Image Super-resolution via Sparse Representation

Jianchao Yang, John Wright, Yi Ma

Coordinated Science Laboratory
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

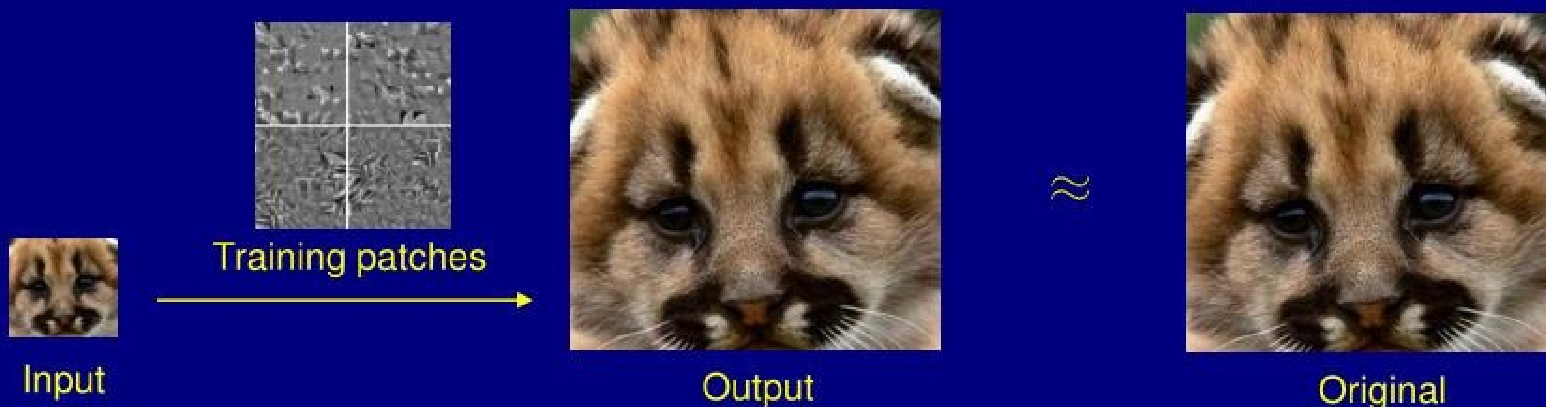
SIAM Imaging Science Symposium, July 7, 200

OUTLINE

- Super-resolution as sparse representation in dictionary of raw image patches
- Solution via ℓ^1 -norm minimization
- Global consistency, feature selection
- Experiments:
  - Qualitative comparison to previous methods
  - Quantitative comparison to previous methods
- Conclusions and Discussions

LEARNING-BASED SUPER-RESOLUTION – Problem formulation

Problem: given a single low-resolution input, and a set of pairs (high- and low-resolution) of training patches sampled from similar images, reconstruct a high-resolution version of the input.



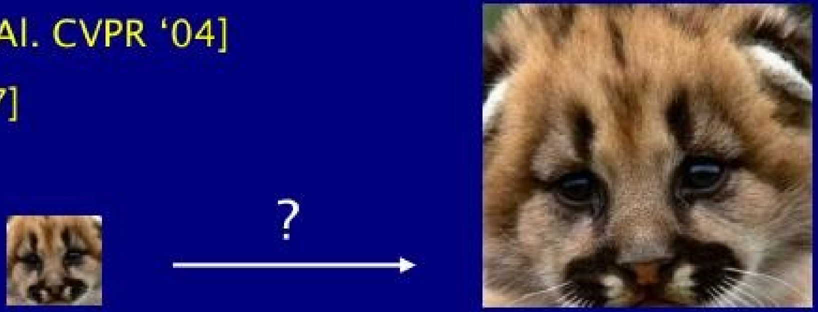
Advantage: more widely applicable than reconstructive (many image) approaches.

Difficulty: single-image super-resolution is an extremely ill-posed problem.

LEARNING-BASED SUPER-RESOLUTION – Prior work

How should we regularize the super-resolution problem?

