Compressive Sensing: Opportunities and pitfalls for Computer Vision

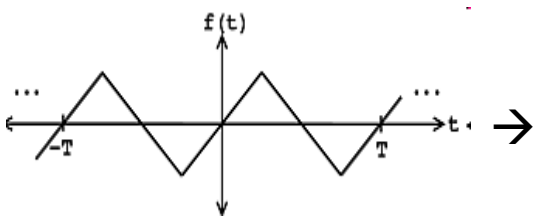
Rama Chellappa

Joint work with
Vishal Patel
Jai Pillai
Dikpal Reddy
Aswin C. Sankaranarayanan
Ashok Veeraraghavan
Dr. Volkan Cevher (Rice)

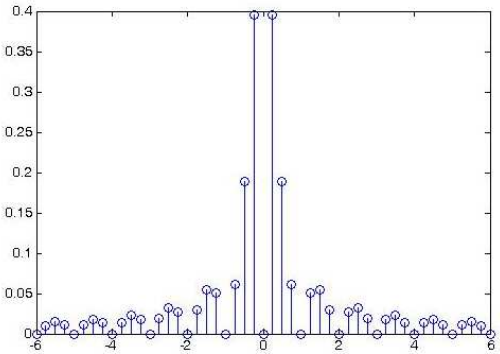
Outline

- Compressive sensing (CS)
  - Basics
  - Restricted isometry property (RIP), Johnson Lindenstrauss (JL) lemma.
  - Recovery algorithms
- Three examples
  - Reconstruction from gradient fields using *l-1* optimization
  - Sparsity-induced iris recognition and cancelability using random projections.
  - CS and graphics
  - Other applications
- Concluding remarks
  - Breakthroughs will come in integrated compressive sensing and processing
  - l-1 optimization and sparse representations by themselves can take you only so far in solving computer vision problems!

Sparsity: Motivation



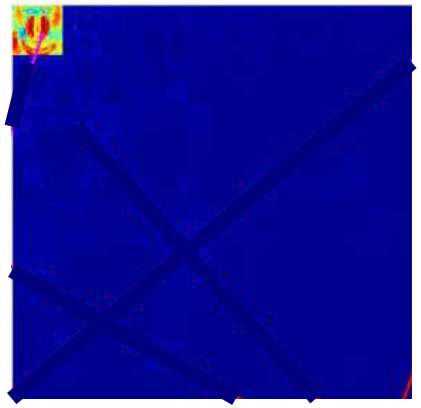
→



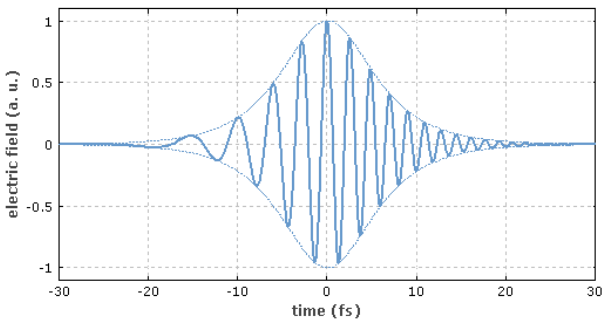
Periodic signal **Sparse in Fourier**

Background subtracted images

→



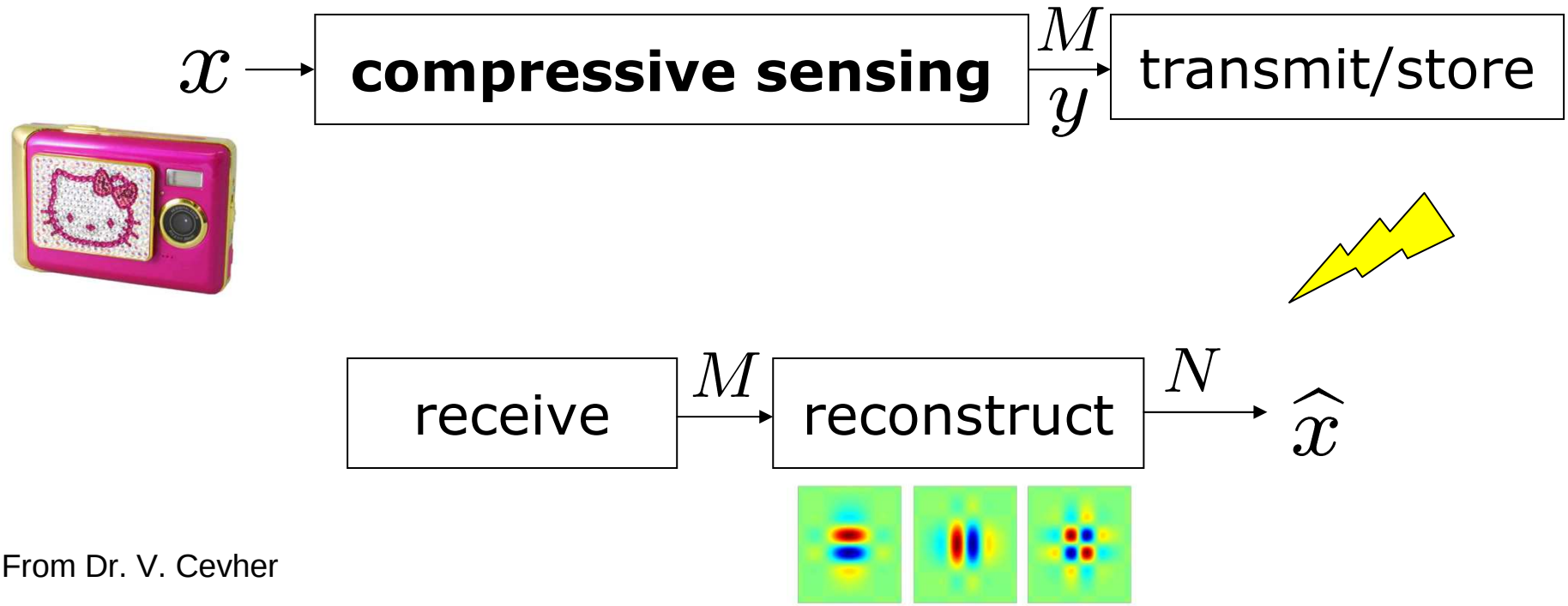
Images **Wavelet tree**



Chirp Signals → Sparse in Short Time Fourier Transform

Compressive sensing

- Directly acquire "***compressed***" data
- Replace samples by more general "measurements" $K \approx \underline{M} \ll N$

