IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 10, 2019, accepted July 14, 2019, date of publication July 16, 2019, date of current version August 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929303

A Novel Image Fusion Framework Based on Sparse Representation and Pulse Coupled Neural Network

LI YIN[1], MINGYAO ZHENG[2,3], GUANQIU QI[4], ZHIQIN ZHU[3], FU JIN[2], AND JAESUNG SIM[5]

[1]Chongqing Key Laboratory of Translational Research for Cancer Metastasis and Individualized Treatment, Chongqing University Cancer Hospital and Chongqing Cancer Institute, Chongqing Cancer Hospital, Chongqing 400030, China
[2]Key Laboratory for Biorheological Science and Technology of Ministry of Education (Chongqing University), Chongqing University Cancer Hospital and Chongqing Cancer Institute, Chongqing Cancer Hospital, Chongqing 400044, China
[3]College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[4]Computer Information Systems Department, Buffalo State College, Buffalo, NY 14222, USA
[5]Department of Mathematics and Computer Information Science, Mansfield University of Pennsylvania, Mansfield, PA 16933, USA

Corresponding author: Fu Jin (jfazj@126.com)

ABSTRACT Image fusion techniques are applied to the synthesis of two or more images captured in the same scene to obtain a high-quality image. However, most of the existing fusion algorithms are aimed at single-mode images. To improve the fusion quality of multi-modal images, a novel multi-sensor image fusion framework based on non-subsampled shearlet transform (NSST) is proposed. First, the proposed solution uses NSST to decompose source images into high- and low-frequency components. Then, an improved pulse coupled neural network (PCNN) is proposed to process high-frequency components. Thus, the feature extraction effect of the high-frequency component is meliorated. After that, a sparse representation (SR) based measure, including compact dictionary learning and Max-L1 fusion rule, is designed to enhance the detailed features of the low-frequency component. Finally, the final image is obtained by the reconstruction of high- and low-frequency components via NSST inverse transformation. The proposed method is compared with several existing fusion methods. The experiment results show that the proposed algorithm outperforms other algorithms in both subjective and objective evaluation.

INDEX TERMS Multi-sensor fusion, NSST, PCNN, sparse representation, dictionary learning, image fusion.

I. INTRODUCTION

Image fusion techniques [36], [38] process multiple images from the same scene, extract useful information, and integrate them into a high-quality image. The integrated image is more informative than source images. Meanwhile, the corresponding visual effects are also improved for the observation in human visual system. Image fusion is commonly used in different areas, such as multi-focus image fusion [25], [49], infrared-visual image fusion [39], and medical image fusion [29], [42]. In recent years, image fusion has gradually become a hot research topic, and various image fusion algorithms have been proposed. These algorithms can be roughly divided into two categories: spatial and transform domain based algorithms [46].

The associate editor coordinating the review of this manuscript and approving it for publication was Anubha Gupta.

Spatial domain based algorithm can be classified into three types: pixel-based, block-based, and region-based algorithm [8], [20]. Pixel-based image fusion algorithm directly averages the pixel values of all source images. Susheela discussed and compared various pixel-based image fusion techniques [34]. According to the comparisons, the average algorithm was simple and fast, but it weakened the image contrast, blurred the edge information to a certain extent, and lost some important information in source images. In fact, block-based and region-based spatial domain algorithms are increasingly prevalent in recent years. A block-based image fusion method proposed by Li decomposed source images into several blocks first, then the features of each block were calculated individually [15]. However, a single-mode blocking may cause the block effect and lower the fusion quality. A lot of multi-mode block based image fusion algorithms

have been proposed since then. Bai proposed a quadtree-based multi-focus image fusion that used a weighted focus-measure as well as an effective quadtree decomposition strategy [3]. In the proposed solution, source images are decomposed into blocks with optimal sizes in a quadtree structure. According to the region shape, region-based algorithms segmented source images into different areas first, and then each segmented area was processed separately. A spatial-domain based multi-focus image fusion through gradient-based decision map construction and mathematical morphology was proposed [2]. The proposed algorithm used the weighted kernel of image gradient. Based on free boundary conditions, morphological operations and active contour model are applied to the boundary adjustment of focus and defocusing regions to get the real boundary. However, the data size of multi-focus source images is large. For spatial domain based algorithms, the process of data analysis is time-consuming. Furthermore, most of block-based and region-based algorithms are aimed at multi-focus image fusion. They are not good at processing other types of source image.

Transform domain based algorithms extract transform coefficients from source images first. Then, the extracted coefficients are processed according to some criteria, and inverted to the fused image. Transform domain based algorithms show the superiority among numerous image fusion algorithms. Sparse representation (SR) and multi-scale transformation (MST) are two typical transform domain based algorithms. With the rapid development of SR-based techniques, many SR-based algorithms have been proposed, and showed good performance [21], [28]. SR-based algorithms train an over-complete dictionary with a large number of multi-modality images first, and then use the sparse coefficients extracted from trained dictionary to represent each source image [30]. Finally, the fused image is obtained by combining maximum sparse coefficients with the trained dictionary. Aishwarya proposed an image fusion method that integrated discrete wavelet transform (DWT) [1]. High- and low-frequency subbands of source images were obtained by using DWT decomposition first, and then fused separately. It used SR to get the fused high-frequency subbands, in which more structural features of source images are preserved. However, redundancy and loss of residues are two main issues in SR-based methods. As another type of widely used method, there are many kinds of MST-based solution, such as wavelet transform (WT) [45], nonsubsampled contourlet transform (NSCT) [7], [19], [51] and nonsubsampled Shearlet transform (NSST) [33], etc. In most of MST-based fusion methods, source images are decomposed to obtain high- and low-frequency subbands first. Then the obtained subbands are merged following the selective fusion rules. Finally, a fused image is reconstructed by applying inverse MST to the merged subbands.

The mathematical properties of WT are similar to human visual system (HVS) model, so WT has been broadly applied to image processing. Zhang proposed a medical image fusion method based on wavelet theory that disassembled source images into different frequency subbands to preserve more information [45]. Classical WT can show resolution and locality, but cannot show directivity and anisotropy. NSCT can make up for the shortcomings of WT. The decomposition process of NSCT consists of multi-scale and multi-direction decomposition. Non-downsampling pyramid filter banks (NSPFB) and non-downsampling directional filter banks (NSDFB) are applied to multi-scale and multi-direction decomposition respectively. Yin proposed an image fusion algorithm by NSCT-PCNN (pulse coupled neural network) transform, which improves the information loss during the sub-sampling process [43]. However, nonsubsampled directional filter bank of NSCT exists in the restricted directions.

In order to improve the preservation of details and energy, this paper proposes a new image fusion framework for non-subsampled shearlet transform (NSST) based on the transform domain image fusion algorithm. PCNN model is used to distinguish the absolute values of high-frequency coefficients in different source images. According to the SR-based measure, such as principal component analysis (PCA) based dictionary learning, and Max-L1 fusion rule, the low-frequency fusion rule is formulated to enhance the detailed features. There are two main contributions of this paper as follows:

- PCNN model is used to distinguish the absolute values of high-frequency coefficients for the enhancement of high-frequency detail retention. The adaptive parameter simplification has a fast convergence speed and a small number of iterations, which can reduce the computational complexity of discrete PCNN model [11].
- It applies PCA-based dictionary learning to the activity level measurement of low-frequency coefficients. The proposed SR-based measure preserves more energies and details of source image, and also decreases the complexity of dictionary learning.

The remaining sections of this paper are structured as follows: Section II introduces NSST and PCNN; Section III discusses the proposed NSST-based multi-mode fusion framework; Section IV compares and analyzes the results of comparative experiments; and Section V concludes this paper.

II. RELATED WORK

A. NON-SUBSAMPLED SHEARLET TRANSFORM

As a multi-scale and multi-directional analysis method, NSST is able to obtain an optimal representation for source images in various directions and angles [18]. Being different from the traditional shearlet transform (ST), shift-invariance as an upgraded version is implemented in NSST. Therefore, the occurrence of ringing phenomenon in Gibbs type could be reduced in NSST. Additionally, NSST can also improve the computational efficiency of ST. As mentioned, NSST can effectively retain the spatial details and structural information of source images, so it is widely used in image fusion [37].

NSST achieves shift-invariant, multi-scale, and multi-direction by using non-subsampled pyramid filters (NSPFs) and shift-invariant shearlet filters(SFs). As shown in Fig. 1,

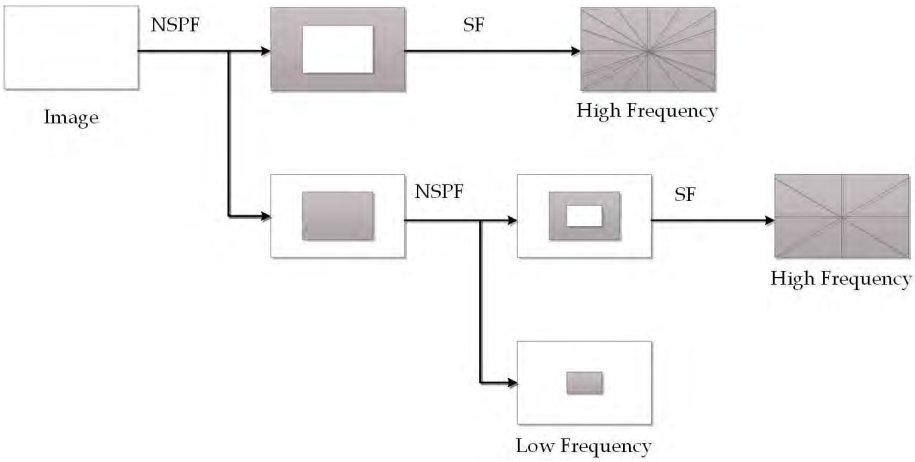


FIGURE 1. Two-level NSST Decomposition.

three same-size subbands can be obtained as input in two-level NSST decomposition, which include two high-frequency bands and one low-frequency band. NSPF is used to realize multi-scale decomposition in NSST. It removes down-samplers from laplacian pyramid first, and then carries on the corresponding up-sampling. In each decomposition level, source images are decomposed by NSPF to obtain a low- and high-pass component respectively first. Then, NSPF iterative decomposition is applied to the obtained low-frequency component to achieve multi-scale decomposition. For each NSPF, the ideal frequency support area for the kth stage low-pass filter is $[-(\pi/2^k)(\pi/2^k)]^2$.

