Image Super-Resolution via Sparse Representation

Jianchao Yang, John Wright, Thomas Huang and Yi Ma

accepted by IEEE Trans. on Image Processing 2010

Presented by known 2010/4/20
Super-Resolution

Techniques that enhance the resolution of an imaging system

Low-cost imaging sensor  High-definition display
Previous Approaches

- Register multiple low-resolution images
  - insufficient # of input images
  - ill-conditioned registration
  - unknown blurring operators

- Interpolation based
  - Simple interpolation → over-smooth
  - Using prior → limited in modeling textures
Previous Approaches (2)

Machine learning based

- Learn the co-occurrence of low and high resolution image patches

- 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]

Need enormous patch pairs
Fix k neighbors

Input
Trained patches
≈
Output
Original
Key Idea

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

* Patches are overlapped
Key Idea

- **Input LR image**
  - Cut into patches

- **For each patches**
  - \( \alpha \) is sparse

**Training**

- **Input LR image**
  - Cut into patches

- **Patch pairs**

- **LR dictionary**

- **HR dictionary**

**Testing**

- **Patch pairs**

- **Input LR image**
  - Cut into patches

- **Output HR image**
  - \( \alpha \) is sparse
Notation

- $X$ the high resolution image
- $Y$ the low resolution image
- $X$ the high resolution image patch
- $Y$ the low resolution image patch
- $D^n$ dictionaries for high resolution
- $D^\ell$ dictionaries for low resolution
- bold lower case letter: vector $\mathbf{v}$
- unbold upper case letter: matrix $\mathbf{M}$
- unbold lower case letter: scalar $s$
Problem and Constraint

Problem Formulation
- Given
  - low resolution image $Y$
  - Trained dictionaries $D_h$, $D_\ell$
- To recover high resolution image $X$

Assumption & Constraints
- Reconstruction constraint
  \[ Y = S H X \]
- Sparse prior
  \[ x \approx D_h \alpha \quad , \quad y \approx D_\ell \alpha \]
  for some $\alpha \in \mathbb{R}^K$ with $||\alpha||_0 \ll K$
Why Sparse?

- Natural images may generally be describe in terms of a small number of structure primitives (edges, lines) [Field, 1994]
- Filter images with a set of log-Gabor filters and collecting histograms of the result $\rightarrow$ high kurtosis [Field 1993]

Proposed Approaches

- **Generic image**
  - local sparse prior (recover local detail)
  - → global constraint (remove artifact)

- **Face image (Face Hallucination)**
  - global constraint from face prior
  - → local sparse prior
Local Sparse Prior

- find a sparse representation $\alpha$ correspond to $D_\ell$ for each low-resolution image patch $y$

$$\min ||\alpha||_0 \text{ s.t. } ||FD_\ell \alpha - Fy||_2^2 \leq \epsilon$$

- $F$ : feature extraction operator

- recover $\alpha$ by $\ell^1$-norm (assume the coefficients are sufficiently sparse)

[Donoho 2006]

$$\min ||\alpha||_1 \text{ s.t. } ||FD_\ell \alpha - Fy||_2^2 \leq \epsilon$$

- Reformulate by Lagrange Multiplier

$$\min ||FD_\ell \alpha - Fy||_2^2 + \lambda ||\alpha||_1$$
Sparse Prior (neighbor)

Consider neighbor patches
\[ \min \| \alpha \|_1 \text{ s.t. } \| FD_\ell \alpha - Fy \|_2^2 \leq \varepsilon_1 \]
\[ \| PD_\bar{h} \alpha - w \|_2^2 \leq \varepsilon_2 \]

- \( P \): extract overlap region
- \( w \): previous reconstruct on the overlap

Reformulate:
\[ \alpha^* = \arg \min_{\alpha} \| \tilde{D} \alpha - \tilde{y} \|_2^2 + \lambda \| \alpha \|_1, \]
\[ \tilde{D} = \begin{bmatrix} FD_\ell \\ \beta PD_\bar{h} \end{bmatrix}, \quad \tilde{y} = \begin{bmatrix} Fy \\ \beta w \end{bmatrix}, \beta = 1 \]

\[ \Rightarrow x = D_\bar{h} \alpha^* \Rightarrow X_0 \]
Global Constraint

to ensure the reconstructed image $X_0$ satisfy global constraint

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

Solve by back-projection method

$$X_{t+1} = X_t + ((Y - SHX_t) \uparrow s) * p$$

$s$: up-sample scale factor

$p$: “back-projection” filter (類似deblur)