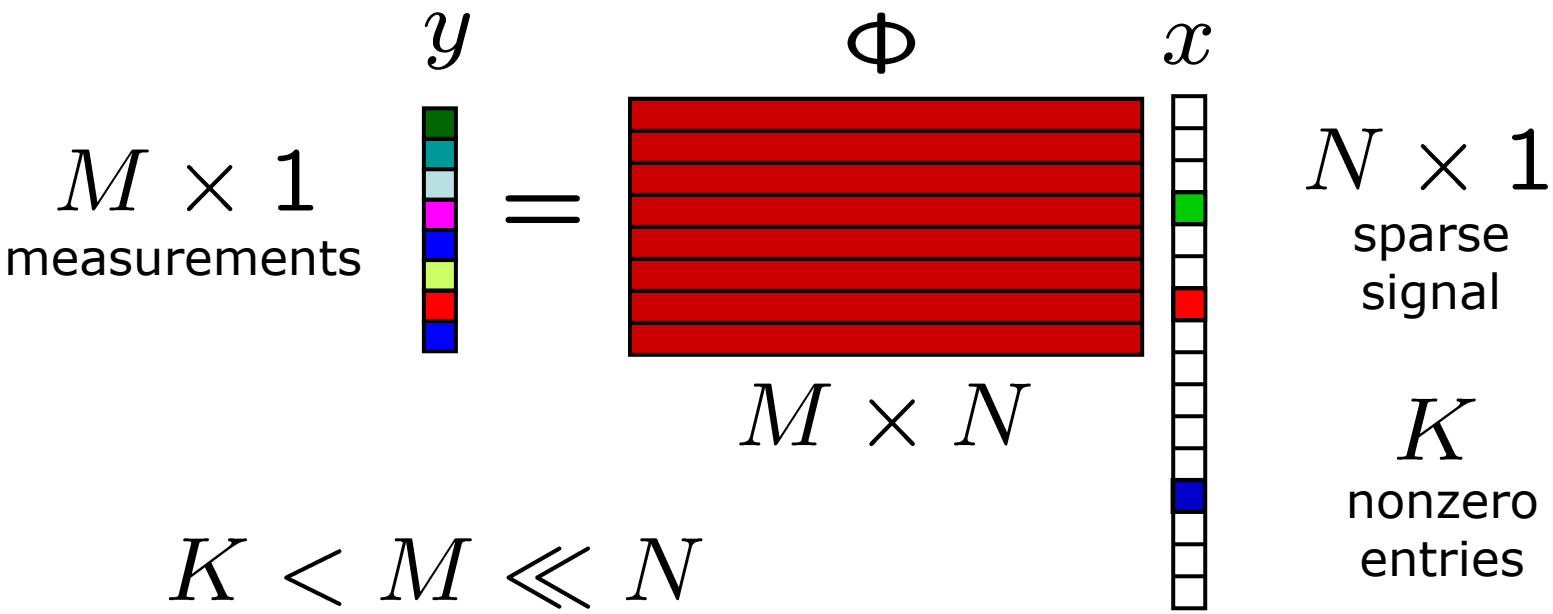


From Dr. V. Cevher

Compressive sampling

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

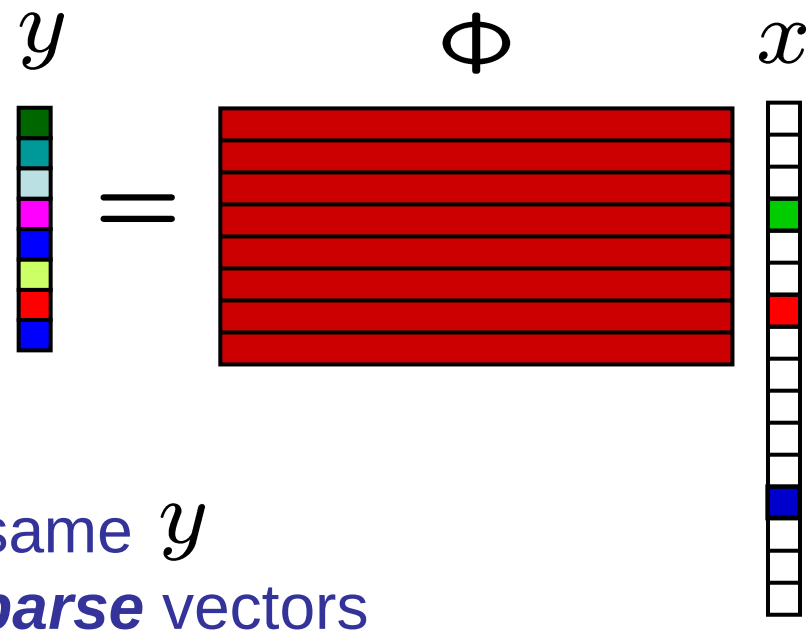
$$y = \Phi x$$



$$K < M \ll N$$

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)

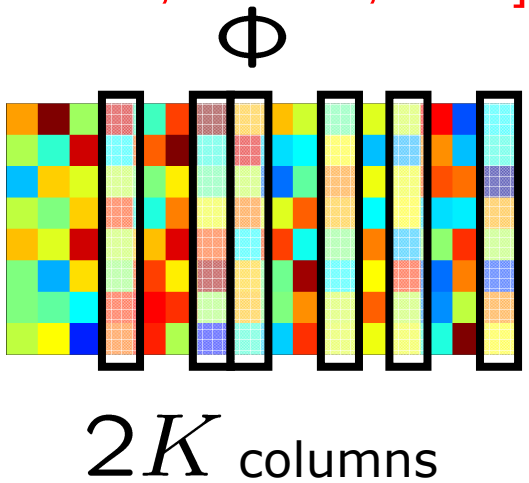
A matrix Θ is said to satisfy the RIP of order K with constants $\delta_K \in (0,1)$ if

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

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

- RIP is a sufficient condition for Basis Pursuit algorithms 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.
- Other conditions for l_0-l_1 equivalence include incoherence, phase transition diagrams [Donoho, 2004] and many more.
- RIP-p property, where the ℓ-2 norm is replaced by the ℓ-p norm

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

CS signal recovery

- Recovery: given $y = \Phi x$

  (ill-posed inverse problem) find x

  (sparse)

$$\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$$

- ℓ_2 fast, wrong
- 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

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]

From Dr. Cevher

software @
dsp.rice.edu/cs

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

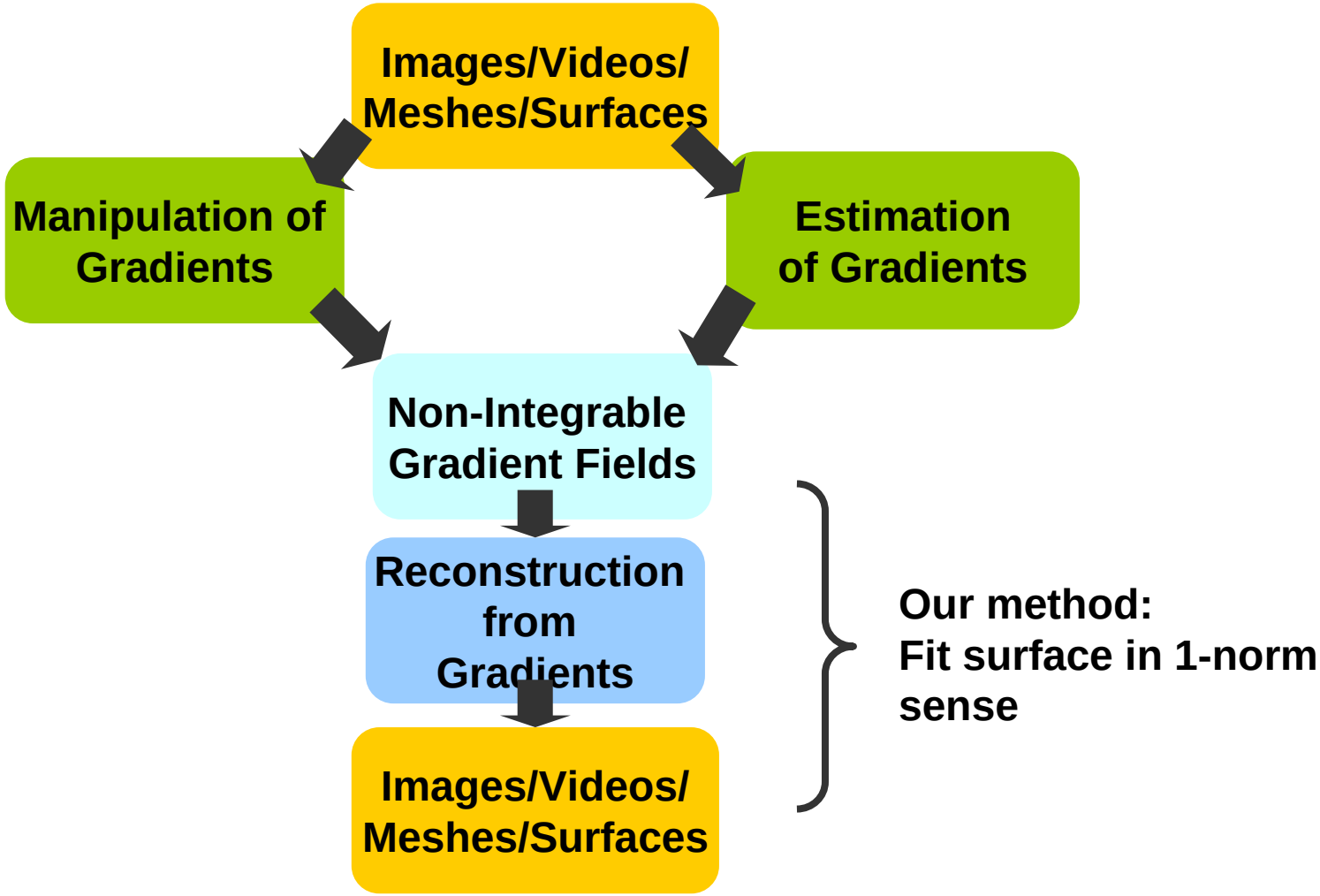
$$\min \|x\|_1 \, s.t \|y - \Phi x\|_2 \leq \varepsilon$$

From Dr. Cevher

Recent developments in computer vision/graphics

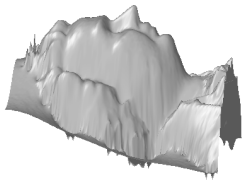
- Single pixel camera [Duarte, et al, 2008]
- Compressive sensing for background subtraction [Cevher et al, ECCV 2008]
- Face recognition [Wright et al, 2009]
- Compressive Sensing of Reflectance Fields
  - Peers et al, ACM Transactions on Graphics, 2009)
- Compressive SAR imaging
- Sparsity-induced algorithms for iris, shape from gradients, super resolution…
- Floodgates have been opened!

