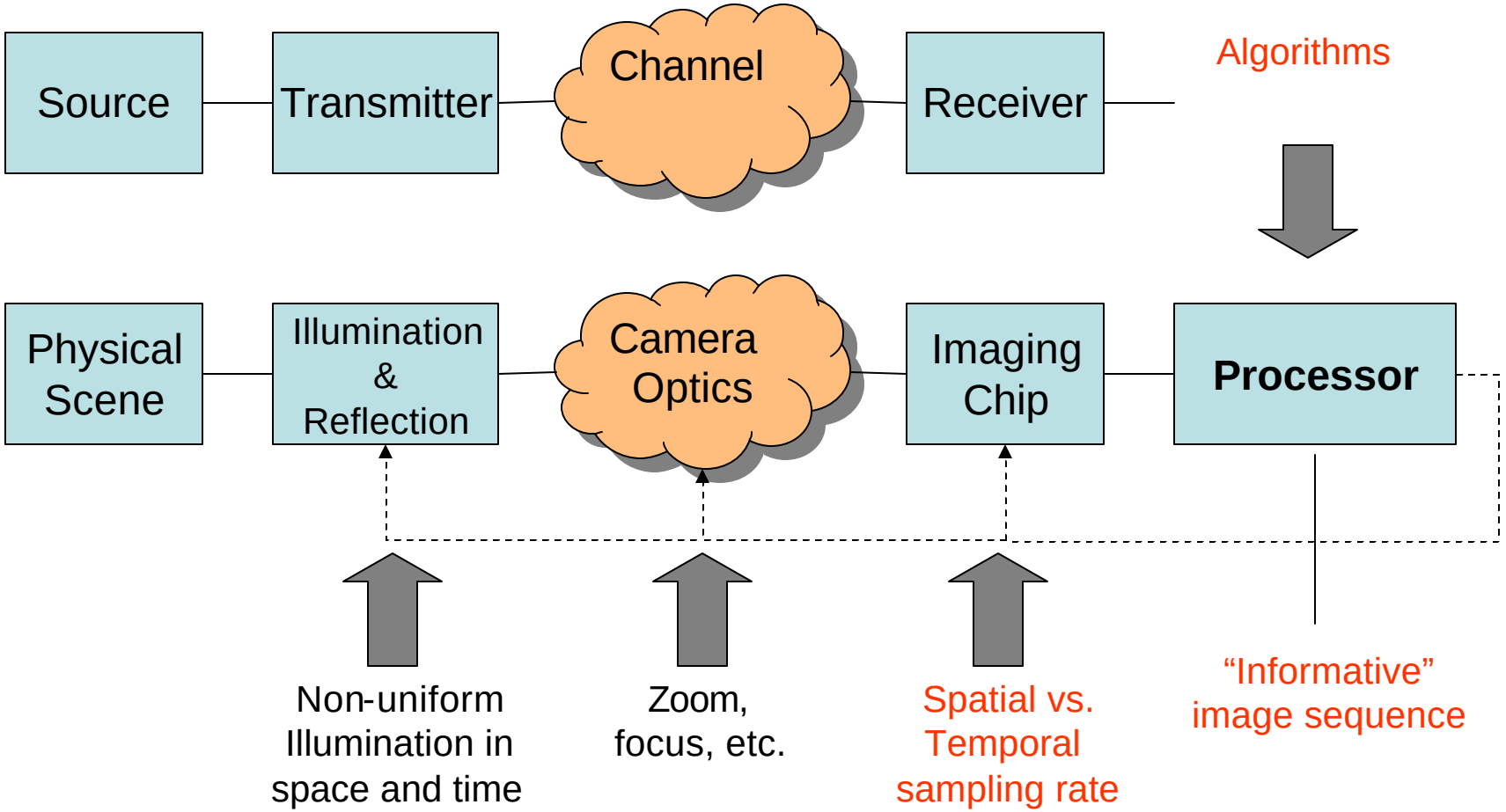
Toward Integration of Sensing and Processing for Improved Resolution

Peyman Milanfar

EE Department
University of California, Santa Cruz
milanfar@ee.ucsc.edu
http://www.soe.ucsc.edu/~milanfar

AFOSR Program Review, June 3, 2004

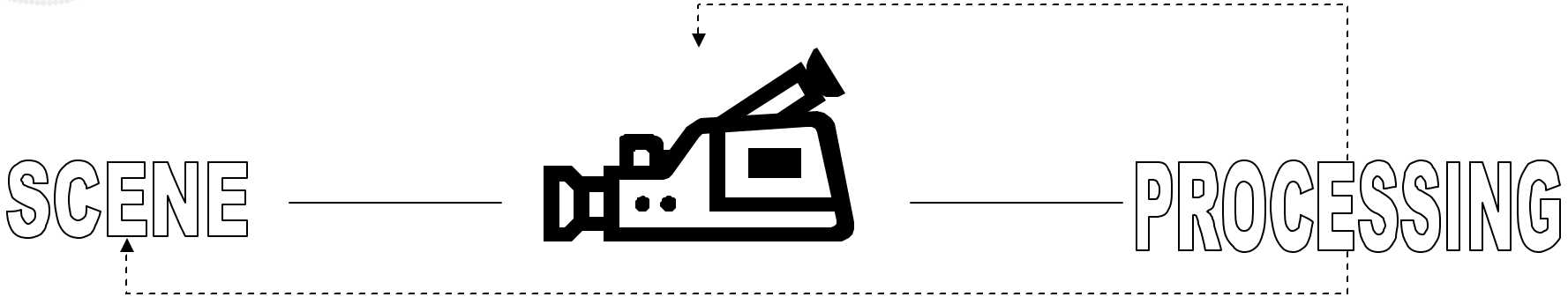
Information and Feedback for Smart Imaging



Some Informative Analogies

- Imaging (Inv. Probs.)
  - Point-spread function
  - Deconvolution
  - Occlusion
  - “Multi-frame” imaging
  - Image registration
  - Resolution Limits
  - ….
- Communication
  - Channel response
  - Equalization
  - Interference
  - Multi-antenna systems
  - Time-delay estimation
  - “Capacity”
  - ….

Key Questions to Consider



What are the limits to the performance of

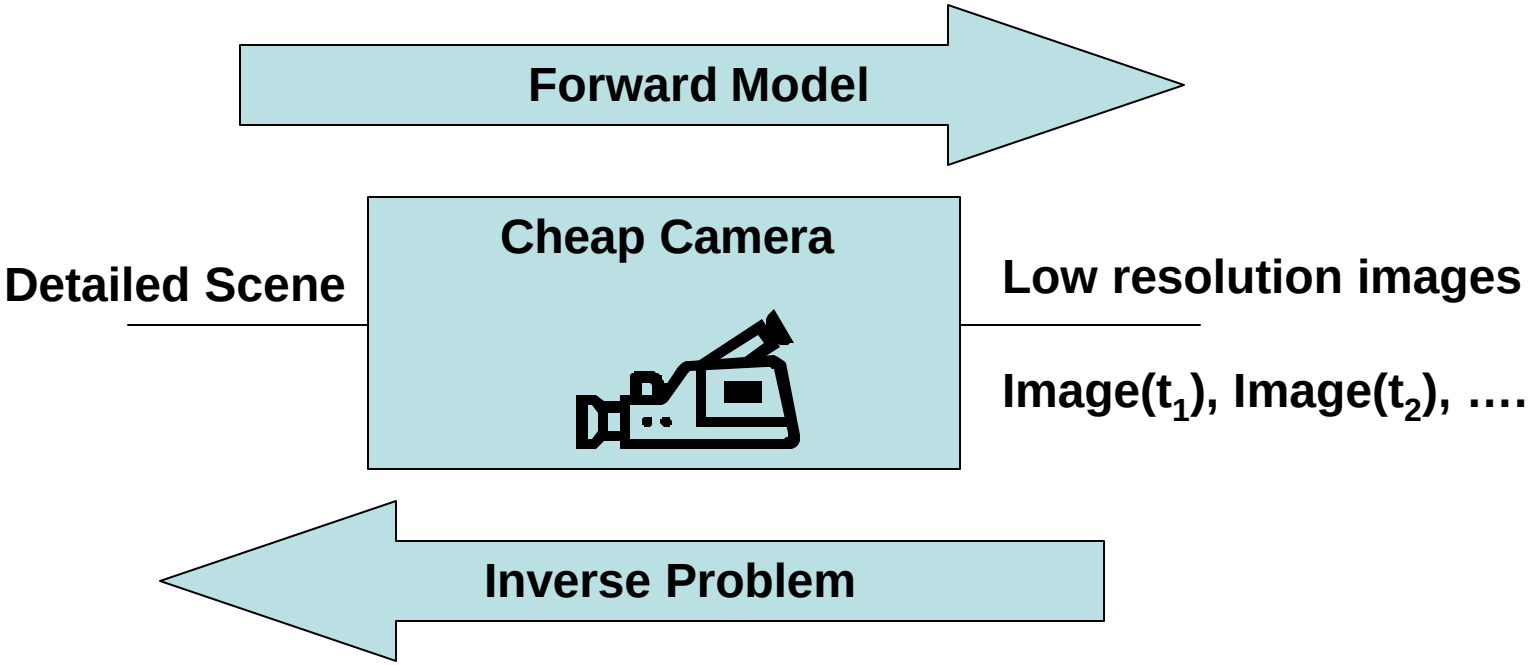
1. The bandwidth-limited, noisy camera?
2. The processing algorithm applied to the sensor output?

