Compressive Sampling based Single-Image Super-resolution Reconstruction by dual-sparsity and Non-local Similarity Regularizer

Shuyuan Yang, Min Wang, Yaxin Sun, Fenghua Sun, Licheng Jiao

A R T I C L E   I N F O

Article history:
Received 1 September 2010
Available online 20 February 2012
Communicated by R.C. Guido

Keywords:
Single-Image Super-Resolution
Reconstruction
Compressive sampling
Dual-sparsity
Non-local similarities
Patch-pixel-collaboration

A B S T R A C T

Recent development on Compressive Sampling (or compressive sensing, CS) theory suggests that High-Resolution (HR) images can be correctly recovered from their Low-Resolution (LR) version under mild conditions. Inspired by it, we proposed a CS based Single-Image Super-resolution Reconstruction (SISR) framework that exploits the dual-sparsity and non-local similarity constraints of images. This new framework relies on the idea that LR image patch can be regarded as the compressive measurement of its corresponding HR patch, and a sufficiently sparse coding of HR patch under some dictionary will make an accurate recovery of HR patch from its measurement possible. In order to adaptively tune the dictionary that can well represents the underlying HR patches, we reduce the SISR to a dual-sparsity constrained optimization problem with dual variables. Moreover, the pixel-based recovery is incorporated as another regularization term to exploit the image non-local similarities, which is very helpful in preserving edge sharpness. The optimization is implemented in a patch-pixel-collaboration and iterative manner, via the Singular Value Decomposition (SVD) and Orthogonal Matching Pursuit (OMP) algorithm. Experiments are taken on some natural images, remote sensing images and medical images, and the results show that our proposed method can not only provide one possible way of recovering HR image under the CS framework, but also generate HR images that are competitive or even superior in quality to images produced by other similar SISR methods.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

High-Resolution (HR) images are desired in remote sensing, medical imaging, biometrics identification and so on. To increase the images resolution, we can either reduce the pixel size by sensor manufacturing techniques or increase the chip size of sensors, which are severely constrained by the physical limitation of imaging systems (Park et al., 2003). Therefore, one should turn to algorithmic techniques to achieve resolution enhancement. Single-Image Super-resolution Reconstruction (SISR) uses signal processing techniques to obtain a HR image \( X \) from an observed degraded Low-Resolution (LR) image \( Y : Y = HBX + v \), where \( H, B, v \) represent the downsampling operator, blurring operator, and the additive noise respectively (Elad and Feuer, 1997; Katagglados and Molina, 2007). Depending on the communities involved, it goes by different names including resolution enhancement (Keren et al., 1988), HR image reconstruction (Stark and Osskoui, 1989), and Super-Resolution (SR) reconstruction (Tipping and Bishop, 2003; Tipping and Bishop, 2003), which have proved to be efficient in many applications (Greenspan et al., 2002; Kennedy et al., 2006).

The classical SISR schemes can be divided into two catalogs: regularization-based methods (Ng et al., 2007; Chan et al., 2007) and interpolation-based methods (Hou and Andrews, 1978; Li et al., 2001; Rajan and Chaudhui, 2001). It is well known that SISR is a severely ill-posed problem whose solution is not unique because of the insufficient number of LR images, ill-conditioned registration and unknown blurring operators. Therefore one can add various regularization items to further stabilize the inversion of this ill-posed problem (Ng et al., 2007; Chan et al., 2007), which are named as regularization-based methods. However, their performance degrades rapidly when the magnification factor becomes large (Baker and Kanade, 2002). Another catalog is interpolation-based methods including bilinear interpolation, bicubic interpolation (Keys, 1981) and some improved versions (Li et al., 2001; Dai et al., 2007; Sun et al., 2008). Interpolation-based methods are effective in preserving the edges in the zoomed image, while limited in modeling the visual complexity of the real images.

Another popular catalog, examples-based methods (Freeman et al., 2000; Roweis and Saul, 2000; Sun et al., 2003; Chang et al., 2004), is developed recently. It believes that HR image can be predicted by learning the co-occurrence relationship between a set of LR example patches and their corresponding HR version. A representative work was proposed by Chang et al. (2004), which...
assumed that a LR image and its corresponding HR image have similar local geometry. It is extended and developed as a sparse representation based scheme, which attracts increasing interest recently (Freeman et al., 2002; Elad and Aharon, 2006; Mairal et al., 2008; Yang et al., 2008; Yang et al., 2010). In paper (Yang et al., 2008; Yang et al., 2010), the authors generate dictionaries by randomly sampling raw patches from training images of similar statistical nature, which is shown to lead to state-of-the-art result. However, in order to obtain an accurate reconstruction, a large number of example patches are needed to establish two dictionaries (HR and LR dictionaries). Therefore, when the number of atoms in the dictionary decreases, the reconstruction result also degrades remarkably.