Gradient domain processing: vision and graphics



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

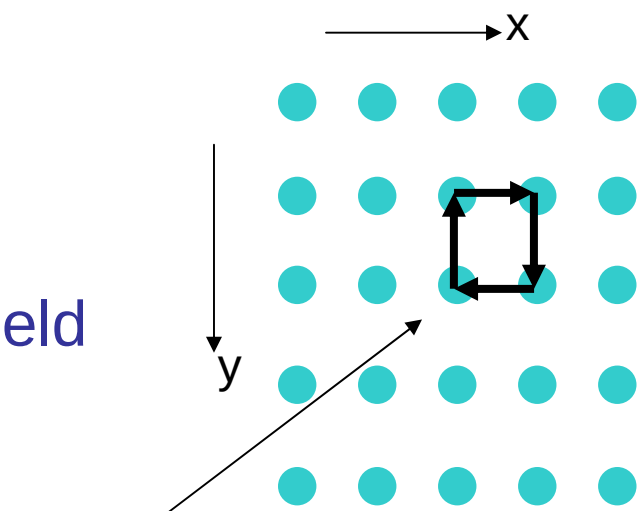
Curl: $Curl(\nabla S) = \nabla \times \nabla S$

Integrability: Conservative vector field

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

$$\nabla \times \nabla S = 0$$

$$Curl(\nabla S) = S_{yx} - S_{xy} = 0$$



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

Reddy, Agrawal and Chellappa
CVPR 2009

Discrete domain

- $S(x, y)$ surface of size $H \times W$ vectorized to give s

- Gradients g obtained from estimation/manipulation

$$g = Ds + e \quad \text{of size} \quad (H-1)W + (W-1)H$$

D: Gradient operation On **s**

- Non-integrable gradient field: Non-zero curl values d

$$Cg = CDs + Ce$$
$$d = Ce$$

Underdetermined system

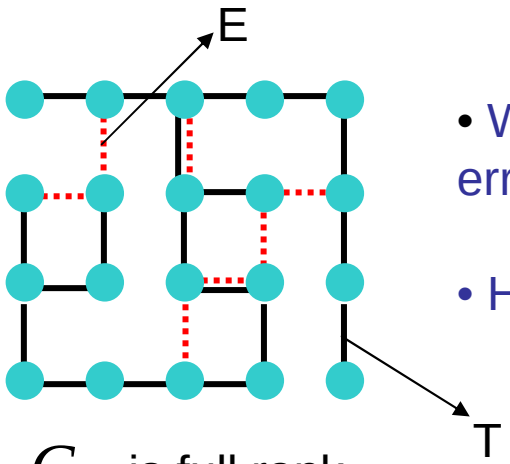
Curl matrix C of size $(H-1)(W-1) \times (H-1)W + (W-1)H$

$$C = \begin{pmatrix} -1 & 1 & & & & 1 & -1 & & & \\ & -1 & 1 & & & & 1 & -1 & & \\ & & & & & & & & & \\ & & & -1 & 1 & & & & 1 & -1 \end{pmatrix}$$

Reddy, Agrawal and Chellappa
CVPR 2009

Interpretation

- Graph analogy: 2D grid as planar graph
  - Nodes as surface values
  - Edges as gradient errors
  - Minimal set is the spanning tree



C_{T^C} is full rank

- Would like to find the spanning tree (T) which has least errors on its edges

- HW-1 edges in spanning tree (equal to N-M)

- If gradient errors in edges E, is C_E full rank?
- What $E_S \subset E$ such that C_{E_S} is full rank?
- Restricted Isometry Property of C?

Reddy, Agrawal and Chellappa
CVPR 2009

CS and reconstruction from gradient fields

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

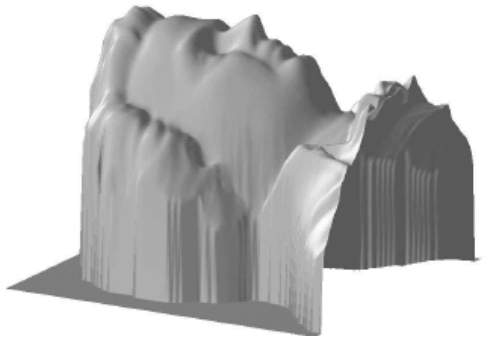
RIP -1: Berinde, et al, 2008

Reddy, Agrawal and Chellappa
CVPR 2009

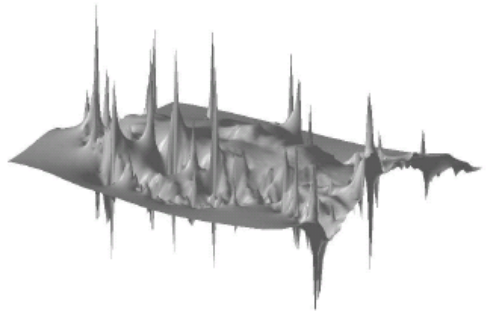
l_1- minimization

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

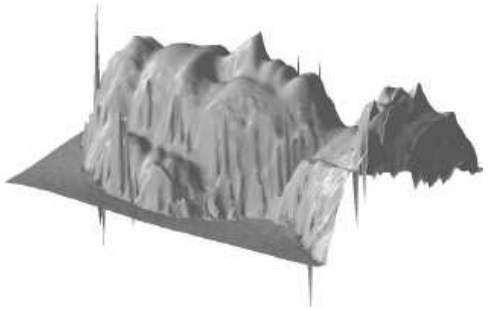
- Corrects outliers. Performs well in noise too.
- Good error confinement property locally



