Optical Engineering

SPIEDigitalLibrary.org/oe

Multimodal image fusion with joint sparsity model

Haitao Yin
Shutao Li

Multimodal image fusion with joint sparsity model

Haitao Yin
Shutao Li
Hunan University
College of Electrical and Information Engineering
Changsha, 410082, China
E-mail: shutao_li@yahoo.com.cn

Abstract. Image fusion combines multiple images of the same scene into a single image which is suitable for human perception and practical applications. Different images of the same scene can be viewed as an ensemble of intercorrelated images. This paper proposes a novel multimodal image fusion scheme based on the joint sparsity model which is derived from the distributed compressed sensing. First, the source images are jointly sparsely represented as common and innovation components using an over-complete dictionary. Second, the common and innovations sparse coefficients are combined as the jointly sparse coefficients of the fused image. Finally, the fused result is reconstructed from the obtained sparse coefficients. Furthermore, the proposed method is compared with some popular image fusion methods, such as multiscale transform-based methods and simultaneous orthogonal matching pursuit-based method. The experimental results demonstrate the effectiveness of the proposed method in terms of visual effect and quantitative fusion evaluation indexes.
[DOI: 10.1117/1.3584840]

Subject terms: image fusion; sparse representation; joint sparsity model; multiscale transform.

Paper 100908RR received Nov. 3, 2010; revised manuscript received Apr. 10, 2011; accepted for publication Apr. 12, 2011; published online Jun. 1, 2011.

1 Introduction

Recently, with extraordinary advances in sensor technology, numerous imaging sensors have been developed in military and civilian applications. The images provided by different sensors of one scenario often present complementary information. An image fusion technique can integrate information from different sensors into a single image. The fused image can preserve the relevant information and reduce the uncertainty and redundancy. Compared with the image provided by the individual sensor, the fused image has several benefits: broadened the spatial and temporal resolution, improved reliability, and increased robustness.[1–3]

The fusion of multimodal images has attracted more attention due to the increasing demands of surveillance purposes and clinical applications.[4] Each image provided by different modal imaging sensors has different features. For example, in the surveillance area, the discernible background of a scene can be provided in the visible image, while some special objects appear in the infrared image. Combining the visible and infrared images, the fused image can predicate the localization of dangerous objects with respect to the background.[5] In clinical application, MR-T1 gives the details of anatomical structures, while MR-T2 shows the contrast between normal and abnormal tissues. Fusing the MR-T1 and MR-T2 images is helpful for diagnosing diseases and reducing the storage cost.[6]

Among the presented pixel level fusion methods, multiscale transform-based image fusion methods are popularly used, including the discrete wavelet transform (DWT),[7] stationary wavelet transform (SWT),[8] and dual-tree complex wavelet transform (DTCWT).[9] Recently developed curvelet transform,[10] ridgelet transform,[11] contourlet transform,[12] and the nonsubsampled contourlet transform (NSCT)[13] can capture geometrical structure of images which are also used to image fusion. The region based fusion methods are also effective methods which can reduce the effect of noise, blurring effects, and misregistration. A novel region based image fusion method was proposed in Ref. 14 based on the high boost filtering, which can extract accurate segmentation. In Ref. 15, a nonparametric and region-based image fusion method was presented using the Bootstrap sampling principle, which reduces the dependence effect of pixels. The basic assumption of the multiscale transform based image fusion methods is that the salient features of the source images are linked to the coefficients of multiscale decomposition.[16] Two factors affect the performance of the multiscale transform-based fusion methods: the type of multiscale decomposition and the decomposition levels. For example, a large decomposition level may appear necessary. But, an overlarge decomposition level may bring about the block effects. Up to now, the decomposition levels are chosen through the experience. To effectively extract the underlying information of the source images, the sparse representation-based image fusion method is designed in Refs. 17 and 18. Sparse representation assumes that the signals can be represented as a linear combination of given atoms.[19–21] These atoms consist of an over-complete dictionary. Through sparse coding and the over-complete dictionary, the source images have a one-to-one correspondence with the sparse coefficients. Thus, in the sparse representation-based image fusion method, the sparse coefficients are treated as the salient features of the source images. Sparse coding guarantees the robustness of the sparse representation-based image fusion method. The orthogonal matching pursuit (OMP) is used to solve the sparse representation problem in Ref. 17. Different from OMP, the simultaneous OMP (SOMP) can address the simultaneous sparse approximation problem that sparsely decomposes multiple input signals at once. The scheme stated in Ref. 18 adopts the SOMP to solve the sparse representation problem.

0091-3286/2011/$25.00 © 2011 SPIE

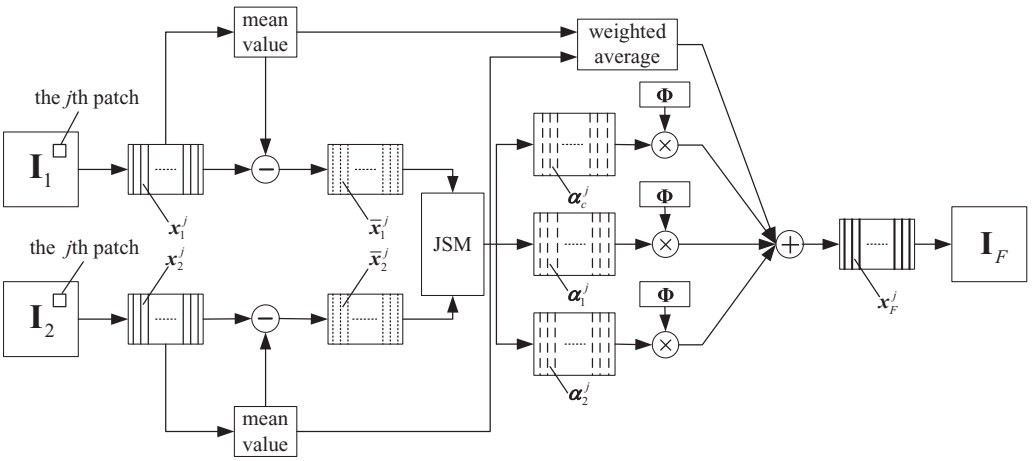

Fig. 1 Diagram of JSM-based image fusion method.

As we know, the difficulty of image fusion is how to separate the complementary information among the source images. The major contribution of this paper is to propose a novel image fusion scheme based on the joint sparsity model (JSM), which can separate the complementary information effectively. Based on the distributed compressed sensing theory,[22] different images from the various sensors of the same scene can be grossly sparsely represented as two features: the common features and the innovations features. We assume that the sparse coefficients are the salient information.[17,18] The innovation features separated through the JSM are the complementary information of different source images.

This paper is organized as follows. In Sec. 2, we briefly review the basic theory of sparse representation and JSM. An image fusion scheme based on JSM is designed in Sec. 3. Experimental results and comparisons are presented in Sec. 4. We conclude with discussions in Sec. 5.

2 Sparse Representation and Joint Sparsity Model

2.1 *Sparse Representation*

Sparse representation is a powerful tool for signal representation which has been applied to many image processing areas, such as denoising,[20] super-resolution,[23] text detection,[24] etc. To be more precise, let $\mathbf{\Phi} \in \mathbb{R}^{n\times m}$ $(n < m)$ denote the over-complete dictionary. The columns $\{\boldsymbol{\phi}_i\}_{i=1}^{m}$ of dictionary $\mathbf{\Phi}$ are regarded as the atoms. Based on the theory of sparse representation, every signal vector $\boldsymbol{x} \in \mathbb{R}^n$ can be represented as a sparse linear combination of these atoms. That is, the signal $\boldsymbol{x}$ can be calculated as $\boldsymbol{x} = \mathbf{\Phi}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^m$is a sparse coefficient vector which has small nonzero entries. Due to the dictionary $\mathbf{\Phi}$ with more columns than rows, the equation $\boldsymbol{x} = \mathbf{\Phi}\boldsymbol{\alpha}$ is under-determined. So, there are infinite solutions satisfying $\boldsymbol{x} = \mathbf{\Phi}\boldsymbol{\alpha}$. However, under sparsity criteria, the sparsest solution α^* is well defined which is the solution of the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \|\boldsymbol{x} - \mathbf{\Phi}\boldsymbol{\alpha}\|_2 \le \varepsilon, \tag{1}$$

where $\|\cdot\|_0$ denotes the ℓ_0 norm counting the number of nonzero elements, and $\varepsilon \ge 0$ is the error tolerance. The fundamental ingredient is the dictionary $\mathbf{\Phi}$ which is often chosen as the discrete cosine transform, wavelets, and learning from the training database,[21] etc.

