

A multi-source image fusion algorithm based on gradient regularized convolution sparse representation

WANG Jian^{1,2,*}, QIN Chunxia¹, ZHANG Xiufei¹, YANG Ke¹, and REN Ping¹

1. School of Electronics and Information Engineering, Northwestern Polytechnical University, Xi'an 710072, China;
2. No. 365 Institute, Northwestern Polytechnical University, Xi'an 710065, China

Abstract: Image fusion based on the sparse representation (SR) has become the primary research direction of the transform domain method. However, the SR-based image fusion algorithm has the characteristics of high computational complexity and neglecting the local features of an image, resulting in limited image detail retention and a high registration misalignment sensitivity. In order to overcome these shortcomings and the noise existing in the image of the fusion process, this paper proposes a new signal decomposition model, namely the multi-source image fusion algorithm of the gradient regularization convolution SR (CSR). The main innovation of this work is using the sparse optimization function to perform two-scale decomposition of the source image to obtain high-frequency components and low-frequency components. The sparse coefficient is obtained by the gradient regularization CSR model, and the sparse coefficient is taken as the maximum value to get the optimal high frequency component of the fused image. The best low frequency component is obtained by using the fusion strategy of the extreme or the average value. The final fused image is obtained by adding two optimal components. Experimental results demonstrate that this method greatly improves the ability to maintain image details and reduces image registration sensitivity.

Keywords: gradient regularization, convolution sparse representation (CSR), image fusion.

DOI: 10.23919/JSEE.2020.000027

1. Introduction

Image fusion technology is very useful for getting images that are more conducive to human visual perception and computer processing [1]. The simplest way is the pixel-level fusion. Wang et al. [2] proposed a fast weighted guided filter image fusion algorithm. Yang et al. [3] presented a fusion framework based on block-matching and 3D (BM3D) multi-scale transform. Ma et al. [4] proposed a fusion algorithm for the infrared and visible images by

using a generative adversarial network, termed as Fusion-GAN. In the above methods, if the representation method is translational and when a small geometrical is distorted or the image registration is out of alignment, the output results of different source images will not show the same features, so non-registered areas in the fusion image are prone to pseudo Gibbs phenomenon. In fact, Blanc and Zhang discussed the related methods that can be insensitive to image misregistration in an image fusion [5,6].

Recently, with the development of compressed sensing, sparse representation (SR) [7] and dictionary learning provide new tools for the image fusion. Image fusion refers to the process of integrating multiple images of the same scene collected by different sensors into a comprehensive image for observation or further processing [8,9]. The SR is a new image representation theory that can select a combination of sparse linear atoms in a given overcomplete dictionary to describe an image or image block. It has been successfully applied to many image processing problems, such as denoising, interpolation and recognition [10].

Due to the sparse capability of the SR, the image fusion method based on the SR [11–15] has received more and more attention. Liu et al. [11] proposed a convolution SR (CSR) image fusion framework and the multi-focus image fusion, and the multi-modal image is decomposed into a base layer and a detail layer. Yang et al. [12] established a sparse decomposition model under the discrete cosine transformation (DCT) dictionary, and used the orthogonal matching pursuit (OMP) method for the SR. Its advantages are better image fusion and low computational complexity, but its sparse ability is poor, and it is too dependent on the geometric features of the image. Li et al. [13] merged the sparse constraints to ensure that the sparse coefficients obtained have group sparsity. Chen et al. [14] added gradient sparsity constraints to the sparse solution, which makes the sparse coefficients reflect sharp edges accurately. In order to take advantage of the correlations presented in the ima-

Manuscript received October 12, 2019.

*Corresponding author.

This work was supported by the National Natural Science Foundation of China (61671383), and Shaanxi Key Industry Innovation Chain Project (2018ZDCXL-G-12-2; 2019ZDLGY14-02-02; 2019ZDLGY14-02-03).

ge, Yang and Li [15] used simultaneous OMP (SOMP) to jointly decompose segments from multiple sources on the same dictionary, making non-zero coefficients from different sources appear in the same location. A complete dictionary has been learned from a large number of training samples similar to the input image [16–19], thereby increasing the adaptability of the representation. However, there is only one general dictionary that does not accurately reflect the complex structure in the input image. Therefore, Kim et al. [20] divided the cluster training samples into a number of structural groups and then trained a specific sub-dictionary on each group. This makes each sub-dictionary best suited to a particular structure, and the entire dictionary has a strong representation. Wang et al. [21] constructed a spectral dictionary and a spatial detail dictionary for the fusion of multi-spectral and panchromatic images, respectively. Zhang et al. [22] designed an efficient multi-focus image fusion method based on the SR, which takes into account the detailed information of each image block and its spatial neighborhood. In [23], a multi-focus image fusion algorithm based on the discrete wavelet transform (DWT) and the SR is proposed, which can better preserve the focus area of the source image and reduce artifacts. A joint sparse model was designed in [24–26] to retain as much texture detail information as possible.

In the above image fusion algorithms, the SR only reflects the global characteristics of the image, ignoring the local features, thus affecting the final fusion effect. In addition, the computational complexity of dictionary learning in the above SR is high. Therefore, this paper proposes a multi-source image fusion algorithm for the gradient regularization CSR, which is not sensitive to image misregistration and is better than the traditional method considering subjective visual consequent and objective evaluation.

2. Gradient regularized CSR

The traditional mathematical expression of the block SR is

$$\arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{D}\mathbf{x} - \mathbf{s}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (1)$$

where \mathbf{D} is a dictionary matrix, \mathbf{x} is a sparse coefficient, \mathbf{s} represents the input source image and λ is the regularized parameter.

$$\begin{cases} \mathbf{D} = (\mathbf{D}_0 \ \mathbf{D}_1 \ \cdots \ \mathbf{D}_{m-1}) \\ \mathbf{x} = (\mathbf{x}_0 \ \mathbf{x}_1 \ \cdots \ \mathbf{x}_{m-1})^T \end{cases} \quad (2)$$

The CSR [27] is commonly used for noise reduction with a convolution basis. The result of the whole image is an optimal single value, which can preserve the details of the image well, especially with translational invariance. The mathematical expression is as follows:

$$\arg \min_{\mathbf{x}_m} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m - \mathbf{s} \right\|_2^2 + \lambda \sum_m \alpha_m \|\mathbf{x}_m\|_1 \quad (3)$$

where \mathbf{d}_m ($m \in \{1, 2, \dots, M\}$) is a set of M dictionary filter sets; M is the dimension of the filter; $*$ denotes convolution; \mathbf{x}_m is a set of sparse coefficient sets; and α_m represents the coefficient weight of a set of the l_1 norm.

