IT 530: Image Representation and Analysis

Course Instructor: Ajit Rajwade

Autumn 2012

M Tech (semester 3)/PhD elective

Course Website

- http://intranet.daiict.ac.in/~ajit_r/IT530_Autumn2012.html

Pre-requisites

- Background in image processing AND machine learning (student should have taken the M. Tech. courses in Digital Image Processing and Pattern Recognition at DAIICT).

- In case of questions about pre-requisites, ask me.

Course Objectives

- To survey contemporary topics in image representation and analysis
- To broaden students' knowledge in the allied areas of image processing, computer vision and machine learning
- To (hopefully) stimulate research ideas
- This course is NOT an introductory course in image processing or computer vision

Evaluations

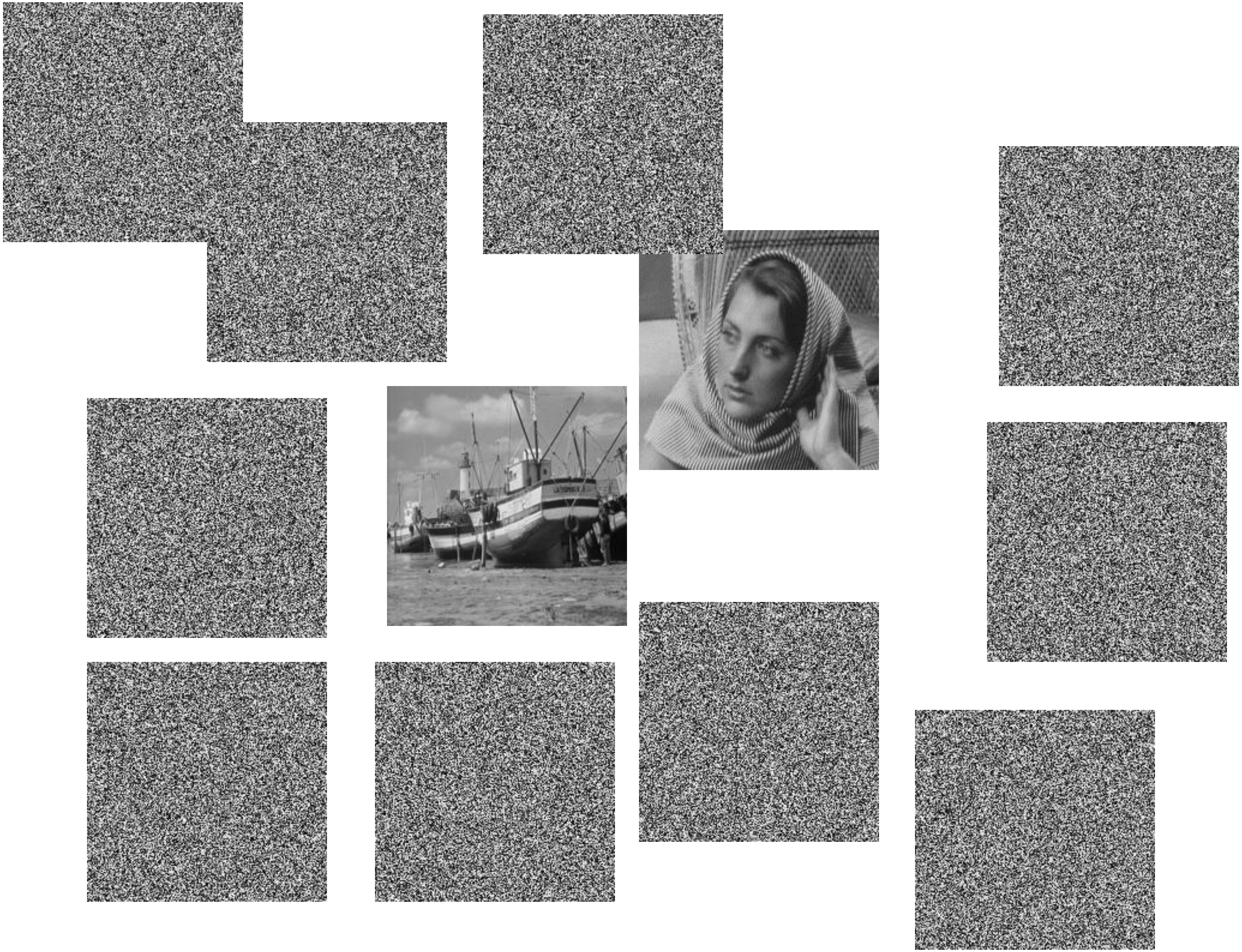
- Each student to present a set of 2-3 related papers. Duration of presentation ~ 1 hour. Everyone is encouraged to ask questions and have an informal discussion.
- Quality of presentation and clarity of answers: 60%
- Final examination (written): 20%
- Two/three small assignments involving MATLAB programming: 20%
- Students are expected to read papers assigned as reading material for lectures, as well as papers marked out for student presentations.

