SUPER-RESOLUTION IMAGE RECONSTRUCTION – METHODS AND LESSONS LEARNED

Sally Wood
Thomas J. Bannan Professor in Electrical Engineering, Santa Clara University

IEEE SCV Signal Processing Society Chapter
13 December 2016

Outline

- What is super-resolution image processing?
- What are some application areas?
- Background and definitions
- Motivations and methods for multi-image super-resolution
- Lessons learned and relationship to newer methods of single image super-resolution

WHAT IS SUPER-RESOLUTION IMAGE PROCESSING?

What Is Computational Super-resolution?

- Two interpretations:
  - Increasing resolution beyond optical diffraction limit
  - Increasing pixel density and spatial frequency content beyond the limit of the sensor array pixel size
- Increasing the spatial resolution of an image beyond the resolution of the straightforward image acquisition procedure for that application

Single or Multiple Image super-resolution?

- Multiple image super-resolution – added information to improve resolution comes from new observations
  - Combine information from multiple images
  - Need given or derived information about relationship of images
  - Assumptions about depth of field to avoid stereo effects from lateral shift of perspective for multiple cameras
  - Assumption about lack of motion of subjects for single camera time sequence
  - Compute estimate the higher resolution image that is most likely or is most consistent with observations or minimizes expected image error
- Single image super-resolution – added information to improve resolution comes from a-priori information
  - Specific statistical or structural information or assumptions
  - Library of training data for specific context

Application Variability

- What is the application context?
- What is the objective of super-resolution processing?
- Why can't a higher resolution image be acquired directly?
- What is the cost of errors?

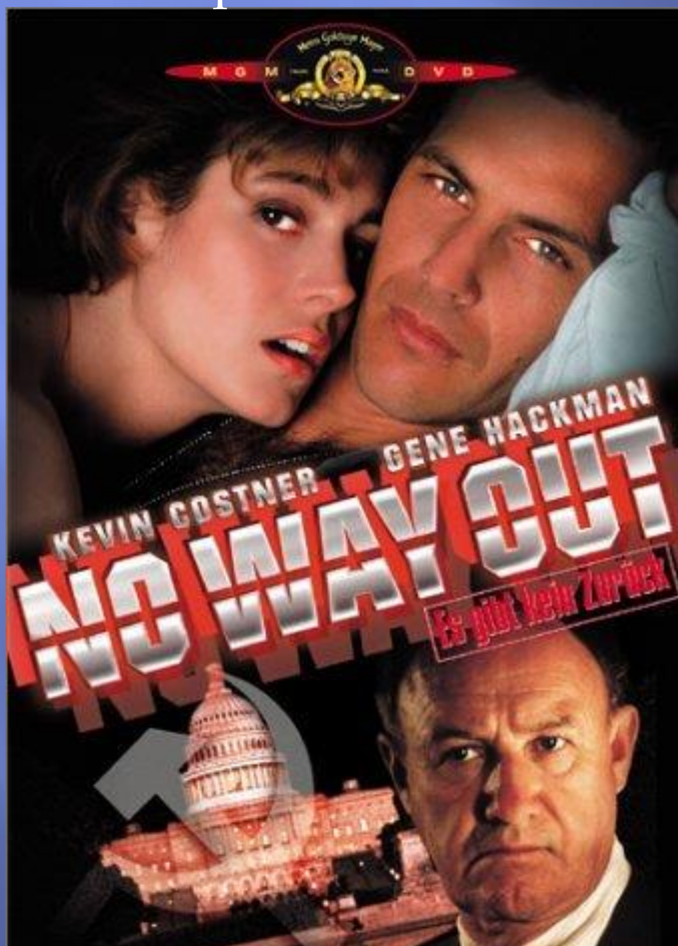
WHAT ARE SOME APPLICATION AREAS?

Who Needs Super-resolution?

- Improve spatial resolution of images for improved viewing satisfaction without improving camera
  - Lower camera cost, size
  - Need for flat form factor prevents use of zoom lens
- Improved estimate of quantitatively meaningful information in an image
  - Identification of structures, features, alphanumeric data
  - Fusion applications in which different spectral bands have different spatial resolution
  - Improve utility of medical images without increasing cost, acquisition time, patient dosage

Super-resolution With Only One Blurred LR Image?

Problem of image restoration or denoising, not super-resolution.
For example:



From promotion for the movie:

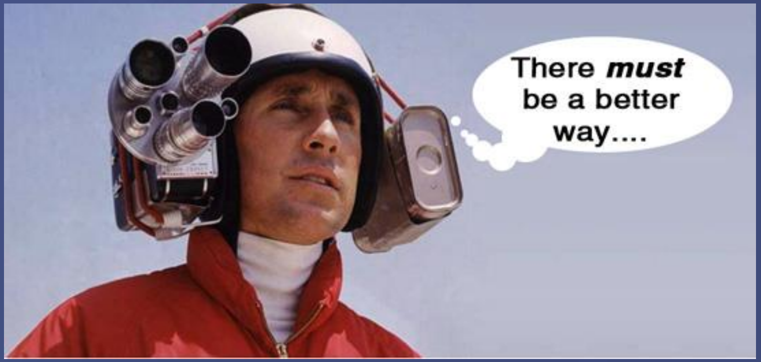
Story Line:

Tom Farrell is a navy officer who gets posted at the Pentagon and is to report to the secretary of defense David Brice. He starts an affair with Susan Atwell not knowing that she is Brice's mistress. When Susan is found dead, Tom is assigned to the case of finding the killer who is believed to be a KGB mole! Tom could soon become a suspect when a Polaroid negative of him was found at Susan's place.

He now has only a few hours to find the killer before the computer regenerates the photo.

Why Not Just Get a Better Camera?

- Could get a camera with a good telephoto lens
- Cost:
  - The cost of a better camera with better optical zoom capability might be greater than the cost of adding computational capability.
- Size:
  - The increased length of the optical path for increased zoom capability might cause problems
    - Flat wall mounted cameras for museum surveillance
    - Helmet mounted cameras for first responders
- Flexibility
  - select ROI for increased resolution
  - May mot need increased resolution over the whole image



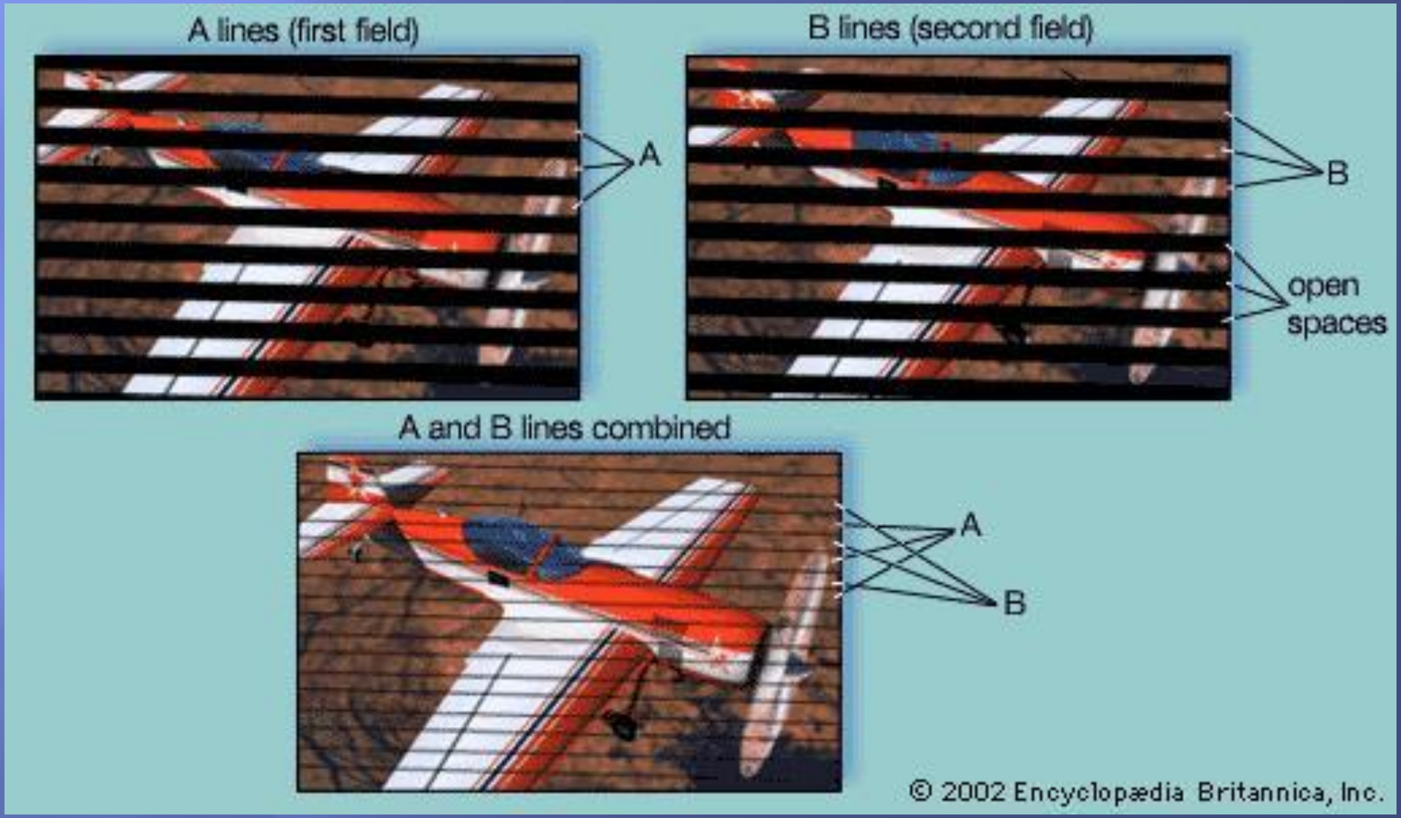
Application Areas

- Consumer applications – digital cameras –
  - Create high resolution images computationally rather than purchasing a more expensive or larger camera
- Surveillance or security monitoring
  - Get high resolution images from a camera limited in size, form factor, weight
  - Use widespread distributed camera networks
- In general motivated either by desire for improved performance from existing technology or by design needed for special constraints

Examples in Common Use

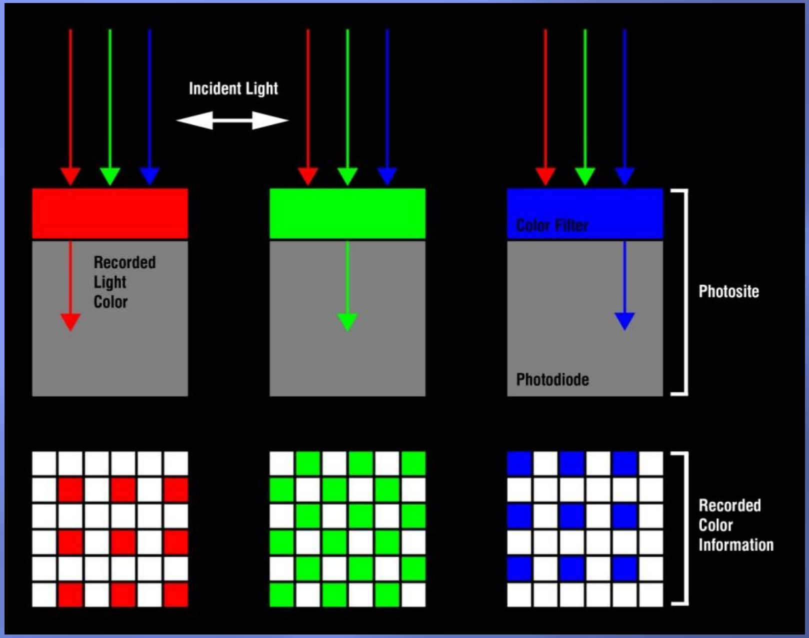
- Interlaced TV images of SDTV –
  - Each frame had even lines or odd lines so each new frame updated half the image lines. Produced 2X resolution of full frame update with same data rate
- Bayer pattern of color images
  - 50% of pixels are green, 25% red, 25% blue
- Compound eye vision
- Fusion of multispectral images – registration and resampling
  - Visible and lower resolution IR
  - Geoscience (Wang 2005) fuse low-resolution multispectral images (LRMI) with high-resolution panchromatic images (HRPI).

Interlaced SDTV Images

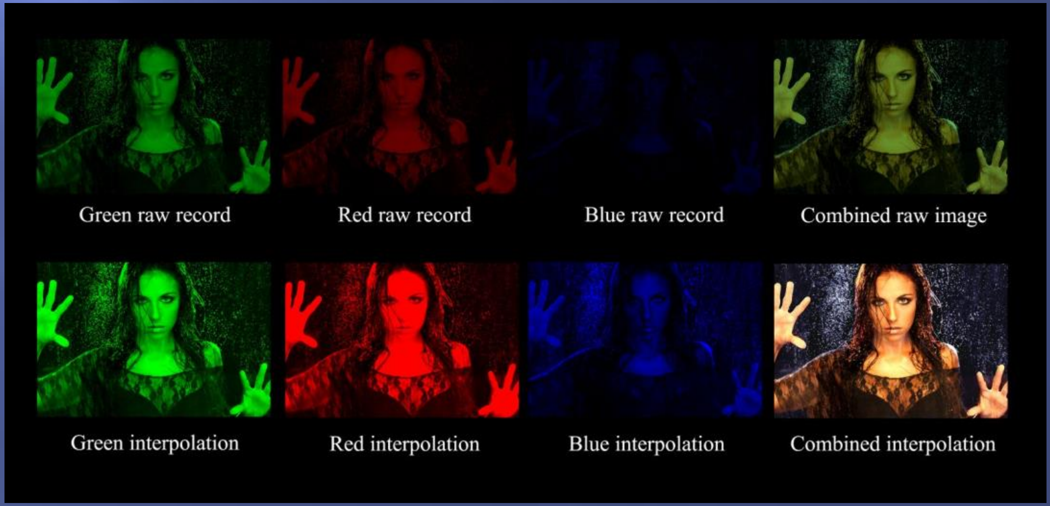


- xhttps://www.britannica.com/technology/television-technology/images-videos/Interlaced-scanning-for-standard-television-display-The-first-field-made/775