The method decomposes the signal by the convolution sum of the filter and the corresponding characteristic response. The formula (3) is added to a gradient of the l_2 norm regularization coefficient map to represent the low-frequency components of the image [28]. Considering the edge smoothing effect of the norm gradient regularization, a total variation regularization (TVR) alternative is adopted and (3) can be got as

$$\begin{aligned} & \arg \min_{\mathbf{x}_m} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m - \mathbf{s} \right\|_2^2 + \lambda \sum_m \alpha_m \|\mathbf{x}_m\|_1 + \\ & \frac{\mu}{2} \sum_m \beta_m \|\sqrt{(\mathbf{g}_0 * \mathbf{x}_m)^2 + (\mathbf{g}_1 * \mathbf{x}_m)^2}\|_2^2 \end{aligned} \quad (4)$$

where β_m represents the coefficient of the l_2 norm; μ is the regularized parameter; \mathbf{g}_0 and \mathbf{g}_1 are gradient filters along the rows and columns of the image, and $\mathbf{g}_0 = [-1 \ 1]$, $\mathbf{g}_1 = [-1 \ 1]^T$.

3. Alternating direction multiplier method (ADMM) algorithm of gradient regularization

The ADMM is a dual convex optimization algorithm, which can be decomposed into several problems to be solved alternately [29]. A linear operator \mathbf{D}_m is defined, \mathbf{G}_l ($l \in \{0, 1\}$) satisfies $\mathbf{D}_m \mathbf{x}_m = \mathbf{d}_m * \mathbf{x}_m$ and $\mathbf{G}_l \mathbf{x}_m = \mathbf{g}_l * \mathbf{x}_m$, so that the final term of (4) is transformed into

$$\begin{aligned} & \frac{\mu}{2} \sum_m \beta_m \|\sqrt{(\mathbf{g}_0 * \mathbf{x}_m)^2 + (\mathbf{g}_1 * \mathbf{x}_m)^2}\|_2^2 \\ & \frac{\mu}{2} \sum_m \beta_m (\|\mathbf{G}_0 \mathbf{x}_m\|_2^2 + \|\mathbf{G}_1 \mathbf{x}_m\|_2^2) \end{aligned} \quad (5)$$

because $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}_l$ ($l \in \{0, 1\}$) are block matrices.

$$\begin{aligned} \boldsymbol{\Gamma}_l &= \begin{pmatrix} \sqrt{\beta_0} \mathbf{G}_l & 0 & \cdots & 0 \\ 0 & \sqrt{\beta_1} \mathbf{G}_l & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\beta_{m-1}} \mathbf{G}_l \end{pmatrix}, \\ \boldsymbol{\alpha} &= \begin{pmatrix} \alpha_0 \mathbf{I} \\ \alpha_1 \mathbf{I} \\ \vdots \\ \alpha_{m-1} \mathbf{I} \end{pmatrix} \end{aligned} \quad (6)$$

where \mathbf{I} is the identity matrix.

The final item of (5) can be further transformed as

$$\begin{aligned} \frac{\mu}{2} \sum_m \beta_m \|\sqrt{(\mathbf{g}_0 * \mathbf{x}_m)^2 + (\mathbf{g}_1 * \mathbf{x}_m)^2}\|_2^2 = \\ \frac{\mu}{2} (\|\mathbf{G}_0 \mathbf{x}\|_2^2 + \|\mathbf{G}_1 \mathbf{x}\|_2^2). \end{aligned} \quad (7)$$

Carry on the Fourier transform for (4) and we have

$$\begin{aligned} \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{Dx} - \mathbf{s}\|_2^2 + \lambda \|\boldsymbol{\alpha} \odot \mathbf{x}\|_1 + \\ \frac{\mu}{2} (\|\boldsymbol{\Gamma}_0 \mathbf{x}\|_2^2 + \|\boldsymbol{\Gamma}_1 \mathbf{x}\|_2^2) \end{aligned} \quad (8)$$

where \odot denotes element-wise multiplication.

The auxiliary matrices \mathbf{y}_0 , \mathbf{y}_1 and \mathbf{y}_2 are introduced, and (8) is transformed into

$$\begin{aligned} \arg \min_{\mathbf{x}, \mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2} \frac{1}{2} \|\mathbf{Dx} - \mathbf{s}\|_2^2 + \lambda \|\boldsymbol{\alpha} \odot \mathbf{y}_2\|_1 + \\ \frac{\mu}{2} (\|\mathbf{y}_0\|_2^2 + \|\mathbf{y}_1\|_2^2), \\ \text{s.t. } \begin{pmatrix} \boldsymbol{\Gamma}_0 \mathbf{x} \\ \boldsymbol{\Gamma}_1 \mathbf{x} \\ \mathbf{x} \end{pmatrix} - \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = 0. \end{aligned} \quad (9)$$

The dual variable and Lagrange multipliers \mathbf{u}_0 , \mathbf{u}_1 and \mathbf{u}_2 are introduced, and the constrained optimization problem of (9) is transformed into the form of non-constrained optimization problem through iteration:

$$\begin{aligned} \mathbf{x}^{(j+1)} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{Dx} - \mathbf{s}\|_2^2 + \frac{\rho}{2} \|\boldsymbol{\Gamma}_0 \mathbf{x} - \mathbf{y}_0^{(j)} + \mathbf{u}_0^{(j)}\|_2^2 + \\ \frac{\rho}{2} \|\boldsymbol{\Gamma}_1 \mathbf{x} - \mathbf{y}_1^{(j)} + \mathbf{u}_1^{(j)}\|_2^2 + \frac{\rho}{2} \|\mathbf{x} - \mathbf{y}_2^{(j)} + \mathbf{u}_2^{(j)}\|_2^2, \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbf{y}^{(j+1)} = \arg \min_{\mathbf{y}} \lambda \|\boldsymbol{\alpha} \odot \mathbf{y}\|_1 + \frac{\rho}{2} \|\boldsymbol{\Gamma}_0 \mathbf{x}^{(j+1)} - \mathbf{y}_0 + \mathbf{u}_0^{(j)}\|_2^2 + \\ \frac{\rho}{2} \|\boldsymbol{\Gamma}_1 \mathbf{x}^{(j+1)} - \mathbf{y}_1 + \mathbf{u}_1^{(j)}\|_2^2 + \frac{\rho}{2} \|\mathbf{x}^{(j+1)} - \mathbf{y}_2 + \mathbf{u}_2^{(j)}\|_2^2, \\ \mathbf{u}_i^{(j+1)} = \mathbf{u}_i^{(j)} + \mathbf{x}^{(j+1)} - \mathbf{y}_i^{(j+1)}, \quad i = \{0, 1, 2\} \end{aligned} \quad (11) \quad (12)$$

where $\rho > 0$ is the penalty parameter; j is the number of iteration.