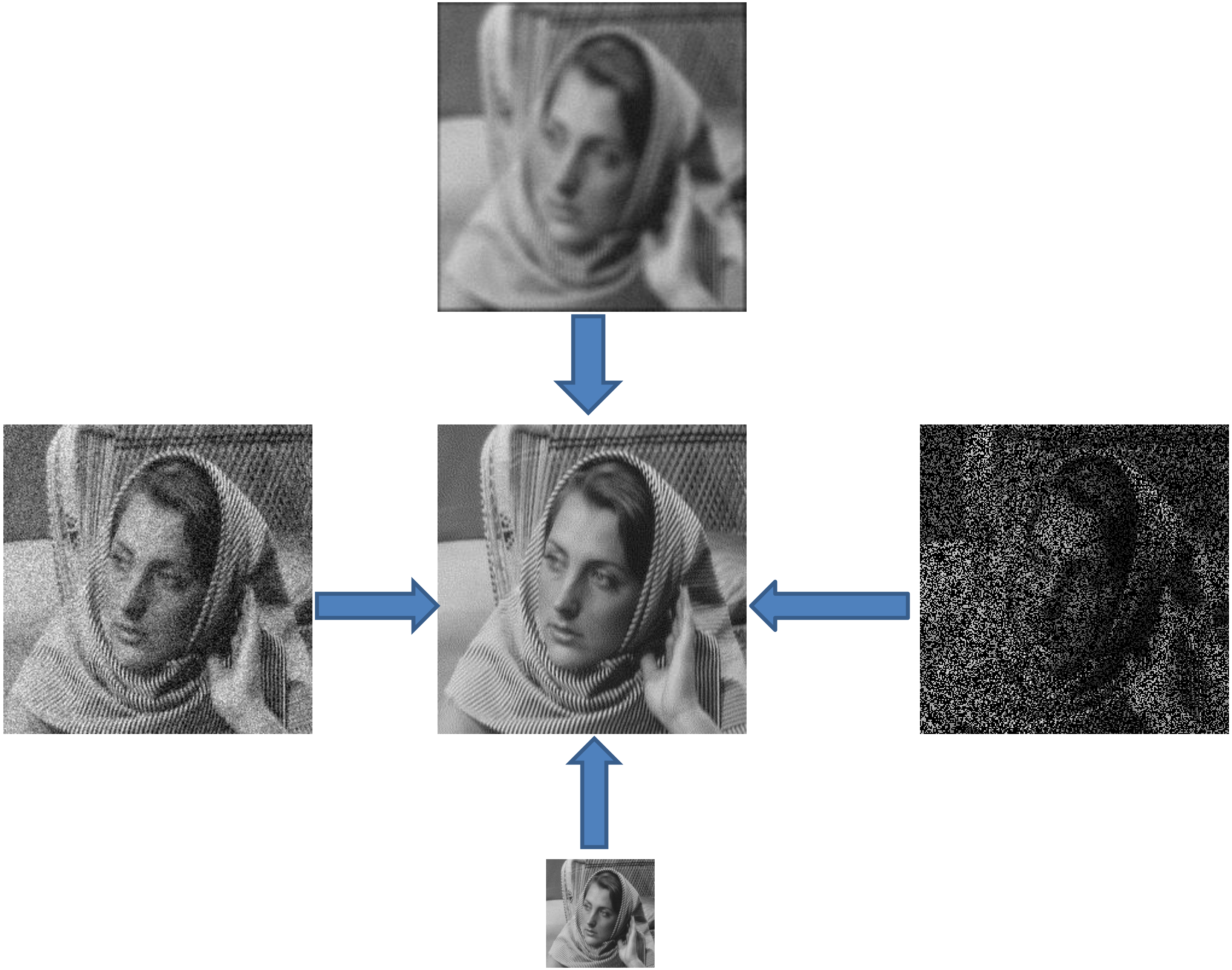
Tentative List of Course Topics

- **Statistics of natural images** with applications in image denoising, compression, etc.
- **Dictionary learning** (techniques of efficient image representation): Applications in denoising, inpainting, deblurring and classification
- **Compressive Sensing** and practical compressive imaging systems
- Non-local self-similarity of images

Statistics of Natural Images

- Number of possible 200 x 200 images (of 255, i.e. 8 bit intensity levels) = 256^40000 = 2^320000 = 10^110000.
- This is several trillion times the number of atoms in the universe (10^90).
- Only a tiny subset of these are plausible as natural images.





