Multi-modality medical image fusion based on separable dictionary learning and Gabor filtering☆

Qiu Hu [a,b], Shaohai Hu [a,b,*], Fengzhen Zhang [c]

[a] *Institute of Information Science, Beijing Jiaotong Universitys, Beijing, 100044, China*
[b] *Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, 100044, China*
[c] *Key Laboratory of Solar Activity, National Astronomical Observatory, Chinese Academy of Science, Beijing, 100101, China*

ARTICLE INFO

ABSTRACT

Sparse representation (SR) has been widely used in image fusion in recent years. However, source image, segmented into vectors, reduces correlation and structural information of texture with conventional SR methods, and extracting texture with the sliding window technology is more likely to cause spatial inconsistency in flat regions of multi-modality medical fusion image. To solve these problems, a novel fusion method that combines separable dictionary optimization with Gabor filter in non-subsampled contourlet transform (NSCT) domain is proposed. Firstly, source images are decomposed into high frequency (HF) and low frequency (LF) components by NSCT. Then the HF components are reconstructed sparsely by separable dictionaries with iterative updating sparse coding and dictionary training. In the process, sparse coefficients and separable dictionaries are updated by orthogonal matching pursuit (OMP) and manifold-based conjugate gradient method, respectively. Meanwhile, the Gabor energy as weighting factor is utilized to guide the LF components fusion, and this further improves the fusion degree of low-significant feature in the flat regions. Finally, the fusion components are transformed to obtain fusion image by inverse NSCT. Experimental results demonstrate the more competitive results of the proposal, leading to the state-of-art performance on both visual quality and objective assessment.

1. Introduction

Medical image is becoming one of the most important assistants in clinical research and disease diagnosis in recent years. However, technology limitation significantly reduces the quality of acquired image, and this further affects the accuracy of medical image interpretation and analysis. By integrating complementary information of multi-modality images, the obtained fusion image is more comprehensive to describe target details, and this improves the accuracy of diagnosis [1]. For example, magnetic resonance imaging (MRI) can provide detailed information on soft tissue, while a dense structure such as a bone is well visualized by computed tomography (CT) technique. The fusion image obtained by integrating complementary information of them both can simultaneously describe soft tissue and bone information.

Multi-scale transform (MST) technology has been widely used to increase the sparsity of fusion image. To obtain fusion image, source images are decomposed and reconstructed at different scales and directions with MST-based methods. Dual-tree complex wavelet transform (DTCWT) [2], Laplace pyramid transform (LP) [3], curvelet transform (CVT) [4], nonsubsampled shearlet transform (NSST) [5] and nonsubsampled contourlet transform (NSCT) [6] are the conventional MST

methods for image fusion. Li compares the performance of different MST methods, and verifies that NSCT performs best in multi-modality medical image fusion [7]. In recent years, many novel MST-based fusion methods have been proposed. Du proposes a local laplacian filtering (LLF)-based multi-scale decomposition approach and the fusion high frequency (HF) components are based on the information of interest (IOI) strategy [8]. Das and Kundu introduce a fuzzy-adaptive reduced pulse coupled neural network (RPCNN)-based fusion strategy in NSCT domain [9]. Qiu proposes a sparse representation and the dynamic group sparsity recovery (DGSR)-based fusion strategy in NSST domain, and the HF components are fused by the absolute-maximum rule [10]. Xia combines sparse representation with pulse coupled neural network (PCNN) in NSCT domain for medical image fusion [11]. Yin proposes a parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain (NSST-PAPCNN)-based medical image fusion strategy [12]. Obviously, deep learning shows effectiveness in multi-modality image fusion. To overcome the difficulty in designing robust activity level measurements and weight assignment strategies, Liu employs a Siamese convolutional neural network to generate weight map for medical image fusion [13]. Hermessi proposes

a fully convolutional Siamese model trained with directional NSST features in a similarity learning fashion for the fusion of multi-modality image [14].

The applications of sparse representation (SR) in image de-noising [15], image fusion [16], voxel selection [17] and compressed sensing [18] have proved that SR is effective for image processing. SR expresses the natural sparsity of image by training as few sparse co-efficients as possible. This principle is consistent with the physiological characteristics of human visual system (HVS) [19]. Researches have demonstrated that the SR-based fusion methods generally outperform conventional MST in multi-focus image fusion [20]. However, the SR-based methods also have drawbacks. First of all, the sparse representation ability of over-complete dictionary may not be sufficient to handle fine details, therefore deviations between source and fusion images are existed to some extent. The high redundant dictionary is sensitive to random noise, and may cause visual artifacts in fusion results [21]. Secondly, the sliding window technology employed in SR-based methods inevitably reduces correlation of adjacent patches, and will result in spatial inconsistency of fusion image. Different from the high similarity between multi-focus source images, differences of brightness and texture at the same region between multi-modality medical source images are significant in general. A small change of pixels in the flat region with less texture information will cause different brightness and texture of adjacent fusion patches, and this will result in discontinued structure of fusion result. Thus, SR is not sufficient to achieve satisfactory multi-modality medical image fusion result. A popular special consistency based fusion approach is formulating an energy function. Gabor filter distinguishes and represents image texture with frequency and direction representation, and has good invariance to noise and transformation [22]. Ramamoorthy attempts to take Gabor energy as a fusion criterion for medical image [23], and this strategy achieves good fusion results. Inspired by this, it is of great significance to design and deep research the fusion scheme integrating SR and Gabor filter with complementary advantages.

Combining the complementary advantages of SR and Gabor filter, a novel multi-modality medical image fusion method in NSCT domain is proposed. In the proposal, NSCT is as the fusion framework. A novel SR method so-called the analytic separable dictionary learning (ASeDiL) method is firstly exploited to fuse the HF components decomposed by NSCT, where ASeDiL was mainly used for image de-noising in [24], while Gabor energy is as weighting factor to guide the low frequency (LF) components fusion. The main contribution of the proposal can be summarized as: On one hand, the ASeDiL-based method not only increases the sparsity of image decomposed by NSCT, but increases dimensions of texture extraction without adding dictionary redundancy, furthermore, the effective noise suppression performance of ASeDiL improves the accuracy of texture extraction. On the other hand, Gabor energy improves the spatial inconsistency of medical fusion image caused by sparse representation.

2. Related work

2.1. Separable dictionary learning algorithm

Conventional sparse representation algorithm, such as k-singular value decomposition (K-SVD) [25], employs over-complete dictionary trained by image samples to obtain the over-complete matrix reflecting image features. The objective function of dictionary learning is defined as

$$\arg \min_{D,\alpha} \|\alpha\|_0 \quad s.t. \quad \|y - D\alpha\|_2^2 \le \varepsilon \tag{1}$$

where $y \in R^m$ means column vector of image sample $Y \in R^{\sqrt{m} \times \sqrt{m}}$, $D \in R^{m \times n}(m < n)$ means over-complete dictionary, $\alpha \in R^n$ means the sparse coefficient obtained by dictionary training and ε means tolerance of reconstruction error. $\|\cdot\|_0$ means the number of non-zero elements in vector (i.e. the sparsity), and $\|\cdot\|_2^2$ means the inner product of vector.

