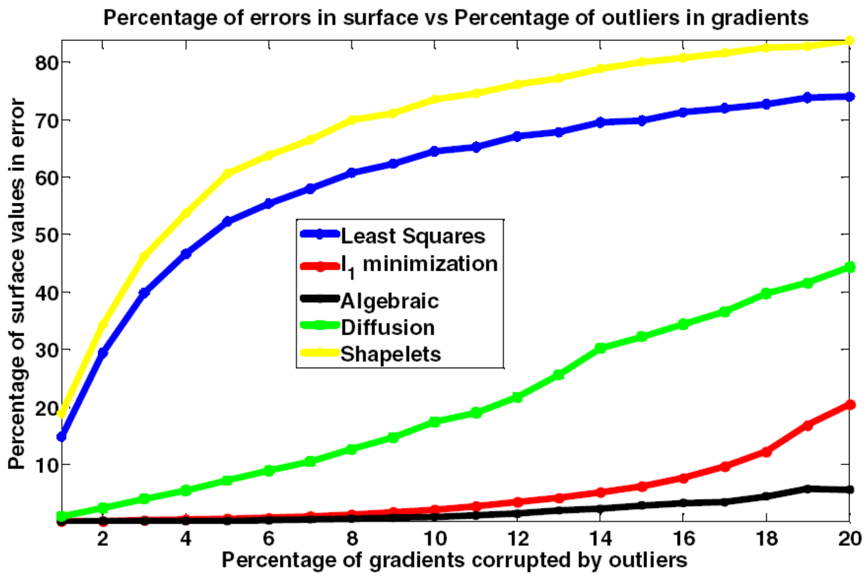




Profile through the middle

l_1 minimization

Comparison



Local error confinement

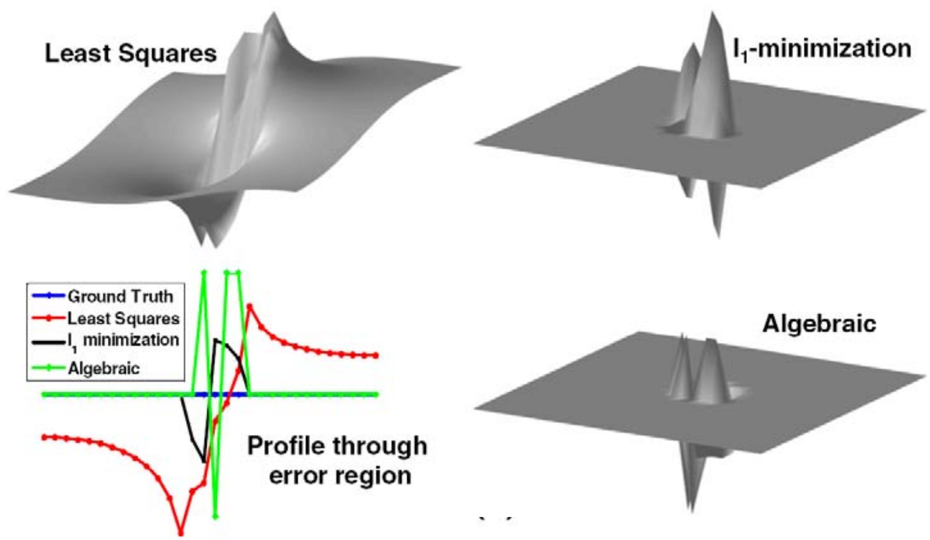


Image recovery: problem setup

An image x is said to be K-sparse in gradient if $\| TV(x) \|_0 = K$

where $TV(x)_{n,m} = \sqrt{p(n,m)^2 + q(n,m)^2}$

with $p(n,m) = x(n,m+1) - x(n,m)$

and $q(n,m) = x(n+1,m) - x(n,m)$

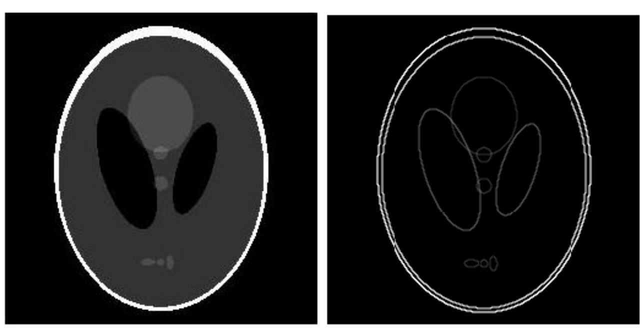
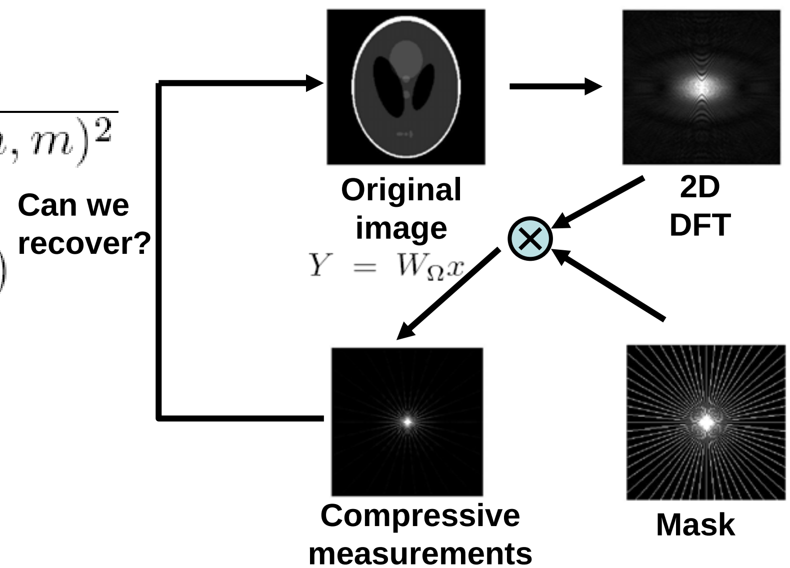


Figure 1: Shepp-Logan Phantom and its gradient



Basis pursuit (BP) solution

- Find the sparsest solution x that satisfies Fourier constraint:

$$x_{rec} = \arg\min_{x' \in \mathbb{C}^N} \| x \|_0 \quad \text{s. t.} \quad Y = W_\Omega x'$$

- Relax combinatorial problem into a convex optimization one:

$$x_{rec} = \arg\min_{x' \in \mathbb{C}^N} \| x \|_1 \quad \text{s. t.} \quad Y = W_\Omega x'$$

- Theoretical guarantees: RIP, incoherence, …

Total variation (TV) minimization

- In the spirit of BP, find the sparsest image in the total variation sense that satisfies our Fourier constraint
- In other words, solve (Candes, Romberg, Tao, 2006):

$$Y = W_\Omega x$$

- How can we use a greedy pursuit algorithm such as MP and OMP to recover sparse gradient images? (Maleh, Gilbert, Stauss, 2007)

$$x_{rec} = \arg \min_{x' \in \mathbb{C}^N} \| TV(x') \|_1 \quad \text{s. t.} \quad W_\Omega x' = W_\Omega x$$