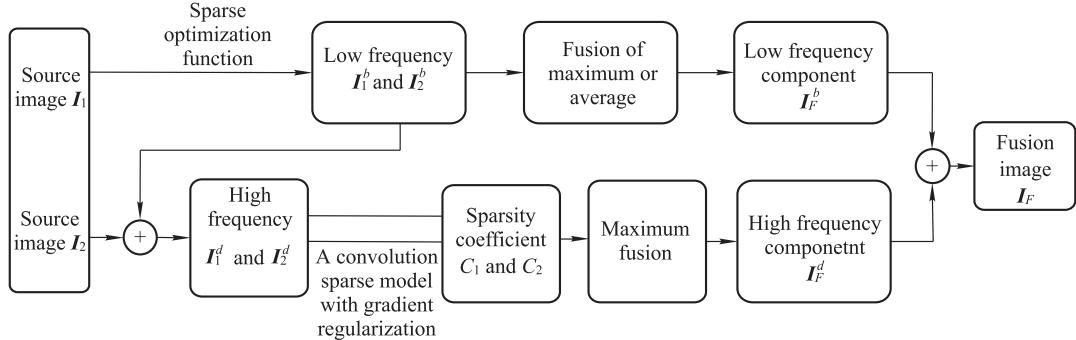


Fig. 1 Multi-source fusion algorithm based on gradient regularized CSR

Equation (11) is given by

$$\mathbf{y}_m^{(j+1)} = \mathbf{S}_{\alpha_m \lambda / \rho} (\mathbf{x}_m^{(j+1)} + \mathbf{u}_m^{(j)}) \quad (13)$$

where $\mathbf{S}_{\alpha_m \lambda / \rho} (\mathbf{x}_m^{(j+1)} + \mathbf{u}_m^{(j)})$ is the specific expression of the function $\mathbf{S}_\gamma(\mathbf{u})$ as

$$\mathbf{S}_\gamma(\mathbf{u}) = \text{sign}(\mathbf{u}) \odot \max(0, |\mathbf{u}| - \gamma) \quad (14)$$

where $\widehat{\mathbf{D}}_m$, $\widehat{\mathbf{x}}_m$, $\widehat{\boldsymbol{\Gamma}}_0$, $\widehat{\boldsymbol{\Gamma}}_1$, $\widehat{\mathbf{y}}_0$, $\widehat{\mathbf{y}}_1$, $\widehat{\mathbf{y}}_2$, $\widehat{\mathbf{u}}_0$, $\widehat{\mathbf{u}}_1$, $\widehat{\mathbf{u}}_2$ and $\widehat{\mathbf{s}}$ are used in the discrete Fourier transform (DFT) domain to represent \mathbf{D}_m , \mathbf{x}_m , $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, \mathbf{y}_0 , \mathbf{y}_1 , \mathbf{y}_2 , \mathbf{u}_0 , \mathbf{u}_1 , \mathbf{u}_2 and \mathbf{s} respectively, to solve the problem of a low computational efficiency of (10).

$$\begin{aligned} \arg \min_{\widehat{\mathbf{x}}} \frac{1}{2} \|\widehat{\mathbf{D}} \widehat{\mathbf{x}} - \widehat{\mathbf{s}}\|_2^2 + \frac{\rho}{2} \|\widehat{\boldsymbol{\Gamma}}_0 \widehat{\mathbf{x}} - \widehat{\mathbf{y}}_0 + \widehat{\mathbf{u}}_0\|_2^2 + \\ \frac{\rho}{2} \|\widehat{\boldsymbol{\Gamma}}_1 \widehat{\mathbf{x}} - \widehat{\mathbf{y}}_1 + \widehat{\mathbf{u}}_1\|_2^2 + \frac{\rho}{2} \|\widehat{\mathbf{x}} - \widehat{\mathbf{y}}_2 + \widehat{\mathbf{u}}_2\|_2^2, \end{aligned} \quad (15)$$

$$\begin{cases} \widehat{\mathbf{D}} = (\widehat{\mathbf{D}}_0 \ \widehat{\mathbf{D}}_1 \ \cdots \ \widehat{\mathbf{D}}_{m-1}) \\ \widehat{\mathbf{x}} = (\widehat{\mathbf{x}}_0 \ \widehat{\mathbf{x}}_1 \ \cdots \ \widehat{\mathbf{x}}_{m-1})^\top \end{cases} \quad (16)$$

The formula (15) is a problem of quadratic optimization. The partial derivative of $\widehat{\mathbf{x}}$ is obtained and the partial derivative is 0.

$$\begin{aligned} (\widehat{\mathbf{D}}^H \widehat{\mathbf{D}} + \rho \widehat{\boldsymbol{\Gamma}}_0^H \widehat{\boldsymbol{\Gamma}}_0 + \rho \widehat{\boldsymbol{\Gamma}}_1^H \widehat{\boldsymbol{\Gamma}}_1 + \rho \mathbf{I}) \widehat{\mathbf{x}} = \\ (\widehat{\mathbf{D}}^H \widehat{\mathbf{s}} + \rho (\widehat{\mathbf{y}}_2 - \widehat{\mathbf{u}}_2 + \widehat{\boldsymbol{\Gamma}}_0 (\widehat{\mathbf{y}}_0 - \widehat{\mathbf{u}}_0) + \widehat{\boldsymbol{\Gamma}}_1 (\widehat{\mathbf{y}}_1 - \widehat{\mathbf{u}}_1))) \end{aligned} \quad (17)$$

The matrix $\widehat{\mathbf{D}}$ is composed of M diagonal matrices of $N \times N$, where M is the dimension of the filter, N is the dimension of the image \mathbf{s} , $\widehat{\mathbf{D}}^H \widehat{\mathbf{D}}$ is a symmetric matrix of $MN \times MN$, and $\widehat{\boldsymbol{\Gamma}}_0^H \widehat{\boldsymbol{\Gamma}}_0$ and $\widehat{\boldsymbol{\Gamma}}_1^H \widehat{\boldsymbol{\Gamma}}_1$ are the diagonal matrices respectively. Equation (17) can be solved by the Sherman-Morrison equation.

