



Simultaneous image fusion and super-resolution using sparse representation

Haitao Yin, Shutao Li*, Leyuan Fang

College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Article history:

Received 16 August 2011
Received in revised form 20 December 2011
Accepted 10 January 2012
Available online 25 January 2012

Keywords:

Image fusion
Image super-resolution
Sparse representation

ABSTRACT

Given multiple source images of the same scene, image fusion integrates the inherent complementary information into one single image, and thus provides a more complete and accurate description. However, when the source images are of low-resolution, the resultant fused image can still be of low-quality, hindering further image analysis. To improve the resolution, a separate image super-resolution step can be performed. In this paper, we propose a novel framework for simultaneous image fusion and super-resolution. It is based on the use of sparse representations, and consists of three steps. First, the low-resolution source images are interpolated and decomposed into high- and low-frequency components. Sparse coefficients from these components are then computed and fused by using image fusion rules. Finally, the fused sparse coefficients are used to reconstruct a high-resolution fused image. Experiments on various types of source images (including magnetic resonance images, X-ray computed tomography images, visible images, infrared images, and remote sensing images) demonstrate the superiority of the proposed method both quantitatively and qualitatively.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

With the recent advances in imaging sensors, multiple images with different features can now be acquired from the same scene. By integrating the inherent complementary information, image fusion can thus yield a more accurate and complete description [1,2]. The most well-known image fusion approach is based on multiresolution analysis, such as the discrete wavelet transform [3], complex wavelet transform [4], nonsubsampled contourlet transform (NSCT) [5,6], multiscale directional bilateral filter [7] and contourlets [8]. The source images are first decomposed, and the resultant coefficients are fused by various fusion rules. Finally, by performing the inverse transformation of multiresolution analysis, the fused image can be reconstructed from the fused multiresolution coefficients. Different from the multiresolution analysis approach, Yang and Li recently proposed a novel and competitive approach based on the use of sparse representations [9,10].

However, in many applications, the source images have limited resolution. For example, in medical imaging, the resolution is constrained by a trade-off among resolution, signal-to-noise ratio and acquisition speed. In remote sensing, the obtained images also have low resolution because of the limited transmission bandwidth. Consequently, with low-resolution source images as input, the image produced by image fusion also has a low resolution.

To improve the resolution, a separate image super-resolution step has to be performed.

Super-resolution aims to generate a high-resolution image from one or more low-resolution images. Popular interpolation techniques include Bilinear, Bicubic, and edge-guided image interpolation [11]. The interpolation methods are simple and have low computation cost. However, they are not good at reconstructing high-frequency details. Another category of image super-resolution methods is based on learning. The fundamental issue of learning-based image super-resolution is how to define the relationships between the high-resolution and low-resolution images. Popular techniques include the use of the Markov random field [12,13] and locally linear embedding (LLE) [14]. In [12], the relationships between the high-frequency parts at different resolution levels are modeled as a Markov network. Belief propagation is then used to reconstruct the high-resolution image. In [13], Sun et al. used a primitive manifold with low intrinsic dimensionality. In [14], Chang et al. proposed an approach based on the LLE, with the assumption that the high-resolution image patches and low-resolution patches form manifolds of the same geometry structure. The K -nearest-neighbor strategy is then used to reconstruct the high-resolution image. However, fixing the number of neighbors may lead to blurred edges. To alleviate this problem, Yang et al. [15] proposed an approach based on sparse representation, with the assumption that the high-resolution and low-resolution images share the same set of sparse coefficients. This also has the added advantage that its computational complexity is lower than that in [14].

As mentioned above, traditional approaches generate a high-resolution fused image by performing image fusion and image

* Corresponding author. Tel.: +86 731 88822924.

E-mail addresses: haitao_yin@yahoo.cn (H. Yin), shutao_li@yahoo.com.cn (S. Li), fangleiyuan@gmail.com (L. Fang).

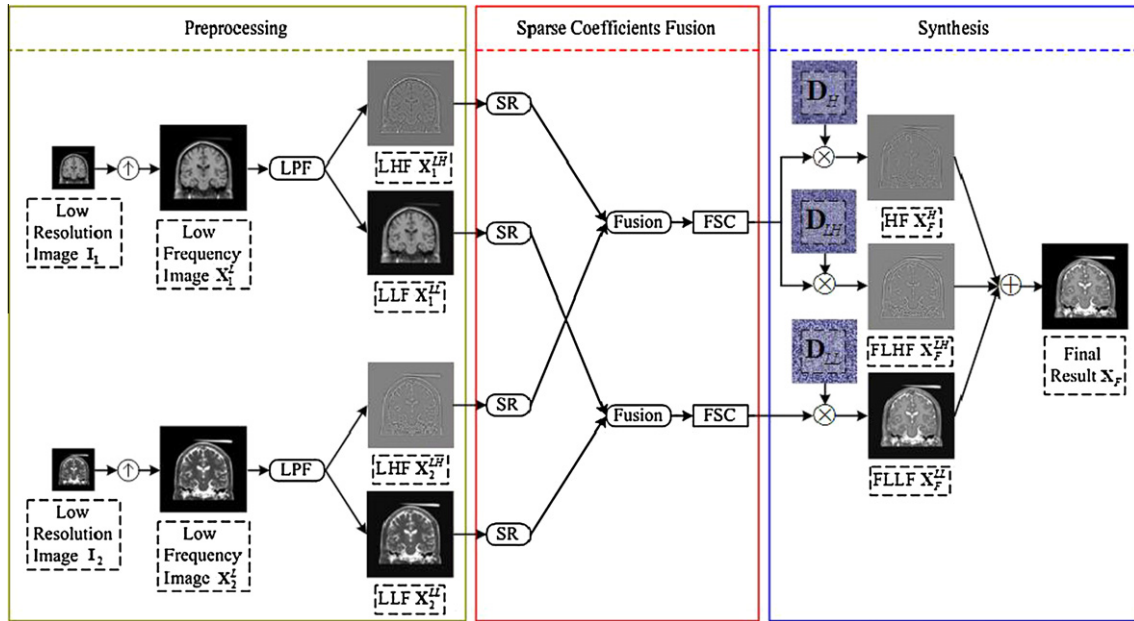


Fig. 1. Overview of the proposed method. LPF: low-pass filter. LHF: low-high frequency component. LLF: low-low frequency component. SR: sparse representation. FSC: fused sparse coefficients. HF: reconstructed high frequency. FLHF: fused low-high frequency component. FLLF: fused low-low frequency component. D_H : dictionary for HF. D_{LH} : dictionary for LHF. D_{LL} : dictionary for LLF.

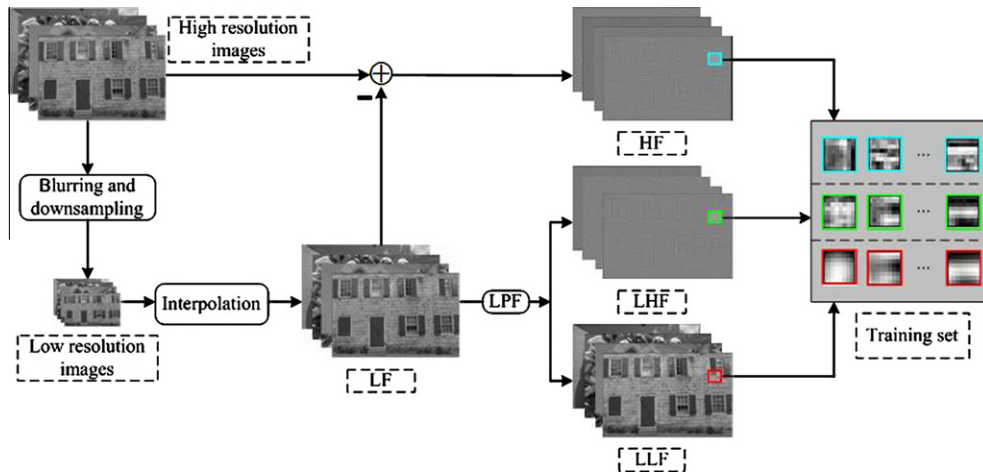


Fig. 2. Procedure to construct the training sets. LF: low frequency component of high resolution images. LPF: low-pass filter. HF: high frequency component of high resolution images. LHF: high frequency component of LF. LLF: low frequency component of LF.

super-resolution separately. However, if one performs image super-resolution first (and then followed by image fusion), any artifacts created during super-resolution will be propagated to the fusion step, and consequently reduce the quality of the fused image. On the other hand, if one performs image fusion first, the artifacts introduced during fusion are propagated to image super-resolution step, and even be magnified further.

Note that the image fusion and super-resolution may have some same foundations. For example, sparse representation not only can be used as the feature extraction in image fusion, but also can be used as prior restraint in image super-resolution. This paper proposes a novel approach that performs image fusion and super-resolution simultaneously. Comparing with the traditional approaches, artifacts will not be propagated as in a two-step approach. The proposed approach is based on sparse representation, and consists of three steps: preprocessing, sparse coefficient fusion, and synthesis. In the preprocessing step, the low-resolution source images are upscaled and decomposed into high- and

low-frequency components. In the sparse coefficient fusion step, these components are decomposed via sparse coding into sparse representations, which are then fused using image fusion rules. In the synthesis step, the final high-resolution image is reconstructed from the fused high-frequency component, the fused low-frequency component and the reconstructed high-frequency component.

