Learning-Based Image Super-Resolution

Zhouchen Lin
zhoulin@microsoft.com

Microsoft Research Asia

Nov. 8, 2008

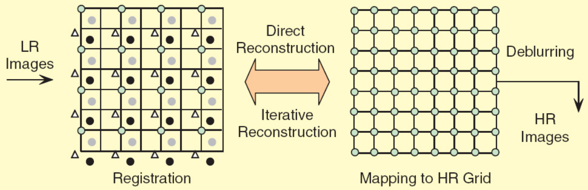
Outline

What is Super-Resolution?

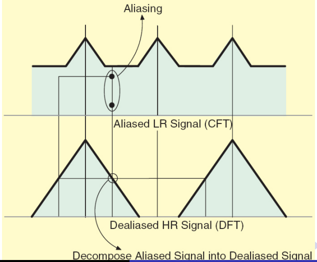
- SR is a technique that increases image/video details
- SR vs. Interpolation and Enhancement

Classification of SR Algorithms (1)

- Interpolation-based: register + interpolate + deblur

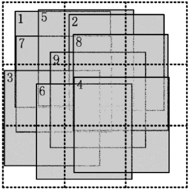


- Frequency-based: dealias



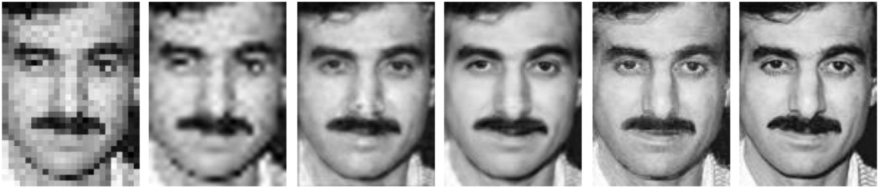
Classification of SR Algorithms (2)

- Reconstruction-based: register + weak prior + solve a linear system



$$L=PH+E$$

- Learning-based: knowledge + inference



(a) Input 24*32 (b) Cubic B-Spline (c) Baker et al. (d) Liu et al. (e) Our method (f) Original 96*128

Advantages & Disadvantages of Learning-Based SR

- \+ Require fewer low-res images, even single image!
- \+ Achieve higher magnification factors (MF)
- \+ Faster
- \+ More versatile, e.g., style transfer
- – Work with fixed MFs
- – Performance unpredictable

Input Train: generic Train: noise Train: rects

Classification of Learning-Based SR Algorithms (1)

Based on applications:

- For general images
- For specific images
    - only face/text images

Classification of Learning-Based SR Algorithms (2)

Based on implementations:

- Indirect maximum a posteriori (MAP)

$$\mathbf{H} = \arg\max_{\mathbf{H}} P\left(\left\{\dot{\mathbf{L}}_i\right\}_{i=1}^{N} \middle| \dot{\mathbf{H}}\right) P\left(\dot{\mathbf{H}}\right)$$

  - Local: Infer the HR image patch by patch
  - Global: Infer the coefficients of the bases for the HR image
- Direct MAP

$$\mathbf{H} = \arg\max_{\mathbf{H}} P\left(\dot{\mathbf{H}} \middle| \left\{\dot{\mathbf{L}}_i\right\}_{i=1}^{N}\right)$$

  - Local only

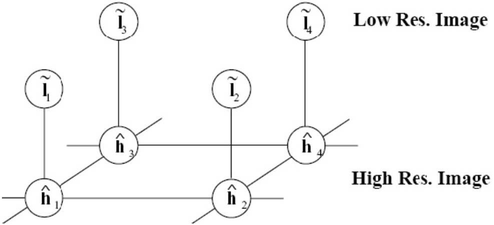
Representative Learning-Based SR Algorithms (1)

Local indirect maximum a posteriori (MAP):
Freeman & Pasztor. Learning Low-Level Vision. ICCV 1999.