Color Images fom Bayer Pattern Sensors

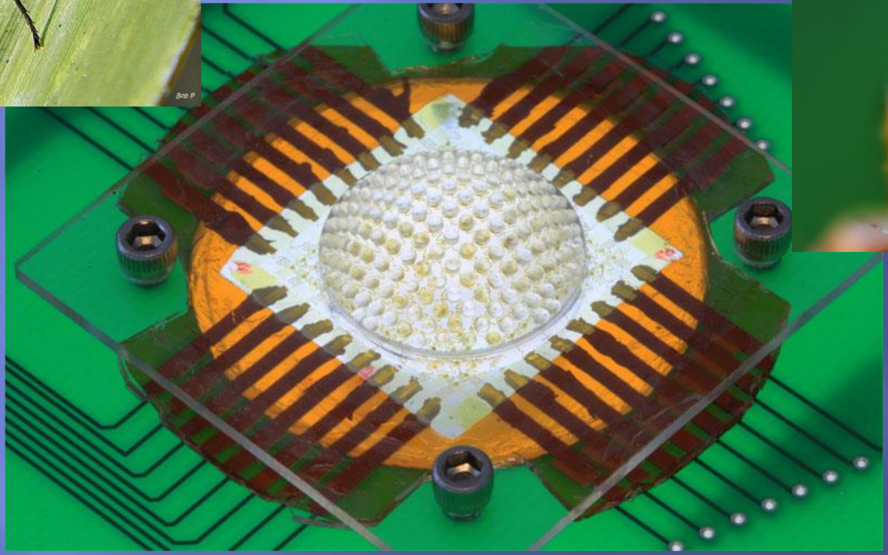


Each pixel sensor is covered by a red, green, or blue filter



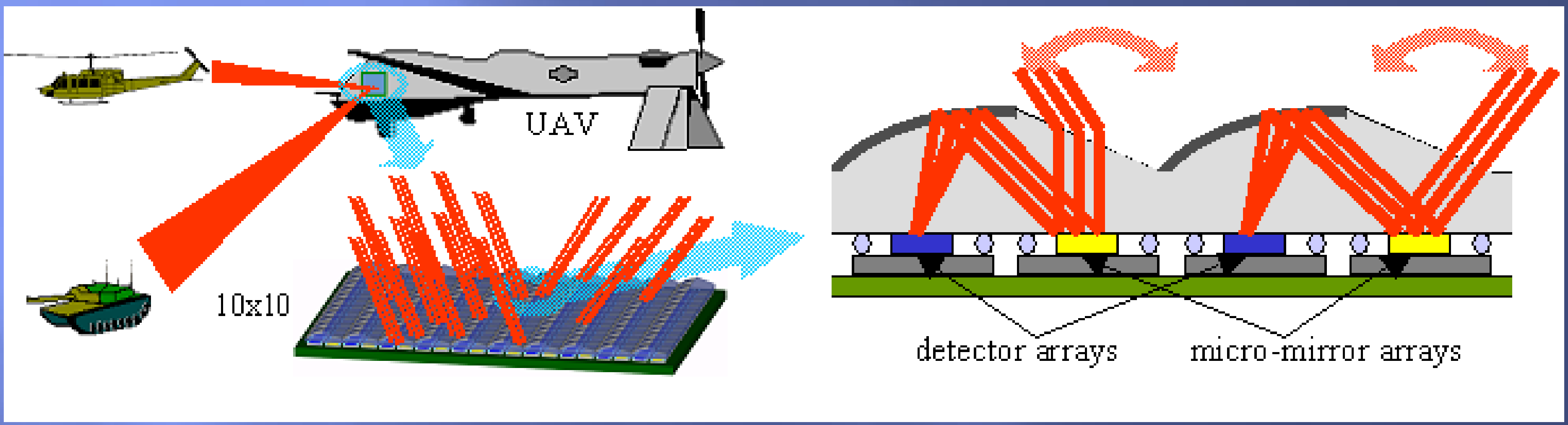
- http://www.creativeplanetnetwork.com/news/news-articles/dv101-raw-deal-what-does-it-mean-record-raw-imagery/423392

Compound Fly Eye



- • http://phenomena.nationalgeographic.com/2013/05/02/insect-eye-digital-camera-sees-what-you-just-did/
- http://commons.wikimedia.org/wiki/File:Eyes_of_a_Holcocephala_fusca_Robber_Fly.jpg
- http://commons.wikimedia.org/wiki/File:Thru_The_Eyes_Of_Ruby_%28the_fly%29_%288219315716%29.jpg

A Special Purpose Flat Camera Design



- Array of small sub-imagers with fields of view that can be controlled by micro-mirrors
- Local variance can control density of coverage

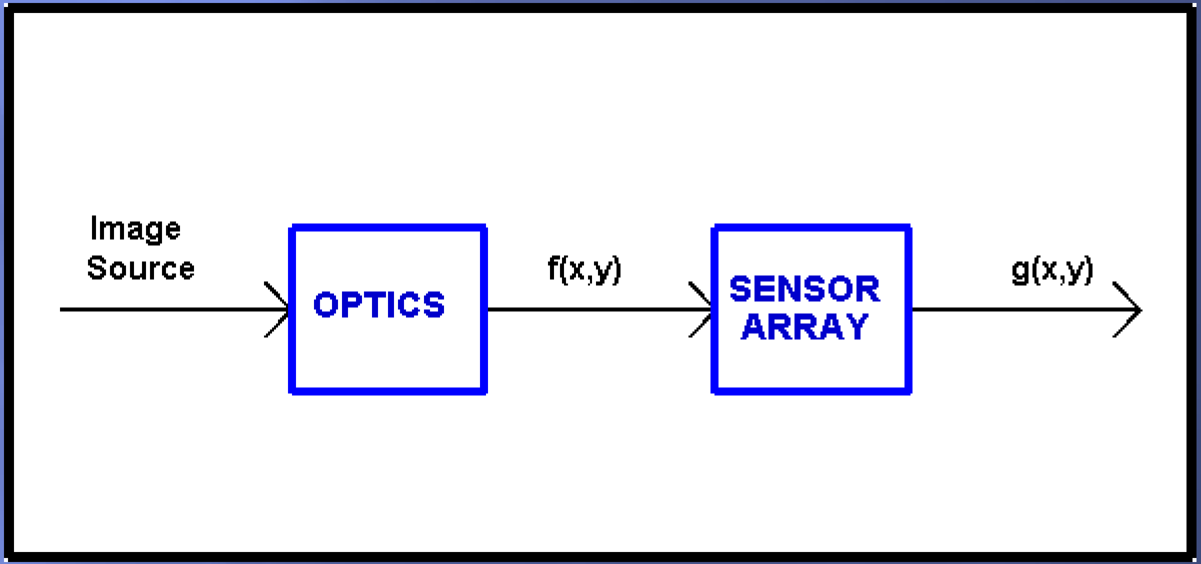
BACKGROUND AND DEFINITIONS

High Resolution (HR) Images from Multiple Low Resolution (LR) Images

- Multiple lower resolution images can be captured with sub-pixel shifts
  - From a single camera in motion
    - Trajectory or vibration
  - From multiple cameras at the same time
    - Multiple imagers may have dynamically controlled directions
- Using a model for image formation, the information in the LR images can be combined to estimate the HR image
- Ill-posed, but interesting problem
- Basic Assumptions:
  - Aliased high frequency content is present in LR images, not eliminated to create a band limited signal
  - There is no significant motion in the scene while LR images are captured
  - There is no compression or space varying response that would cause LR image combination problems

Model for Image Formation

f(x,y) is light field on sensor plane
g(x,y) is f(x,y) blurred by the sensor response function. This is sampled to provide digital LR image



Optics:

- Magnification
- Spatial frequency response
- May be space variant

Sensor array

- Pixel spacing
- Pixel size
- Dynamic range
- Noise characteristics

Simple Model of Sensor Response

- Assume pixel centered at (x_0, y_0) has response $A(x, y, x_0, y_0)$. If shift invariant the response is $g(x_0, y_0) = \iint_{-\infty}^{\infty} f(x, y)A(x_0 - x, y_0 - y)\, dxdy$
- LR image is obtained by sampling the function of continuous variables, g(x,y), at intervals determined by the pixel spacing, w.
- In this model $A(x, y\ )$ will always be zero for $|x| > 0.5w$ or $|y| > 0.5w$.
- Due to the fill factor of the array, the pixel will respond to an area smaller than wxw.

Relationship of LR Image to Desired HR Image

- Let all two dimensional HR and LR image be stored in column vector form.
- $\underline{g}_k = H_k \underline{f}$ relates the k^th LR image $\underline{g}_k$ to the desired HR image $\underline{f}$ through the observation partial matrix H_k.
  - This matrix is sparse and includes the blurring of the HR image due to integration over the sensor surface and relative position information.
  - H_k = D$B_k M_k$ where M represent positioning and possibly warping transformations, B represents the blur of the sensor, and D represents the downsampling to the sensor pixel structure

Combining LR Images

- For M LR images, ideally:

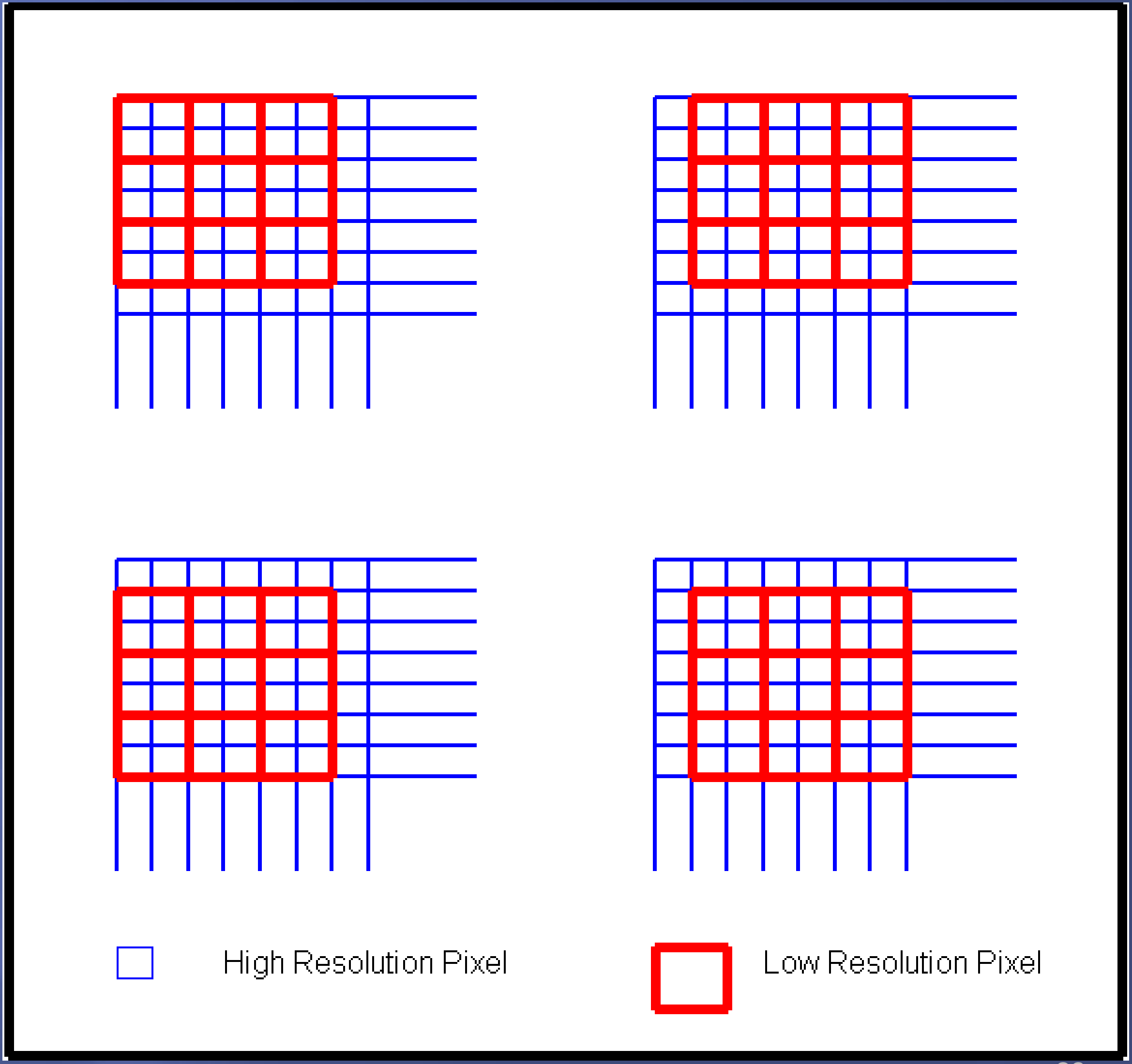
$$\underline{g} = \begin{bmatrix} \underline{g}_1 \\ \underline{g}_2 \\ \vdots \\ \underline{g}_M \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_M \end{bmatrix} \underline{f} = H\underline{f}$$

- When noise is added to the equation

$$\underline{g} = H\underline{f} + \underline{v}$$

Array for 2x Improvement

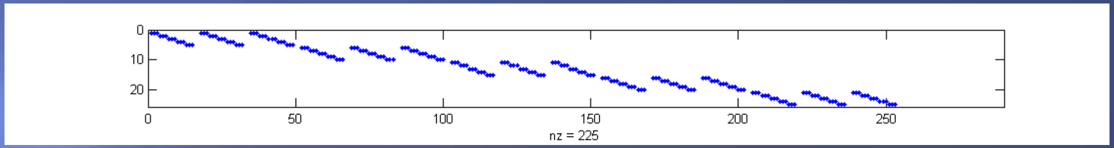
- Each LR pixel is the sum of four desired HR pixels
- Four LR pixel arrays are shown at four different position offsets



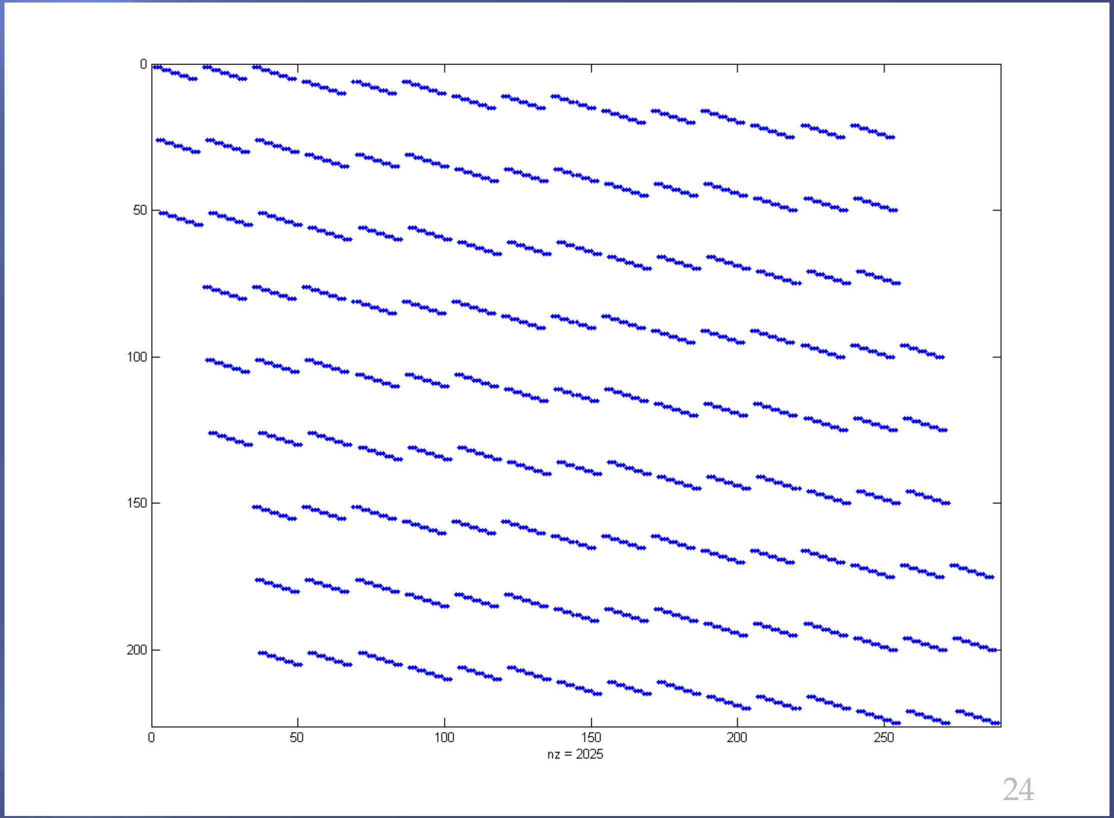
Observation Partial Matrix

Desired Improvement is 3X. Nine LR images are used, each of which is 5x5 pixels.