The remainder of this paper is organized as follows. Section 2 briefly reviews the approach of sparse representation. Section 3 presents the proposed scheme. In Section 4, experimental results on various types of source images acquired are reported. Finally, we conclude this paper in Section 5.

2. Sparse representation

In the past decades, sparse representation has become an important tool for image denoising, compression, and super-resolution [16]. The main idea of sparse representation is that a given signal



Fig. 3. Images for training the dictionaries.

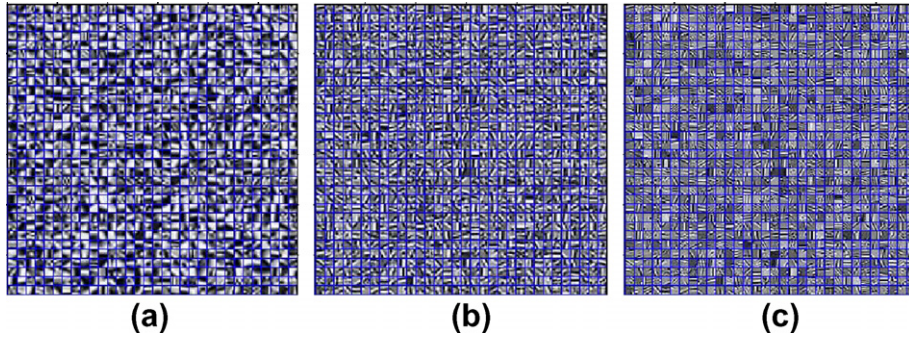
Fig. 4. Learned dictionaries corresponding to the magnification factor of 2: (a) dictionary \mathbf{D}_L ; (b) dictionary \mathbf{D}_{LH} ; and (c) dictionary \mathbf{D}_H .

Table 1

Combination of methods that will be compared with the proposed method.

Abbreviations	Fusion method	Super-resolution method
NSCT1	NSCT	Bicubic
SR1	SR	Bicubic
SOMP1	SOMP	Bicubic
NSCT2	NSCT	SRSR
SR2	SR	SRSR
SOMP2	SOMP	SRSR

$\mathbf{x} \in \mathbb{R}^n$ can be represented by a linear combination of a few atoms in an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{n \times m}$ ($n < m$). In other words, the signal \mathbf{x} can be expressed as $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^m$ is the coefficient vector with only a few nonzero elements. The sparsest $\boldsymbol{\alpha}$ can be obtained by solving the following optimization problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \leq \varepsilon, \quad (1)$$

where $\varepsilon \geq 0$ is an error tolerance, and $\|\cdot\|_0$ denotes the ℓ_0 -norm (which counts the number of nonzero entries).

However, the problem (1) is known to be intractable. To obtain a computationally practical yet provably correct solution, two relaxation approaches are commonly used [17]. The first one is based on convex optimization, which replaces the ℓ_0 -norm in (1) with the ℓ_1 -norm [18], yielding

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \leq \varepsilon. \quad (2)$$

Introducing the Lagrange multiplier λ , problem (2) can be transformed into the Lasso problem

$$\min_{\boldsymbol{\alpha}} \lambda \|\boldsymbol{\alpha}\|_1 + \frac{1}{2} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2. \quad (3)$$

The second approach is based on greedy algorithm, which iteratively updates the estimated sparse coefficients by choosing one or several atoms from the dictionary. A representative greedy algorithm with low computational complexity is the orthogonal matching pursuit (OMP) [19], which iteratively updates the estimated sparse coefficients by choosing the most relevant atom.

3. Simultaneous image fusion and super-resolution

3.1. Framework

The proposed framework for simultaneous image fusion and super-resolution is shown in Fig. 1. It consists of three major steps.

- (1) *The preprocessing step.* First, two low resolution source images $\mathbf{I}_1, \mathbf{I}_2$ are upsampled (by a given factor) using Bicubic interpolation. The resultant images (denoted \mathbf{X}_1^L and \mathbf{X}_2^L) contain low-frequency components of the underlying high-resolution source images. Using low-pass filtering, images $\mathbf{X}_1^L, \mathbf{X}_2^L$ are further decomposed into high-frequency components $\mathbf{X}_1^{LH}, \mathbf{X}_2^{LH}$ (low-high frequency (LHF) components) and low-frequency components $\mathbf{X}_1^{LL}, \mathbf{X}_2^{LL}$ (low-low frequency (LLF) components).
- (2) *The sparse coefficient fusion step.* Instead of directly using the whole images, we consider small, overlapping image patches in each source image. Each patch is of size $\sqrt{n} \times \sqrt{n}$, and is

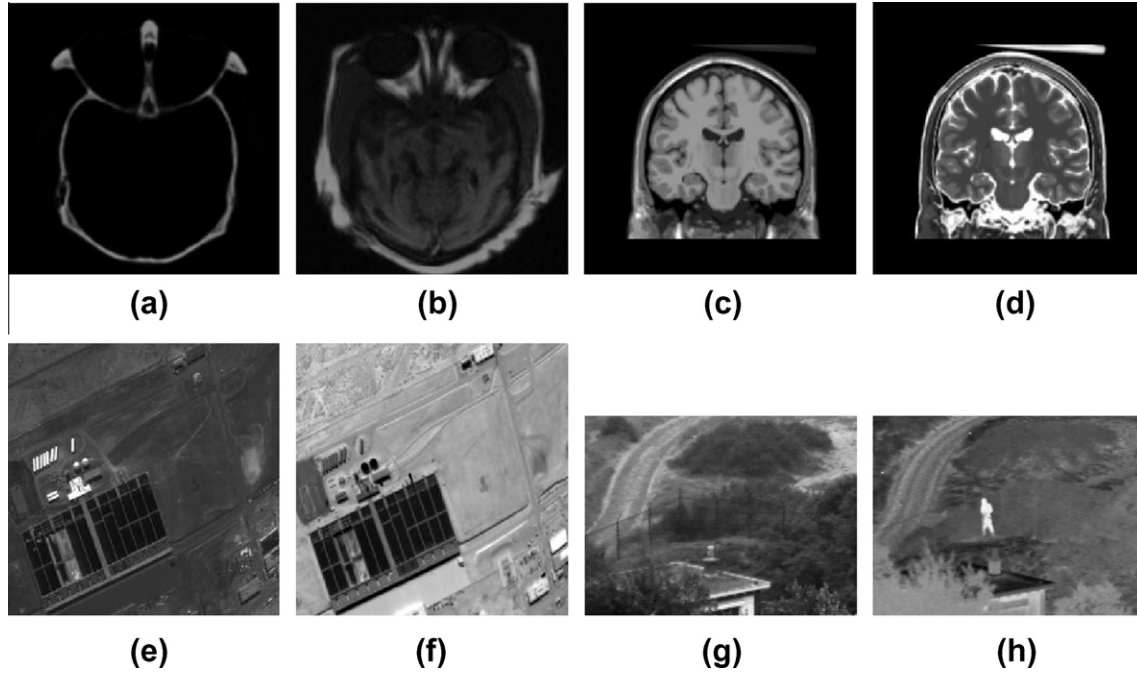


Fig. 5. Four pairs of original high resolution source images: (a) CT image; (b) MR image; (c) MR-T1 image; (d) MR-T2 image; (e) and (f) are a pair of remote sensing images; (g) visible image; and (h) infrared image.

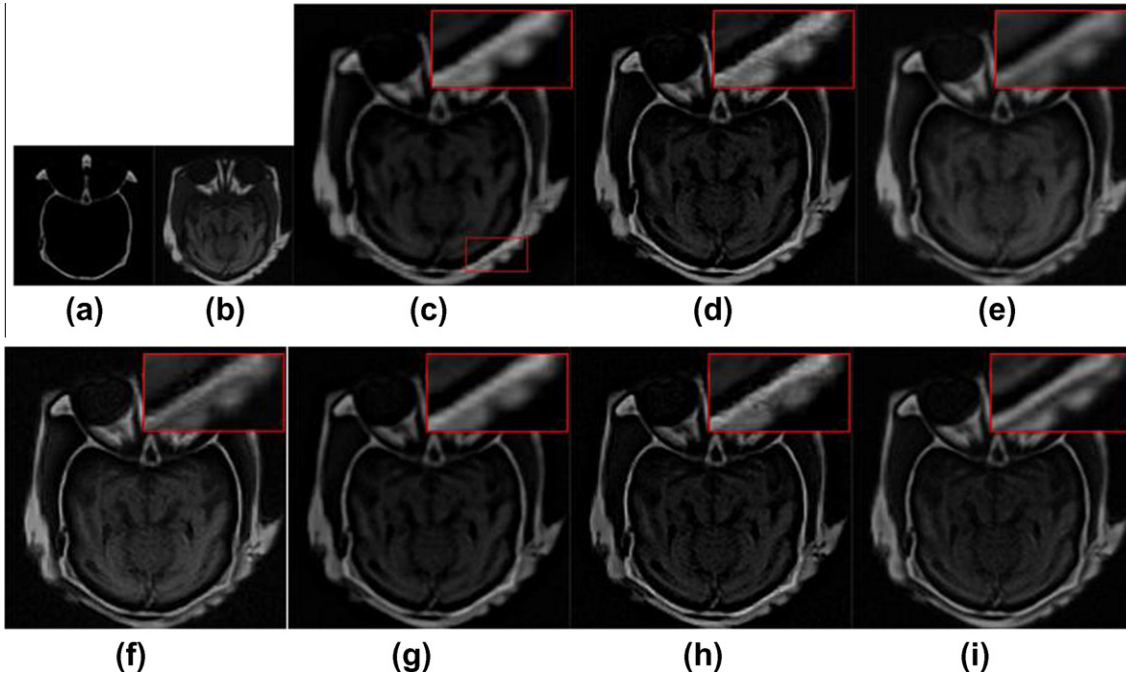


Fig. 6. Low-resolution CT and MR images and the results obtained (super-resolution factor = 2): (a) Low resolution CT image; (b) low resolution MR image; (c) NSCT; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method. The top right panel in each image shows magnification of a part containing bone structure and tissue structure information.