4. Multi-source fusion with gradient regularized CSR

The principle of the algorithm in this paper is shown in Fig. 1.

To begin with, the sparse optimization function is used to decompose the source image into high-frequency and low-frequency components. Sparse coefficients of high frequency components are obtained by a gradient regularized CSR model. Then, use the maximum value strategy. Low-frequency components are obtained by an extreme or average value fusion strategy.

5. Improved algorithm

If the image registration is out of alignment, the output results of different source images will not show the same features. Blanc and Zhang discussed the use of image registration in the image fusion [5,6]. Thus, researchers study the image registration problem separable with the image fusion, and many effective methods for image registration have been proposed. The gradient regularization method considers the consistency between image blocks, and uses sparse coding with the convolution to achieve the SR of the entire image, that is, the sum of the convolutions of the filter and the corresponding feature response can increase the robustness of regional registration misalignment. When the area step sizes in the horizontal direction and the vertical direction are one pixel, the SR of the image using the algorithm in this paper has translation invariance. In addition, using the maximum and average strategies can also alleviate the above-mentioned regional misregistration problems.

The principle of the multi-source image fusion algorithm based on gradient regularization of the CSR is as follows:

Suppose k registered source images \mathbf{I}_k ($k \in \{1, 2, \dots, K\}$) are geometrically registered, and there are a set of dictionary filters \mathbf{d}_m ($m \in \{1, 2, \dots, M\}$). The multi-source image fusion algorithm of the gradient regularized CSR proposed consists of four parts.

5.1 Two-scale image decomposition

The source image \mathbf{I}_k is decomposed into a high-frequency component \mathbf{I}_k^d and a low-frequency component \mathbf{I}_k^b . This two-scale decomposition method has been applied to many image fusion methods [30]. The high-frequency part of the image mainly reflects the detailed information, while the low-frequency part mainly represents the spectral information in the image, which is obtained by

$$\arg \min_{\mathbf{I}_k^b} \|\mathbf{I}_k - \mathbf{I}_k^b\|_F^2 + \eta(\|\mathbf{g}_x * \mathbf{I}_k^b\|_F^2 + \|\mathbf{g}_y * \mathbf{I}_k^b\|_F^2) \quad (18)$$

where \mathbf{I}_k is the source image input; $*$ represents convolution; η is the regularized parameter and set to 5; \mathbf{I}_k^b is the low frequency part; \mathbf{g}_x and \mathbf{g}_y are gradient filters along image rows and columns, and $\mathbf{g}_x = [-1 \ 1]$, $\mathbf{g}_y = [-1 \ 1]^T$

respectively. The formula (18) is a Tikhonov regularization problem whose regularization is a used method of regularization of ill-posed problems and can be solved effectively through the fast Fourier transform (FFT):

$$\arg \min_{\mathbf{I}_K^B} \|\mathbf{I}_K - \mathbf{I}_K^B\|_F^2 + \eta(\|\mathbf{G}_X \mathbf{I}_K^B\|_F^2 + \|\mathbf{G}_Y \mathbf{I}_K^B\|_F^2). \quad (19)$$

The formula (19) is a problem of quadratic optimization. The standard approach is the least squares linear regression. However, if no \mathbf{I}_K^B satisfies (19) or more than one \mathbf{I}_K^B does, that is, the solution is not unique – the problem is said to be ill-posed. To solve the above optimization problems, it should solve the partial derivative of \mathbf{I}_K^B set to 0:

$$\mathbf{I}_K^B = \frac{\mathbf{I}_K}{1 + \eta(\mathbf{G}_X^2 + \mathbf{G}_Y^2)}. \quad (20)$$

The inverse frequency Fourier transform is performed by (20) to obtain the low frequency component, and the high frequency component can be obtained by subtracting the low frequency component from the source image.

$$\mathbf{I}_k^d = \mathbf{I}_k - \mathbf{I}_k^b \quad (21)$$

5.2 Multi-source image fusion algorithm with high frequency component

The high frequency part reflects the details of image edge, texture and so on, the sparse coefficient graph $\mathbf{c}_{k,m}$ ($m \in \{1, 2, \dots, M\}$) can be obtained by

$$\arg \min_{\mathbf{c}_{k,m}} \frac{1}{2} \left\| \sum_{m=1}^M \mathbf{d}_m * \mathbf{c}_{k,m} - \mathbf{I}_k^d \right\|_2^2 + \lambda \sum_m \alpha_m \|\mathbf{c}_{k,m}\|_1 + \frac{\mu}{2} \sum_m \beta_m \|\sqrt{(\mathbf{g}_0 * \mathbf{c}_{k,m})^2 + (\mathbf{g}_1 * \mathbf{c}_{k,m})^2}\|_2^2 \quad (22)$$

where $\mathbf{c}_{k,m}$ is a sparse coefficient set of l_1 and l_2 norms; \mathbf{I}_k is the detail layer of the source image. The formula (22) can be solved by the gradient regularization ADMM algorithm. The last term of (3) can be transformed into

$$\begin{aligned} & \frac{\mu}{2} \sum_m \beta_m \|\sqrt{(\mathbf{g}_0 * \mathbf{c}_{k,m})^2 + (\mathbf{g}_1 * \mathbf{c}_{k,m})^2}\|_2^2 = \\ & \frac{\mu}{2} \sum_m \beta_m (\|\mathbf{G}_0 \mathbf{C}_{k,m}\|_2^2 + \|\mathbf{G}_1 \mathbf{C}_{k,m}\|_2^2) = \\ & \frac{\mu}{2} (\|\boldsymbol{\Gamma}_0 \mathbf{C}\|_2^2 + \|\boldsymbol{\Gamma}_1 \mathbf{C}\|_2^2) \end{aligned} \quad (23)$$