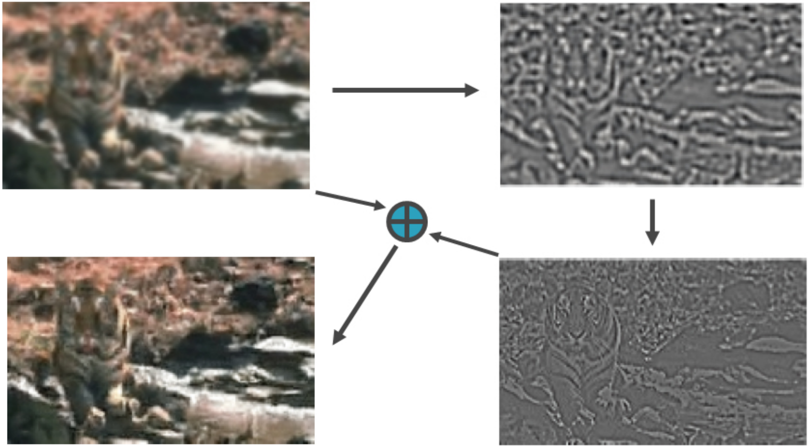
$$\mathbf{H} = \bar{\mathbf{L}} + \hat{\mathbf{H}}$$

$$\hat{\mathbf{H}} = \arg\max_{\hat{\mathbf{H}}} P\left(\tilde{\mathbf{L}}|\hat{\mathbf{H}}\right) P\left(\hat{\mathbf{H}}\right)$$

$$P\left(\tilde{\mathbf{L}}|\hat{\mathbf{H}}\right) = \prod_k P\left(\tilde{\mathbf{L}}_k|\hat{\mathbf{H}}_k\right), \quad P\left(\hat{\mathbf{H}}\right) = \prod_{\hat{\mathbf{H}}_j \in \mathcal{N}(\hat{\mathbf{H}}_i)} P\left(\hat{\mathbf{H}}_i|\hat{\mathbf{H}}_j\right)$$



Exemplar Result



Representative Learning-Based SR Algorithms (2)

Global indirect MAP: face hallucination only
Gunturk et al. Eigenface-Domain Super-Resolution for Face Recognition. IEEE T. Image Processing, 2003.

$$\mathbf{h} = \arg\max_{\mathbf{h}} P\left(\{\mathbf{l}_i\}_{i=1}^{N} \middle| \mathbf{h}\right) P(\mathbf{h})$$

$$P\left(\{\mathbf{l}_i\}_{i=1}^{N} \middle| \mathbf{h}\right) \sim \exp\left(-\sum_{i=1}^{N} \xi_i^t \mathbf{Q}^{-1} \xi_i\right), \quad \xi_i = \mathbf{l}_i - \mathbf{F}_l^t \mathbf{P}^{(i)} \mathbf{F}_h \mathbf{h} - \eta$$

$$P(\mathbf{h}) \sim \exp\left(-(\mathbf{h} - \mu_{\mathbf{h}})^t \Lambda^{-1} (\mathbf{h} - \mu_{\mathbf{h}})\right)$$

Exemplar Result

Representative Learning-Based SR Algorithms (3)

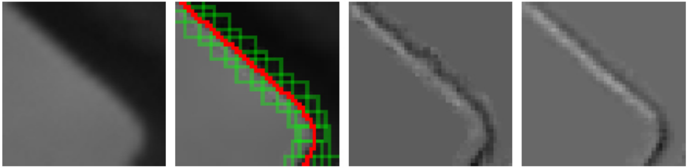
Local direct MAP:

Sun, Tao, and Shum. Image Hallucination with Primal Sketch Priors. CVPR 2003.

$$\mathbf{H} = \bar{\mathbf{L}} + \mathbf{H}^p$$

$$\mathbf{H}^p = \arg\max_{\mathbf{H}^p} P\left(\mathbf{H}^p|\bar{\mathbf{L}}\right)$$

$$P\left(\mathbf{H}^p|\bar{\mathbf{L}}\right) \approx \prod_k P(C_k|\bar{\mathbf{L}}), \quad P(C_k|\bar{\mathbf{L}}) \sim \prod_i^{n_k-1} \Psi(B_i^l, B_{i+1}^l) \prod_i^{n_k} \Phi(B_i^l, B_i^h)$$



(a) (b) (c) (d)

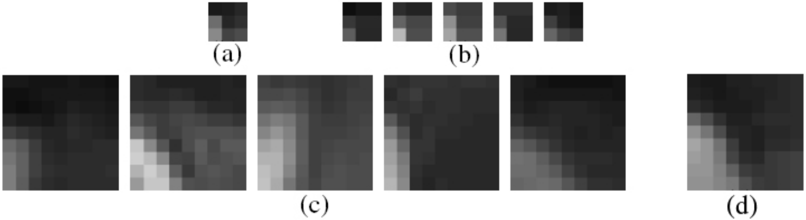
Exemplar Result

Representative Learning-Based SR Algorithms (4)

Using manifold learning techniques:

Chang, Yeung, and Xiong. Super-Resolution Through Neighbor Embedding. CVPR 2004.