Accordingly, the ideal support of equivalent high-pass filter is the complement of high-pass filter at the corresponding region $[-(\pi/2^{k-1}), (\pi/2^{k-1})]^2 \backslash [-(\pi/2^k), (\pi/2^k)]^2$. The equivalent filters of a k-level cascading NSPFB are given in Eq. 1,

$$H_1\left(Z^{2^{n-1}I}\right)\prod_{K=0}^{n-2} H_0\left(Z^{2^k I}\right), \quad 1 \le n \le K$$
$$\prod_{K=0}^{n-2} H_0\left(Z^{2^k I}\right), \quad n = K+1 \quad (1)$$

where $H_0(z)$ and $H_1(z)$ denote the low-pass filter and the corresponding high-pass filter respectively. As shown in Fig. 1, the multi-direction representations of high-pass components are obtained by SF. SF is mapped from pseudo-polarization grid system to Cartesian coordinate system in NSST. The whole process can be performed directly through the two-dimension convolution. It effectively avoids the down-sampling operation, and then realizes the shift-invariance of NSST. After applying SF to all m levels, the high-frequency subbands are obtained by NSPF. 2^m directional subband coefficients, that have the same size as source images, are generated. Thus, this process achieves the multi-direction decomposition of NSST. NSST, as a full multi-scale, multi-direction, and shift-invariant expansion, can reflect the features of source images well in image fusion. After the decomposition of NSPF and SF, source images can be decomposed into high- and low-pass bands that provide theoretical bases for subsequent image fusion.

B. SPARSE REPRESENTATION AND DICTIONARY LEARNING

1) SPARSE REPRESENTATION

As the main idea of sparse representation theory, a linear combination of the lowest possible number of atoms or transform bases in an over-complete dictionary is used to represent image signals. Since the number of atoms in an over-complete dictionary is greater than the signal dimension, it enables a flexible signal representation. In the definition, R is number field, $y \in R^n$ represents the vector of image signal in n dimension, and $\Phi \in R^{n\times m}$ is an $n \times m$ over-complete dictionary, in which every column vector can be treated as an atom of this over-complete dictionary. According to the sparse representation theory, the atoms that are the column vectors of an over-complete dictionary can represent corresponding image signal vector by linear combination, i.e., $y = \Phi\alpha$, where $\alpha \in R^m$ denotes a sparse coefficient vector ($m > n$). The sparse coefficient vector can be obtained by solving the optimization problem of Eq. 2:

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \ s.t. \ \|y - \Phi\alpha\|_2^2 < \varepsilon \quad (2)$$

where $\|\alpha\|_0$ is the count of non-zero items in α, and ε is the bounded error. The image signal is represented by using a superposition of the smallest possible number of atoms in the over-complete dictionary. The optimal solution of Eq. 2 can be obtained by greedy approximation methods, such as matching tracking (MP) [12], or orthogonal matching tracking (OMP) [22].

2) DICTIONARY LEARNING

The key issue of SR-based image reconstruction and fusion is the selection of over-complete dictionary.

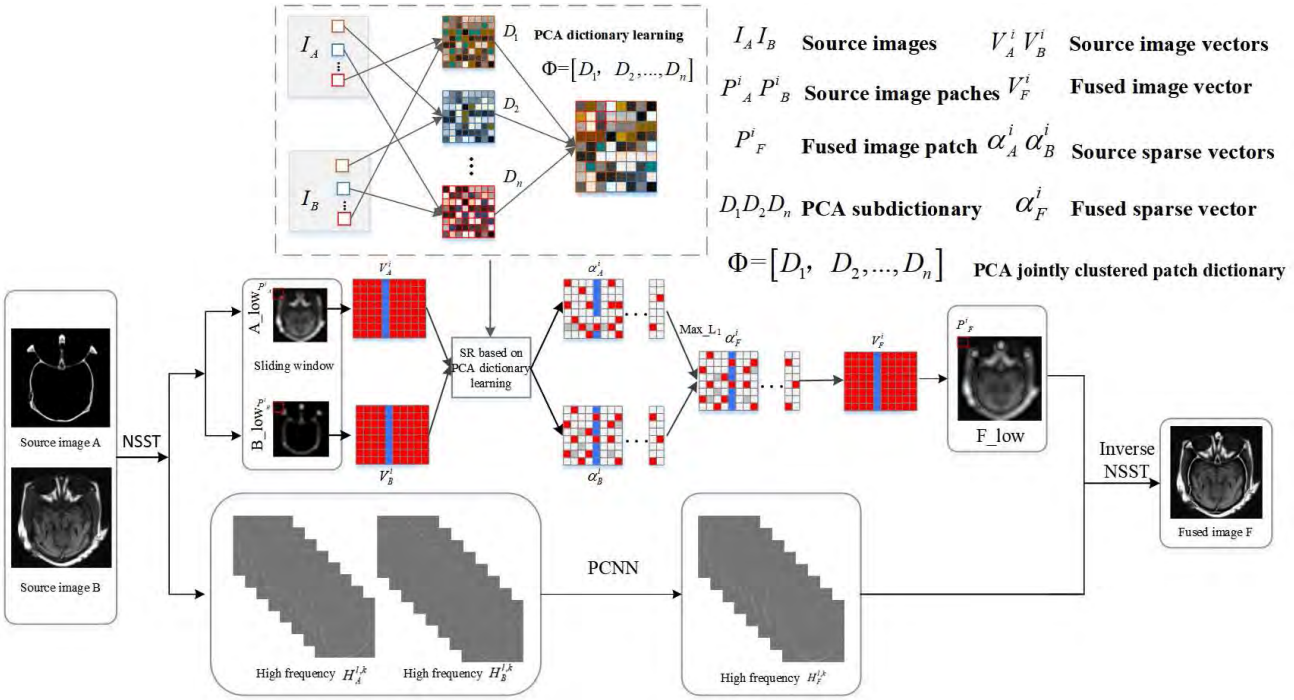


FIGURE 2. The Proposed Image Fusion Framework.

Mostly, the over-complete dictionary uses a fixed dictionary, such as DCT or wavelet transform. Although a fixed dictionary can be easily implemented, its performance is limited to the type of data and application. To obtain a non-restricted dictionary from input images, Aharon learned a K-SVD dictionary from a set of training images and adaptively updated it by related SVD operations [12]. K-SVD has better performance in many image reconstruction methods than other fixed dictionary based methods. Comparing to fixed dictionary based methods, K-SVD can extract patches from globally trained dictionaries or adaptively trained dictionaries. It makes K-SVD perform well in many image reconstruction methods.

However, K-SVD has a high computational complexity in practical applications. As an iterative learning process, K-SVD needs to process a large amount of training images. The high computational complexity limits the dimensions of K-SVD dictionary. Lessons learned from different dictionary learning methods and clustering methods are used as learning motivations. So a dictionary learning method based on image patch clustering is proposed for multi-modal image fusion. The proposed method clusters similar image patches in all source images to compose a complete dictionary. Trained sub-dictionaries can be combined to describe all multi-modal image signals by adding image blocks from all source images. In addition, K-SVD over-complete dictionary is highly structured. All multi-modal image signals can be described by using a learned dictionary, but it is redundant. It is necessary to reduce the computational complexity of image fusion by eliminating the redundancies in learned dictionary (reducing the dictionary size) [12], [19].

III. PROPOSED FRAMEWORK

A. OVERVIEW

The proposed image fusion framework is shown in Fig. 2, which includes four components: image decomposition, high- and low-pass band fusion, and image reconstruction. First source images are decomposed into high- and low-frequency bands by NSST. Then, PCNN model is used to realize the fusion of high-frequency coefficients. Next, according to the fusion rules of SR measure, the proposed solution integrates principal component analysis (PCA) based dictionary learning and Max-L1 fusion rule to achieve both energy preservation and detail extraction of low-frequency bands. Finally, the fused image is obtained by inverse transform of NSST.

B. FUSION RULE OF HIGH-FREQUENCY BANDS

1) IMPROVED PCNN

In order to reduce the computational complexity of discrete PCNN [11] in the fusion of high-frequency coefficients, a simplified PCNN model with automatic parameter settings was proposed [10], [43], which is described as follows:

$$F_{ij}[n] = S_{ij} \tag{3}$$

$$L_{ij}[n] = V_L \sum_{kl} W_{ijkl} Y_{kl}[n-1] \tag{4}$$

$$U_{ij}[n] = e^{-af} U_{ij}[n-1] + F_{ij}[n]\left(1 + \beta L_{ij}[n]\right) \tag{5}$$

$$Y_{ij}[n] = \begin{cases} 1, & if\ U_{ij}[n] > E_{ij}[n-1] \\ 0, & otherwise \end{cases} \tag{6}$$

$$E_{ij}[n] = e^{-a_c} E_{ij}[n-1] + V_E Y_{ij}[n] \tag{7}$$

where $F_{ij}[n]$ and $L_{ij}[n]$ are feeding input and linking input of neuron at position (i, j) in nth iteration respectively. V_L is a parameter that links the input magnitude. $U_{ij}[n]$ consists of two items. The fist item $e^{-af}U_{ij}[n-1]$ is the attenuation of its previous value, where af denotes an exponential decay coefficient. The section item $F_{ij}[n]\left(1+\beta L_{ij}[n]\right)$ denotes a nonlinear modulation of $F_{ij}[n]$ and $L_{ij}[n]$. $Y_{ij}[n]$ denotes the output model of PCNN, which has two statues: excited ($Y_{ij}[n]=1$) and unexcited ($Y_{ij}[n]=0$).