- One LR image — $H_1=$

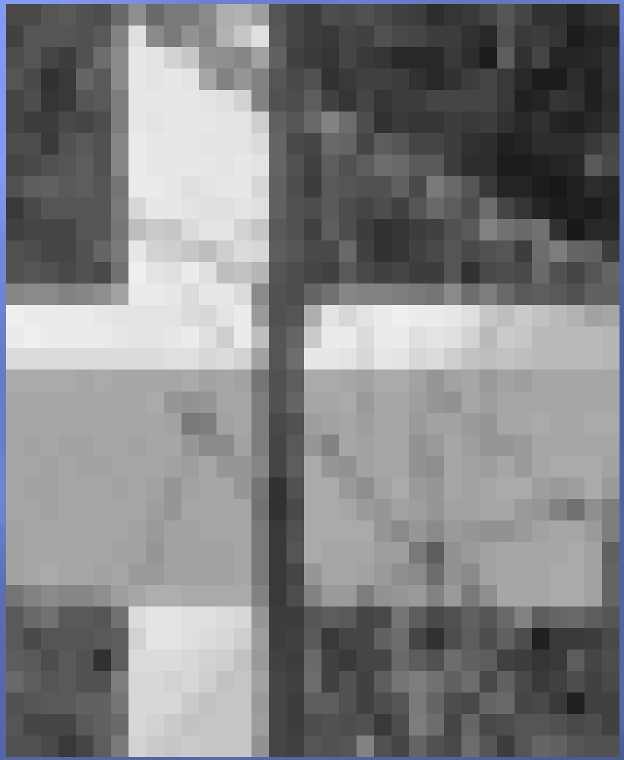


- 3x3 set of LR images — $H=$

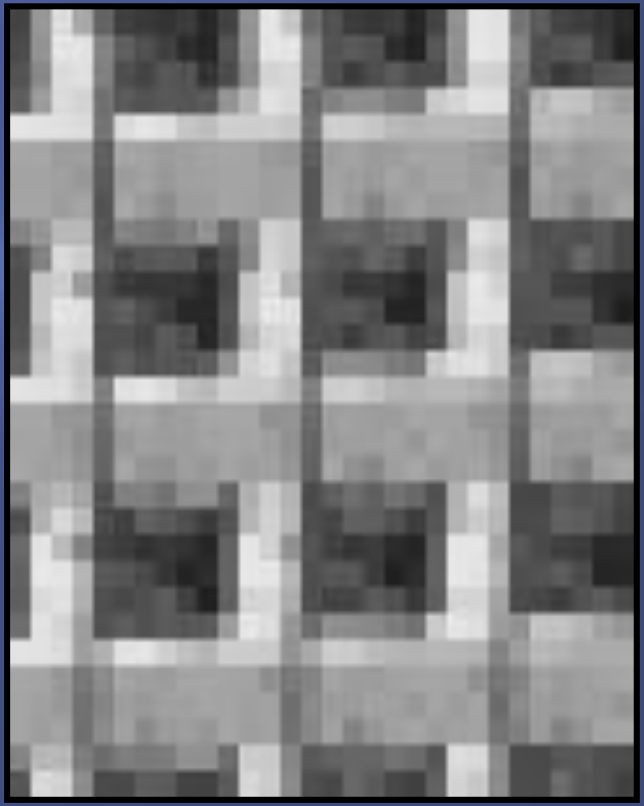


Example Image Segment shown as 3x3 LR array

Full Resolution



Array of LR Images

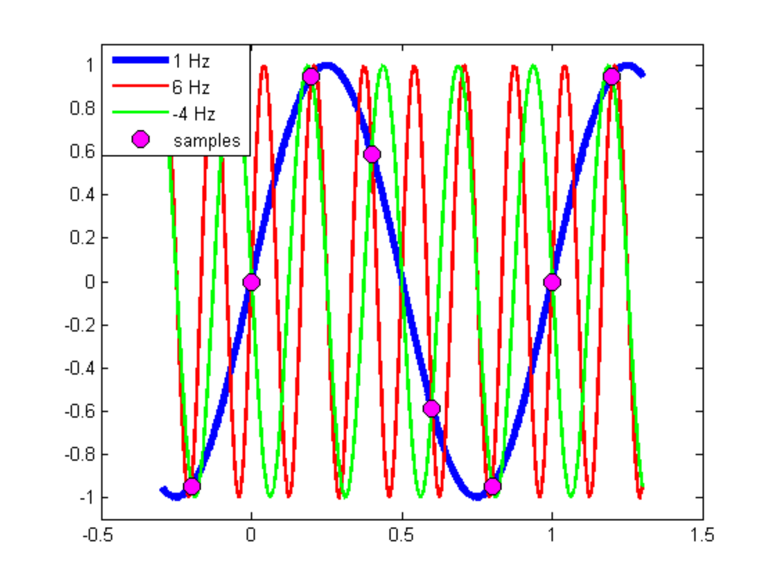


LR Images Must Have Aliased HF Information

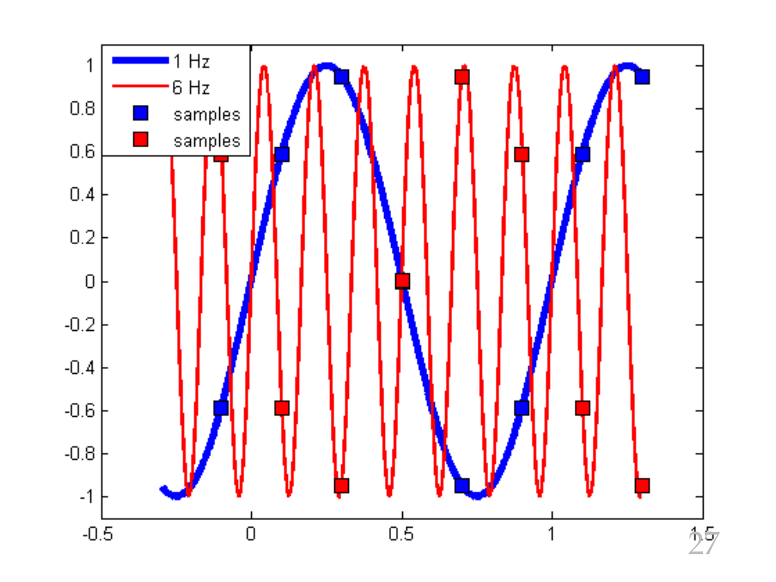
- If aliased high frequency information has been lost, then super resolution methods can not extract it
  - Optics must produce f(x,y) with high frequencies that are undersampled by g(x,y).
  - If $A(x,y) = rect\left(\frac{x}{a}\right) rect\left(\frac{y}{a}\right)$, then sampling is far from the ideal sampling assumed in time signal processing.
  - $A(u,v) = a^2 sinc(au) sinc(av)$ will attenuate high frequencies and null spatial frequencies at multiples of $\left(\frac{1}{a}\right)$.
- If some aliased information has been retained AND accurate registration of offset LR images is possible, then aliased information can be extracted.

One Dimensional Aliasing

- Three different frequencies have the same sample sequence

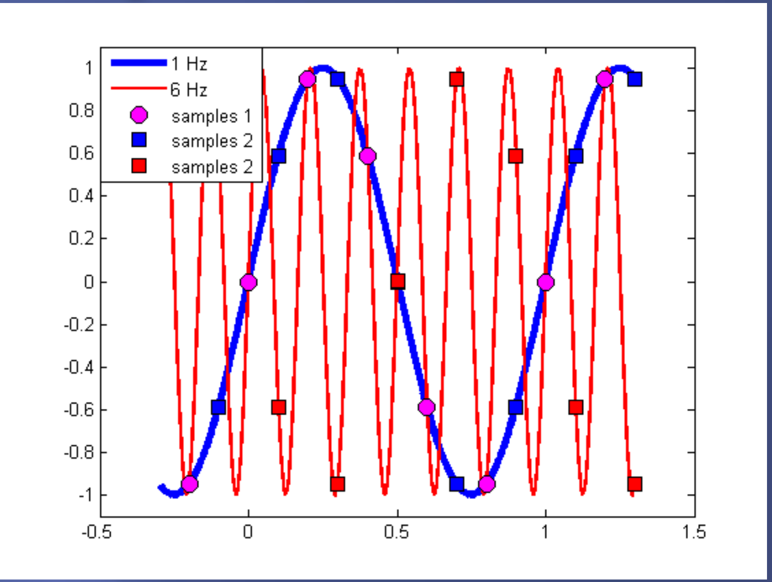


- A second set of sample sequences offset by half the sampling interval doubles the effective sampling frequency – frequencies with identical sample sequences above now have different sample sequences

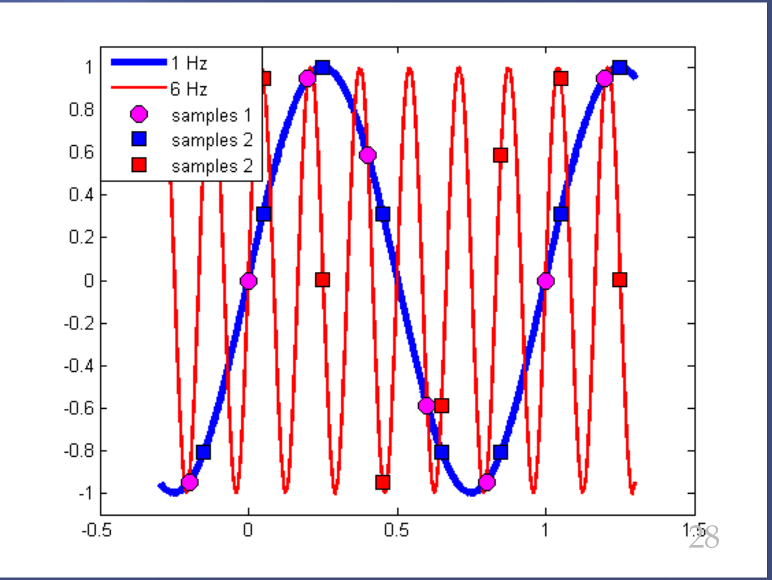


Combined LR Sequences

- Combined sample sequences can distinguish the two frequencies



- Offset of second sequence does not have to be exactly half the original sample sequence



Aliasing in an Image

- Aliasing causes patterns that are visually interpreted incorrectly. Aliased component are suppressed for viewing, but must be retained for super-resolution computations
- Patterns depend on relative position of camera and scene objects and may change dramatically if either moves
- https://svi.nl/AntiAliasing

MOTIVATIONS AND METHODS FOR MULTI-IMAGE SUPER-RESOLUTION

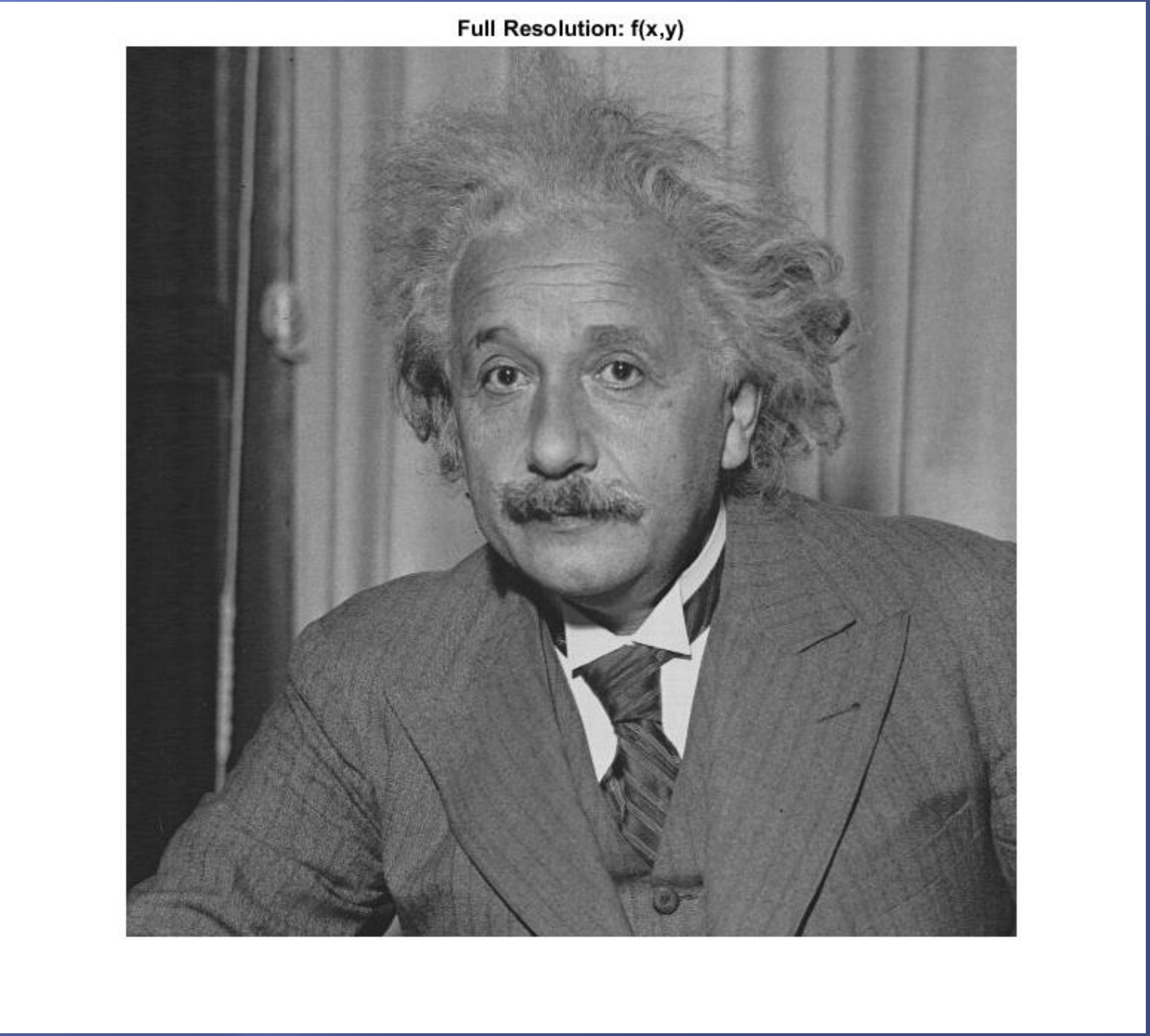
Demonstrations Using Constructed Examples

- HR image is blurred and downsampled at specified rate
- Relative position of HR images is known
- Shifts are equally spaced fractions of LR pixels