Ground truth



Least squares

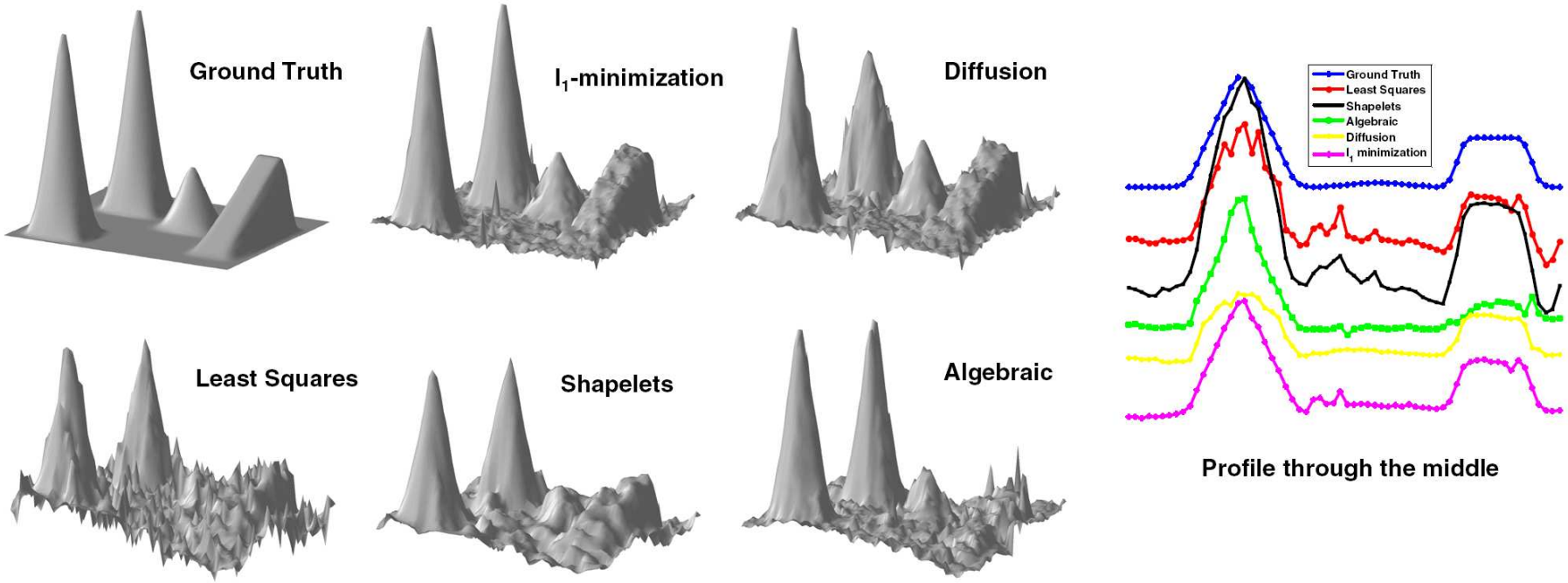


1-norm fit

Reddy, Agrawal and Chellappa
CVPR 2009

l_1 minimization

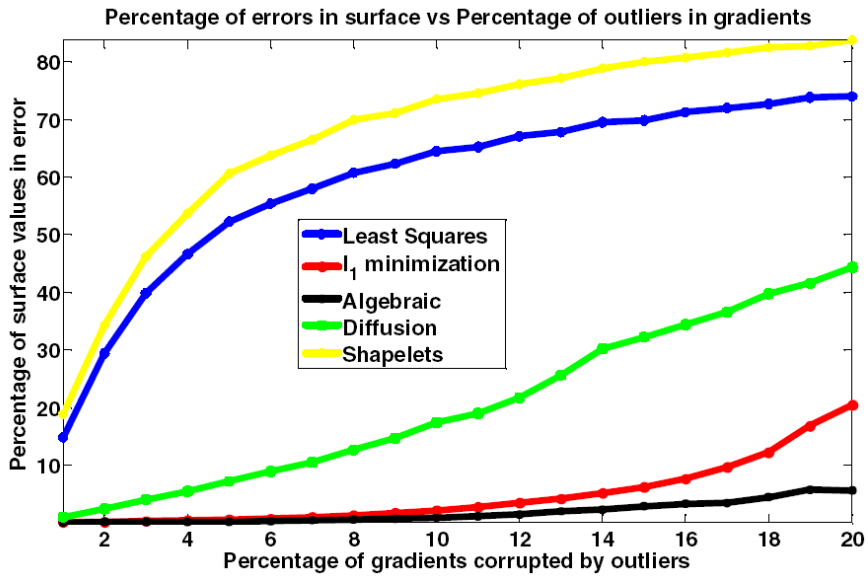
Under noise and outliers



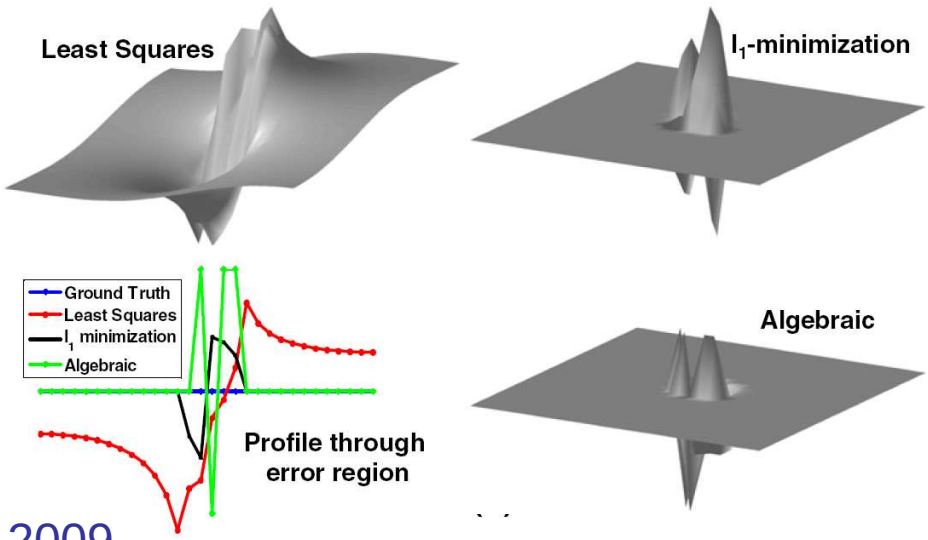
Reddy, Agrawal and Chellappa
CVPR 2009

l_1 minimization

Comparison



Local error confinement



Reddy, Agrawal
and Chellappa, CVPR 2009

Face recognition via sparse representations

- Automatic face recognition algorithm robust to occlusion, expressions and disguise.
- Represent the test face as a sparse linear combination of the training faces.
- Estimate the class of the test image from the sparse coefficients.
- Can identify and reject non face images.
- Performance affected by illumination variations and misalignment.
- John Wright et al. PAMI 2009

Formulation

- Let v_{ij} be the jth training image in the ith class.
- Matrix A contains the training face images as its columns

$$A \doteq [A_1, A_2, \ldots, A_k] = [\boldsymbol{v}_{1,1}, \boldsymbol{v}_{1,2}, \ldots, \boldsymbol{v}_{k,n_k}]$$

- The test image y can be written as a linear combination of the faces of the correct class.
- So if y belongs to the ith class

$$\boldsymbol{y} = \alpha_{i,1}\boldsymbol{v}_{i,1} + \alpha_{i,2}\boldsymbol{v}_{i,2} + \cdots + \alpha_{i,n_i}\boldsymbol{v}_{i,n_i}$$

- It can be written as $y = Ax_0$

$$\boldsymbol{x}_0 = [0, \cdots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0]^T$$

John Wright et al, PAMI 2009

Formulation

- Since the number of classes is high, the coefficient vector is sparse.
- So it can be recovered by solving the Basis Pursuit problem

$$\hat{\boldsymbol{x}}_1 = \arg\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{y}.$$

- The non zero coefficients in the sparse coefficient vector will correspond to the true class.
- Compute the residuals obtained when the test face is represented using each class separately.

$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\,\delta_i(\hat{\boldsymbol{x}}_1)\|_2$$

Where $\delta_i(x)$ is non zero only for the ith class columns.

- The class which minimizes this residual error is the estimated class of the test vector.
- Best results on Yale B data base.
- Robust to occlusion and disguise.
- John Wright et al. PAMI 2009

Handling registration and illumination

- Wagner et al. CVPR 2009
- 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.
  - Restrictive
- The sparse representation based algorithm is performed on this modified gallery.

Iris recognition - 1

- 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.

Pillai et al, BTAS 2009

Iris recognition - 2

- We developed a sparse representation-based algorithm for iris image selection and recognition.
  - Can perform selection and recognition in a single step.
  - Can handle wide variety of distortions like blur, occlusion, specular reflections and segmentation errors.
  - Is robust to occlusions common in iris images.
- Sparsity in representation is a measure of image quality and can be used for iris image selection.
- Iris is divided into multiple sectors to take care of problems due to occlusion.
- Sparsity is better for iris than faces!
- Pillai et al, BTAS 2009

Basic formulation

- Assume L classes and n images per class in gallery. $\mathbf{D}_k = [\mathbf{x}_{k1}, ..., \mathbf{x}_{kn}]$
- The training images of the kth class is represented as

$$\begin{aligned}\mathbf{D} &= [\mathbf{D}_1, ..., \mathbf{D}_L] \in \mathbb{R}^{N \times (n.L)} \\ &= [\mathbf{x}_{11}, ..., \mathbf{x}_{1n} | \mathbf{x}_{21}, ..., \mathbf{x}_{2n} | ..... | \mathbf{x}_{L1}, ..., \mathbf{x}_{Ln}]\end{aligned}$$

- Dictionary D is obtained by concatenating all the training images
- The unknown test vector can be represented as a linear combination of the training images as

$$\mathbf{y} = \sum_{i=1}^{L} \sum_{j=1}^{n} \alpha_{ij} \mathbf{x}_{ij}$$

Pillai, BTAS 2009

Basic formulation

- In a more compact form

$$\mathbf{y} = \mathbf{D}\alpha.$$

$$\alpha = [\alpha_{11}, ..., \alpha_{1n} | \alpha_{21}, ..., \alpha_{2n} | ..... | \alpha_{L1}, ..., \alpha_{Ln}]^T$$

- We make the assumption that the test image can be written as a linear combination of the training images of the correct class alone.
- So the coefficient vector α is sparse.
- Hence α can be recovered by Basis Pursuit as

$$\hat{\alpha} = \arg \min_{\alpha' \in \mathbb{R}^N} \| \alpha' \|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{D}\alpha'.$$

Pillai et al, BTAS 2009

Sparsity for image selection

- When the test image is well acquired, the coefficient vector α will be sparse.
- A measure of sparsity is the Sparse Concentration Index SCI, defined by

$$SCI(\alpha) = \frac{\frac{L.\max \|\Pi_i(\alpha)\|_1}{\|\alpha\|_1} - 1}{L - 1}.$$

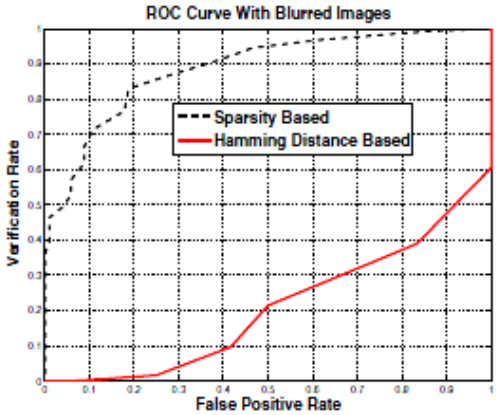
- SCI measures the fraction of the energy present in the "best" class.
- So well acquired images will have high SCI.
- Hence reject the images having low SCI value.
- Pillai et al, BTAS 2009

Selection and recognition algorithm

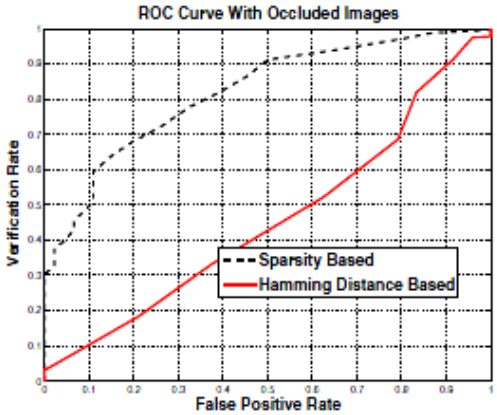
- Given the gallery, construct the dictionary D by arranging the training images as its columns.
- Using the test image, by Basis Pursuit, obtain the coefficient vector α.
- Obtain the Sparsity Concentration Index (SCI).
- Compare SCI with a threshold to reject the poorly acquired images.
- Find the reconstruction error while representing the test image with coefficients of each class separately.
- Select the class giving the minimum reconstruction error.
- Pillai et al, BTAS 2009

Results

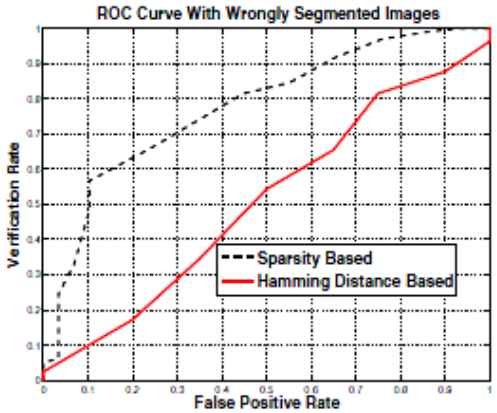
Image Selection – ROC curves



(a)