In this PCNN model, there are five parameters, such as V_L, af, β, a_c, and V_E. In order to simplify the parameter settings, the formula proposed in [5], [43] is used to calculate these parameters.

$$af = \log\left(1/\sigma\ (S)\right) \tag{8}$$

$$\lambda = \beta V_L = \frac{\left(S_{\max}/S'\right)-1}{6} \tag{9}$$

$$V_E = e^{-af} + 1 + 6\lambda \tag{10}$$

$$a_c = \ln\left(\frac{\frac{V_E}{S'}}{\frac{1-e^{-3af}}{1-e^{-af}} + 6\lambda e^{-af}}\right) \tag{11}$$

where $\sigma\ (S)$ denotes the standard deviation of input signal S. S' and S_{max} mean the normalized threshold of Otsu's method and the maximum intensity of input signal respectively.

2) FUSION OF HIGH-FREQUENCY BANDS

PCNN model is applied to the fusion of high-frequency bands. According to the PCNN model proposed in [5], [43], the absolute value of high-frequency band is used as the network input, that is $F_{ij}[n] = \left|H_S^{l,k}\right|, S \in \{A, B\}$. The activity level of high-frequency coefficient is iteratively measured by the total number of excitation. The firing time is accumulated by the following steps added to the end of each iteration.

$$T_{ij}(n) = T_{ij}(n-1) + Y_{ij}(n) \tag{12}$$

The excitation of each neuron is ordered as $T_{ij}(N)$ first, where N denotes the number of iterations [43], [44]. Then, PCNN excitation times of the high-frequency components of source images are calculated as: $H_A^{l,k}$ and $H_B^{l,k}$ respectively. Therefore, the fused high-frequency components can be obtained by using Eq. 22. A high-frequency coefficient with a large number of excitations is selected as the fusion coefficient [43]. The fusion algorithm of high-frequency components is shown in Alg. 1.

$$H_F^{l,k}(i,j) = \begin{cases} H_A^{l,k}(i,j), & \text{if } T_{A,ij}^{l,k}(N) \geq T_{B,ij}^{l,k}(N) \\ H_B^{l,k}(i,j), & \text{otherwise} \end{cases} \tag{13}$$

Algorithm 1 Fusion of High-frequency Bands

Input:
high-frequency bands $\left\{H_A^{l,k}, H_B^{l,k}\right\}$, the number of directions at each decomposition level $K(l), l \in [1, L]$, the number of PCNN N iterations

Output:
fused high-frequency band $H_F^{l,k}(i,j)$

Outer Loop
for $l = 1$ to L **do**
  for $k = 1$ to $K(l)$ **do**
    Initialize the PCNN model:$Y_{ij}[0]=0$, $U_{ij}[0]=0$, $E_{ij}[0]=0$, $T_{ij}[0]=0$ and $F_{ij}[n]=\left|H_S^{l,k}\right|, S \in \{A, B\}$.
    Estimate the PCNN parameters using Eq. 17 - 20.
    for $n = 1$ to N **do**
      Calculate the PCNN model by Eq. 3 - 16.
    end for
    Merge $H_A^{l,k}$ and $H_B^{l,k}$ using Eq. 4 to obtain $H_F^{l,k}(i,j)$.
  end for
end for

C. FUSION RULE OF LOW-FREQUENCY BANDS

1) JOINT PATCH CLUSTERING

The pixels of source image can be clustered by image features, such as local intensity, or gradient that describes the low-layer structure. In the proposed fusion algorithm, the local regression weights of steering kernel (SKR) are clustered as image features [12]. The vectorized steering kernel of the ith pixel of the kth source image can be defined by the steering covariance matrix as follows:

$$w_k^i = \left[\ldots w_k^{ij} \ldots\right]^T \tag{14}$$

$$w_k^{ij} = \frac{\sqrt{dec\left(c_j\right)}}{2\pi h_{steer}^2} \exp\left[-\frac{\left(x_k^i - x_k^j\right)^T c_j \left(x_k^i - x_k^j\right)}{2h_{steer}^2}\right], \quad for \quad j = 1, 2, \ldots, p \tag{15}$$

where x_k^i represents a spatial location of the ith pixel of I_k, h_{steer} represents a global smoothing parameter for adjusting the kernel support. c_j denotes the steering covariance matrix that can be estimated by the local gradient of neighboring patch centered at the jth pixel. p denotes the number of pixels in the kernel, and its value is decided by the feature vector w_k^i. The details of SKR feature extraction can be found in [12].

To jointly cluster all the pixels from k different source images, all the feature vectors are concatenated into a single feature matrix, i.e., $w = [w_1, w_2, \ldots, w_k]$, where $w_k = \left[\ldots w_k^i \ldots\right]$. K-means clustering is implemented by using this single feature matrix as input. By minimizing the l_2 distance of each feature vector to the class center, each low-frequency image can be segmented into mutually disjointed regions, as shown in Eq. 16:

$$I_k = \bigcup_{c=1}^{C} \{i | i \in \Omega_c\} \tag{16}$$

where Ω_c denotes the cth region, and c is the total number of clusters. Clustering algorithm is applied to the feature vector sets that consist of feature vectors obtained from different source images, so pixels with similar features are classified into the same cluster. A more informative dictionary that

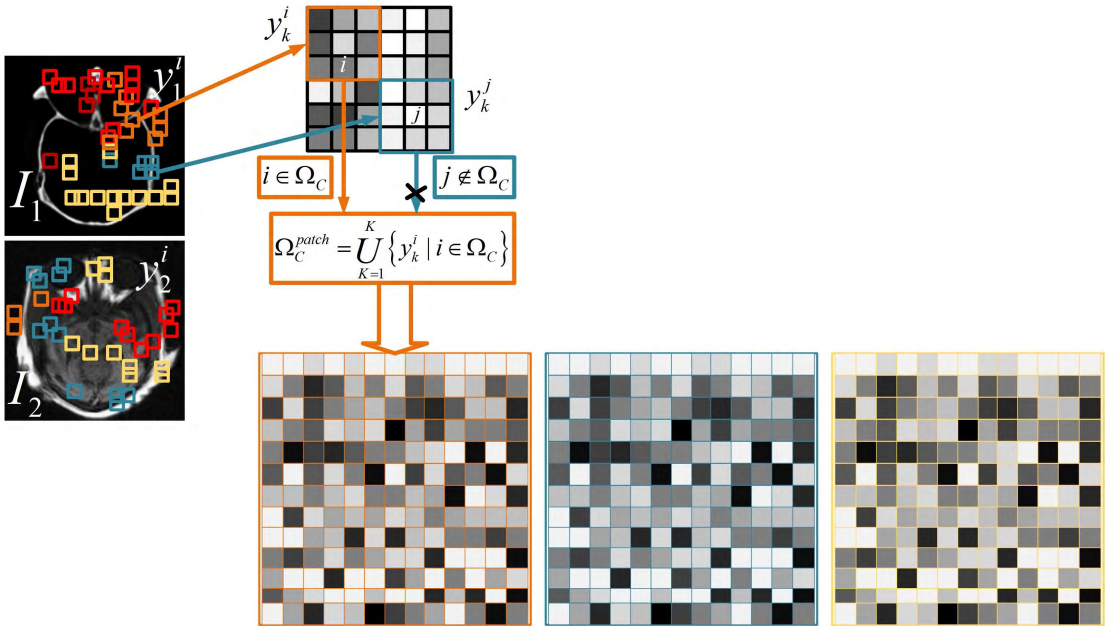


FIGURE 3. Generation of joint patch clusters.

contains all the image structure information of source images is learned by utilizing the joint clustering blocks from all source images. During the joint clustering, SKR can reflect the local structure features of source image. Therefore, SKR as an image feature is used in clustering. Eq. 14 and 15 calculate SKR eigenvector to obtain the feature matrix, and use K-mean algorithm for clustering. The center pixel refers to the center pixel of decomposed image block [12]. Therefore, it extracts all the central pixels that are classified into overlapping patches of the same cluster. Then all the patches are grouped to obtain joint patch clusters, as shown in Fig. 3. The generation of a joint patch cluster is shown as follows:

$$\Omega_C^{patch} = \bigcup_{k=1}^{K} \left\{ y_k^i | i \in \Omega_c \right\} \tag{17}$$

where Ω_C^{patch} denotes the cth joint patch cluster, y_k^i represents a patch from the kth source image, and its center pixel i belongs to Ω_c.