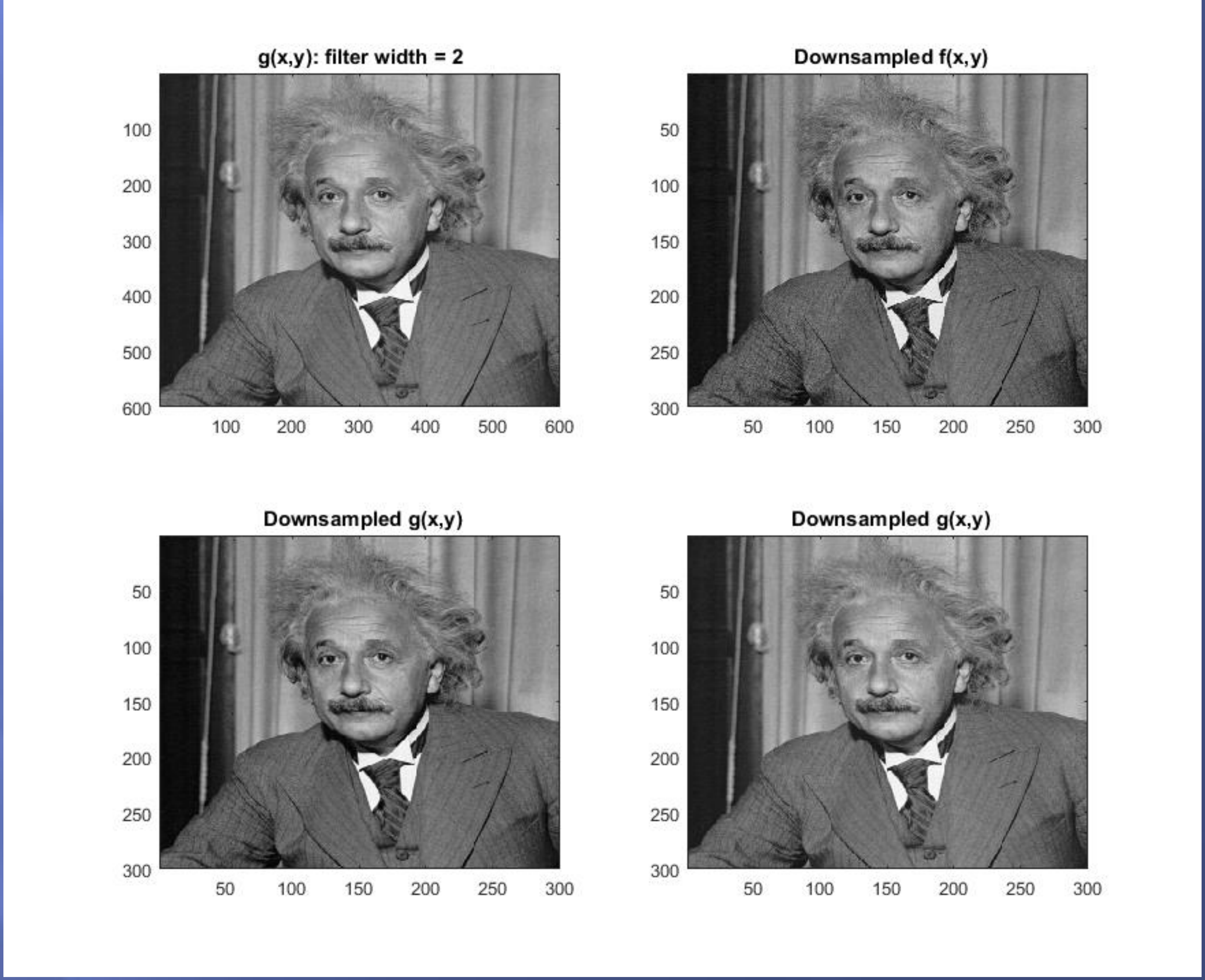
Example 1

Einstein Photo 600x600 pixels

https://th.physik.uni-frankfurt.de/~jr/physpiceinstein.html



Pixel in g(x,y) includes 4 pixels from f(x,y)

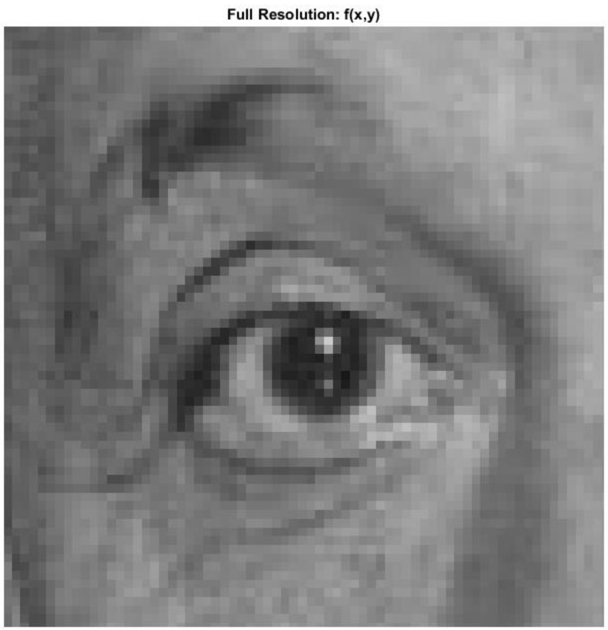


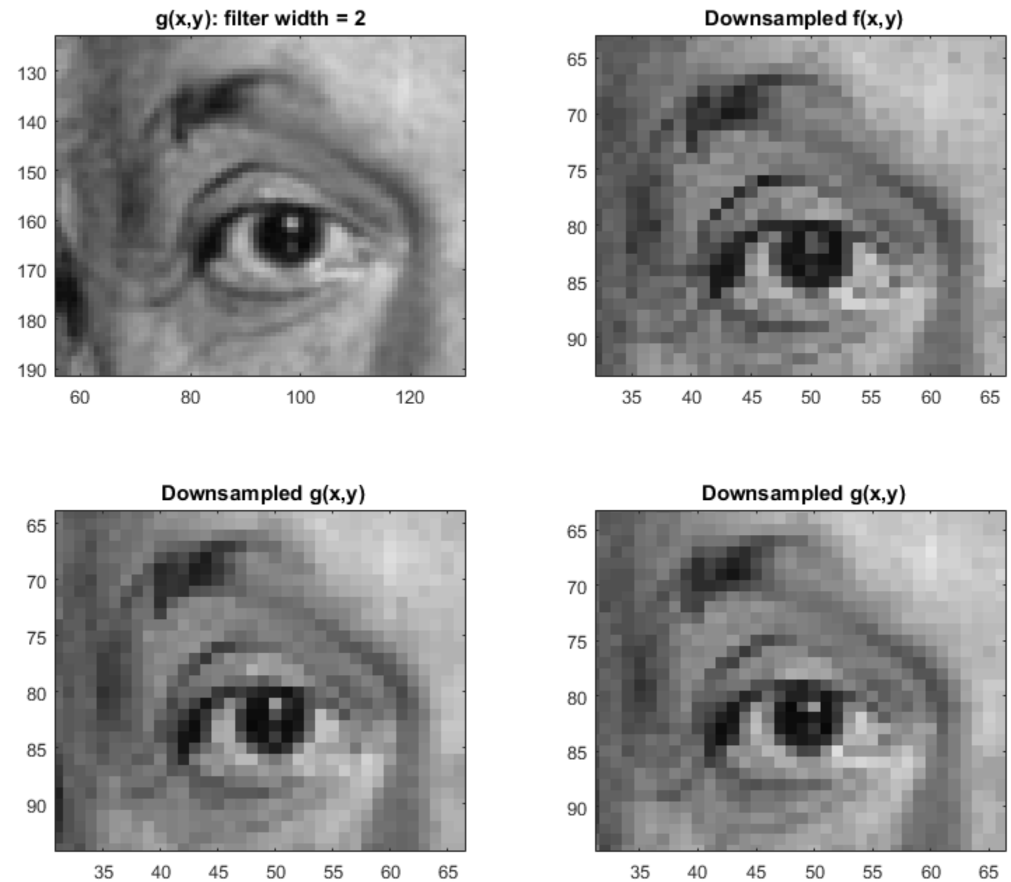
Four 300x300 downsampled versions of g(x,y) can be aligned or interpolated to form 600x600 g(x,y)

Detail of Eye 2X Reduction

- Original

Reduced Resolution





Detail of pupil in original is reduced in g(x,y)

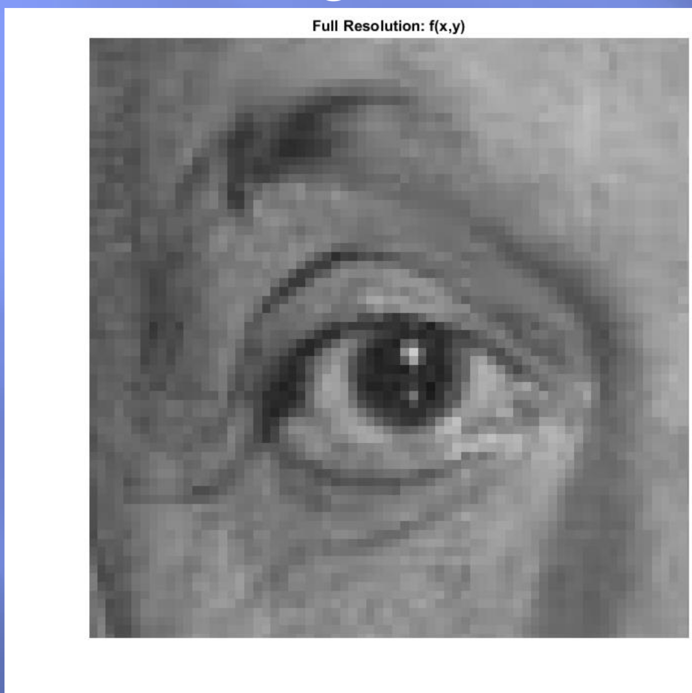
Reconstruction

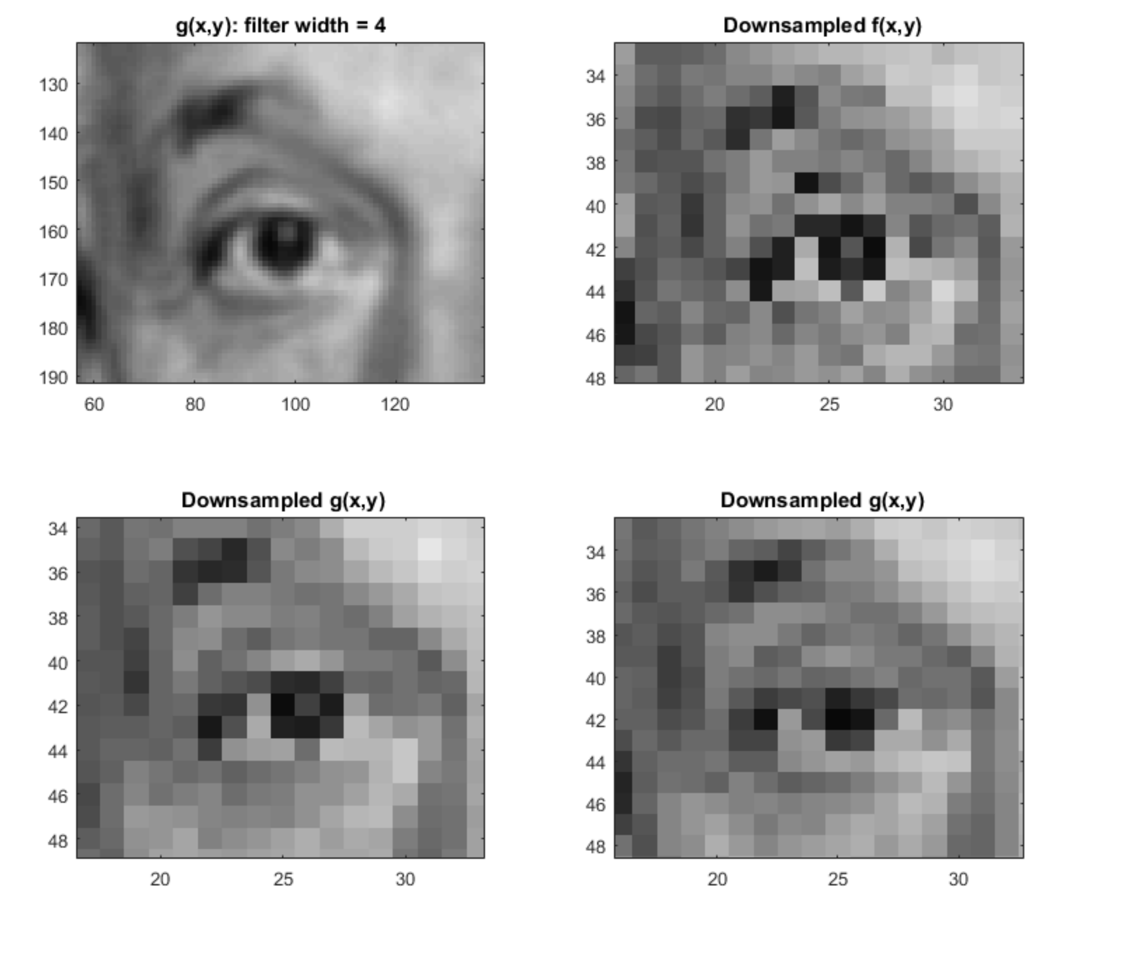
- Four 300x300 down-sampled version of g(x,y) can be combined by registration and interpolation to get 600x600 g(x,y).
- To get 600x600 f(x,y) from g(x,y) need to deblur or restore image to undo the blur of the larger pixel size.
- If downsampled versions of f(x,y) were available, f(x,y) could be recovered without deblurring. However, this is unlikely. Only downsampled versions of g(x,y) are produced by the lower resolution imaging system