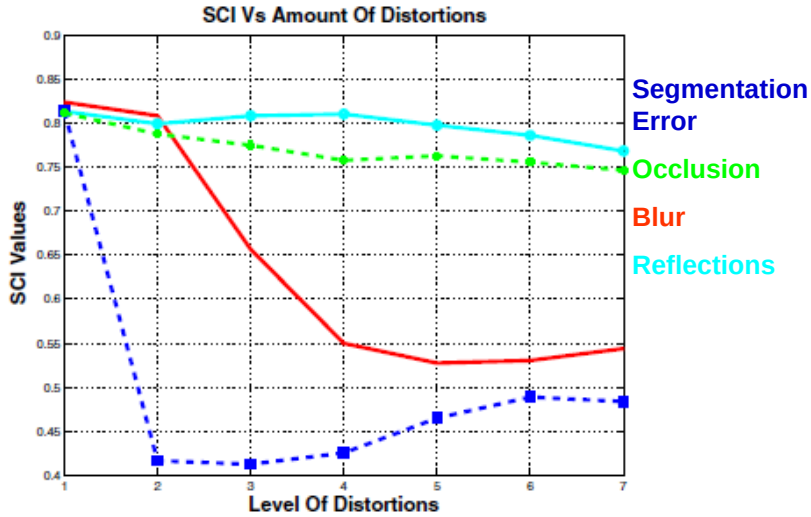


(b)



(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, Submitted for ICASSP 2010

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

- Thus distance between two higher dimensional vectors remain same even after projecting to a lower dimensional space using certain mappings.
- 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

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

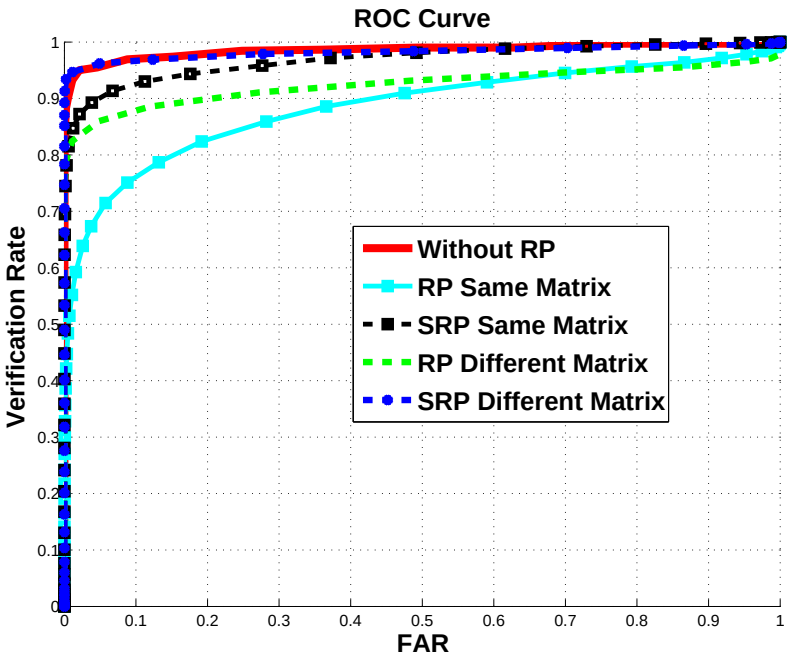
- Recognition is performed on the vector y.
- Direct Random Projections on iris images give poor results due to :
  - Occlusion due to eyelids act as outliers corrupting the whole data after the linear transform.
  - Combines both the good and bad regions of the iris image.
  - Pillai et al, Submitted to ICASSP 2010.

Sectored random projections (SRP)

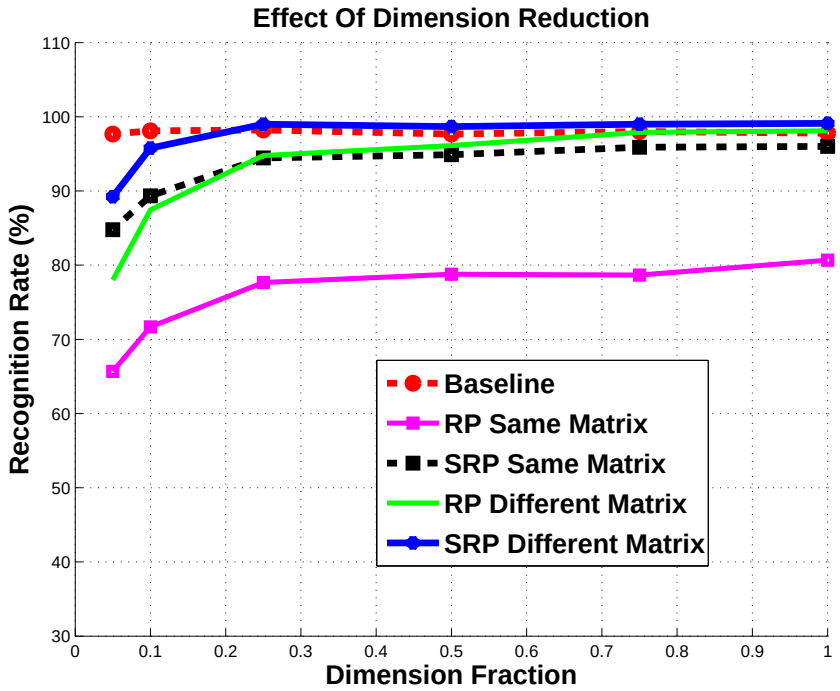
- Apply RP to different iris sectors separately.
- Advantages
  - Bad regions cannot corrupt the whole image.
  - Computations reduced from N^2 to Nk, K <<N.
- How SRP meets the Cancelability Requirements ?
  - **Performance** - does not drop after applying SRP.
  - **Non-Invertibility -** due to RP and Dimension Reduction.
  - **Revocability** – Apply a new RP if the old patterns are lost.
  - **Different Applications** – Assign a different matrix for each application.
  - **Compatibility** – Only a single matrix multiplication stage has to be added to existing algorithms.
  - Pillai et al, Submitted to ICASSP 2010.

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, Submitted to ICASSP 2010.

Model-based compressive sensing for reflectance fields

A Hybrid Subspace Sparse Signal Model

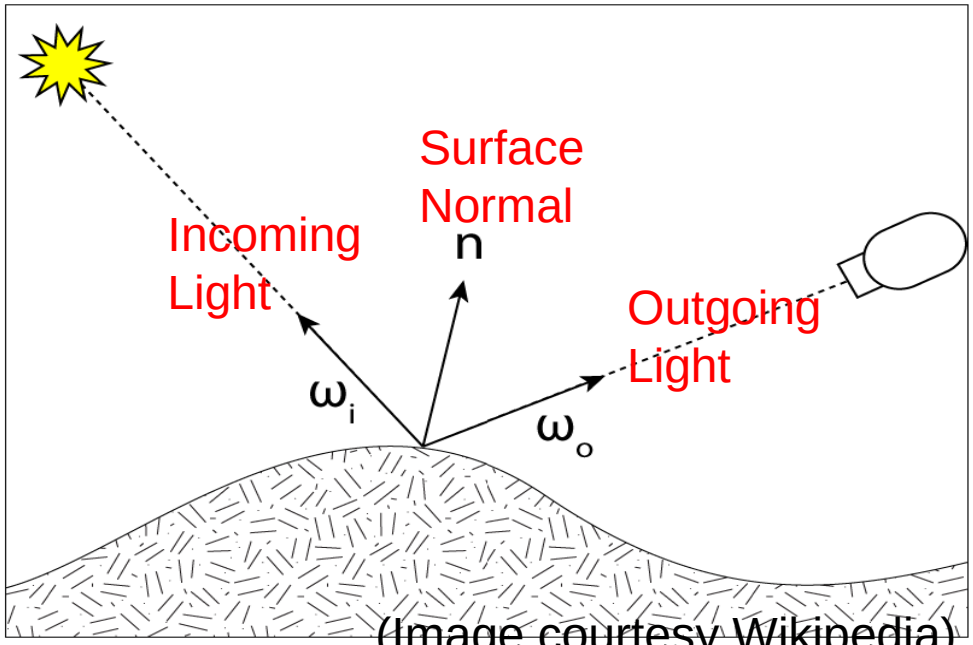
A. Sankaranarayanan and A. Veeraraghavan,

2009.