Unfortunately, problem (1) is a combinatorial problem, which is NP-hard in general. Some suboptimal algorithms are used to solve problem (1) in practice, such as, greedy algorithm,[25] nonconvex relaxation,[26] and convex relaxation.[27] Under this certain condition, these suboptimal algorithms are equivalent to the original problem (1). Due to faster computation speed and lower complexity, the greedy algorithms are widely studied.

2.2 *Joint Sparsity Model*

In this sub-section, joint sparsity of an ensemble under the same basis and the related JSM are reviewed. The joint sparsity model indicates that different signals from the various sensors of the same scene form an ensemble. All signals in one ensemble have a common sparse component, and each individual signal owns an innovation sparse component.[22] We use $\mathbf{\Gamma} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\Lambda\}$ to denote an ensemble of signals, where $\boldsymbol{x}_i \in \mathbb{R}^n$, $i = 1, 2, \ldots, \Lambda$. In JSM, the signals in an ensemble can be sparsely represented as

$$\boldsymbol{x}_i = \mathbf{\Phi}\boldsymbol{\alpha}_c + \mathbf{\Phi}\boldsymbol{\alpha}_i, \quad i = 1, 2, \ldots, \Lambda, \tag{2}$$

where $\mathbf{\Phi} \in \mathbb{R}^{n\times m}$ $(n < m)$ is an over-complete dictionary, $\boldsymbol{\alpha}_c \in \mathbb{R}^m$ is the common sparse coefficients for all signals, and $\boldsymbol{\alpha}_i \in \mathbb{R}^m$ denotes the innovation sparse coefficients for the i'th individual signal. Furthermore, the ensemble and its jointly sparse representation can be formulated as

$$\boldsymbol{X} = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_\Lambda^T\right]^T \in \mathbb{R}^{\Lambda n},$$

$$\boldsymbol{X} = \mathbf{D}\Theta,$$

$$\mathbf{D} = \begin{pmatrix} \mathbf{\Phi} & \mathbf{\Phi} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Phi} & \mathbf{0} & \mathbf{\Phi} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Phi} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Phi} \end{pmatrix} \in \mathbb{R}^{\Lambda n\times(\Lambda+1)m},$$

where $\mathbf{0} \in \mathbb{R}^{n\times m}$ is a zero matrix, and the sparse coefficients vector is $\Theta = [\boldsymbol{\alpha}_c^T, \boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \ldots, \boldsymbol{\alpha}_\Lambda^T]^T \in \mathbb{R}^{(\Lambda+1)m}$.

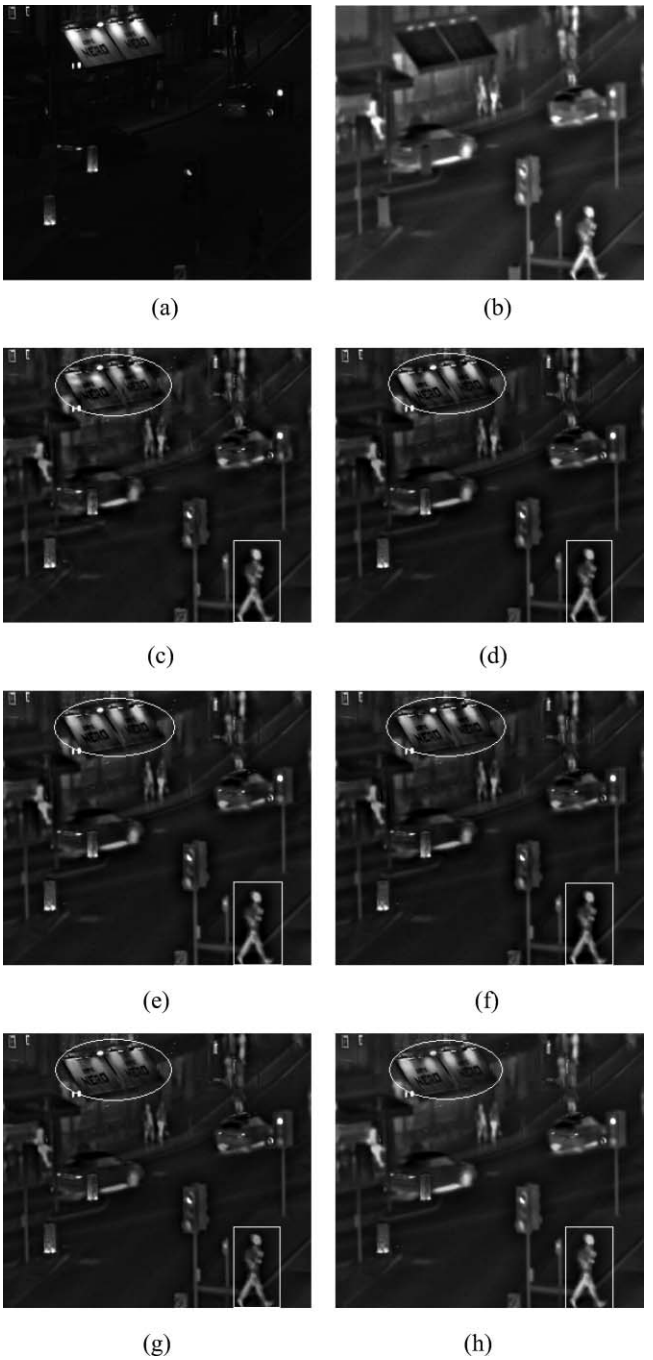

Fig. 2 The Bristol Queen's road source images and fused images of different fusion algorithms. (a) Visible image; (b) infrared image; (c) fused image of DWT; (d) fused image of NSCT; (e) fused image of SWT; (f) fused image of DTCWT; (g) fused image of SOMP; (h) fused image of JSM.

Combining the theory of sparse representation, the sparsest vector Θ^* is the minimum of the following optimization problem:

$$\min_{\Theta} \|\Theta\|_0 \quad \text{subject to} \quad \|\boldsymbol{X} - \mathbf{D}\Theta\|_2 \leq \varepsilon, \tag{3}$$

where $\varepsilon \geq 0$ is the error tolerance. Many effective algorithms can solve problem (3). In this paper, we apply the OMP algorithm[25] to solve it. The main steps of OMP are summarized as follows:

Table 1 The objective evaluation of various methods for the Bristol Queen's road images.

Methods	MI	Q_E	Q_W	Q_0	$Q^{AB/F}$	Entropy
DWT	2.1241	0.5912	0.7640	0.5945	0.5545	6.4726
NSCT	2.1572	0.6414	0.8008	0.6896	0.6525	6.4610
SWT	2.1532	0.6342	0.7927	0.6745	0.6408	6.4601
DTCWT	2.1392	0.6258	0.7849	0.6551	0.6127	6.4249
SOMP	2.1680	0.5733	0.7391	0.6576	0.6188	6.2408
JSM	**2.3228**	**0.7080**	**0.8573**	**0.7691**	**0.6702**	**6.6806**

Algorithm: OMP

Input: a signal $\boldsymbol{X} \in \mathbb{R}^{\Lambda n}$ and a matrix $\mathbf{D} \in \mathbb{R}^{\Lambda n \times (\Lambda+1)m}$

Output: a sparse coefficient vector $\Theta \in \mathbb{R}^{(\Lambda+1)m}$

Initialize: $k = 0$, $\Theta^0 = 0$, the residual $\boldsymbol{r}^0 = \boldsymbol{X} - \mathbf{D}\Theta^0 = \boldsymbol{X}$, and the index set $\Omega^0 = \varnothing$.