Detail of Eye 4x Reduction

- Original

Reduced Resolution



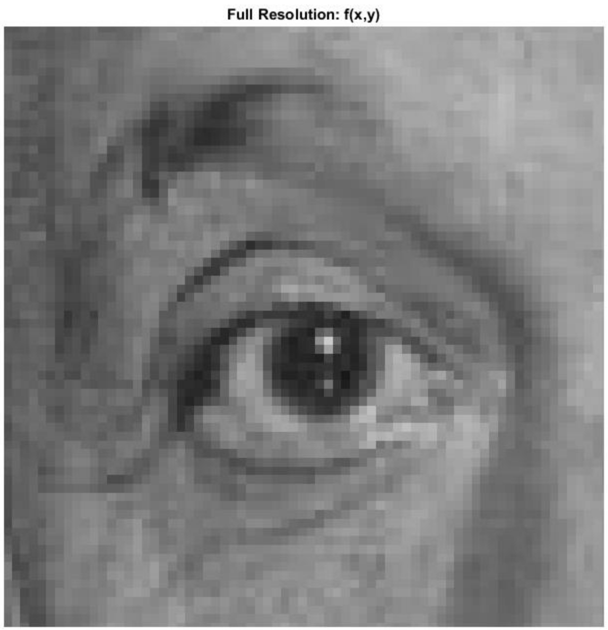


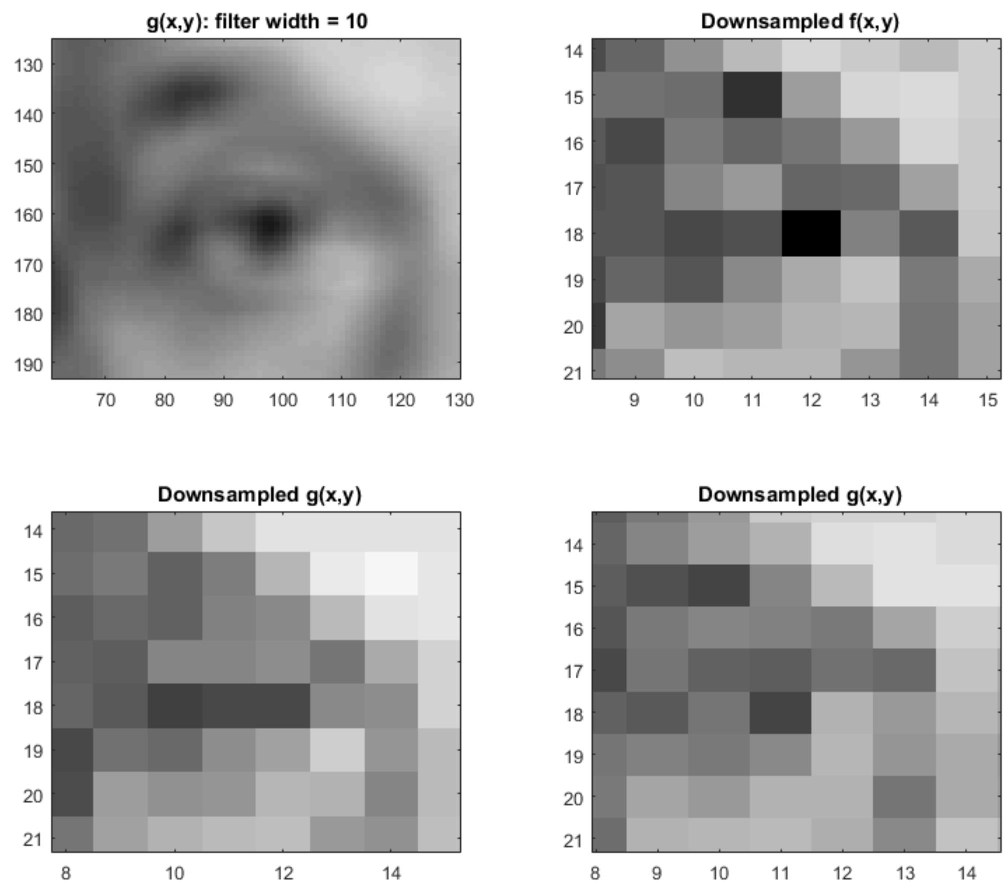
Highlight in pupil in original is missing in g(x,y)

Detail of Eye 10x Reduction

- Original

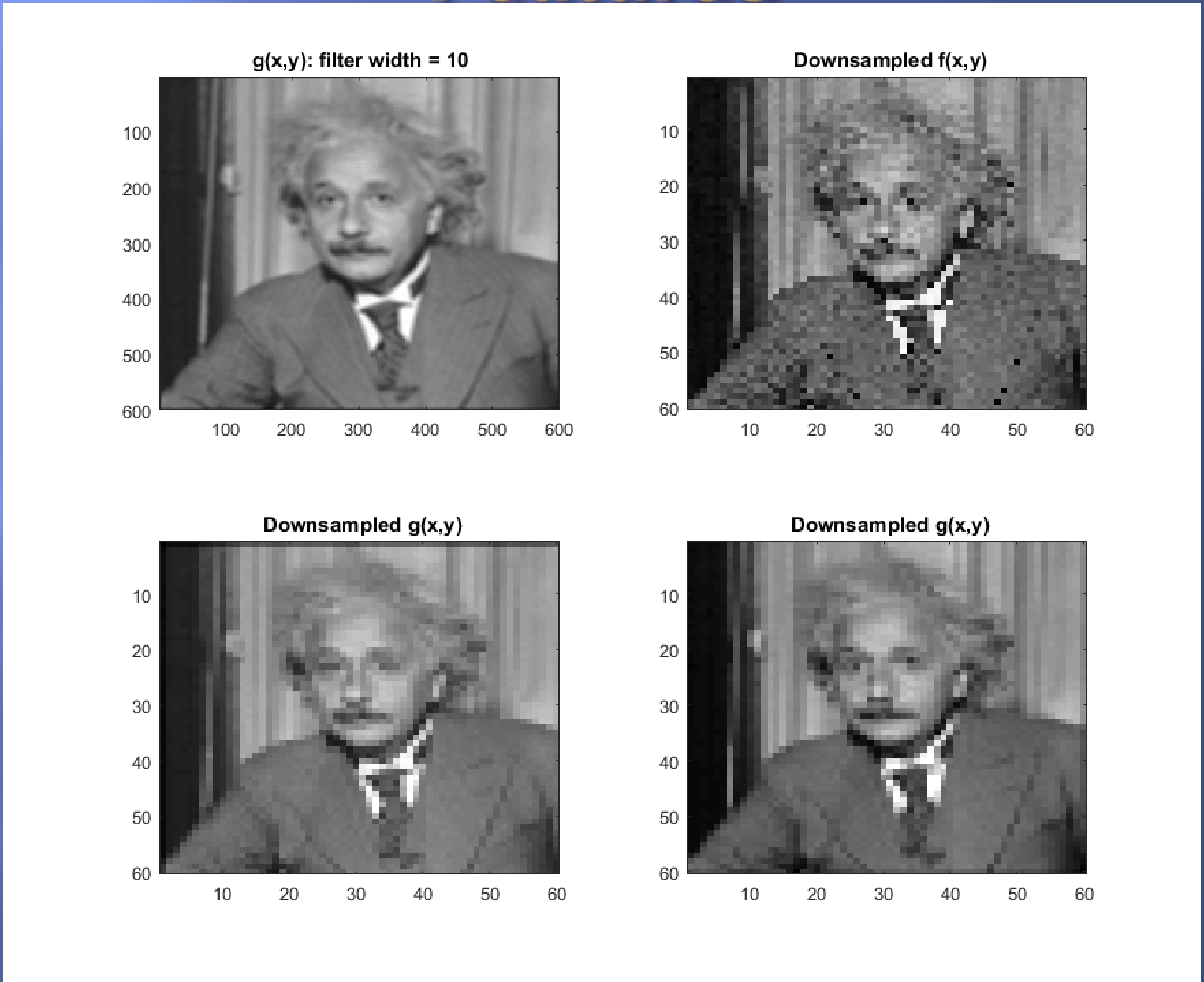
Reduced Resolution



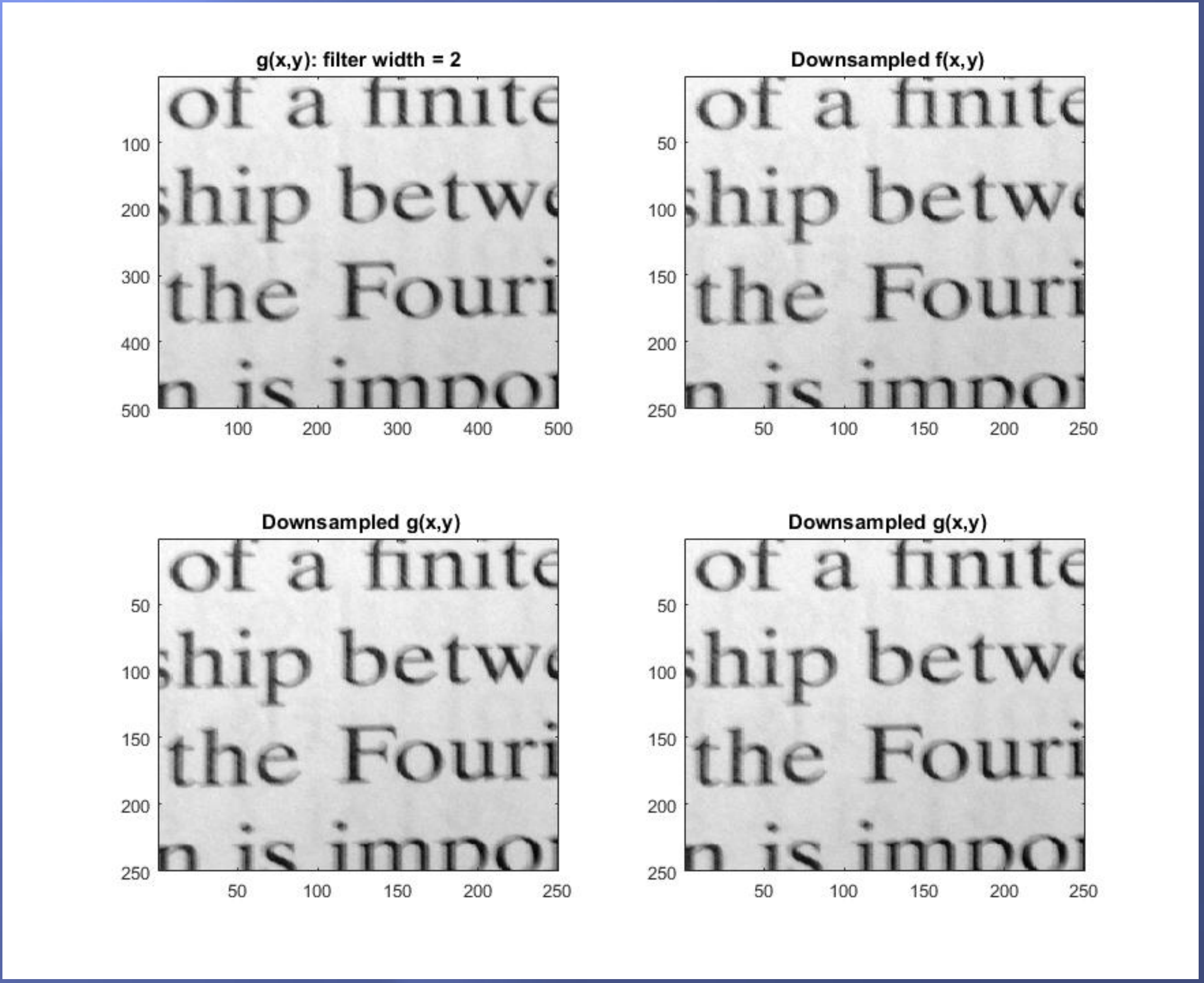


Detail of eye in original is missing in g(x,y)