Capturing surface properties

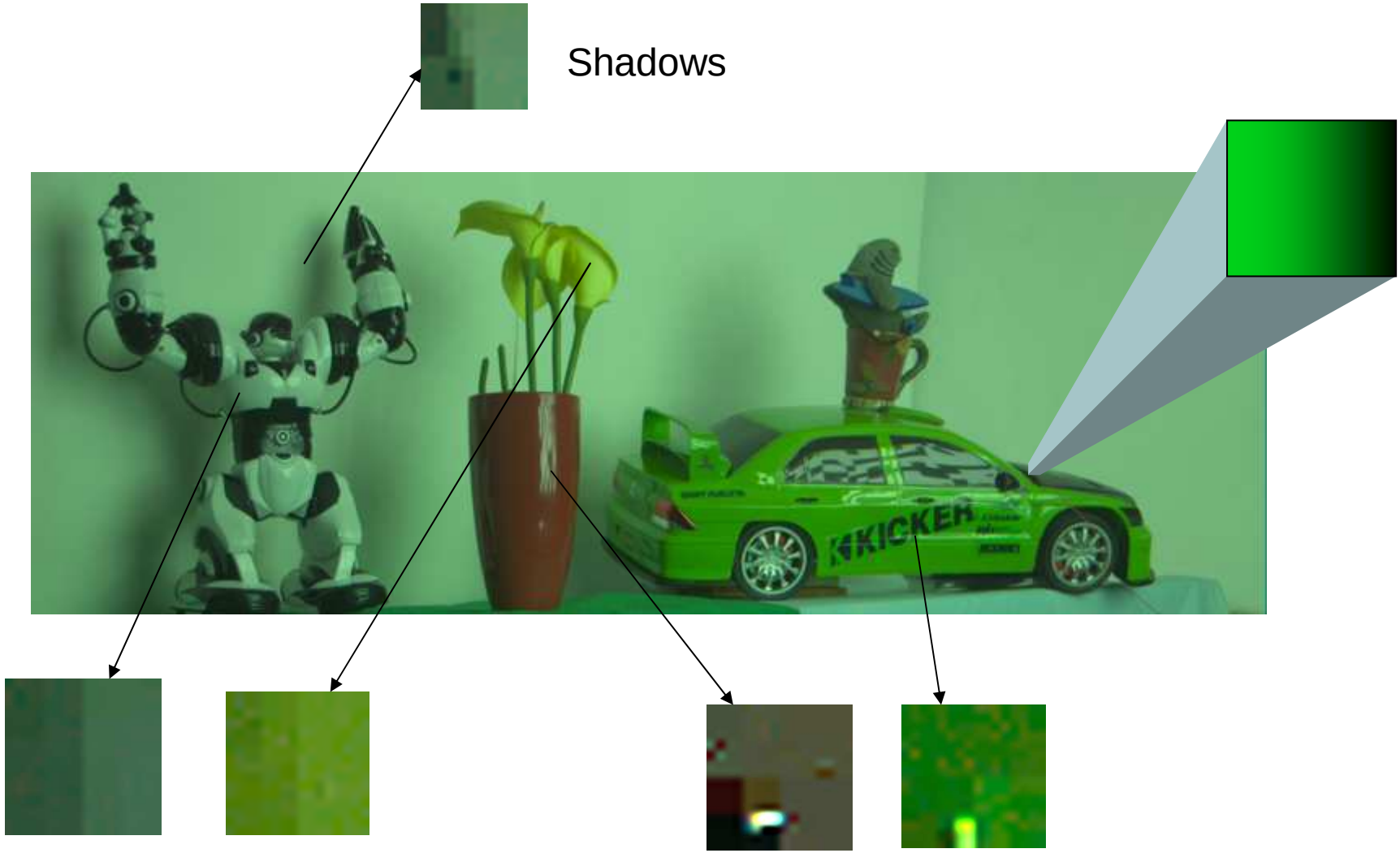
- Bidirectional Reflectance Function
  - Two incident angles, Two observation angles
  - Measures amount of “light” in the outgoing direction given unit incident irradiance.
  - 4D function characterizing surface properties
  - Acquisition ?

- Reflectance field
  - Fix outgoing direction



(Image courtesy Wikipedia)

Reflectance field of a scene



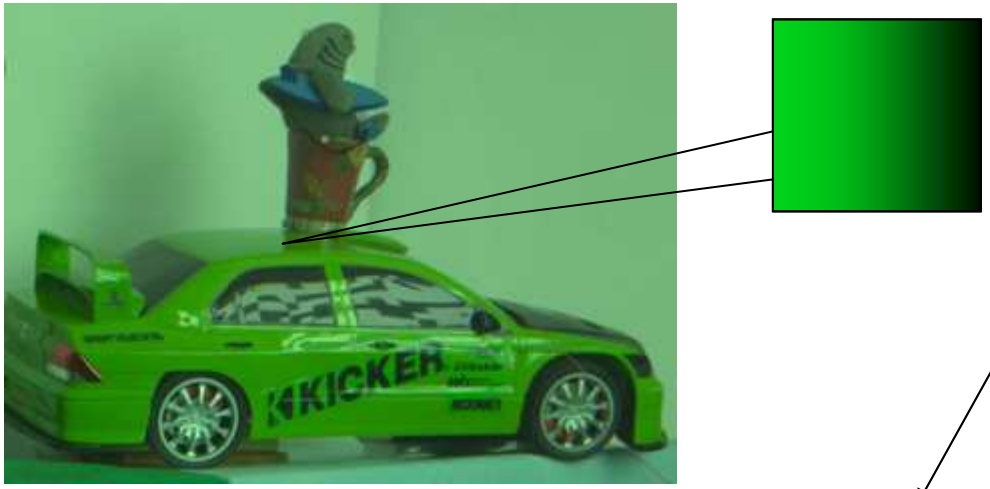
Compressive acquisition

Premise: BRDFs are inherently redundant and compressive

- Measurements are inherently costly!
- However, we can exploit the compressibility of the signal to solve for the Reflectance field from an under-determined set of equations.
- Solution: Take compressive measurements
- Use an “/1-inversion” to obtain the RF.

(Peers et al, ACM ToG, 2008)

Reflectance field acquisition



RF characterizes how appearance of a scene changes with changing scene illumination.

$$c_i = t_i^T L$$

Intensity profile at a pixel

Reflectance transfer function

Lighting conditions

When t_i is sparse in Basis B we can directly apply compressive sensing on Random multiplexed lighting L.

Solving for each pixel independently leads to incoherent reconstructions.
Need to enforce spatial coherency in reconstruction.
Peers et al , ACM ToG, 2009

Enforcing spatial coherency

$$c_i = t_i^T L$$

Image formation at each pixel

$$C = TL$$

Stacking individual pixels to get the image formation for the scene

$$C = T B B^T L$$

B is the sparsifying basis for T and exploits the redundancy in the RF of an individual scene point

$$PC = (PTB)(B^T L) = \hat{T}\hat{L}$$

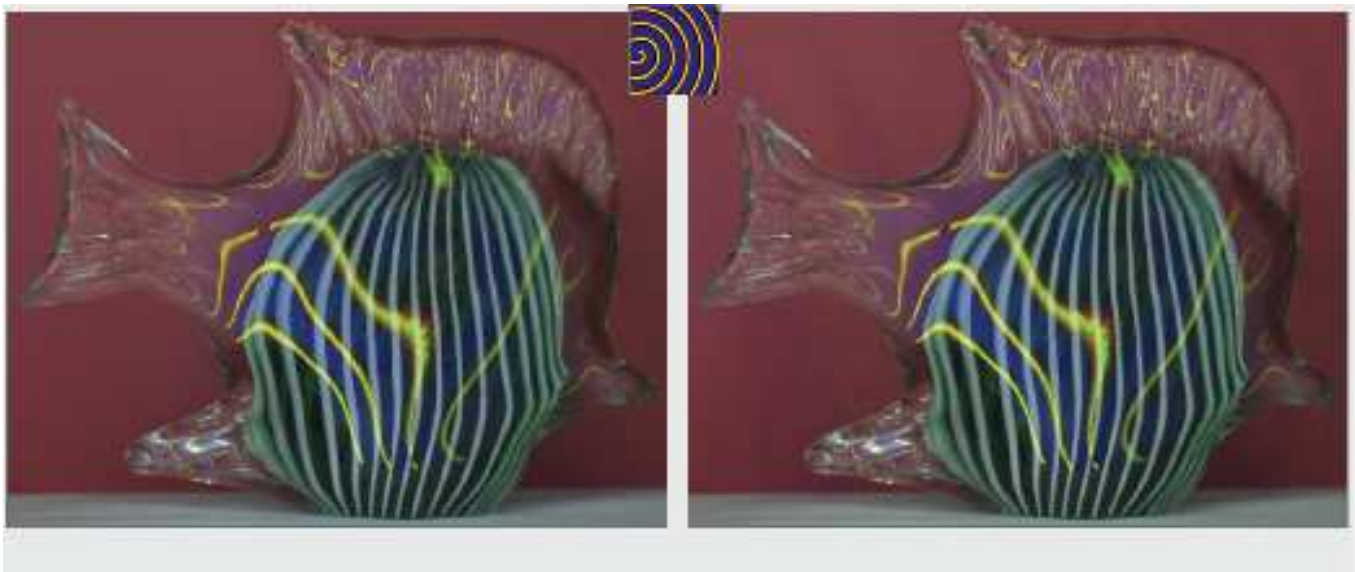
P is the spatial coherency basis. P is dependant on the scale at we choose to enforce spatial correlations. At different scales we have different “P” matrices.

Peers et al, ACM ToG, 2009

Compressive sensing set up (Peers et al 2009)

- A single video camera and a CRT monitor as a controllable high-resolution light field emitter.
- Radiometric calibration of the camera and the emitter
- A 2-D image constructed from each measurement vector from measurement ensemble (Gaussian random Haar wavelet coefficients) is emitted from the CRT monitor.
- A photograph of the scene is taken under this illumination contributing to a column in of the observation matrix.
- Has issues with measurement noise, quantization noise, etc.
- The authors (Peers et al, ACM ToG 2009) present novel solutions to these problems.