2) DICTIONARY LEARNING

Fig. 3 illustrates an over-complete dictionary built directly from a joint patch cluster. Although the over-complete dictionary contains more image patches with similar structures from source images, it may cause the redundancy in learning dictionary [12], [19]. So the dimensionality reduction is required for the redundant over-complete dictionary. Recently, many methods have been proposed to reduce the dimensionality of over-complete dictionary to obtain a compact dictionary [35], [40]. The proposed solution combines the main components of each joint patch cluster to learn a compact and informative dictionary. Since only a few PCA bases can approximate the patches in a same cluster well, the most useful top p principal components are selected to form a sub-dictionary [12], as shown in Eq. 18:

$$D_c = \left[d_1, d_2, \ldots, d_p \right], \ s.t. \ p = \arg\max_{p} \left\{ \sum_{j=p+1}^{n} L_j > \sigma \right\} \tag{18}$$

where D_c represents the sub-dictionary for cth cluster of p eigenvectors. L_j denotes the eigenvalue of the corresponding jth eigenvector d_j. Feature values are arranged in descending order, i.e., $L_1 > L_2 > \ldots > L_n > 0$. δ can be used to approximate p [12]. To avoid the over-fitting of image signal, δ is usually proportional to the signal dimension. However, if the δ value is too high, the number of atoms in the constructed sub-dictionary may not be sufficient. Image signals reconstructed using such a sub-dictionary are over-smoothed. Therefore, it is necessary to set the δ value that can use the minimum number of atoms to correctly represent the image signal. In the proposed fusion framework, it sets $\sigma = 0.95$. As shown in Fig. 4, the PCA base is selected as the atom of corresponding subsequence to describe the basic structure of each cluster. Once all the sub-regions of joint block clusters are obtained, they are merged into a single dictionary, as shown in Eq. 19:

$$\Phi = [D_1, D_2, \ldots, D_C] \tag{19}$$

The combined dictionary Φ in Eq. 19 is composed by the main components of clustered patch from source images, and also referred to an aggregated patch dictionary. Comparing with fixed DCT dictionary and k-SVD learned dictionary, principal component analysis (PCA) based dictionary proposed in this paper is more compact, but the corresponding patch cluster still contains the most informative components. Therefore, the proposed PCA-based dictionary can reduce

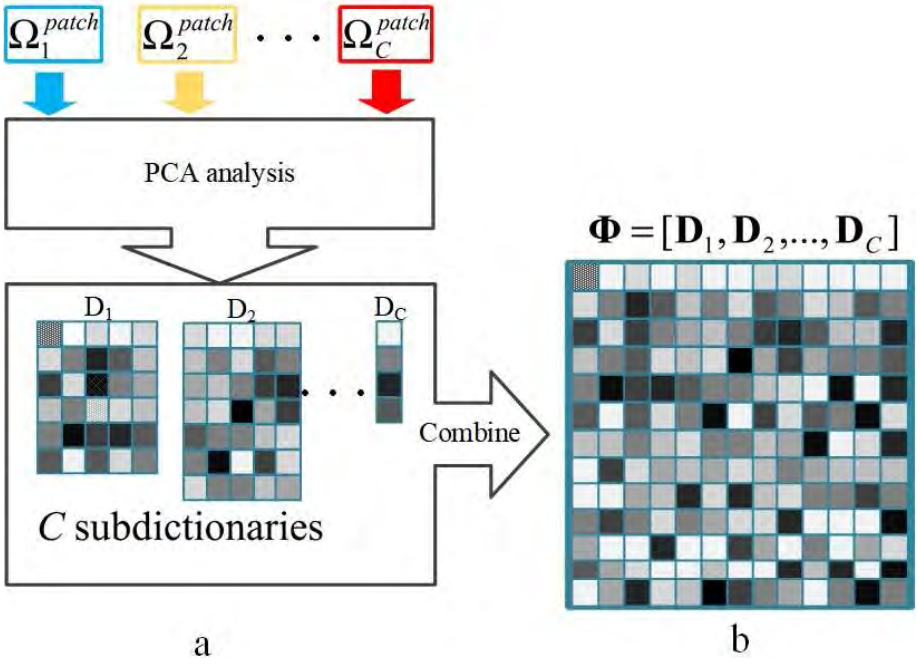


FIGURE 4. (a) Procedure of proposed dictionary learning method, (b) Jointly clustered patch dictionary.

computational complexity, when it achieves the excellent reconstruction performance.

3) FUSION OF LOW-FREQUENCY BANDS

The fusion strategy of low-frequency bands also has an important impact on the final fusion quality. To achieve the detail extraction and energy preservation of low-frequency images, an SR-based fusion method is applied to each low-pass NSST band.

Step 1: The low-frequency images L_A and L_B are decomposed into $\sqrt{n} \times \sqrt{n}$-size image blocks using a fixed-size sliding window. Suppose there are T patches denoted as $\{P_A^i\}_k^T$ and $\{P_B^i\}_k^T$ in L_A and L_B respectively.

Step 2 : For each position i, it rearranges $\{P_A^i, P_B^i\}$ into column vectors $\{v_A^i, v_B^i\}$. Then it normalizes each vector's mean value to zero to obtain $\{V_A^i, V_B^i\}$ by Eq. 20:

$$V_A^i = v_A^i - \bar{v}_A^i \cdot W \tag{20}$$

where W denotes an $n \times 1$ vector. $\bar{v}_A^i$ is the mean values of all the elements in v_A^i. V_B^i is obtained through the same way as V_A^i.

Step 3: It calculates the sparse coefficient vectors $\{\alpha_A^i, \alpha_B^i\}$ of $\{V_A^i, V_B^i\}$ using simultaneous orthogonal matching pursuit (SOMP) algorithm by Eq. 21:

$$\alpha_A^i = \arg\min_{\alpha_A^i} \|\alpha\|_0 \quad s.t. \left\| y_{A,L}^i - \Phi\alpha \right\|_2^2 < \varepsilon \tag{21}$$

where ε denotes the bounded representation error, $\|\alpha\|_0$ denotes the number of non-zero items in α, $y_{A,L}^i$ represents the low-frequency component L_A. Φ is the learned dictionary. Similarly, α_B^i can also be obtained by using Eq. 21.

Step 4: It merges α_A^i and α_B^i by Max-L1 rule to obtain the fused sparse coefficients (as shown in Eq. 22).

$$\alpha_F^i = \begin{cases} \alpha_A^i, & if \left\|\alpha_A^i\right\|_1 > \left\|\alpha_B^i\right\|_1 \\ \alpha_B^i, & otherwise \end{cases} \tag{22}$$

The fused result of v_A^i and v_B^i is calculated by Eq. 23.

$$V_F^i = \Phi\alpha_F^i + \bar{v}_F^i \cdot W \tag{23}$$

where the merged mean value $\bar{v}_F^i$ is obtained by Eq. 24.

$$\bar{v}_F^i = \begin{cases} \bar{v}_A^i & \text{if} \quad \alpha_F^i = \alpha_A^i \\ \bar{v}_B^i & \text{otherwise} \end{cases} \tag{24}$$

Step 5: All the original image patches in $\{P_A^i\}_k^T$ and $\{P_B^i\}_k^T$ are iterated according to the above Step 2-4 processes to obtain all the fused vectors $\{V_A^i\}_{i=1}^T$. For each V_F^i, it is reshaped into an $\sqrt{n} \times \sqrt{n}$ size patch P_F^i. Then P_F^i is plugged into its original position in L_F. The low-pass fused result L_F is obtained by averaging the accumulation times of each pixel's value in L_F. The fusion algorithm of low-frequency components is shown in Alg. 2.

Once the low- and high-frequency fused band L_F and H_F are obtained, it applies the corresponding inverse NSST to L_F and H_F to reconstruct the final fused image I_F. The main steps of proposed medical image fusion method are summarized in Alg. 3.

Algorithm 2 Fusion of Low-frequency Bands

Input:

low-frequency bands $\{L_A, l_B\}$, the number of low-pass sunbands patches T.

Output:

fused low-frequency band L_F.

Outer Loop
for each image **do**
  for $j = 1$ to T **do**
    Extract image features to obtain eigenvector d_j by Eq. 14 and 15.
    Cluster feature vectors using PCA to obtain sub-dictionary D_C by Eq. 18.
  end for
  Aggregate all sub-dictionaries using Eq. 19 to obtain a combined dictionary Φ.
end for
for each low-frequency image **do**
  Use sliding window decomposition to get patches $\left\{P_A^i\right\}_k^T$ and $\left\{P_B^i\right\}_k^T$
end for
for $i = 1$ to T **do**
  Rearrange $\left\{P_A^i, P_B^i\right\}$ into column vectors $\left\{v_A^i, v_B^i\right\}$.
  Calculate the sparse coefficient vectors $\left\{\alpha_A^i, \alpha_B^i\right\}$ using SOMP algorithm by Eq. 21.
  Merge α_A^i and α_B^i by Max-L1 rule to obtain the fused sparse coefficients.
  The fused result of v_F^i is calculated by Eq. 23.
  Reshape v_F^i to obtain P_F^i.
end for
Merge $\left\{P_F^i\right\}_{i=1}^T$ to obtain fused low-frequency band L_F.

Algorithm 3 Proposed NSST-Based Multi-Sensor Image Fusion Framework

Input:

the number of NSST decomposition levels L, the number of directions at each decomposition level $K(l), l \in [1, L]$, source images $\left[I_A\ I_B\right]$.

Output:

fused image I_F.

Outer Loop
for each image **do**
  Perform NSST decomposition on $\left[I_A\ I_B\right]$ to obtain high- and low-frequency components of image $\left\{H_A^{l,k}, H_B^{l,k}\right\}$ and $\{L_A, L_B\}$.
end for
Fuse low-frequency bands by Alg. 2.
Fuse high-frequency bands by Alg. 1.
Obtain a fused image by inverse NSST I_F $\xleftarrow{\text{inverse NSST}}$ $\left\{L_F, H_F^{l,k}(i,j)\right\}$.

IV. EXPERIMENTS AND ANALYSES

A. EXPERIMENT PREPARATION

It uses 104 pairs of medical images, 10 pairs of multi-focus images, 10 pairs of infrared-visible images and 23 pairs of Panchromatic and Multi-spectral Satellite Images in comparative experiments to test the fusion performance. The resolutions of test images are 256×256, 240×320 and 512×512 respectively. The proposed fusion algorithm trains all the source images to obtain a PCA-based dictionary. Medical image pairs are obtained from http://www.med.harvard.edu/aanlib/home.html. Infrared-visible and gray-level multi-focus image pairs were collected by Liu *et al.* [22] and can be downloaded from quxiaobo.org. All the experiments are run in Matlab 2014a on an Intel(R) Core(TM)i7-7700k CPU @ 4.20GHz desktop with 16.00 GB RAM.