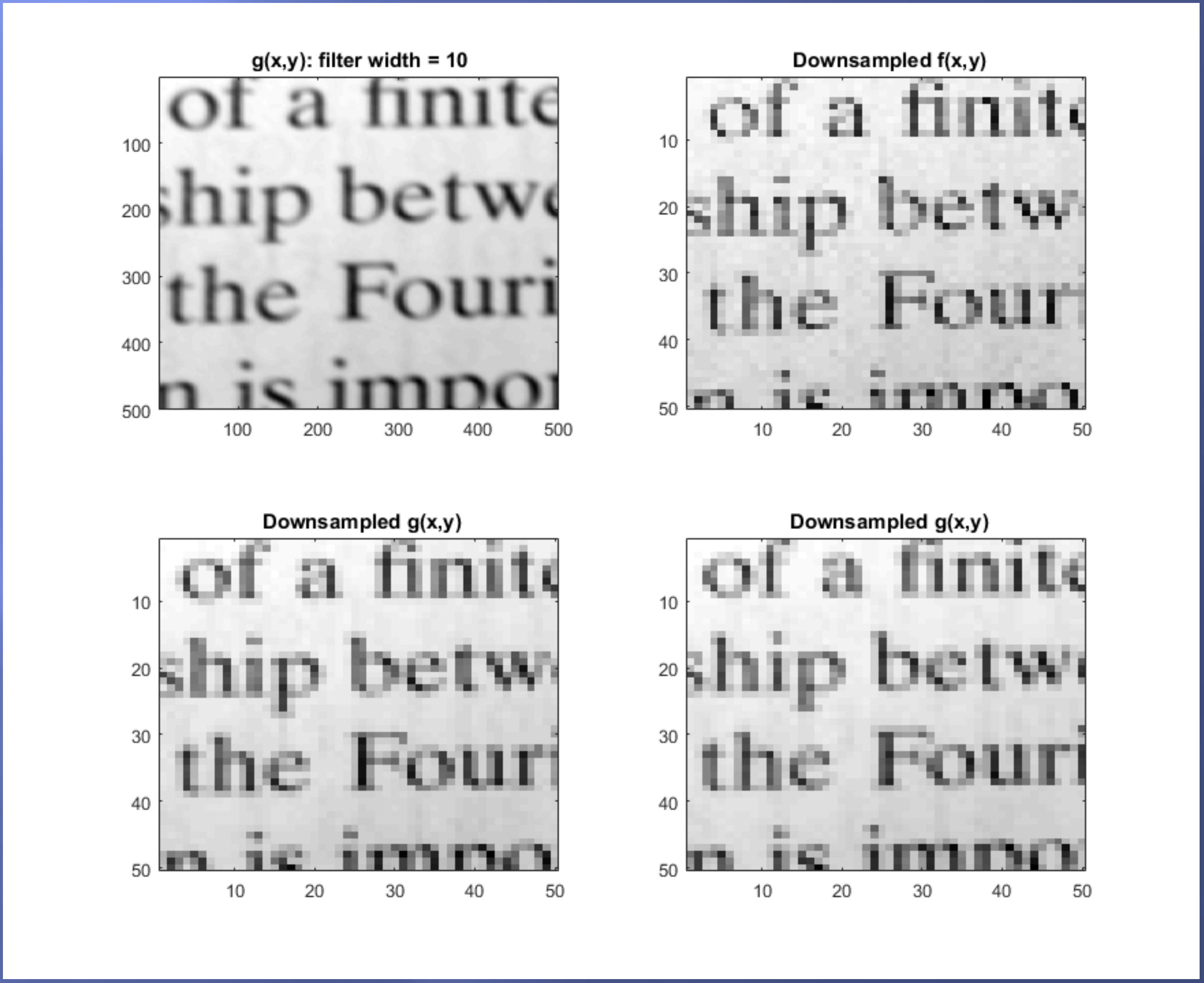
10X Reduction Showing larger Features



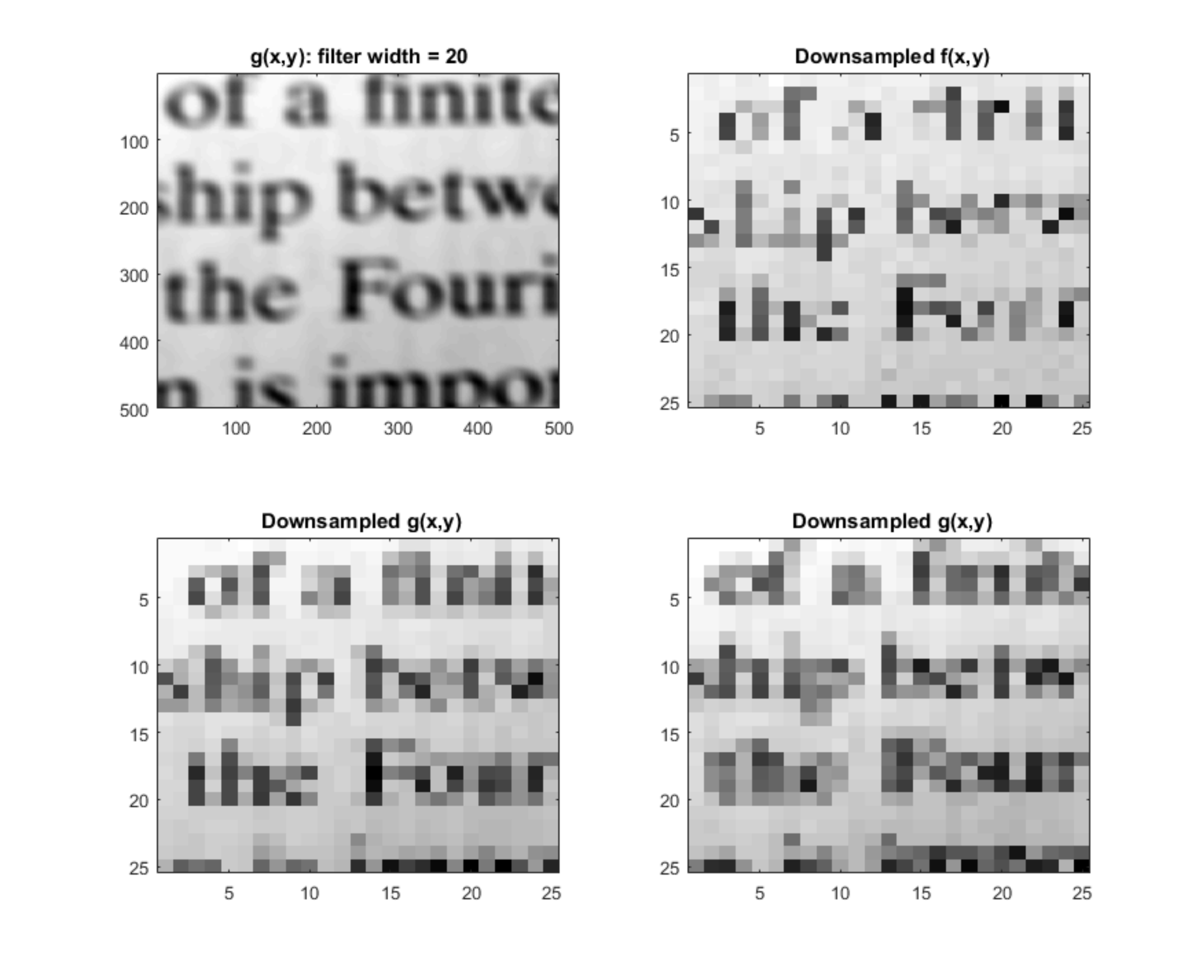
Example 2 – 2x Reduction



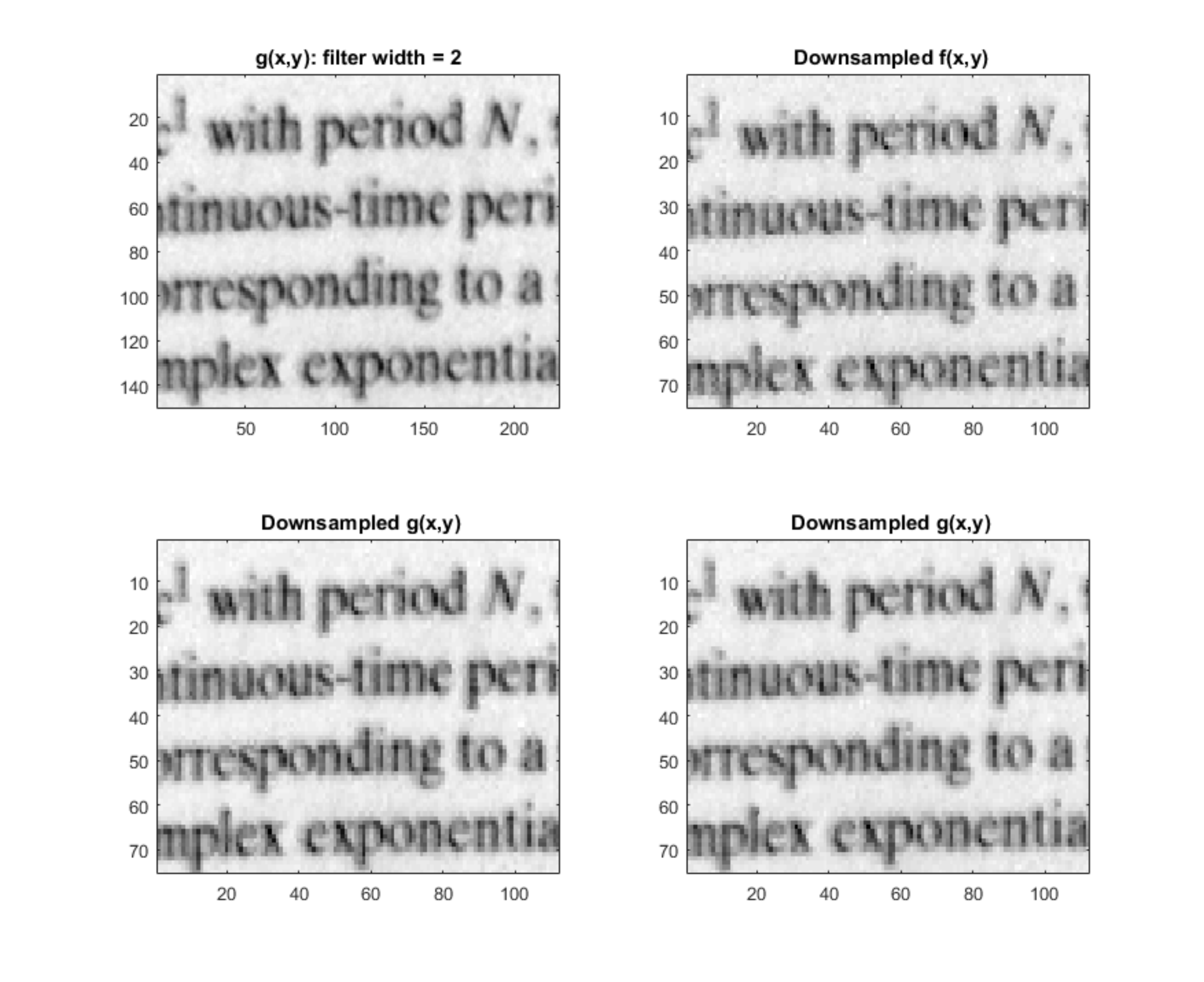
Example 2 – 10x Reduction



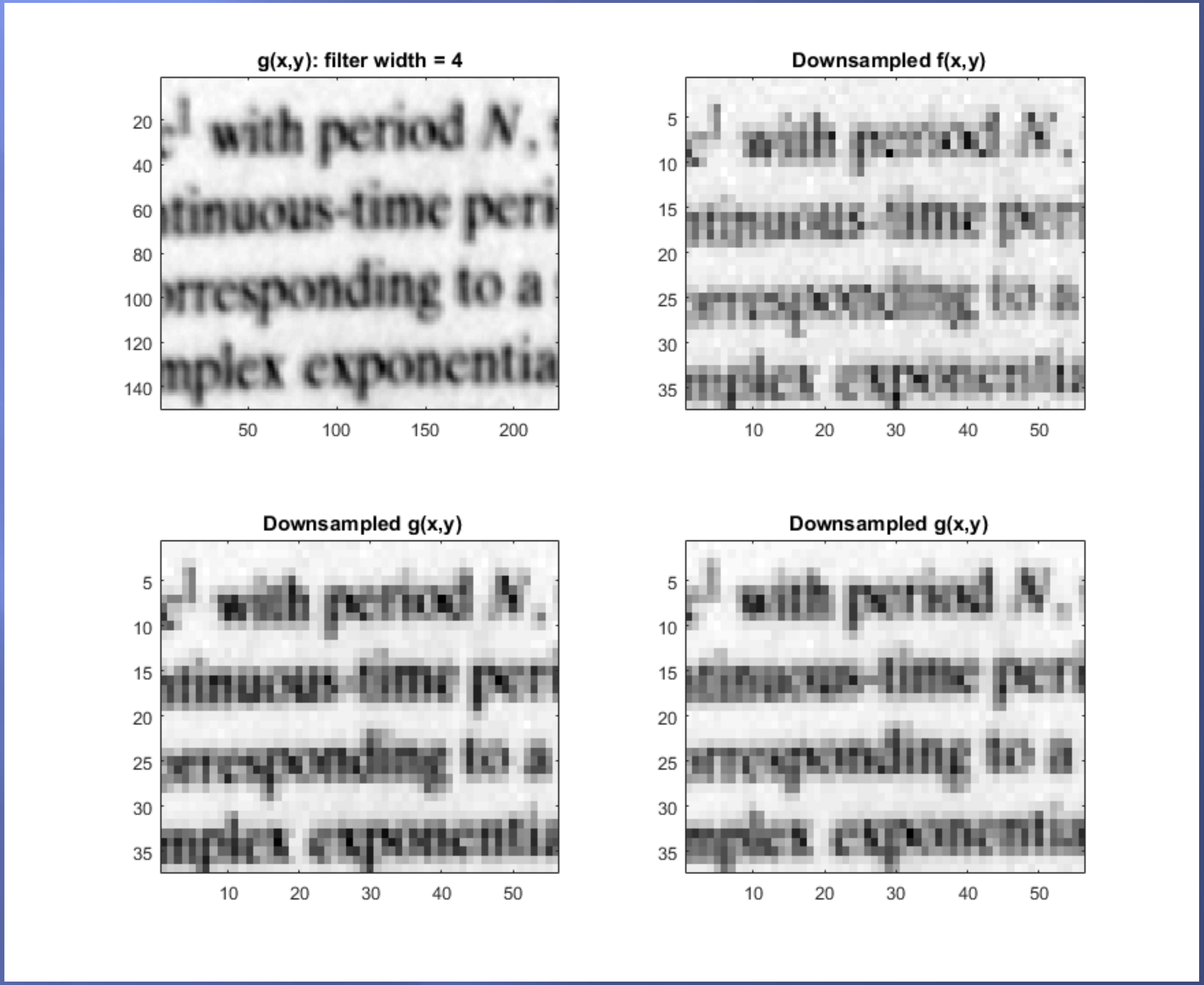
Example 2 – 20x Reduction



Example 3 – 2x Reduction



Example 3 – 4x Reduction



Demonstration Summary

- In all cases full g(x,y) image created from multiple offset down-sampled LR images shows more coherent detail than individual LR images.
  - Depending on the image content and the rate of down-sampling, some LR images contain most of the edge and structure information
  - Creating g(x,y) requires registration and interpolation.
- Creating the higher level of detail in f(x,y) requires deblurring.

Challenges for Super-resolution Image Reconstruction from Multiple Images

- LR image registration
- Image noise
- Modeling information and sensitivity to incorrect parameters
- Computational burden

How Are LR Images Aligned?

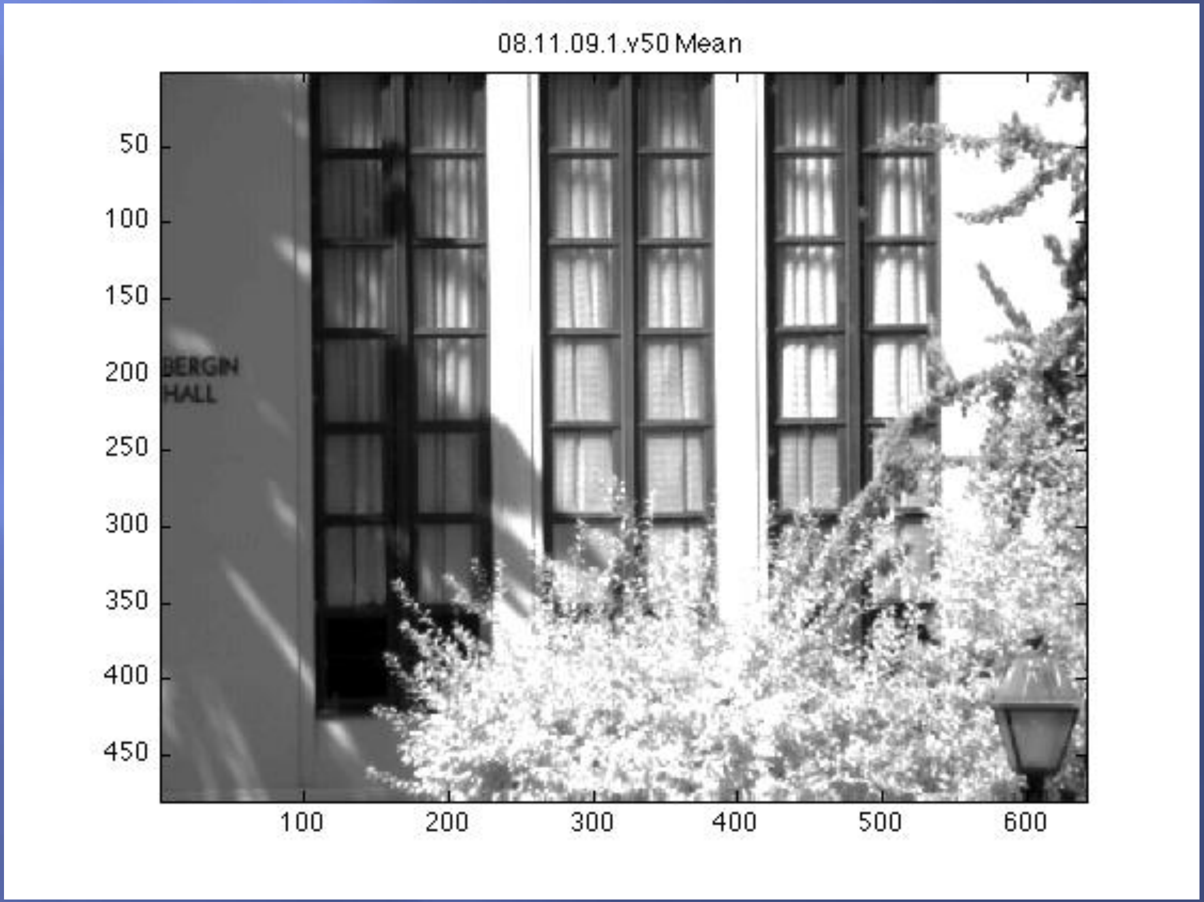
- An imaging system that acquires LR images and controls sub pixel offsets, would have known offsets. This would require a precise special purpose image acquisition system.
- Otherwise registration is complicated because LR images must have aliased image content if HR image is to be restored
- Space domain alignment – could use correlation for sub-pixel registration
  - Because of aliasing, edges and significant points are represented in same way in LR images and sub pixel shift estimation is difficult
- Frequency domain (Vanderwalle, 2005):
  - Use linear phase difference in frequency domain to estimate offset
  - However, aliased component will have a different offset. Assume that higher frequency aliased components have lower amplitudes.
  - Rotation can cause problems

Image Noise

- Many sources of image noise
  - Amplifier noise and measurement noise
  - Quantization noise – easily modeled as a uniformly distributed noise over normalized quantization interval $\frac{1}{2^b}$
    - Typically b=8 for image sensors
  - Poisson noise
- If Poisson noise dominates, then if N_{max} is the maximum count associated with maximum output level g_{max}, let $\alpha = g_{max} / N_{max}$.
  - $sd(g) = \alpha \cdot sd(N)$ and $var(g) = \alpha \cdot g$.
  - At $g = g_{max}$, the variance is $\alpha\, g_{max} = (g_{max})^2 / N_{max}$.
- For denser pixel arrays with lower values of N_{max}, the noise variance will be higher.