In this paper, we indicate another solution for SISR via the recent developed Compressed Sensing (or Compressed Sensing, CS) theory, which provides a possible way of recovering sparse signals from their projection onto a small number of random vectors (Candès and Tao, 2006; Donoho, 2006). Under the framework of CS, Y can be regarded as the low dimensional projection of the HR image X: Y = HBX + v, where Φ is corresponding to a degradation matrix in SISR. According to Donoho (2006), if v = 0 and X is sparsely represented under a dictionary D: X = Dz (z is a sparse coefficient vector with few non-zero elements) and Y has sufficient number of measurements, z can be correctly recovered from Y = ΦX = ΦDz by solving such an l₀-norm optimization problem,

\[
\mathbf{z} = \arg \min \{ \| \mathbf{x} \|_0 \} \quad \text{s.t.} \quad \mathbf{y} = \Phi \mathbf{D} \mathbf{z}
\]

and then the HR image \( \mathbf{X} \) can be estimated by \( \mathbf{X} = \mathbf{D} \mathbf{z} \). Inspired by this idea, paper (Fan, 2009) proposed a wavelet domain and CS based SISR method. However, an accurate reconstruction of signals largely depends on whether the employed sparse domain can well represent the underlying signal (W.S Dong, 2011). Obviously the contents may vary significantly across different image patches. Although the analytically designed wavelet dictionary is characteristic of rapid implementation, it cannot adaptively represent arbitrary patches that are corresponding to unknown patterns.

Learning a sparse dictionary from example patches is a prospecting solution to adaptively representation in this framework (Yang et al., 2011). In this paper we introduce machine learning technology into the CS based SISR framework and propose a new dual-sparsity and non-local similarity regularization method. This new method relies on the idea that LR image patch can be regarded as the compressive measurement of its corresponding HR patch, and a sufficiently sparse representation of HR patch under an appropriate dictionary will make an accurate recovery of HR patch possible. Two sparsities, i.e., the sparsity of HR image under the dictionary, and the sparsity of example HR image patches under the dictionary, are both addressed to reduce the SISR to a dual-sparsity regularization optimization problem with dual variables. Unlike the traditional dictionary learning approach (Elad and Aharon, 2006), we propose an optimization method that can simultaneously tune the dictionary and sparse coefficients. Moreover, the pixel based recovery is incorporated as another regularization term to exploit the non-local similarities of images. The optimization is implemented in a patch-pixel-collaboration and iterative manner, via the Singular Value Decomposition (SVD) and Orthogonal Matching Pursuit (OMP) algorithm. Experiments are taken on some natural images, remote sensing images and medical images, and the incoherence of the learned dictionary and sensing matrix is also investigated. The results show that our proposed method can not only provide one possible way of recovering HR image under the CS framework, but also generate HR images that are competitive or even superior in quality to images produced by other similar SISR methods.

The rest of this paper is organized as follows. Section 2 details our proposed CS based SISR approach. In Section 3, experiments are taken to compare the proposed method with related schemes. The conclusions are finally summarized in Section 4.


In this section, we firstly depict the proposed CS based SISR approach in detail, and then reduce the SISR as a dual-sparsity and dual-variable optimization problem. The non-local self-similarity regularizer used in the optimization algorithm is also explained.


The patches based reconstruction is adopted in our method, that is, the LR image patches in Y are processed in raster-scan order, from left to right and top to bottom, and then sequentially recovered. Given a \( p \times p(p \in \mathbb{Z}) \) LR patch taken from Y: \( y_{LR} \in \mathbb{R}^{m \times n} \) (m = p²), our aim is to reconstruct its corresponding \( q \times q(q \in \mathbb{Z}) \) HR patch \( y_{HR} \in \mathbb{R}^{n \times n} \). The amplify factor is denoted as \( q = p^2 \). Ignoring the blurring operator and the noise, we can describe the relationship between \( y_{LR} \) and \( y_{HR} \) by

\[
y_{LR} = H y_{HR}, y_{HR} \in \mathbb{R}^{m \times n}
\]

A simple example of the matrix \( H \in \mathbb{R}^{16 \times 16} \) is as follows,

\[
H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

when \( p = 2, q = 4, m = 4, n = 16 \). The degradation matrix \( H \) is corresponding to a downsampling operator, as shown in Fig. 1.

According to the recent developed compressive sampling theory (Donoho, 2006), it is capable of recovering \( y_{HR} \) from \( y_{LR} \) under the sparsity prior of \( x_{HR} \). That is, \( x_{HR} \) can be represented as a sparse linear combination by an over-complete dictionary \( D_1 \) in \( \mathbb{R}^{n \times n} \) that is not coherent with the measurement (or sampling) matrix \( H \), i.e.,

\[
x_{HR} = D_1 z
\]

Here the “sparsity” of the decomposition coefficient \( z \in \mathbb{R}^n \) means \( \| z \|_0 = S < n < K \), and K is the number of elements (or atoms) in the dictionary \( D_1 \). Under this sparsity assumption, \( x_{HR} \) can thus be reconstructed by taking only \( m \geq O(S) \) measurements. As soon as the sparse coefficient \( z \) is determined by (5),

\[
\begin{align*}
\min_{z} & \| z \|_0 \\
\text{s.t.} & \ y_{LR} = H x_{HR} = H D_1 z
\end{align*}
\]

an estimation of \( x_{HR} \) can be obtained using (4).