1) OBJECTIVE EVALUATION METRICS

As the reference image (ground truth) does not exist in practice, it is not an easy task to quantitatively evaluate the quality of a fused image. In recent years, many image fusion metrics have been proposed. But none of them can be universally applied to any fusion scenario. It usually needs several metrics to make a comprehensive evaluation. In this paper, eight popular metrics are employed to quantitatively evaluate the performance of different fusion methods, which are Q^{TE} [6], [24], [26], Q^{IE} [24], [41], $Q^{\mathrm{AB/F}}$ [27], [50], Q^{P} [24], [47], Q^{MI} [24], Q^{Y} [24], [41], Q^{CB} [4], [24], and Q^{VIF} [32], [50].

Q^{TE} [6], [24], [26] and Q^{IE} [24], [41] evaluate the Tsallis entropy and nonlinear correlation information entropy (NCIE) of fused image respectively. $Q^{\mathrm{AB/F}}$ [27], [50] and Q^{P} [24], [47] are used to measure the edge information. $Q^{\mathrm{AB/F}}$ [27], [50] is a gradient-based quality index. Q^{P} [24], [47] is a phase congruency based image fusion metric. Q^{MI} [24] and Q^{Y} [24], [41] measure the similarity between the fused image and source images. Q^{MI} [24]is a quantitative measure of the mutual dependence of two variables. Q^{Y} [24], [41] is a structural similarity based metric for fusion assessment. Q^{CB} [4], [24] and Q^{VIF} [32], [50] evaluate the human visualization performance of fused images. Q^{CB} [4], [24] is a human perception inspired quality metric for image fusion. Q^{VIF} [32], [50] is the ratio between distorted test image information and referencing image information.

B. EXPERIMENT RESULTS OF SIX POPULAR FUSION METHODS

In this section, six popular fusion methods are used to compare with the proposed NSST-based fusion framework. They are adaptive spare representation (ASR) based image fusion framework proposed by Liu and Wang [23], multi-channel medical image fusion (CT) method proposed by Zhu *et al.* [50], multi-modal image fusion framework integrated joint patch clustering based dictionary learning (KIM) proposed by Kim *et al.* [12], multi-scale transform (NSCT) method proposed by Liu *et al.* [22], MST- and SR-based (MST-SR) fusion framework proposed by Liu *et al.* [22], and image fusion with parameter-adaptive pulse coupled neural

network(PA-PCNN) proposed by Yin *et al.* [43]. Fusion results obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed methods are shown in Fig. 5 - 13. In the comparative experiments, the fusion time of ASR, CT, MST-SR and the proposed method does not include the time of dictionary training. Since KIM used joint patch clustering based dictionary learning, its fusion time includes the dictionary learning time. NSCT and PA-PCNN applied multi-scale decomposition to image fusion rather than sparse representation based dictionary learning, so it has a short processing time.

1) EXPERIMENT RESULTS OF MEDICAL IMAGES

Modern medical images, such as computed tomography (CT), magnetic resonance (MRI), single photon emission tomography (SPECT), positron emission tomography (PET), reflect human body information from different angles [48], [50]. However, a single image cannot provide enough comprehensive information for medical diagnosis. Due to the insufficient information, even if a doctor is good at spatial imagination and corresponding inferences, the diagnosis results may not be reliable. Medical image fusion techniques solve this issue. An informative image can be obtained by the fusion of two or more medical images from the same scene. It can effectively reduce the uncertainty and unreliability in clinical diagnosis and treatment [13].

Figure 5, 6, and 7 are three examples of medical image fusion. As shown in Figure 5, both CT and KIM obtain a fused image with high edge brightness. The fused image obtained by NSCT has a poor contrast. Compared with MST-SR, the fused images obtained by the proposed method and PA-PCNN method have moderate brightness and clear edge details. In Figure 6, the fused image obtained by KIM is blurred, and has poor performance in human visual system.

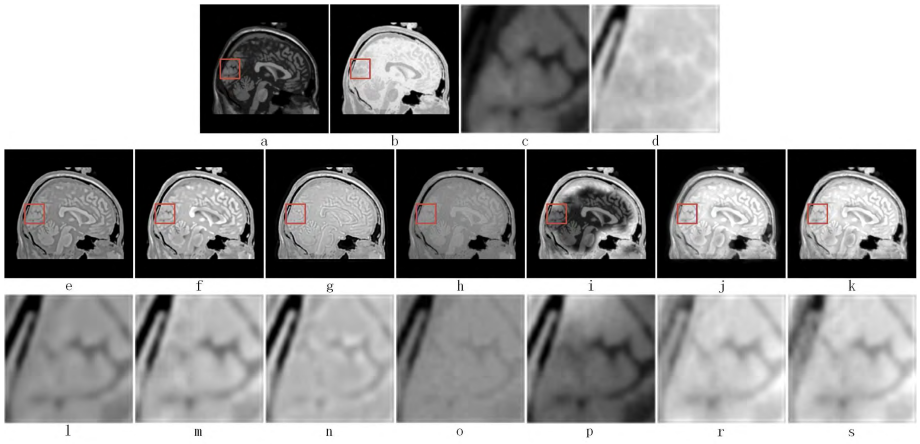


FIGURE 5. Medical Image Fusion Experiments of CT-MRI. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

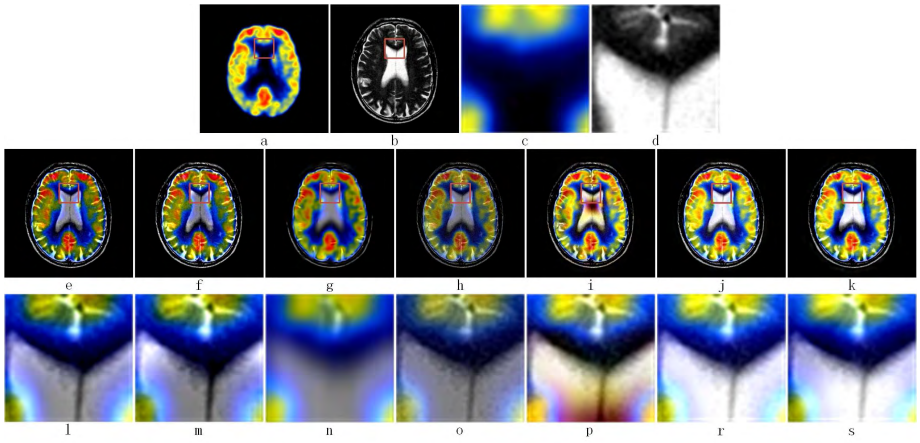


FIGURE 6. Medical Image Fusion Experiments of MRI-PET. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

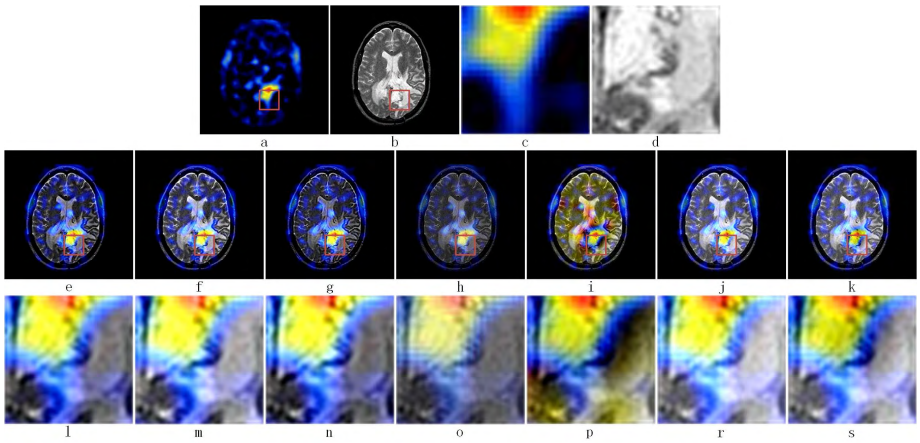


FIGURE 7. Medical Image Fusion Experiments of MRI-SPECT. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

TABLE 1. Objective evaluations of medical image fusion experiments.

	Q^{TE}	Q^{IE}	$Q^{AB/F}$	Q^{P}	Q^{MI}	Q^{Y}	Q^{CB}	Q^{VIF}	time
ASR	0.8179	0.8058	0.2907	0.5659	1.6756	0.3345	0.7598	0.3024	25.1233s
CT	0.7503	0.8061	0.3337	0.5432	2.0960	0.3838	0.7206	0.3140	25.8972s
KIM	0.8020	0.8057	0.2953	0.5530	1.6504	0.3397	**0.7634**	0.2925	44.2311s
NSCT	0.7307	0.8055	0.3108	0.4514	1.6451	0.3547	0.6919	0.2752	**7.5721s**
MST-SR	0.7584	0.8057	0.3048	0.4795	1.6701	0.359	0.6518	0.2951	9.1670s
PA-PCNN	0.7612	0.8059	0.3231	0.5004	1.7155	0.3643	0.7036	0.3033	12.5014s
Proposed	**0.8256**	**0.8092**	**0.5495**	**0.5989**	**2.4399**	**0.5150**	0.5498	**0.4534**	8.6580s

The image fused by NSCT has low brightness, and the one obtained by MST-SR has high brightness. In comparison with ASR and CT, the proposed methodnd PA-PCNN method have better performance in detail preservation. As shown in Figure 7, the image obtained by CT is bright. The details of fused images obtained by ASR and KIM are poorly preserved. The image fused by NSCT has weak contrast, and the color of the image obtained by MST-SR is distorted. Compared with other fusion methods, the proposed method and PA-PCNN method are more conducive to observation. Shown as Figure5, 6, and 7, the proposed method and PA-PCNN method perform excellent in detail preservation and visual effect. It is difficult to distinguish the difference between the proposed method and PA-PCNN method from the human visual system.