ordered as an n -dimensional column vector. Let the sets of patches in \mathbf{X}_1^{LH} , \mathbf{X}_2^{LH} , \mathbf{X}_1^{LL} and \mathbf{X}_2^{LL} be $\{\mathbf{x}_{1i}^{LH}\}_{i=1}^N$, $\{\mathbf{x}_{2i}^{LH}\}_{i=1}^N$, $\{\mathbf{x}_{1i}^{LL}\}_{i=1}^N$ and $\{\mathbf{x}_{2i}^{LL}\}_{i=1}^N$, respectively, where N is the number of patches in one image. Using dictionaries \mathbf{D}_{LH} and $\mathbf{D}_{LL} \in \mathbb{R}^{n \times m}$ for the high- and low-frequency components (the learning of these dictionaries will be discussed in Section 3.2), respectively, sparse coefficients of the patches can be obtained via OMP [19] by solving the following problems:

$$\begin{aligned} \hat{\alpha}_{si}^{LH} = \arg \min_{\alpha_{si}^{LH}} \|\alpha_{si}^{LH}\|_0 \quad \text{subject to} \quad \|\mathbf{x}_{si}^{LH} - \mathbf{D}_{LH} \alpha_{si}^{LH}\|_2^2 \leq \varepsilon, \\ i = 1, 2, \dots, N, \quad s = 1, 2, \end{aligned} \quad (4)$$

and

$$\begin{aligned} \hat{\alpha}_{si}^{LL} = \arg \min_{\alpha_{si}^{LL}} \|\alpha_{si}^{LL}\|_0 \quad \text{subject to} \quad \|\mathbf{x}_{si}^{LL} - \mathbf{D}_{LL} \alpha_{si}^{LL}\|_2^2 \leq \varepsilon, \\ i = 1, 2, \dots, N, \quad s = 1, 2, \end{aligned} \quad (5)$$

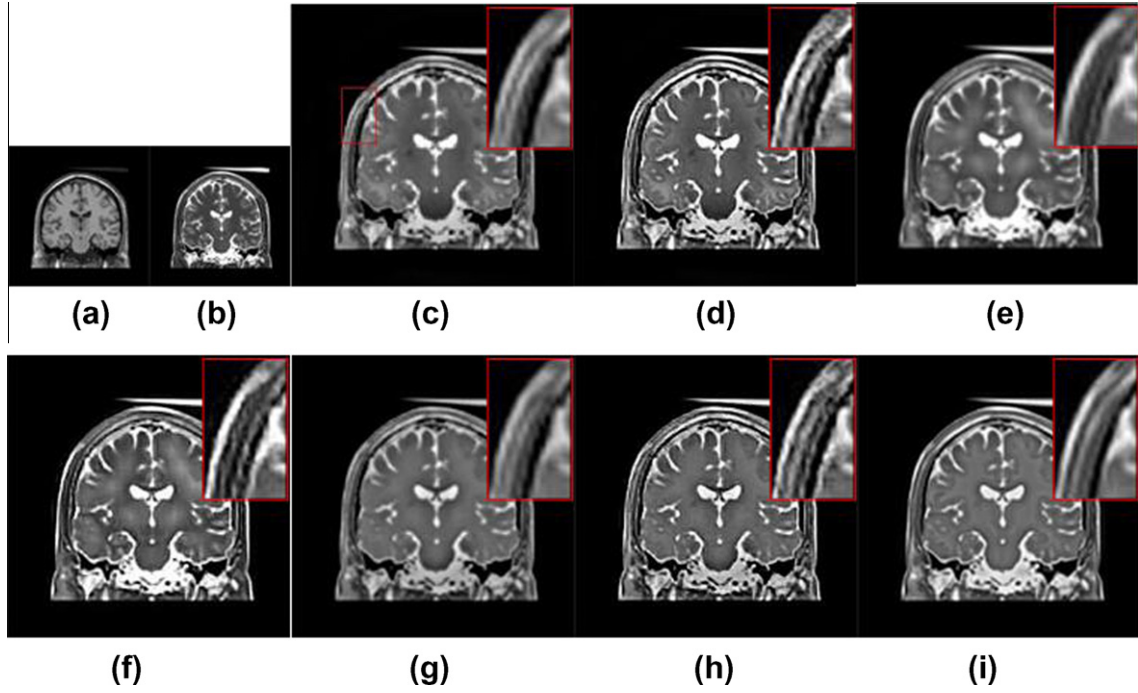


Fig. 7. Low resolution MR images and the results obtained (super-resolution factor = 2): (a) Low resolution MR-T1 image; (b) low resolution MR-T2 image; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method. The top right panel in each image shows magnification of a part containing skulls and marrows information.

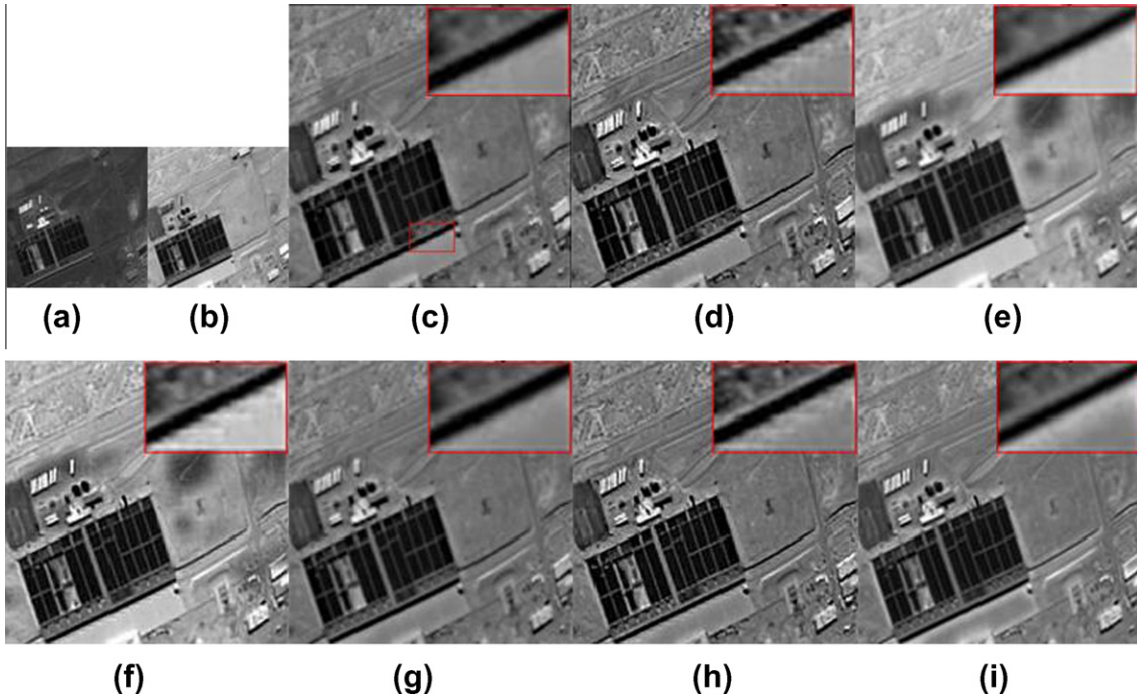


Fig. 8. Low-resolution remote sensing images and the results obtained (super-resolution factor = 2): (a) and (b) Two low resolution remote sensing images; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method. The top right panel in each image shows magnification of a part containing an edge of the building.

where $s \in \{1, 2\}$ denotes the sth source image. Next, the obtained sparse coefficients $\{\tilde{\alpha}_{1i}^{LH}\}_{i=1}^N$, $\{\tilde{\alpha}_{2i}^{LH}\}_{i=1}^N$, $\{\tilde{\alpha}_{1i}^{LL}\}_{i=1}^N$ and $\{\tilde{\alpha}_{2i}^{LL}\}_{i=1}^N$ are fused using the following fusion rule

$$\alpha_{Fi}^{LH} = \tilde{\alpha}_{si}^{LH}, \quad \hat{s} = \arg \max_{s \in \{1, 2\}} \left\{ \|\tilde{\alpha}_{si}^{LH}\|_2 \right\}, \quad i = 1, 2, \dots, N, \quad (6)$$

$$\alpha_{Fi}^{LL} = \tilde{\alpha}_{si}^{LL}, \quad \hat{s} = \arg \max_{s \in \{1, 2\}} \left\{ \|\tilde{\alpha}_{si}^{LL}\|_2 \right\}, \quad i = 1, 2, \dots, N. \quad (7)$$