What is Compressive Sensing?

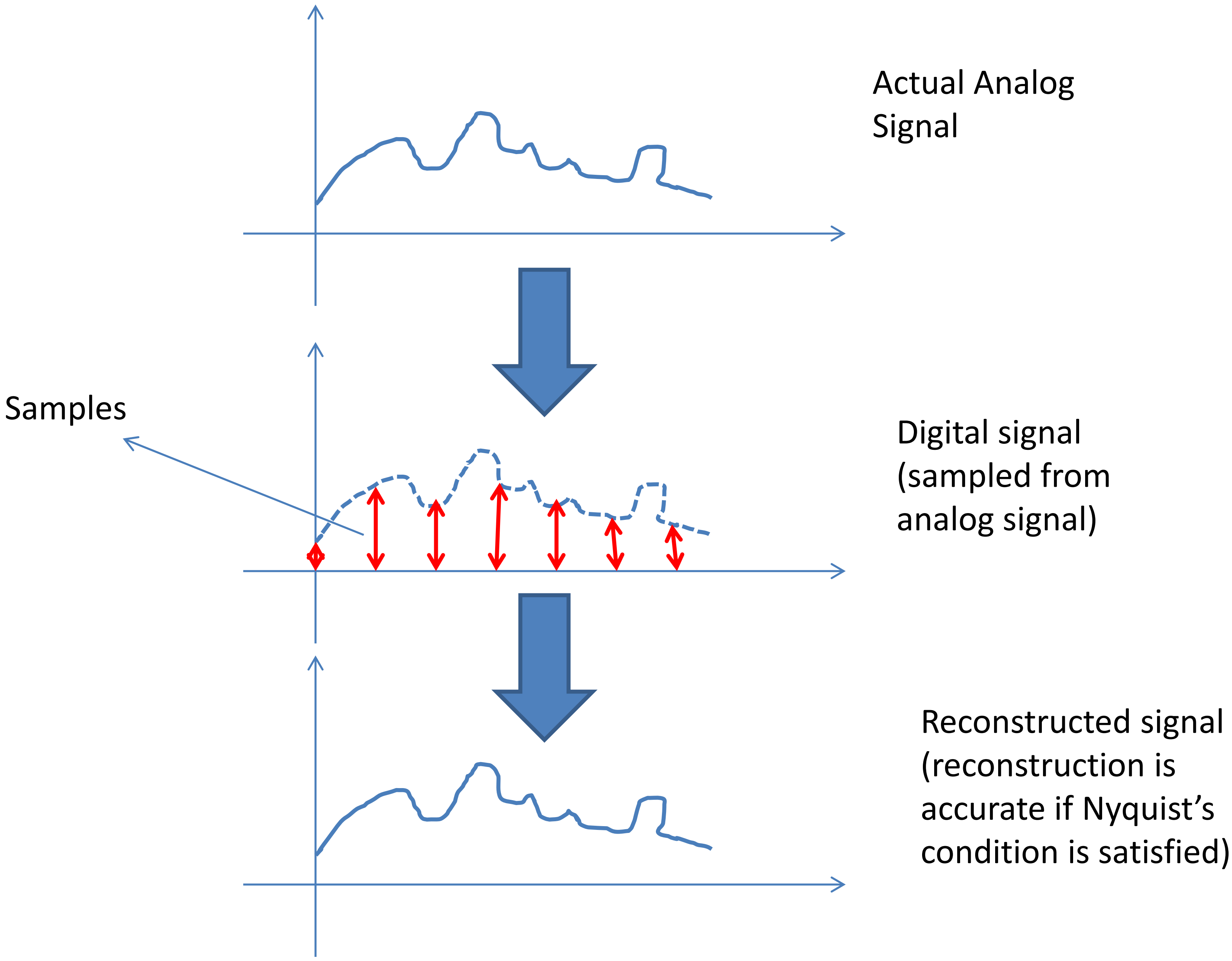
- Standard imaging devices first acquire images and then compress them (i.e. discard the remaining portion).
- E.g.: a standard digital camera captures a full image (N x M pixels) and then compresses it using (say) JPEG algorithm.
- Wasteful! Why not acquire images already in a compressed format?

What is Compressive Sensing?

- Numerous applications: MRI, acquisition of videos, acquisition of hyperspectral images, motion deblurring in videos, and many others.
- Very contemporary and popular topic in signal and image processing.