Iterate: increment k by 1 and repeat the following steps:

1) Calculate $\boldsymbol{g}^k = \mathbf{D}^T \boldsymbol{r}^{k-1}$, find one index i^k corresponding to the highest magnitude of $\boldsymbol{g}^k$, and update the index set $\Omega^k = \Omega^{k-1} \bigcup \{i^k\}$.
2) Compute the new estimate Θ^k, such that $\Theta^k = \arg\min_{\Theta} \left\| \boldsymbol{X} - \mathbf{D}_{\Omega^k}\Theta \right\|_2^2$, where $\mathbf{D}_{\Omega^k}$ is submatrix of $\mathbf{D}$ restricted in the set Ω^k.
3) Update the residual $\boldsymbol{r}^k = \boldsymbol{X} - \mathbf{D}\Theta^k$.
4) If the stopping criterion holds, output the solution Θ^k.

3 Proposed Image Fusion Scheme

In this section, before presentation of the proposed image fusion scheme, we first introduce preprocessing of the source images and joint sparsity used in the proposed method.

3.1 *Joint Sparsity Model for Image Fusion*

Since pixel level image fusion depends on the local information of source images, we consider the image patches, and the sliding window technique is used to traverse the whole images from left top to right bottom. Let the size of the image patch be $\sqrt{n} \times \sqrt{n}$, and the j'th patch is ordered lexicographically as a column vector $\boldsymbol{x}_i^j \in \mathbb{R}^n$, where $i \in \{1, 2, \ldots, \Lambda\}$denotes the i'th source image, and Λ is the number of source images. Then, $\boldsymbol{\Gamma}_j = \{\boldsymbol{x}_1^j, \boldsymbol{x}_2^j, \ldots, \boldsymbol{x}_\Lambda^j\}$ consists of an ensemble. Given an over-complete dictionary $\boldsymbol{\Phi} \in \mathbb{R}^{n \times m}$ $(n < m)$, i'th signal of j'th ensemble can be represented as

$$\boldsymbol{x}_i^j = \boldsymbol{\Phi}\boldsymbol{\alpha}_c^j + \boldsymbol{\Phi}\boldsymbol{\alpha}_i^j, \tag{4}$$

where $\boldsymbol{\alpha}_c^j$, $\boldsymbol{\alpha}_i^j \in \mathbb{R}^m$are the sparse coefficients. Thus, $\boldsymbol{\Phi}\boldsymbol{\alpha}_i^j$ is the complementary information. Using the same analysis of JSM presented in Sec. 2.2, the j'th ensemble $\boldsymbol{\Gamma}_j$ can be rewritten as

$$\boldsymbol{X}^j = \mathbf{D}\Theta^j, \tag{5}$$

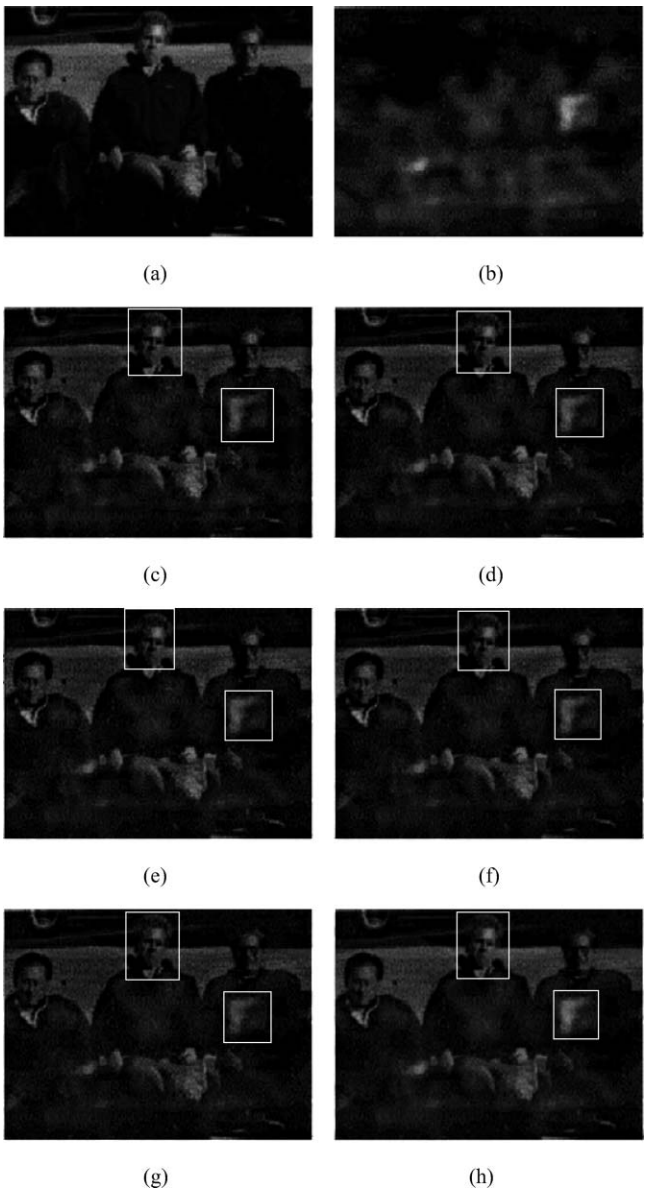

Fig. 3 The visual and MMW source images and fused images of different fusion algorithms. (a) Visible image; (b) MMW image; (c) fused image of DWT; (d) fused image of NSCT; (e) fused image of SWT; (f) fused image of DTCWT; (g) fused image of SOMP; (h) fused image of JSM.

where

$$\boldsymbol{X}^j = \left[\left(\boldsymbol{x}_1^j\right)^T, \left(\boldsymbol{x}_2^j\right)^T, \ldots, \left(\boldsymbol{x}_\Lambda^j\right)^T \right]^T \in \mathbb{R}^{\Lambda n},$$

$$\mathbf{D} = \begin{pmatrix} \boldsymbol{\Phi} & \boldsymbol{\Phi} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\Phi} & \mathbf{0} & \boldsymbol{\Phi} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Phi} \end{pmatrix} \in \mathbb{R}^{\Lambda n \times (\Lambda+1)m},$$

$$\Theta^j = \left[\left(\boldsymbol{\alpha}_c^j\right)^T, \left(\boldsymbol{\alpha}_1^j\right)^T, \left(\boldsymbol{\alpha}_2^j\right)^T, \ldots, \left(\boldsymbol{\alpha}_\Lambda^j\right)^T \right]^T \in \mathbb{R}^{(\Lambda+1)m}.$$

Table 2 The objective evaluation of various methods for the visual and MMW images.

Methods	MI	Q_E	Q_W	Q_0	$Q^{AB/F}$	Entropy
DWT	2.2246	0.6746	0.7673	0.5981	0.5578	5.9863
NSCT	2.2530	0.7587	0.8429	0.6425	0.6480	5.9665
SWT	2.2564	0.7215	0.8230	0.6448	0.6406	5.9781
DTCWT	2.2432	0.7415	0.8316	0.6391	0.6180	6.0084
SOMP	2.2716	0.7439	0.8438	0.6769	0.6462	6.0315
JSM	**2.3712**	**0.7960**	**0.8956**	**0.8039**	**0.6866**	**6.5831**

If the source images are contaminated by noise, the formulation of the j'th ensemble $\boldsymbol{\Gamma}_j$ should be slightly modified as

$$\boldsymbol{X}^j = \mathbf{D}\Theta^j + n, \tag{6}$$

where n is assumed as the unknown noise. The sparsest vector Θ^j can be calculated through the optimization problem (3) with appropriate error tolerance ε. In this paper, we assume that the additional noise is the zero mean Gaussian noise with standard deviation σ. Then, ε can be set as $\sqrt{\Lambda n} \cdot C \cdot \sigma$,[28] where $C > 0$ is a constant. The parameter C was chosen as 1.15 empirically for the image patch with size 8×8 in Ref. 20.