Can we improve the overall performance by

1. Optimizing sensing and processing together using the bounds above?
2. Actively probing the medium of interest so as to extract the most information from the ensemble sensor/processor?

Resolution as a measure of information in images

$$L(x,y,t) = \downarrow_N \left[H(x,y,t) * PSF(x,y) \right] + noise$$

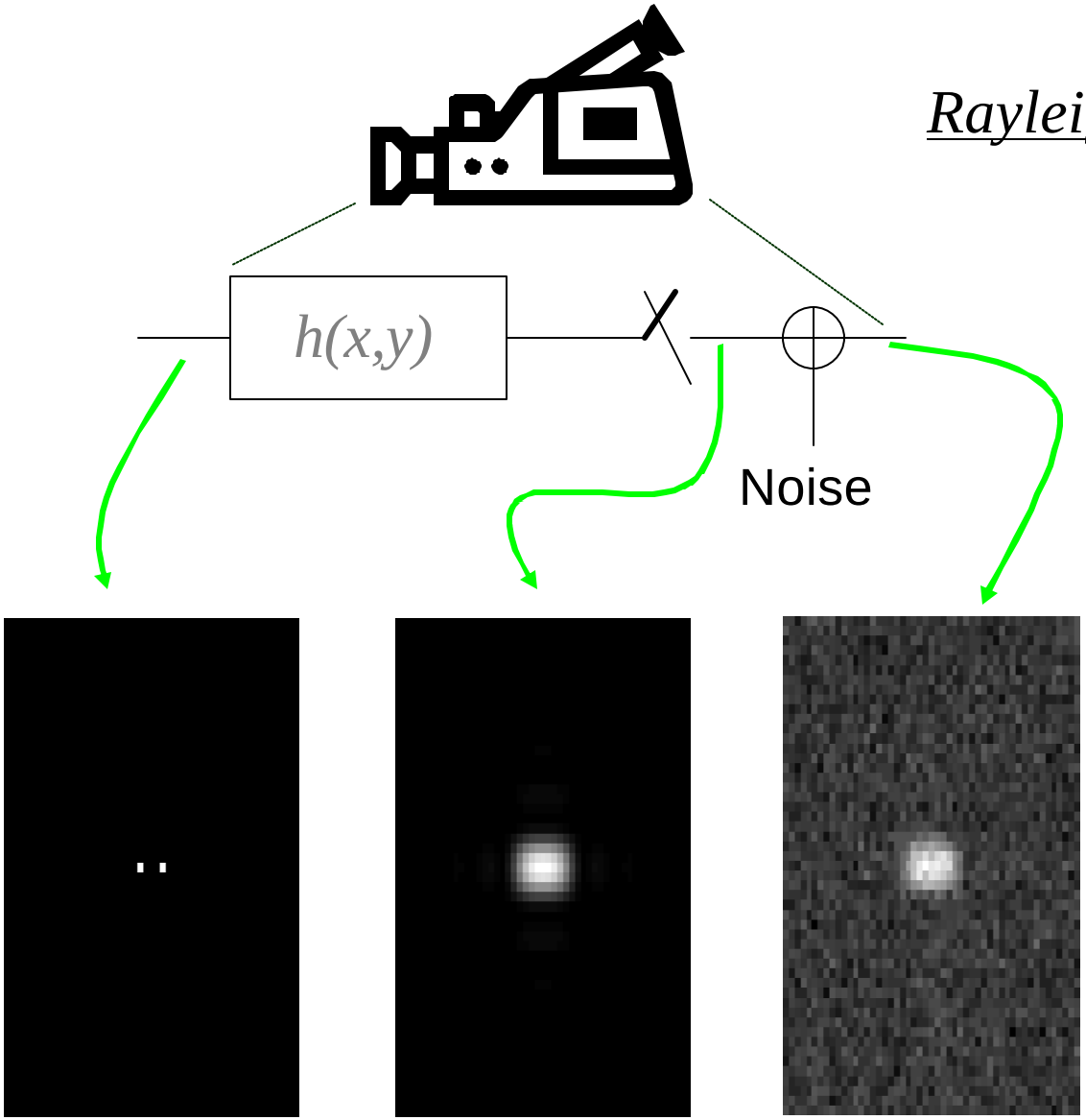


Compute $H(x,y,t)$

Sensing : Resolution Limits of a Camera

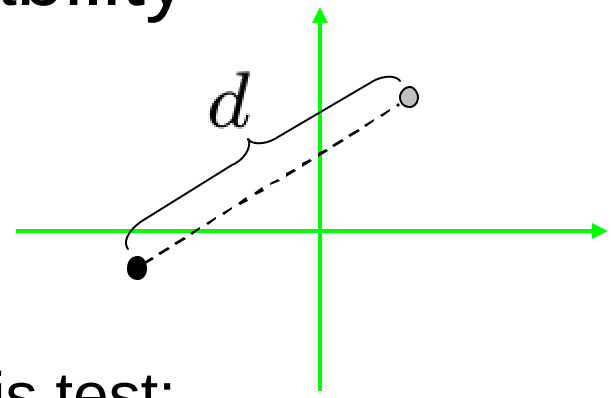
Imaging Closely-Spaced Point Sources

Rayleigh's limit isn't.



Performance Limit: Required SNR for Resolvability

- Point sources:



- Composite statistical hypothesis test:

$$\begin{cases} \mathcal{H}_0 : d = 0 \\ \mathcal{H}_1 : d > 0 \end{cases} \Rightarrow \text{SNR} \approx \frac{C}{N^2 d^4}$$

$$\begin{cases} \mathcal{H}_0 : \alpha = \beta \\ \mathcal{H}_1 : \alpha \neq \beta \end{cases} \Rightarrow \text{SNR} \approx \frac{C_2}{N^2 (\alpha - \beta)^2 d^2}$$

Constants depends on the camera, the required false alarm and detection rates. (Optimize!)

- GLRT is the UMP test here.

Many Extensions

- Multi-frame scenarios:
  - Motion Estimation
- More general camera models
- Perturbation Analysis.
  - How does the scaling law change if model if there exists model mismatch?

Processing :Multi-frame Resolution Enhancement (Super-resolution)

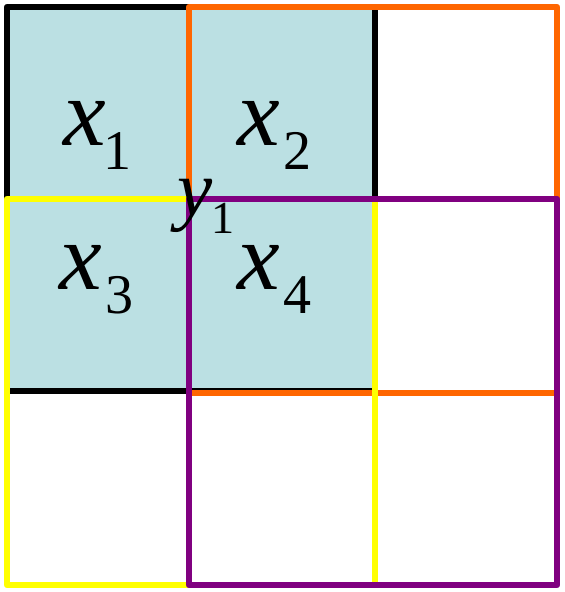
Resolution Enhancement Idea

• Given multiple low-resolution moving images of a scene (a video), generate a high resolution image (or video).

• DIVERSITY: “Collect multiple views to gain spatial resolution”

Resolution Enhancement Model

- A simple model relating the low-resolution blurry image to the high resolution crisper image.