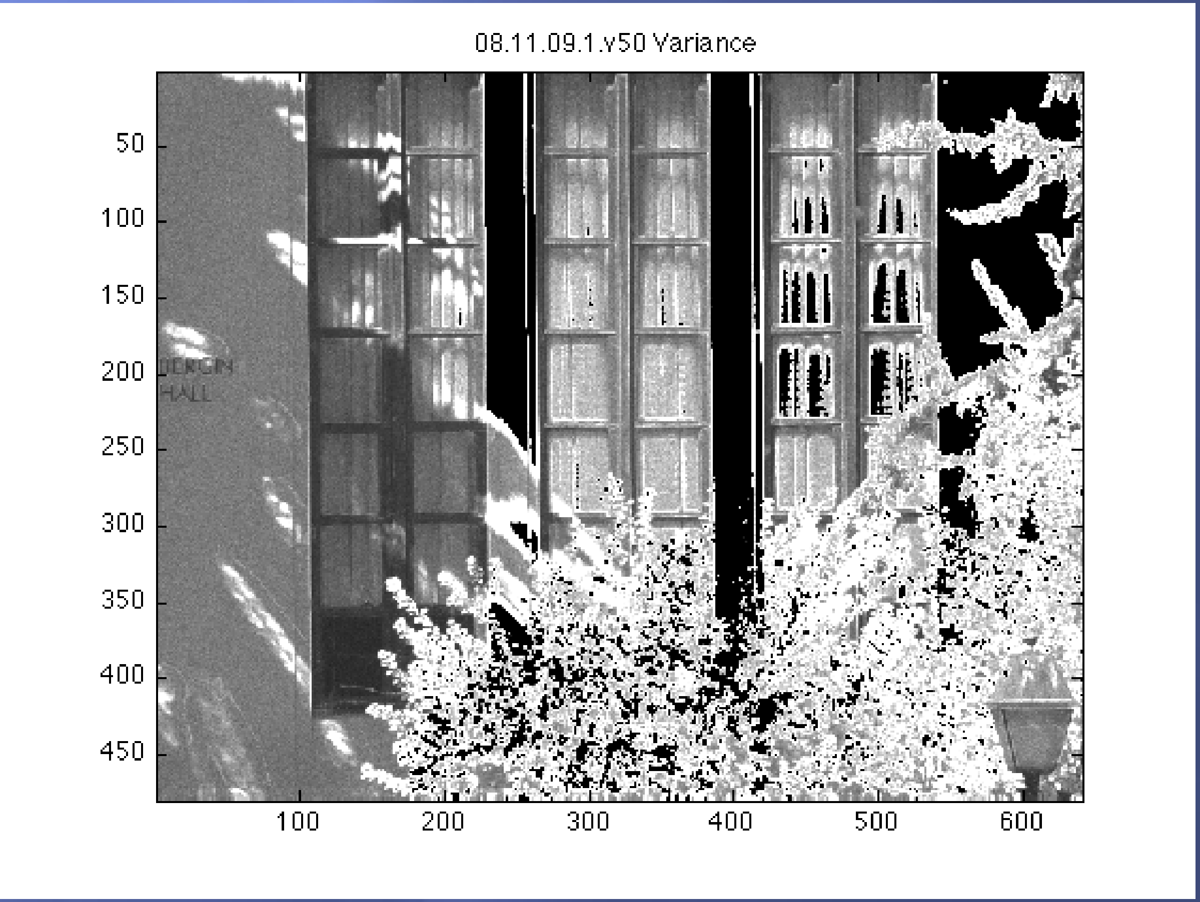
Mean Value of 50 images

- From 50 images of same scene



Variance of 50 Images

- From 50 images of same scene



Observations

- Averaged image has less variation due to noise
- Variance image looks very much like mean image
  - Higher than expected variance is associated with motion in bushes and trees
  - Low variance (zero variance) is associated with saturation. In the variance image zero variance areas are black.

Computation

How Can HR Images Be Computed?

- Use model of the image acquisition for the LR images in terms of the desired HR images
  - Optics, Sensors, LR image registration
  - $\underline{g} = H\underline{f} + \underline{v}$
- Use computational methods from signal and image processing and medical image reconstruction from projection measurements
  - Statistical methods: Kalman filter, maximum likelihood
  - Iterative methods: predicted error backprojection, POCS

Backprojection Methods

- **Simple backprojection –**

$$\underline{f}_B = H^T \underline{g} = H^T \left(H \underline{f} + \underline{v} \right)$$

  - **Simple computation**
  - **Close to least sqares result for very high noise levels**
- **Iterative back projection of prediction error - Kaczmarz method or ART. For individual LR pixels**

$$\Delta g_i = g_i - \underline{h}_i^T \cdot \underline{f}^i$$

$$\underline{f}^{i+1} = \underline{f}^i + \frac{\Delta g_i \underline{h}_i}{\left| \underline{h}_i \right|^2}$$

  - **New $\underline{f}^{i+1}$ is consistent with g_i. Relaxation coefficient may be used for partial correction for better convergence behavior.**
  - **Many iterations needed. May not converge.**

Statistical Estimation

- **Linear update with Gaussian assumptions. For zero mean:**

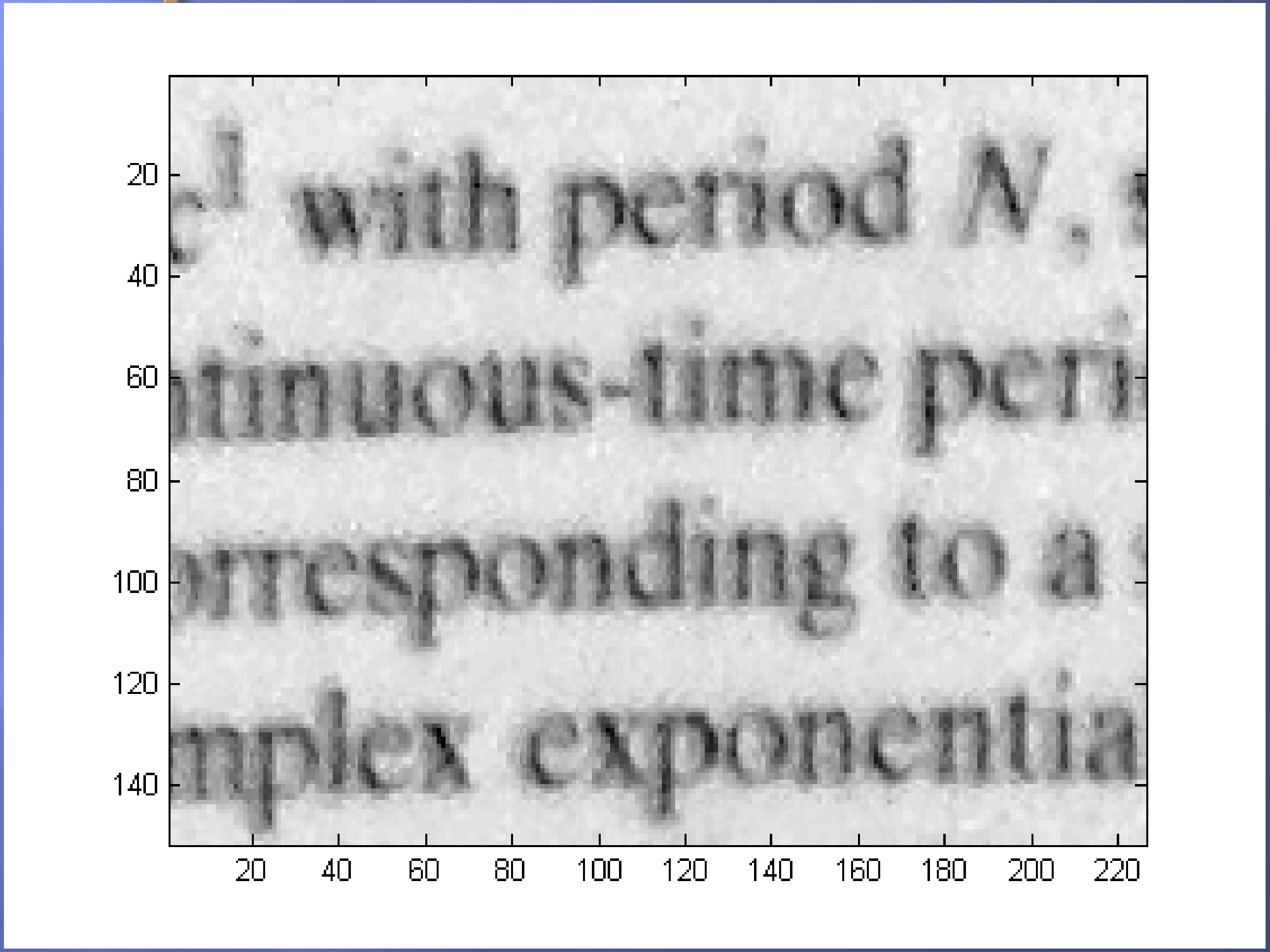
$$\underline{\hat{f}} = \boldsymbol{K}\underline{\boldsymbol{g}}$$

$$\boldsymbol{K} = \boldsymbol{P_0 H^T}\left(\boldsymbol{H P_0 H^T} + \boldsymbol{R_{vv}}\right)^{-1} = \boldsymbol{R_{fg}}\left(\boldsymbol{R_{gg}}\right)^{-1}$$

$$\xi = (I - KH)P_0(I - KH)^T + K\boldsymbol{R_{vv}}K^T$$

  - Potentially large computational requirement
  - Can use circulant matrix approximations (Milanfar) for more efficient computation.

Compression Inconsistency



JPEG compression: sharp edges create ringing high frequency content that is visually plausible but not correct and not consistent as image scene is shifted relative to the pixel grid.

Summary of Challenges

- Accurate registration of low resolution images to generate correct H matrices is difficult.
- High noise level from most image sensors
- Low dynamic range of input image values. (Note that the human visual system can generally distinguish about 6.5 bits of gray level.)
- Computational requirements
- Aliased high frequency content is attenuated by averaging of each sensor. Null space is created by sensor spacing.
- If image data is already compressed, almost all aliased high frequency content is eliminated or modified in a way that makes it not useful

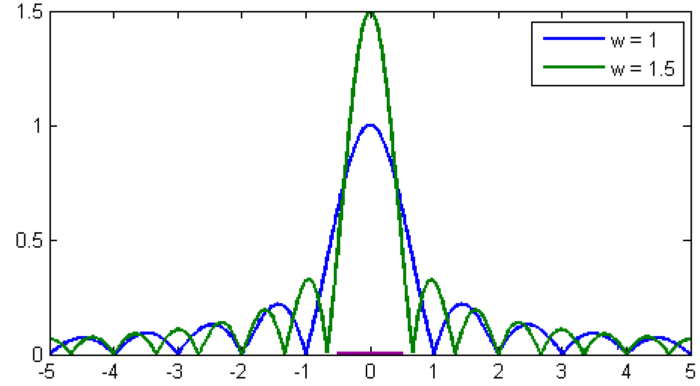
Computational Strategies

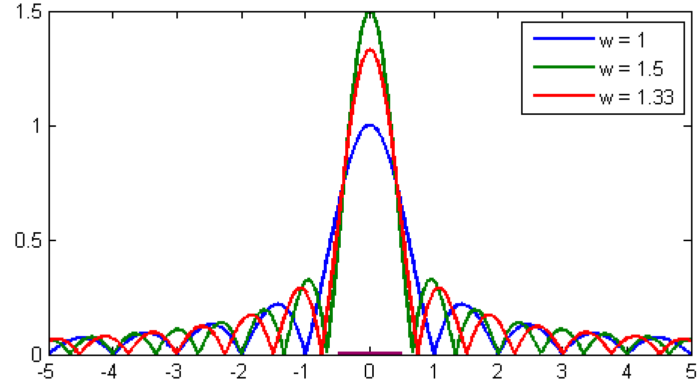
- H matrices are sparse and structured.
- Circulant approximations and frequency domain implementations can lead to efficient computation with large matrices (Vandewalle, Milanfar) .
- Local reconstruction of small tiles of the desired HR image has many benefits.
  - Less sensitivity to model mismatch
  - More manageable computation
- Potential problems:
  - Space varying optics creates problems in aligning LR image content – need accurate warping model
  - Compression creates artifacts that are not consistent in LR images

Diversity of Imaging Systems to Reduce Null Space

- Use of magnification and orientation diversity can reduce impact of null space due to regular and uniform sensor spacing.
- For a square pixel model, $A(u,v) = a^2 sinc(au) sinc(av)$ will attenuate high frequencies and null spatial frequencies at multiples of $\left(\frac{1}{a}\right)$.
  - If some LR image sensors are rotated, lines of zero response in frequency will not overlap.
  - If some LR images have lower magnification, the null frequencies will be different.

Continuous Frequency Response of Diverse Sensor Array

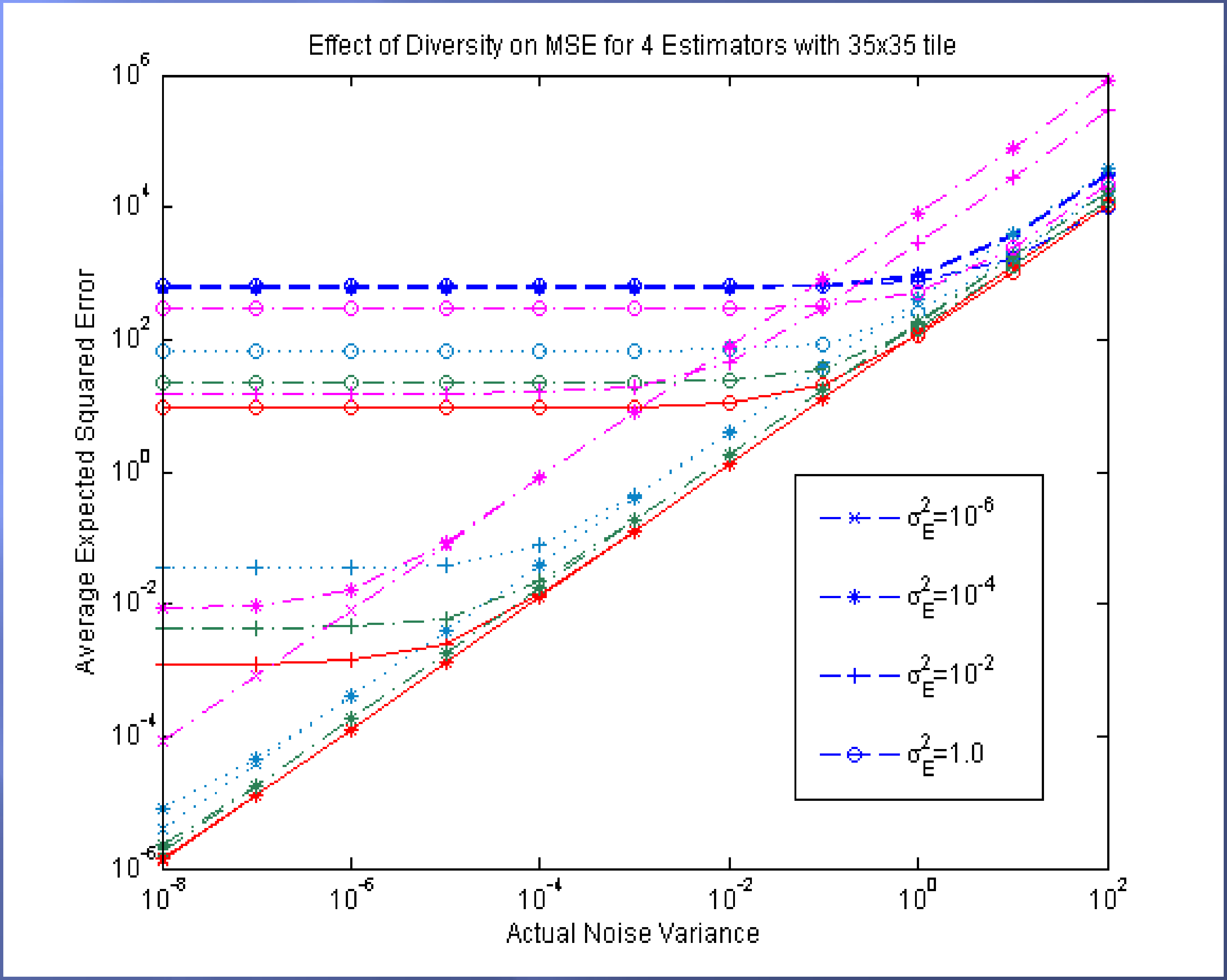




- Two different pixel widths provide non zero response at first null frequencies.
- Three different widths

Effect of Sub-imager Geometry

- Performance of five sub-imager geometries is compared for 35x35 pixel tile for diverse and uniform arrays.
  - Geometry is indicated by color.
  - The mse of four estimators using four different assumed noise variances is compared for each geometry. Noise level is indicated by marker shape.
  - Horizontal axis is actual noise level.
  - With diversity of sensor size and/or orientation, expected error is reduced to a level controlled by the measurement noise level, not the null space.