An observation

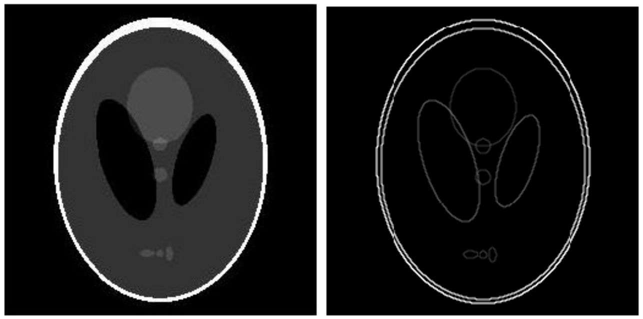
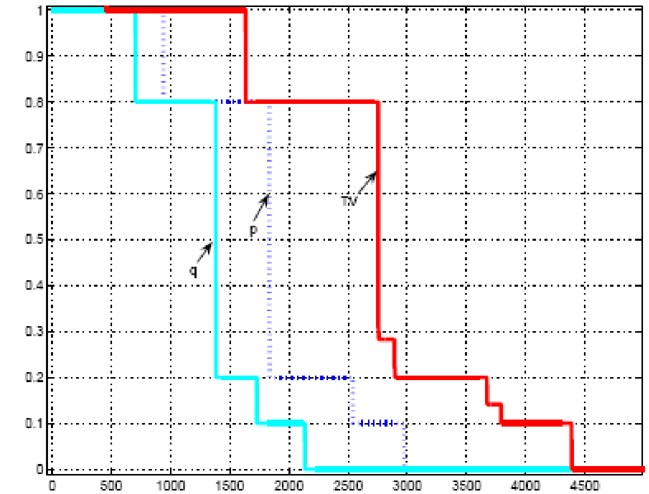


Figure 1: Shepp-Logan Phantom and its gradient



- Instead of reconstructing by the TV minimization, recover the image by separately reconstructing the gradients and then solving for the image.
- Requires a far fewer number of measurements than the TV minimization!

Least squares integration method (A. C. Gilbert

Generate a reconstruction $\tilde{X}$ of X given approximations $\widetilde{\partial_x X}$ and $\widetilde{\partial_y X}$ of $\partial_x X$ and $\partial_y X$, respectively. Here, $\partial_x X$ and $\partial_y X$ are the gradients in the horizontal and vertical directions, respectively.

$$\begin{aligned} \tilde{X} &= \operatorname*{argmin}_{Y} \left\| \partial_x Y - \widetilde{\partial_x X} \right\|_2^2 + \left\| \partial_y Y - \widetilde{\partial_y X} \right\|_2^2 \\ &+ \beta \left\| \partial_x Y \right\|_2^2 \beta \left\| \partial_y Y \right\|_2^2 + \lambda \left\| \mathcal{F}_\Omega Y - \mathcal{F}_\Omega X \right\|_2^2 \end{aligned}$$

In the Fourier domain

$$\begin{aligned} \hat{\hat{X}} &= \operatorname*{argmin}_{\hat{Y}} \left\| \left(1 - e^{-2\pi i \omega_1 / N}\right) \hat{Y} - \widehat{\widetilde{\partial_x X}} \right\|_2^2 + \left\| \left(1 - e^{-2\pi i \omega_2 / N}\right) \hat{Y} - \widehat{\widetilde{\partial_y X}} \right\|_2^2 \\ &+ \beta \left(\left\| \left(1 - e^{-2\pi i \omega_1 / N}\right) \hat{Y} \right\|_2^2 + \left\| \left(1 - e^{-2\pi i \omega_2 / N}\right) \hat{Y} \right\|_2^2 \right) \\ &+ \lambda \left\| \left(\hat{Y} - \hat{X} \right) 1_\Omega \right\|_2^2 . \end{aligned} \tag{3.9}$$

Proposition III.1. *The least squares problem (3.9) can be solved element-wise by the following formula:*

$$\hat{\hat{X}}_{\omega_1,\omega_2} = \frac{\left(1 - e^{2\pi i \omega_1 / N}\right) \widehat{\widetilde{\partial_x X}}_{\omega_1,\omega_2} + \left(1 - e^{2\pi i \omega_2 / N}\right) \widehat{\widetilde{\partial_y X}}_{\omega_1,\omega_2} + \lambda \hat{X}_{\omega_1,\omega_2} 1_\Omega}{(1+\beta)\left(\left|1 - e^{-2\pi i \omega_1 / N}\right|^2 + \left|1 - e^{-2\pi i \omega_2 / N}\right|^2 \right) + \lambda 1_\Omega} . \tag{3.10}$$

Similar to enforcing integrability using Fourier Basis while reconstructing depth form gradients. Frankot and Chellappa, IEEE Trans. PAMI, 1988)

Sparse gradient-based image reconstruction

$$p(n,m) = x(n,m+1) - x(n,m)$$

$$q(n,m) = x(n+1,m) - x(n,m)$$

$$Y = W_\Omega x$$

$$P = (e^{\frac{j2\pi k_2}{l}} - 1)Y$$

$$Q = (e^{\frac{j2\pi k_1}{l}} - 1)Y$$

$$P = W_\Omega p$$

$$Q = W_\Omega q$$

$$p_{rec} = \arg\min_{p'} \| p' \|_1 \quad \text{s. t.} \quad \| W_\Omega p' - P \|_2 \leq \varepsilon$$

$$q_{rec} = \arg\min_{q'} \| q' \|_1 \quad \text{s. t.} \quad \| W_\Omega q' - Q \|_2 \leq \varepsilon$$

Integrate gradients

Poisson Solver

$$x_{rec}$$

Poisson solver

- The gradient field (p,q) of an image should be integrable (i.e. the integral along any closed curve should be equal to zero).
- This is often not the case when inherent noise during the estimation process contaminates (p,q).
- Have to enforce integrability.
- Frankot Chellappa PAMI 1988, Simchony et al. 1990, Agrawal et al. ICCV 2005, ECCV 2006, Reddy, et al, CVPR 2009.