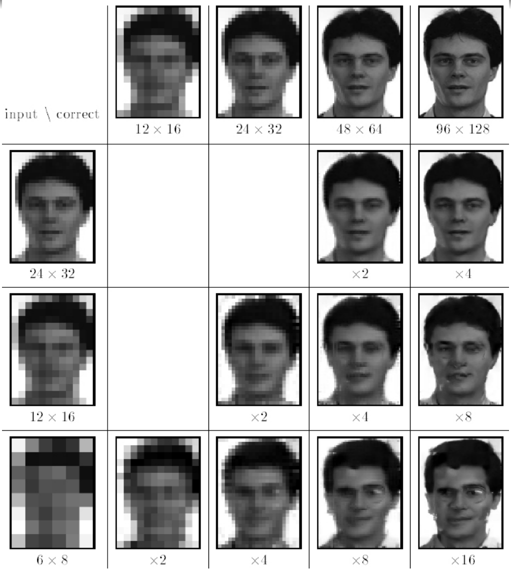
1. For each LR patch



2. Enforcing local compatibility and smoothness constraints between adjacent HR patches.

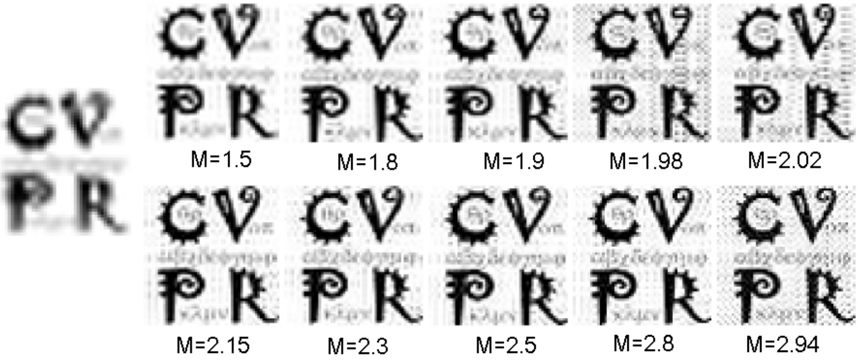
Exemplar Result

Do limits exist for Learning-Based SR?



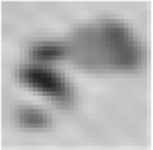
Related Work

Limits of Reconstruction-Based SR Algorithms [2]



What are limits of SR?

- Good SR result: close to the ground truth

- Average performance

Abstract Model of Learning-Based SR

- An SR Algorithm: a mapping s from LR image (low-dim space) to HR image (high-dim space)
- Average performance: expected risk

$$R(s) = \int r(h, s(d(h)))p(h)\mathrm{d}h$$

r: risk function, d: downsampling operator, $p(h)$: distribution

Problem Formulation

$$R(s) = g_s(N, m) = \left(\frac{1}{mN}\tilde{g}_s(N, m)\right)^{1/2}$$

$$\tilde{g}_s(N, m) = \int_{\mathbf{h}} \|\mathbf{h} - s(\mathbf{D}\mathbf{h})\|^2 p_h(\mathbf{h}) d\mathbf{h}$$

N: image size, m: magnification factor, $\mathbf{D}$: downsampling matrix

- Does not help if compute $g_s(N, m)$ for a particular s.
- Find lower bound $b(N, m)$ for $g_s(N, m)$ that is valid for all s.
- Lower bound is indefinite if no assumption on $p_h(\mathbf{h})$ is made.

Statistics of General Natural Images

- The distribution of HR images (HRI) is not concentrated around several HRIs, and the distribution of LR images (LRI) is not concentrated around several LRIs either.
- Smoother LRIs have a higher probability than nonsmooth ones.

Statistics of specific class of images is unclear.