"PSF"

$$y_1 = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + e_1$$

$$y_2 = 0 \cdot x_1 + a_1 x_2 + 0 \cdot x_3 + a_3 x_4 + e_2$$

$$y_3 = 0 \cdot x_1 + 0 \cdot x_2 + a_1 x_3 + a_2 x_4 + e_3$$

$$y_4 = 0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + a_1 x_4 + e_4$$

The Mathematical Model

k-th frame $\underline{\mathbf{y}}_k = A_k \underline{\mathbf{x}} + \underline{e}_k \quad for \quad 1 \le k \le p$

$$A_k = DHF(v_k)$$

Downsampling — Blurring — Warping

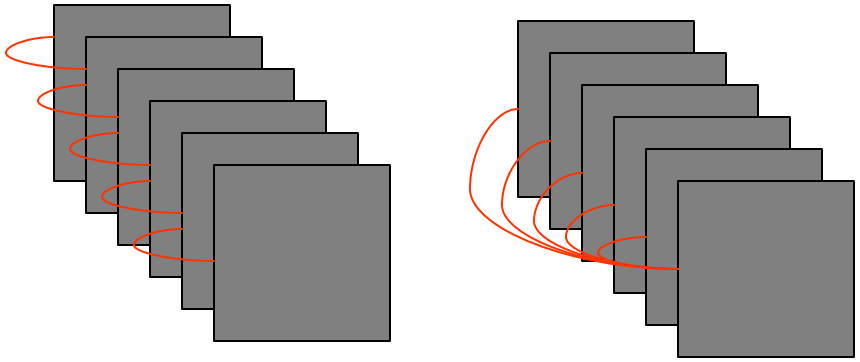
$$\underline{\mathbf{y}} = A(\underline{\mathbf{v}})\underline{\mathbf{x}} + e$$

- Statistical Estimation problem
- The system is typically underdetermined and ill-conditioned.
  - Need N^2 frames for factor of N enhancement.
- Model is uncertain, and sensitive to unknown parameters.
- Computational complexity is a major concern.

Insights from the CRB (optimize!)

$$\mathrm{Cov}\left(\begin{bmatrix}\hat{\mathbf{v}}\\ \hat{\mathbf{x}}\end{bmatrix}\right) \geq J^{-1}\left(\begin{bmatrix}\mathbf{v}\\ \mathbf{x}\end{bmatrix}\right) = \begin{bmatrix} J_{vv}^{-1} & J_{vx}^{-1} \\ J_{xv}^{-1} & J_{xx}^{-1}\end{bmatrix}$$

- Given multiple frames of video, conventional motion estimation methods are not best:
  - pairwise estimation
  - with a reference frame



- Justifies “bundle adjustment”: maximize mutual correlation between all combinations of frames.

Insights from the CRB

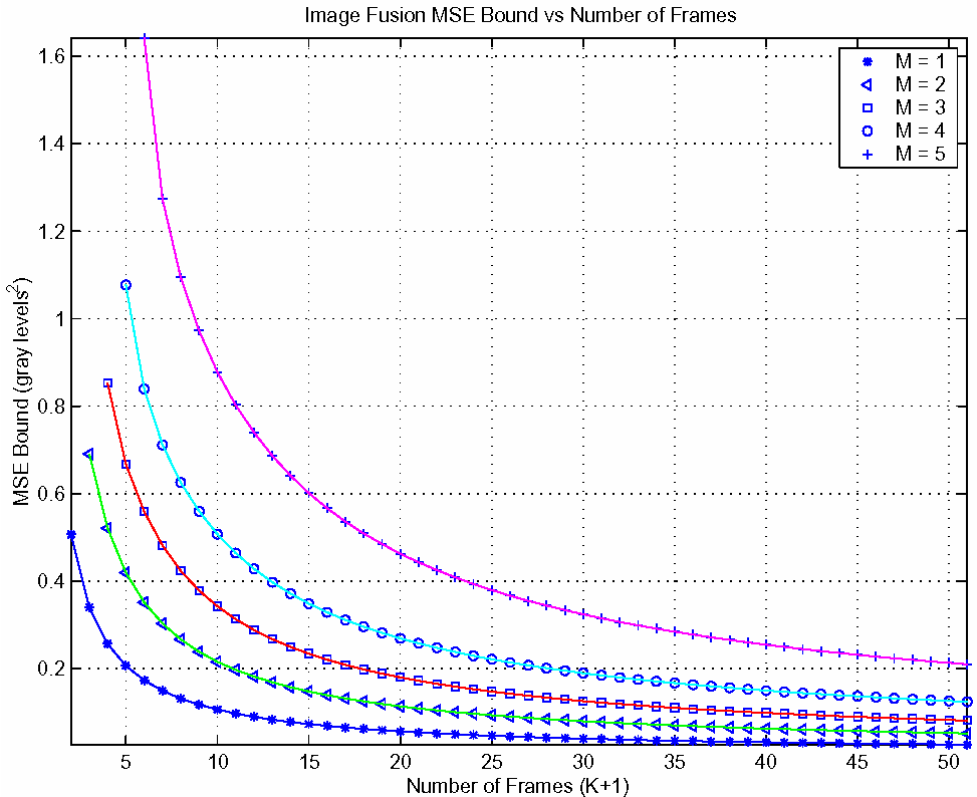
$$\mathrm{Cov}\left(\begin{bmatrix}\hat{\mathbf{v}}\\ \hat{\mathbf{x}}\end{bmatrix}\right) \geq J^{-1}\left(\begin{bmatrix}\mathbf{v}\\ \mathbf{x}\end{bmatrix}\right) = \begin{bmatrix} J_{vv}^{-1} & J_{vx}^{-1} \\ J_{xv}^{-1} & J_{xx}^{-1} \end{bmatrix}$$

- Motion estimation and image upscaling (interpolation) are deeply connected
  - Can't really do one without the other.
- Specific selection of motions matters
  - We can specify which are "best"

Insights from the CRB

$$\mathrm{Cov}\left(\begin{bmatrix}\hat{\mathbf{v}}\\ \hat{\mathbf{x}}\end{bmatrix}\right) \geq J^{-1}\left(\begin{bmatrix}\mathbf{v}\\ \mathbf{x}\end{bmatrix}\right) = \begin{bmatrix} J_{vv}^{-1} & J_{vx}^{-1} \\ J_{xv}^{-1} & \boxed{J_{xx}^{-1}} \end{bmatrix}$$

- More frames give diminishing returns.
  - How to make best use of your imaging resources?



The Optimization Problem

Data Info: Builds robustness to model uncertainty

$$\{\hat{\underline{\mathbf{x}}}, \hat{\underline{\mathbf{v}}}\} = \underset{\underline{X}}{\operatorname{argmin}} \Big[\left\| A(\mathbf{v})\underline{\mathbf{x}} - \underline{\mathbf{y}} \right\|_1 + l \sum_{l=-P,\ m=0}^{P} \sum^{P} a^{m+l} \left\| \underline{\mathbf{x}} - S_x^l S_y^m \underline{\mathbf{x}} \right\|_1 \Big]$$

L1 Prior: Incorporates multiscale model of edges

$0 < a < 1$

The Iterative Solution:

- The cost function can be minimized using a 2-step approach.
  - Fusion of frames: “shift and add”
  - Simultaneous deblurring and interpolation