Gradient descent 2 version (?!)

$$X^* = \arg \min_X \|SHX - Y\|^2 + c\|X - X_0\|^2$$

$$X_{t+1} = X_t + \nu[H^T S^T (Y - SHX_t) + c(X_t - X_0)]^{13}$$

$H$: blurring filter

$S$: down-sampling filter
Generic image SR(algo 1)

1. **local** sparse prior (recover local detail)

\[
\alpha^* = \arg \min_{\alpha} \| \tilde{D} \alpha - \tilde{y} \|_2^2 + \eta \| \alpha \|_1,
\]

\[
\tilde{D} = \begin{bmatrix}
FD
\beta PD_h
\end{bmatrix}, \quad \tilde{y} = \begin{bmatrix}
Fy
\beta w
\end{bmatrix}, \quad \beta = 1
\]

\[\rightarrow x = D_h \alpha^* \Rightarrow X_0\]

2. **global** constraint (remove artifact)

\[
X^* = \arg \min_X \| X - X_0 \| \quad \text{s.t.} \quad SHX = Y
\]
Global Optimization

Simultaneously solve the coefficients of all the patches

\[ X^* = \arg\min_{X, \{ \alpha_{ij} \}} \left\{ \| SHX - Y \|_2^2 + \lambda \sum_{i,j} \| \alpha_{ij} \|_0 \right\} + \gamma \left[ \sum_{i,j} \| D_h \alpha_{ij} - P_{ij} X \|_2^2 + \tau \rho(X) \right] \]

\[ \min \| \alpha \|_1 \text{ s.t. } \| FD_\ell \alpha - Fy \|_2^2 \leq \varepsilon \]

\[ X^* = \arg\min_X \| X - X_0 \| \text{ s.t. } SHX = Y \]
Proposed Approaches

- **Generic image**
  - local sparse prior (recover local detail)
  - \(\rightarrow\) global constraint (remove artifact)

- **Face image (Face Hallucination)**
  - global constraint from face prior
  - \(\rightarrow\) local sparse prior
Face Prior

Using Non-negative matrix factorization (NMF)

\[
\arg \min_{U,V} \| A - UV \|_F^2 \quad \text{s.t.} \quad U \geq 0, \quad V \geq 0
\]

- \( A(n \times m) \): training face images (column)
- \( U(n \times r) \): trained face basis
- \( V(r \times m) \): coef correspond to each train face
Using Face Prior

- to find $p(X|Y)$
  - find Maximum a posteriori (MAP)
    $$X^* = \arg \max_X p(Y|X)p(X)$$
  - find coef $c^*$ that satisfy:
    $$c^* = \arg \min_c \|SHUc - Y\|_2^2 + \eta \rho(Uc) \quad \text{s.t.} \quad c \geq 0$$

  $S$: down-sampling filter
  $H$: blurring filter
  $U$: trained face basis

  medium resolution image $\hat{X} = UC^{18}$
Face image SR (algo 2)

1. **global** constraint from face prior

\[ c^* = \arg \min_c \| SHUc - Y \|^2_2 + \eta \| \Gamma U c \|_2 \quad \text{s.t.} \quad c \geq 0. \]

\[ \rightarrow \hat{X} = Uc^* \]

2. **local** sparse prior

for each patch \( y \) from \( \hat{X} \):

\[ \alpha^* = \arg \min_{\alpha} \| \tilde{D}\alpha - \tilde{y} \|^2_2 + \eta \| \alpha \|_1, \]

\[ \tilde{D} = \begin{bmatrix} FD_{\ell} \\ \beta PD_{\hat{h}} \end{bmatrix}, \quad \tilde{y} = \begin{bmatrix} Fy \\ \beta w \end{bmatrix}, \quad \beta = 1 \]

\[ \rightarrow x = D_{\hat{h}}\alpha^* \quad \Rightarrow \quad X^* \]
Learning Dictionary Pairs

