(Super) Resolution: Statistical Definition, Computation, and Fundamental Limits

Peyman Milanfar*
EE Department
University of California, Santa Cruz

*Joint work with Sina Farsiu, Miki Elad, Dirk Robinson, Morteza Shahram, Hiro Takeda

International Conference on Super-Resolution Imaging
Hong Kong, August 28-31, 2005
Inverse Problems

- Noise sensitivity
- Non-uniqueness
- Numerical conditioning

Milanfar et al. EE Dept, UCSC
Inverse Problem of Interest

\[ LR(x, y, t) = \downarrow_N [HR(x, y, t) * PSF(x, y)] + noise \]

Forward Model

Detailed Scene

Cheap Camera

Low resolution images

Image(t_1), Image(t_2), ....

Inverse Problem

Compute \( HR(x, y, t) \)
Information in Imaging

Communication System:

Source → Transmitter → Channel → Receiver → Sink

Optical Imaging:

Physical Scene → Illumination & Reflection → Camera Optics → Imaging Chip → Processor

Non-uniform Illumination in space and time
Lens Engineering, zoom, focus, etc.
Space/time sampling rate, Foveation
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”
  - ...

Milanfar et al. EE Dept, UCSC
Agenda

• Sensor model and limitations
• Processing algorithms
• SR Performance limits
• Further extensions and directions
Sensing: Resolution Limits of a Canonical Image Sensor
Pinhole Camera

- The image of two sources is the incoherent sum of PSFs, representing the effect of the diffraction.

- When the point sources are “too close”, according to the Rayleigh criterion these two point sources are not resolvable.
Imaging Closely-Spaced Point Sources

Rayleigh’s limit isn’t.
Resolution: Stochastic Problem

- Point sources:

- Measured Signal:

\[ \alpha h(x_k - p_k, y_l - p_y) + \beta h(x_k + q_k, y_l + q_y) + w(x_k, y_l) \]

Resolution as a composite statistical hypothesis test:

\[ \mathcal{H}_0 : d = 0 \quad \text{One peak is present} \]
\[ \mathcal{H}_1 : d > 0 \quad \text{Two peaks are present} \]
Optimal Solution and a Scaling Law

• "Information capacity": What is the minimum SNR required to detect the presence of two point sources with high confidence?

\[
\text{SNR} \approx \frac{C}{N d^4}
\]

Depends on the sensor PSF, the required false alarm and tolerable error rates. (Optimize!)

Number of samples at focal plane array


M. Shahram, and P. Milanfar, "Statistical and Information-Theoretic Analysis of Resolution in Imaging", to appear in *IEEE Transactions on Information Theory*

Milanfar et al. EE Dept, UCSC
Example for super-critical sampling

What happens when there is aliasing?

Milanfar et al. EE Dept, UCSC
Example for sub-critical sampling (50% below Nyquist, two frames)

Message: Things can get a lot worse, but not impossible!

Milanfar et al. EE Dept, UCSC
Processing: Multi-frame Resolution Enhancement (Super-resolution)
Why Spatial Resolution Enhancement?

- To obtain an alias-free, “diffraction limited” image we need 4 pixels covering the Airy disk:

- That is: radius of the Airy disk must match the pixel dimensions.
Motivation: SuperResolution Goes to Hollywood

CBS Program “Numb3rs” Episode from March 11, 2005
Overcoming Sensor Limitations by Processing

The Idea: “Diversity” + Aliasing

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

Data Courtesy USAF
Resolution Enhancement Model

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

\[
\begin{align*}
    y_1 &= a_1 f_1 + a_2 f_2 + a_3 f_3 + a_4 f_4 + e_1 \\
    y_2 &= 0 \cdot f_1 + a_1 f_2 + 0 \cdot f_3 + a_3 f_4 + e_2 \\
    y_3 &= 0 \cdot f_1 + 0 \cdot f_2 + a_1 f_3 + a_2 f_4 + e_3 \\
    y_4 &= 0 \cdot f_1 + 0 \cdot f_2 + 0 \cdot f_3 + a_1 f_4 + e_4
\end{align*}
\]
Low vs High Res Pixels

x2 enhancement:
Need 4 frames.
The Mathematical Model

\[ y_k = A_k \mathbf{f} + e_k \quad \text{for} \quad 1 \leq k \leq p \]

\[ A_k = DHF(v_k) \]

• 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
The Optimization Problem

\[ \{ \hat{f}, \hat{v} \} = \arg\min_{f, v} \left[ \left\| A(v)f - y \right\|_1 + \right. \]

\[ \left. \lambda \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{\left|m+|l|\right|} \left\| f - S_x^l S_y^m f \right\|_1 \right] \]

0 < \alpha < 1

Data Info: Builds robustness to model uncertainty

L1 Prior: Incorporates multiscale model of edges

Why this $L_1$ prior?