Some technical details about Compressed Sensing

- Goes beyond the **Nyquist rate** of sampling (as dictated by Shannon's Sampling theorem)
- **Shannon's theorem:** A band-limited signal with maximum frequency ***B*** can be reconstructed from its samples provided the sampling rate is more than ***2B*** samples per second (i.e. Nyquist rate).
- The above reconstruction is accurate.
- Reconstruction proceeds using the *sinc interpolant* (Whittaker-Shannon interpolation formula)


Actual Analog
Signal
Samples
Digital signal
(sampled from
analog signal)
Reconstructed signal
(reconstruction is
accurate if Nyquist's
condition is satisfied)

Compressive sensing: some applications

- Note – for a band-limited signal, you can do perfect reconstruction with just 2B samples (finite number). Without the assumption of band-limitedness, polynomial interpolation will require many more samples for perfect reconstruction.
- But with CS, you can do perfect reconstruction with **much fewer samples** than dictated by Shannon's theorem (we will see later on how).

CS theory allows you to acquire a 512 x 512 image in the form of a tiny array of numbers – each number being the dot product of the image with a **randomly generated code.** This technology is called as the **single pixel camera**.

Original image

Reconstruction with 20%

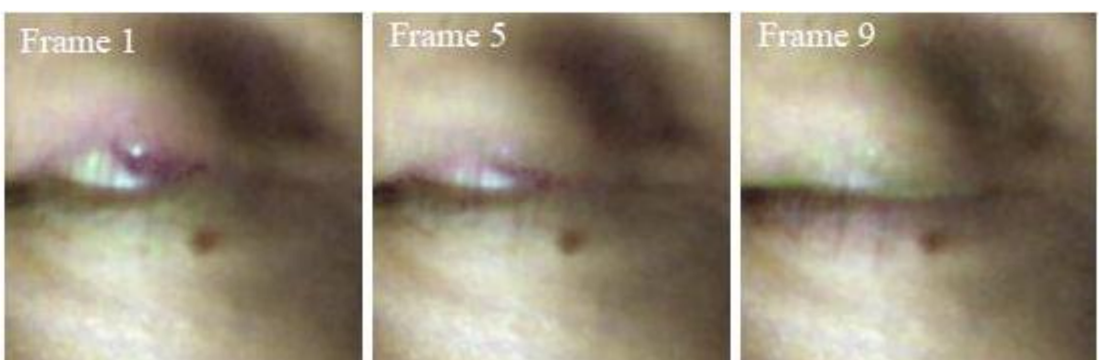
Reconstruction with 40%

http://dsp.rice.edu/cscamera

Imagine you take a 60 fps video-camera. CS theory allows you to make simple changes to the hardware of the camera to increase its frame rate from 60 fps to 540 fps.

Snapshot measured by a CS-based video-camera (in unit time)

In the same amount of time, a standard video-camera will measure one frame, which is a simple average of 9 frames – causing motion blur

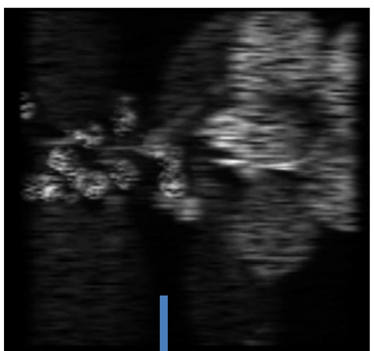


Reconstructed Frames (3 out of 9)

Reconstructed frames obtained using the measured snapshot and a simple CS recovery algorithm