Numerical examples

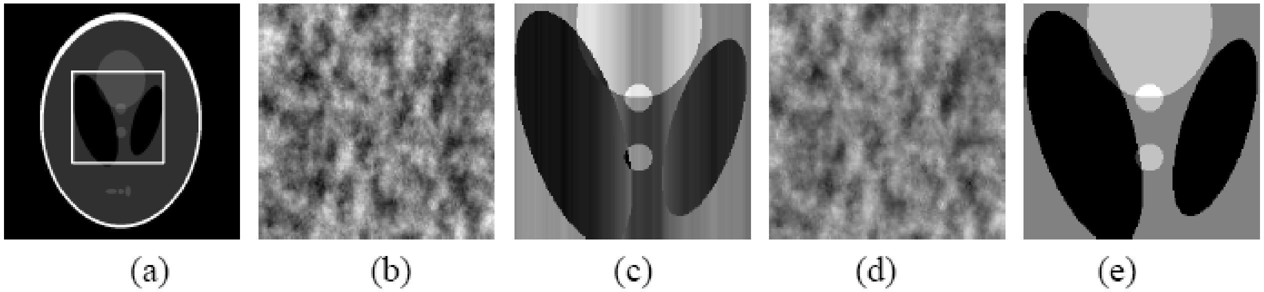


Figure 2: (a) 512×512 Shepp-Logan Phantom example. Reconstructed by (b) zero-filling, (c) GradientOMP, (d) TV method, and (e) our proposed method.

94.6% random undersampling

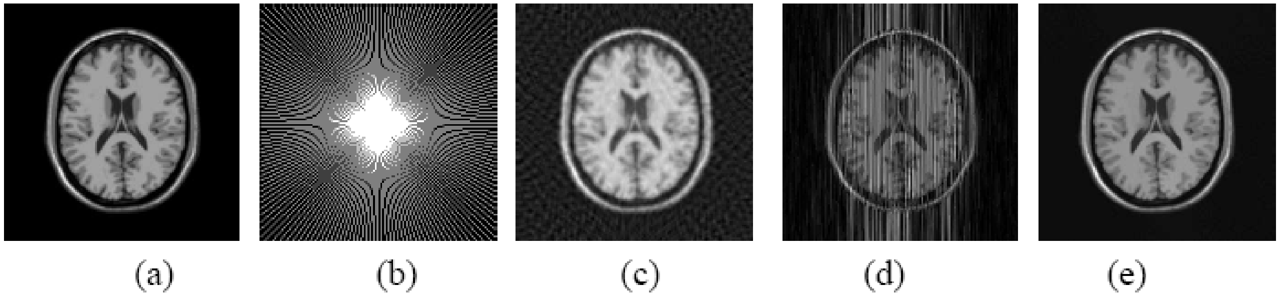


Figure 3: (a) Original MRI image. (b) Radial k-space trajectories. Reconstructed by (c) zero-filling, (d) GradientOMP, and (e) our method.

68.5% random undersampling

Experiments

(a) (b) (c) (d)

Figure 4: (a) Noisy 128×128 phantom image (SNR 20dB). (b) Sampling pattern. Reconstructed image by (c) GradientOMP and (c) by our method.

65% random undersampling

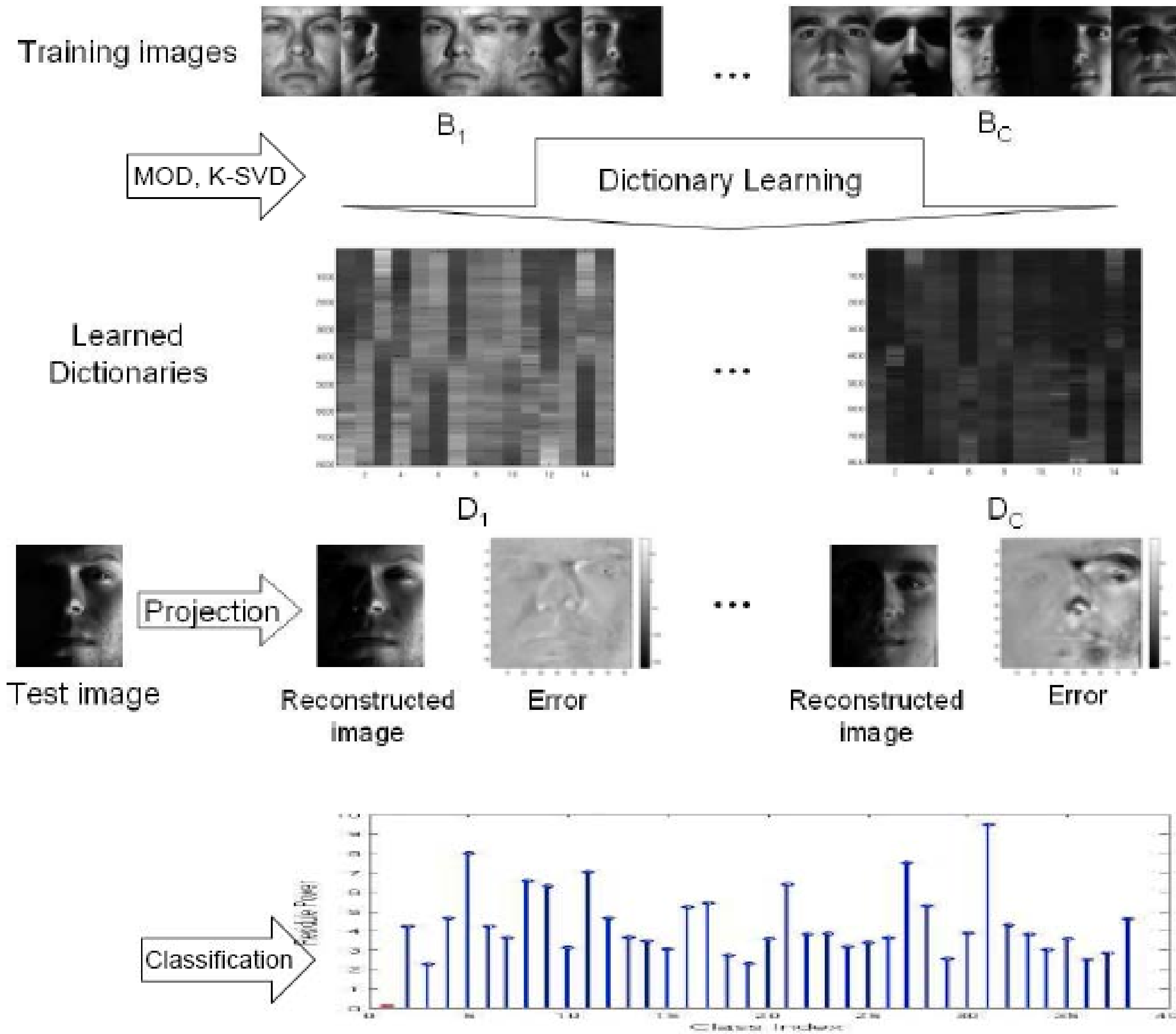
Face recognition via sparse representations

- Automatic face recognition algorithm robust to occlusion, expressions and disguise – John Wright et al, PAMI 2009.
- Represent the test face as a sparse linear combination of the training faces.
- Estimate the class of the test image from the sparse coefficients.
- Alignment is handled by finding the best transformation between the training and test images.
- Illumination variation is taken care of by including training images under varying illuminations - Wagner et al. CVPR 2009

Dictionary-based face recognition

- Can work with only one training image per class.
- Sparse representation is found by learning a dictionary for each class using KSVD.
- Uses a relighting approach to deal with the illumination problem.