Goal

to learn compact dictionary pairs

Single Dictionary Training

\[ X = \{x_1, x_2, \ldots, x_t\} \quad \text{t 個 patches} \]

\[ D = \arg \min_{D,Z} \|X - DZ\|_2^2 + \lambda \|Z\|_1 \]

s.t. \[ \|D_i\|_2 \leq 1, i = 1, 2, \ldots, K \]

Joint Dictionary Training

\[ D^h = \arg \min_{\{D^h, Z\}} \|X^h - D^hZ\|_2^2 + \lambda \|Z\|_1 \]

\[ D^l = \arg \min_{\{D^l, Z\}} \|Y^l - D^lZ\|_2^2 + \lambda \|Z\|_1 \]
Joint Dictionary Training

\[ D_h = \arg \min_{\{D^h, Z\}} \|X^h - D^h Z\|_2^2 + \lambda \|Z\|_1 \]
\[ D_l = \arg \min_{\{D^l, Z\}} \|Y^l - D^l Z\|_2^2 + \lambda \|Z\|_1 \]

\[ [D_h, D_l] = \arg \min_{\{D^h, D^l, Z\}} \left( \frac{1}{N} \|X^h - D^h Z\|_2^2 + \frac{1}{M} \|Y^l - D^l Z\|_2^2 + \lambda \left( \frac{1}{N} + \frac{1}{M} \right) \|Z\|_1 \right) \]

\[ [D_h, D_l] = \arg \min_{\{D^h, D^l, Z\}} \left( \|X_c - D_c Z\|_2^2 + \hat{\lambda} \|Z\|_1 \right) \]

\[ X_c = \begin{bmatrix} \frac{1}{\sqrt{N}} X^h \\ \frac{1}{\sqrt{M}} Y^l \end{bmatrix}, \quad D_c = \begin{bmatrix} \frac{1}{\sqrt{N}} D^h \\ \frac{1}{\sqrt{M}} D^l \end{bmatrix}, \quad \hat{\lambda} = \lambda \left( \frac{1}{N} + \frac{1}{M} \right) \]

\( N: \# \text{ dimension of high res patches} \]
\( M: \# \text{ dimension of low res patches} \]

21
Dictionary Training

To Solve: \[ D = \arg\min_{D,Z} \|X - DZ\|_2^2 + \lambda\|Z\|_1 \]

s.t. \[ \|D_i\|_2^2 \leq 1, i = 1, 2, \ldots, K \]

1. Initialize \( D \) with Gaussian random matrix
2. Fix \( D \), update \( Z \)
   \[ Z = \arg\min_{Z} \|X - DZ\|_2^2 + \lambda\|Z\|_1 \]
3. Fix \( Z \), update \( D \)
   \[ D = \arg\min_{D} \|X - DZ\|_2^2 \]
   s.t. \[ \|D_i\|_2^2 \leq 1, i = 1, 2, \ldots, K \]
4. Iterate 2,3 until converge
Features for low res patches

\[
\min ||\alpha||_1 \text{ s.t. } ||FD_\ell \alpha - Fy||_2^2 \leq \epsilon_1 \\
||PD_h \alpha - w||_2^2 \leq \epsilon_2
\]

 Goal

- to ensure the computed coefficients fit the most relevant part of the low-resolution images

 Idea

- using some kind of high pass filter

 Method

- apply 1\textsuperscript{st} and 2\textsuperscript{nd} order derivatives on up-sampled(2 倍) whole low-resolution image
- separate into patches
Experiment

- Single image super-resolution
  - Generic super-resolution
  - Face super-resolution
- Effect of Dictionary Size
- Robustness to Noise
Experiment setup

➤ Generic image
  • Low resolution patch
    • 3x3 (up to 6x6) overlap 1 pixel
  • High resolution patch
    • 9x9 overlap 3 pixel

➤ Face image
  • 5x5 for both low and high resolution

➤ Dictionary
  • Train from 100,000 patches
  • Size 1024
<table>
<thead>
<tr>
<th>Input</th>
<th>bicubic</th>
<th>neighbor-embedding</th>
<th>proposed</th>
<th>origin</th>
<th>bicubic</th>
<th>back-projection</th>
<th>neighbor-embedding</th>
<th>soft edge prior</th>
<th>proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Face Super-resolution

- Input
- two-step generic
- bicubic back-proj
- proposed origin
- NMF+biilter filter
Effect of dictionary size

<table>
<thead>
<tr>
<th>Images</th>
<th>Bicubic</th>
<th>D256</th>
<th>D512</th>
<th>D1024</th>
<th>D2048</th>
<th>Raw Patches</th>
</tr>
</thead>
</table>

2 version (?!)