In the process of dictionary learning in Eq. (1), block images are converted into vectors, and this can be viewed as extracting texture information from image on only one dimension. Furthermore, the correlation and structural information of adjacent blocks have also been reduced. To overcome the aforementioned deficiencies of dictionary learning, separable dictionary principle is employed in the proposal to extract texture information. On premise of ensuring the quality of image reconstruction, by minimizing the ℓ_0 norm of sparse coefficients, the objective function of separable dictionary learning algorithm is defined as

$$\arg \min_{D_A, \{s_j\}_{j=1}^N, D_B} \frac{1}{N} \left(\sum_{j=1}^N \left\| S_j \right\|_0 \right)$$

$$s.t. \ D_A S_j D_B^T = Y_j, S_j \in R^{n \times n}, \forall j = 1, \dots, N \tag{2}$$

where S means sparse matrix, N is the number of image samples. $D_A \in R^{\sqrt{m} \times n}$ and $D_B \in R^{\sqrt{m} \times n}$ are the over-complete separable dictionaries with same size, then $n \ge m$. The over-complete dictionary D is obtained by the Kronecker product of D_A and D_B.

In Eq. (1), it is unreasonable to update dictionary and sparse co-efficient synchronously for the big differences in numerical value and meaning between dictionary and sparse coefficient. Then sparse coding and dictionary updating in the proposal are updated by orthogonal matching pursuit (OMP) and manifold-based conjugate gradient algorithms, respectively. The steps of separable dictionary learning algorithm are as follows.

2.1.1. Sparse coding

To get the sparsest representation under current dictionary, the objective function of sparse coding is obtained through Eq. (2) as

$$\arg \min_{\{s_j\}_{j=1}^N} \frac{1}{N} \left(\sum_{j=1}^N \left\| S_j \right\|_0 \right) s.t. \left\| D_A S_j D_B^T - Y_j \right\|_F^2 < \varepsilon \ \forall j = 1, \dots, N \tag{3}$$

where $\|\cdot\|_F^2$ means the inner product operation of matrices. The sparse coding process is consisted of four parts:

① Select the most relevant dictionary atom: in OMP algorithm, maximizing the absolute value of inner product between dictionary atom $D_{i,j}$ and the current residual $R \in R^{\sqrt{m} \times \sqrt{m}}$ as $(i, j) \leftarrow \arg\max \left| \left\langle R, D_{i,j} \right\rangle \right|$ to point out the position of the most relevant dictionary atom.

② Update support set: the support set is defined as $\Lambda_k = \left\{ (i_1, j_1), (i_2, j_2), \dots, (i_k, j_k) \right\}$ after the kth iteration and the corresponding separable dictionaries are $D_{A_{\Lambda_k}} = [a_{i_1}, a_{i_2}, \dots, a_{i_k}]$ and $D_{B_{\Lambda_k}} = [b_{j_1}, b_{j_2}, \dots, b_{j_k}]$, where a_i is the ith column of D_A and b_j is the jth column of D_B.

③ Update sparse matrix S: update sparse matrix by the least squares method as

$$S = \arg \min \|R\|_F^2 = \arg \min tr(RR^T)$$
$$= \arg \min \left(s_{\Lambda_k}^T W s_{\Lambda_k} - 2H s_{\Lambda_k} \right) \tag{4}$$

where s_{Λ_k} is consisted of the elements in sparse matrix S corresponding to Λ_k, and W and H are calculated as

$$W = \begin{bmatrix} tr\left(D_{i_1 j_1} D_{i_1 j_1}^T \right) & \cdots & tr\left(D_{i_1 j_1} D_{i_k j_k}^T \right) \\ \vdots & \ddots & \vdots \\ tr\left(D_{i_k j_k} D_{i_1 j_1}^T \right) & \cdots & tr\left(D_{i_k j_k} D_{i_k j_k}^T \right) \end{bmatrix} \quad H = \begin{bmatrix} tr\left(Y D_{i_1 j_1}^T \right) \\ \vdots \\ tr\left(Y D_{i_k j_k}^T \right) \end{bmatrix}$$

The partial derivation of s_{Λ_k} in Eq. (4) is calculated as

$$s_{\Lambda_k} = W^{-1} H \tag{5}$$

④ Update residual: the reconstructed image sample under the current dictionary is estimated as $\hat{Y} = D_{A_{\Lambda_k}} diag\left(s_{\Lambda_k} \right) D_{B_{\Lambda_k}}$, then the updated residual $R = Y - \hat{Y}$ is obtained after k iterations.

④ returns to ① to perform iterative cycles, and when $\left\| D_A S_j D_B^T - Y_j \right\|_F^2 > \varepsilon$, the iteration is terminated and the optimal sparse matrix S is obtained.

2.1.2. Update separable dictionaries

When S is obtained by sparse coding, the objective function of separable dictionary updating is defined as

$$\arg \min_{D_A, D_B} \frac{1}{N} \left(\sum_{j=1}^{N} \left\| D_A S_j D_B^T - Y_j \right\|_F^2 \right) \tag{6}$$

Combining the constraints of dictionary, that is, the ℓ_2 norm of dictionary atoms equals to 1 and no correlation of atoms, log function is employed to fit the full rank and column irrelevance of separable dictionary. Then, Eq. (6) is rewritten as

$$f(D_A, D_B) \leftarrow \arg \min_{D_A, D_B} \frac{1}{N} \left(\sum_{j=1}^{N} \left\| D_A S_j D_B^T - Y_j \right\|_F^2 \right)$$
$$+ \lambda[r(D_A) + r(D_B)] + \beta[h(D_A) + h(D_B)] \tag{7}$$

where $f(\cdot)$ means the reconstructed objective function, λ and β are the fitting parameters, and $h(D_A), r(D_A), h(D_B), r(D_B)$ are defined as

$$h(D_A) = -\frac{1}{m \log(m)} \log \det(\frac{1}{n} D_A D_A^T)$$
$$r(D_A) = -\sum_{k<l} \log(1 - ((D_{A_k})^T (D_{A_l}))^2)$$
$$h(D_B) = -\frac{1}{m \log(m)} \log \det(\frac{1}{n} D_B D_B^T)$$
$$r(D_B) = -\sum_{k<l} \log(1 - ((D_{B_k})^T (D_{B_l}))^2)$$

Projecting $V = (D_A, D_B)$, consisted of the separable dictionaries D_A and D_B, onto the Riemannian manifold [26], and Eq. (7) can then be rewritten as

$$\arg \min_{V \in M} f(V) \tag{8}$$

where M means the Riemannian sub-manifold of some Euclidean spaces, and V is one point on M. The conventional steepest descent method is slow to converge the objective function $f(V)$ into minimum, and then the conjugate gradient method [27] as the optimization algorithm is utilized in the proposal. Firstly, a geodesic $\Gamma(V, H, t)$ is the smooth shortest curve from $V \in M$ in the direction of H on tangent space $T_V M$ to another point on M, and the corresponding linear search step size is calculated as

$$\eta^{(i)} \leftarrow \arg \min_{i \geq 0} f(\Gamma(V^{(i)}, H^{(i)}, t)) \tag{9}$$

