Single-Image Super-Resolution: Coupled-dictionary learning vs model-based processing

Pier Luigi Dragotti
Imperial College London

September 4, 2017

Special thanks to J. Huang and X. Deng for helping with the preparation of the slides
Outline

• This talk is about
  
  • Image processing applications that are useful to compare data-driven signal processing vs model-based signal processing
  
  • Shallow learning but not deep-learning
  
  • Overview of key notions used in sparsity-driven signal processing methods
Problem Statement

A visual scene is turned into a **digital** image by a camera.
Can we overcome the limitation of the camera and, given the pixels, obtain a sharper image with increased resolution?
The problem of enhancing the resolution of a single image is known as **Single-Image Super-Resolution**.
Problem Statement

- Single Image Super-Resolution differs from traditional Image Super-Resolution because only one low-resolution image is available
- Highly Ill-posed
- Two main approaches:
  - **Data-driven Approach**: learn the high-resolution image from a database of low-resolution (LR) and high-resolution (HR) pairs
  - **Model-Based Approach**: use priors on the property of natural images to estimate the HR image from the observed LR one.
Data-Driven SISR

• The **key insight** in the data-driven approach is that images or patches of images have a sparse representation in a (redundant) dictionary.

• Example of dictionaries include wavelets, union of bases or *learned* dictionaries.

*Figure*: Cameraman is reconstructed using only 8% of the wavelet coefficients.
Bases and Overcomplete Dictionaries

a) Orthogonal Basis

b) Biorthogonal Basis

c) Frame
Bases and Overcomplete Dictionaries: Matrix Interpretation

- Assume the ‘atoms’ \( \{ \varphi_i \} \) are finite dimensional column vectors of size \( N \)

- Stack them one next to the other to form the **synthesis** matrix \( M \):

\[
\begin{bmatrix}
\varphi_1 & \cdots & \varphi_i & \cdots
\end{bmatrix}
\]

- If \( M \) is square and invertible then \( \{ \varphi_i \}_{i=1}^N \) is a basis (of \( \mathbb{R}^N \) or \( \mathbb{C}^N \)).

- If \( M \) is ‘fat’ but has \( N \) linearly independent columns it form a redundant or overcomplete dictionary.
Data-Driven Super Resolution

- The **key insight** in the data-driven approach is that images or patches of images have a sparse representation in a redundant dictionary.
- The dictionary is usually learned.

\[
x = Dz + n
\]

where
- \(x\) is the original signal
- \(D\) is the dictionary
- \(z\) is the sparse representation of the signal
- \(n\) is the noise
- \(m > n\) indicates that the dictionary is fat

**Introduction**

Sparse representation problem
- Find a \(K\)-sparse signal \(x\) from noisy observation \(y\)
- Dictionary \(D\) is a fat matrix

\[
\|x\|_0 = K
\]

**Sparsity**

- The sparsity \(K\) is the number of non-zero coefficients in the sparse representation

**Dictionary Learning**

- The dictionary \(D\) is learned from a large dataset of images
- The goal is to find a dictionary that allows the representation of images as a sparse linear combination of its atoms
Dictionary Learning

• Learn dictionaries by alternating between
  • Learning the sparse representations given the dictionaries (sparse coding step)
  • Update the dictionary given the sparse representations (dictionary update step)

\[
\begin{bmatrix}
X \\
\vdots
\end{bmatrix} = \begin{bmatrix}
D \\
\vdots
\end{bmatrix} \begin{bmatrix}
Z
\end{bmatrix}
\]
Sparse Coding Step

• Given $D$, learn the sparse representations $z_i$

• Sparse Representation Algorithms:
  
  • Greedy algorithms:
    • Matching Pursuit (MP)
    • Orthogonal Matching Pursuit (OMP)
    • ...

  • Convex Relaxation Algorithms:
    • Basis Pursuit (BP)
    • ....
Dictionary Update Step

- Given the sparse representations update the dictionary.
  - Many possible approaches, k-SVD (Aharon-Elad:06) is the most used.

\[
\begin{bmatrix}
\vdots \\
\end{bmatrix}
\begin{bmatrix}
\vdots \\
\end{bmatrix} =
\begin{bmatrix}
\vdots \\
\end{bmatrix}
\begin{bmatrix}
\vdots \\
\end{bmatrix} 
\]

- Find \(d_i\) and \(z_i^T\) that minimize \(\|E_i - d_i z_i^T\|\).
- This is achieved by taking the SDV of \(E_i\) (with a small caveat to keep \(z_i^T\) sparse).
Data-Driven Super-Resolution