Dictionary-based face recognition



Classification result among several classes

Learning dictionaries - KSVD

Objective: Find the best dictionary to represent the samples $\mathbf{B} = [\mathrm{x}_1, \cdots, \mathrm{x}_m]$ as sparse compositions, by solving the following optimization problem:

$$\arg\min_{\mathbf{D},\Gamma} \|\mathbf{B} - \mathbf{D}\Gamma\|_F^2 \text{ subject to } \forall i \ \|\gamma_i\|_0 \leq T_0.$$

Input: Initial dictionary $\mathbf{D}^{(0)} \in \mathbb{R}^{N \times P}$, with normalized columns, signal matrix $\mathbf{B} = [\mathrm{x}_1, \cdots, \mathrm{x}_m]$ and sparsity level T_0.

Output: Trained dictionary $\mathbf{D}$ and sparse representation matrix Γ.

Procedure:

Set $J = 1$. Repeat until convergence:

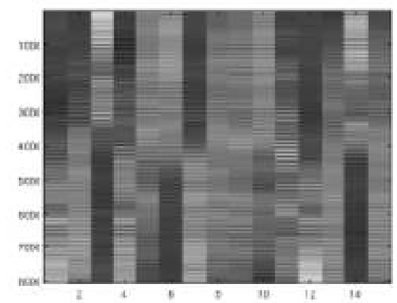
- *Sparse coding stage*: Use any pursuit algorithm to compute the sparse representation vectors γ_i for each signal $[\mathrm{x}_1, \cdots, \mathrm{x}_m]$.
- *Dictionary update stage*: For each column $k = 1, \cdots, P$in $\mathbf{D}^{(J-1)}$ update by
  - Define the group of examples that use this atom, $\omega_k = \{i | 1 \leq i \leq P, \gamma_T^k(i) \neq 0\}$.
  - Compute the overall representation error matrix, E_k, by

    $$\mathrm{E}_k = \mathbf{B} - \sum_{j \neq k} \mathrm{d}_j \gamma_T^j.$$
  - Restrict E_k by choosing only the columns corresponding to ω_k and obtain E_k^R.
  - Apply SVD decomposition $\mathrm{E}_k^R = \mathrm{U}\Delta\mathrm{V}^T$. Select the updated dictionary column $\hat{\mathrm{d}}_k$to be the first column of U. Update the coefficient vector γ_R^k to be the first column of V multiplied by $\Delta(1,1)$.
- Set $J = J + 1$.

Given a set of training examples, $\mathbf{B} = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$, find the dictionary that leads to the best representation for each member in this set, under strict sparsity constraints.

Training faces

KSVD



Learned dictionary

Ahron, et al, IEEE Trans. SP, Nov. 2006.

Outlier rejection

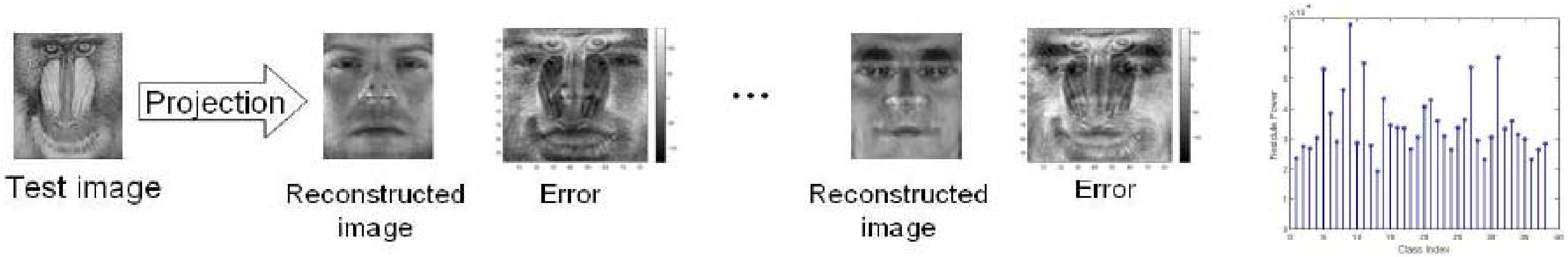


Figure: An invalid test image and the resulting residuals.

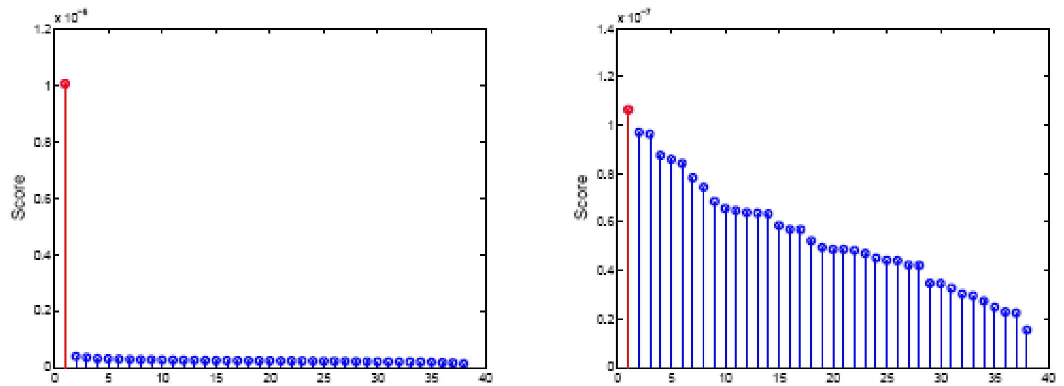


Figure: The score values corresponding to a valid sample (left) and an invalid sample (right).

Define score s_{yi} of the test image $\mathbf{y}$ to the i^{th} class as $s_{yi} = \frac{1}{\|\mathbf{r}^i(\mathbf{y})\|_2^2}$. Sort score values in decreasing order such that $s'_{y1} \geq s'_{y2} \geq \cdots \geq s'_{yC}$. Use $\lambda_y = \frac{s'_{y1}}{s'_{y2}}$ to reject outliers.

Dealing with the illumination problem

- Robust albedo estimation (Biswas et al. PAMI 2009)
  - Estimate albedo
  - Relight images with different light source direction
  - Use relighted images for training

Lambert's Cosine Law

$$I = \rho \max(\boldsymbol{n}^T \boldsymbol{s}, 0)$$

Image Albedo Surface Normal Light Source

Input Image

Estimated Albedo

Fig. 4. Examples of the original images (first column) and the corresponding relighted images with different light source directions from the PIE data set.