Then, the next iteration point $V^{(i+1)} = \Gamma(V^{(i)}, H^{(i)}, \eta^{(i)})$ can be obtained. $\Gamma(V^{(i)}, H^{(i)}, t) = (\Gamma(D_A, H_{D_A}, t), \Gamma(D_B, H_{D_B}, t))$, $\Gamma(D_A, H_{D_A}, t) = [r(D_{A_1}, H_{D_{A_1}}, t), \ldots, r(D_{A_n}, H_{D_{A_n}}, t)]$ and $\Gamma(D_B, H_{D_B}, t) = [r(D_{B_1}, H_{D_{B_1}}, t), \ldots, r(D_{B_n}, H_{D_{B_n}}, t)]$, and circle $r(d, h, t)$ is defined as

$$r(d, h, t) = \begin{cases} d & if \ \|h\|_2 = 0 \\ d \cos(t \|h\|_2) + h \frac{\sin(t \|h\|_2)}{\|h\|_2} & otherwise \end{cases}$$

The Riemannian gradient G of the current point $V^{(i+1)}$, that is, the orthogonal projection of the gradient $\nabla f(V^{(i+1)})$ onto tangent space $T_V M$, can then be calculated as

$$G(V^{(i+1)}) = \Pi_{T_V M} \left(\nabla f \left(V^{(i+1)} \right) \right) = \nabla f(V^{(i+1)})$$
$$- V^{(i+1)} diag \left(\left(V^{(i+1)} \right)^T \left(\nabla f \left(V^{(i+1)} \right) \right) \right) \tag{10}$$

At this point, the corrected $H^{(i+1)}$ based on the previous search direction $H^{(i)}$ is calculated as

$$H^{(i+1)} = -G^{(i+1)} + \delta \tau(H^{(i)}, V^{(i)}, H^{(i)}, \eta^{(i)}) \tag{11}$$

where $\tau(H^{(i)}, V^{(i)}, H^{(i)}, \eta^{(i)})$ is defined as

$$\tau(\xi, d, h, t) = \xi - \frac{\xi^T h}{\|h\|_2} (\xi \|h\|_2 \sin(t \|h\|_2) + h(1 - \cos(t \|h\|_2)))$$

Iterative cycles of Riemannian gradient G and search direction H back to Eq. (9), and the updated dictionary $V = (D_A, D_B)$ can then be obtained.

More details of the separable dictionary learning algorithm are available in [24].

2.2. Gabor filter

Gabor filter, discriminating and characterizing texture of image by frequency and direction representation, is effective for the description of image energy distribution [28]. Employing Gaussian kernel function $g(x, y)$ of pixel point (x, y) modulated by a sine wave, the Gabor filter of m_1th scale and n_1th direction is defined as

$$g_{m_1 n_1}(x', y') = a^{-m_1} g(x, y), \quad a > 1, m_1 = 1, 2, \ldots, m_0, n_1 = 1, 2, \ldots, n_0$$
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = a^{-m_1} \begin{bmatrix} \cos \theta_{n_1} & \sin \theta_{n_1} \\ -\sin \theta_{n_1} & \cos \theta_{n_1} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \theta_{n_1} = \frac{n_1 \pi}{n_0} \tag{12}$$

where a means the scale parameter, m_0 and n_0 are the total scales and directions, respectively. Then, LF component is convolved with Gabor filter to obtain the Gabor coefficient

$$C_g(x, y, m, n) = \iint L(x', y') g(x - x', y - y') dx' dy' \tag{13}$$

where $L(x, y)$ is the input matrix of LF component. At this point, the Gabor energy of all scales m_0 and directions n_0 can be calculated as

$$E_g(x, y) = \sum_{m=1}^{m_0} \sum_{n=1}^{n_0} \left| C_g(x, y, m, n) \right|^2 \tag{14}$$

3. Proposed fusion method

3.1. Overview

The schematic diagram of the proposed multi-modality medical image fusion method is shown in Fig. 1. As can be seen from Fig. 1, the fusion scheme is divided into four steps: NSCT decomposition, HF components fusion, LF components fusion, and NSCT reconstruction. The proposal takes two source images as an example, and the principle can be directly extended to multiple source images.

3.2. Detailed fusion scheme

3.2.1. NSCT decomposition

Decomposing both source images Y_1 and Y_2 into K scales and L directions, and the multi-scale decomposition coefficients $\{L^{Y_1}, H_{kl}^{Y_1}\}$ and $\{L^{Y_2}, H_{kl}^{Y_2}\}$ can be obtained. L^{Y_1} and L^{Y_2} represent the LF components of source images, $H_{kl}^{Y_1}$ and $H_{kl}^{Y_2}$ represent the HF components at $k \in [1, K]$ scale and $l \in [1, L]$ direction of source images.

3.2.2. HF components fusion

The HF components H^{Y_1} and H^{Y_2} decomposed by NSCT are adopted as the image samples to obtain sparse matrices S_1 and S_2 by OMP algorithm, respectively, and the updated dictionary $V = (D_A, D_B)$ is obtained by the manifold-based conjugate gradient algorithm. The over-complete dictionary leads to non-unique solution of sparse matrix. The sparsest set of solutions will make the target vector most similar to the image sample. The fused sparse coefficient can be obtained by the "Maximum Rule" as

$$S = \max(S_1, S_2) \tag{15}$$

Then, the fused HF component obtained by sparse reconstruction is calculated as
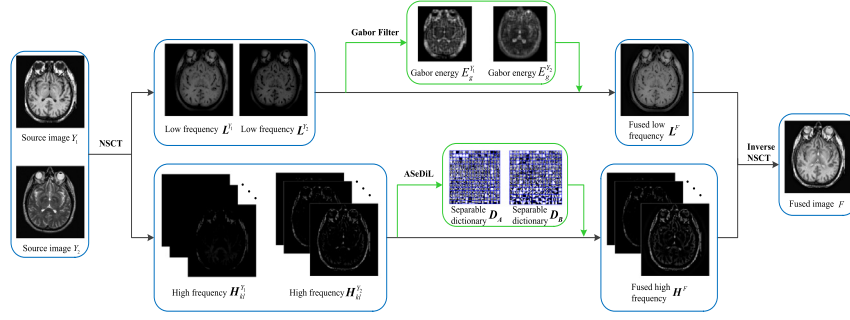
$$H^F = D_A S D_B^T \tag{16}$$

Fig. 1. The schematic diagram of the proposed medical image fusion method.

3.2.3. LF components fusion

Since most of image energy is reserved in the LF component, the parameters of image energy as a guidance of LF components fusion is a beneficial attempt. Gabor filter has good invariance to noise and transformation while representing image energy [29]. Therefore, LF components are fused by the guidance of Gabor filter in the proposal.