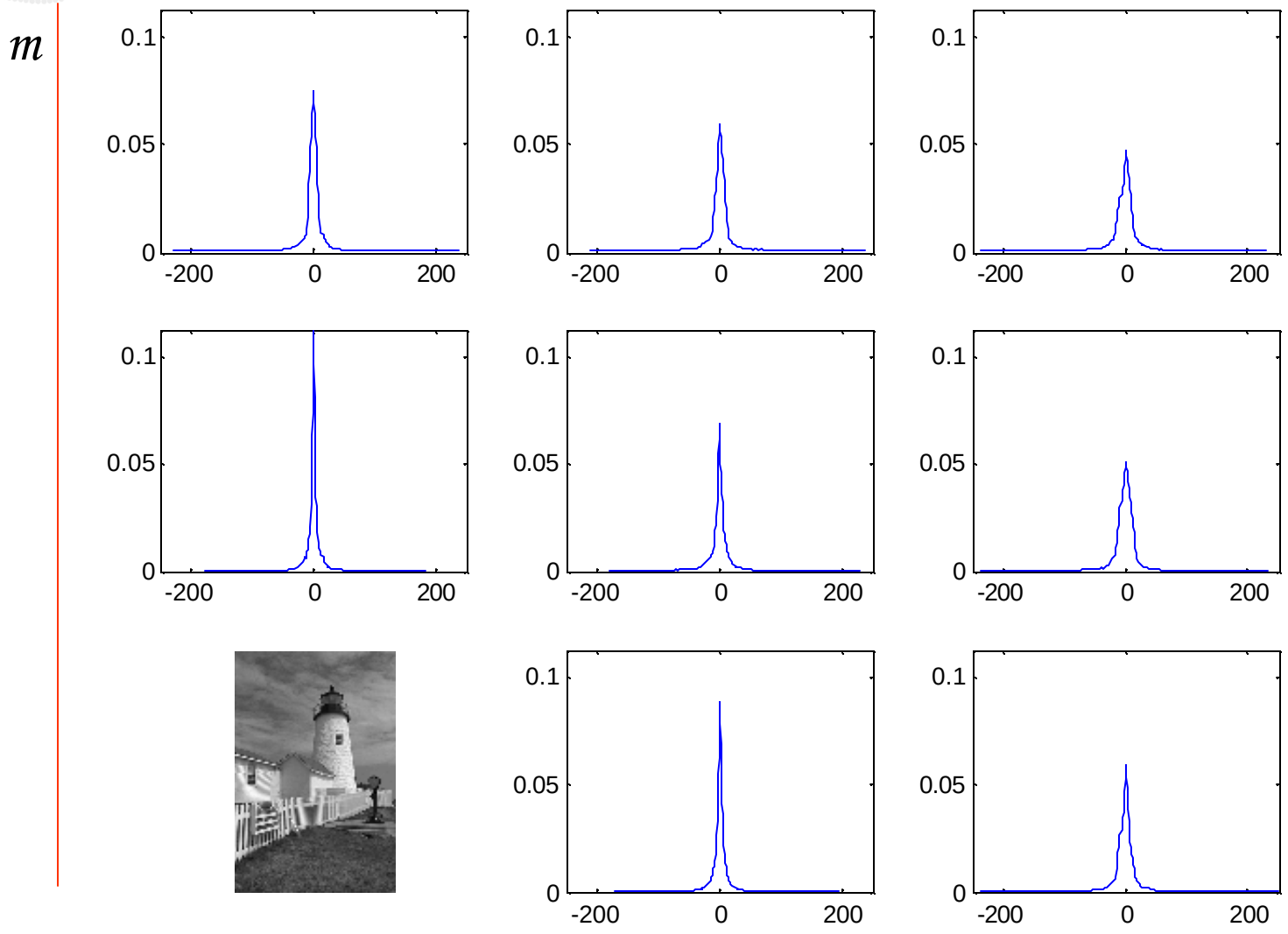
“Fast and Robust Multi-Frame Super-resolution”,
S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, accepted for publication in *IEEE Trans. on Image Processing, January 2004*

Why this L_1 prior?

$$\mathbf{e}_{l,m} = \underline{\mathbf{x}} - S_x^l S_y^m \underline{\mathbf{x}}$$

What do these look like?

Histograms of $\mathbf{e}_{l,m} = \underline{\mathbf{x}} - S_x^l S_y^m \underline{\mathbf{x}}$



Some Examples of SR

Result from Webcam

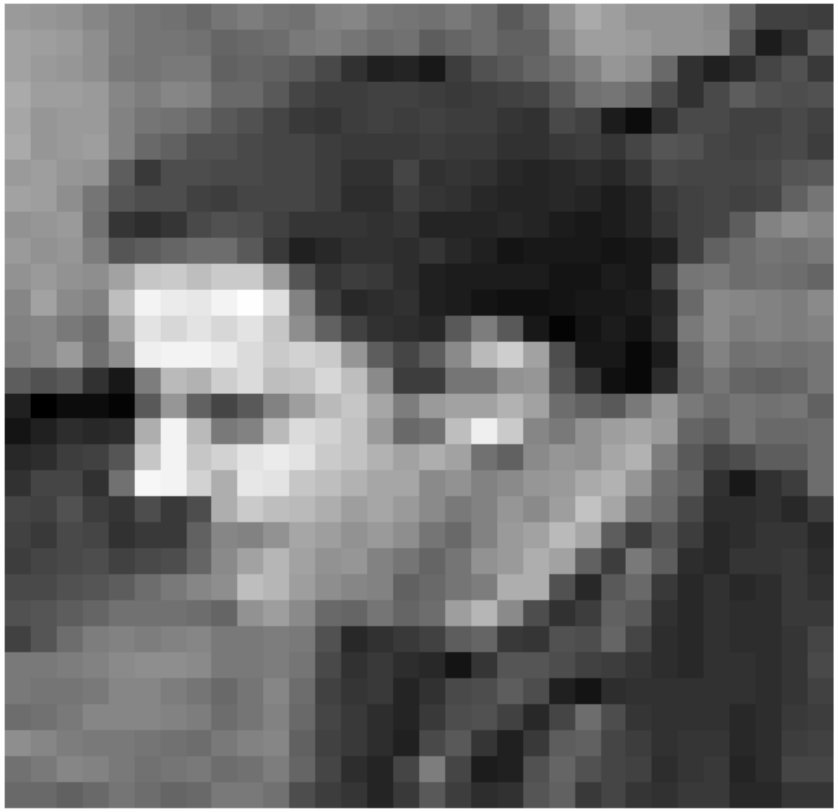
Overcoming Sensor Limitations by Processing.

MPEG Surveillance Video

Data Courtesy Vigilant Technology

Milanfar et al. EE Dept, UCSC

Resolution enhancement of a face from a sequence captured by a surveillance camera

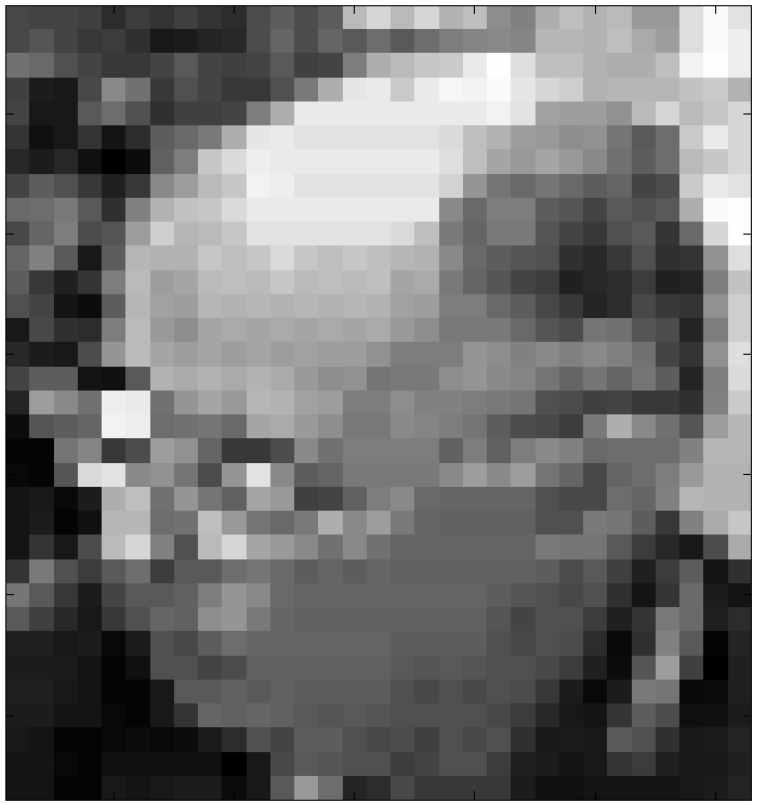


One Low-resolution Frame

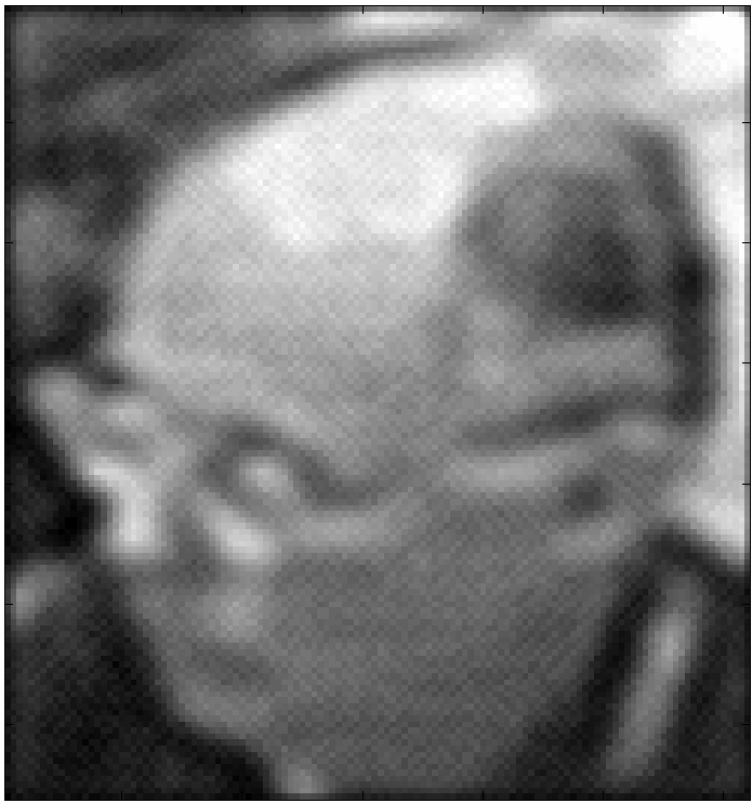
Output High-resolution Still Frame

Milanfar et al. EE Dept, UCSC

Resolution enhancement of a face from a sequence captured by a surveillance camera