Experimental results (Yale B)

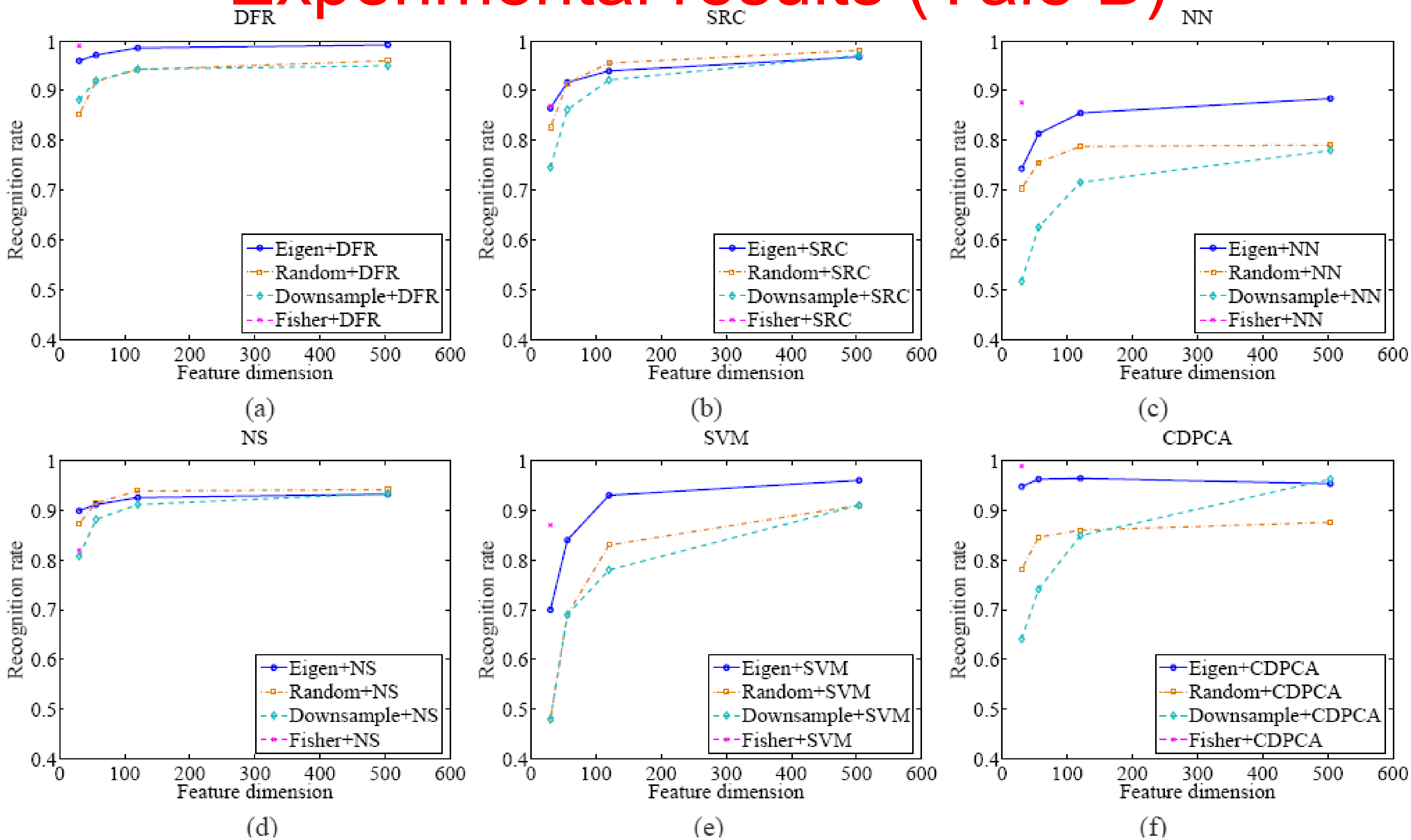


Fig. 5. Performance comparison on the Extended Yale B database with various features, feature dimensions and methods. (a) Our method (DFR) (b) SRC [5] (c) NN [5] (d) NS [5] (e) SVM [5] (f) CDPCA.

Results on other data sets not shown due to lack of time

- **DFR – 99.17 %**
- **SRC – 98.1 %**
- **CDPCA – 98.83 %**

Experimental results

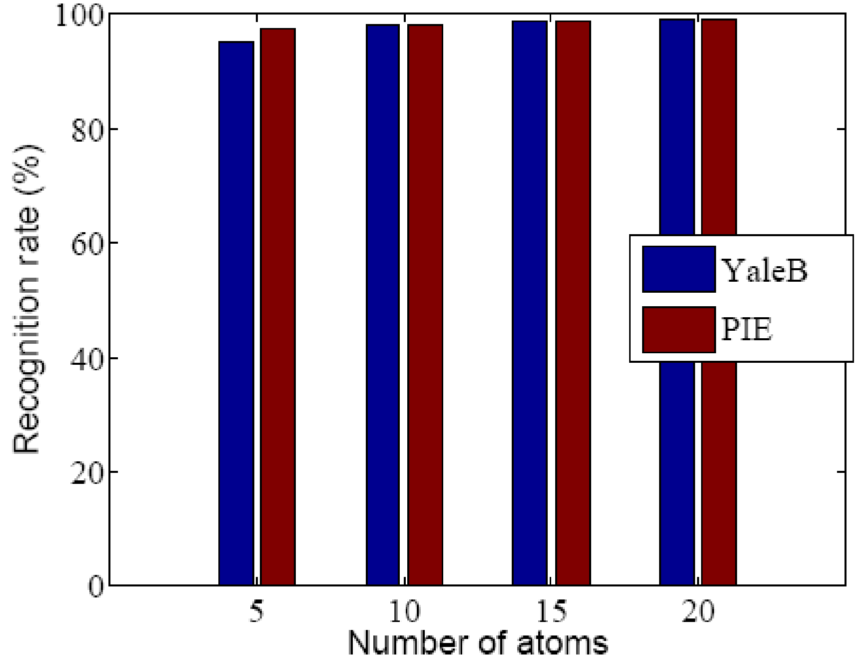


Fig. 8. Recognition rate vs. number of dictionary atoms on Extended Yale B and PIE data sets.

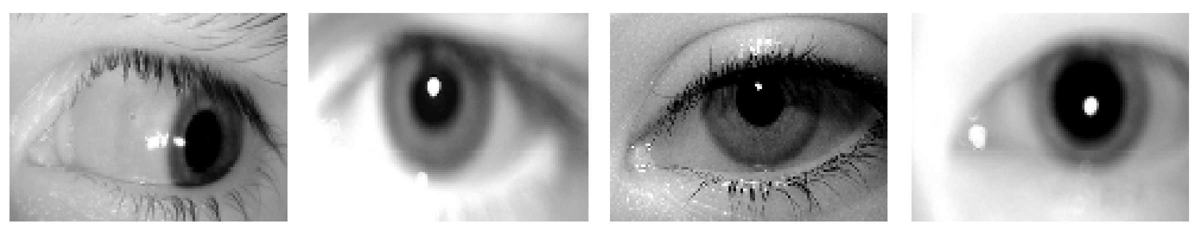
TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT METHODS WITH RESPECT TO THE NUMBER OF TRAINING SAMPLES PER SUBJECT.