3.2 *Proposed Method*

Without the loss of generality, we assume that there are two gray source images $\mathbf{I}_1$, $\mathbf{I}_2 \in \mathbb{R}^{N\times M}$, which are geometrically registered. The diagram of the proposed JSM-based image fusion method is depicted in Fig. 1. The detailed steps are summarized as follows:

1. Using the sliding window technique, the source images $\mathbf{I}_1$, $\mathbf{I}_2$ are divided into patches with size $\sqrt{n}\times\sqrt{n}$, and all the patches are lexicographically ordered as n-dimensional column vectors. There are $\left(N-\sqrt{n}+1\right)\times\left(M-\sqrt{n}+1\right)$ patches for each source image.
2. For the j'th patches $\boldsymbol{x}_i^j$ $(i=1,2)$ of the two source images, first the mean values m_i^j $(i=1,2)$ of the

Table 3 The objective evaluation of various methods for the MR-T1 and MR-T2 images.

Methods	MI	Q_E	Q_W	Q_0	$Q^{AB/F}$	Entropy
DWT	2.3856	0.3736	0.5433	0.3521	0.5281	6.7987
NSCT	2.4656	**0.4621**	0.6368	0.5033	0.6552	6.6011
SWT	2.4515	0.4513	0.6106	0.4377	0.6468	6.6087
DTCWT	2.4105	0.4526	0.6222	0.4190	0.6130	6.7342
SOMP	2.4890	0.4575	0.6399	0.5017	0.6527	6.5337
JSM	**2.5604**	0.4606	**0.6794**	**0.5080**	**0.6569**	**7.0341**

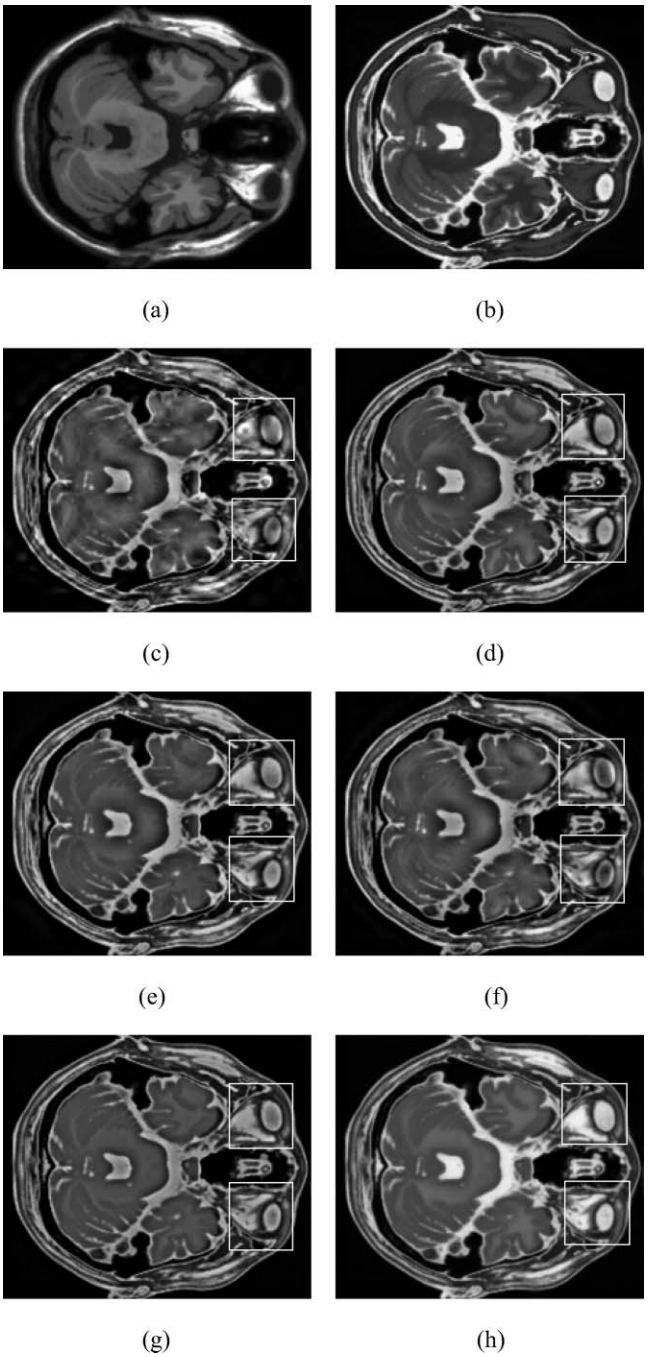

Fig. 4 The MR-T1 and MR-T2 source images and fused images of different fusion algorithms. (a) MR-T1; (b) MR-T2; (c) fused image of DWT; (d) fused image of NSCT; (e) fused image of SWT; (f) fused image of DTCWT; (g) fused image of SOMP; (h) fused image of JSM.

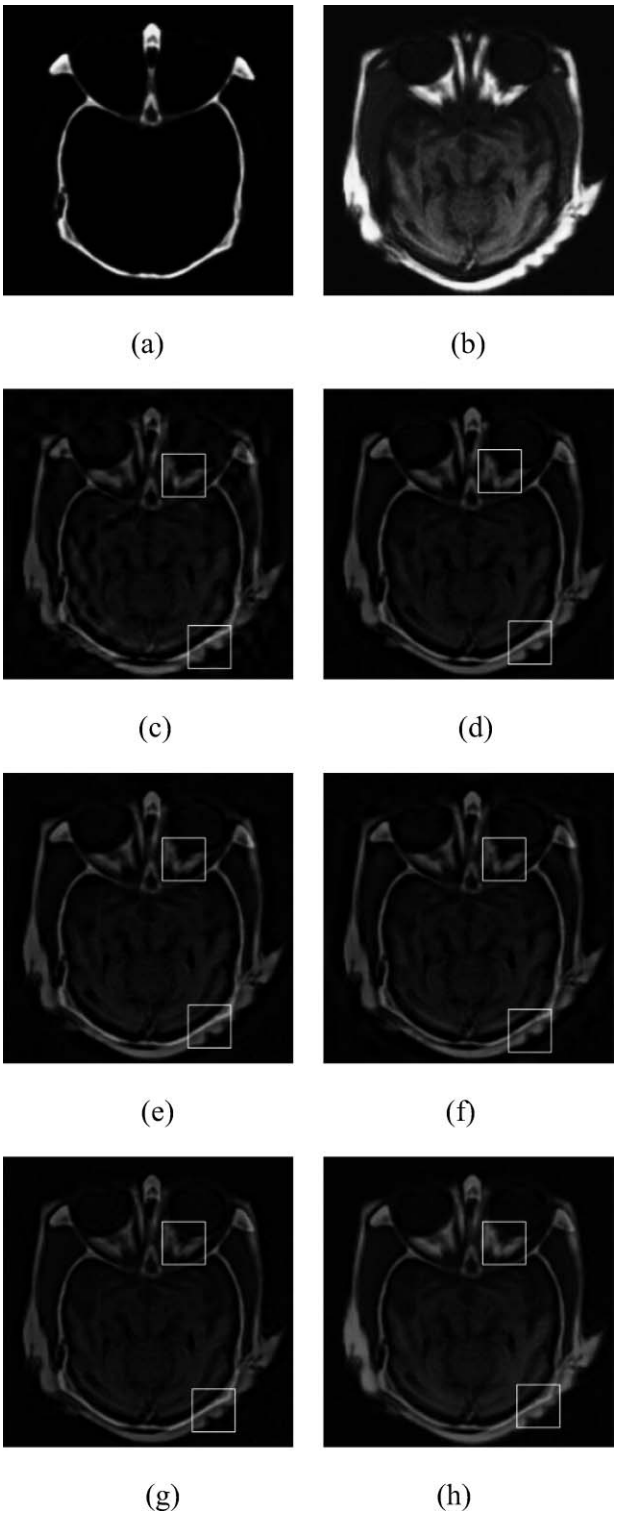

Fig. 5 The CT and MR source images and fused images of different fusion algorithms. (a) CT; (b) MR; (c) fused image of DWT; (d) fused image of NSCT; (e) fused image of SWT; (f) fused image of DTCWT; (g) fused image of SOMP; (h) fused image of JSM.

patches are subtracted. $\bar{\boldsymbol{x}}_i^j \ \ (i=1,2)$ are obtained. Then, by optimization problem (3), $\bar{\boldsymbol{x}}_i^j \ \ (i=1,2)$ are jointly sparsely represented as the common part $\boldsymbol{\alpha}_c^j$ and two innovation parts $\boldsymbol{\alpha}_1^j,\ \boldsymbol{\alpha}_2^j$.