![Fig. 1. The downsampling operation that is corresponding to H.](image-url)
2.2. Dual-sparsity and dual-variable optimization

The sparsity or compressibility shown in (5) is the first sparsity prior used in our method. In order to generate an over-complete dictionary \(D_h\) that can well represents the underlying HR patches, we need another sparsity assumption to adaptively tune the dictionary from a set of example HR image patches. In this section, we will reduce the learning of dictionary \(D_h\) as another sparsity-oriented optimization problem. Recent research on image statistics suggests that image patches can be well represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary (Elad and Aharon, 2006; Aharon et al., 2006; Mairal et al., 2008). Under this assumption, the HR image patches set \(Q_h = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n\) sampled from some training HR images can be represented as a sparse linear combination in a dictionary \(D_h = [d_1, d_2, \ldots, d_K] \in \mathbb{R}^{K \times n}\) with the sparse coefficient vectors \(x_i \in \mathbb{R}^K\) and \(\|x_i\|_0 < K\). The objective of designing \(D_h\) is to make the reconstruction error over \(Q_h\) be minimal under the sparsity assumption, i.e.,

\[
\begin{align*}
\min_{\mu, D_h} & \|\mu\|_0 \\
\text{s.t.} & \quad Q_h = D_h \mu
\end{align*}
\]  

(7)

with the sparse coefficients \(\mu = [\mu_1, \mu_2, \ldots, \mu_n]\) is the coefficients matrix. This is the second sparsity prior in the reconstruction.

Combining (5) and (7) together, the recovery of HR image can be reduced to such a dual-sparsity constraint optimization problem,

\[
\begin{align*}
\min_{\mu, D_h} & \|\mu\|_0 \\
\text{s.t.} & \quad Q_h = D_h \mu; \\
& \quad y_{\text{LR}} = HD_h \mu
\end{align*}
\]  

(8)

From it we can see that it is a complex non-convex optimization problem with multiple variables \((x, \mu, D_h)\). Therefore, we exchange the dual-sparsity objective function and constraints in (8), and reformulate it as (9),

\[
\begin{align*}
\min_{\mu, D_h} & \|y_{\text{LR}} - HD_h \mu\|_2^2 \\
\text{s.t.} & \quad \|\mu\|_0 < \tau_1; \quad \|Q_h - D_h \mu\|_2^2 < \tau_2
\end{align*}
\]  

(9)

Here \(\|\cdot\|_2\) is called the Frobenius norm and is defined as \(\|A\|_2 = \sqrt{\sum_{i,j} A_{ij}^2}\). \(\tau_1, \tau_2\) are two threshold about the sparsity.

Suppose that there are totally \(P\) patches extracted from \(Y\), given the \(j\)-th LR patch \((y_{\text{LR}})_j = 1, \ldots, P\), the reconstruction of its HR version \(x_{\text{LR}}(j)\) can be written as,

\[
\begin{align*}
\min_{\mu, D_h} & \|y_{\text{LR}} - HD_h \mu\|_2^2 \\
\text{s.t.} & \quad \|\mu\|_0 < \tau_1; \quad \|Q_h - D_h \mu\|_2^2 < \tau_2
\end{align*}
\]  

(10)

Combining the two objective functions together, denoting \(z = [x_1, x_2, \ldots, x_Q] \in \mathbb{R}^{K \times Q}\), \(\bar{z} = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_P] \in \mathbb{R}^{K \times P}\), \(y_{\text{LR}}(j) = H y_{\text{LR}}(j)\), \(\bar{y} = [y_{\text{LR}}, y_{\text{LR}}(1), \ldots, y_{\text{LR}}(P)] = [Q_h, H^T y_{\text{LR}}(1), \ldots, H^T y_{\text{LR}}(P)] \in \mathbb{R}^{K \times (Q+P)}\) \(H^T\) is the pseudo-inverse of matrix \(H\), we can reduce the dual-sparsity optimization problem with multiple variables in (10) to a dual-sparsity and dual variable optimization problem in (11),

\[
\begin{align*}
\min_{D_h, z} & \|\bar{y} - H D_h z\|_2^2 \\
\text{s.t.} & \quad \|z\|_0 < \tau
\end{align*}
\]  

(11)

Here \(\tau = \tau_1 + \tau_2\).

2.3. Non-local self-similarity regularizer

In fact there are often many repetitive image structures (or self-similarity) in an image, especially for natural images. Such non-local redundancy is very helpful to improve the quality of reconstructed images. The self-similarity of image patches has been well known and successfully increases the efficiency of many
image processing tasks (Buades et al., 2005). In this section, we will introduce another non-local regularizer based on self-similarity of images into (11), which is very helpful in preserving edge sharpness in image recovery.