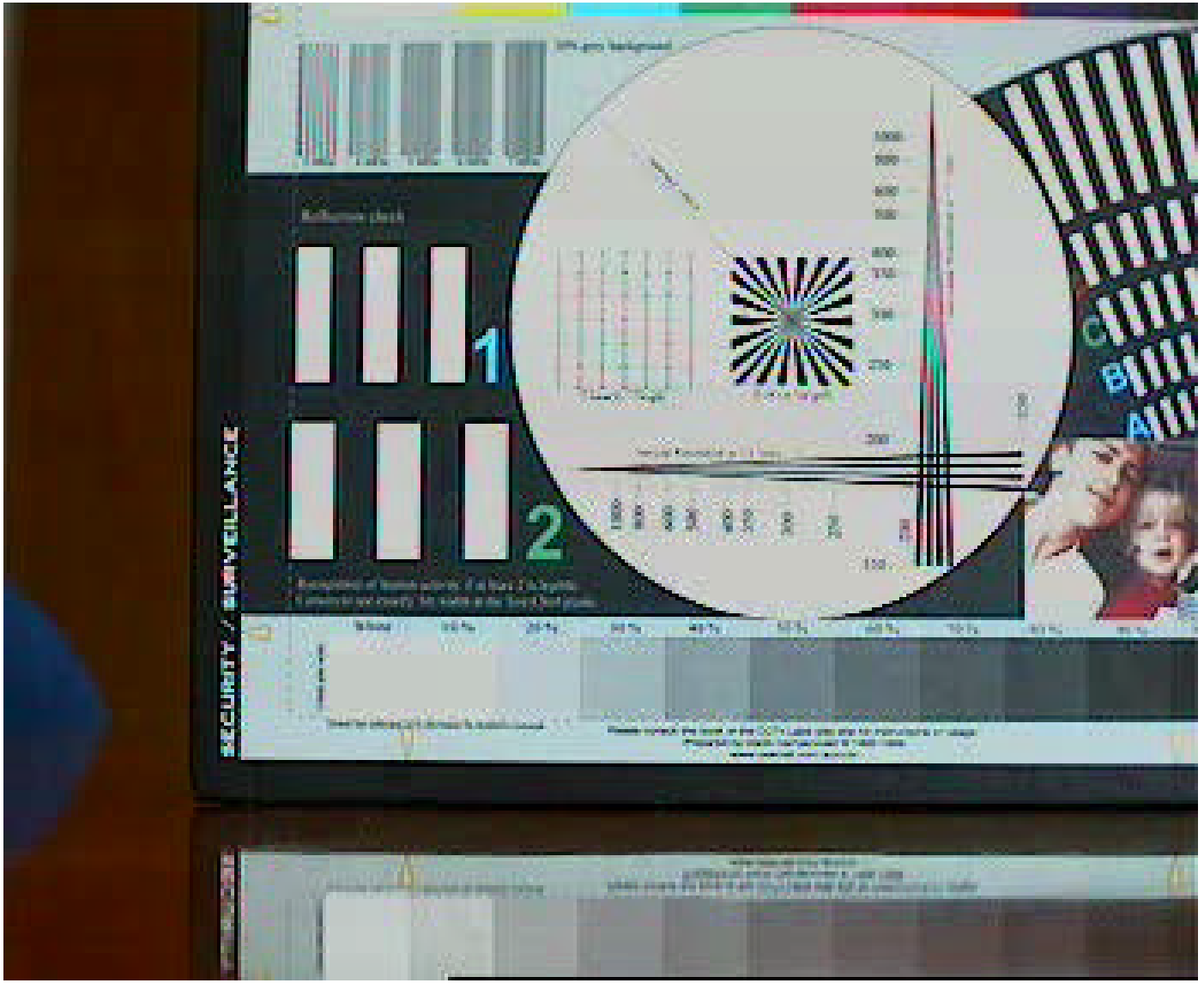


One Low-Resolution Frame



Output High-resolution Still Frame

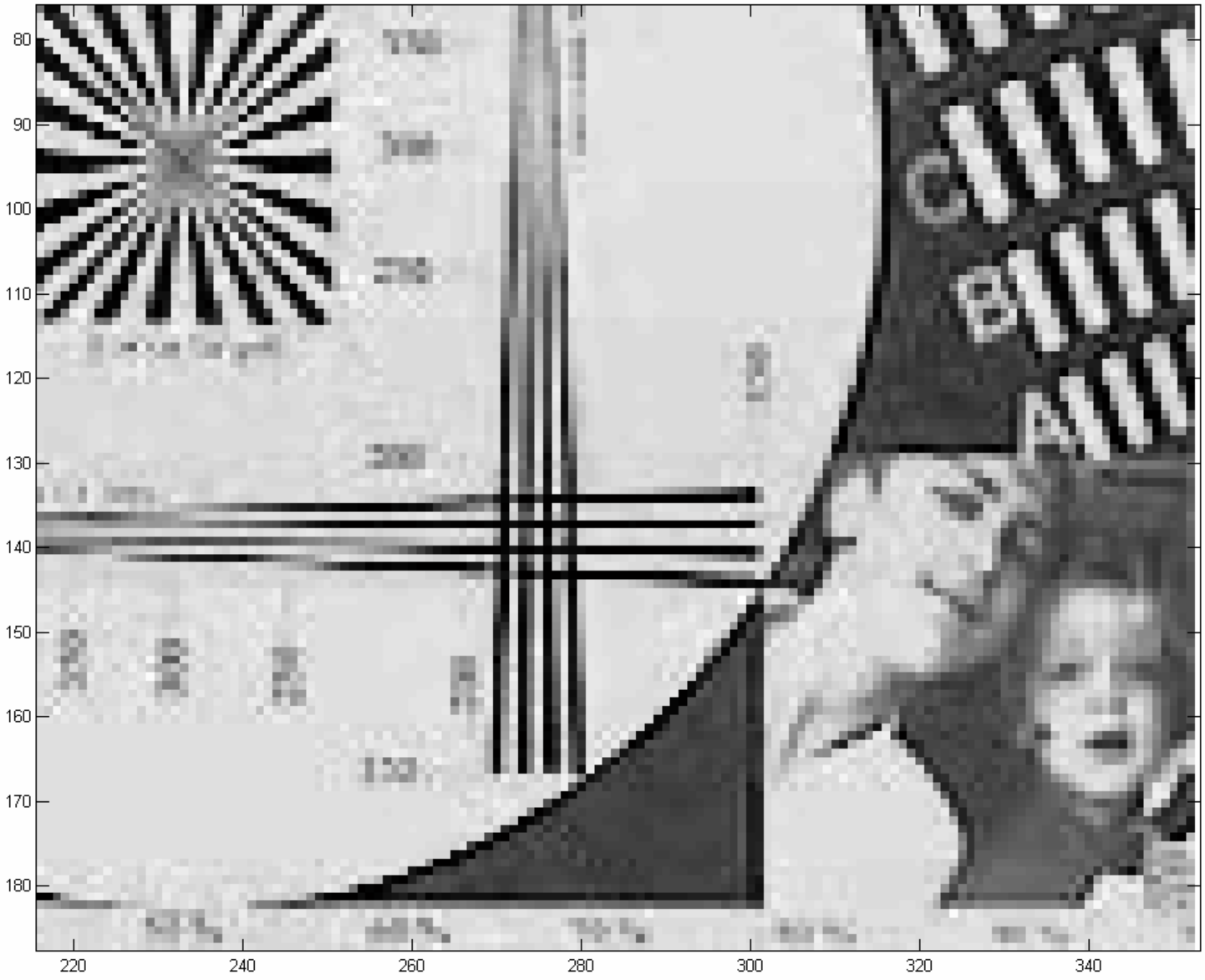
Milanfar et al. EE Dept, UCSC



Data Courtesy Vigilant Technology

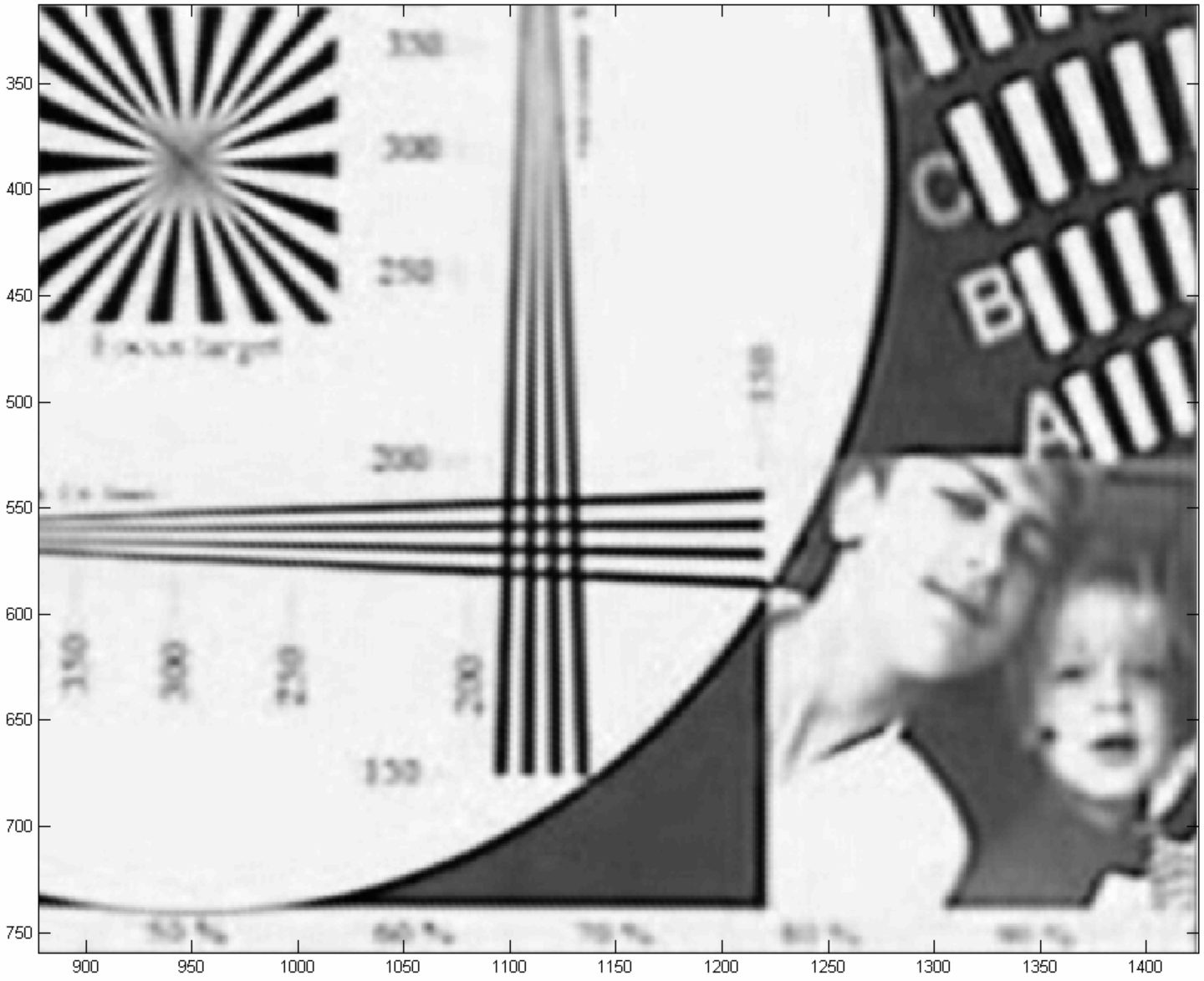
Milanfar et al. EE Dept, UCSC

Detail Before



Data Courtesy Vigilant Technology

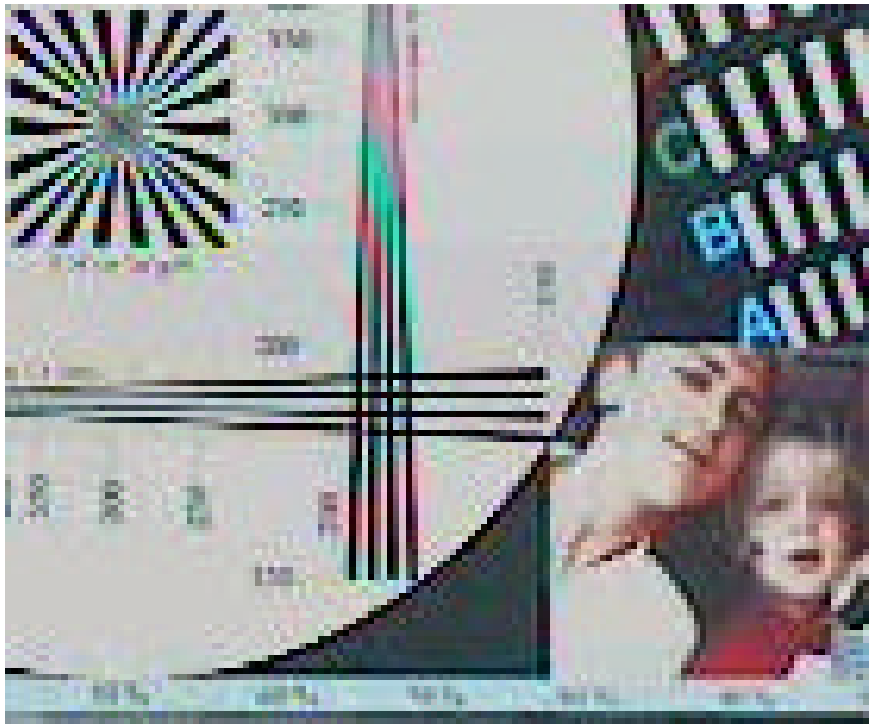