3. Combining the given over-complete dictionary $\boldsymbol{\Phi}$, the j'th patch of the fused image $\mathbf{I}_F$ is obtained through

$$\boldsymbol{x}_F^j = \boldsymbol{\Phi}\boldsymbol{\alpha}_c^j + \boldsymbol{\Phi}\boldsymbol{\alpha}_1^j + \boldsymbol{\Phi}\boldsymbol{\alpha}_2^j + \tau m_1^j + (1-\tau)\, m_2^j, \qquad (7)$$

where $\tau = 1/(1+\exp\{-\beta(\|\bar{\boldsymbol{x}}_1^j\|_2 - \|\bar{\boldsymbol{x}}_2^j\|_2)\}),\ (\beta > 0)$ is the parameter for fusing the mean values.

4. At last, each patch $\boldsymbol{x}_F^j$ is reshaped as a block with size $\sqrt{n}\times\sqrt{n}$. Then, these blocks are embedded into the image $\mathbf{I}_F$ at the designated location. Due to the

Table 4 The objective evaluation of various methods for the CT and MR images.

Methods	MI	Q_E	Q_W	Q_0	$Q^{AB/F}$	Entropy
DWT	2.1598	0.4025	0.6257	0.4301	0.5884	5.4903
NSCT	2.2369	0.5257	0.7208	0.6334	0.7441	5.3722
SWT	2.2338	0.5278	0.7113	0.6261	0.7342	5.3991
DTCWT	2.1920	0.4720	0.6779	0.5412	0.6733	5.4827
SOMP	2.2955	0.4916	0.6838	0.6594	0.7366	5.3832
JSM	**2.3878**	**0.5770**	**0.7962**	**0.7748**	**0.7699**	**5.6389**

sliding technique, several block values constitute one pixel value. The average processing is applied to each pixel location and the final fused image $\mathbf{I}_F$ is obtained.

When the source image is a color image, we first apply the intensity-hue-saturation (IHS) transform[29] to convert the color image from the RGB space into the intensity (I), hue (H), and saturation (S) space. Then, the intensity components are fused by the proposed method. Ultimately, the final RGB image is obtained by the inverse IHS transform of the fused intensity, the hue, and saturation components.

For the noisy source images, the proposed method can perform image denoising and fusion simultaneously. Following the procedures of the proposed method, the stopping error of OMP is set as $\sqrt{2n} \cdot C \cdot \sigma$, according to the analyses presented in Sec. 3.1.

4 Experiments

In this section, experiments on several different category source images are presented for testing the proposed method. Because the parameters have an important influence on the performance of various tested image fusion methods, the parameters for different methods are firstly presented. Then, six evaluation criteria applied to assess the quality of the fused images are given. Finally, the performance of the JSM-based fusion method is demonstrated in comparison with the SOMP based fusion method and various popular multiscale transform-based methods including DWT, NSCT, SWT, and DTCWT.

Table 5 The statistical results of various methods for 20 pairs' source images.

Methods	MI	Q_E	Q_W	Q_0	$Q^{AB/F}$	Entropy
DWT	2.2533	0.5890	0.7630	0.6990	0.5639	**7.1529**
NSCT	2.2836	0.6424	0.8003	**0.7669**	0.6328	7.1010
SWT	2.2780	0.6371	0.7963	0.7529	0.6270	7.1206
DTCWT	2.2718	0.6369	0.7955	0.7494	0.6230	7.0989
SOMP	2.2914	0.6299	0.7924	0.7683	0.6201	7.0501
JSM	**2.3443**	**0.6517**	**0.8028**	0.7586	**0.6395**	7.0687

4.1 *Experimental Setting*

The over-complete dictionary $\mathbf{\Phi}$ is a fundamental ingredient for sparse representation. To better fit the sparity, the learning-based dictionary has the potential to outperform the predetermined dictionaries. We use the K-SVD algorithm[21] to train the dictionary through solving the below optimization problem

$$\min_{\mathbf{\Phi},\Xi} \|\mathbf{X} - \mathbf{\Phi}\Xi\|_F^2 \quad \text{subject to} \quad \forall i, \quad \|\boldsymbol{x}_i\|_0 \le T, \tag{8}$$

where $\mathbf{X} = \{\boldsymbol{x}_i\}_{i=1}^N$ is the patches' family from the image database, Ξ is a sparse coefficient matrix, and T is a natural number controlling the sparsity level. In this experiment, the miscellaneous package of the USC-SIPI image database[30] is used as the training image database, and T is set as 5. The trained over-complete dictionary is applied in both the SOMP- and JSM-based methods.

The patch size has an important influence on the performance of sparse representation. As a rule thumb, 8×8 is widely used in image denoising by sparse representation.[20] So, we also set the patch size at 8×8 in the JSM based image fusion method. Due to this patch size, the size of learned over-complete dictionary $\mathbf{\Phi}$ is 64×256. In the following experiments, two source images are used. So, the size of matrix $\mathbf{D}$ is 128×768. For the OMP applied to solve the joint sparsity model (3), the stopping residual error must be fixed beforehand. We set this stopping error at 0.001 for the clean source images. The parameters of the SOMP-based image fusion method are all set as the default values (see Ref. 18) for the noise-free case, and the over-complete dictionary is the same as the dictionary used in the JSM-based method. In the noise case, the stopping errors of JSM and SOMP based methods are set as $1.15\times\sqrt{128}\sigma$,[20] where $\sigma \ge 0$ is the noise intensity. The type of filter and the number of decomposition levels are two crucial parameters for the multiscale transform based methods. Next, these two parameters are presented for every tested multiscale transform-based method. For the DWT- and SWT-based methods, the wavelet basis Biorthogonal "bior(2,2)" is used. The "Near-Symmetric 13-19 tap filter" and "quarter sample shift orthogonal 18-18 tap filter" are employed as the first level filter and other levels filters, respectively, for the DTCWT-based method. The number of decomposition levels for DWT-, SWT-, and DTCWT-based methods are all four. For the NSCT based method, the pyramid filter is set as the "pyrexc" filter which is derived from 1D using the maximally flat mapping function with 2 vanishing moments, and the "vk" filter derived from the McClellan transform is applied as the directional filter. For the NSCT-based method, the number of directions for each level from coarse to fine is {4, 8, 8, 16}. According to the comparison results presented in Ref. 31, the selection of filters above and the decomposition levels are considered the best for each multi-scale transform based method. The fusion rule of multiscale transform-based methods applied in this paper is the max-abs rule.

In our experiments, the source images are registered. When the images are mis-registered, the registration techniques designed in Refs. 32 and 33 can be used.