For each local patch in an image especially natural image, we can find the similar patches in the whole image according to Gaussian neighborhood (in practice, in a sufficiently large area around G) (Buades et al., 2005). The value of the i-th pixel in xHR, XR, can then be regarded as a mean of the values of all points whose Gaussian neighborhood looks like the neighborhood of xHR. That is,

\[
x_{HR}^i = \sum_{(x_{HR}) \in G} x_{HR}^i w
\]

where \(x_{HR}^i\) is the value of the i-th pixel in the patch \(x_{HR}\) belongs to its neighborhood G, and w is the corresponding connected weights, which is calculated by

\[
w = \exp \left(-\frac{\|x_{HR}^i - x_{HR}\|^2}{h^2}\right)
\]

The formula (12) amounts to say that each recovered value \(x_{HR}^i\) in \(x_{HR}\) is a mean of the values of all points \(x_{HR}^i\) whose Gaussian neighborhood looks like the neighborhood of \(x_{HR}\). It makes a systematic use of all possible self-predictions the image can provide, in the spirit of self-similarity of images.

Assume that the number of pixels in the non-local area G is L, i.e., \(|G| = L\), and denote the neighbor vector of \(x_{HR}^i\) as \(x_{HR} \in \mathbb{R}^{L-1}\) and the corresponding weight vector as \(w \in \mathbb{R}^{L-1}\), the formula (12) can be rewritten as

\[
x_{HR} = Nx \times w.
\]

Then the i-th HR patch \((x_{HR})_i\) can be described as,

\[
(x_{HR})_i = (Nx^1 \times w^1) \times \cdots \times (Nx^r \times w^r)
\]

where \((Nx)_i\) and \((w)_i\) are the neighborhood matrix and weighted matrix of patch \((x_{HR})_i\) respectively. Both the arrangement of \((Nx)_i\), and the calculation of \((w)_i\), are determined by the other HR patch \((x_{HR})_i\) in the non-local area G, that is, \((x_{HR})_i = t(x_{HR})_i\), \((x_{HR})_i \in G\), where t is denoted by a non-local means operation in G. The recovered HR image can then be described as

\[
X = [(x_{HR})_1, \ldots, (x_{HR})_n] = D_h[x_1, \ldots, x_n] = D_hz = T(D_hz) = T(X)
\]

where T represents a non-local means operation on X. Because \(z = [B \cdot x] \in \mathbb{R}^{L \times (Q \times P)}\), so we have

\[
z = zM
\]

where M = \(\left[\frac{[0_{Q \times P}]}{[I_{Q \times P}]}\right] \in \mathbb{R}^{L \times (Q \times P)}\). This constraint (16) can be served as a new regularizer term in the dual-sparsity optimization problem (11). Adding the formula (16) on (11) and let F = 2, we get

\[
\min_{D_h, z} \{||Y - D_hz||^2_2 + \lambda \|X - D_hzM\|^2_2\}
\]

s.t., \(\|z\|_0 \leq \tau\)

An overview of the framework of our proposed compressive sampling based SISR method is shown in Fig. 2. It employs statistical machine learning technology to learn a compact dictionary from HR example patches instead of directly taking them as a dictionary, so the optimized dictionary can capture the inner structure of underlying HR image patches. Therefore, it can avoid the establishment of two dictionaries from a large amount of example patches (Yang et al., 2008; Yang et al., 2010). Moreover, a non-local self-similarity assumption on the image is also considered in our method, which is helpful in preserving edge sharpness in image recovery.

2.4. Procedure of the proposed algorithm

In order to solve (18), we introduce a new variable \(X = D_hz \rightarrow D_hz\) and use the Lagrange multiplier to change (18) to a three variable optimization problem,

\[
\min_{D_h, z} \{||Y - D_hz||^2_2 + ||X - T(X)||^2_2 + \lambda ||X - D_hzM||^2_2\}
\]

s.t., \(\|z\|_0 \leq \tau\)

The procedure of this dual-sparsity and non-local regularizer based super resolution image reconstruction algorithm can be described as below:

**Algorithm 2.**

**Input:** The training HR example patches set \(Q_h\), a LR image \(Y\), magnification factor \(a\) and the degradation matrix \(H\).

**Output:** HR image \(X\).

**Step 1.** Extract patches \([(y_{HR})_i]\) from LR image \(Y\), which is taken starting from the upper-left corner (perhaps with some pixel overlap in each direction) to obtain \(Y = [Q_h, H^T(y_{HR})_1, \ldots, H^T(y_{HR})_n]\).

**Step 2.** Initialize \(X(0)\) by a bicubic interpolation algorithm;

**Step 3.** Repeat the following steps to complete the recovery of \(X\):

**Step 3.1** Update \(D_h\) and \(z\) using Algorithm 2;

**Step 3.2** Update \(X(t) = D_hz(t)|z(t)M\) (patches based updation);

**Step 3.3.** Update \(X(t+1) = T(X(t))\) (pixels based updation);

Until convergence.