Using Gabor filter to select appropriate coefficients to fuse the LF components decomposed by NSCT, the corresponding fusion rule is defined as

$$
L^F(x,y) =
\begin{cases}
L^{Y_1}(x,y), & E_g^{Y_1}(x,y) - E_g^{Y_2}(x,y) > T \\
L^{Y_1}(x,y) \times w^{Y_1} + L^{Y_2}(x,y) \times w^{Y_2}, & \left| E_g^{Y_1}(x,y) - E_g^{Y_2}(x,y) \right| \le T \\
L^{Y_2}(x,y), & E_g^{Y_1}(x,y) - E_g^{Y_2}(x,y) < T
\end{cases}
$$

$$(17)$$

where $E_g^{Y_1}$ and $E_g^{Y_2}$ are the Gabor energies obtained by Eq. (14), and $w^{Y_1} = E_g^{Y_1}/(E_g^{Y_1} + E_g^{Y_2})$ and $w^{Y_2} = E_g^{Y_2}/(E_g^{Y_1} + E_g^{Y_2})$ are the corresponding weight coefficients. $L^F(X,Y)$ means the fusion LF component of pixel point (X,Y). T means the Gabor threshold. If the difference value of Gabor energy is larger than threshold T, the pixel with larger Gabor energy is selected as the fusion result. In other cases, the weighting factor consisted of Gabor energy is adopted to fuse the LF components to ensure the stability of fusion result.

3.2.4. NSCT reconstruction

Inverse NSCT transformation is performed to obtain the fusion image F.

Algorithm 1 summarizes the proposed multi-modality medical image fusion method.

4. Experiments

4.1. Experimental parameters setting

4.1.1. Source images

To verify the performance of the proposal, our fusion method is tested on 127 groups of brain anatomical CT and MR images and compared with 7 popular image fusion methods. MR imaging techniques can provide 4 kinds of anatomical images which are all considered in our research. Due to space constraints, only 5 groups of the 127 groups CT and MR source images tested in this paper are shown in Fig. 2, and all of them are obtained from the Whole Brain Atlas Medical Image Database created by Harvard Medical School. The spatial resolution of all source images is set to 256×256, and image registration has been realized for each group.

4.1.2. Methods for comparison

The performance of proposal is compared with 7 popular image fusion methods: the NSST-based method [5], the NSCT-based method [6], the convolutional sparse representation (CSR)-based method [30], the convolutional sparsity morphological component analysis (CSMCA)-based method [31], the NSCT-RPCNN-based method [9], the sparse representation in NSCT domain (NSCT-SR)-based method [32], and the NSST-PAPCNN-based method [12]. Among them, NSST and NSCT are the typical representatives in MST-based image fusion. CSR and NSCT-SR belong to sparse representation methods for multi-modality image fusion, and they have universal applicability for multiple types of image fusion. CSMCA, NSCT-RPCNN and NSST-PAPCNN belong to the medical image fusion methods, and CSMCA and NSST-PAPCNN have been proposed within one year. The dataset and parameters setting required for separable dictionary training in the proposal are consistent with the ASeDiL method in [24]. The other comparison methods adopt default parameters given by the corresponding authors. The source codes of CSR,[1] CSMCA,[2] NSCT-RPCNN,[2] NSCT-SR[2] and NSST-PAPCNN[2] are available.

4.1.3. Objective evaluation metrics

To evaluate the performance of the above fusion methods in quantitative, four accepted objective evaluation metrics are adopted in experiments, that is, Piella–Heijmans's similarity based metric Q_E [33], spatial frequency (SF), universal image quality index (UIQI) [34] and mutual information [35]. SF is the statistical characteristic based evaluation metric, and it is used to reflect the clarity of image. Q_E, UIQI and MI belong to the evaluation metric based on the relationship between fusion result and source images. Q_E reflects the similarity among intensity, contrast and structure in comprehensive. UIQI is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. MI reflects the amount of information in source images contained in fusion image. The larger values of these metrics imply the better fusion results. For more information about these metrics, please refer to the related resources.

4.1.4. Experimental environments

The above experiments are carried out at the environments of the computer configuration with 3.3 GHz CPU and 16.0 GHz RAM, and the MATLAB version of R2017a and Win7 64-bit operation system.

4.2. Effectiveness of ASeDiL algorithm

As aforementioned, ASeDiL can effectively suppress the interference of random noise to the texture features of reconstructed image. Fig. 3 shows the comparisons of noise suppression performance between ASeDiL and the conventional K-SVD method. It can be seen that

---

[1] https://sites.google.com/site/yuliu316316.
[2] https://sites.google.com/site/wodrsdas/.

Algorithm 1 The medical image fusion method of the proposal

Inputs: sources: Y_1 and Y_2

parameters: scale number of NSCT decomposition: K, direction number of decomposition in each scale: L, step size in the HF components fusion: η, threshold of Gabor filter in the LF components fusion: T

Step 1: NSCT decomposition

01: For each of the source images $Y = [Y_1, Y_2]$
02:     Perform NSCT decomposition to get $\{L^Y, H_{kl}^Y\}$, $k \in [1, K]$, $l \in [1, L]$
03: End

Step 2: HF components fusion

04: For each of the scales $k = 1 : K$
05:     For each of the directions $l = 1 : L$
06:         For each of the HF components $H_{kl}^{Y_1}, H_{kl}^{Y_2}$
07:             Initialize separable dictionary model: $R^{(0)} = H_{kl}^Y, V^{(0)} = (D_A, D_B), G^{(0)} = H^{(0)} = -(\nabla f(V^{(0)}))$
08:             Using sparse coding steps ① ② ③ ④ for cyclic iterations, updating sparse matrix S
09:             Cyclic iterations using (9)-(11), updating the dictionary V
10:         End
11:            Using (15), (16) to obtain the fused HF component H^F
12:     End
13: End

Step 3: LF components fusion

14: For each of the LF components L^{Y_1}, L^{Y_2}
15:     Calculate the Gabor energy of LF component L^Y using (14)
16: End
17: Using (17) to fuse L^{Y_1} and L^{Y_2} to obtain L^F

Step 4: NSCT reconstruction

18: Perform NSCT inverse transform of $\{L^F, H^F\}$ to get F

Output: The fusion image F
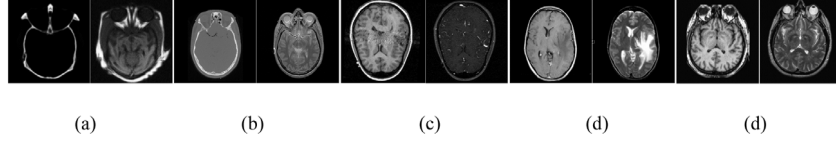

(a)   (b)   (c)   (d)   (d)

Fig. 2. 5 groups of multi-modality medical image tested in this paper.

under the same Gaussian noise, the texture structure reconstructed by ASeDiL is clearer and less interfered by noise. Fig. 4 shows a more intuitive reflection of the reconstructed image in Fig. 3 with the relationship between peak signal-to-noise ratio (PSNR) and superimposed noise σ_n, and further verifies the conclusions of Fig. 3. The curve "Original image" means the unprocessed noised image.

HF component decomposed by NSCT contains most of the significant information such as edge texture, and random noise of source images is mainly contained in HF component. Figs. 3 and 4 have proved that ASeDiL has good suppression performance of random noise and retains clearer texture information. Therefore, ASeDiL is effective as the HF components fusion criterion in the proposal.