Detail After

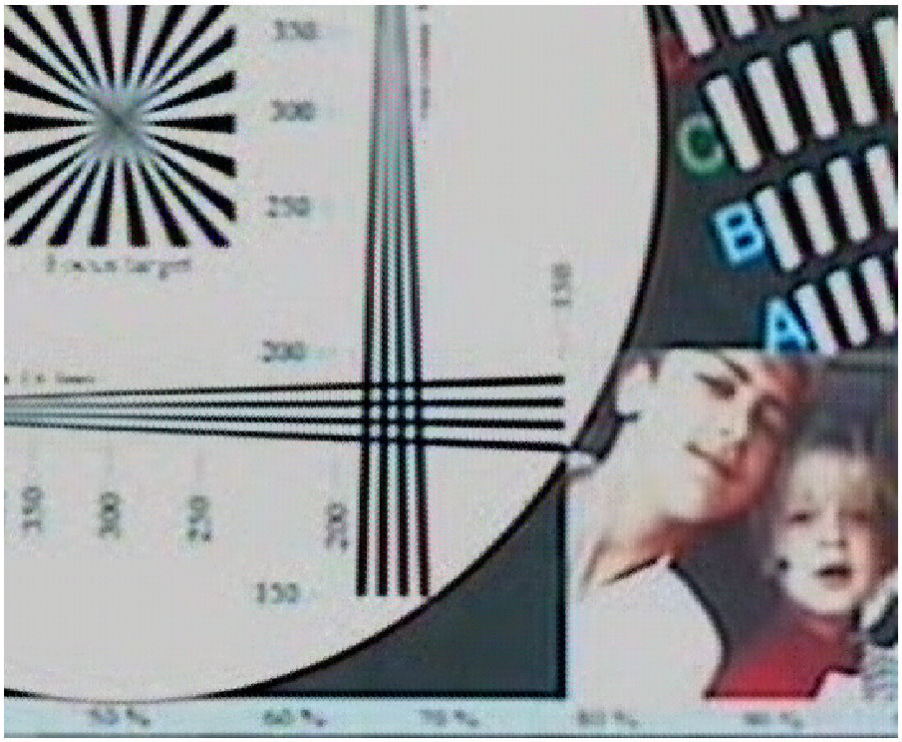


Data Courtesy Vigilant Technology

Simultaneous Color Super-Resolution/Demosaicing

G	R	G	R	G	R
B	G	B	G	B	G
G	R	G	R	G	R
B	G	B	G	B	G
G	R	G	R	G	R
B	G	B	G	B	G





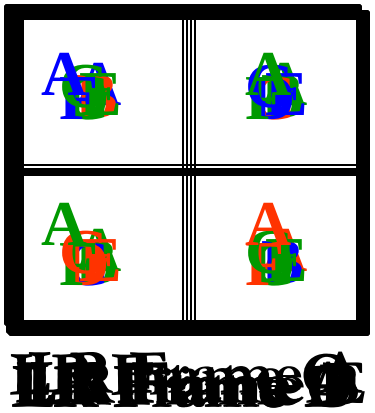
Milanfar et al. EE Dept, UCSC

Resolution enhancement in color from a video sequence captured by a Pyro 1394 webcam

