

Image Super-Resolution via Sparse Representation

Stanley Chan

Purdue University

ECE/STAT 695 Spring 2018

Image Super-resolution



Fig. 1. Reconstruction of a raccoon face with magnification factor 2. Left: result by our method. Right: the original image. There is little noticeable difference visually even for such a complicated texture. The RMSE for the reconstructed image is 5.92 (only the local patch model is employed).

- Low resolution input
- Want high resolution output

Basic Setup

Model:

$$\mathbf{y} = \mathbf{S}\mathbf{H}\mathbf{D}\boldsymbol{\alpha} \quad (1)$$

- $\mathbf{y} \in \mathbb{R}^{m \times 1}$: low resolution image
- $\mathbf{S} \in \mathbb{R}^{m \times n}$: Sampling matrix
- $\mathbf{H} \in \mathbb{R}^{n \times n}$: Blurring matrix
- $\mathbf{D} \in \mathbb{R}^{n \times p}$: Dictionary
- $\boldsymbol{\alpha} \in \mathbb{R}^{p \times 1}$: sparse coefficient vector

Questions:

- How to learn dictionary?
- Can we incorporate information from low resolution VS high resolution?
- Can we incorporate feature extraction methods?

Some Accessories 1

Image to Patch:

- Partition \mathbf{y} into patches $\mathbf{y}_1, \dots, \mathbf{y}_N$
- Patch extraction operator \mathbf{R}_i :

$$\mathbf{y}_i = \mathbf{R}_i \mathbf{y}, \quad i = 1, \dots, N$$

- \mathbf{R}_i consists of rows of the identity matrix
- Overlap is allowed
- To extract overlap between \mathbf{y}_i and \mathbf{y}_{i-1} :

$$\mathbf{w}_i = \mathbf{P}_i \mathbf{y}, \quad i = 1, \dots, N$$

Some Accessories 2

Feature Extraction:

- Extract gradient features of \mathbf{y}_i
- Gradient filter:

$$\begin{aligned}\mathbf{f}_1 &= [-1, 0, 1], & \mathbf{f}_2 &= [-1, 0, 1]^T \\ \mathbf{f}_3 &= [1, 0, -2, 0, 1], & \mathbf{f}_4 &= [1, 0, -2, 0, 1]^T.\end{aligned}$$

- Define convolution matrices $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$.
- Concatenate to make $\mathbf{F} = [\mathbf{F}_1; \mathbf{F}_2; \mathbf{F}_3; \mathbf{F}_4]$.
- Extracted features:

$$\mathbf{F}\mathbf{y}_i = \mathbf{F}\mathbf{D}\alpha_i$$

- \mathbf{f}_j ($j = 1, \dots, 4$) is a gradient operator. Mean is removed.

Local Representation

- Construct two dictionaries: $\mathbf{D}_h, \mathbf{D}_l$.
- For \mathbf{D}_l , do

$$\underset{\alpha_i}{\text{minimize}} \quad \|\alpha_i\|_0 \quad \text{subject to} \quad \|\mathbf{F}\mathbf{D}_l\alpha_i - \mathbf{F}\mathbf{y}_i\|_2^2 \leq \epsilon. \quad (2)$$

- \mathbf{F} is a feature selection matrix.

This minimization says:

- Want α_i to be sparse.
- Since \mathbf{y}_i is low resolution, let us use \mathbf{D}_l .
- We also want to use the feature selection method. So we use \mathbf{F} .
- Since $\|\cdot\|_0$ is hard, we can switch to $\|\cdot\|_1$:

$$\underset{\alpha_i}{\text{minimize}} \quad \|\mathbf{F}\mathbf{D}_l\alpha_i - \mathbf{F}\mathbf{y}_i\|_2^2 + \lambda\|\alpha_i\|_1 \quad (3)$$

High Resolution

For $i = 1, \dots, N$, do

$$\begin{aligned} \hat{\alpha}_i = & \arg \min_{\alpha} \|\alpha\|_1 \\ & \text{subject to} \quad \|\mathbf{F}\mathbf{D}_l\alpha - \mathbf{F}\mathbf{y}_i\|_2^2 \leq \epsilon_1 \\ & \|\mathbf{P}_i\mathbf{D}_h\alpha - \mathbf{P}_i\mathbf{D}_h\hat{\alpha}_{i-1}\|_2^2 \leq \epsilon_2. \end{aligned} \quad (4)$$

- $\mathbf{D}_h\alpha_i$ is the high-resolution patch
- $\|\mathbf{P}_i\mathbf{D}_h\alpha - \mathbf{P}_i\mathbf{D}_h\hat{\alpha}_{i-1}\|_2^2$ ensures high resolution overlap region is consistent
- Raster scan image from left/right, top/down
- Final output: $\mathbf{x}_i = \mathbf{D}_h\hat{\alpha}_i + \mu_i$
- μ_i is the mean of \mathbf{y}_i

Enforce Global Consistency

First do the above sparse representation reconstruction

$$\hat{\mathbf{x}}^0 = \text{Sparse Representation}(\mathbf{y})$$

Then do

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{S}\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + \rho \|\mathbf{x} - \hat{\mathbf{x}}^0\|^2. \quad (5)$$

- Ensures consistency with physical model
- Original paper: Use gradient descent
- Can be solved in closed-form using polyphase decomposition [Chan et al. 2017]

Face Hallucination

Can do better than natural image:

- Face is special
- Can be pre-trained to yield good initial guesses

Let $\hat{\mathbf{x}}^0$ be initial guess (roughly super resolved image).