(3) *The synthesis step.* From the fused sparse coefficients $\{\alpha_{Fi}^{LH}\}_{i=1}^N$ and $\{\alpha_{Fi}^{LL}\}_{i=1}^N$, the fused vectors can be reconstructed as $\mathbf{x}_{Fi}^{LH} = \mathbf{D}_{LH} \alpha_{Fi}^{LH}$, $\mathbf{x}_{Fi}^{LL} = \mathbf{D}_{LL} \alpha_{Fi}^{LL}$, $i = 1, 2, \dots, N$. Reshaping these into $\sqrt{n} \times \sqrt{n}$ patches, and averaging the resultant patches, we obtain the fused high-frequency component \mathbf{X}_F^{LH} and fused low-frequency component \mathbf{X}_F^{LL} . Recall that during

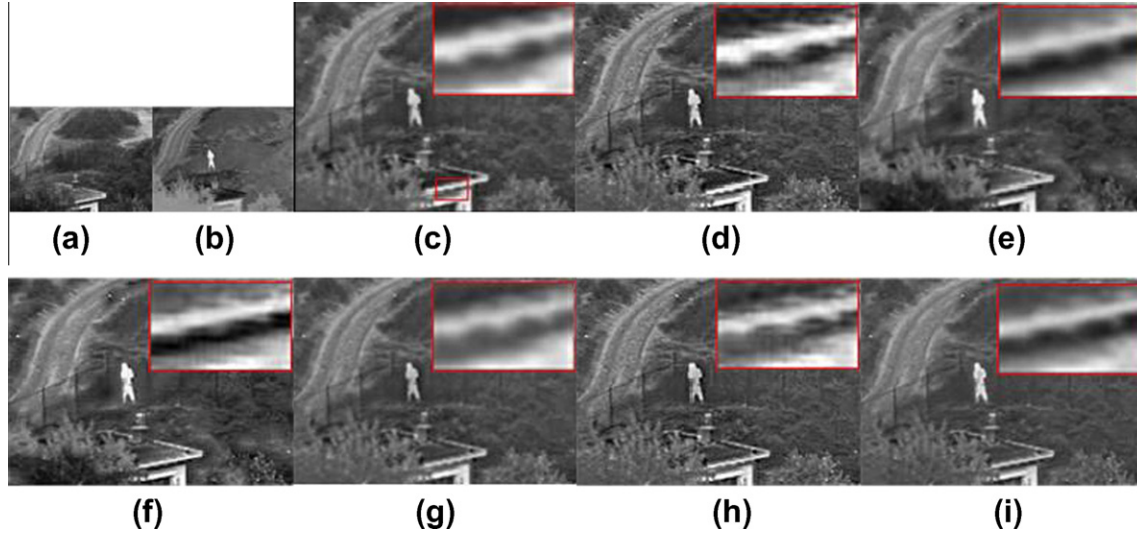


Fig. 9. Low-resolution visible/infrared images and the results obtained (super-resolution factor = 2): (a) Low resolution visible image; (b) low resolution infrared image; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method. The top right panel in each image shows magnification of a part containing an edge of the eave.

Table 2
Quantitative comparison of the results in Figs. 6–9.

Metrics	Methods						
	NSCT1	NSCT2	SR1	SR2	SOMP1	SOMP2	Proposed
Fig. 6							
Q_0	0.5188	0.4807	0.7468	0.6806	0.5388	0.5106	0.6946
Q_E	0.5119	0.4596	0.5126	0.5380	0.4831	0.4619	0.6473
Q_W	0.7465	0.7285	0.7640	0.7775	0.7167	0.7160	0.8178
$Q^{AB/F}$	0.6509	0.5476	0.6542	0.5855	0.6616	0.5527	0.7263
Fig. 7							
Q_0	0.2159	0.2253	0.2083	0.2225	0.2034	0.2248	0.2429
Q_E	0.3200	0.4340	0.3118	0.4315	0.2931	0.4455	0.4604
Q_W	0.5917	0.6536	0.5934	0.6387	0.5647	0.6581	0.6821
$Q^{AB/F}$	0.5179	0.4808	0.5348	0.5201	0.5105	0.4841	0.6117
Fig. 8							
Q_0	0.6119	0.5660	0.5611	0.5593	0.5967	0.5624	0.6307
Q_E	0.4672	0.4767	0.5548	0.5976	0.4559	0.4727	0.5821
Q_W	0.7092	0.7031	0.7796	0.7900	0.7016	0.7030	0.7738
$Q^{AB/F}$	0.4907	0.4667	0.4676	0.4786	0.4767	0.4488	0.5491
Fig. 9							
Q_0	0.5583	0.5087	0.4874	0.5050	0.5375	0.5208	0.5285
Q_E	0.4650	0.4136	0.4045	0.4094	0.4291	0.4213	0.4712
Q_W	0.6851	0.6430	0.6709	0.6672	0.6606	0.6526	0.6852
$Q^{AB/F}$	0.3827	0.3464	0.3746	0.3639	0.3676	0.3385	0.4132

preprocessing, \mathbf{X}_1^L and \mathbf{X}_2^L are regarded as low-frequency components of the underlying high-resolution source images. Hence, the fused image $\mathbf{X}_F^L = \mathbf{X}_1^L + \mathbf{X}_2^L$ can be regarded as the low-frequency component of the desired high-resolution fused image. As for its high-frequency component \mathbf{X}_F^H , we assume that images \mathbf{X}_1^H and \mathbf{X}_2^H share the same sparse representation with respect to dictionaries \mathbf{D}_H and \mathbf{D}_{LH} (where \mathbf{D}_H is used for representing \mathbf{X}_1^H). Therefore, the patches of \mathbf{X}_F^H can be obtained as $\mathbf{x}_{Fi}^H = \mathbf{D}_H \boldsymbol{\alpha}_{Fi}^H$, $i = 1, 2, \dots, N$, and subsequently the high-frequency component \mathbf{X}_F^H is obtained by averaging the patches \mathbf{x}_{Fi}^H ($i = 1, 2, \dots, N$). Finally, the final result is computed as $\mathbf{X}_F = \mathbf{X}_F^H + \mathbf{X}_F^L + \mathbf{X}_F^H$.

The whole procedure thus allows the integration of complementary information inherent in the multiple source images, while at the same time increases the resolution of the image.

3.2. Dictionaries learning

As discussed in Section 3.1, three dictionaries \mathbf{D}_H , \mathbf{D}_{LH} and \mathbf{D}_{LL} are used for \mathbf{X}_F^H , \mathbf{X}_F^{LH} and \mathbf{X}_F^{LL} , respectively. Since a learned dictionary is more effective than pre-constructed dictionaries (such as discrete cosine transformation and wavelet [20]), we will consider in this section how to learn \mathbf{D}_H , \mathbf{D}_{LH} and \mathbf{D}_{LL} .

First, we introduce the procedure used to construct the training sets (Fig. 2). Each high-resolution image \mathbf{I} is blurred and downsampled (with a user-defined factor) to generate a low-resolution image. This is then upsampled back to the original size using Bicubic interpolation. Since the resultant image lacks high-frequency component, it can be regarded as the low-frequency (LF) component of \mathbf{I} ; while the missing high-frequency (HF) component can be obtained by subtracting this LF component from \mathbf{I} . The LF is further decomposed into its high-frequency (denoted LHF (low-high frequency)) and low-frequency (denoted LLF (low-low frequency)) components, where the LLF is obtained by applying the Gaussian low-pass filter (LPF) (with size 5×5 and $\sigma = 0.5$) on LF, and the LHF obtained as the difference between LF and LLF. Patches (each of size $\sqrt{n} \times \sqrt{n}$) are extracted from HF, LHF and LLF to create training sets $\{\mathbf{z}_i^H\}_{i=1}^K$, $\{\mathbf{z}_i^{LH}\}_{i=1}^K$ and $\{\mathbf{z}_i^{LL}\}_{i=1}^K$ (where K is the number of samples) for subsequent learning of the dictionaries \mathbf{D}_H , \mathbf{D}_{LH} and \mathbf{D}_{LL} .