One Low-Resolution Frame

Output High-resolution Still Frame

Data Fusion of Color-filtered Images



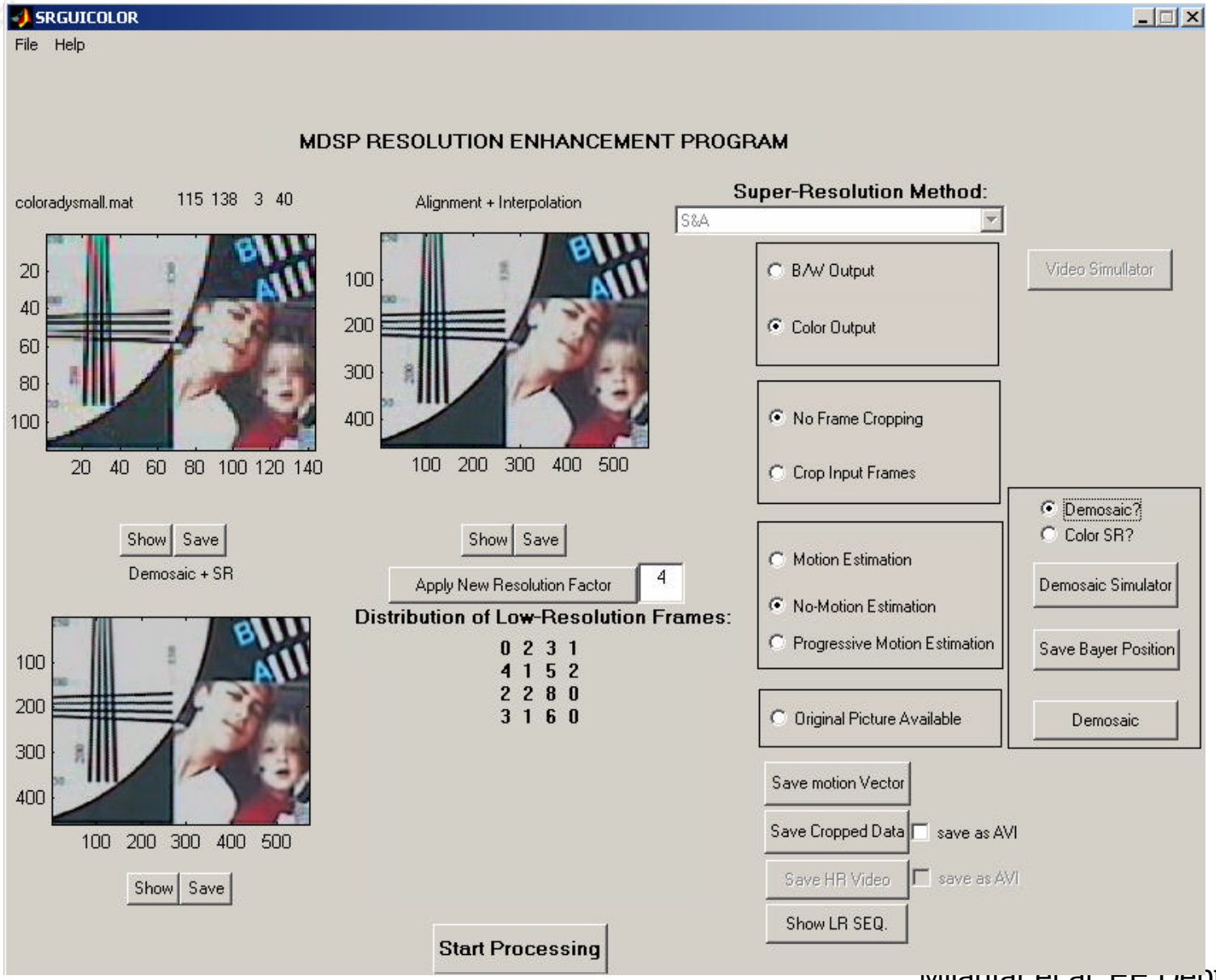
A		D	A		D
	B	E		B	E
	G C	F		G C	F
A		D	A		D
	B	E		B	E
	G C	F		G C	F

Factor of 3 Resolution Enhancement

Dynamic Super-Resolution

- With all the pieces in place, Dynamic Superres becomes possible
  - Video-to-Video

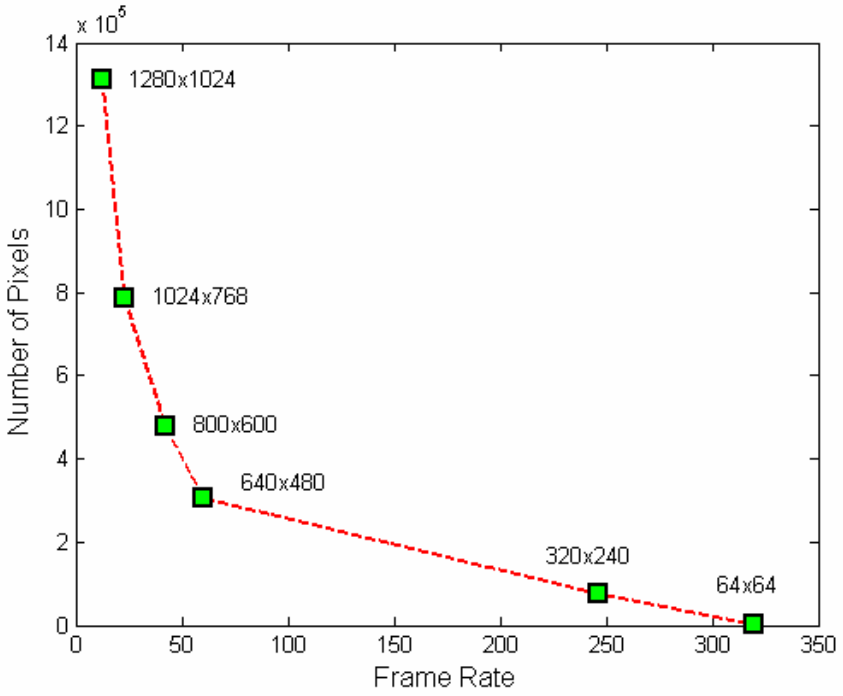
Software


SRGUICOLOR
File Help
MDSP RESOLUTION ENHANCEMENT PROGRAM
coloradysmall.mat 115 138 3 40
Alignment + Interpolation
Super-Resolution Method:
S&A
B/W Output
Color Output
Video Simulator
No Frame Cropping
Crop Input Frames
Demosaic?
Color SR?
Show Save
Show Save
Demosaic + SR
Apply New Resolution Factor 4
Motion Estimation
No-Motion Estimation
Progressive Motion Estimation
Demosaic Simulator
Save Bayer Position
Distribution of Low-Resolution Frames:
0 2 3 1
4 1 5 2
2 2 8 0
3 1 6 0
Original Picture Available
Demosaic
Save motion Vector
Save Cropped Data
save as AVI
Save HR Video
save as AVI
Show LR SEQ.
Show Save
Start Processing