Super-Resolution Model:

• One postulates that HR patches and LR patches admit a common sparse representation $z_i$:

$$x_{i}^{LR} = D^{LR}z_{i}$$

$$x_{i}^{HR} = D^{HR}z_{i}$$

$$x_{i}^{LR} = Ax_{i}^{HR} = AD^{HR}z_{i} = D^{LR}z_{i}$$
Data-Driven Super-Resolution: Algorithm

Algorithm:

• Given $D^{LR}$ and $D^{HR}$ and the LR image to enhance

• Patches $x_i^{LR}$ are extracted

• Using $D^{LR}$ and sparse coding methods (e.g., OMP) the sparse vectors $z_i$ are retrieved

• HR patches are then given by $x_i^{HR} = D^{HR}z_i$

• Open question: How do we learn $D^{LR}$ and $D^{HR}$?
Data-Driven Super-Resolution: Training

Start with an external dataset of images (e.g., BSD 300 dataset)

Extract pairs of LR and HR patches

- Down sampling
- Bicubic interpolation
- Patch extraction

LR patches

HR patches
Data-Driven Super-Resolution: Training

Training:

1. Given $x_i^{LR}$, learn $D^{LR}$ and $z_i$ using K-SVD

2. Given $x_i^{HR}$ and $z_i$ compute $D^{HR}$ directly
Data-Driven Super-Resolution: Example

Algorithm:

• Given $D^{LR}$ and $D^{HR}$ and the LR image to enhance

• Patches $x_i^{LR}$ are extracted

• Using $D^{LR}$ and sparse coding methods the sparse vectors $z_i$ are retrieved

• HR patches are then given by $x_i^{HR} = D^{HR}z_i$

Key references include [Yang et al. 2010, 2012], [Zeyde et al. 2010], [Timofte et al. 2014]
Data-Driven Super-Resolution

• Preliminary Observations:
  • Based on a (simple) modelling assumption: sparsity
  • Most of the complexity is shifted to the training stage

• Given more complex datasets, how much more modelling is required to achieve good performance in the application at hand?
Multimodal Depth Image Super-Resolution

LR Depth Image → Estimated HR Depth Image

HR Color Image → Estimated HR Depth Image
Depth Image Super-Resolution Model

Model:

- HR depth patches and LR depth patches admit a common sparse representation.
- Colour images are made of texture + piecewise smooth 2-D signals.
- The texture is *unique* but the piecewise smooth signal is in *common* with the depth image.
- So HR depth, LR depth and HR colour have something in common but also unique features.
Depth Image Super-Resolution Model

Model:

- HR depth, LR depth and HR colour have something in common but also unique features
- This requires a more sophisticated dictionary model
- We split the dictionary into two parts, one describes the common features and the other the features unique to each modality

\[
\begin{bmatrix}
X^l \\
X^h \\
Y
\end{bmatrix} =
\begin{bmatrix}
\Psi_c^l & \Psi^l & 0 \\
\Psi_c^h & \Psi^h & 0 \\
\Phi_c & 0 & \Phi
\end{bmatrix}
\begin{bmatrix}
Z \\
U \\
V
\end{bmatrix}
\]
Multi-Modal Dictionary Learning

- Learn dictionaries by alternating between
  - Learning the sparse representations given the dictionaries (sparse coding step)
  - Update the dictionary given the sparse representations (dictionary update step)
- Dictionaries are updated iteratively

\[
\begin{bmatrix}
X^l \\
X^h \\
Y
\end{bmatrix}
= 
\begin{bmatrix}
\Psi_c^l & \Psi^l & 0 \\
\Psi_c^h & \Psi^h & 0 \\
\Phi_c & 0 & \Phi
\end{bmatrix}
\begin{bmatrix}
Z \\
U \\
V
\end{bmatrix}
\]
Multi-Modal Super-Resolution: Algorithm

Algorithm:

- Given the learned dictionaries, the LR depth image to enhance and the HR colour image
- Extract patches $x_i^{LR}$ and $y_i$
- Using sparse coding methods (e.g., OMP), retrieve the sparse vectors $z_i, u_i$ and $v_i$
- HR depth patches are then given by: $X^h = [\Psi_c^h \ \Psi^h] \begin{bmatrix} Z \\ U \end{bmatrix}$

Key references include [Rodrigues et al. 2016], [Song, Deng et al. 2017], [Deng et al. 2017]
Multi-Modal Super-Resolution: Results