Recall that in the synthesis step, \mathbf{X}_F^H is reconstructed based on the assumption that \mathbf{X}_F^H and \mathbf{X}_F^{LH} have the same sparse representations with respect to the dictionaries \mathbf{D}_H and \mathbf{D}_{LH} . To ensure this, dictionaries \mathbf{D}_H and \mathbf{D}_{LH} are learned by solving the following optimization problem

$$\{\mathbf{D}_H, \mathbf{D}_{LH}, \mathbf{A}\} = \arg \min_{\{\mathbf{D}_H, \mathbf{D}_{LH}, \mathbf{A}\}} \sum_{i=1}^K \|\mathbf{z}_i^H - \mathbf{D}_H \boldsymbol{\alpha}_i\|_2^2 + \sum_{i=1}^K \|\mathbf{z}_i^{LH} - \mathbf{D}_{LH} \boldsymbol{\alpha}_i\|_2^2$$

subject to $\forall i \|\boldsymbol{\alpha}_i\|_0 \leq \tau_1$, (8)

where $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_K]$ is the matrix containing the sparse coefficients, and τ_1 controls the sparsity level. By introducing auxiliary variables $\mathbf{Z}^H = [\mathbf{z}_1^H, \mathbf{z}_2^H, \dots, \mathbf{z}_K^H] \in \mathbb{R}^{n \times K}$, $\mathbf{Z}^{LH} = [\mathbf{z}_1^{LH}, \mathbf{z}_2^{LH}, \dots, \mathbf{z}_K^{LH}] \in \mathbb{R}^{n \times K}$, $\mathbf{Z} = [\mathbf{Z}^H]^T, [\mathbf{Z}^{LH}]^T]^T \in \mathbb{R}^{2n \times K}$ and $\mathbf{D} = [\mathbf{D}_H]^T, [\mathbf{D}_{LH}]^T]^T \in \mathbb{R}^{2n \times m}$, (where T denotes matrix transpose), problem (8) can be equivalently transformed to

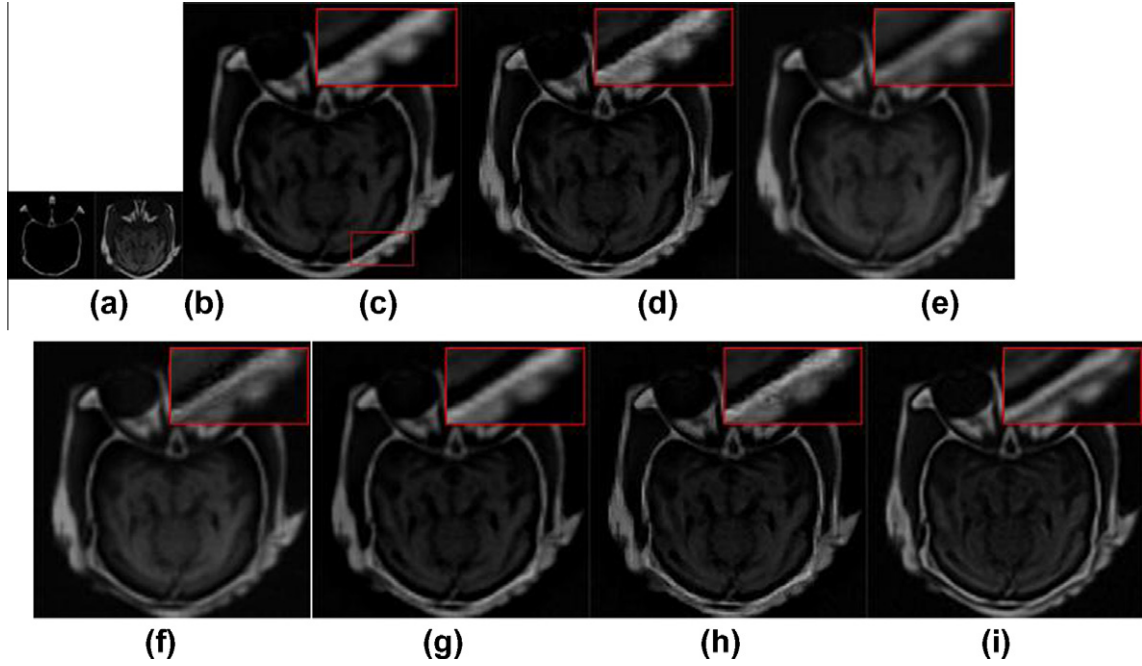


Fig. 10. Low-resolution CT and MR images and the results obtained (super-resolution factor = 3): (a) Low resolution CT image; (b) low resolution MR image; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method.

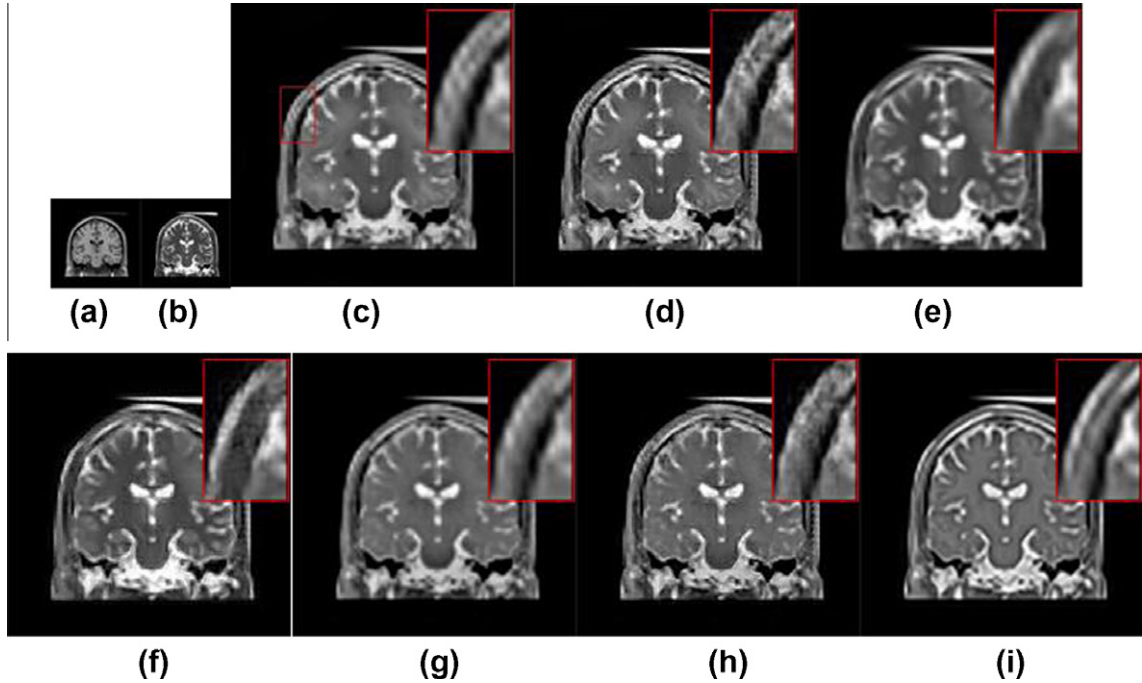


Fig. 11. Low-resolution MR images and the results obtained (super-resolution factor = 3): (a) Low resolution MR-T1 image; (b) low resolution MR-T2 image; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method.

$$\{\mathbf{D}, \mathbf{A}\} = \arg \min_{\{\mathbf{D}, \mathbf{A}\}} \|\mathbf{Z} - \mathbf{D}\mathbf{A}\|_F^2 \quad \text{subject to} \quad \forall i \|\alpha_i\|_0 \leq \tau_1. \quad (9)$$

On the other hand, dictionary $\mathbf{D}_{LL} \in \mathbb{R}^{n \times m}$ can be learned from $\{\mathbf{z}_i^{LL}\}_{i=1}^K$ by solving the following optimization problem

$$\{\mathbf{D}_{LL}, \mathbf{A}_{LL}\} = \arg \min_{\{\mathbf{D}_{LL}, \mathbf{A}_{LL}\}} \|\mathbf{Z}^{LL} - \mathbf{D}_{LL}\mathbf{A}_{LL}\|_F^2 \quad \text{subject to} \quad \forall i \|\alpha_i^{LL}\|_0 \leq \tau_2, \quad (10)$$

where $\mathbf{Z}^{LL} = [\mathbf{z}_1^{LL}, \mathbf{z}_2^{LL}, \dots, \mathbf{z}_K^{LL}] \in \mathbb{R}^{n \times K}$, $\mathbf{A}_{LL} = [\alpha_1^{LL}, \alpha_2^{LL}, \dots, \alpha_K^{LL}] \in \mathbb{R}^{m \times K}$, and τ_2 is another parameter controlling the sparsity level. In this paper, both problems (9) and (10) are solved by the popular dictionary learning algorithm K-SVD [21]. It alternates between two steps: sparse coding and dictionary updating. In the sparse coding step, any pursuit method (such as OMP and basis pursuit) can be used; whereas in the dictionary updating step, the dictionary atoms are updated one-by-one by singular value decomposition (SVD).