No. of training images	DFR	SRC	NS	CDPCA
1	**75.89%**	42.37%	36.13%	5.52%
2	**84.71%**	37.20%	46.36%	26.22%
3	**85.18%**	37.45%	52.40%	30.25%

Iris recognition

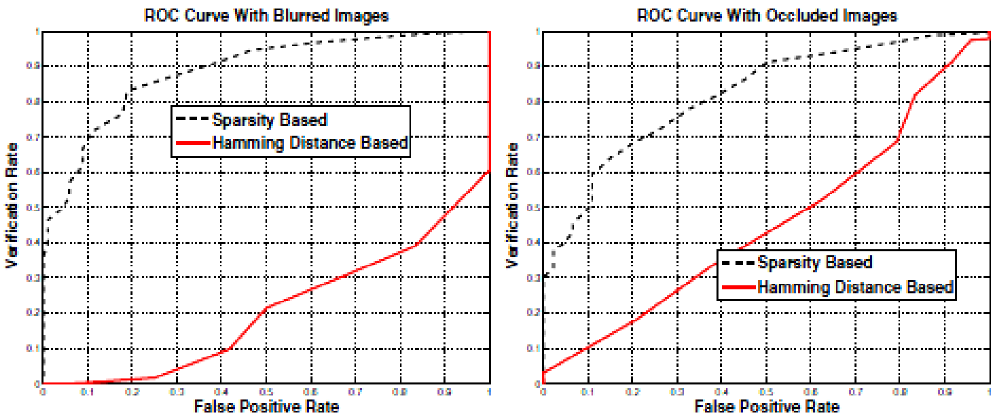
- Recognize a person from the texture features on his/hers iris image [Pillai et al, BTAS 2009]
- Existing algorithms [Daugman 93] give high recognition rates.
- Iris images acquired from a partially cooperating user often suffer from
  - Specular reflections, segmentation error, occlusion & blur.
- We developed a sparse representation-based algorithm for iris image selection and recognition - To appear in PAMI

Pillai et al, BTAS 2009

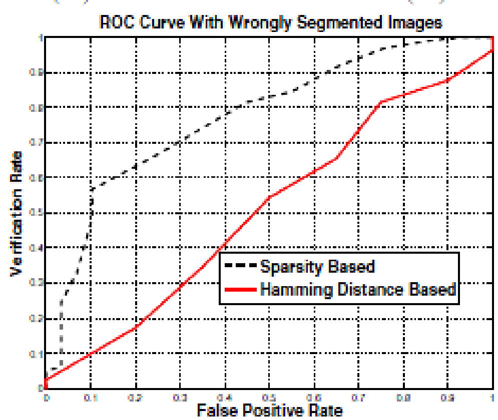


Results

Image Selection – ROC curves

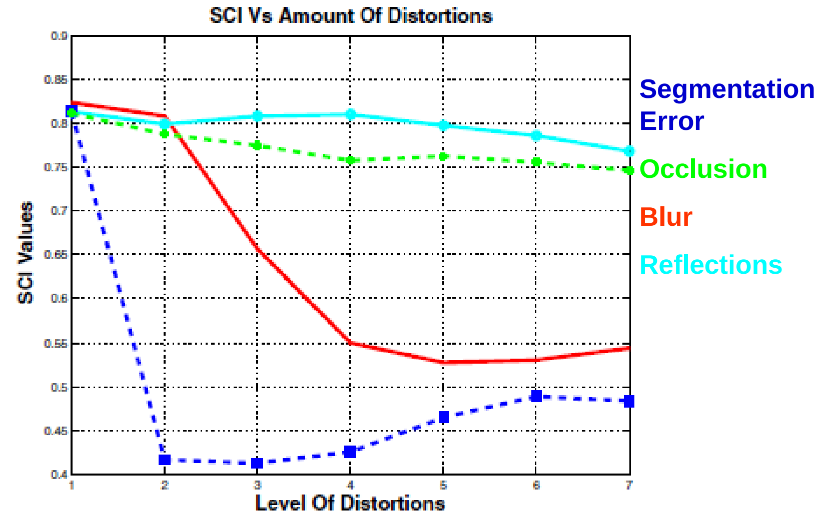


(a) (b)



(c)

SCI Variations with distortions



Recognition Rates On ND-IRIS-0405 Dataset.

Image Quality	NN	Masek's Implementation	Our Method
Good	98.33	97.5	99.17
Blured	95.42	96.01	96.28
Occluded	85.03	89.54	90.30
Seg. Error	78.57	82.09	91.36

Pillai et al, BTAS 2009

Sectored random projections For cancelable iris biometrics

- Need For Cancelability
  - Iris patterns are unique to each person.
  - Iris patterns cannot be re-issued if stolen.
  - Different patterns required for different applications.
- Cancelable Biometrics – Apply a revocable and non invertible transformation on the original one [Ratha et al, 2001, Teoh et al, 2006, Hao et al, 2006]
- Requirements
  - Performance should be retained.
  - Should be non-invertible and revocable.
  - Different codes for different applications.
- Pillai et al, ICASSP 2010, PAMI 2011.

Random projections for cancelability

- Random Projections (RP) can be utilized due to Johnson Lindenstrauss (JL) lemma [1984].

Lemma 1. *(Johnson-Lindenstrauss) Let* $\epsilon \in (0,1)$ *be given. For every set* S *of* $\sharp(S)$ *points in* $\mathbb{R}^N$*, if* n *is a positive integer such that* $n > n_0 = O\left(\frac{\ln(\sharp(S))}{\epsilon^2}\right)$*, there exists a Lipschitz mapping* $f : R^N \rightarrow R^n$ *such that*

$$(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \leq \|f(\mathbf{u})-f(\mathbf{v})\|^2 \leq (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \quad (1)$$

for all $\mathbf{u}, \mathbf{v} \in S$.

- One such mapping is projection using a Random Matrix.

Random projections

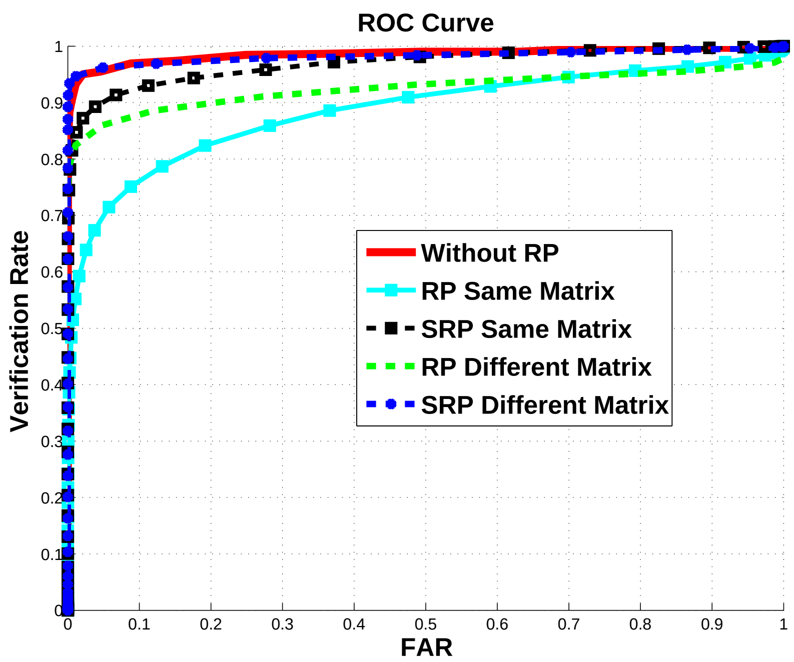
- N dimensional Gabor features of the iris image g is projected randomly onto a subspace of dimension n as follows