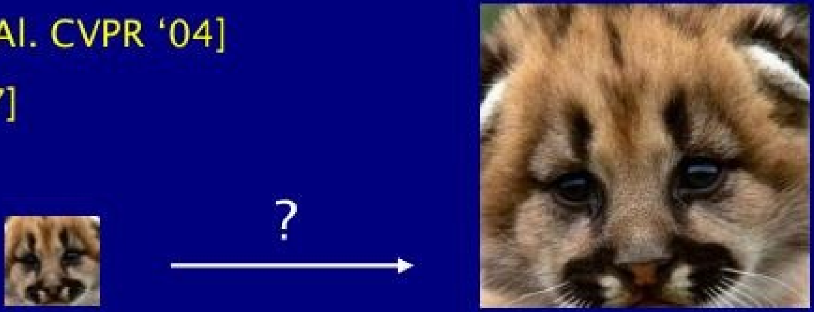
- –Markov random field [Freeman et. Al. IJCV '00]
- –Primal sketch prior [Sun et. Al. CVPR '03]
- –Neighbor embedding [Chang et. Al. CVPR '04]
- –Soft edge prior [Dai et. Al. ICCV '07]



LEARNING-BASED SUPER-RESOLUTION – Prior work

How should we regularize the super-resolution problem?

- –Markov random field [Freeman et. Al. IJCV ‘00]
- –Primal sketch prior [Sun et. Al. CVPR ‘03]
- –Neighbor embedding [Chang et. Al. CVPR ‘04]
- –Soft edge prior [Dai et. Al. ICCV ‘07]



Our approach:

High-resolution patches have a sparse linear representation with respect to an overcomplete dictionary of patches randomly sampled from similar images.

$x \in \mathbb{R}^D$ output high-resolution patch $\qquad D_h \in \mathbb{R}^{D\times n}$ high-resolution dictionary

$$\boxed{x \approx D_h \alpha_0 \quad \text{for some } \alpha_0 \in \mathbb{R}^n \quad \text{with } \|\alpha_0\|_0 \ll n}$$

LINEAR SPARSE REPRESENTATION – SR as Compressed Sensing

We do not directly observe the high resolution patch x, but rather (features of) its low-resolution version:

$D_\ell \doteq L D_h \in \mathbb{R}^{d \times n}$ dictionary of low-resolution patches.

downsampling / blurring operator

The input low-resolution patch $y \in \mathbb{R}^d$ satisfies

$$
\begin{aligned}
y &= Lx \\
&\approx L D_h \alpha_0 = D_\ell \alpha_0
\end{aligned}
$$

d linear measurements of sparse coefficient vector α_0 !

LINEAR SPARSE REPRESENTATION – SR as Compressed Sensing

If we can recover the sparse solution α_0 to the underdetermined system of linear equations $y = D_\ell \alpha$, we can reconstruct x as $D_h \alpha_0$

Formally, we seek the sparsest solution:

$$\hat{\alpha}_0 = \arg\min \|\alpha\|_0 \quad \text{subj} \quad y = D_\ell \alpha$$

convex relaxation

$$\hat{\alpha}_1 = \arg\min \|\alpha\|_1 \quad \text{subj} \quad y = D_\ell \alpha$$

This problem can be efficiently solved by linear programming. In many circumstances it recovers the sparsest solution α_0 [Donoho 2006 CPAM].

ALGORITHM DETAILS – Enforcing patch consistency

Combining local (patch) estimates:

Sample 3 x 3 low resolution patches y on a regular grid.

Allow 1 pixel overlap between adjacent patches.

Enforce agreement between overlapping high-resolution reconstructions.

> Simultaneous solution for $\{\alpha\}$ for all patches: large, but sparse convex program. Still too slow in practice.

Fast approximation: compute α for each patch in raster scan order, enforce consistency with previously computed patch solutions:

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \qquad \text{subj} \qquad \|Fy - FD_\ell \alpha\|_2^2 \le \varepsilon_1^2$$

$$\sum_{\alpha'} \|T'(FD_h\alpha') - T(FD_h\alpha)\|_2^2 \le \varepsilon_2^2.$$

T, T': select overlap between patches

F : linear feature extraction operator

ALGORITHM DETAILS – Feature extraction

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \qquad \text{subj} \qquad \|Fy - FD_\ell\alpha\|_2^2 \leq \varepsilon_1^2$$

$$\sum_{\alpha'} \|T'(FD_\hbar\alpha') - T(FD_\hbar\alpha)\|_2^2 \leq \varepsilon_2^2.$$

F : linear feature extraction operator

Here, F concatenates first and second image partial derivatives, computed from a bicubic interpolation of the low-resolution input.

Emphasizes the part of the signal that is most relevant for human perception and for predicting the high-resolution output.

Transforms usual ℓ^2 fidelity criterion into a more perceptually meaningful Mahalanobis distance.