Image courtesy of M. Rodrigues
Multi-Modal Super-Resolution: Results

Super-resolving hyper-spectral images with the aid of RGB images

Image courtesy of M. Rodrigues
Image Super-Resolution: Model-Based Approach

- Sampling and Resolution Enhancement are heavily connected through wavelet multi-resolution analysis
- The acquisition process can be modelled as low-pass filtering followed by sampling
- In a camera the low-pass filtering is due to the lenses and is modelled with the point spread function
Point Spread Function and Splines

• In a camera the low-pass filtering is due to the lenses and is modelled with the point spread function
• The point spread function in a camera behaves like a spline function
Acquisition Process and Wavelet Decomposition

• The acquisition process remove the fine details of the image
• Since the low-pass filter is a spline, the acquisition process can be interpreted as a process that removes the wavelet coefficients at fine scales
• **Key insight**: Exploit the dependency across scale of the wavelet coefficients to retrieve the lost details.

(a) The high-resolution image 'Peppers'
(b) Low-pass and high-pass subbands of a 2-level 2D wavelet transform of (a)
(c) We only have access to the low-pass subband of the 2-level 2D wavelet transform in (b)
Wavelet Decomposition and Multiresolution
Wavelet Decomposition and Multiresolution

[Images of a person with a camera and a diagram related to wavelet decomposition]
Modelling of Dependencies Across Scales

- We model lines of images as piecewise regular functions defined as the combination of a piecewise polynomial signal and a globally smooth function that lies in shift-invariant subspace:

\[ \alpha \]

\[ \beta \]

\[ + \]

\[ \gamma \]

\[ \delta \]
Modelling of Dependencies Across Scales

- We model lines of images as piecewise regular functions defined as the combination of a piecewise polynomial signal and a globally smooth function that lies in shift-invariant subspace:

\[
x(t) = p(t) + r(t) = p(t) + \sum_n y_n \varphi(t/T - n)
\]

Note that we assume: \[\langle \varphi(t), \tilde{\varphi}(t - n) \rangle = \delta_n \]
Modelling of Dependencies Across Scales

In the wavelet domain, the detail coefficients are only due to the piecewise polynomial signal

$$x(t) = p(t) + r(t)$$

$$= \sum_{n=-\infty}^{\infty} y_{J,n}^p \varphi_{J,n}(t) + \sum_{m=-\infty}^{J} \sum_{n=-\infty}^{\infty} d_{m,n}^p \psi_{m,n}(t)$$

$$+ \sum_{n=-\infty}^{\infty} y_{J,n}^r \varphi_{J,n}(t)$$

$$= \sum_{n=-\infty}^{\infty} \underbrace{(y_{J,n}^p + y_{J,n}^r)}_{y_{J,n}} \varphi_{J,n}(t) + \sum_{m=-\infty}^{J} \sum_{n=-\infty}^{\infty} d_{m,n}^p \psi_{m,n}(t).$$
Reconstruction of Piecewise Smooth Signals

**Key Insight:**
- The residual can be recovered using traditional linear reconstruction methods.
- Piecewise polynomial signals are continuous sparse signals and can be recovered using sparse sampling theory (i.e., finite rate of innovation theory [DragottiVB:07, UriguenBD:13]).

![Diagram of signal processing](image)

**Note that** we assume: \( \langle \varphi(t), \tilde{\varphi}(t - n) \rangle = \delta_n \)
Exact Reconstruction of Piecewise Polynomial Signals

Piecewise polynomial signals are continuous sparse signals and can be recovered using sparse sampling theory (i.e., finite rate of innovation theory [DragottiVB:07, UriguenBD:13]).
Reconstruction of Piecewise Smooth Signals

- remove the contribution of the reconstructed polynomial part $\hat{p}(t)$ from the samples $y_n$.
- reconstruct the residual $\hat{r}(t)$ by classical linear reconstruction.

\[ T \]
\[ x(t) \quad \tilde{\varphi}(t) \quad y_n \quad \text{FRI reconstruction} \quad \hat{p}(t) \]
\[ T \quad y_n \quad p_n \quad \varphi(t) \quad \hat{r}(t) \]
Numerical Results

Fig. 15. Our method is able to accurately recover a piecewise smooth signal from its approximation coefficients. (a) The original high-resolution piecewise smooth signal and its wavelet decomposition. (b) The linear reconstruction (22.7dB) and its wavelet decomposition. (c) TV reconstruction (28.9dB) and its wavelet decomposition. (d) Our reconstruction (47.9dB) and its wavelet decomposition.