4.3. Analysis of algorithm parameters

The fusion performance of the proposal is mainly affected by two parameters, that is, the moving step size η of sliding window in HF components fusion and the Gabor threshold T in LF components fusion. To analyze the impact of the two parameters on fusion performance, the aforementioned objective evaluation metrics (i.e. Q_E, SF, UIQI and MI) and 4 groups of source image (i.e. (a) CT and MRI, (c) MR-TI and MRA, (d) MR-Gd and MR-T2, and (e) MR-T1 and MR-T2) in Fig. 2 are adopted in this section. For the sake of clarity, only a set of results for each parameter are shown, and other parameters are set to perform well. (Analyze the impact of η with $T = 1$ and analyze the impact of T with $\eta = 1$)

The first two rows in Fig. 5 show the impact of η on the objective fusion performance. With the increase of η, the objective evaluation

metrics are smaller gradually and tend to be stable. The best objective performance is obtained when $\eta = 1$. The last two rows in Fig. 5 show the impact of T on the objective fusion performance. It can be seen that the increase of T reduces the objective evaluation metrics and the best objective performance is obtained when $T = 1$.

Based on the earlier discussions, the setting $\{\eta = 1, T = 1\}$ is adopted in all of the following experiments.

4.4. Comparison to other fusion methods

This section compares the proposal to the other fusion methods in the following two aspects: visual quality and objective assessment.

4.4.1. Visual quality

Aiming at the aforementioned 5 groups of medical source image in Fig. 2, 4 groups are chosen for comparisons. In addition, two representative close-ups (marked as left view and right view) of each image (mainly from the perspective of image processing, such as contrast, texture structure and brightness, etc.) are enlarged for better comparison.

Fig. 6 shows the fusion results of CT and MRI images. Compared with other methods, there is significant luminance loss of soft tissue structure in NSST, NSCT, CSR and CSMCA methods [see the right views in Fig. 6(c), (d), (e) and (f)]. NSCT-SR method reduces the luminance distortion, but there is a loss of local texture information [see Fig. 6(h)]. NSCT-RPCNN and NSST-PAPCNN methods perform well in containing the texture and luminance information, but there is a
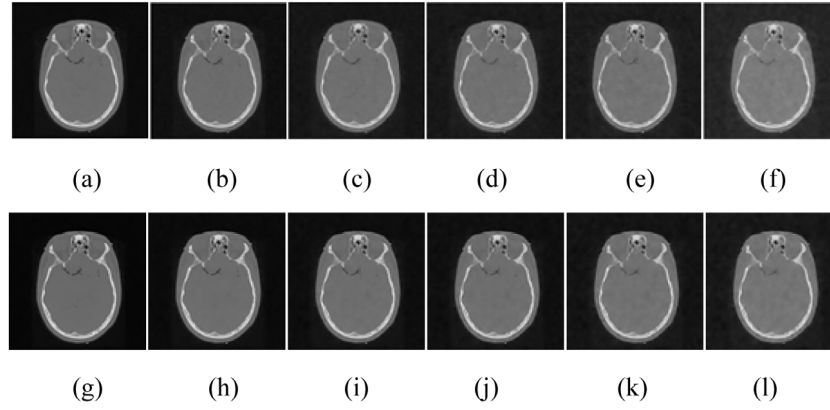
Fig. 3. Noise suppression performance comparisons between K-SVD [(a)–(f)] and ASeDiL [(g)–(l)] methods ($\eta = 1$): (a) $\sigma_n = 5$ dB; (b) $\sigma_n = 10$ dB; (c) $\sigma_n = 15$ dB; (d) $\sigma_n = 20$ dB; (e) $\sigma_n = 25$ dB; (f) $\sigma_n = 30$ dB; (g) $\sigma_n = 5$ dB; (h) $\sigma_n = 10$ dB; (i) $\sigma_n = 15$ dB; (j) $\sigma_n = 20$ dB; (k) $\sigma_n = 25$ dB; (l) $\sigma_n = 30$ dB.
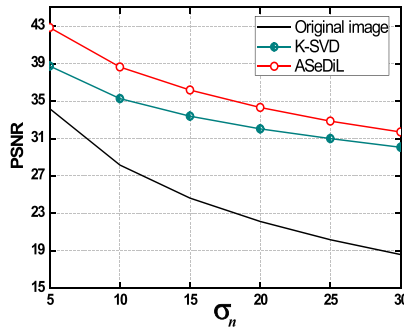

Fig. 4. PSNR as a function of σ_n with the reconstruction images in Fig. 3.

tendency of blurring local edge regions for NSCT-RPCNN method [see the left view in Fig. 6(g)], and NSST-PAPCNN method causes obvious artificial texture at the edge of fusion image [see Fig. 6(i)]. In contrast, the proposal preserves texture structure and luminance information better, while avoiding the artificial texture generation [see Fig. 6(j)].

Fig. 7 shows the fusion results of MR-T1 and MRA images. It can be seen that the fusion results of NSST, NSCT, CSR and CSMCA methods still facing luminance loss problem [see Fig. 7(c), (d), (e) and (f)]. The detailed texture information of the fusion image becomes blurred after using NSCT-RPCNN method [see the right view in Fig. 7(g)]. NSCT-SR method preserves the complete texture information, but some local region has obvious intensity loss [see the left view in Fig. 7(e)]. The fusion result of NSST-PAPCNN method is still outstanding, but the texture details of some local region are blurred [see the right view in Fig. 7(i)]. In contrast, the clarity of texture structure in proposal is better [see Fig. 7(j)].

Fig. 8 shows the fusion results of MR-Gd and MR-T2, and significant information such as luminance and texture are mainly contained in MR-T2 source image. It can be seen that NSST, NSCT and CSR methods cause obvious luminance loss, which is consistent with Fig. 7 [see Fig. 8(c), (d) and (e)]. It is worth noting that the luminance distortion of CSMCA method is significantly improved in the fusion process between MR-Gd and MR-T2 images [see Fig. 8(f)]. NSCT-SR method weakens region correlation of adjacent patches employing sliding window technology, and causes inconsistent local space [see Fig. 8(h)]. Although NSCT-RPCNN and NSST-PAPCNN methods have no obvious luminance information loss, there exists local region ambiguity and contrast distortion [see Fig. 8(g) and (i)]. There is no obvious luminance loss of proposal with clear detail texture and high contrast [see Fig. 8(j)].

Fig. 9 shows the fusion results of MR-T1 and MR-T2. Obviously, luminance information loss still exists in NSST, NSCT and CSR methods

[see the left view in Fig. 9(c), (d) and (e)]. CSMCA method preserves luminance and significant texture information well in MR-T1 and MR-T2 images fusion process [see the left view in Fig. 9(f)]. NSCT-RPCNN and NSST-PAPCNN methods have better fusion results of luminance and texture information, but the contrast distortion appears in both methods and the contour of left eye becomes blurred [see the left views in Fig. 9(g) and (i)]. The proposal performs best in terms of luminance and detail texture retention [see Fig. 9(i)].

4.4.2. Objective assessment

To reflect the fusion performance of the above methods more intuitively in different types of medical source images, 4 groups of source image selected in the visual quality are statistically compared in Fig. 10 with different evaluation metrics. Under the same evaluation metric, it can be seen that the objective performance of the proposed method is outstanding in different types of source images. Besides the 5 groups of medical image in Fig. 2, another 123 groups of brain anatomical CT and MR images are tested to further verify the superiority of the proposed method, including 60 groups of CT and MRI medical images and 63 groups of MR-T1 and MR-T2 medical images. Tables 1 and 2 further statistically average the objective metrics of the proposed method with the 123 groups source images, and the bold font indicate the arrangement of top three. It can be seen that the scores of the proposed method are in the top three under different objective metrics, and this indicates the fusion performance of the proposed method is effective for different type medical image.