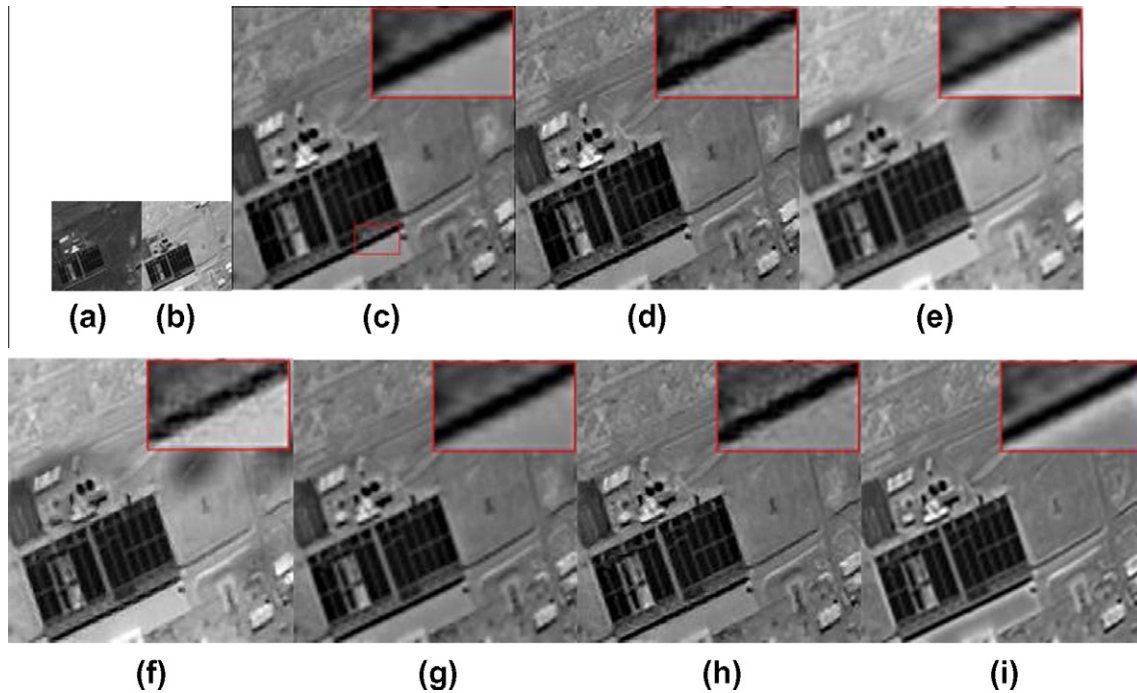


Fig. 12. Low-resolution remote sensing images and the results obtained (super-resolution factor = 3): (a) and (b) Two low resolution remote sensing images; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method.

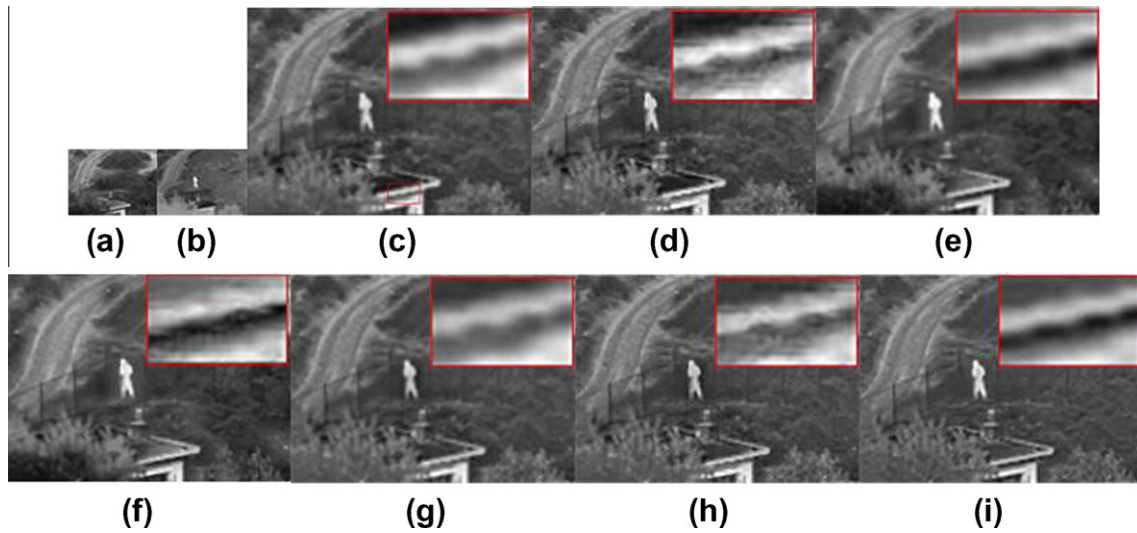


Fig. 13. Low-resolution visible/infrared images and the results obtained (super-resolution factor = 3): (a) Low resolution visible image; (b) low resolution infrared image; (c) NSCT1; (d) NSCT2; (e) SR1; (f) SR2; (g) SOMP1; (h) SOMP2; and (i) proposed method.

4. Experiments

4.1. Experimental settings

Fig. 3 shows six high-quality images (of size 768×512), downloaded from the Kodak website [22]. Using the procedure in Fig. 2, six thousand patches taken from these six images are used to learn the dictionaries, \mathbf{D}_H , \mathbf{D}_{LH} and \mathbf{D}_{LL} . As in [9,10,21], the patch size is set to 8×8 . The sizes of the learned dictionaries, \mathbf{D}_H , \mathbf{D}_{LH} and \mathbf{D}_{LL} are set to 64×1024 accordingly. The K-SVD algorithm [21] is used to solve the dictionary learning problems (9) and (10). A large number of experiments show that if and the number of atoms (m) obeys $\frac{m}{n} \approx 0.005$ ($i = 1, 2$), the learned dictionaries are good. In our experiments, the dictionary size is 1024. Therefore, the

and τ_2 in (9) and (10) are set to 6. Furthermore, OMP is used in the sparse coding stage of K-SVD, and the initial dictionaries are constructed by the randomly selected image patches. Fig. 4 shows the three learned dictionaries (corresponding to a super-resolution magnification factor of 2), where each atom is shown as a 8×8 block. A visual comparison suggests that the features represented in \mathbf{D}_{LL} , \mathbf{D}_{LH} and \mathbf{D}_H are from coarse to fine. Instead of giving the global error of OMP for solving the problems (4) and (5), six atoms are allocated for each patch-representation which is the same as the setting of dictionary learning procedures. In the following experiments, Bicubic interpolation is used during the preprocessing step and training set construction step. Other interpolation methods, such as Bilinear interpolation and Nearest-Neighbor interpolation, can also be used.

Table 3
Quantitative comparison of the results in Figs. 10–13.

Metrics	Methods						
	NSCT1	NSCT2	SR1	SR2	SOMP1	SOMP2	Proposed
Fig. 10							
Q_0	0.4321	0.4402	0.5845	0.5927	0.4327	0.4461	0.6235
Q_E	0.3951	0.4265	0.3006	0.3706	0.3586	0.4092	0.5764
Q_W	0.7056	0.7219	0.6479	0.6837	0.6707	0.6899	0.8029
$Q^{AB/F}$	0.5303	0.4536	0.4595	0.4215	0.5203	0.4496	0.6393
Fig. 11							
Q_0	0.1772	0.1891	0.1583	0.1746	0.1638	0.1816	0.2095
Q_E	0.1894	0.2513	0.1533	0.2181	0.1681	0.2381	0.3084
Q_W	0.4951	0.5410	0.4723	0.5235	0.4671	0.5252	0.6069
$Q^{AB/F}$	0.3545	0.3500	0.3358	0.3388	0.3360	0.3435	0.4611
Fig. 12							
Q_0	0.4810	0.5162	0.4378	0.4767	0.4684	0.5034	0.5202
Q_E	0.3506	0.4018	0.4000	0.4522	0.3342	0.3854	0.4582
Q_W	0.6521	0.6763	0.7033	0.7276	0.6375	0.6610	0.7222
$Q^{AB/F}$	0.3300	0.3378	0.3002	0.3101	0.3210	0.3305	0.3929
Fig. 13							
Q_0	0.4863	0.4946	0.4157	0.4352	0.4643	0.4787	0.4720
Q_E	0.3623	0.3874	0.2977	0.3282	0.3385	0.3697	0.3997
Q_W	0.6211	0.6389	0.5957	0.6190	0.5981	0.6209	0.6483
$Q^{AB/F}$	0.3058	0.2903	0.2849	0.2719	0.2918	0.2805	0.3304

Since existing methods do not perform image fusion and super-resolution simultaneously, the proposed method is compared to traditional strategies that perform image fusion and image super-resolution separately. For the image fusion step, three image fusion methods will be used, namely (1) method [9], which is based on sparse representation (SR); (2) method [10], which is based on simultaneous orthogonal matching pursuit (SOMP); and (3) method [5], which is based on NSCT. For fair comparison, the patch sizes in SR and SOMP are the same as that in the proposed method (i.e., 8×8). The global stopping error criteria of SR and SOMP are set to 0.1. The dictionaries in SR and SOMP are learned by K-SVD. The “pyrex” and “dmaxflat6” are applied as the pyramid filter and the orientation filter in NSCT. The decomposition level is 4 and the number of directions from the coarser scale to the finer scale is 4, 8, 8, and 16. As for the image super-resolution step, Bicubic interpolation and method [15] (which is based on sparse representation (SRSR)) will be used. The parameter setting of SRSR follows that in [15]. Note that if we perform image super-resolution before fusion, the upscaled source images produced by super-resolution will increase the computation cost of the fusion process. Hence, in the experiments, we perform image fusion, and then super-resolution for the traditional strategy. Table 1 summarizes the six combinations of methods.

To evaluate the proposed method, we perform experiments on both downsampled low-resolution images and naturally low-resolution images. In the former case, high-resolution images are first downsampled to create low-resolution source images, while the original images are used as reference images for performance evaluation with the fused result. In the latter case, the proposed method is performed on naturally low-resolution source images.