where \mathbf{C} is the block matrix, and $\mathbf{C} = (C_0 \ C_1 \ \dots \ C_n)^T$; $\mathbf{C}_{k,m} m \in \{1, 2, \dots, M\}$ is sparse coefficient maps obtained by solving the CSR model with the method in [27]. The formula (22) is transformed by Fourier transform:

$$\arg \min_{\mathbf{C}} \frac{1}{2} \|\mathbf{D}\mathbf{C} - \mathbf{I}_K^D\|_2^2 + \lambda \|\boldsymbol{\alpha} \odot \mathbf{C}\|_1 +$$

$$\frac{\mu}{2}(\|\boldsymbol{\Gamma}_0\mathbf{C}\|_2^2 + \|\boldsymbol{\Gamma}_1\mathbf{C}\|_2^2), \quad (24)$$

$$\begin{aligned} & \arg \min_{\mathbf{x}, \mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2} \frac{1}{2} \|\mathbf{DC} - \mathbf{I}_K^D\|_2^2 + \lambda \|\boldsymbol{\alpha} \odot \mathbf{y}_2\|_1 + \\ & \quad \frac{\mu}{2}(\|\mathbf{y}_0\|_2^2 + \|\mathbf{y}_1\|_2^2), \\ & \text{s.t. } \begin{pmatrix} \boldsymbol{\Gamma}_0\mathbf{C} \\ \boldsymbol{\Gamma}_1\mathbf{C} \\ \mathbf{C} \end{pmatrix} - \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = 0 \end{aligned} \quad (25)$$

where \mathbf{I}_K^D is the result of FFT of \mathbf{I}_k^d .

By using dual variables and introducing Lagrange multipliers $\mathbf{u}_0, \mathbf{u}_1$ and \mathbf{u}_2 , the constrained optimization problem of (25) is transformed into an unconstrained optimization problem by iteration. The constrained optimization problem of (25) is transformed into the form of non-constrained optimization problem by iteration:

$$\mathbf{C}^{(j+1)} = \arg \min_{\mathbf{C}} \frac{1}{2} \|\mathbf{DC} - \mathbf{I}_K^D\|_2^2 + \frac{\rho}{2} \|\boldsymbol{\Gamma}_0\mathbf{C} - \mathbf{y}_0^{(j)} + \mathbf{u}_0^{(j)}\|_2^2 +$$

$$\frac{\rho}{2} \|\boldsymbol{\Gamma}_1\mathbf{C} - \mathbf{y}_1^{(j)} + \mathbf{u}_1^{(j)}\|_2^2 + \frac{\rho}{2} \|\mathbf{C} - \mathbf{y}_2^{(j)} + \mathbf{u}_2^{(j)}\|_2^2 \quad (26)$$

$$\begin{aligned} \mathbf{y}^{(j+1)} &= \arg \min_y \lambda \|\boldsymbol{\alpha} \odot \mathbf{y}\|_1 + \frac{\rho}{2} \|\boldsymbol{\Gamma}_0\mathbf{C}^{(j+1)} - \mathbf{y}_0 + \mathbf{u}_0^{(j)}\|_2^2 + \\ & \quad \frac{\rho}{2} \|\boldsymbol{\Gamma}_1\mathbf{C}^{(j+1)} - \mathbf{y}_1 + \mathbf{u}_1^{(j)}\|_2^2 + \frac{\rho}{2} \|\mathbf{C}^{(j+1)} - \mathbf{y}_2 + \mathbf{u}_2^{(j)}\|_2^2 \end{aligned} \quad (27)$$

where $\mathbf{u}_i^{(j+1)} = \mathbf{u}_i^{(j)} + \mathbf{C}^{(j+1)} - \mathbf{y}_i^{(j+1)}$. Equation (27) is given as

$$\mathbf{y}_m^{(j+1)} = \mathbf{S}_{\alpha_m \lambda / \rho} (\mathbf{C}_m^{(j+1)} + \mathbf{u}_m^{(j)}). \quad (28)$$

Therefore, we can set $\widehat{\mathbf{C}}_m$ and $\widehat{\mathbf{I}}_K^D$ as \mathbf{C}_m and \mathbf{I}_K^D in the DFT domain where $\widehat{\mathbf{D}} = (\widehat{\mathbf{D}}_0 \quad \widehat{\mathbf{D}}_1 \quad \dots \quad \widehat{\mathbf{D}}_{m-1})$, and $\widehat{\mathbf{C}} = (\widehat{\mathbf{C}}_0 \quad \widehat{\mathbf{C}}_1 \quad \dots \quad \widehat{\mathbf{C}}_{m-1})^T$. Using the convolution theorem in the DFT domain, (26) can be transformed into

$$\begin{aligned} & \arg \min_{\widehat{\mathbf{C}}} \frac{1}{2} \|\widehat{\mathbf{DC}} - \widehat{\mathbf{I}}_K^D\|_2^2 + \frac{\rho}{2} \|\widehat{\mathbf{I}}_0 \widehat{\mathbf{C}} - \widehat{\mathbf{y}}_0 + \widehat{\mathbf{u}}_0\|_2^2 + \\ & \quad \frac{\rho}{2} \|\widehat{\mathbf{I}}_1 \widehat{\mathbf{C}} - \widehat{\mathbf{y}}_1 + \widehat{\mathbf{u}}_1\|_2^2 + \frac{\rho}{2} \|\widehat{\mathbf{C}} - \widehat{\mathbf{y}}_2 + \widehat{\mathbf{u}}_2\|_2^2. \end{aligned} \quad (29)$$

The formula (29) is a problem of quadratic optimization as (19). When the partial derivative of $\widehat{\mathbf{C}}$ is 0,

$$\begin{aligned} & (\widehat{\mathbf{D}}^H \widehat{\mathbf{D}} + \rho \widehat{\mathbf{I}}_0^H \widehat{\mathbf{I}}_0 + \rho \widehat{\mathbf{I}}_1^H \widehat{\mathbf{I}}_1 + \rho \mathbf{I}) \widehat{\mathbf{C}} = \\ & \widehat{\mathbf{D}}^H \widehat{\mathbf{I}}_K^D + \rho (\widehat{\mathbf{y}}_2 - \widehat{\mathbf{u}}_2 + \widehat{\mathbf{I}}_0 (\widehat{\mathbf{y}}_0 - \widehat{\mathbf{u}}_0) + \widehat{\mathbf{I}}_1 (\widehat{\mathbf{y}}_1 - \widehat{\mathbf{u}}_1)). \end{aligned} \quad (30)$$

The matrix $\widehat{\mathbf{D}}$ is composed of M diagonal matrices of $N \times N$, and M is the dimension of filters, N is the dimension of the source image \mathbf{I}_k^d , $\widehat{\mathbf{D}}^H \widehat{\mathbf{D}}$ is a symmetric matrix of $MN \times MN$, $\widehat{\mathbf{I}}_0^H \widehat{\mathbf{I}}_0$ and $\widehat{\mathbf{I}}_1^H \widehat{\mathbf{I}}_1$ are diagonal matrices.