Table 1 lists the objective assessments of seven fusion methods on medical images. The proposed method has good performance in seven evaluation metrics, including Q^{TE}, Q^{IE}, $Q^{AB/F}$, Q^{P}, Q^{MI}, Q^{Y}, and Q^{VIF}. Since the image fused by proposed method is similar to the original image, the proposed method performs well on the preservation of structured and detailed information. KIM has the highest score on Q^{CB} that means it has the best performance in human perception among all seven fusion methods. ASR has the poor performance on all metrics. The human visual effect is poor on KIM. In addition, the proposed method spends the shortest processing time. Therefore, the proposed method has the best overall performance on medical image fusion.

2) EXPERIMENT RESULTS OF MULTI-FOCUS IMAGES

Due to the limited focus range of optical system, it is difficult to focus all the objects in the same scene by one shot. When an object is in the focus range, it can be clearly displayed, otherwise it is blurred. In this way, it is difficult to get a comprehensive and in-depth understanding of a scene through an image. Multi-focus image fusion, as an important branch of image fusion, merges multiple images with different focal lengths into one image. It can retain the clear part of each image to compose an informative image for subsequent image processing [14], [17].

Two multi-focus image fusion examples are displayed in Figure 8 and 9. As shown in Figure 8, the fused image obtained by CT achieves good visibility, but has poor performance in the detailed texture of leaves. The edge details of images fused by ASR and NSCT are not clear. In Figure 9, the results of ASR and CT show poor performance in detail preservation. Although the fused image obtained by KIM is bright, the detailed information is weak. Compared to NSCT method, the proposed method obtains

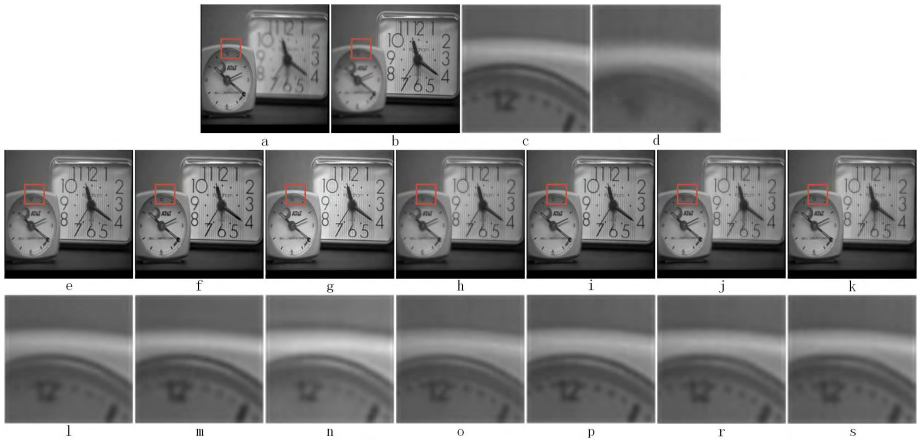


FIGURE 8. Multi-focus Image Fusion Experiments - 1. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

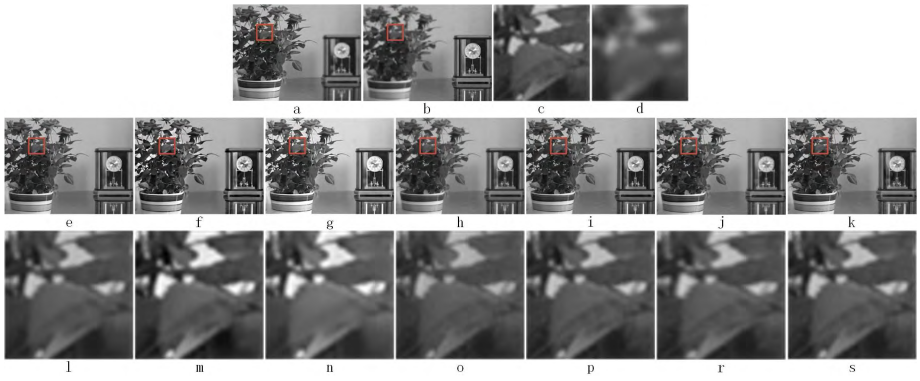


FIGURE 9. Multi-focus Image Fusion Experiments - 2. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

TABLE 2. Objective evaluations of multi-focus image fusion experiments.

	Q^{TE}	Q^{IE}	$Q^{AB/F}$	Q^{P}	Q^{MI}	Q^{Y}	Q^{CB}	Q^{VIF}	time
ASR	0.9044	0.8271	0.7445	**0.8938**	4.6238	0.8300	0.6816	0.7637	17.3716s
CT	0.7982	0.8244	0.5719	0.7156	4.3370	0.7772	0.6048	0.6603	25.7182s
KIM	0.7836	0.8226	0.6794	0.8258	4.0904	0.7909	0.6288	0.7636	61.3477s
NSCT	0.8875	0.8261	**0.8401**	0.8600	4.5298	0.9115	0.6917	0.7591	11.6990s
MST-SR	0.8102	0.8224	0.7031	0.8652	4.2181	0.8917	**0.7478**	0.7308	13.7548s
PA-PCNN	0.9049	0.8278	0.7122	0.7997	4.5787	0.8687	0.6932	0.7963	12.3593s
Proposed	**0.9871**	**0.8319**	0.7200	0.8263	**4.9980**	**0.9129**	0.7104	**0.8028**	**9.7779s**

a fused image that has clear edge information. The fused images of MST-SR, PA-PCNN and the proposed method have good performance in human visual system. It is difficult to figure out the difference between three fused images by human eyes. Therefore, objective metrics are used to measure the fusion performance.

As shown in Table 2, compared with other methods, the proposed method performs well in several evaluation metrics, including Q^{TE}, Q^{IE}, Q^{MI}, Q^{Y}, and Q^{VIF}. It indicates that the proposed method performs well in Tsallis entropy and nonlinear correlation entropy. For Q^{P}, ASR has good performance on phase consistency. NSCT has the best performance in $Q^{AB/F}$, which means it performs well in quality index. MST-SR has a good human perception in terms of Q^{CB}. Although the proposed method has slightly weaker performance than other methods in some metrics, it has good

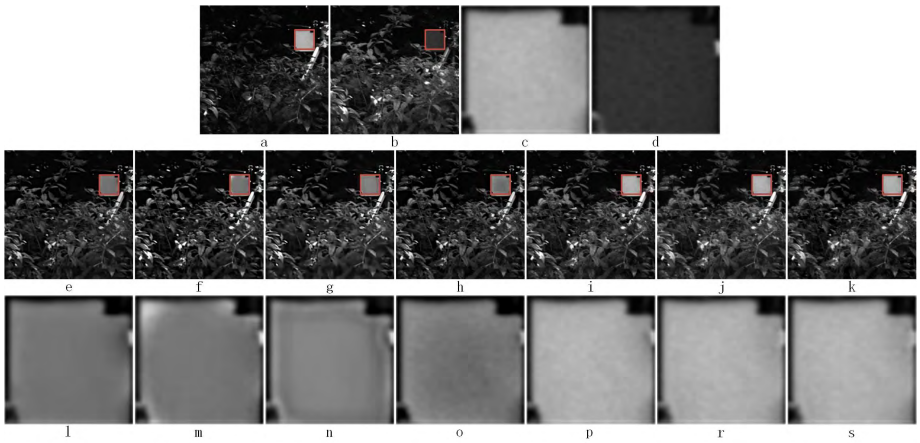


FIGURE 10. Infrared-visible Image Fusion Experiments -1. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

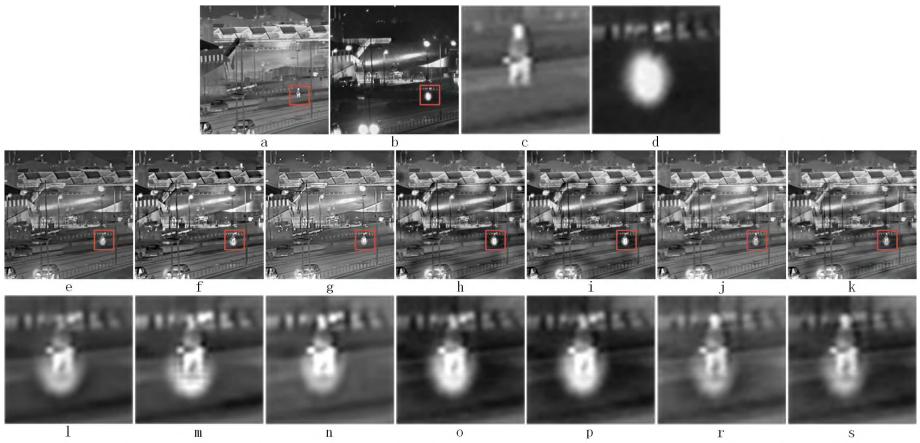


FIGURE 11. Infrared-visible Image Fusion Experiments -2. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

performance in the preservation of information and details as well as visualization. Moreover, the proposed method has the shortest processing time.

3) EXPERIMENT RESULTS OF INFRARED-VISIBLE IMAGES

Infrared images can well identify the target. Visible light images can provide good details of the scene, in which the target is located. Image fusion techniques make the full use of complementary information and spatio-temporal correlation of visible and infrared images to achieve engineering requirements. High-quality and comprehensive information can be obtained by integrating the corresponding information from multiple images. At present, image fusion techniques have been widely used in battlefield evaluation, target recognition, and other related fields [16].

Figure 10 and 11 show two examples of infrared-visible image fusion. In Figure 10, the edge details of images obtained by ASR and CT are poor. The image fused by KIM has high brightness, but low contrast. The image obtained by NSCT has high contrast, but its edge details are blurred. Compared with MST-SR, the proposed method has better performance in detail preservation. In addition, pedestrian details show that the proposed method performs well in detail contrast and brightness. In Figure 11, ASR method has poor performance in the preservation of image details. The edge brightness of images obtained by CT and KIM is high. Compared with NSCT, the image fused by the proposed method has higher brightness and contrast. The local contrast of fused image obtained by MST-SR is lower than that obtained by PA-PCNN and Proposed. In addition, pedestrian details show that the proposed method and PA-PCNN perform well in detail contrast and brightness. It is difficult to distinguish the difference between the two fusion results by human visual observation.