- Let’s look at pixel differences across scales

\[ I_{l,m} = f - S_x S_y f \]
Histograms of $I_{l,m} = f - S_x^l S_y^m f$
Data Courtesy  Vigilant Technology

Milanfar et al. EE Dept, UCSC
Before
After: 4x
Detail Before

Data Courtesy  Vigilant Technology
Data Courtesy  Vigilant Technology
Security Camera (before/after)

60 input frames
Processing Limits: Statistical Bounds on Super-Resolution Performance
Review Basic Formulation

• Consider a sequence of noisy, translating images \( \{y_k\} \) over time.

\[
y_k = \text{Translate} \left( y_{k-1}, v_{k-1,k} \right) + \text{error}
\]

Frame-to-frame motion vectors

• Image formation model:

\[
y_k = \text{Sample}[f(x, y, t_k) * h(x, y)] + \text{noise}
\]

Aliasing

Point-spread function

Milanfar et al. EE Dept, UCSC
Fusion of Multiple Video Frames

- **Reconstruction Problem**: Given the frames, estimate the high resolution image \( f(x, y, t) \). (Superresolution)
  
  - Implicit problem: Estimate the motion vectors from aliased images

\[
f_k = \text{Translate} \left( f_j, v_{j,k} \right) + \text{error}
\]

\[
f_k = \text{Sample} \left[ f(x, y, t_k) * h(x, y) \right] + \text{noise}
\]

Milanfar et al. EE Dept, UCSC
Registration of Multiple Aliased Images

- **Motion Problem**: Given the frames, estimate vectors \( \{v_{j,k}\} \)

  \[
  \text{Implicit problem: Estimate underlying high resolution image from aliased data}
  \]

\[
\begin{align*}
\mathbf{f}_k &= \text{Translate} (\mathbf{f}_j, v_{j,k}) + \text{error} \\
\mathbf{f}_k &= \text{Sample}[f(x, y, t_k) * h(x, y)] + \text{noise}
\end{align*}
\]

Registering Aliased Images: A Very Poorly Understood Problem
How well can the problem be solved?

Estimation approach: Look at the Fisher Information (hence the Cramer-Rao bound).

\[
J\left(\{v_{j,k}\}, f\right) = \begin{bmatrix}
J_{vv} & J_{fv} \\
J_{fv}^T & J_{ff}
\end{bmatrix}
\]

- \(J_{vv}\) - Depends on the set of motions (sampling offsets) and the amount of texture energy in the signal
- \(J_{ff}\) - Depends only on the set of motions

Milanfar et al. EE Dept, UCSC
CRB for Aliased Image Registration

Using Schur decomposition, the CRB for aliased image registration is:

$$\text{Cov}\left(\left\{ v_{j,k} \right\} \right) \geq \left( J_{vv} - J_{fv} J_{ff}^{-1} J_{vf}^T \right)^{-1}$$

With just a pair of aliased images, the FIM is generically singular, hence unbiased pairwise registration of aliased images is essentially impossible. (Not so in absence of aliasing!)

Registering **Sets** of Images

CR Bound (per frame) for multi-frame image registration.
Insights gained:

- How much information is lost by needing to estimate the motion vectors?

- How many frames to get a decent answer?

We expect a 10-20% loss in MSE performance. (A lot!)
Further Extensions: Color, Dynamics, Algorithmic Improvements
Color Super-Resolution

• Two types of input to consider:
  – Raw CFA data
  – Full RGB fields

• Unified Treatment
Simultaneous Demosaicing and Super-Resolution

Bayer Filtered Motion Sequence

Single-Frame Demosaicing

OLD

Image “fusion”

NEW

Hi-resolution Demosaicing

Milanfar et al. EE Dept, UCSC
Characteristics of Color Algorithm

I. Robust to the data noise and motion estimation errors ($L_1$ Norm).
II. Sharp edges in luminance component ($L_1$).
III. Minimize artifact in the chrominance component ($L_2$ Norm).
IV. Similar edge location-orientation in all color bands.


RGB Color Security Camera

24 input frames
RGB Color Super-Resolution

40 input frames, resolution enhancement factor of x4

Milanfar et al. EE Dept, UCSC
Demosaiced from 1-CCD CFA Camera

24-Frame Demosaicing and Reconstruction x3

Data courtesy of Technion
Dynamic Super-Resolution

- Naïve Approach:

- Right Approach:

Dynamic Super-Resolution

- Adapted for color
- Improved robustness
- Different implementation

Video-to-Video Example I

LR Video

HR Outcome

Milanfar et al. EE Dept, UCSC
Software

MotionDSP (Milanfar, Farsiu, Elad)
Further Refinements

• Computationally, it still makes the most sense to solve the motion/fusion problems in series.
  – Need extremely accurate motion estimation.
  – Need excellent filtering, interpolation.
Better Motion Estimation

• Almost all motion estimation algorithms today deal with the case of only two (consecutive) frames at a time.