Table II

<table>
<thead>
<tr>
<th>Sampling Kernel</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bior 4.4</td>
<td></td>
</tr>
<tr>
<td>Linear spline</td>
<td></td>
</tr>
</tbody>
</table>

Our improved method in Sec. IV-B with the linear reconstruction method and some of the state-of-the-art algorithms, we show the upsampling results of two different downsampling kernels in terms of PSNR and SSIM (structural similarity index [31]) in Table II and Table III respectively. Visual comparisons on 'Zebra' (downsampling kernel of bior4.4) and 'Comic' (downsampling kernel of linear spline) are shown in Fig. 16 and Fig. 17 respectively. Note that for self-learning method [17], we use a third-party implementation [32] and we cannot guarantee the implementation duplicates the original results. For other methods we ensure the comparison is fair because we modified the blurring kernel in the source codes to the specific kernel we use (bior4.4/linear spline), and for dictionary-based methods of [15] and [16], the dictionary was re-trained with the same kernel used in upsampling process.

Our proposed method is universal in the sense that it works with different blurring kernels. The results demonstrate that our proposed basic (fast) method, with no learning involved, outperforms other reconstruction-based algorithms, e.g. total variation [6], contourlet [7] and even one of the dictionary learning methods [15]. The improved method – FRESH is robust and outperforms state-of-the-art methods in different categories. Moreover, our method leads to visually pleasant edges.

2) Upsampling of Images Taken With a Camera: Finally, we show that the proposed algorithm is also able to upsample the images taken with a real camera, where the blurring due to lens is not exactly a scaling function as assumed previously but can still be modeled as a spline. We demonstrate in Fig. 18 that the algorithm achieves visually good performance for upsampling factor of 4. In the following result, the original photographs are taken with Canon 400D, and its point spread function is modeled by the fifth order spline. The upsampling is performed only on the luminance component of the input image and the chrominance component are simply upscaled by bicubic interpolation.

C. Computation Complexity and Discussions

Upsampling an image of size $N \times N$ to $2^K N \times 2^K N$ with the basic algorithm proposed in Sec. IV-A requires number

- Algorithm capable of **increasing the resolution of digital images up to 4X**.
- Based on applying the 1-D resolution enhancement algorithm along several directions of the image.
- The upsampled images are merged using wavelet theory.
- Self-learning further improves performance.
- **Accurately retrieve fine details** lost during the acquisition process.

[WeiD:TIP16]
FRESH: FRI-Based Single-Image Super-Resolution

Input image
128x128 pixels

Linear upsampling along columns
256x128 pixels

FRI upsampling along rows
256x256 pixels
FRESH: FRI-Based Single-Image Super-Resolution

Input image: 128x128 pixels
Linear upsampling along rows: 128x256 pixels
FRI upsampling along columns: 256x256 pixels

Original input image

Decomposition of image upsampled along rows

Decomposition of image upsampled along columns

High-res image after inverse decomposition
256x256 pixels

FRI upsampling of main and secondary diagonals of low-res image

Fusion of upsampled images based on their dominant gradient
FRESH Results: Real Data

Low-res input
64 x 64 pixels

Final result
256 x 256 pixels
Single-Image Super-Resolution: Numerical Comparisons

(a) Original
(b) Linear (25.9dB)
(c) Timofte (27.3dB)
(d) FRESH (27.7dB)
The HR depth image is reconstructed from a LR depth image with a registered HR intensity/color image as guidance. Since depth and intensity images are two modalities of the same scene, they must have strong structural similarities.

**Six Dictionaries**

<table>
<thead>
<tr>
<th>$\psi^l_c$</th>
<th>$\psi^h_c$</th>
<th>$\Phi^h_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi^l_u$</td>
<td>$\psi^h_u$</td>
<td>$\Phi^h_u$</td>
</tr>
</tbody>
</table>

Dictionaries learned from external datasets

Reconstructed HR depth image
Multimodal Depth Super-Resolution: Numerical Comparisons

- Method (h) is a combination of FRESH and multimodal dictionary learning
- Method (g) is based on deep learning
Conclusions

• The notion of sparsity is still essential to develop and understand both model-based or data-driven methods.

• Data-driven algorithms based on shallow (or deep) learning still use a modelling assumption (sparsity).

• Model-based approaches are competitive when they can reflect closely the nature of the data (e.g., depth images), but lack flexibility.

• Model-based algorithm can always be combined with data-driven methods to yield algorithms with best performance.


On Data-Driven Single-Image Super Resolution

On Coupled Dictionary Learning for Multi-Modal Data