For a quantitative evaluation of the fused result obtained on downsampled low-resolution images and the original high-resolution image, the following four metrics will be used.

- (1) Q_0 [23], which models image distortion by a combination of correlation loss, luminance distortion, and contrast distortion. It is defined as

$$Q_0(\mathbf{X}, \mathbf{Y}) = \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2}, \quad (11)$$

where $\mathbf{X} = \{\mathbf{x}_i | i = 1, 2, \dots, M\}$ and $\mathbf{Y} = \{\mathbf{y}_i | i = 1, 2, \dots, M\}$ are the original image and test image, respectively, M is the number of pixels, \bar{x} and \bar{y} are the mean values of \mathbf{X} and \mathbf{Y} , σ_x^2 and σ_y^2 denote the deviation of \mathbf{X} and \mathbf{Y} , and σ_{xy} represents the covariance between \mathbf{X} and \mathbf{Y} . Here, since we have two source images (denoted \mathbf{A} and \mathbf{B}), we define Q_0 as

$$Q_0(\mathbf{A}, \mathbf{B}, \mathbf{F}) = (Q_0(\mathbf{A}, \mathbf{F}) + Q_0(\mathbf{B}, \mathbf{F}))/2. \quad (12)$$

- (2) Q_W and Q_E [24], which are modified versions of Q_0 with different weighting strategies. Q_W is defined as

$$Q_W(\mathbf{A}, \mathbf{B}, \mathbf{F}) = \sum_{w \in W} (w)(\lambda(w)Q_0(\mathbf{A}, \mathbf{F}|w) + (1 - \lambda(w))Q_0(\mathbf{B}, \mathbf{F}|w)), \quad (13)$$

where $\lambda(w)$ is relative importance of image \mathbf{A} compared to image \mathbf{B} in the window w , and $c(w)$ is the normalized overall salience of the window w .

- (3) Q_E is defined as

$$Q_E(\mathbf{A}, \mathbf{B}, \mathbf{F}) = Q_W(\mathbf{A}, \mathbf{B}, \mathbf{F}) \cdot Q_W(\mathbf{A}', \mathbf{B}', \mathbf{F}')^\alpha, \quad (14)$$

where \mathbf{A}' , \mathbf{B}' and \mathbf{F}' are the edge images corresponding to the image \mathbf{A} , \mathbf{B} and \mathbf{F} , respectively, and α is the contribution of the edge images compared to the original images.

- (4) $Q^{AB/F}$ [25] is defined as

$$Q^{AB/F} = \frac{\sum_{n=1}^N \sum_{m=1}^M Q^{AF}(n, m)w^A(n, m) + Q^{BF}(n, m)w^B(n, m)}{\sum_{i=1}^N \sum_{j=1}^M (w^A(i, j) + w^B(i, j))}, \quad (15)$$

where $Q^{AF}(n, m) = Q_g^{AF}(n, m)Q_\alpha^{AF}(n, m)$, and the $Q_g^{AF}(n, m)$, $Q_\alpha^{AF}(n, m)$ denote the edge strength and orientation preservation values at the pixel (n, m) , respectively, the definition of $Q^{BF}(n, m)$ is the same as $Q^{AF}(n, m)$, $w^A(n, m)$ and $w^B(n, m)$ denote the significance of $Q^{AF}(n, m)$ and $Q^{BF}(n, m)$, respectively.

The ranges of Q_0 , Q_E , Q_W , and $Q^{AB/F}$ are in $[0, 1]$. The larger the value, the better the fused result.

4.2. Experimental results on downsampled images

Four pairs of source images, including CT, MR, remote sensing image, visible/infrared image, (Fig. 5¹) are used in the experiments. Experiments are performed on a PC with Pentium dual-core 2.93 GHz CPU and 2 GB RAM, operating under MATLAB 7.10.

In this section, we first create artificial low-resolution source images by downsampling the original high-resolution source images in Fig. 5 by a factor of 2. These low-resolution images are then fused and upscaled by a factor of 2 to produce the fused image. The first experiment is performed on the fusion of a CT image, which shows the bone structure (Fig. 6a), and a MR image, which exhibits the soft tissues structure (Fig. 6b). Fig. 6c–i show the results obtained by NSCT1, NSCT2, SR1, SR2, SOMP1, SOMP2, and the proposed method, respectively. As can be seen, the results of NSCT1, SR1 and SOMP1 have blurred edges (Fig. 6c, e and g); while jaggy artifacts are observed in the results of NSCT2, SR2 and SOMP2 (Fig. 6d, f and h). In comparison, the proposed method (Fig. 6i) produces sharper edges, and details of the bone and tissues are better preserved.

The second experiment is performed on the fusion of the (low-resolution) MR-T1 and MR-T2 images of the brain (Fig. 7a and b). These images are generated along the parallel and perpendicular axes, respectively, and contain information in the horizontal and vertical directions. Again, the results of NSCT1, SR1 and SOMP1 (Fig. 7c, e and g) exhibit blurred edges, as Bicubic interpolation leads to a loss in high-frequency details. In comparison, the results

¹ Fig. 5a, b, and e–h are downloaded from an image fusion website [26]. Fig. 5c and d are downloaded from Brainweb [27].

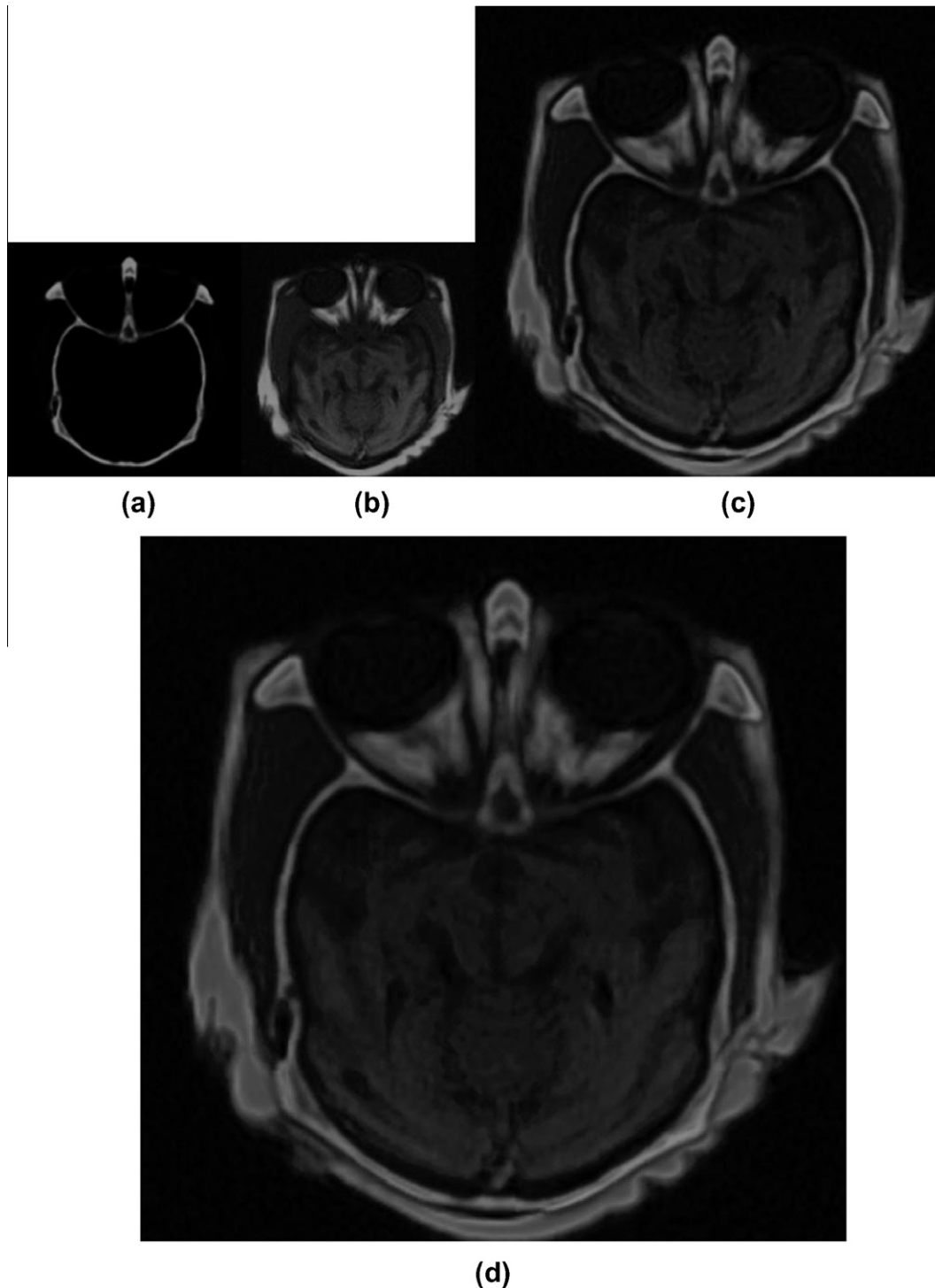


Fig. 14. Real source images and the results obtained by the proposed method for different magnification factors: (a) Real CT image; (b) real MR image; (c) result with magnification factor 2; and (d) result with magnification factor 3.

of NSCT2, SR2 and SOMP2 (Fig. 7d, f and h) show clearer skulls and marrows. This is consistent with the results in [15] that image super-resolution based on sparse representation can generate sharper edges than Bicubic interpolation. However, artifacts generated in the image fusion step are magnified during super-resolution. Thus, artifacts are introduced in the gray matter and white matter for the results of NSCT2, SR2 and SOMP2 (Fig. 7d, f and h). On the other hand, the proposed method (Fig. 7i) exhibits the best visual quality in terms of edge preservation of the brain structure.

