Image Super-Resolution Via Sparse Representation

Soumendu Majee
Omar Elgendy

Super Resolution

- Image super resolution is a special case of image interpolation.
- Given a low resolution image, estimate its high resolution version.
- It’s an ill-posed problem:
  - Have to incorporate prior information about the image
    - Locally smooth assumption
    - Trained dictionary
Signal Model

- Observed low resolution image: $Y$
- Unknown high resolution image: $X$
- $Y$ can be written as the blurred and down-sampled version of $X$.

\[ Y = SHX. \]

Where, $S$ is a down-sampling operator and $H$ is a blurring filter.
The patches $x$ of the high resolution image $X$ can be represented as sparse linear combination in a dictionary $D_h$:

$$x \approx D_h \alpha \quad \text{for some } \alpha \in \mathbb{R}^K \text{ with } \|\alpha\|_0 \ll K.$$

The sparse representation, $\alpha$ is recovered by representing the patches from the low resolution image $Y$ with respect to a low resolution dictionary, $D_l$ contained within $D_h$. 
The two dictionaries $D_h$, $D_l$ are trained to have the same sparse representation for each high-resolution and low-resolution patch pair.

For every low-resolution patch, find its sparse representation in $D_l$ as:

$$\min ||\alpha||_1 \quad \text{s.t.} \quad ||FD_l\alpha - Fy||_2^2 \leq \epsilon,$$

where, $F$ is a linear feature extraction operator to get perceptually meaningful constraint.
Adjacent patch compatibility

- Enforcing sparse representation of each local patch does not guarantee adjacent patches to be compatible.

- Scan patches in raster order and enforce compatibility by:

  \[
  \min \| \alpha \|_1 \quad \text{s.t.} \quad \| FD_l \alpha - F y \|_2^2 \leq \epsilon_1 \\
  \| PD_h \alpha - w \|_2^2 \leq \epsilon_2
  \]

  where, \( P \) extracts the overlap between current and previous patch and contains the overlapped reconstructed values of previous patch.

- Or more concisely:

  \[
  \min_{\alpha} \| \tilde{D} \alpha - \tilde{y} \|_2^2 + \lambda \| \alpha \|_1
  \]

  where \( \tilde{D} = [ FD_l \quad \beta PD_h ] \) and \( \tilde{y} = [ F y \quad \beta w ] \).
Reconstructing the high resolution patch

- After the optimal sparse representation $\alpha^*$ has been found from the low resolution patch and previous reconstructions, the high resolution patch can be found as:

$$x = D_h \alpha^*.$$
The previous reconstruction might produce images that are inconsistent with the blurred down-sampled signal model. This is rectified by projecting the previous reconstruction, $X_0$ onto the solution space of the signal $Y = SHX$, by computing:

$$X^* = \arg \min_X \|SHX - Y\|_2^2 + c\|X - X_0\|_2^2.$$
Full Super-resolution algorithm

Algorithm 1 (SR via Sparse Representation).

1: **Input:** training dictionaries $D_h$ and $D_l$, a low-resolution image $Y$.
2: **For** each $3 \times 3$ patch $y$ of $Y$, taken starting from the upper-left corner with 1 pixel overlap in each direction,
   - Compute the mean pixel value $m$ of patch $y$.
   - Solve the optimization problem with $\tilde{D}$ and $\tilde{y}$ defined in (8): $\min_{\alpha} \| \tilde{D}\alpha - \tilde{y} \|_2^2 + \lambda \| \alpha \|_1$.
   - Generate the high-resolution patch $x = D_h \alpha^*$. Put the patch $x + m$ into a high-resolution image $X_0$.
3: **End**
4: Using gradient descent, find the closest image to $X_0$ which satisfies the reconstruction constraint
   \[ X^* = \arg \min_X \| SHX - Y \|_2^2 + c \| X - X_0 \|_2^2. \]
5: **Output:** SR image $X^*$.  

Super Resolution of Face Images

- Face image resolution enhancement is useful in surveillance scenarios since large distance between camera and the person makes the resolution of the face part very low.