**Step 4.** Output the high resolution image \(X\).
and the $k^{th}$ row in the sparse coefficients $\tilde{a}$ as $\tilde{a}^k$. Then $d_k\tilde{a}^k$ can be separated from $D_k\tilde{a}$, as described as follows:

$$
\|V - D_k\tilde{a}\|^2 = \|V - \sum_{i=1}^{K} d_i\tilde{a}^i\|^2 = \|(V - \sum_{i=1}^{K} d_i\tilde{a}^i) - d_k\tilde{a}^k\|^2 = \|E_k - d_k\tilde{a}^k\|^2
$$

(23)

where $E_k = (\tilde{V} - \sum_{i=1}^{K} d_i\tilde{a}^i)$ is the approximation error when the $k^{th}$ atom is removed. The SVD of $E_k$ will produce the closest rank-1 matrix that minimizes the error in (23). Perform the SVD on $E_k; E_k = UDV^T$ and denote the maximum singular value as $\Sigma_{11}$, and denote its corresponding vectors in $U$ and $V$ as $u_1$ and $v_1$ respectively. Therefore, taking $d_k = u_1$ and $\tilde{a}^k = \Sigma_{11}v_1$ can eliminate the largest component of the error matrix $E_k$, and thus will effectively minimize the error as defined in (23). However, such an operation will be a mistake, because the new vector $\tilde{a}^k$ is very likely to be filled, since in such an updation of $d_k$ and $\tilde{a}^k$, we do not enforce the sparsity constraint. Therefore a simple remedy scheme is used here. Define $\omega_k$ as the group of indices pointing to examples that use the atom $d_k$, i.e., those where $\tilde{a}^k(i) \neq 0$. Thus, $\omega_k := \{i|1 \leq i \leq K, \tilde{a}^k(i) \neq 0\}$

(24)

Define $\Omega_k$ as a matrix of size $(P\times Q)\times|\omega_k|$, with ones on the $(\omega_k(i))_{th}$ entries, and zeros elsewhere. When multiplying $\tilde{a}^k = \Omega_k\omega_k$, this shrinks the row vector $\tilde{a}^k$ by discarding the zero entries, resulting with the row vector $\tilde{a}^k = \Omega_k\omega_k$. Similarly, the multiplication $V_k^T = V_k\Omega_k$ creates a matrix of size $n \times |\omega_k|$ that includes a subset of the examples that are currently using the atom $d_k$. The same effect happens with $E_k^T = E_k\Omega_k$ implying a selection of error columns that correspond to examples that use the atom $d_k$.

The procedure of the dictionary learning algorithm based on dual-sparsity can be described by Algorithm 2:

**Algorithm 2. Dictionary Learning based on Dual-sparsity**

- **Step 1. Initialize** the dictionaries $D_k$ randomly;
- **Step 2. Repeat** the following steps to complete the learning of dictionary $D_k$:
  - **Step 2.1. Under the condition of fixed dictionary**, we solve (22) using OMP algorithm to obtain $\tilde{a}$;
  - **Step 2.2. Update the dictionary $D_k$**:
    - For $k = 1$ to $K$
      - (2.2.1) Calculate $\omega_k$ by the formula (24) and calculate $E_k^T = E_k\Omega_k$;
      - (2.2.2) Perform SVD on $E_k^T$;
      - (2.2.3) Update the dictionaries and their sparse coefficients using largest singular value of $E_k^T$ and the corresponding singular vectors:
    
    $$d_j = u_1; \tilde{a}^k = \Sigma_{11}v_1;$$

(25)

*Until convergence*.

### 3. Simulational experiments

In this section, we give several illustrative examples to demonstrate the effectiveness of our proposed method. All the simulations are conducted in MATLAB 7.0 on PC with Intel Core2/1.8GHz 1G. The test LR images include 256 × 256 natural image, remote sensing and medical images, and we aim to recover their 512 × 512 HR images from them and some HR images examples.

#### 3.1. Training patches and some quantitative guidelines about the recovery result

In our method, some training images are used to generate example HR patches, including twenty natural/six remote sensing/six medical images. Fig. 3 shows eight natural images contain scenery, animals and plants, three remote sensing images contain water, building, farmland, and three medical images contains CT anatomy of the brain.

The natural images is of size 512 × 512, and the remote sensing images and medical images are of size 1024 × 1024. In the establishment of training patches set, $(p \times q) \times (p \times q)$ image patches are taken from these images. In order to evaluate the performance of SISR algorithms, some experiments are designed and the results are compared by visual quality subjectively and some guidelines in estimating the reconstituted images including the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and mean structural similarity (MSSIM) (Wang et al., 2004; Alan et al., 2008; Sumohana et al., 2008).

Define the PSNR of the recovery image:

$$
\text{PSNR} = 10 \times \log(255^2/\text{MSE})(\text{db})
$$

(26)

where $\text{MSE} = \sum_{i=1}^{n} \sum_{j=1}^{m} (f_{i,j} - f_{i,j})^2/(m_1 \times m_2)$, and $f$ represents the original $m_1 \times m_2$ image and $f'$ represents the recovered image. The higher of PSNR is, less distortion of the recovered image.