Theorem 1: Lower Bound of the Expected Risk

$$\tilde{g}_s(N,m) = \int_{\mathbf{h}} \|\mathbf{h} - s(\mathbf{Dh})\|^2 p_h(\mathbf{h}) d\mathbf{h}$$

Theorem 1: $\tilde{g}_s(N,m)$ is lower bounded by $\tilde{b}(N,m)$, where

$$\tilde{b}(N,m) = \frac{1}{4}\mathrm{tr}\left[(\mathbf{I} - \mathbf{UD})\Sigma(\mathbf{I} - \mathbf{UD})^t\right] + \frac{1}{4}\|(\mathbf{I} - \mathbf{UD})\bar{\mathbf{h}}\|^2$$

$\mathbf{U}$: upsampling matrix, $\mathbf{DU} = \mathbf{I}$
Σ: covariance matrix, $\bar{\mathbf{h}}$: mean

Sketch of Proof (1)

$$\tilde{g}_s(N, m) = \int_{\mathbf{h}} \|\mathbf{h} - s(\mathbf{D}\mathbf{h})\|^2 p_h(\mathbf{h}) d\mathbf{h}$$

Choose **Q** and **V** such that $\begin{pmatrix} \mathbf{D} \\ \mathbf{Q} \end{pmatrix} (\mathbf{U} \quad \mathbf{V}) = \mathbf{I}$. Denote $\mathbf{M} = (\mathbf{U} \quad \mathbf{V})$. Perform transform $\mathbf{h} = \mathbf{M} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$, then

Sketch of Proof (2)

$$\tilde{g}(N,m) = \int\limits_{\mathbf{x},\mathbf{y}} \left\| (\mathbf{U} \quad \mathbf{V}) \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} - s(\mathbf{x}) \right\|^2 p_{x,y}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

$$= \int\limits_{\mathbf{x}} p_x(\mathbf{x}) V(\mathbf{x}) \mathrm{d}\mathbf{x},$$

$$p_{x,y}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \right) = |\mathbf{M}| p_h \left(\mathbf{M} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \right)$$

$$V(\mathbf{x}) = \int\limits_{\mathbf{y}} ||\mathbf{V}\mathbf{y} - \phi(\mathbf{x})||^2 \, \tilde{p}_y(\mathbf{y}|\, \mathbf{x}) \, \mathrm{d}\mathbf{y}$$

$$\phi(\mathbf{x}) = s(\mathbf{x}) - \mathbf{U}\mathbf{x}$$

Sketch of Proof (3)

$$V(\mathbf{x}) = \int\limits_{\mathbf{y}} \left|\left|\mathbf{V}\mathbf{y} - \phi\left(\mathbf{x}\right)\right|\right|^2 \tilde{p}_y\left(\mathbf{y}|\,\mathbf{x}\right)\mathrm{d}\mathbf{y}$$

$$\phi_{opt}(\mathbf{x};\tilde{p}_y) = \mathbf{V}\int\limits_{\mathbf{y}} \mathbf{y}\tilde{p}_y\left(\mathbf{y}|\,\mathbf{x}\right)\mathrm{d}\mathbf{y}$$

$$V(\mathbf{x}) = \int\limits_{\mathbf{y}} \left|\left|\mathbf{V}\mathbf{y}\right|\right|^2 \tilde{p}_y\left(\mathbf{y}|\,\mathbf{x}\right)\mathrm{d}\mathbf{y} - \left|\left|\phi_{opt}(\mathbf{x};\tilde{p}_y)\right|\right|^2$$

$$\left|\left|\phi_{opt}(\mathbf{x};\tilde{p}_y)\right|\right|^2 \le \frac{3}{4}\frac{\int_{\mathbf{y}} \left|\left|\mathbf{V}\mathbf{y}\right|\right|^2 p_{x,y}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right)\mathrm{d}\mathbf{y}}{p_x(\mathbf{x})}$$