Integrating Sensing and Processing

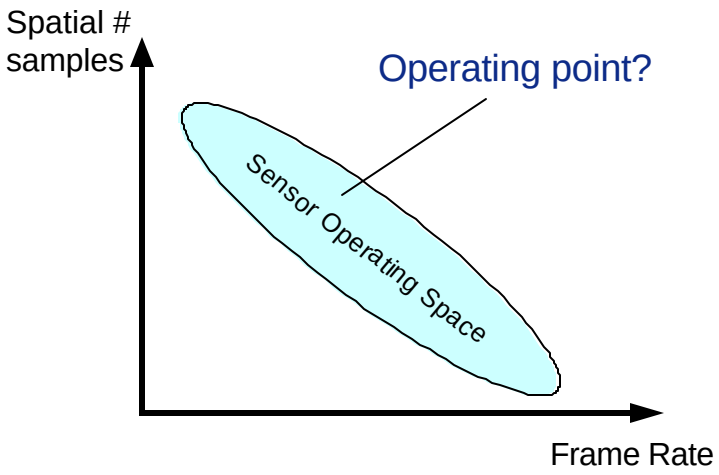
Tradeoffs and limitations

- Spatial and temporal sampling trade-off in imaging devices due to physical limitations such as:
  - Light integration time
  - Sensor timing
  - Analog circuit delays



Tradeoffs and limitations

- Adapt the sensor according to the scene for optimal usage of the available bandwidth and sensor capabilities
- Stream the data to a post processor to create high spatio-temporal resolution sequence



Measuring content effectively and efficiently

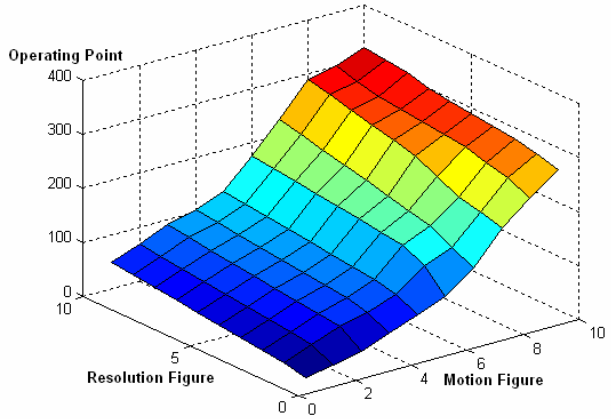
- What is an effective measure of true content (“edge-iness”) in an image that is
  - Robust to noise
  - Stable under aliasing
  - Fast to compute

$$R(\underline{\mathbf{x}}) = \sum_{l=-P,}^{P} \sum_{m=0}^{P} a^{|m|+|l|} \left\| \underline{\mathbf{x}} - S_x^l S_y^m \underline{\mathbf{x}} \right\|_1$$

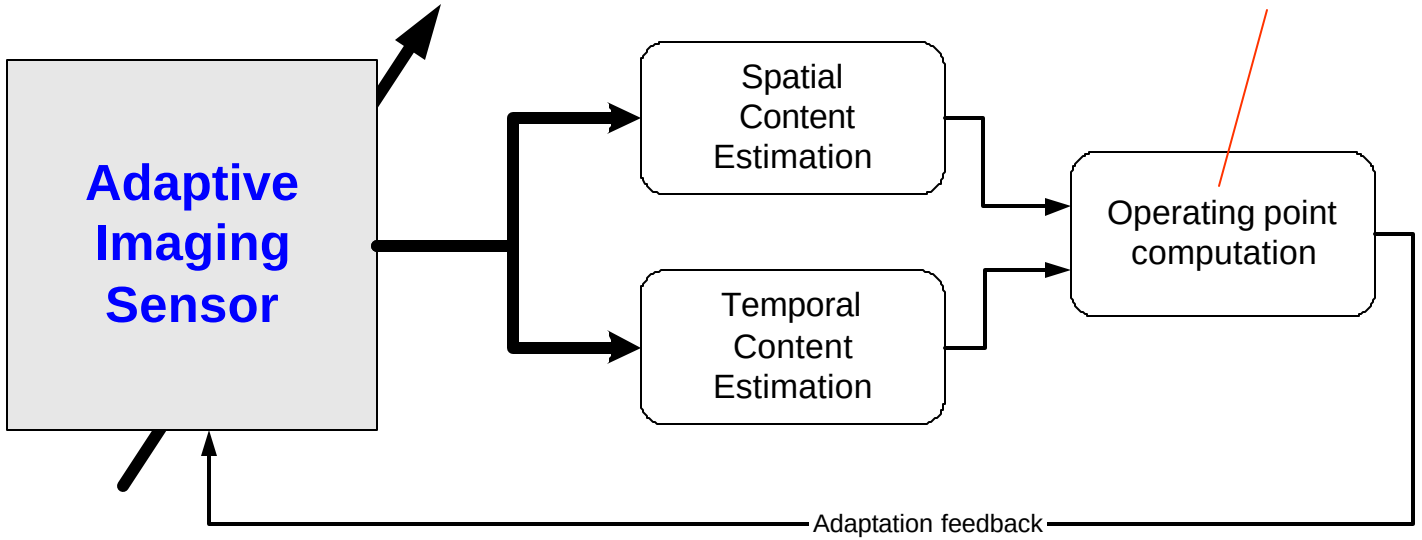
- Basically is the log-likelihood for the multiscale prior model used earlier

- Can also be used in time direction.

Basic feedback mechanism

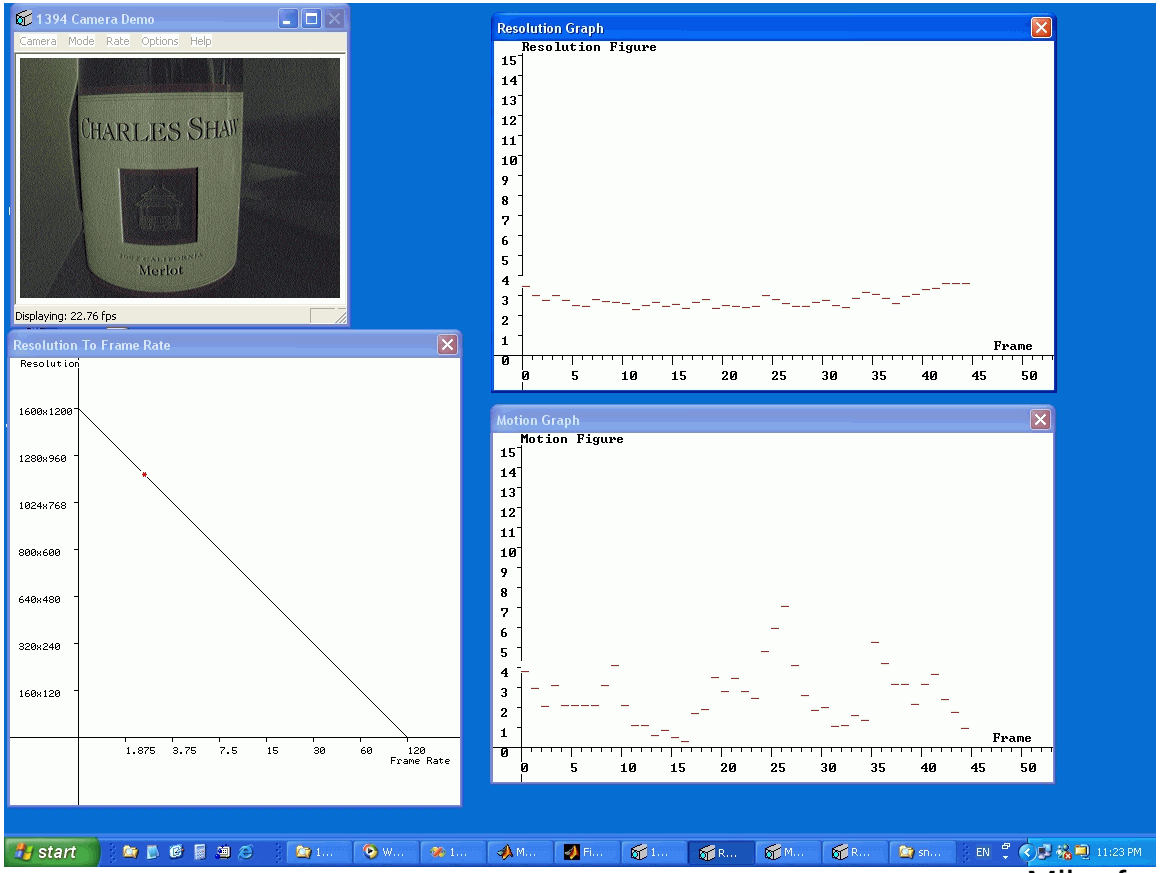


Currently a lookup table



Real Experiments

- A real time setup using Pyro IEEE1394 camera.
- Powerful tool to develop algorithms



What is next: Closing the loop