The solving method of (30) can be solved by the Sherman Morrison equation.

For a linear system mathematical expression,

$$(\mathbf{J} + \mathbf{aa}^H) \mathbf{x} = \mathbf{o} \quad (31)$$

where \mathbf{J} represents a diagonal matrix; $\mathbf{a} \in \mathbb{R}^n$ is the column vector. $\mathbf{o} \in \mathbb{R}^{n \times n}$ is the offset matrix, by using the Sherman Morrison equation which is as follows:

$$(\mathbf{A} + \mathbf{uv}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{uv}^H \mathbf{A}^{-1}}{1 + \mathbf{u}^H \mathbf{A}^{-1} \mathbf{v}} \quad (32)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an invertible matrix; $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are column vectors.

In combination with (31) and (32) it can be derived that $(\mathbf{J} + \mathbf{aa}^H)^{-1}$. Since $\mathbf{a}^H \mathbf{J}^{-1} \mathbf{o}$ is a scalar, then it can be obtained that

$$\mathbf{x} = \frac{1}{\rho} \left(\mathbf{o} - \frac{\mathbf{a}^H \mathbf{o}}{\rho + \mathbf{a}^H \mathbf{a}} \mathbf{a} \right), \quad \rho = \mathbf{J} \mathbf{I}^{-1}. \quad (33)$$

Equation (30) is similar to (31), and (30) can be obtained by using the Sherman Morrison equation:

$$\begin{aligned} \widehat{\mathbf{C}} &= (\rho \widehat{\mathbf{I}}_0^H \widehat{\mathbf{I}}_0 + \rho \widehat{\mathbf{I}}_1^H \widehat{\mathbf{I}}_1 + \rho \mathbf{I})^{-1} \\ & \left\{ [\widehat{\mathbf{D}}^H \widehat{\mathbf{I}}_K^D + \rho (\widehat{\mathbf{y}}_2 - \widehat{\mathbf{u}}_2 + \widehat{\mathbf{I}}_0^H (\widehat{\mathbf{y}}_0 - \widehat{\mathbf{u}}_0) + \widehat{\mathbf{I}}_1^H (\widehat{\mathbf{y}}_1 - \widehat{\mathbf{u}}_1))] - \right. \\ & \left. \frac{\widehat{\mathbf{D}}^H [\widehat{\mathbf{D}}^H \widehat{\mathbf{I}}_K^D + \rho (\widehat{\mathbf{y}}_2 - \widehat{\mathbf{u}}_2 + \widehat{\mathbf{I}}_0^H (\widehat{\mathbf{y}}_0 - \widehat{\mathbf{u}}_0) + \widehat{\mathbf{I}}_1^H (\widehat{\mathbf{y}}_1 - \widehat{\mathbf{u}}_1))] \widehat{\mathbf{D}}}{[\rho \widehat{\mathbf{I}}_0^H \widehat{\mathbf{I}}_0 + \rho \widehat{\mathbf{I}}_1^H \widehat{\mathbf{I}}_1 + \rho \mathbf{I}] + \widehat{\mathbf{D}}^H \widehat{\mathbf{D}}} \right\}. \end{aligned} \quad (34)$$

Then, the inverse Fourier transform of $\widehat{\mathbf{C}}$ is performed to obtain $\mathbf{c}_{k,m}$. Assume that $\mathbf{c}_{k,1:M}(x, y)$ represents the content of $\mathbf{c}_{k,m}$ at the location (x, y) in the space domain, and $\mathbf{c}_{k,1:M}(x, y)$ is an M dimension vector. According to the SR fusion method adopted in [12,15], the l_1 norm of $\mathbf{c}_{k,1:M}(x, y)$ is used as the activity level measurement of the source image. The activity level diagram $\mathbf{A}_k(x, y)$ is obtained by

$$\mathbf{A}_k(x, y) = \|\mathbf{c}_{k,1:M}(x, y)\|_1. \quad (35)$$

In order to make this method insensitive to registration misreading, the window-based average strategy is applied to obtaining the final level of activity map for $A_k(x, y)$:

$$\overline{\mathbf{A}}_k(x, y) = \frac{\sum_{p=-r}^r \sum_{q=-r}^r \mathbf{A}_k(x+p, y+q)}{(2r+1)^2} \quad (36)$$

where r determines the size of the window. The larger the value of r , the more stable the image registration misreading will be. At the same time, however, some small details may be lost. In the multi-focus image, the edge of

the object in the multi-source image has different definitions, which makes the edge position of each source image not exactly the same. Therefore, a relatively large r is more suitable for the multi-focus image fusion. In the multi-modal image fusion, a relatively small r is more suitable for the multi-modal image fusion because of the small scale details in the source image. Using the “maximum or minimum strategy” according the multi-focus or multi-modal image respectively, the combined coefficients diagram are

$$\mathbf{c}_{F,1:M}(x, y) = \mathbf{c}_{k^*,1:M}(x, y), \quad (37)$$

where $k^* = \arg \max_k (\overline{\mathbf{A}}_k(x, y))$ for multi-focus images or $k^* = \arg \min_k (\overline{\mathbf{A}}_k(x, y))$ for multi-modal images.