Sketch of Proof (4)

$$
\begin{aligned}
\tilde{g}(N,m) &= \int_{\mathbf{x}} p_x(\mathbf{x}) \left(\int_{\mathbf{y}} ||\mathbf{V}\mathbf{y}||^2 \, \tilde{p}_y\left(\mathbf{y}|\,\mathbf{x}\right) \mathrm{d}\mathbf{y} - \left|\left|\phi_{opt}(\mathbf{x};\tilde{p}_y)\right|\right|^2 \right) \mathrm{d}\mathbf{x} \\
&\geq \frac{1}{4} \int_{\mathbf{x}} p_x(\mathbf{x}) \int_{\mathbf{y}} ||\mathbf{V}\mathbf{y}||^2 \, \tilde{p}_y\left(\mathbf{y}|\,\mathbf{x}\right) \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x} \\
&= \frac{1}{4} \int_{\mathbf{x},\mathbf{y}} ||\mathbf{V}\mathbf{y}||^2 \, p_{x,y}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \frac{1}{4} \int_{\mathbf{h}} ||\mathbf{V}\mathbf{Q}\mathbf{h}||^2 \, p_h(\mathbf{h})\mathrm{d}\mathbf{h} \\
&= \frac{1}{4}\mathrm{tr}\left((\mathbf{I}-\mathbf{U}\mathbf{D})\Sigma(\mathbf{I}-\mathbf{U}\mathbf{D})^t\right) + \frac{1}{4}\left|\left|(\mathbf{I}-\mathbf{U}\mathbf{D})\bar{\mathbf{h}}\right|\right|^2
\end{aligned}
$$

Estimating the Lower Bound via Sampling

Theorem 2: If we sample $M(p,\varepsilon) = \dfrac{(C_1 + 2C_2)^2}{16p\varepsilon^2}$ HRIs independently, then with probability of at least $1-p$, $|\hat{\tilde{b}}(N,m) - \tilde{b}(N,m)| < \varepsilon$.

$\hat{\tilde{b}}(N,m)$ is the value of $\tilde{b}(N,m)$ estimated from real samples,
$C_1 = \sqrt{E\left(\left|\left|(\mathbf{I}-\mathbf{UD})(\mathbf{h}-\bar{\mathbf{h}})\right|\right|^4\right) - \mathrm{tr}^2\left[(\mathbf{I}-\mathbf{UD})\Sigma(\mathbf{I}-\mathbf{UD})^t\right]}$,
and $C_2 = \sqrt{\bar{\mathbf{b}}^t\Sigma\bar{\mathbf{b}}}$, $\bar{\mathbf{b}} = (\mathbf{I}-\mathbf{UD})^t(\mathbf{I}-\mathbf{UD})\bar{\mathbf{h}}$.

Sketch of Proof (1)

$$\left|\hat{\tilde{b}}(N,m) - \tilde{b}(N,m)\right| \leq \frac{1}{4}|\xi - \mathbb{E}\xi| + \frac{1}{2}|\eta - \mathbb{E}\eta|$$

$$\xi = \mathrm{tr}(\mathbf{B}\hat{\hat{\Sigma}}_M), \quad \eta = \bar{\mathbf{b}}^t\hat{\hat{\mathbf{h}}}_M$$

$$\hat{\hat{\Sigma}}_M = \frac{1}{M}\sum_{k=1}^{M}(\hat{\mathbf{h}}_k - \bar{\mathbf{h}})(\hat{\mathbf{h}}_k - \bar{\mathbf{h}})^t, \quad \hat{\hat{\mathbf{h}}}_M = \frac{1}{M}\sum_{k=1}^{M}\hat{\mathbf{h}}_k$$

$$\mathbf{B} = (\mathbf{I} - \mathbf{UD})^t(\mathbf{I} - \mathbf{UD}), \quad \bar{\mathbf{b}} = \mathbf{B}\bar{\mathbf{h}}$$

Sketch of Proof (2)

$$\left|\hat{\tilde{b}}(N,m) - \tilde{b}(N,m)\right| \leq \frac{1}{4}|\xi - \mathbb{E}\xi| + \frac{1}{2}|\eta - \mathbb{E}\eta|$$

$$P(|\xi - \mathbb{E}\xi| \geq \delta) \leq \frac{\text{var}\xi}{\delta^2} = \frac{C_1^2}{M\delta^2}$$

$$P(|\eta - \mathbb{E}\eta| \geq \delta) \leq \frac{\text{var}\eta}{\delta^2} = \frac{C_2^2}{M\delta^2}$$