RMS error
Robustness to noise

The equation of sparse prior can be formulated as a probabilistic model

\[ \alpha^* = \arg \min \lambda \| \alpha \|_1 + \frac{1}{2} \| \tilde{D} \alpha - \tilde{y} \|_2^2 \]

\[ = \arg \min \frac{\lambda}{\sigma^2} \| \alpha \|_1 + \frac{1}{2\sigma^2} \| \tilde{D} \alpha - \tilde{y} \|_2^2 \]

\[ = \arg \max P(\alpha) \cdot P(\tilde{y} | \alpha, \tilde{D}) \]

\[ P(\alpha) = \frac{1}{2b} \exp(-\frac{\| \alpha \|_1}{b}), \quad b = \frac{\sigma^2}{\lambda} \]

\[ P(\tilde{y} | \alpha, \tilde{D}) = \frac{1}{2\sigma^2} \exp(-\frac{1}{2\sigma^2} \| \tilde{D} \alpha - \tilde{y} \|_2^2) \]

Laplacian

Gaussian

2 version

(?!)

<table>
<thead>
<tr>
<th>Noise Levels / Gaussian $\sigma$</th>
<th>0</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicubic</td>
<td>9.873</td>
<td>10.423</td>
<td>11.037</td>
<td>11.772</td>
</tr>
<tr>
<td>Our method</td>
<td><strong>8.359</strong></td>
<td><strong>9.240</strong></td>
<td><strong>10.454</strong></td>
<td><strong>11.448</strong></td>
</tr>
</tbody>
</table>
Conclusion

A novel approach toward single image super-resolution based on sparse representations are presented

- jointly trained coupled dictionaries from high- and low-resolution image patch pairs.
- enforced the compatibilities among adjacent patches both locally and globally

Experimental results demonstrate the effectiveness of the sparsity as a prior for patch-based super-resolution both for generic and face images.
Future Works

➤ Determine the **optimal dictionary size** for natural image patches

➤ Tighter connections to the **theory of compressed sensing**
  - conditions on the appropriate patch size
  - conditions on the appropriate features
  - approaches for training the coupled dictionaries
THANK YOU ~😊~
Lagrange Multiplier

平常要求 $f(x,y)$ 的極值

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$ 求解

若 $x,y$ 的值被 $g(x,y)=0$ 所限制

改求 $f(x,y) + \lambda g(x,y)$ 的極值

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} = 0$$ 求解

$$\frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} = 0$$

$$g(x,y) = 0$$
Solution to $\ell^0$, $\ell^1$, $\ell^2$ sparse

\[ x_1 = \arg \min_x ||x||_1 \text{ s.t. } \tilde{y} = \tilde{A}x \]
\[ \tilde{y} = \tilde{A}x \]

1-$\ell^0$ ball

\[ x_1 = \arg \min_x ||x||_1 \text{ s.t. } ||\tilde{y} - \tilde{A}x|| \leq \epsilon \]
\[ \tilde{y} = \tilde{A}x \]

1-$\ell^0$ ball

Equivalence breakdown point (EBP)

\[ x_2 = \arg \min_x ||x||_2 \text{ s.t. } \tilde{y} = \tilde{A}x \]
\[ \tilde{y} = \tilde{A}x \]

1-$\ell^1$ ball

1-$\ell^0$ ball

1-$\ell^2$ ball
How to estimate EBP?

**Theorem:**
如上圖,C是一個1-\(\ell^1\) ball, 若P(=AC)是k-neighborly, 且y=Ax0中x0是k-sparse, 則以\(\ell^1\) -minimization求出來的解即為x0, 且是唯一解

**Definition:**
- If P is k-neighborly \(\rightarrow\) C上(k-1)D表面都可以對到P上

**Properties**
- If x0 is k-sparse \(\rightarrow\) 與1-\(\ell^1\) ball的(k-1)D 表面有交點
- 這些表面都是simplex (n-D simplex, n+1 vertices)

放大\(\ell^1\) ball C, 則P也會跟著放大, 當P的表面碰到y的時候的x, 就是x1也就是x0