4.2 *Evaluation Criteria*

Six evaluation criteria, i.e., mutual information (MI), Q_0, Q_W, Q_E, $Q^{AB/F}$, and entropy are considered to quantitatively

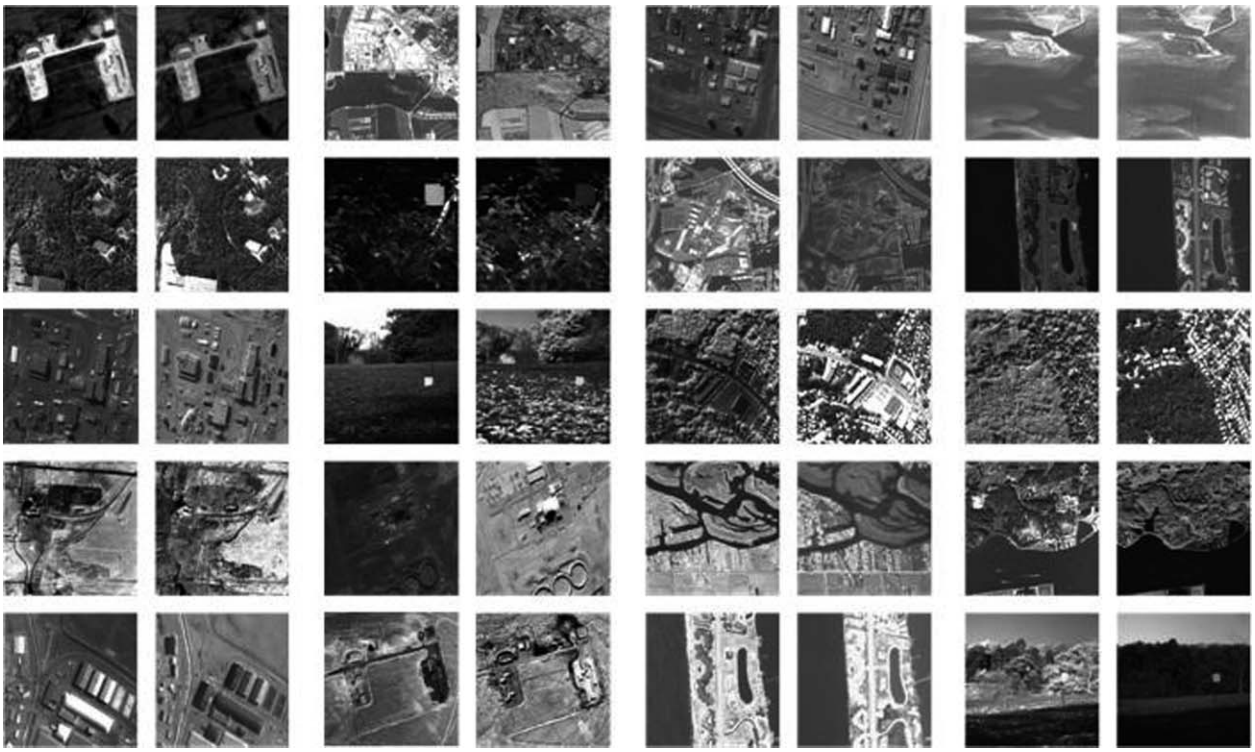

Fig. 6 20 pairs' source images.

evaluate the performance of the tested image fusion methods in this paper.

The mutual information (MI) (Ref. 34) determines the statistical information redundancy between two random variables. If we view the source images and the fused image as random variables, MI can reflect how much information from source images is transferred into the fused image. The Q_0 (Ref. 35) evaluates the distortion of the fused image by combining three factors: loss of correlation, luminance distortion, and contrast distortion. The Q_W and Q_E (Ref. 36) are defined based on the Q_0. The Q_W and Q_E compute how much of the salient information of source images is transferred into the fused image. The difference between Q_W and Q_E is: Q_W gives more weight at the windows with higher saliency, while Q_E is an edge-dependent fusion quality index. The $Q^{AB/F}$ (Ref. 37) evaluates the edge information transferred from the source images into the fused image by applying the Sobel edge operator to compute the edge strength and orientation information at each pixel. It should be as close to 1 as possible. Entropy measures the amount of information overall in the image. The larger value for the above six evaluation criteria implies the better fusion result.

4.3 *Experimental Results of Clean Images*

To illustrate the performance of the JSM-based method, some clean source images provided by various sensors are used to compare with the SOMP-, DWT-, NSCT-, SWT-, and DTCWT-based fusion methods.

Figures 2(a) and 2(b) show a pair of source images which are supplied by Lewis et al.[38] Figure 2(a) is the visible image of "Bristol Queen's road" providing the background of a road and the clear signboard of a coffee shop, while Fig. 2(b) is the infrared image which shows the outline of pedestrians and cars. Figures 2(c)–2(h) are the fused images by various tested methods. It is clear that Figs. 2(c)–2(g) have more black shadow over the signboard than Fig. 2(h); some black artifacts exist around the pedestrian for all the tested methods, but the range of Figs. 2(c)–2(g) is larger than Fig. 2(h). The results of the objective assessment are presented in Table 1. The bold is the best results of individual evaluation criteria. Obviously, our proposed method is superior to others for all six criteria.

Figures 3(a) and 3(b) are a pair of visual and 94 GHz millimeter wave (MMW) images respectively. The visual image shows the outline and the appearance of people, while the MMW image displays the existence of a gun. The fused images by various methods are depicted in Figs. 3(c)–3(h). From the fused results, there is a gun beneath the clothes of the people located on the right. The people's faces, located in the middle, and the gun of Fig. 3(h) are clearer than those in Figs. 3(c)–3(g). What is more, the information of the face is from the visual image, while the information of the gun derived from the MMW image. Intuitively, more information of the source images is transferred into the fused image by the JSM-based method than others. To evaluate this visual inspection objectively, the values of six evaluation criteria are listed in Table 2, where the highest MI value of the JSM based method is the effective evidence of visual inspection. For Q_E, Q_W, Q_0, $Q^{AB/F}$, and entropy, the JSM also have the best results.

Figure 4 illustrates the source images and the comparison of the fused results on a pair of MR-T1 and MR-T2 images. Figures 4(a) and 4(b) are the MR-T1 and MR-T2 source images, respectively. MR-T1 provides the longitudinal relaxation difference, while MR-T2 shows the transverse relaxation difference. The tissues of the result image, Fig. 4(h), are much clearer than the others. In Table 3, the values of the evaluation criteria of the six methods are presented.