SSIM measures the structural similarity between the original and recovered images, which is defined as:

$$
\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\mu_x^2 + \mu_y^2 + \sigma_x^2 + \sigma_y^2 + c_1}.
$$

(27)

where $\mu_x, \mu_y$ are the intensity measure, comparability measure and structure measure respectively; $\sigma_x, \sigma_y$ represent the variance of the original and recovered image respectively; $\alpha_x, \gamma$ are the weighted coefficients of intensity, comparability and structure measure respectively. SSIM is related with the distortion of the visual sensing. Dividing the original and recovered image into $L$ blocks, we can obtain a series of image blocks $(x_i, y_i; i=1, 2, \ldots, L)$. The MSSIM is defined as:

$$
\text{MSSIM} = \prod_{i=1}^{L} \frac{\text{SSIM}(x_i, y_i)}{\text{SSIM}(x_i, y_i)}
$$

(28)

The higher of SSIM and MSSIM are, much similar of the structure of the recovered image to the original image.

#### 3.2. Experiment 1: experiments on natural images

In this test, $2 \times 2$ patch in LR images is considered locally each time together, with an overlap of one pixel between adjacent LR patches. In order to maintain the continuity of a HR block with its neighbors, correspondingly 2af × 2af = 4 × 4 patches with an overlap of af pixels between adjacent HR patches are extracted from training samples. We compare our proposed CS based SISR approach by dual-sparsity and Non-Local Similarity Regularizer (here we name it as CS_DS-NLS), with Cubic, Bilinear interpolation method, Yang’s example-based method (Yang et al., 2010) that shows to lead to state-of-the-art SISR result, the compressive sampling method with fixed dictionary generated by Discrete Cosine Transform (DCT) (here we name it as CS_DCT) and compressive sampling method with only dual-sparsity constraints (here we name it as CS_DS).

The interpolation methods are implemented by the Image Processing Toolbox in Matlab7.0 Software. The example-based method (Yang et al., 2010) uses randomly selective raw patches to form pair dictionaries and Hongkang Lee’s package http://www.eecs.berkeley.edu/~yang/ is used in our test. Both 100’000 and 10’000 training patches are used for learning HR and LR dictionaries respectively, with the size of dictionary being 1024. The difference between the fixed dictionary based method and our
method is the employment of a discrete cosine dictionary and a solely sparsity, and there are also 1024 atoms in the DCT dictionary.

In our proposed CS_DS_NLS method, we take 10'000 patches randomly from the training samples and 50, 500 atoms are adopted respectively, that is, \( Q = 10'000; K = 50, 500; \) \( h = 10; \varepsilon = 0.1 \times K; \) \( \lambda \) is experimentally determined by varying it from 0 to 1, with equal space 0.05, and the best result is adopted. In CS_DS method, \( Q = 10'000; K = 50, 500. \) Table 1 showed the PSNR, SSIM, MSSIM of the reconstructed Lena, Peppers and Boat images by different methods, where bold indicate the best result. From the table we can see that Yang’s method and our proposed method outperformed the Bicubic, Bilinear interpolation methods remarkably about the three guidelines.

Fig. 4(a) plots the original Lena image and Fig. 4(b)–(f) plot the reconstructed Lena images by Bicubic-, CS_DCT-, CS_DS-, CS_DS_NLS-based methods and Yang’s methods respectively, and Fig. 4(g)–(l) show the amplification of Fig. 3(a)–(f) respectively. From it we can see that the edges of images obtained by interpolation methods are illegible. CS_DCT-based method outputs the worst result, which can be explained by the reason that DCT dictionary cannot sparsely represent the image. The reconstruction result of Yang’s method outperforms CS_DS when \( Q/K = 100'000/1024. \) However, the number of training patches in CS_DS is only one tenth of Yang’s method. When 500 atoms are used in CS_DS, it outputs comparable result with that of Yang’s. The CS_DS based method is comparable with that of Yang’s when \( K = 500. \) When the NLS regularizer is added in CS_DS_NLS method, we can see that the edge sharpness of the reconstructed image is well preserved compared with that of other methods, which is consistent with the numerical criterions.

In the above experiment, we do not consider the existence of noise \( v. \) To investigate the robustness of our algorithm, in this experiment we added Gaussian noise (with the mean and variance being 0 and 5 respectively) to the LR images. In CS_DS- and CS_DS_NLS-based methods, there are 10’000 example patches and 500 atoms in the dictionary. In Yang’s method, 10’0000 patches and 1024 atoms are used. Table 2 shows the numerical results of the reconstructed images of different methods, where bold indicate the best result. From it we can see that when noise exists, the variation of the performance of CS_DS and CS_DS_NLS is less than the Bicubic interpolation and Yang’s method.

Fig. 5 shows the reconstruction results of Boat image, where Fig. 5(a) is the original image and Fig. 5(b)–(d) are the result of CS_DS-, CS_DS_NLS-based methods and Yang’s method. Fig. 4(e)–(h) show the amplification of the local region in Fig. 5(a)–(d) respectively. From it we can see that the reconstruction image of
### 3.3. Experiment 2: experiments on remote sensing and medical images

We investigate the performance of our proposed method on the remote sensing images came from the USC_SIPi image database and the true medical CT images. There are 6 training images for two classes of images, from which HR example patches are randomly generated. We let $K = 50$ in CS_DCT-, CS_DS-, CS_DS_NLS-based methods and Yang’s method, and let $Q = 10'000$ in CS_DS-, CS_DS_NLS-based methods and Yang’s method. The other experiments condition is the same with that of experiment on natural images.