TABLE 3. Objective evaluations of infrared-visible image fusion experiments.

	Q^{TE}	Q^{IE}	$Q^{AB/F}$	Q^{P}	Q^{MI}	Q^{Y}	Q^{CB}	Q^{VIF}	time
ASR	0.5859	0.8099	0.6850	**0.7808**	2.6183	0.7823	0.6464	0.5518	19.6469s
CT	0.5121	0.8083	0.6618	0.7361	2.2963	0.7768	0.6041	0.5547	26.1473s
KIM	0.5343	0.8089	0.6228	0.5805	2.4381	0.7449	0.6386	0.5498	60.2463s
NSCT	0.5528	0.8092	0.7585	0.7795	2.4808	0.8517	0.6407	0.5861	10.7854s
MST-SR	0.5391	0.8093	0.7602	0.7153	2.4713	0.8698	0.6456	0.5806	12.3094s
PA-PCNN	0.5026	0.8096	0.7290	0.5894	2.4308	0.8477	0.6081	0.4841	12.5042s
Proposed	**0.5871**	**0.8103**	**0.7959**	0.7209	**2.6897**	**0.8991**	**0.6597**	**0.5947**	**9.0942s**

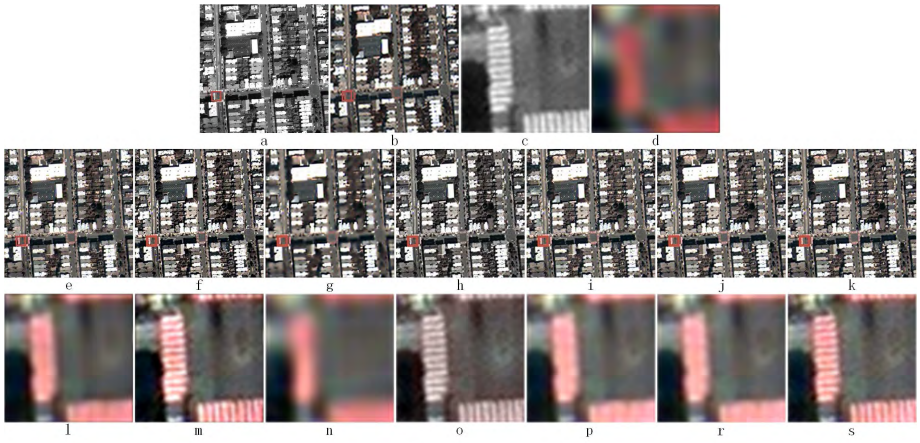


FIGURE 12. Panchromatic and Multi-spectral Satellite Images Fusion Experiments -1. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

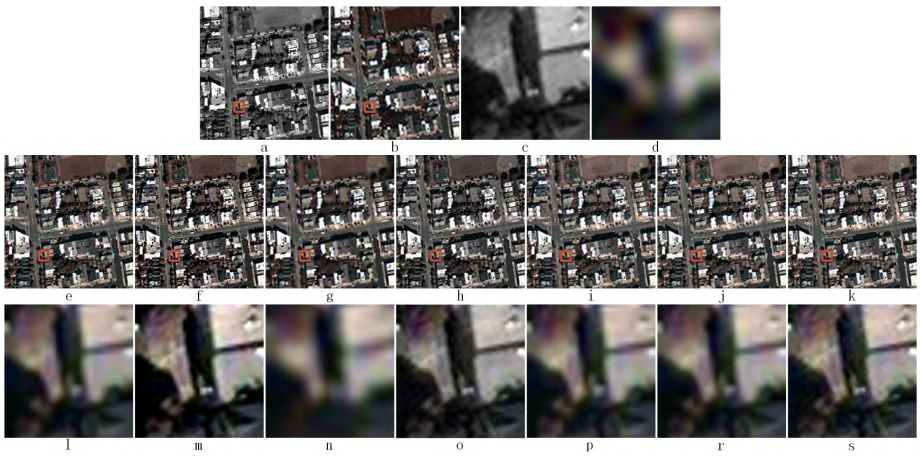


FIGURE 13. Panchromatic and Multi-spectral Satellite Images Fusion Experiments -2. (a) and (b) are two source images. (e), (f), (g),(h), (i), (j)and(k)are the fused images obtained by ASR, CT, KIM, NSCT, MST-SR, PA-PCNN and the proposed method respectively. (c), (d), (l), (m), (n), (o), (p), (q), (r) and (s)correspond to the local amplified images of (a), (b), (e), (f), (g),(h), (i), (j)and(k)respectively.

Table 3 shows the objective assessments of seven fusion methods on infrared-visible images. The image obtained by ASR has the highest Q^P, which is consistent with the high phase of original image. The proposed method has the highest value on other metrics, especially in Q^{IE}, $Q^{AB/F}$, Q^{MI}, Q^Y and Q^{VIF}. So the proposed method has high structural similarity with original image. At the same time, it is also prominent in edge preservation. In addition, the proposed method takes much less time in fusion process than other methods. Therefore, the proposed method is superior to other methods in infrared-visible image fusion.

TABLE 4. Objective evaluations of panchromatic and multi-spectral satellite images fusion Experiments.

	Q^{TE}	Q^{IE}	$Q^{AB/F}$	Q^{P}	Q^{MI}	Q^{Y}	Q^{CB}	Q^{VIF}	time
ASR	0.5095	0.8102	0.6488	0.5148	2.6840	0.8022	0.6913	**0.6437**	14.0568s
CT	0.5188	0.8111	0.6956	0.6174	2.6907	0.8846	0.6061	0.5852	24.7157s
KIM	0.4604	0.8087	0.4358	0.3522	2.4172	0.6588	0.5523	0.4939	46.5652s
NSCT	0.5513	0.8118	0.7978	0.6279	2.8994	0.9332	**0.6980**	0.6039	**2.4987s**
MST-SR	0.5042	0.8101	0.6961	0.5114	2.6561	0.8278	0.6896	0.6071	12.4256s
PA-PCNN	0.5171	0.8105	0.7031	0.5301	2.7249	0.8204	0.6883	0.6209	6.5809s
Proposed	**0.5737**	**0.8133**	**0.8020**	**0.6426**	**3.0243**	**0.9406**	0.6912	0.6045	5.9580s

4) EXPERIMENT RESULTS OF PANCHROMATIC AND MULTI-SPECTRAL SATELLITE IMAGES

There is an inverse relationship between spatial and spectral resolution in optical remote sensing [9]. Earth observation satellites provide MS images with high spatial resolution and panchromatic images with low spatial resolution. The panchromatic and multi-spectral satellite images fusion method can effectively integrate the spatial details of panchromatic image and the spectral details of multi-spectral image, thereby changing the detection and map update [31].

Figures 12 and 13 are two examples of full-color and multi-spectral satellite image fusion. It can be seen from the two figures that the energy is lost during the fusion process of ASR, KIM, MST-SR and PA-PCNN, so that the fusion results have different degrees of detail blur. Especially in Figures 13(n), it can be seen that the building texture of the KIM fusion result is lost. CT, NSCT, and the proposed method details are well preserved. However, as can be seen from the partial enlargement, the saturation of the NSCT fusion results is lower than other methods. Furthermore, the local contrast of the CT fusion results is lower than the proposed method.

Table 4 shows the objective assessments of seven fusion methods on panchromatic and multi-spectral satellite images. The images obtained by ASR and MST have the highest Q^{VIF} and Q^{CB} respectively, which indicated that the best performance in human perception among all seven fusion methods. The proposed method has the highest value on other metrics, especially in Q^{TE}, Q^{IE}, $Q^{AB/F}$,Q^{P}, Q^{MI} and Q^{Y}. So the proposed method has high structural similarity with original image. At the same time, it is also prominent in edge preservation. Although, the Q^{VIF} and Q^{CB} scores of the proposed method ranked third in all seven methods, the overall performance of the fusion method proposed in this paper is better than other methods in panchromatic and multi-spectral satellite images fusion.

According to the above experiments, the effectiveness of proposed image fusion framework is confirmed by subjective visual effects and objective evaluation metrics.

V. CONCLUSION AND DISCUSSION

A general image fusion method based on non-subsampled shearlet transform (NSST) is proposed. The fusion framework integrates NSST, pulse coupled neural network (PCNN), and sparse representation (SR) based measure to improve the visual quality of fused images. Specifically, the framework applies NSST to achieving high- and low-frequency image decomposition. The high-frequency coefficients are fused by PCNN. In the improved PCNN model, all the parameters can be adaptively calculated according to input frequency bands, so it has a faster convergence speed. According to SR-based methods, such as principal component analysis (PCA) based dictionary learning and Max-L1 fusion rule, the low-frequency fusion rule is formulated to enhance the detailed features. The energy and details of source images can be preserved well by the low-frequency fusion rule. Meanwhile, it also reduces the complexity of learned dictionary. Finally, the fused image is obtained by inverse transform of NSST.

Experiment results show that the proposed fusion framework is superior to KIM and MST fusion methods in terms of human visual perception and objective evaluation metrics. Experiment results also indicate that the proposed fusion method is superior in multi-modal medical image fusion. In future, the parameter selection strategy of PCNN model in the proposed method will be improved to further enhance the algorithm efficiency and extend its usage in medical applications.

REFERENCES

[1] N. Aishwarya, S. Abirami, and R. Amutha, "Multifocus image fusion using discrete wavelet transform and sparse representation," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw.*, 2016, pp. 2377–2382.

[2] X. Bai, M. Liu, Z. Chen, P. Wang, and Y. Zhang, "Multi-focus image fusion through gradient- based decision map construction and mathematical morphology," *IEEE Access*, vol. 4, pp. 4749–4760, 2016.