$$\mathbf{y} = \Phi \mathbf{g} \in \mathbb{R}^n$$

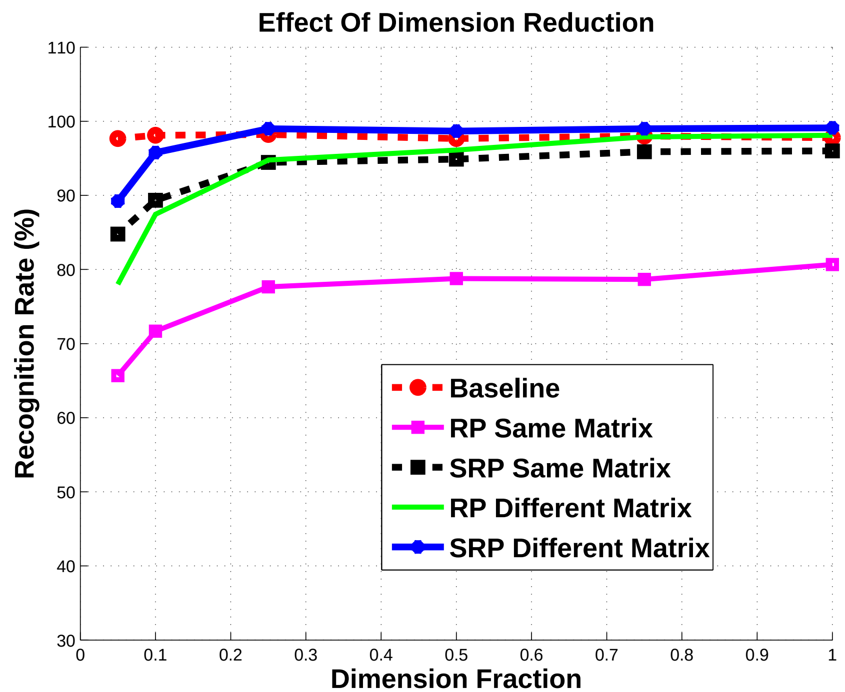
- Recognition is performed on the vector y.
- Direct Random Projections on iris images give poor results due to :
  - Occlusion due to eyelids act as outliers corrupting the whole data after the linear transform.
  - Combines both the good and bad regions of the iris image.

Results – Recognition performance

Performance Comparison



Effect Of Dimension Reduction



Observations

- SRP performs close to the original system without random projections.
- Only minor drop in performance due to dimension reduction upto 30% of the original dimension.
- Dimension reduction further improves non invertibility due to the non zero dimension of the null space.
- Pillai et al, ICASSP 2010.

Other works not discussed

- CS for video interpretation
- Compressive wireless arrays for bearing estimation of sparse sources in angle domain (Cevher et al, ICASSP 2008).
- Compressed sensing for multi-view tracking and 3-D voxel reconstruction (Reddy et al, ICIP 2008).
- Compressive SAR imaging (Patel, et al, 2008)

Remarks

- Compressive sensing cannot and will not solve all computer vision problems.
- May provide better solutions in many cases.
- Renewed interest in *l*-1-based methods!
- Compressive sensor design will lead the way.
  - Coded aperture imaging
  - Algorithms have to be integrated with sensing to reap the full benefits.
  - Tempting to replace all *l* -2 regularization methods with *l* -1.
- Ask if you can do without CS before attempting to do with CS!
- Check the Rice University website

Compressive sensing for video processing

- Background subtraction and tracking
  - Cevher et al, ECCV 2008, Reddy, et al, ICIP 2008.
- Registration
- Acquisition of dynamic scenes
  - Dynamic texture models (Turaga, et al, submitted for ECCV 2010)

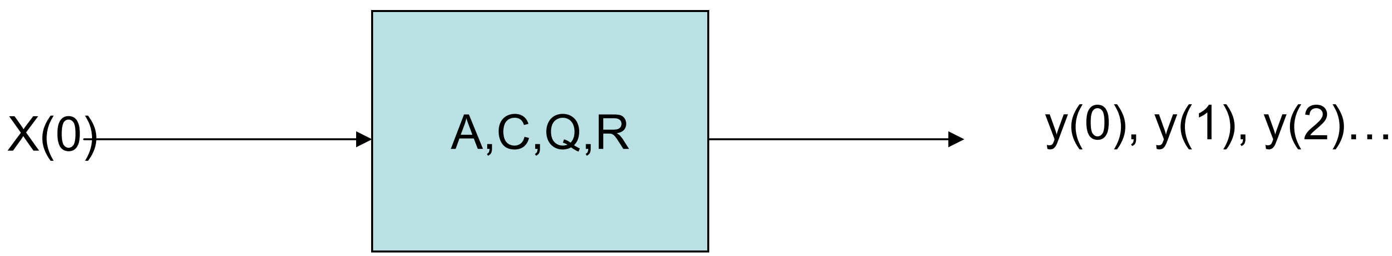
LDS model

- Video modeled as first order Gauss Markov process

$y(t) = C\,x(t) + w(t)$

$x(t+1) = A\,x(t) + v(t)$

x(t) d x 1 hidden state vector
A: d x d transition matrix
C: N x d observation matrix



Why LDS models?

- Useful for modeling video sequences
  - Soatto
- Have been used for unsupervised video clustering
  - Turaga, et al CVIU 2008
- Have been used for human activity recognition
- Orthonomral C matrices enable geodesic distances for classification
  - Turaga, et al, CVPR 2008, PAMI 2011.