Complete feature vector for each low-resolution patch is 384 dimensional.

SUPERRESOLUTION VIA SPARSITY – Algorithm pseudocode

Input: training dictionaries D_h and D_ℓ, low-resolution image Y.

for each 3×3 patch y of Y, taken in raster scan order with 1 pixel overlap,

- $\hat{\alpha} = \arg\min \lambda \|\alpha\|_1 + \frac{1}{2}\|\tilde{D}\alpha - \tilde{y}\|_2^2$
- Place high resolution patch $x = D_h\hat{\alpha}$ in X_0.

end

Using back-projection, find the closest image to X_0 satisfying the reconstruction constraint:

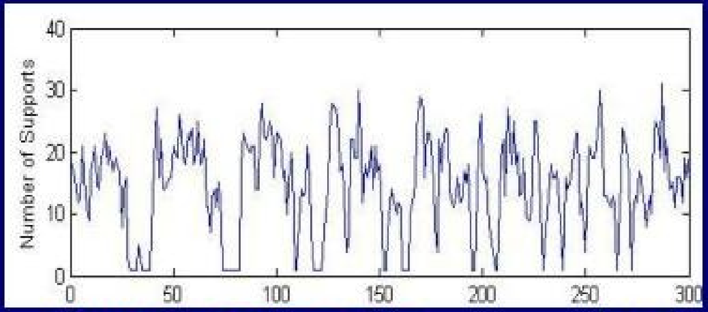
$$X^* = \arg\min_X \|X - X_0\| \quad \text{s.t.} \quad LX = Y.$$

Output: super-resolution image X^*.

RELATIONSHIP TO PREVIOUS WORK – Adaptivity, simplicity

Adaptivity of representation

ℓ^1-minimization automatically selects the smallest number of training samples that can represent the input.



Number of nonzero coefficients, $\|\hat{\alpha}\|_0$

Rectifies overfitting and underfitting issues inherent in fixed-neighbor methods (e.g., Neighbor Embedding [Chang CVPR '04]).

Simplicity of dictionary

Sparsity in fixed bases (wavelet, curvelet), or learned bases (K-SVD, alternating minimization) has been applied extensively to image compression, denoising, inpainting, and more recently to classification and categorization.

For superresolution, sparse representation in simple bases of randomly sampled patches already performs competitively.

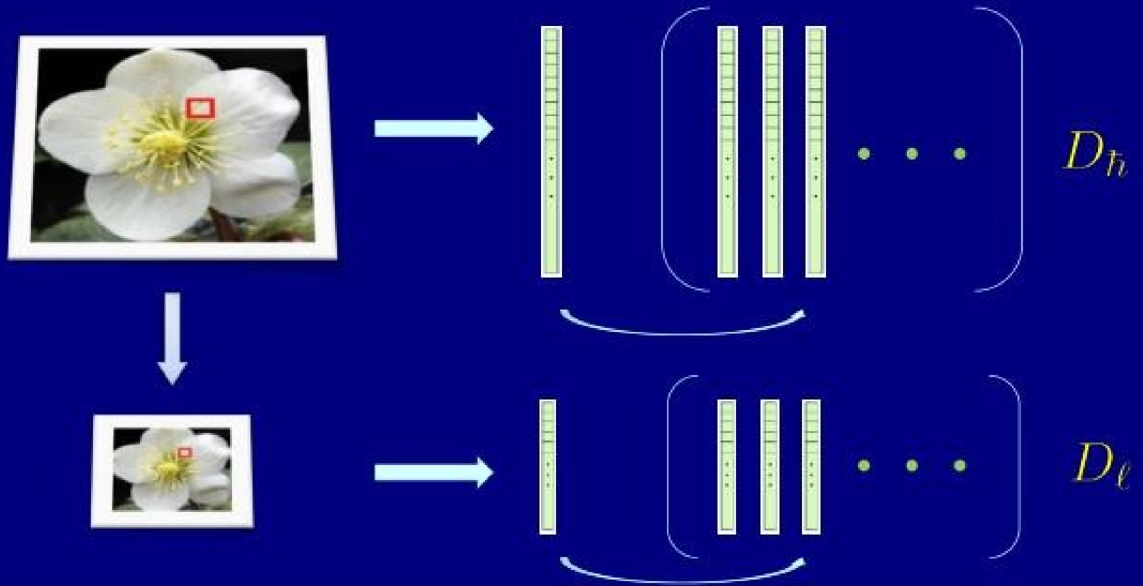
EXPERIMENTAL SETUP: Dictionary preparation

Two training sets:

Flower images -- smooth textures, sharp ed

Animal images -- high-frequency textures

Randomly sample 100,000 high-resolution / low-resolution patch pairs from each set of training images: $D_\ell \in \mathbb{R}^{384 \times 100,000}$.