5. Conclusion and forecast

A novel multi-modality medical image fusion method in NSCT domain is proposed, and the main advantages of the proposal include the two points. On one hand, the ASeDiL-based method not only increases the sparsity of image decomposed by NSCT, but increases dimensions of texture extraction without adding dictionary redundancy, furthermore, the effective noise suppression performance of ASeDiL improves the accuracy of texture extraction. On the other hand, Gabor energy improves the spatial inconsistency of medical fusion image caused by sparse representation. A large number of experiments on 127 groups of multi-modality medical image verify the effectiveness of the proposed method. Compared with 7 representative fusion methods, the visual quality and objective assessment show that the proposed method has reached the state-of-art performance. However, the iterative processing of sparse representation is time-consuming and this will be our future work to deal with this problem.

CRediT authorship contribution statement

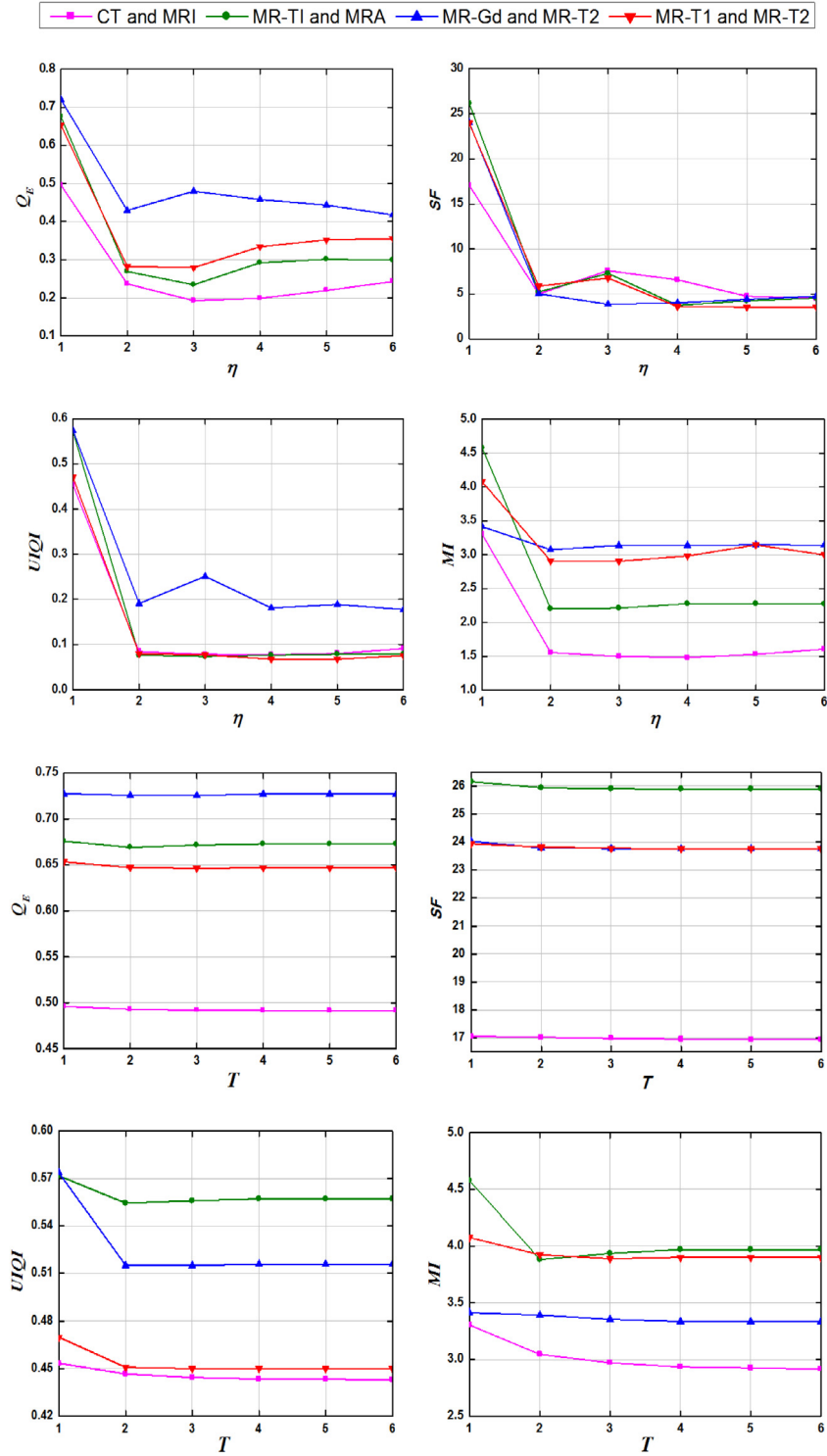**Qiu Hu:** Writing - original draft. **Shaohai Hu:** Validation. **Fengzhen Zhang:** Methodology.

Fig. 5. Objective performance of the proposal under different random parameters.

Table 1
Objective quality evaluation of CT/MRI medical image fusion.

Index	NSST	NSCT	CSR	CSMCA	NSCT-RPCNN	NSCT-SR	NSST-PAPCNN	Proposed
Q_E	0.7673	0.7767	0.7648	0.7573	**0.7792**	**0.7803**	0.7689	**0.7771**
SF	**22.9454**	**22.0173**	20.2508	21.8924	20.8459	22.0026	18.6284	**22.1142**
UIQI	0.6081	0.637	**0.7255**	0.7025	**0.7081**	0.6381	0.6043	**0.7353**
MI	2.0631	2.1124	**2.3573**	**2.6003**	1.8797	2.1127	1.8462	**2.2317**
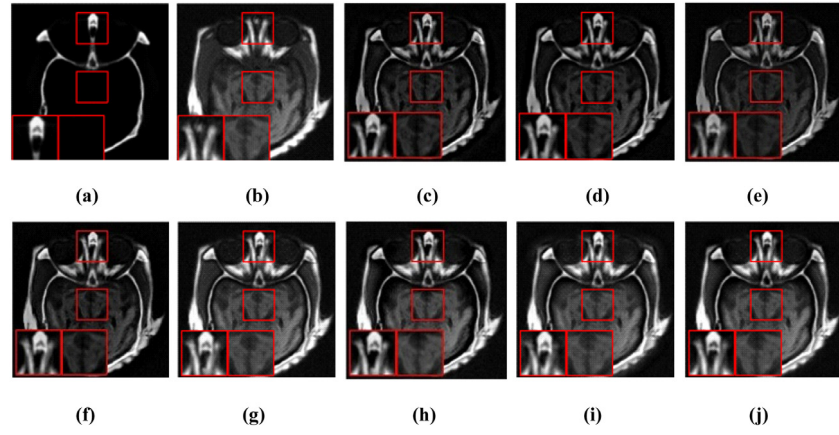
Fig. 6. The fusion results of CT and MRI images. (a) CT; (b) MRI; (c) NSST; (d) NSCT; (e) CSR; (f) CSMCA; (g) NSCT-RPCNN; (h) NSCT-SR; (i) NSST-PAPCNN; (j) Proposed.
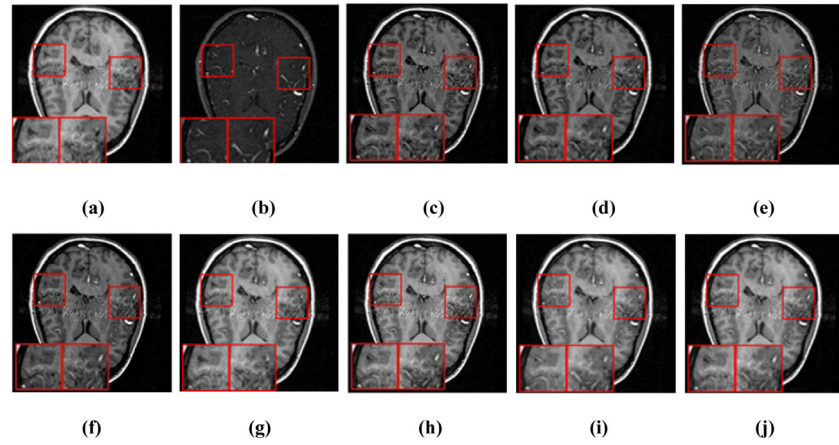