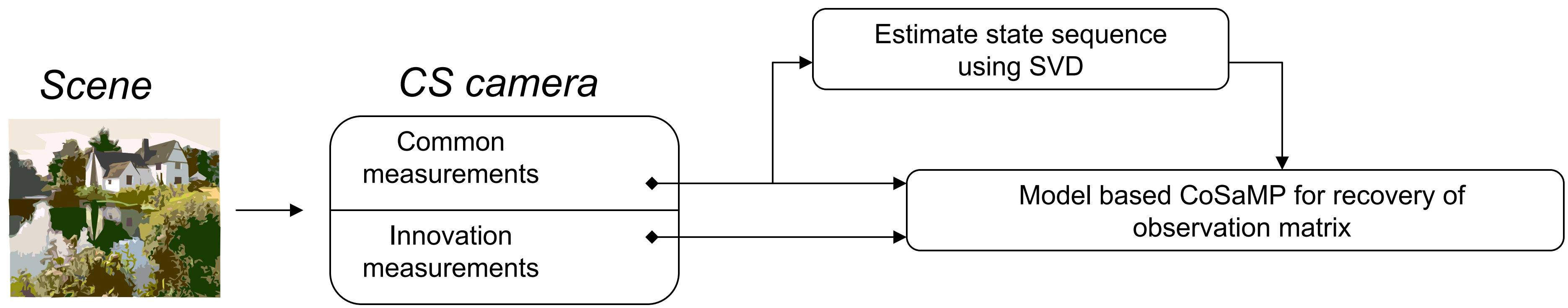
Measurement strategy

- Take two sets of measurements at each time-instant – one common, another that varies with time

$$\mathbf{z}_t = \begin{pmatrix} \check{\mathbf{z}}_t \\ \tilde{\mathbf{z}}_t \end{pmatrix} = \begin{bmatrix} \check{\Phi} \\ \tilde{\Phi}_t \end{bmatrix} \mathbf{y}_t = \Phi_t \mathbf{y}_t$$

Common measurements: Same measurement matrix

Model driven sensing of dynamic scenes



Solve for the state-sequence and the observation matrix

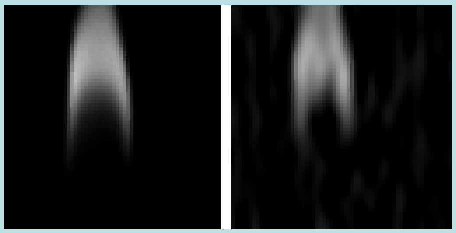
- The state-sequence can be estimated upto a linear transform by SVD of common measurements.

$$[\check{\mathbf{z}}]_{1:T} = \left[\, \check{\mathbf{z}}_1 \; \check{\mathbf{z}}_2 \; \cdots \; \check{\mathbf{z}}_T \,\right] = \check{\Phi} C \left[\, \mathbf{x}_1 \; \mathbf{x}_2 \; \cdots \; \mathbf{x}_T \,\right] = \check{\Phi} C [\mathbf{x}]_{1:T}$$

- Assume columns of C are sparse in a basis ψ, then solve for the following problem by structured CoSAMP.

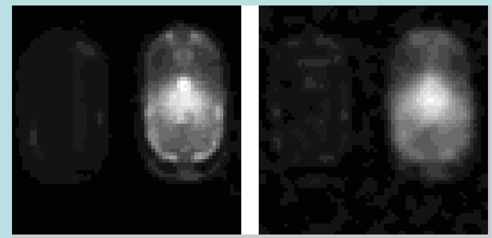
$$\min \sum_{k=1}^{d} \|\Psi^T \mathbf{c}_k\|_1, \text{ subject to } \|\mathbf{z}_t - \Phi_t C \hat{\mathbf{x}}_t\|_2 \leqslant \epsilon, \forall t$$

Reconstruction results



(Ground truth) (Reconstruction)

Candle sequence
- 1024 frames captured at 1000 fps
- Resolution 64 x 64
- Measurement rate: 1.2 %
- State dimension: 15



(Ground truth) (Reconstruction)

LED sequence
- 500 frames
- Resolution 64 x 64
- Measurement rate: 2.1 %
- State dimension: 7

Video classification results

- UCSD dataset consists of 254 videos of length 50 frames capturing traffic of three types – light, moderate, heavy. We performed a classification experiment of the videos into these three categories.
- 4 different train-test scenarios provided with the dataset.
- Classification is performed using the subspace-angles based metric with a nearest-neighbor classifier on the LDS parameters.
- The experiment was performed using the parameters estimated directly without reconstructing the frames. For comparison, we also perform the same experiments with fitting the LDS model on the original frames (oracle LDS).

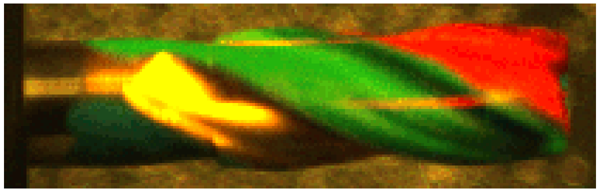
Classification results

Classification results on the traffic databases for two different values of state space dimension d. Results are over a database of 254 videos, each of length 50 frames at a resolution of 64 x 64 pixels under a measurement rate of 4%.

	d=5				d=10			
	Expt. 1	Expt. 2	Expt. 3	Expt. 4	Expt. 1	Expt. 2	Expt. 3	Expt. 4
Oracle LDS	85.71	85.93	87.5	92.06	77.77	82.81	92.18	80.95
CS-LDS	84.12	87.5	89.06	85.71	85.71	73.43	78.10	76.10

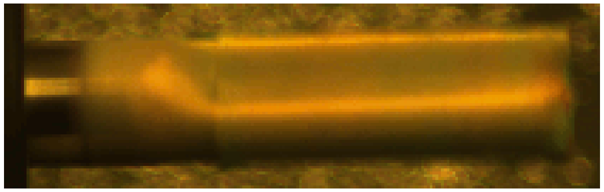
Rotating Mill Tool captured by Dragonfly2

Mill tool rotating at 50Hz



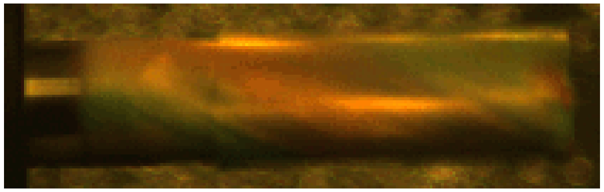
Reconstructed Video at 2000fps

Mill tool rotating at 50Hz



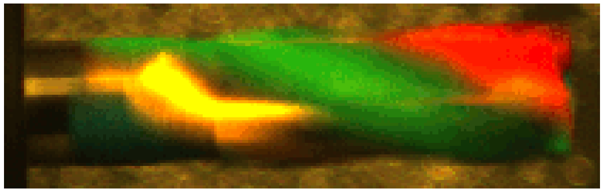
Normal Video: 25fps

Mill tool rotating at 50Hz



Coded Strobing Video: 25fps

Mill tool rotating at 150Hz

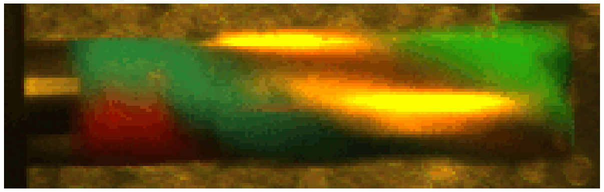


Reconstructed Video at 2000fps

Mill tool rotating at 100Hz

Reconstructed Video at 2000fps

Mill tool rotating at 200Hz



Reconstructed Video at 2000fps