Finally, the fusion results of the high-frequency part are reconstructed:

$$\mathbf{I}_F^d(x, y) = \sum_{m=1}^M \mathbf{d}_m * \mathbf{c}_{F,m}(x, y). \quad (38)$$

5.3 Multi-source image fusion algorithm with low-frequency components

The low-frequency part mainly represents the spectral information in the image. For the low-frequency components of the image, different fusion methods are suitable for different types of fusion images. For the multi-focus image fusion, the most important thing is to extract the details of the source image. Because some details still exist in the basic layer, for the multi-focus image fusion, the “maximum” fusion strategy is selected to extract some details existing in the basic layer, and the fusion result expression of the low-frequency component is

$$\mathbf{I}_F^b(x, y) = \mathbf{I}_{k^*}^b(x, y), \quad (39)$$

$$k^* = \arg \max_k (\overline{\mathbf{A}}_k(x, y)). \quad (40)$$

However, the fusion strategy of “maximum” may lead to the inconsistency of human vision for the multi-modal medical image, because the gray value of the same location may be very different. The average fusion strategy can not only preserve the texture details of the source image, but also meet the consistency of the human visual system:

$$\mathbf{I}_F^b(x, y) = \frac{1}{K} \sum_{k=1}^K \mathbf{I}_k^b(x, y). \quad (41)$$

5.4 Two-scale image reconstruction

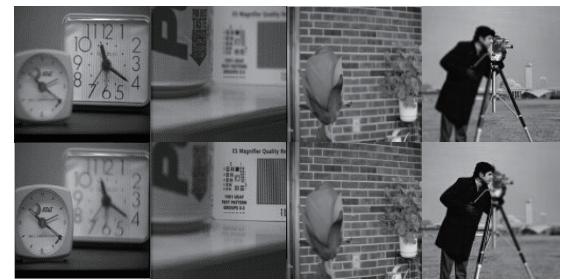
After getting the different low- and high-frequency components respectively according to different images, the high-frequency $\mathbf{I}_F^d(x, y)$ and low-frequency $\mathbf{I}_F^b(x, y)$ components are reconstructed to get the fused image $\mathbf{I}_F(x, y)$:

$$\mathbf{I}_F(x, y) = \mathbf{I}_F^b(x, y) + \mathbf{I}_F^d(x, y). \quad (42)$$

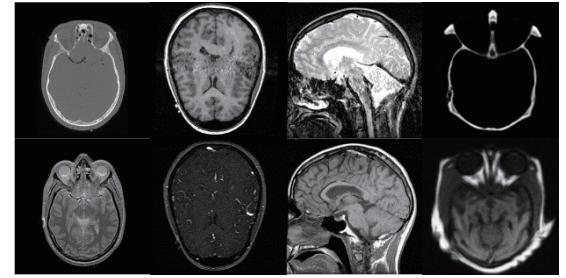
6. Results and analysis

The experimental platform for this paper is the notebook with Intel(R) Core(TM) CPU i7-3610QM CPU@2.30 GHz, Memory 8.0 GB and 64-bit Win7 operating system. All the algorithms we mentioned are achieved with Matlab2014b.

In order to verify the feasibility and effectiveness of the algorithm, four groups of multi-focus images (Fig. 2(a)), multi-modal medical images (Fig. 2(b)) and infrared and visible images (Fig. 2(c)) are used. The size of each source image is 256×256. In the multi-source image fusion data used in this paper, all the source images are registered, which means that the objects in all images are geometrically aligned. The proposed algorithm compared with the fusion algorithm is based on non-subsampled contourlet transform (NSCT) [31], double-tree complex wavelet transform (DTCWT) [32], guide filter (GFF) [33,34], image matting (IM) [35], SR [36], curvature transform and SR (CVT-SR) [37,38], and the combination of NSCT and pulse coupled neural network (NSCT-PCNN) [39] to confirm the performance respectively. The number of decomposition layers of DTCWT, NSCT and NSCT-PCNN is set to four, and the numbers of directions corresponding to decomposition on each layer are 4, 8, 8 and 16. The “averaging” mode is applied to low-frequency subbands and the “maximum” absolute value is applied to high-frequency subbands. The experiments compare subjective evaluation indicators and objective evaluation indicators.



(a) Multi-focus image group



(b) Multi-modal medical image group



(c) Infrared and visible image group

Fig. 2 Source image groups

6.1 Experimental analysis and parameter set

The size of the sliding window block is 23×23 in the multi-focus image fusion, as shown in Table 1. When $\mu = 0.001$, $\lambda = 0.001$, the quality of the fusion algorithm is the best, which determines the regularization parameter λ of the l_1 norm and the regularization parameter μ of the gradient l_2 norm. When the values of μ and λ are selected, the size r of the activity sliding window block is to be determined. When r is 23, the four objective evaluation indices above are the best, as shown in Table 2. For the multi-modal medical image fusion and the infrared and visible images fusion, the value of the regularization parameter λ of the l_1 norm and the regularization parameter μ of the gradient l_2 norm are determined. When μ and λ are fixed, then determine the size of the sliding window block r , as shown in Table 3. When r is 3, the image fusion performance meets the human visual system.

Table 1 Average objective evaluation indicators of multi-focus images when sliding window r is fixed and μ and λ change

μ	λ	MI	$Q^{AB/F}$	Q^Y	PSNR/dB
0.1	0.1	6.753 2	0.615 9	0.814 6	29.326 7
0.01	0.01	7.984 9	0.726 0	0.905 6	30.062 4
0.001	0.001	8.857 3	0.747 5	0.966 9	32.061 4
0.000 1	0.000 1	8.854 4	0.747 2	0.966 9	31.432 5

Table 2 Average objective evaluation indicators of multi-focus images when sliding window r changes and μ and λ are fixed

r	MI	$Q^{AB/F}$	Q^Y	PSNR/dB
9	8.476 9	0.742 7	0.937 5	30.321 5
11	8.613 9	0.746 2	0.942 0	30.345 6
13	8.721 5	0.747 1	0.952 8	30.354 6
15	8.766 2	0.747 3	0.955 8	31.253 7
17	8.755 2	0.747 1	0.957 1	31.341 6
19	8.799 8	0.747 0	0.963 0	32.032 4
21	8.834 8	0.747 2	0.965 9	31.036 8
23	8.857 3	0.747 5	0.966 9	32.061 4
25	8.830 7	0.746 5	0.966 4	31.072 5

Compared with the uncertainty of human visible analysis, objective evaluation indices such as the mutual information (MI), the information structure similarity Q^Y , the

peak signal-to-noise ratio (PSNR) and the edge holding degree $Q^{AB/F}$ are introduced to evaluating the performance of different fusion methods quantitatively.