So with probability at least $1 - p$,

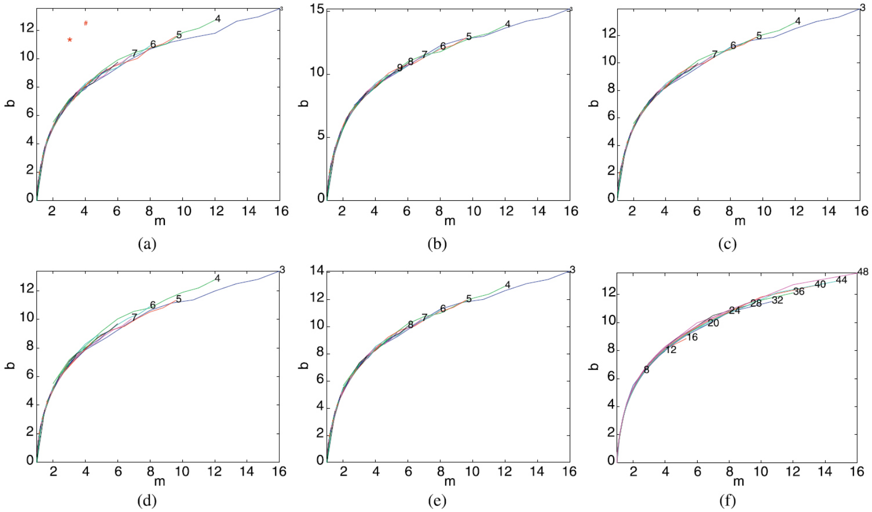
$$\left|\hat{\tilde{b}}(N,m) - \tilde{b}(N,m)\right| \leq \frac{C_1}{4\sqrt{Mp}} + \frac{C_2}{2\sqrt{Mp}}$$

Estimating the Limits via the Lower Bound

If at a particular MF m, $b(N, m)$ is larger than a threshold T, then at this MF no SR algorithm can effectively recover the original HRI:

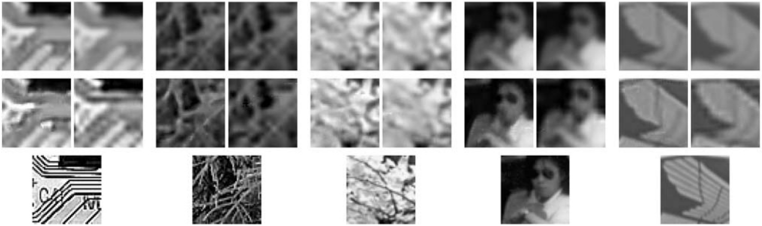
$$\text{limit} \leq b^{-1}(T)$$

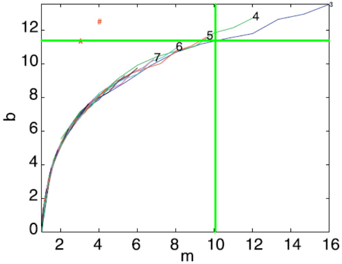
Experiments



Estimating the Limits

$T = 11.1$ is a large enough threshold.





Considering the Noise

To take noise into account, $\tilde{g}(N, m)$ should be changed to

$$\tilde{g}'(N, m) = \int_{\mathbf{h},\mathbf{n}} \left|\left|\mathbf{h} - s\left(\mathbf{Dh} + \mathbf{n}\right)\right|\right|^2 p_{h,n}\left(\begin{pmatrix} \mathbf{h} \\ \mathbf{n} \end{pmatrix}\right) \mathrm{d}\mathbf{h}\mathrm{d}\mathbf{n},$$

Accordingly,

$$\begin{aligned}\tilde{b}'(N, m) = \ & \frac{1}{4}\mathrm{tr}\left[(\mathbf{I} - \mathbf{UD})\Sigma(\mathbf{I} - \mathbf{UD})^t\right] \\ & + \frac{1}{4}\mathrm{tr}\left(\mathbf{U}\Sigma_n\mathbf{U}^t\right) + \frac{1}{4}\left|\left|(\mathbf{I} - \mathbf{UD})\bar{\mathbf{h}} - \mathbf{U}\bar{\mathbf{n}}\right|\right|^2.\end{aligned}$$

Future Work and Open Problems

- Tighter upper bound of the limits
- Limits of SR algorithms for specific image classes
- How to represent and incorporate the prior more effectively?
- How to make the algorithms scalable with the MF?
- What is the relationship between the SR performance and the training samples?
  - How to choose optimal training samples?

References

Zhouchen Lin et al. Limits of Learning-Based Superresolution Algorithms. Int'l J. Computer Vision, 2008.

Zhouchen Lin et al. Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation, IEEE T. PAMI, 2004.

Questions?