QUALITATIVE COMPARISON: Flower, zoom by 3x (flower dictionary)

Low-resolution input:

Bicubic

Neighbor embedding
[Chang CVPR '04]

Our method

Original

QUALITATIVE COMPARISON: Girl, zoom by 3x (flower dictionary)

Low-resolution input

Bicubic

Neighbor embedding
[Chang CVPR ‘04]

Our method

Original

QUALITATIVE COMPARISON: Parthenon, zoom by 3x (flower dictionar

Input Image

Bicubic

Neighbor embedding

Our method

QUALITATIVE COMPARISON: Raccoon, zoom by 3x (animal dictionary)

Low-resolution input:

Bicubic

Neighbor embedding
[Chang CVPR '04]

Our method

FURTHER EXAMPLES: zoom by 3x

Input:

Output:

QUALITATIVE COMPARISON: girl, zoom by 4x (flower dictionary)

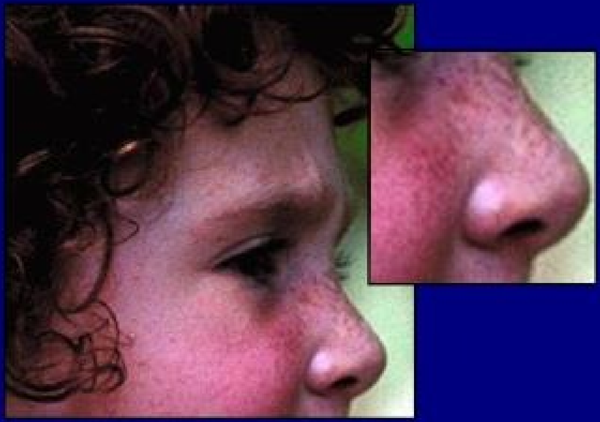
Input, upsampled

Bicubic

MRF / BP
[Freeman IJCV '00]

Soft edge prior
[Dai ICCV '07]

Our method



Original

QUANTITATIVE COMPARISON: RMS error

Image		Bicubic	Neighborhood embedding	Our method
Flower		3.51	4.20	***3.23***
Girl		5.90	6.66	***5.61***
Parthenon		12.74	13.56	***12.25***
Raccoon		9.74	9.85	***9.19***

Our approach outperforms bicubic interpolation and neighbor embedding on all examples tested.

OTHER APPLICATIONS: Face Hallucination

Fig. 2. Results of our algorithm compared to other methods. From left to right columns: low resolution input; bicubic interpolation; back projection; sparse coding via NMF followed by bilater filtering; sparse coding via NMF and Sparse Representation; Original.

CONCLUSIONS

- Assumption:
  - High-resolution image patches have sparse representation in a dictionary of patches randomly sampled from similar images.

- Super-resolution as sparse representation:
  - Observe a small number of linear measurements (in this case the low-resolution image)
  - Recover sparse representation via ℓ^1 minimization
  - Framework can incorporate overlap constraints, ect.

- Implications:
  - Surprisingly good results (competitive with state-of-the-art) with a very simple algorithm
  - Randomly sampled patches provide an effective dictionary

FUTURE PROBLEMS

- Connections to compressed sensing: minimum patch size or feature space dimension to recover sparse representation?

- How much data: How many training samples are required to sparsify natural image categories? How restricted does the category have to be for the sparse representation to be recoverable by – minimization?

- Combining dictionaries from multiple classes: simultaneous supervised image segmentation and super-resolution.

REFERENCES & ACKNOWLEDGMENT

- References:
  - *"Image super-resolution as sparse representation of raw image patches,"* CVPR 2008

- People:
  - **Prof. Thomas Huang**, ECE, University of Illinois
  - **Prof. Yi Ma**, ECE, University of Illinois
  - **Jianchao Yang**, PhD student, ECE, University of Illinois
- Funding:
  - NSF EHS-0509151
  - NSF CCF-0514955
  - ONR YIP N00014-05-1-0633
  - NSF IIS 0703756

THANK YOU

Questions, please?