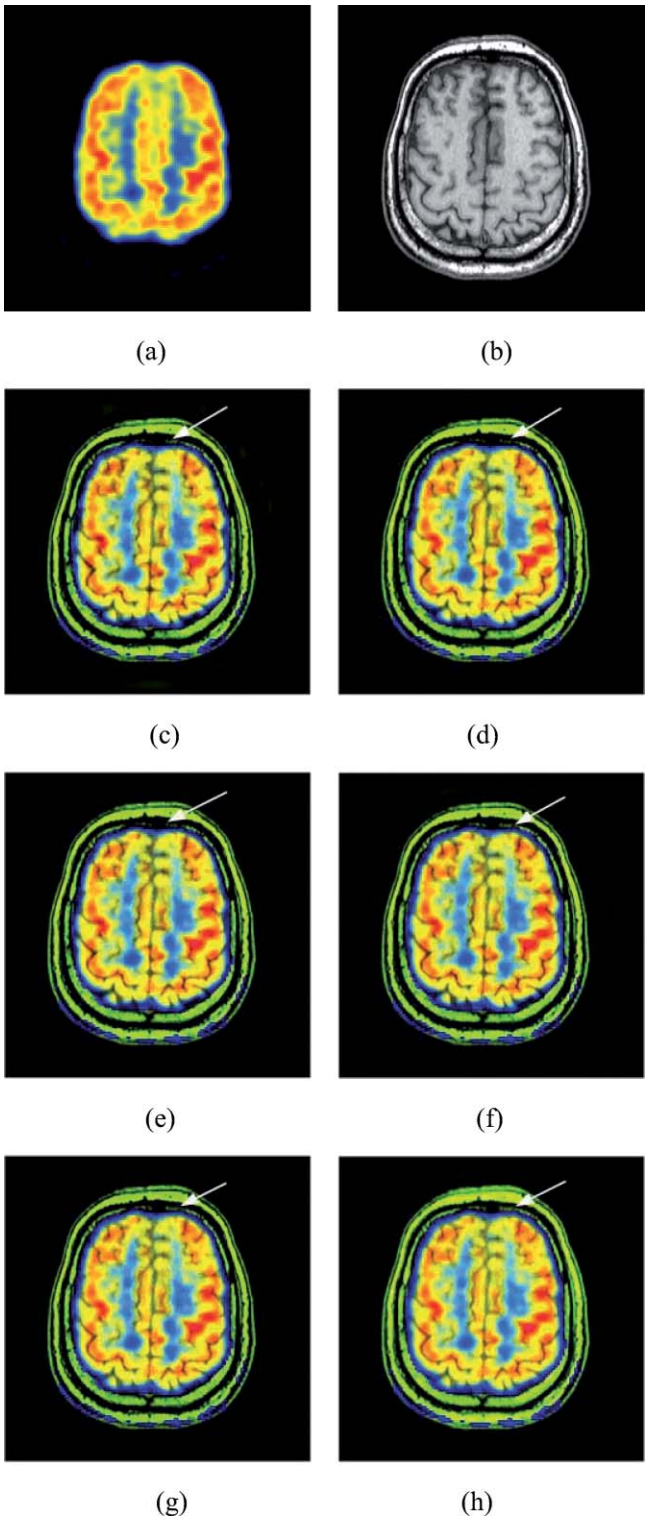

Fig. 7 The PET-FDG and MR-T1 source images and fused images of different fusion algorithms. (a) PET-FDG; (b) MR-T1; (c) fused image of DWT; (d) fused image of NSCT; (e) fused image of SWT; (f) fused image of DTCWT; (g) fused image of SOMP; (h) fused image of JSM.

From Table 3, we can note that the fusion performance of the JSM based method is better than others in terms of evaluation criteria including MI, Q_W, Q_0, and $Q^{AB/F}$, and entropy. As to the Q_E, the NSCT based method is somewhat better than our method.

Figure 5 shows the results of the CT and MR images. Figures 5(a) and 5(b) are the source images. The fused results of the proposed method and compared methods are presented in Figs. 5(c)–5(h). Careful inspection of the fused results reveal that the DWT-based method produces some artificial. The proposed method exhibits better visual quality with much clearer tissue and skeletal features than compared methods. The evaluation criteria of the tested methods are listed in Table 4. From Table 4, we can see that the proposed method obtains the best objective evaluations.

In order to demonstrate the universality of the proposed method, an experiment on larger image sets is presented. Figure 6 shows 20 pairs of source images. Each pair of source images are fused by the considered fusion methods, and the evaluation criteria of each pair experiment are calculated. Then, the statistical results of these 20 pairs' experiments are obtained through averaging. Table 5 exhibits the statistical results. From the Table 5, we observe that the JSM-based method achieves the best results in four of the six evaluation criterias, i.e., MI, Q_E, Q_W, and $Q^{AB/F}$. As to Q_0, the NSCT based method achieves a slightly better result than the JSM based method. The statistical results demonstrate the superiority of the proposed method.

Finally, a pair of PET-FDG and MR-T1 medical images is considered. Figure 7(a) is the PET-FDG image of a normal brain which provides the functional information with low spatial resolution, while Fig. 7(b) is the MR image of a normal brain which shows the high resolution of the structural information without the functional information. In this study, the PET-FDG image is treated as a color image. All the tested methods are carried out following the below processes: first, the PET-FDG image is transformed into the IHS space; next, fusing the intensity component and the MR-T1 image by various compared methods, a new intensity component is obtained; at last, the final fused results are acquired by the inverse IHS transform of the new intensity component and the hue and saturation components of the PET-FDG image into the RGB space. Figures 7(c)–7(h) are the fused results. At the target marked by an arrow, the JSM-based and the SOMP-based methods provide more details than the other methods. Table 6 lists the values of the six evaluation criteria for all compared methods. For the PET-FDG and the fused results being the color images, we first calculate the values of the evaluation criteria for divided R, G, and B space, then the evaluation indexes displayed in Table 6 are the average of the R, G, and B components. From the data in Table 6, we

Table 6 The objective evaluation of various methods for the PET-PDF and MR-T1 images.

Methods	MI	Q_E	Q_W	Q_0	$Q^{AB/F}$	Entropy
DWT	2.6339	0.3421	0.5720	0.2457	0.4328	0.6706
NSCT	2.6438	0.3589	0.5865	0.2526	0.4555	0.6740
SWT	2.6407	0.3557	0.5843	0.2529	0.4525	0.6740
DTCWT	2.6408	0.3488	0.5781	0.2507	0.4463	0.6743
SOMP	2.6625	0.3587	0.5874	0.2625	0.4675	0.6748
JSM	**2.6715**	**0.3752**	**0.6013**	**0.2630**	**0.4773**	**0.7018**

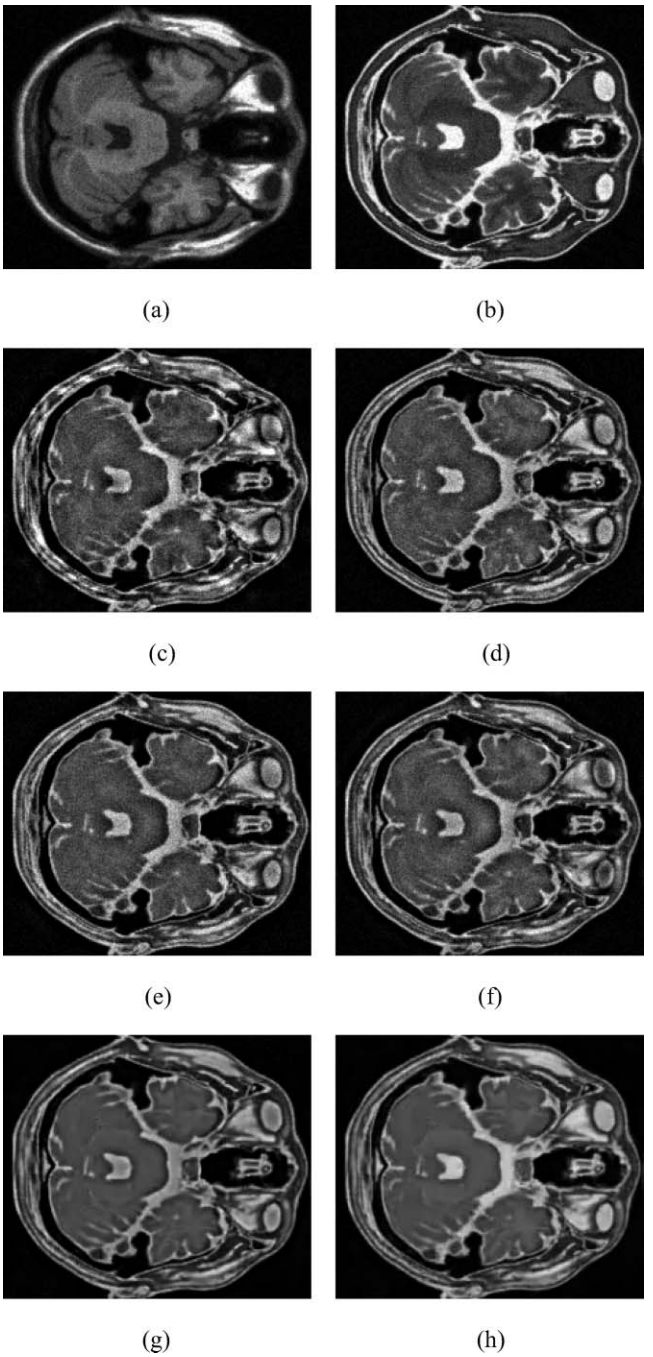

Fig. 8 The noisy MR-T1 and MR-T2 source images with $\sigma = 20$ and fused images of different fusion methods. (a) Noisy MR-T1; (b) noisy MR-T2; (c) fused image of DWT; (d) fused image of NSCT; (e) fused image of SWT; (f) fused image of DTCWT; (g) fused image of SOMP; (h) fused image of JSM.

can conclude that the JSM based method is obviously better than the other methods.