Fig. 7. The fusion results of MR-T1 and MRA images. (a) MR-T1; (b) MRA; (c) NSST; (d) NSCT; (e) CSR; (f) CSMCA; (g) NSCT-RPCNN; (h) NSCT-SR; (i) NSST-PAPCNN; (j) Proposed.
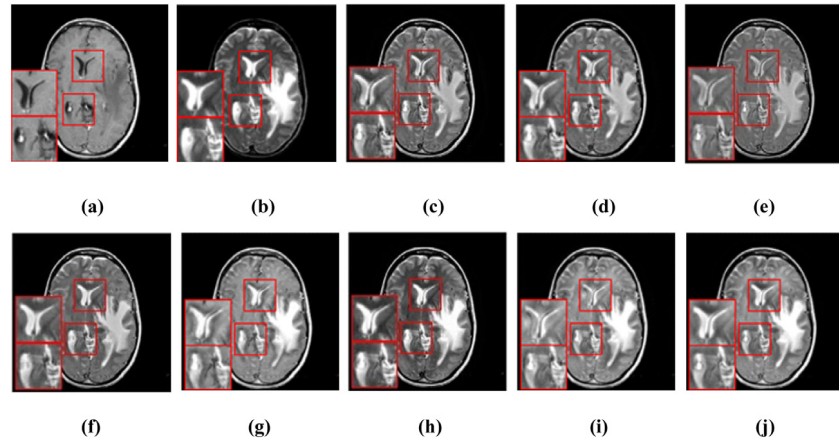

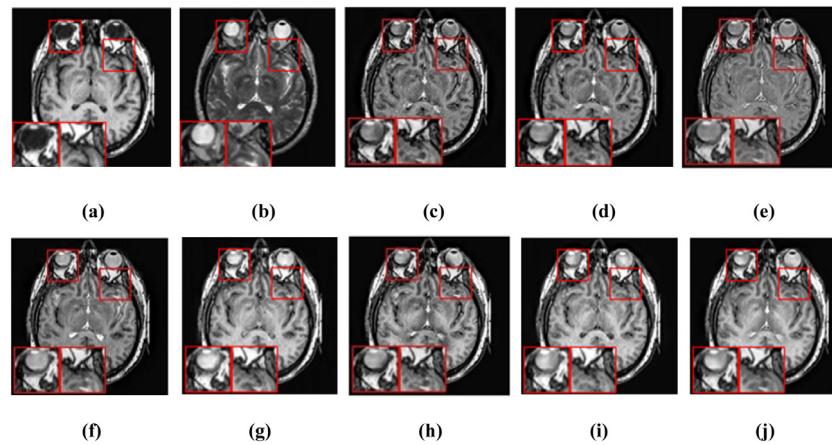
Fig. 8. The fusion results of MR-Gd and MR-T2 images. (a) MR-Gd; (b) MR-T2; (c) NSST; (d) NSCT; (e) CSR; (f) CSMCA; (g) NSCT-RPCNN; (h) NSCT-SR; (i) NSST-PAPCNN; (j) Proposed.

Table 2
Objective quality evaluation of MR-T1/MR-T2 medical image fusion.

Index	NSST	NSCT	CSR	CSMCA	NSCT-RPCNN	NSCT-SR	NSST-PAPCNN	Proposed
Q_E	0.6902	0.7188	0.7397	0.7411	0.7531	**0.7562**	**0.7543**	**0.7569**
SF	25.6768	24.5484	21.7924	24.3952	**26.4242**	**26.2591**	25.7767	25.7582
UIQI	0.4866	0.4819	0.6657	**0.6849**	0.6724	**0.6799**	0.6425	**0.6891**
MI	2.194	2.2555	2.3343	2.4299	2.5757	**2.6953**	**2.7139**	**2.6365**

Fig. 9. The fusion results of MR-T1 and MR-T2 images. (a) MR-T1; (b) MR-T2; (c) NSST; (d) NSCT; (e) CSR; (f) CSMCA; (g) NSCT-RPCNN; (h) NSCT-SR; (i) NSST-PAPCNN; (j) Proposed.
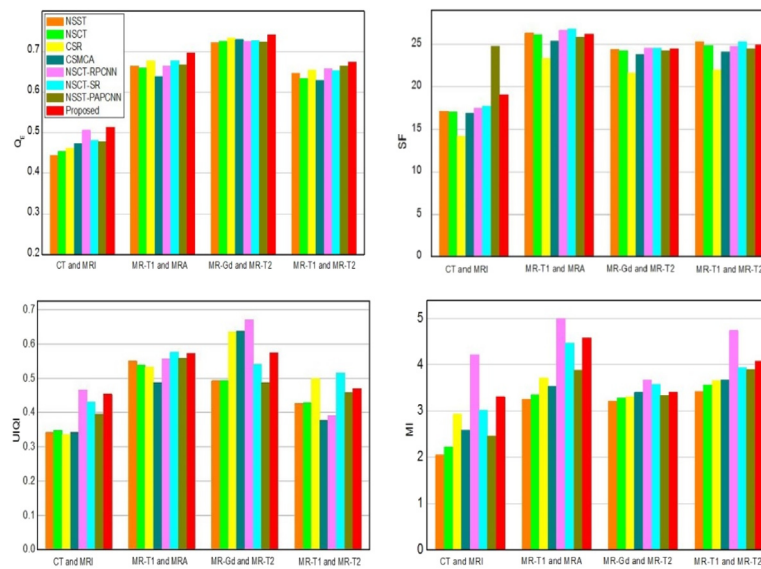

Fig. 10. Objective evaluation of different fusion methods in different type of source image.

Acknowledgments

References

[1] B.C. Porter, D.J. Rubens, J.G. Strang, J. Smith, S. Totterman, K.J. Parker, Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers, IEEE Trans. Med. Imaging 20 (4) (2001) 354–359, http://dx.doi.org/10.1109/42.921484.