The reconstruction result of four test images shown in Fig. 6 by different methods is shown in Table 3, where bold indicate the best result. The visual result of two images is shown in Fig. 7.

From those we can see that CS_DS_NLS outperform the best result among these methods, however, the improvements of these sparsity and sparse coding based methods over the interpolation-based method are not as remarkable as that of natural images. The reason perhaps is the different resolution of training remote sensing images and medical CT images (All the remote sensing images are of different high resolution relatively lower than 10 m x 10 m and the CT images are obtained from different devices). Therefore, the texture, contours, and edges in different remote sensing images and medical CT images manifest various kinds of characteristics, which is negative in the example-based methods.

### 3.4. Experiment 3: coherence analysis of the dictionary and sensing matrix

As mentioned above, in the compressive sampling based SISR methods, an accurate recovery of HR image relies on two conditions, one is the sufficient measurements for an accurate reconstruction, and another is the incoherence between $H$ and $D_h$.

In this experiment, a coherence measure between $H$ and $D_h$ is given by the generalized mutual coherence coefficient:

$$
\mu(H,D_h) := \max_{1 \leq i,j \leq M} \frac{|(h_i,d_j)|}{\|h_i\| \cdot \|d_j\|}
$$

in (D.Carvajalino and Sapiro, 2009). Here $\mu(H,D_h)$ measures the maximal correlation between both matrix elements (see also (Park et al., 2003) for the related definition of mutual coherence of a dictionary, which plays an important role in the success of basis pursuit and the greedy pursuit algorithms). It turns out that when $H$ is Gaussians or ±1 random matrices, then $H$ are largely incoherent with any fixed sparsifying basis with overwhelming probability. However, $H$ cannot satisfy the above condition. In paper (D.Carvajalino and Sapiro, 2009), the authors indicated that as long as the number of measurements $m$ satisfies $m \geq C \cdot \mu(H,D_h) \cdot S \cdot \log N$ (S is the sparsity degree and $N$ is the length of signal) for some positive constant $C$, then, with overwhelming probability, the reconstruction is exact (even actually using $l_0$ for sparse promotion instead of $l_0$, making the optimization problem convex).

We investigate the mutual coherence in CS_DS_NLS with the variation of dictionary size and the number of patches, by taking the Peppers image, remote sensing image in Fig. 6(b) and medical image in Fig. 6(d) as the examples. Considering the random initialization of the dictionary, 10 independent experiments are taken for a pair of parameters (patch size, dictionary size), and the average result is shown in Table 4. From it we can see that when the patch size and dictionary size vary, the mutual coherence coefficient CS_DS_NLS is clearer than that of other methods, which can be explained by the non-local similarity regularizer that leads to an intrinsic filtering or denoising.

### Table 3: Comparison result of different methods on natural images.

<table>
<thead>
<tr>
<th>Image</th>
<th>Methods</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>Bicubic</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Bilinear</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DCT-</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Yang’s</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DS-</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DS_NLS</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSIM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSSIM</td>
</tr>
<tr>
<td>Peppers</td>
<td>Bicubic</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Bilinear</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DCT-</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Yang’s</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DS-</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DS_NLS</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSIM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSSIM</td>
</tr>
<tr>
<td>Boat</td>
<td>Bicubic</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Bilinear</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DCT-</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Yang’s</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DS-</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>CS_DS_NLS</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSIM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSSIM</td>
</tr>
</tbody>
</table>

*Note: The visual result of two images is shown in Fig. 7.*
varies a little. So there exists a constant $C$ that satisfies the high probability reconstruction condition.

It should be mentioned that in our paper we only present the result of $af = 2$ for the reason that CS based SISR method degrades rapidly when the sampling ratio is very low. The compressive ratio is calculated as $m/N$, and in our experiment $af = 2$ and $N = 9m$, so the compressive ratio is $1/9 \approx 11\%$, which is too low to recover the original image under the CS framework.

3.5. Experiment 4: influences of the patch size, dictionary size on our method

In this experiment, we investigate the influences of the patch size and dictionary size on our method. We let $K = 500$ (when patch size is 4, the number of atoms in the dictionary should be larger than 64 to make the dictionary over-complete) in CS_DS and CS_DS_NLS, and 10,000 training sample patches are used. The results of Lena, Boat and Peppers images are listed in Fig. 8, where different patch size/pixel overlap are considered.

From it we can see that our proposed algorithm performs best when the patch size takes 2 and 1 pixel overlap. Its performance degrades when the patch size becomes large while the number of overlap pixel remains unchanged. When the overlap is fixed, the smaller of the patch size, better construction result can be obtained.