Pairwise estimation (“Progressive”)

Fixed reference estimation (“Anchored”)
Constraints on Motion Vectors Across Time

Translation case:

- Linear set of constraints imply that the motion vectors live in a subspace.
- For more general motion, you have nonlinear constraints, → group structure.
- Applicable to any core algorithm.

\[ v_{i,k} = v_{i,j} + v_{j,k} \]
\[ v_{k,j} = -v_{j,k} \]
\[ v_{i,i} = 0 \]

Better Filtering and Interpolation: The Kernel Regression Idea

Data:

\[ y_i = z(x_i) + \varepsilon_i, \quad x_i = [x_1, x_2]^T \]

Local Polynomial Kernel Regression:

\[
\arg\min_{\{\beta_n\}} \sum_{i=1}^{P} \left[ y_i - \beta_0 - \beta_1^T (x_i - x) - \beta_2^T \text{vech}\{(x_i - x)(x_i - x)^T\} - \cdots \right]^2 K_H(x_i - x)
\]
Even Better: Adaptive Kernel Regression

Consider the **Denoising** Problem First:

\[
\arg\min_{\{\beta_n\}} \sum_{i=1}^{P} \left[ y_i - \beta_0 - \beta_1^T (x_i - x_j) - \cdots \right]^2 K_H(x_i - x_j) K_g(y_i - y_j)
\]

Special Case (order=0):

\[
\hat{\hat{x}}(x_j; 0, H, g) = \frac{\sum_{i=1}^{P} K_H(x_i - x_j) K_g(y_i - y_j) y_i}{\sum_{i=1}^{P} K_H(x_i - x_j) K_g(y_i - y_j)}
\]

\[\rightarrow\text{ Bilateral filter} \quad \text{(Gaussian Kernels)} \quad \text{Tomasi ('98) Elad ('01)}\]
Choice of Radiometric Kernel

• $K_g$ implicitly exploits the local gradient information. (e.g. BF, $K_g = \exp\left(-\frac{(y_i - y_j)^2}{\sigma_g^2}\right)$)

• Improved solution is possible by explicit incorporation of orientation info.

Adaptive Kernel Regression

- **Steerable Kernel Regression**

\[
\arg \min_{\{\beta_n\}} \sum_{i=1}^{P} \left[ y_i - \beta_0 - \beta_1^T (x_i - \mathbf{x}) - \cdots \right]^2 \frac{1}{h s_i} K \left( \frac{(x_i - \mathbf{x})^T S_i (x_i - \mathbf{x})}{h^2} \right)
\]

Local Steering matrix

Global Scaling parameter

Proceedings of the Asilomar Conference on Signals and Systems, Oct. 2005,
Pacific Grove, CA
Image Denoising

Original image

Noisy, sigma = 25
Milanfar et al. EE Dept, UCSC
Denoising Results

Bilateral filter, RMSE = 8.66

Adaptive kernel, Steered, order = 0, RMSE = 7.04

Standard kernel, order = 2, RMSE = 10.21

Adaptive kernel, Steered, order = 2, RMSE = 7.03

Gaussian kernel

Milanfar et al. EE Dept, UCSC
What About Fully Adaptive Kernel-based Interpolation?

**The Problem**: At the location \( x \) where we wish to interpolate, there is no pixel value (yet).

**The Solution**: Produce a “pilot”, low-complexity, estimate of the pixel \( y_i \) then apply the more sophisticated adaptive kernel techniques described earlier.

The process can in fact be iterated for further improvement.

Milanfar et al. EE Dept, UCSC
Illustration:

30% of Pixels Retained  Local Constant Kernel Estimate  Adaptive Kernel Estimate

Milanfar et al. EE Dept, UCSC
Upsampling Example: Interpolation from Regular Samples

Standard kernel, order = 2
RMSE = 8.32

Steerable kernel, order = 2
RMSE = 7.59
Interpolation from Irregular Samples

Random Downsampling 85%

Standard kernel, order = 2
RMSE = 9.35

Steerable kernel, order = 2
RMSE = 8.38

EE Dept, UCSC
SR Example

Low resolution video, 8 frames

Local constant estimator, order=0

Local quadratic estimator, order=2

Estimated scene
Super-resolution Example

Resolution enhancement from video frames captured by a commercial webcam (3COM Model No.3719)
Some Final Remarks

• SR is an idea whose time has come.

• Time to seriously consider applications.

• **A prediction**: in 5-7 years, SR will be used routinely in consumer products.
The Life of Super-Resolution so-far

- **A pregant idea:** Super-Res is conceived
  - Yen (’56) and Papoulis (’77) Sampling Theorems

- **Birth:** A first super-resolution algorithm
  - Tsai and Huang (’84)

- **Toddler:** Back-projection methods
  - Peleg, Keren, Schweitzer (’87), Peleg and Irani (’90)

- **Early Education:** Some formal signal processing
  - Bose (’90), Tekalp et al (’92)

- **Pre-teen:** The facts of life
  - Elad (’95), Katsaggelos (’95), Schultz (’95), Foroosh (’95)

- **Teenager:** Getting good with numbers, and learning to learn
  - Nguyen, Milanfar, Golub (’98), Baker (’99)

- **College:** Color, compression, stability, learn to adapt better
  - See Special Issue of EURASIP JASP

- **TODAY:** SR has recently graduated from college.
  - **Time to get a job** and become useful.
  - Or go to graduate school….