4.4 *Experimental Results of Noisy Images*

In this sub-section, the case of noisy source images is investigated. Figures 4(a) and 4(b) are corrupted by the Gaussian noise with zero mean and deviation $\sigma = 20$. The noisy images are depicted in Figs. 8(a) and 8(b). In our experiments, σ denotes the noise intensity. Figures 8(c)–8(h) present the fused results of the tested methods. Based on visual comparison, the fused images of JSM- and SOMP-based method have the least noise and the tissues of Fig. 8(h) are much clearer than the tissues of Fig. 8(g).

All the experiments are completed in the environment of a Pentium dual-core CPU 2.93 GHz with a 2.00 GB RAM PC, operating under MATLAB 7.10. It may take 4 min. for the source images with size 256×256. The proposed method takes more time than the multiscale transform-based methods. Since the patches (vectors) generated by the "sliding window" technique are independent, parallel processing with multicore could decrease running time with the development of hardware. The algorithm can be implemented in C++ with optimization to dramatically increase the speed.

5 Conclusions

In this paper, we have designed a multimodal image fusion method based on the joint sparsity model. The major contribution of this study is that the complementary information of the multimodal images monitoring the same scene can be effectively separated through the jointly sparse decomposition. To demonstrate the performance of the proposed method, experiments on several different category source images engaging in different application fields, such as surveillance, weapon detection, and medical diagnosis, are performed. The experimental results confirm that the proposed method exhibits better fusion results than the DWT-, NSCT-, SWT-, DTCWT-, and SOMP-based methods both in visual inspection and objective evaluation indexes. Note that the over-complete dictionary has the important influence in the efficiency of the sparse representation. The dictionary applied in this paper is trained off-line. The dictionary trained on-line is suitable for different source images and should be investigated further in our future works.

Acknowledgments

The authors would like to thank the anonymous reviewers and editor for their insightful comments and suggestions. This work is supported by the National Natural Science Foundation of China (No. 60871096), the Ph.D. Programs Foundation of Ministry of Education of China (No. 200805320006), the Key Project of Chinese Ministry of Education (2009-120), and the Open Projects Program of National Laboratory of Pattern Recognition.

References

1. Y. Q. Sun, J. W. Tian, and J. Liu, "Novel method on dual-band infrared image fusion for dim small target detection," *Opt. Eng.* **46**(11), 116402 (2007).
2. J. Lanir, M. Maltz, and S. R. Rotman, "Comparing multispectral image fusion methods for a target detection task," *Opt. Eng.* **46**(6), 066402 (2007).
3. Y. C. Tzeng and K. S. Chen, "Image fusion of synthetic aperture radar and optical data for terrain classification with a variance reduction technique," *Opt. Eng.* **44**(10), 106202 (2005).
4. A. A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Inf. Fusion* **8**(2), 114–118 (2007).
5. A. Toet and E. M. Franken, "Perceptual evaluation of different image fusion schemes," *Displays* **24**(1), 25–37 (2003).
6. W. Dou, S. Ruan, Y. Chen, D. Bloyet, and J. Constans, "A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images," *Image Vis. Comput.* **25**(2), 164–171 (2007).
7. H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.* **57**(3), 235–245 (1995).
8. S. Li, J. T. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images," *Inf. Fusion* **3**(1), 17–23 (2002).

9. J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah, "Region based image fusion using complex wavelets," in *Proc. 7th Int. Conf. on Information Fusion*, Stockholm, Sweden, pp. 555–562 (2004).
10. F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion* **8**(2), 143–156 (2007).
11. T. Chen, J. Zhang, and Y. Zhang, "Remote sensing image fusion based on ridgelet transform," in *Proc. Int. on Geoscience and Remote Sensing Symposium*, Seoul, Korea, pp. 1150–1153 (2005).
12. M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.* **14**(12), 2091–2106 (2005).
13. S. Li and B. Yang, "Hybrid multiresolution method for multisensor multimodal image fusion," *IEEE Sens. J.* **10**(9), 1519–1526 (2010).
14. T. Zaveri and M. Zaveri, "A novel region based image fusion method using highboost filtering," in *Proc. of Int. Conf. on Systems, Man, and Cybernetics*, San Antonio, TX, USA, pp. 966–971 (2009).
15. M. Zribi, "Nonparametric and region-based image fusion with Bootstrap sampling," *Inf. Fusion* **11**(2), 85–94 (2010).
16. G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.* **37**(9) 1855–1872 (2004).
17. B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.* **59**(4), 884–892 (2010).
18. B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*.
19. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse coding for natural images," Nature (London) 3**81**(6583), 607–609 (1996).
20. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006).
21. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006).
22. M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *39th Asilomar Conf. on Signals, Systems and Computer*, Pacific Grove, CA, pp. 1537–1541 (2005).
23. J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010).
24. M. Zhao, S. Li, and J. Kwok, "Text detection in images using sparse representation with discriminative dictionaries," *Image Vis. Comput.* **28**(12), 1590–1599 (2010).
25. S. G. Mallat and Z. Zhang, "Matching pursuits and time-frequency dictionaries," *IEEE Trans. Signal Process.* **41**(12), 3397–3415 (1993).
26. I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.* **45**(3), 600–616 (1997).
27. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.* **43**(1), 129–159 (2001).
28. J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.* **17**(1), 53–69 (2008).
29. T. M. Tu, S. C. Su, H. C. Shyu, and P. S. Huang, "Efficient intensity-hue-saturation-based image fusion with saturation compensation," *Opt. Eng.* **40**(5), 720–728 (2001).
30. http://sipi.usc.edu/database/database.cgi?volume=misc.
31. S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Inf. Fusion* **12**(2), 74–84 (2011).
32. E. Coiras, J. Santamaría, and C. Miravet, "Segment-based registration technique for visual-infrared images," *Opt. Eng.* **39**(1), 282–289 (2000).
33. X. Cai and W. Zhao, "Novel image registration method using edge correlation," *Opt. Eng.* **49**(1), 017003 (2010).
34. G. H. Qu, D. L. Zhang, and P. F. Yan, "Information measure for performance of image fusion," *Electron. Lett.* **38**(7), 313–315 (2002).
35. Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002).
36. G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. of Int. Conf. on Image Processing*, Barcelona, Spain, pp. 173–176 (2003).
37. C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.* **36**(4), 308–309 (2000).
38. J. J. Lewis, S. G. Nikolov, C. N. Canagarajah, D. R. Bull, and A. Toet, "Uni-modal versus joint segmentation for region-based image fusion," in *Proc. of the 9th Int. Conf. on Information Fusion*, Florence, Italy, pp. 10–13 (2006).

Haitao Yin received his BS, and MS degrees in applied mathematics from the College of Mathematics and Econometrics, Hunan University, Changsha, China, in 2007 and 2009, respectively. He is currently pursuing a PhD degree at the College of Electrical and Information Engineering, Hunan University, Changsha, China. His research interests include image processing, sparse representation, and pattern recognition.

Shutao Li received his BS, MS, and PhD degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively. In 2001, he joined the College of Electrical and Information Engineering, Hunan University. He is currently a professor with the College of Electrical and Information Engineering, Hunan University. He has authored or co-authored more than 100 referred papers. His professional interests are information fusion, pattern recognition, bioinformatics, and image processing. He served as a member of the Neural Networks Technical Committee from 2007 to 2008. He won two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006.