- Fortunately, human faces have a specific structure that can be exploited in addition to local patch sparsity
Face subspace from NMF

- Face image has several independent parts such as eyes, nose etc
  - Good fit for NMF (Non-negative matrix factorization)

- Given concatenated data matrix, X with each column as a data point, find the best basis matrix U and coefficient matrix V as:

\[
\arg\min_{U,V} \|X - UV\|_2^2 \quad \text{s.t.} \quad U \geq 0, V \geq 0
\]
Increase face resolution by using the face subspace and smooth image assumption:

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

where, \( \Gamma \) is a matrix performing high pass filtering

Estimate the new median resolution image as:

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

Use patch sparsity to increase the resolution further from the medium resolution image
Algorithm 2 (Face Hallucination via Sparse Representation)

1: Input: sparse basis matrix $U$, training dictionaries $D_h$ and $D_l$, a low-resolution aligned face image $Y$.
2: Find a smooth high-resolution face $\hat{X}$ from the subspace spanned by $U$ through:
   - Solve the optimization problem in (16)
   $$\arg\min_{c} \|SHUC - Y\|_2^2 + \eta \|\Gamma U c\|_2 \quad \text{s.t.} \quad c \geq 0.$$
   - $\hat{X} = UC^*$.
3: For each patch $y$ of $\hat{X}$, taken starting from the upper-left corner with 1 pixel overlap in each direction,
   - Compute and record the mean pixel value of $y$ as $m$.
   - Solve the optimization problem with $\tilde{D}$ and $\tilde{y}$ defined in (8):
     $$\min_{\alpha} \|\tilde{D}\alpha - \tilde{y}\|_2^2 + \lambda \|\alpha\|_1.$$
   - Generate the high-resolution patch $x = D_h\alpha^* + m$.
   - Put the patch $x$ into a high-resolution image $X^*$.
4: Output: SR face $X^*$. 
Learning the Dictionary Pair \((D_h, D_l)\)

**Single Dictionary Training**

- **Given:** Set of training examples
  \[ X = \{x_1, x_2, \ldots, x_t\} \]

- **Problem Formulation**
  \[
  D = \arg \min_{D, Z} \|X - DZ\|_2^2 + \lambda \|Z\|_1
  \]
  \[
  \text{s.t. } \|D_i\|_2^2 \leq 1, i = 1, 2, \ldots, K
  \]

  - To enforce sparsity
  - To remove the scaling ambiguity
  - Non-Convex!
  - In both \(D, Z\)
Single Dictionary Training

Solution: Alternative minimization

1) Initial guess: Gaussian random matrix $D$ with each column normalized

2) Fix $D$, update $Z$ by

$$Z = \arg \min_Z \|X - DZ\|_2^2 + \lambda \|Z\|_1$$

3) Fix $Z$, update $D$ by

$$D = \arg \min_D \|X - DZ\|_2^2$$

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

4) Iterate between 2 and 3 until converge
Learning the Dictionary Pair \((D_h, D_l)\)

Joint Dictionary Training

- **Given:**
  - Sampled high resolution image patches \(X^h = \{x_1, x_2, \ldots, x_n\}\)
  - Corresponding low resolution patches \(Y^l = \{y_1, y_2, \ldots, y_n\}\)

- **Goal:** learn \(D_h\) and \(D_l\), so that the sparse representations of \(x_i\) and \(y_i\) are the same \(\forall i = 1, \ldots, n\)

Joint Dictionary Training

- **Problem Formulation**

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

- **Can be rewritten as**

where

\[
\min_{\{D_h, D_l, Z\}} \|X_c - D_c Z\|^2_2 + \hat{\lambda} \|Z\|_1
\]

\[
X_c = \begin{bmatrix} \frac{1}{\sqrt{N}} X^h \\ \frac{1}{\sqrt{M}} Y^l \end{bmatrix} \quad D_c = \begin{bmatrix} \frac{1}{\sqrt{N}} D_h \\ \frac{1}{\sqrt{M}} D_l \end{bmatrix}
\]
Learning the Dictionary Pair \((D_h, D_l)\)

Jianchao Yang, Hao Tang, Thomas Huang, Yi Ma, appeared in TIP’10
Quantitative Comparison:

girl, zoom by 4x (flower dictionary)

Input, upsampled

Bicubic

MRF / BP
[Freeman IJCV ‘00]

Soft edge prior
[Dai ICCV ‘07]

SRSR

Original

## Quantitative Comparison

<table>
<thead>
<tr>
<th>Image</th>
<th>Bicubic</th>
<th>Neighborhood embedding</th>
<th>SRSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flower</td>
<td>3.51</td>
<td>4.20</td>
<td>3.23</td>
</tr>
<tr>
<td>Girl</td>
<td>5.90</td>
<td>6.66</td>
<td>5.61</td>
</tr>
<tr>
<td>Parthenon</td>
<td>12.74</td>
<td>13.56</td>
<td>12.25</td>
</tr>
<tr>
<td>Raccoon</td>
<td>9.74</td>
<td>9.85</td>
<td>9.19</td>
</tr>
</tbody>
</table>

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

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.

Jianchao Yang, Hao Tang, Thomas Huang, Yi Ma, appeared in TIP’10
Thank You
Questions?