4. Conclusion

Example-based SISR is a popular technology that receives increasing interests in recent years. Most existing example-based SISR approaches interpret the “learning” as just a kind of “searching” for the best-matched LR patch, and then “pasting” the corresponding HR component. Therefore, two dictionaries should be established, which have a large amount of atoms and the algorithm
is of high complexity. This paper indicated another possible solution for SISR via the recent developed compressive sampling theory. It only uses a single dictionary to realize the reconstruction, and the machine learning technology is introduced into the CS based SISR framework. Compared with other available SISR schemes, the proposed method has the following characteristics,

1. It realizes the SISR under the framework of compressive sampling theory. If the CS information matrix \( A_s = \Phi D \) satisfy the restricted isometry property (RIP) condition (Donoho, 2006) and the number of measurements \( m \) satisfies \( m \geq C \cdot \mu(\Phi, D) \cdot S \cdot \log N \) (\( S \) is the sparsity of \( X \) and \( N \) is the length of \( X \)) for some positive constant \( C \), then, with overwhelming probability, the reconstruction is exact (even actually using \( l_1 \) for sparse promotion instead of \( l_0 \), making the optimization problem convex).

2. It exploits both the dual-sparsity and non-local similarity constraints of images to recovery the HR image. Under the framework of compressive sampling, the sparsity of underlying image patches is also

---

**Table 3**

Comparison result of different methods on remote sensing and medical images.

<table>
<thead>
<tr>
<th>Image</th>
<th>Measures</th>
<th>Bicubic</th>
<th>CS_DCT</th>
<th>Yang</th>
<th>CS_DS</th>
<th>CS_DS_NLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 5(a)</td>
<td>PSNR (dB)</td>
<td>22.1732</td>
<td>21.5613</td>
<td>23.1212</td>
<td>23.7946</td>
<td>23.9030</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9725</td>
<td>0.9776</td>
<td>0.9814</td>
<td>0.9890</td>
<td>0.9915</td>
</tr>
<tr>
<td></td>
<td>MSSIM</td>
<td>0.9118</td>
<td>0.9150</td>
<td>0.9339</td>
<td>0.9402</td>
<td>0.9421</td>
</tr>
<tr>
<td>Fig. 5(b)</td>
<td>PSNR (dB)</td>
<td>22.5920</td>
<td>21.0399</td>
<td>23.4225</td>
<td>23.5056</td>
<td>23.7337</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9748</td>
<td>0.9760</td>
<td>0.9853</td>
<td>0.9817</td>
<td>0.9840</td>
</tr>
<tr>
<td></td>
<td>MSSIM</td>
<td>0.9109</td>
<td>0.9138</td>
<td>0.9320</td>
<td>0.9333</td>
<td>0.9401</td>
</tr>
<tr>
<td>Fig. 5(c)</td>
<td>PSNR (dB)</td>
<td>29.0098</td>
<td>24.1198</td>
<td>29.5822</td>
<td>29.6657</td>
<td>29.6938</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9952</td>
<td>0.9855</td>
<td>0.9959</td>
<td>0.9970</td>
<td>0.9979</td>
</tr>
<tr>
<td></td>
<td>MSSIM</td>
<td>0.9314</td>
<td>0.9260</td>
<td>0.9347</td>
<td>0.9351</td>
<td>0.9360</td>
</tr>
<tr>
<td>Fig. 5(d)</td>
<td>PSNR (dB)</td>
<td>26.4404</td>
<td>22.1010</td>
<td>27.3438</td>
<td>27.3590</td>
<td>27.3622</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9882</td>
<td>0.9766</td>
<td>0.9901</td>
<td>0.9922</td>
<td>0.9925</td>
</tr>
<tr>
<td></td>
<td>MSSIM</td>
<td>0.9288</td>
<td>0.9121</td>
<td>0.9292</td>
<td>0.9301</td>
<td>0.9303</td>
</tr>
</tbody>
</table>
taken into account in learning an appropriate dictionary for HR image, along with a non-local self-similarity assumption on the image that is very helpful in preserving edge sharpness. So the proposed method can achieve much better results than its counterparts in terms of both PSNR and visual perception.

3. It employs machine learning technology to learn a compact dictionary from HR example patches instead of directly taking them as a dictionary, so the obtained dictionary can capture the inner structure of underlying HR image patches. Consequently the proposed method need less atoms when compared...
with other example-based methods. Experimentally the learned dictionary proves to have low mutual incoherence with the degradation matrix \( \Phi \) to guarantee the accurate reconstruction of HR image patches.

4. Moreover, the proposed method learns one over-complete dictionary for HR patches instead of two dictionaries in paper (Yang et al., 2008; Yang et al., 2010), which shows to lead to state-of-the-art result. Therefore, its implementation is of low complexity.

Acknowledgment

The authors thank the anonymous reviewers for their constructive comments. This work was supported by the National Science Foundation of China under Grant No. 61072108, 60971112, the Program of NCET-10-0668, the Basic Science Research Fund in Xidian University under Grant No. JY10000902041 and China Postdoctoral Science Foundation.

References