[3] X. Bai, Y. Zhang, F. Zhou, and B. Xue, "Quadtree-based multi-focus image fusion using a weighted focus-measure," *Inf. Fusion*, vol. 22, pp. 105–118, Mar. 2015.

[4] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," in *Proc. Conf. Inf. Sci. Syst. (CISS)*, 2009, pp. 518–523.

[5] Y. Chen, S. K. Park, Y. Ma, and R. Ala, "A new automatic parameter setting method of a simplified PCNN for image segmentation," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 880–892, Jun. 2011.

[6] N. Cvejic, C. N. Canagarajah, and D. R. Bull, "Image fusion metric based on mutual information and Tsallis entropy," *Electron. Lett.*, vol. 42, no. 11, pp. 626–627, May 2006.

[7] S. Das and M. K. Kundu, "NSCT-based multimodal medical image fusion using pulse-coupled neural network and modified spatial frequency," *Med. Biol. Eng. Comput.*, vol. 50, no. 10, pp. 1105–1114, 2012.

[8] I. De and B. Chanda, "Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure," *Inf. Fusion*, vol. 14, no. 2, pp. 136–146, 2013.

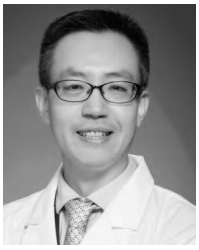
[9] M. González-Audícana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004.
[10] J. L. Johnson, "Pulse-coupled neural nets: Translation, rotation, scale, distortion, and intensity signal invariance for images," *Appl. Opt.*, vol. 33, no. 26, p. 6239, 1994.
[11] J. L. Johnson, M. L. Padgett, and O. Omidvar, "Guest editorial overview of pulse coupled neural network (PCNN) special issue," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 461–463, May 1999.
[12] M. Kim, D. K. Han, and H. Ko, "Joint patch clustering-based dictionary learning for multimodal image fusion," *Inf. Fusion*, vol. 27, pp. 198–214, Jan. 2016.
[13] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, Jul. 2018.
[14] H. Li, X. Li, Z. Yu, and C. Mao, "Multifocus image fusion by combining with mixed-order structure tensors and multiscale neighborhood," *Inf. Sci.*, vols. 349–350, pp. 25–49, Jul. 2016.
[15] H. Li, H. Qiu, Z. Yu, and B. Li, "Multifocus image fusion via fixed window technique of multiscale images and non-local means filtering," *Signal Process.*, vol. 138, pp. 71–85, Sep. 2017.
[16] H. Li, H. Qiu, Z. Yu, and Y. Zhang, "Infrared and visible image fusion scheme based on NSCT and low-level visual features," *Infr. Phys. Technol.*, vol. 76, pp. 174–184, May 2016.
[17] H. Li, Z. Yu, and C. Mao, "Fractional differential and variational method for image fusion and super-resolution," *Neurocomputing*, vol. 171, pp. 138–148, Jan. 2016.
[18] H. Li, J. Zhu, and D. Tao, "Asymmetric projection and dictionary learning with listwise and identity consistency constraints for person re-identification," *IEEE Access*, vol. 6, pp. 37977–37990, 2018.
[19] Y. Li, Y. Sun, X. Huang, G. Qi, M. Zheng, and Z. Zhu, "An image fusion method based on sparse representation and sum modified-Laplacian in NSCT domain," *Entropy*, vol. 20, no. 7, p. 522, 2018.
[20] Y. Li, Y. Sun, M. Zheng, X. Huang, G. Qi, H. Hu, and Z. Zhu, "A novel multi-exposure image fusion method based on adaptive patch structure," *Entropy*, vol. 20, no. 12, p. 935, 2018.
[21] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.
[22] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.
[23] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, 2015.
[24] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.
[25] A. Malviya and S. G. Bhirud, "Multi-focus image fusion of digital images," in *Proc. Int. Conf. Adv. Recent Technol. Commun. Comput.*, 2009, pp. 887–889.
[26] R. Nava, B. Escalante-Ramírez, and G. Cristobal, *Mutual Information Improves Image Fusion Quality Assessments*. Bellingham, WA, USA: SPIE, 2007.
[27] V. Petrović, "Subjective tests for image fusion evaluation and objective metric validation. information fusion," *Inf. Fusion*, vol. 8, no. 2, pp. 208–216, 2007.
[28] G. Qi, W.-T. Tsai, C. J. Colbourn, J. Luo, and Z. Zhu, "Test-algebra-based fault location analysis for the concurrent combinatorial testing," *IEEE Trans. Rel.*, vol. 67, no. 3, p. 802–831, Sep. 2018.
[29] G. Qi, J. Wang, Q. Zhang, F. Zeng, and Z. Zhu, "An integrated dictionary-learning entropy-based medical image fusion framework," *Future Internet*, vol. 9, no. 4, p. 61, 2017.
[30] G. Qi, Z. Zhu, K. Erqinhu, Y. Chen, Y. Chai, and J. Sun, "Fault-diagnosis for reciprocating compressors using big data and machine learning," *Simul. Model. Pract. Theory*, vol. 80, pp. 104–127, Jan. 2018.
[31] R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, Jun. 2016.
[32] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[33] Y. Sun, S. Hu, S. Liu, and W. Sun, "A novel multi-focus image fusion algorithm based on NSST-FRFT," in *Proc. 12th Int. Conf. Signal Process. (ICSP)*, 2014, pp. 780–783.
[34] S. Dahiya, P. K. Garg, and M. K. Jat, "A comparative study of various pixel-based image fusion techniques as applied to an urban environment," *Int. J. Image Data Fusion*, vol. 4, no. 3, pp. 197–213, 2013.
[35] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, Jun. 2014.
[36] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.
[37] A. Vishwakarma and M. K. Bhuyan, "Infrared and visible image fusion using NSST and phase stretch transform," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2017, pp. 472–476.
[38] K. Wang, G. Qi, Z. Zhu, and Y. Chai, "A novel geometric dictionary construction approach for sparse representation based image fusion," *Entropy*, vol. 19, no. 7, p. 306, 2017.
[39] Y. Wang, X. Shi, L. Fei, and Y. Ye, "An effective color image fusion algorithm of visual and infrared night images," in *Proc. Control Decis. Conf.*, 2014, pp. 2974–2978.
[40] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1382–1394, Apr. 2013.
[41] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, no. 2, pp. 156–160, 2008.
[42] W. Yang and J.-R. Liu, "Research and development of medical image fusion," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng.*, Oct. 2013, pp. 307–309.
[43] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.
[44] K. Zhan, J. Shi, J. Teng, Q. Li, M. Wang, and F. Lu, "Linking synaptic computation for image enhancement," *Neurocomputing*, vol. 238, pp. 1–12, May 2017.
[45] H. Zhang and X. Cao, "A way of image fusion based on wavelet transform," in *Proc. IEEE 9th Int. Conf. Mobile Ad-Hoc Sensor Netw.*, Dec. 2013, pp. 498–501.
[46] L. Zhang and M. Ge, "An image enhancement algorithm based on space domain and transform domain," *Electron. Opt. Control*, vol. 21, no. 12, pp. 45–48, 2014.
[47] J. Zhao, R. Laganiere, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *Int. J. Innov. Comput. Inf. Control*, vol. 3, no. 6, pp. 1433–1447, 2007.
[48] Z. Zhu, Y. Chai, H. Yin, Y. Li, and Z. Liu, "A novel dictionary learning approach for multi-modality medical image fusion," *Neurocomputing*, vol. 214, pp. 471–482, Nov. 2016.
[49] Z. Zhu, G. Qi, Y. Chai, and P. Li, "A geometric dictionary learning based approach for fluorescence spectroscopy image fusion," *Appl. Sci.*, vol. 7, no. 2, p. 161, 2017.
[50] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, "A novel multi-modality image fusion method based on image decomposition and sparse representation," *Inf. Sci.*, vol. 432, pp. 516–529, Mar. 2018.
[51] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019.

LI YIN received the Medical master's degree from Chongqing Medical University, in 2013. She is currently a Medical Physicist with Chongqing University Cancer Hospital. She is mainly involved in medical imaging research.

MINGYAO ZHENG was born in 1995. She is currently pursuing the master's degree in control engineering with the Automation College of Chongqing University of Post and Telecommunications. Her main research interest includes image processing.

GUANQIU QI received the Ph.D. degree in computer science from Arizona State University, in 2014. He is currently an Assistant Professor with the Computer Information Systems Department, Buffalo State College, Buffalo, NY, USA. His primary research interests include many aspects of software engineering, such as SaaS (Software-as-a-Service), TaaS (Testing-asa-Service), big data testing, combinatorial testing, and service-oriented computing, and machine learning, and image processing.

ZHIQIN ZHU received the B.E. and Ph.D. degrees from Chongqing University, in 2010 and 2016, respectively. He is currently an Associate Professor with the Automation College of Chongqing University of Post and Telecommunications. He has presided over the funds of National Science Foundation of China and Educational Department of Chongqing. His primary research interests include machine learning and image processing.



FU JIN received the Ph.D. degree from the Shanghai Institute of Applied Physics Chinese Academy Of Sciences, in 2010. He is currently a Medical Physicist with Chongqing University Cancer Hospital. He is mainly involved in radiotherapy physics research.

JAESUNG SIM received the Ph. D. degree in business computer information systems from the University of North Texas, and the M.S. degree in management information systems from Texas Tech University. He is currently the Chair Person with the Department of Computer Information Science, Mansfield University of Pennsylvania. His area of specializations include data base management systems, data warehousing, data mining, e-commerce, cognitive process, and organizational change in information systems. He has published and presented numerous papers at the *American Journal of Engineering Research*, *Review of Business Research*, *Journal of Information and Communication Convergence Engineering*, and academic conferences.