[2] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, Inf. Fusion 8 (2) (2007) 119–130, http://dx.doi.org/10.1016/j.inffus.2005.09.006.

[3] P.J. Burt, E.H. Adelson, The laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540, http://dx.doi.org/10.1109/TCOM.1983.1095851.

[4] F. Nencini, A. Garzelli, S. Baronti, L. Alparonec, Remote sensing image fusion using the curvelet transform, Inf. Fusion 8 (2) (2007) 143–156, http://dx.doi.org/10.1016/j.inffus.2006.02.001.

[5] G. Easley, D. Labate, W.-Q. Lim, Sparse directional image representations using the discrete shearlet transform, Appl. Comput. Harmon. Anal. 25 (1) (2008) 25–46, http://dx.doi.org/10.1016/j.acha.2007.09.003.

[6] A.L. da Cunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: Theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101, http://dx.doi.org/10.1109/TIP.2006.877507.

[7] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transforms for image fusion, Inf. Fusion 12 (2) (2011) 74–84, http://dx.doi.org/10.1016/j.inffus.2010.03.002.

[8] J. Du, W. Li, B. Xiao, Anatomical-functional image fusion by information of interest in local Laplacian filtering domain, IEEE Trans. Image Process 26 (12) (2017) 5855–5866, http://dx.doi.org/10.1109/TIP.2017.2745202.

[9] S. Das, M.K. Kundu, A neuro-fuzzy approach for medical image fusion, IEEE Trans. Biomed. Eng. 60 (12) (2013) 3347–3353, http://dx.doi.org/10.1109/TBME.2013.2282461.

[10] Q. Chenhui, W. Yuanyuan, Z. Huan, X. Shunren, Image fusion of ct and mr with sparse representation in nsst domain, Comput. Math. Methods Med. 2017 (2) (2017) 1–13, http://dx.doi.org/10.1155/2017/9308745.

[11] J. Xia, Y. Chen, A. Chen, Y. Chen, Medical image fusion based on sparse representation and pcnn in nsct domain, Comput. Math. Methods Med. 2018 (5) (2018) 1–12, http://dx.doi.org/10.1155/2017/9308745.

[12] M. Yin, X.N. Liu, Y. Liu, X. Chen, Medical image fusion with parameter-adaptive pulse coupled-neural network in nonsubsampled shearlet transform domain, IEEE Trans. Instrum. Meas. 68 (1) (2018) 49–64, http://dx.doi.org/10.1109/TIM.2018.2838778.

[13] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks, in: 2017 20th International Conference

on Information Fusion, Fusion, 2017, http://dx.doi.org/10.23919/ICIF.2017.8009769.

[14] H. Hermessi, O. Mourali, E. Zagrouba, Convolutional neural network based multimodal image fusion via similarity learning in the shearlet domain, Neural Comput. Appl. 30 (7) (2018) 2029–2045, http://dx.doi.org/10.1007/s00521-018-3441-1.

[15] S. Li, L. Fang, H. Yin, An efficient dictionary learning algorithm and its application to 3-D medical image denoising, IEEE Trans. Biomed. Eng. 59 (2) (2012) 417–427, http://dx.doi.org/10.1109/TBME.2011.2173935.

[16] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Inf. Fusion 13 (1) (2012) 10–19, http://dx.doi.org/10.1016/j.inffus.2010.04.001.

[17] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, Z. Gu, Voxel selection in FMRI data analysis based on sparse representation, IEEE Trans. Biomed. Eng. 56 (10) (2009) 2439–2451, http://dx.doi.org/10.1109/TBME.2009.2025866.

[18] G. Wang, Y. Bresler, V. Ntziachristos, Compressive sensing for biomedical imaging, IEEE Trans. Med. Imag. 30 (5) (2011) 1013–1016, http://dx.doi.org/10.1109/TMI.2011.2145070.

[19] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (6583) (1996) 607–609, http://dx.doi.org/10.1038/381607a0.

[20] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892, http://dx.doi.org/10.1109/tim.2009.2026612.

[21] M. Elad, I. Yavneh, A plurality of sparse representations is better than the sparsest one alone, IEEE Trans. Inform. Theory 55 (10) (2009) 4701–4714, http://dx.doi.org/10.1109/TIT.2009.2027565.

[22] D. Gabor, Theory.of. communication, Theory of communication part 1: the analysis of information, J. Inst. Electr. Eng. III, Radio Commun. Eng. 93 (26) (1946) 429–441, http://dx.doi.org/10.1049/ji-3-2.1946.0074.

[23] M. Ramamoorthy, K. AneesBarvin, Medical image fusion using gabor and gradient measurement, Int. J. Innovative Res. Sci., Eng. Technol. 3 (3) (2014) 2077–2083.

[24] F. Zhang, Y. Cen, R. Zhao, H. Wang, Y. Cen, L. Cui, S. Hu, Analytic separable dictionary learning based on oblique manifold, Neurocomputing 236 (C) (2017) 32–38, http://dx.doi.org/10.1016/j.neucom.2016.09.099.

[25] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322, http://dx.doi.org/10.1109/TSP.2006.881199.

[26] M. Seibert, J. Wörmann, R. Gribonval, M. Kleinsteuber, Separable cosparse analysis operator learning, Signal Process. Conf. (2014) 770–774.

[27] S. Hawe, M. Seibert, M. Kleinsteuber, Separable dictionary learning, IEEE Conf. Comput. Vision Pattern Recognit. (2013) 438–445, http://dx.doi.org/10.1109/CVPR.2013.63.

[28] C. He, Y.F. Zheng, S.C. Ahalt, Object tracking using the gabor wavelet transform and the golden section algorithm, IEEE Trans. Multimed. 4 (4) (2002) 528–538, http://dx.doi.org/10.1109/ROBOT.2001.932851.

[29] N. Dhengre, K.P. Upla, R.D. Trivedi, Multimodal biomedical image fusion: Use of log-Gabor and guided filters with non-subsampled contourlet transform, in: Third International Conference on Image Information Processing, vol. 1, 2015, pp. 6–11, http://dx.doi.org/10.1109/ICIIP.2015.7414710.

[30] Y. Liu, X. Chen, R. Ward, Z. Wang, Image fusion with convolutional sparse representation, IEEE Signal Process. Lett. 23 (12) (2016) 1882–1886, http://dx.doi.org/10.1109/LSP.2016.2618776.

[31] Y. Liu, X. Chen, R. Ward, Z. Wang, Medical image fusion via convolutional sparsity based morphological component analysis, IEEE Signal Process. Lett. 26 (3) (2019) 485–489, http://dx.doi.org/10.1109/LSP.2019.2895749.

[32] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inf. Fusion 24 (1) (2015) 147–164, http://dx.doi.org/10.1016/j.inffus.2014.09.004.

[33] P. G, H. H, A new quality metric for image fusion, Int. Conf. Image Process. 3 (2003) 173–176, http://dx.doi.org/10.1109/ICIP.2003.1247209.

[34] W. Z, B.A. C, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84, http://dx.doi.org/10.1109/97.995823.

[35] Q. G, Z. D, Y. P, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315, http://dx.doi.org/10.1049/el:20020212.