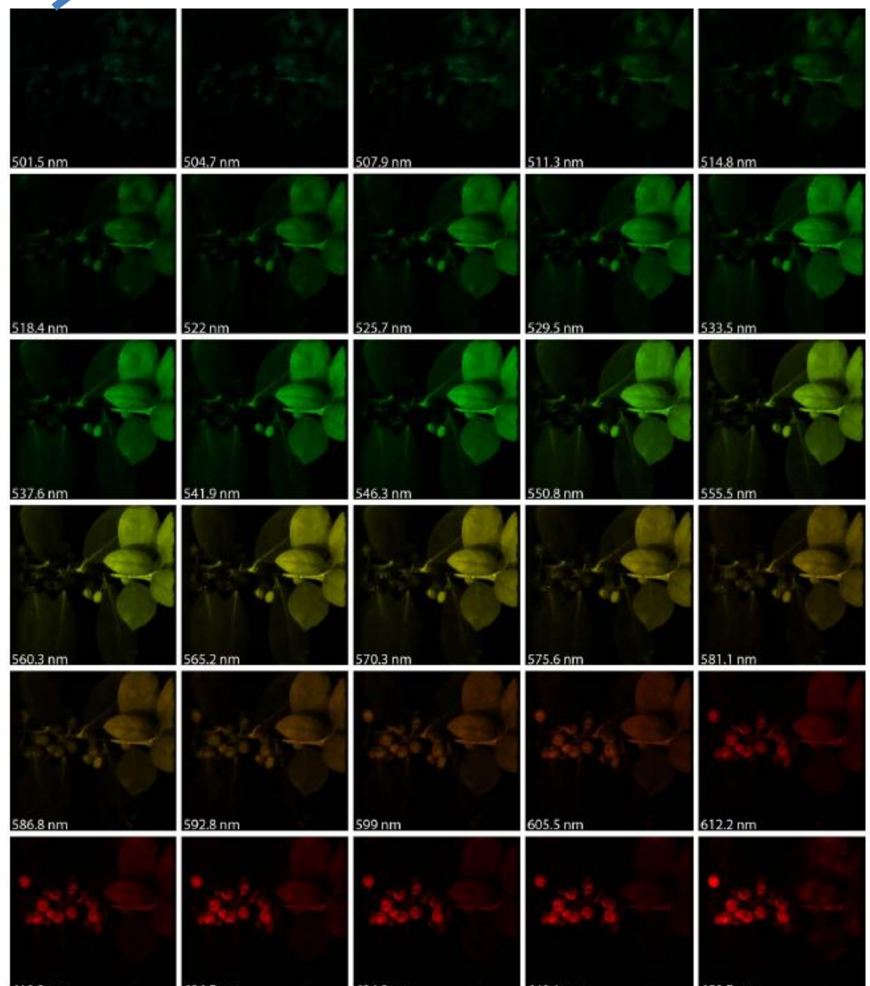
Compressive Acquisition of Hyper-spectral Images

Color (RGB) image of the scene



This is what the camera measures – a single 2D coded snapshot image (a coded superposition of single-channel images)

Reconstruction of the 46 channels of the hyperspectral datacube using the 2D snapshot and a simple recovery algorithm



Fourier Transform

$$F_k = \sqrt{N} \sum_{n=0}^{N-1} f_n e^{-i2\pi kn/N}$$

k-th Fourier coefficient

Fourier coefficient

$$f_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F_k e^{i2\pi kn/N}$$

Value of signal ***f*** at location *n* (***f*** is a vector of size *N*)

(complex) sinusoidal signals

$$f = HF$$

Vector of *N* Fourier coefficients

N x *N* orthonormal matrix (Fourier basis matrix)

Fourier Basis Matrix

$$f = HF \qquad\qquad HH^T = H^T H = I$$

$$\begin{bmatrix} x_0 \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} e^{i2\pi(0)(0)/N} & e^{i2\pi(1)(0)/N} & & e^{i2\pi(N-1)(0)/N} \\ e^{i2\pi(0)(1)/N} & e^{i2\pi(1)(1)/N} & & e^{i2\pi(N-1)(1)/N} \\ & & & \\ & & & \\ & & & \\ e^{i2\pi(0)(N-1)/N} & e^{i2\pi(1)(N-1)/N} & & e^{i2\pi(N-1)(N-1)/N} \end{bmatrix} \begin{pmatrix} F_0 \\ F_1 \\ \cdot \\ \cdot \\ \cdot \\ F_{N-1} \end{pmatrix}$$

n varies across rows (constant over all entries in a given row). *k* varies across columns (constant across all entries in a given column)

2D Fourier Transform

$$F_{uv} = \sqrt{M} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f_{xy} e^{-i2\pi(xu+yv)/M}$$

$$M = N^2$$

F and ***f*** are vectors of length M = NxN

$$f_{xy} = \frac{1}{\sqrt{M}} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} F_{uv} e^{i2\pi(xu+yv)/M}$$

$$f = HF$$

Vector of size M x 1

Matrix of size M x M

Some technical details about Compressed Sensing

- The only assumptions about the signal made by Shannon's theorem: signal is **band-limited** and the **samples are uniformly spaced**.
- But many natural signals (esp. images) have other nice properties – they are sparse in well-known bases like wavelets.

Wavelet basis

Signal/Image $\longleftarrow$ $$y = W\theta$$ $\longrightarrow$ Wavelet coefficients of *y* – most of which turn out to be zero

- Exploiting this property allows PERFECT signal reconstruction with **much fewer samples** than dictated by Shannon's theorem.

Compressed Sensing: What will this course do?

- Enlist key results related to compressive sensing.
- Work through the detailed proof of one key result.
- Study applications – practical compressive sensing systems.
- If time permits, study some extensions (e.g. matrix completion and robust principal components analysis).