For patches $i = 1, \dots, N$, do

$$\begin{aligned} \hat{\alpha}_i = & \arg \min_{\alpha} \|\alpha\|_1 \\ & \text{subject to} \quad \|\mathbf{F}\mathbf{D}_l\alpha - \mathbf{F}\hat{\mathbf{x}}_i^0\|_2^2 \leq \epsilon_1 \\ & \|\mathbf{P}_i\mathbf{D}_h\alpha - \mathbf{P}_i\mathbf{D}_h\hat{\alpha}_{i-1}\|_2^2 \leq \epsilon_2. \end{aligned} \quad (6)$$

- Final output: $\mathbf{x}_i = \mathbf{D}_h\hat{\alpha}_i + \mu_i$
- μ_i is the mean of $\hat{\mathbf{x}}_i^0$

Non-negative Matrix Factorization

Let \mathbf{X} be a dataset; Each column is a sample patch.

$$(\mathbf{U}, \mathbf{V}) = \arg \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{UV}\|_F^2, \quad \text{s.t. } \mathbf{U} \geq 0, \mathbf{V} \geq 0. \quad (7)$$

Two standard approaches:

- Alternating minimization
- Multiplicative update

How to use NMF?

$$\mathbf{c}^* = \arg \min_{\mathbf{c} \geq 0} \|\mathbf{SHUc} - \mathbf{Y}\|^2 + \eta \|\mathbf{\Gamma Uc}\|_2^2 \quad (8)$$

- $\mathbf{\Gamma}$ = “high-pass” matrix
- \mathbf{c} = coefficient vector

Dictionary Learning

- Given training dataset $(\mathbf{x}_i, \mathbf{y}_i)$ for $i = 1, \dots, N$.
- $\mathbf{x}_i \in \mathbb{R}^q$, $\mathbf{y}_i \in \mathbb{R}^p$.
- Solve

$$\hat{\mathbf{D}}_h = \arg \min_{\mathbf{D}_h, \{\alpha_i\}} \sum_{i=1}^N \{ \|\mathbf{x}_i - \mathbf{D}_h \alpha_i\|^2 + \lambda \|\alpha_i\|_1 \}$$

$$\hat{\mathbf{D}}_l = \arg \min_{\mathbf{D}_l, \{\alpha_i\}} \sum_{i=1}^N \{ \|\mathbf{y}_i - \mathbf{D}_l \alpha_i\|^2 + \lambda \|\alpha_i\|_1 \}$$

- How to ensure common sparsity?

$$\underset{\mathbf{D}_h, \mathbf{D}_l, \mathbf{A}}{\text{minimize}} \quad \frac{1}{p} \|\mathbf{Y} - \mathbf{D}_l \mathbf{A}\|^2 + \frac{1}{q} \|\mathbf{X} - \mathbf{D}_h \mathbf{A}\|^2 + \lambda \left(\frac{1}{p} + \frac{1}{q} \right) \|\mathbf{A}\|_1.$$

Results



Fig. 5. Results on an image of the Parthenon with magnification factor 3 and corresponding RMSEs. Top row: low-resolution input, Bicubic interpolation (RMSE: 12.724), BP (RMSE: 12.131). Bottom row: NE (RMSE: 13.556), SE [7] (RMSE: 12.228), and our method (RMSE: **11.817**).

Results



Fig. 3. Results of the flower image magnified by a factor of 3 and the corresponding RMSEs. Left to right: input, Bicubic interpolation (RMSE: 4.066), NE [11] (RMSE: 4.888), our method (RMSE: **3.761**), and the original.



Fig. 4. Results of the girl image magnified by a factor of 3 and the corresponding RMSEs. Left to right: input, Bicubic interpolation (RMSE: 6.843), NE [11] (RMSE: 7.740), our method (RMSE: **6.525**), and the original.

Results

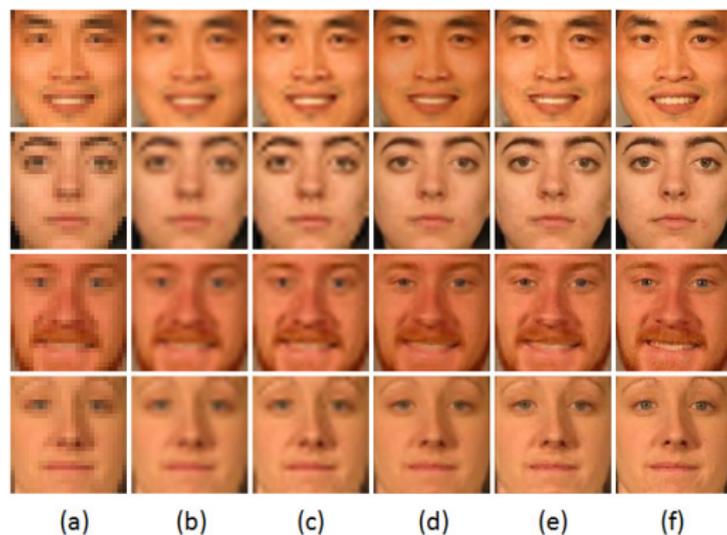


Fig. 8. Results of our algorithm compared to other methods and the corresponding average RMSEs. From left to right columns: (a) low resolution input; (b) Bicubic interpolation (RMSE: 8.024); (c) back projection (RMSE: 7.474); (d) global NMF modeling followed by bilateral filtering (RMSE: 10.738); (e) global NMF modeling and Sparse Representation (RMSE: **6.891**); (f) Original.

Reference

- J. Yang, J. Wright, Y. Huang and Y. Ma, Image Super-resolution via Sparse Representation, IEEE Trans. Image Process. pp.2861-2873, vol. 19, no. 11, Nov. 2010.