Effect of Diversity on MSE for 4 Estimators with 35x35 tile
Average Expected Squared Error
Actual Noise Variance
10^{6}
10^{4}
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-6}
10^{-4}
10^{-2}
10^{0}
10^{2}
$\sigma_E^2=10^{-6}$
$\sigma_E^2=10^{-4}$
$\sigma_E^2=10^{-2}$
$\sigma_E^2=1.0$

Simulated Example

Test image

Selected 35x35 tile

(a)

(b)

Two Reconstruction with Same Number of LR images

All LR at best magnification | Diverse LR Imagers

(c)

(d)

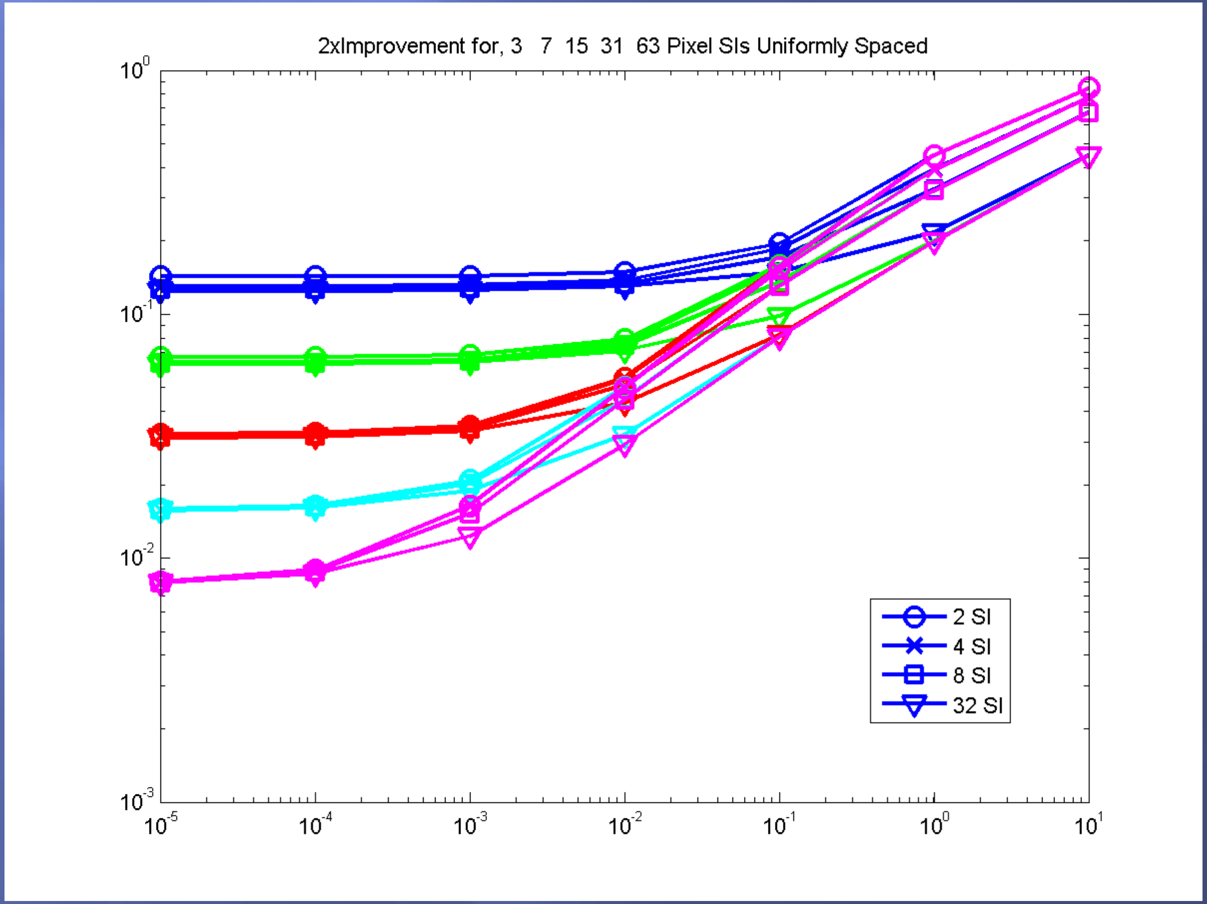
$$\sigma^2 = 0.01$$

Registration Errors

- If registration is perfect, expected error in reconstructed images is reduced when larger tiles are reconstructed from larger LR images. This is due to finer frequency resolution from a larger number of data samples.
- However, if the assumed relative position of the LR images has some variability, the benefit of the larger tiles is lost.

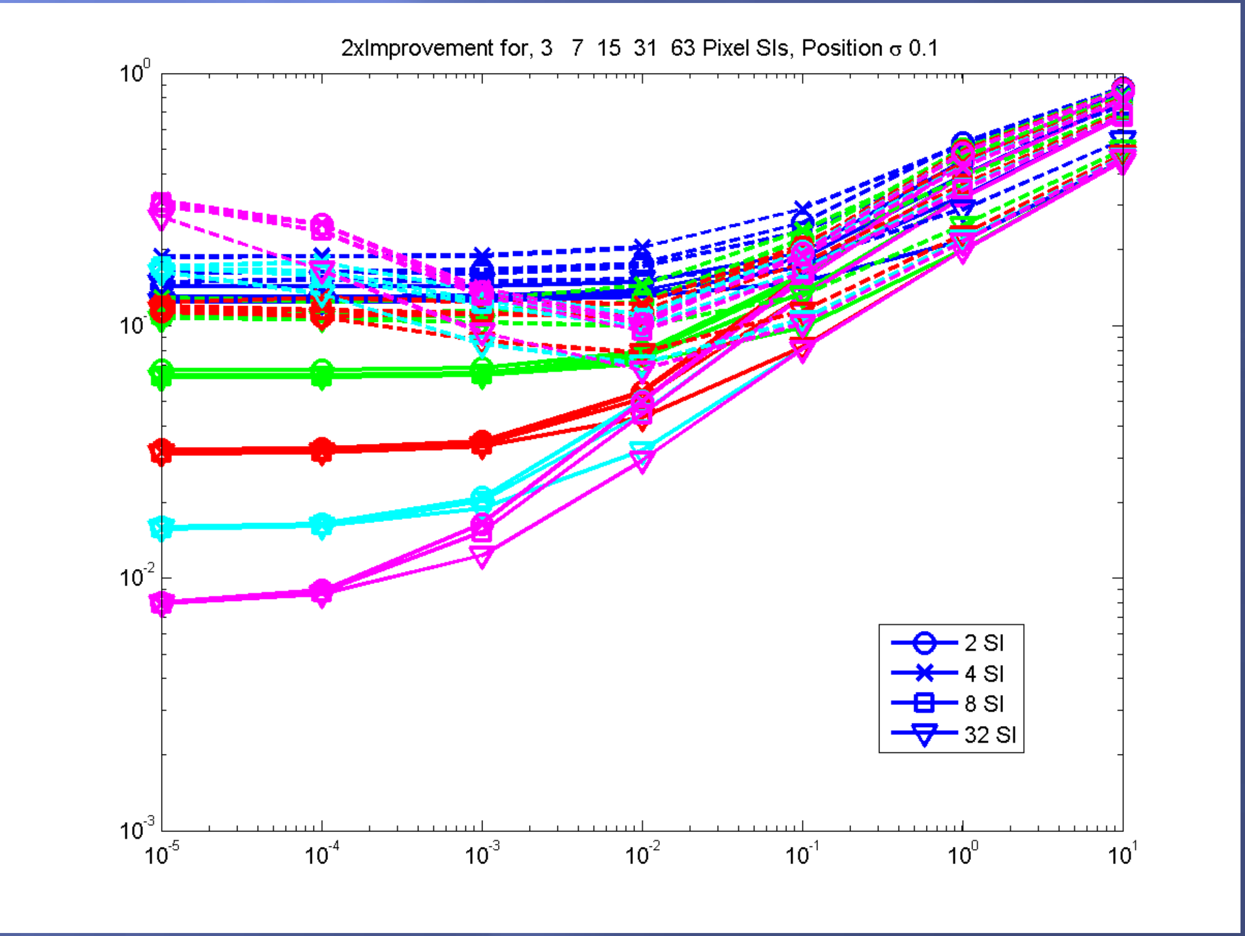
Perfectly known LR Positions

- Marker shape indicates number of LR images and color indicates size of LR image.



Random Registration Error

- Registration error with variance =0.1*desired pixel width shown with dashed lines.



Performance Measures

Objective and Subjective Measures

For most images the precise value of an isolated pixel is not important. In medical imaging a pixel value may quantitatively represent a physical quantity such as X-ray attenuation coefficient

Images vs Man-made Signals

- Man-made signals, e.g. communications, are designed to be efficient with other man-made technology for transmission and reception
- Radar and sonar use man-made signals to actively interrogate the natural surroundings
- Medical images – reconstructed images of biological structure using active or passive measurements
  - Quantitative value of pixels is usually significant as a measured quantity
- Images – passive capture of light patterns through image acquisition system
  - No control of signal content
  - Adapted visual system
  - Exact quantitative value of a pixel is often not important. Value in context of other pixel values is more important.

Mean Squared Error

- $MSE = \frac{1}{N}\sum_{n=0}^{N-1}\left(\hat{f}_n - f_n\right)^2$
- Advantages
  - Well understood and easily computed
- Disadvantages
  - Not strongly correlated with perceptual judgment about quality
    - Adding bias does not change perception of image content, but greatly increases MSE
  - Lowest MSE may still not be a usable image
  - Measures quality of single image

Expected Mean Squared Error

- $EMSE = \frac{1}{N}\left\langle \sum_{n=0}^{N-1}\left(\widehat{f_n} - f_n\right)^2 \right\rangle$
- Advantages
  - Measure of expected error for reconstruction method based on likelihood of occurrence of source images and noise
- Disadvantages
  - Not strongly correlated with perceptual judgment about quality
  - Null space components add large component to EMSE but other more likely content may have low error

Peak Signal to Noise Ratio

- $PSNR = 10\log_{10}\frac{\left(max(|f_n|)\right)^2}{MSE}$
- Advantages
  - Most common measure of image quality; widely used so methods can be compared
- Disadvantages
  - Similar to MSE

Structural Similarity

- Over a local area (e.g. 11x11 pixels) similarity of nonnegative valued x and y is computed for
  - Luminance: $l(x,y) = \frac{2\mu_x\mu_y + C_1}{(\mu_x)^2 + (\mu_y)^2 + C_1}$
  - Contrast: $c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{(\sigma_x)^2 + (\sigma_y)^2 + C_2}$
  - Structure: $s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$
  - SSIM = $l(x,y)\, c(x,y)\, s(x,y)$
- Correlation with perceptual judgment is not much stronger than PSNR.

Subjective Measure

- User preference based on viewing images is averaged over viewers
- Advantages
  - Attempts to measure quality that is relevant to human perception
- Disadvantages
  - Different observers may have significantly difference preferences
  - Different image qualities may be preferred based on image content – e.g. contrast
  - Other factors may affect preference

What is the Purpose of the Image?

- Beautiful HR image – user preference
- Identification of text or numeric content e.g. license plate numbers of ship numbers – use accuracy of identification regardless of the image appearance
- Detection of small objects or motion
- Pixel values represent quantitative measures– e.g. x-ray attenuation – use MSE

- **Identify goal of super-resolution and use attainment of goal as performance measure**

LESSONS LEARNED

Objective of Super-rresolution

- Improved viewing
  - Subjective measure of performance
  - Metrics can be similar to compression
  - Cost of errors
- Quantitative improvement in detection or identification
  - Can use quantitative measures of correctness not based on pixel values
    - Correct identification of structures in images
  - Ideal performance limited by
    - Measurement noise – image pixel variance
    - Uncertainty of image acquisition parameters and model parameters
  - Local computation may be sufficient

Lessons Learned

- Simulations work much better than real experiments. With perfect registration and no noise, large improvements in resolution are possible in theory. But
  - Low levels of noise are not possible even with a lot of averaging. Pixels are becoming smaller and noisier.
  - Accurate registration is very difficult.
    - A single camera in motion must compute position
    - Multiple fixed cameras may have known positions but output must be calibrated

Lessons Learned

- Accurately registered and interpolated LR images can provide meaningful results even if still blurred by the sensor response function. This makes computation more stable.
- Local supper-resolution computation using small image tiles is advantageous.
- Unmodeled sources of error
  - Motion of scene objects
  - Built in camera processing. For example, JPEG compression reduces high frequency content and replaces much of it with something that is visually plausible, but not correct and not consistent in LR images.
- Performance measures should be related to the objective of the super-resolution.

RELATIONSHIP TO NEWER METHODS OF SINGLE IMAGE SUPER-RESOLUTION

Single Image Super-resolution

- Viewing resolution of a single image can always be improved by up-sampling and interpolation – bandlimited, bi-linear, bi-cubic
  - High frequency content is not added by this process. To add helpful high frequency content, need some a-priori assumptions or reference data to select best content
- Data-driven optimization approaches use large data sets to provide information
  - Can use sparse coding (Wang, 2015) and/or deep convolutional networks (Dong, 2014) based on internal image similarities or large training sets to learn a mapping between low-resolution and high-resolution data
  - Avoids ill-conditioned reconstruction and deblurring operations
  - Can be tuned for specific applications

QUESTIONS?