The third experiment is performed on the fusion of two low-resolution remote sensing images (Fig. 8a–b). They exhibit obvious complementary information, especially in the regions containing the buildings. As can be seen, the regions of buildings in the results of NSCT1, SR1 and SOMP1 (Fig. 8c, e and g) are blurred. NSCT2, SR2 and SOMP2 can preserve the complementary information of Fig. 8a and b with some details enhanced, but introduce undesired artifacts (Fig. 8d, f and h). Overall, as can be seen more clearly in the magnified panels, the proposed method (Fig. 8i) provides clearer edges without blur and artifacts.

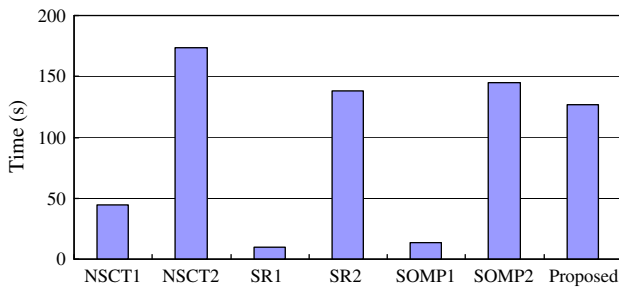


Fig. 15. Time comparison of the various methods.

The forth experiment is performed on the fusion of a low-resolution visible and infrared images. In the visible image (Fig. 9a), the background (such as the tree, house and road) is clear; while in the infrared image (Fig. 9b), the region containing the person is clear. As can be seen, the edges in the results of NSCT1, SR1, and SOMP1 (Fig. 9c, e and g) are blurred; whereas the results of NSCT2, SR2 and SOMP2 (Fig. 9d, f and h) are blocky, which can render further image analysis such as edge detection and target recognition difficult. In contrast, the result obtained by the proposed method (Fig. 9i) has much better visual quality.

Finally, Table 2 shows a quantitative comparison in terms of Q_0 , Q_E , Q_W , and $Q^{AB/F}$. For the CT and MR image pair (Fig. 6), the proposed method performs best in terms of Q_E , Q_W , and $Q^{AB/F}$, and is only slightly outperformed by SR1 on Q_0 . For the MR-T1 and MR-T2 pair (Fig. 7), the proposed method performs best on all four metrics. For the remote sensing image pair (Fig. 8), the proposed method performs best except on Q_E and Q_W . For the visible and infrared image pair (Fig. 9), the proposed method performs best on three out of the four metrics. Hence, overall, the proposed method is superior than the other methods.

In the next set of experiments, we perform super-resolution with a factor of 3. Similar to the previous setup, the low-resolution source images are generated by downsampling the original high-resolution images in Fig. 5 by a factor of 3. Fig. 10–13 shows the images and results obtained by the various methods. As in the previous experiments, the proposed method often generates a high-resolution fused image with little artifacts.

Table 3 shows a quantitative comparison based on Q_0 , Q_E , Q_W , and $Q^{AB/F}$. For the medical images in Figs. 10 and 11, the proposed method outperforms the others on all four metrics. For the remote sensing image pair (Fig. 12), the proposed method performs the best in terms of Q_0 , Q_E and $Q^{AB/F}$, and is only slightly outperformed by SR2 on Q_W . For the visible and infrared image pair (Fig. 13), the proposed method performs best on Q_E , Q_W , and $Q^{AB/F}$. Overall, as in the previous set of experiments, the proposed method is better than the other methods.

4.3. Experimental results on natural low-resolution images

Finally, experiment is performed on a pair of naturally low-resolution CT and MR images (Fig. 14a and b). Fig. 14c and d show the results obtained by the proposed method, with a super-resolution magnification factor of 2 and 3, respectively. As can be seen, the fused image better preserves features in the source images, and little artifacts are introduced. The bone and tissue structures can both be clearly seen. This indicates that the proposed method can integrate complementary information from multiple source images, while at the same time enhances the image resolution effectively.

Fig. 15 shows the running time of the various methods, on the experiments with a super-resolution magnification factor of 2. As can be seen, NSCT1, SR1 and SOMP1 are fast, as Bicubic interpola-

tion is simple. However, as demonstrated before, their fusion performance is inferior. As for the fusion-super-resolution strategies based on sparse representation, the proposed method is the fastest.

5. Conclusions

In this paper, we perform image fusion and super-resolution simultaneously based on sparse representation. The proposed method has two advantages. First, it avoids the propagation of artifacts generated by image fusion or super-resolution as in the traditional two-stage process. Second, the computational complexity is lower than performing image fusion and super-resolution separately. Experiments on different types of source images demonstrate the superiority of our method over existing fusion-super-resolution strategies based on interpolation and sparse representation. In this paper, the sparsity assumption does not take the structure of the sparse signal into account. In the future, we will investigate the specific structures of different sparse signals and further improve the performance of the proposed method.

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their insightful comments and suggestions. This paper is supported by the National Natural Science Foundation of China (No. 61172161), the Key Project of Chinese Ministry of Education (2009-120), the Fundamental Research Funds for the Central Universities, Hunan University, the Scholarship Award for Excellent Doctoral Student granted by Chinese Ministry of Education and the Natural Science Foundation of Hunan Province, China (No. 10JJ6096). The authors would like to thank Professor James T. Kwok for improving the English usage.

References

- [1] A.A. Goshtasby, S. Nikolov, Image fusion: advances in the state of the art, *Information Fusion* 8 (2) (2007) 114–118.
- [2] B.V. Dasarthy, Information fusion in the realm of medical applications – a bibliographic glimpse at its growing appeal, *Information Fusion* 13 (1) (2012) 1–9.
- [3] G. Pajares, J.M. de la Cruz, A wavelet-based image fusion tutorial, *Pattern Recognition* 37 (9) (2004) 1855–1872.
- [4] J.J. Lewis, R.J. O’Callaghan, S.G. Nikolov, D.R. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, *Information Fusion* 8 (2) (2007) 119–130.
- [5] Q. Zhang, B. Guo, Multifocus image fusion using the nonsubsampling contourlet transform, *Signal Processing* 89 (7) (2009) 1334–1346.
- [6] T. Li, Y. Wang, Biological image fusion using a NSCT based variable-weight method, *Information Fusion* 12 (2) (2011) 85–92.
- [7] J. Hu, S. Li, The multiscale directional bilateral filter and its application to multisensor image fusion, *Information Fusion* 13 (3) (2012) 196–206.
- [8] S. Yang, M. Wang, L. Jiao, R. Wu, Z. Wang, Image fusion based on a new contourlet packet, *Information Fusion* 11 (2) (2010) 78–84.
- [9] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, *IEEE Transactions on Instrumentation and Measurement* 59 (4) (2010) 884–892.
- [10] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, *Information Fusion* 13 (1) (2012) 10–19.
- [11] L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, *IEEE Transactions on Image Processing* 15 (8) (2006) 2226–2238.
- [12] W.T. Freeman, E.C. Pasztor, O.T. Carmichael, Learning low-level vision, *International Journal of Computer Vision* 40 (1) (2000) 25–47.
- [13] J. Sun, N.Ni. Zheng, H. Tao, H.Y. Shum, Image hallucination with primal sketch priors, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 729–736.
- [14] H. Chang, D.Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, pp. 275–282.
- [15] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, *IEEE Transactions on Image Processing* 19 (11) (2010) 2861–2873.

- [16] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, first ed., Springer, New York, 2010.
- [17] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solution of systems of equations to sparse modeling of signals and images, *SIAM Review* 51 (1) (2009) 34–81.
- [18] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM Review* 43 (1) (2001) 129–159.
- [19] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing* 41 (12) (1993) 3397–3415.
- [20] J.L. Starck, M. Elad, D.L. Donoho, Image decomposition via the combination of sparse representations and a variational approach, *IEEE Transactions on Image Processing* 14 (10) (2005) 1570–1582.
- [21] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Transactions on Signal Processing* 54 (11) (2006) 4311–4322.
- [22] Kodak Website, <<http://r0k.us/graphics/kodak/>>.
- [23] Z. Wang, A.C. Bovik, A universal image quality index, *IEEE Signal Processing Letters* 9 (3) (2002) 81–84.
- [24] G. Piella, H. Heijmans, A new quality metric for image fusion, In: *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, 2003, pp. 173–176.
- [25] C.S. Xydeas, V. Petrovic, Objective image fusion performance measure, *Electronics Letters* 36 (4) (2000) 308–309.
- [26] Image fusion Website, <www.imagefusion.org>.
- [27] Brainweb, <<http://mouldy.bic.mni.mcgill.ca/brainweb/>>.