Relighting results



(from Compressive measurements)

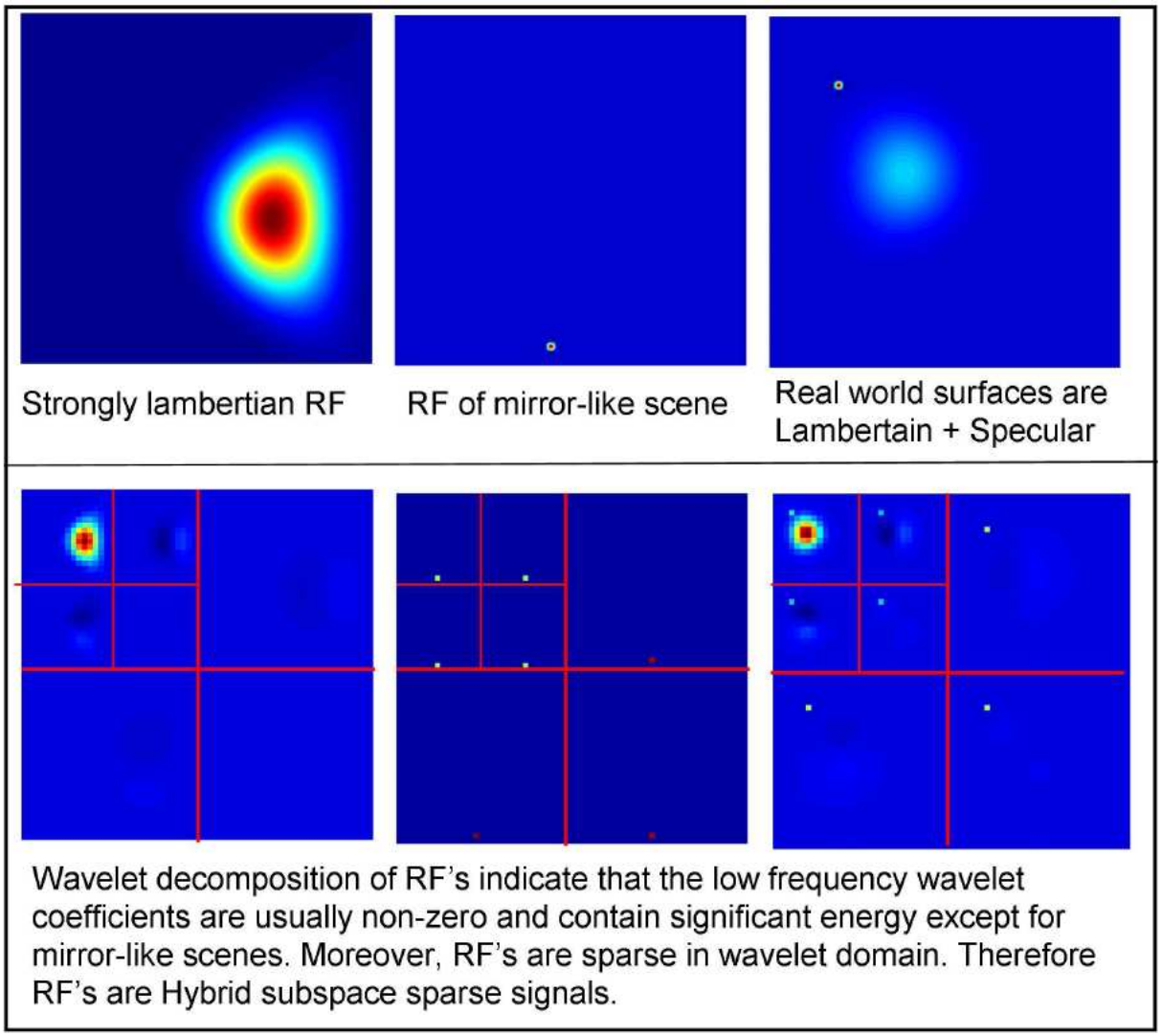
(Ground truth)

(Results courtesy Peers et al. ToG, 2008)

Real Signals have more structure than just Sparsity

Diffuse components of the RF are highly subspace compressible

In contrast, high frequency specularities are sparse.

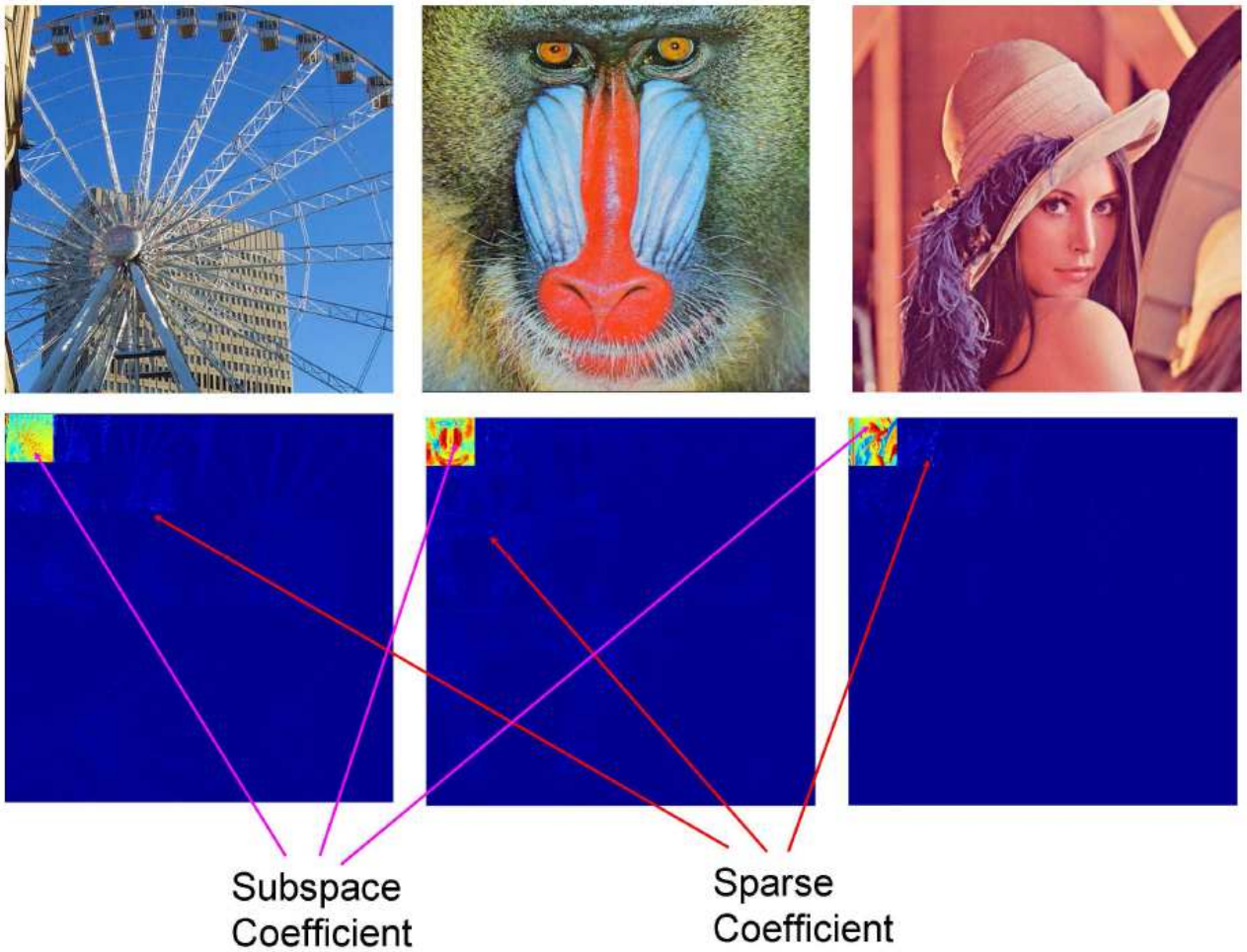


Wavelet decomposition of RF's indicate that the low frequency wavelet coefficients are usually non-zero and contain significant energy except for mirror-like scenes. Moreover, RF's are sparse in wavelet domain. Therefore RF's are Hybrid subspace sparse signals.

Images ...

The same is true of other visual signals such as reflectance fields (RF).

Here, the Lambert's part of the RF is subspace compressible, while the high frequency specular part of the RF is sparse.

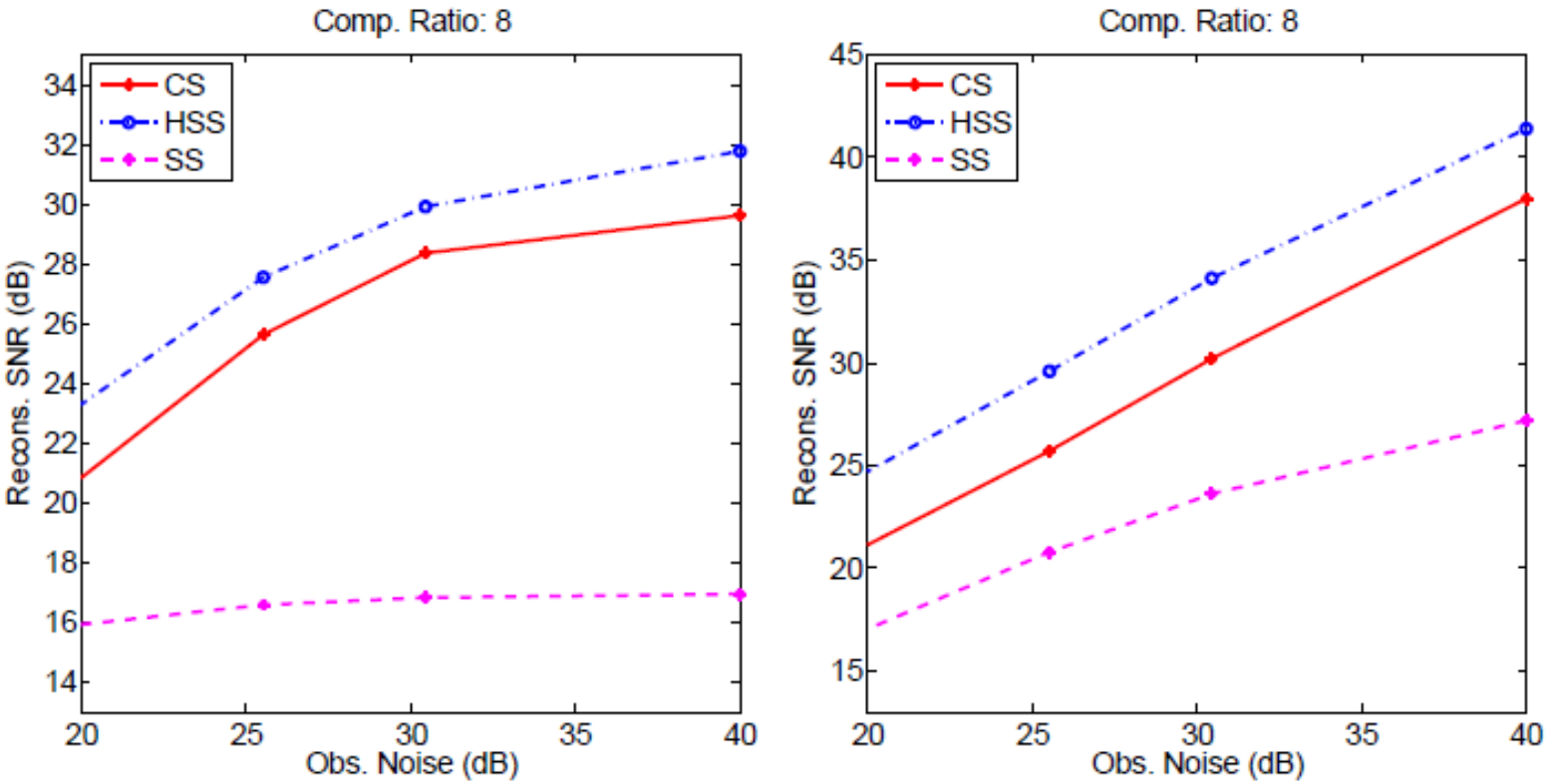


A. Sankaranarayanan and
A. Veeraraghavan, 2009.

General signal models

- Real world signals exhibit richer structure than just sparsity or subspace compressibility.
- A mixture of these two models is a better approximation of the actual signals.
- Merge the traditional theory of subspace sampling with that of compressive sensing to address these hybrid subspace sparse signals.
- A. Sankaranarayanan and A. Veeraraghavan, 2009.

Effectiveness of the hybrid subspace space model



The HSS signal model gives much better approximations at high compression ratios. Here, the signals are 64x64 Reflectance fields over two databases of BRDFs. A. Sankaranarayanan and A. Veeraraghavan, 2009.

Solving under the HSS model

L2 norm for the subspace compressible components

L1 norm for the sparse components

$$\min_{\mathbf{s}} \|\mathbf{s}_p\|_2 + \|\mathbf{s}_c\|_1$$

$$s.t.$$

$$\|\mathbf{y} - A_p\mathbf{s}_p - A_c\mathbf{s}_c\| \leq \epsilon$$

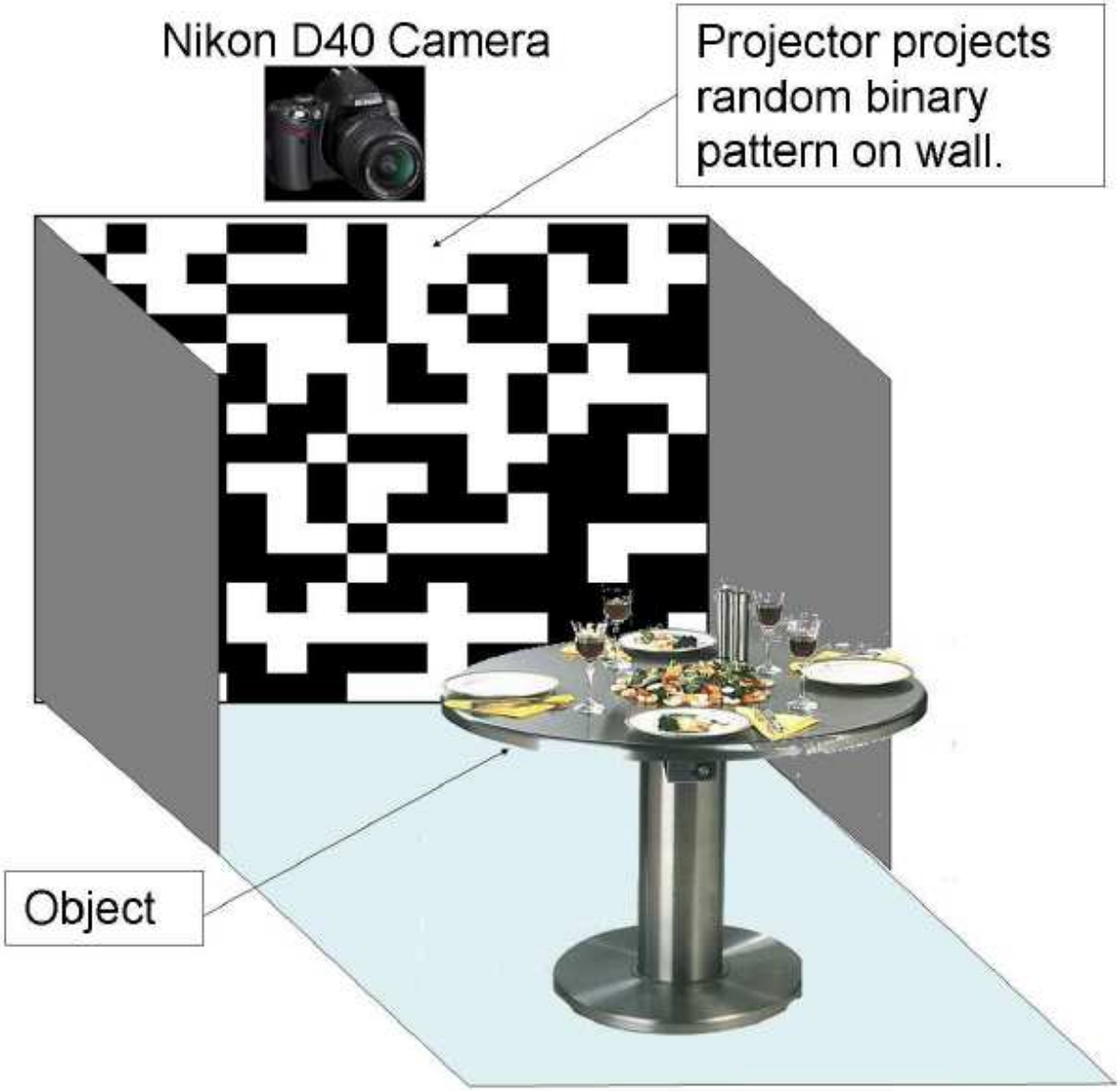
Observation error. Ap and Ac are the basis corresponding to the subspace and sparse components respectively

A. Sankaranarayanan and A. Veeraraghavan, 2009.

Advantages of the HSS Model

- Typically, for a K-sparse signal we need O(K log (N/K)) measurements, which say we approximate with 4K measurements
- Under a HSS model with K1 subspace components and K2 sparse components, we would need (K1 + 4*K2) measurements, as opposed to 4*(K1+K2) as required for a naïve CS scheme.
- Equivalently, this can be recast as better reconstruction SNR for a given compression ratio.
- CONS: Need to know subspace components apriori. This may require domain knowledge or lots of data from which the subspace and sparse projection can be learnt.
- A. Sankaranarayanan and A. Veeraraghavan, 2009.

Compressive acquisition



Compressive measurements

A. Sankaranarayanan and A. Veeraraghavan, 2009.

Scene relighting Results (RF was acquired using 192 measurements)

Relighting pattern

A. Sankaranarayanan and A. Veeraraghavan, 2009.

Other works not discussed

- CS for background subtraction (Cevher et al, ECCV 2008).
- CS and particle filters (Carin, Workshop on Stat. Signal Proc., 2009)
- 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).
- CS for SAR imaging (Patel et al, ICIP 2009).
- …

Remarks

- Compressive sensing cannot and will not solve all computer vision problems.
- May provide better solutions in some cases.
- Renewed interest in *l*-1-based methods!
- Compressive sensor design will lead the way.
  - Coded aperture imaging (MERL)
  - 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!
- Riding another wave!