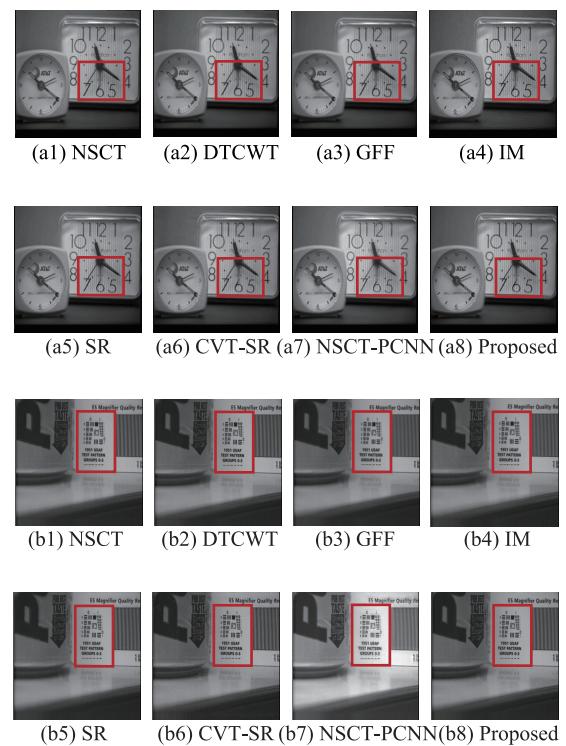
Table 3 Objective evaluation indicators of multi-modal medical and infrared and visible light images when sliding window r changes and μ and λ are fixed

r	MI	$Q^{AB/F}$	Q^Y	PSNR/dB
1	2.878 3	0.732 8	0.741 4	36.573 9
3	2.901 2	0.738 3	0.741 0	37.660 9
5	2.879 8	0.738 1	0.738 2	35.643 7
7	2.868 0	0.726 7	0.732 4	36.273 2
9	2.841 2	0.705 0	0.725 0	36.641 7
11	2.799 0	0.687 3	0.719 8	35.637 8

In order to compare the complexity of the algorithm in this paper, each group of experiments is repeated five times, and the average time consumption calculated is shown in each table.

6.2 Experiment and analysis of multi-focus image fusion

This section mainly uses four sets of multi-focus images as shown in Fig. 2(a). It can be clearly seen from Fig. 3 that the result images have ringing effects through the fusion result graphs based on the DTCWT, CVT-SR and NSCT algorithms, and to a certain extent, the image edge contrast is reduced, as shown in Fig. 3(a1)–Fig. 3(d1), Fig. 3(a2)–Fig. 3(d2), Fig. 3(a6)–Fig. 3(d6).



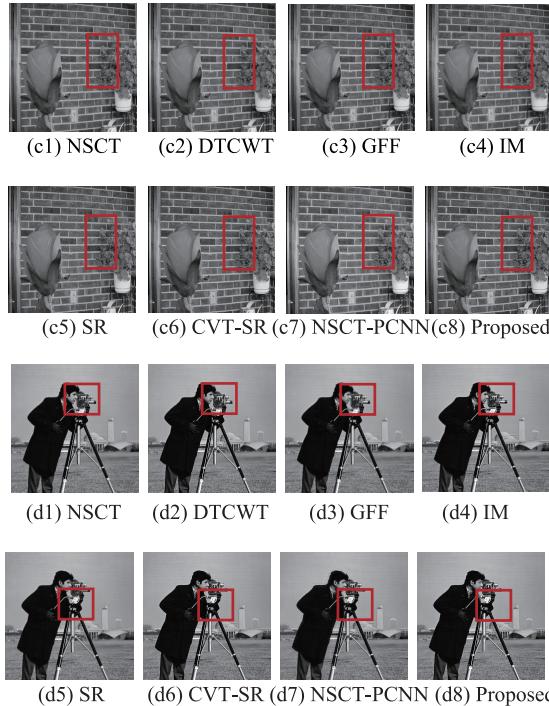


Fig. 3 Multi-focus source images and fusion results of different methods

The result images of the GFF algorithms, as shown in Fig. 3(a3)–Fig. 3(d3), have artifacts in the edge region of the fused images. Although multi-directional filtering based on the IM method has a strong resolution, the fused image will be affected by the severe artifacts generated by the down-sampling operation, so that the contrast of the fused image is significantly reduced, as shown in

Fig. 3(a4)–Fig. 3(d4). The NSCT method has the advantages of multi-scale analysis, the fused image obtained by this method loses part of the edge information, and the false contour of the focus area in these images is obvious, which makes the contour of the focus area blurred. Similarly, the fusion method based on the NSCT-PCNN has a great visual improvement on the merged image, and more significant features can be extracted from the source image, but the result still has artifacts, and the weak edge is insensitive and cannot accurately extract the boundary information of the focus area, as shown in Fig. 3(a7) and Fig. 3(b7). The SR-based algorithm has been always based on local image blocks rather than on the entire image block. Some details are smoothed or even lost in the fused image (Fig. 3(a5)). In contrast, the fused image of the method proposed herein optimally extracts the focus area from the source image by accurately locating the boundary of the focus area. From Fig. 3(a8)–Fig. 3(d8), it can be seen that the contour of the focal region is clear and complete. In addition, the contrast of the fused image obtained by this method is higher than that of other fusion methods, the transition region between the fusion region and the background is natural, and few human influences are introduced in the fusion process, which is convenient for identifying different targets in a complex background.

Objective assessment of four groups of the multi-focus image fusion method is shown in Table 4. The algorithm is able to retain a lot of focus information, and reflects the stability and reliability of the proposed algorithm to a certain extent.

Table 4 Objective indicators of different methods for multi-focus images

Source	Index	NSCT	DTCWT	GFF	IM	SR	CVT-SR	NSCT-PCNN	Proposed
Multi-focus image (a)	MI	7.327 4	7.208 7	8.052 1	8.211 3	7.553 5	7.920 5	7.438 4	8.820 2
	$Q^{AB/F}$	0.730 3	0.720 5	0.742 6	0.747 6	0.736 7	0.697 0	0.679 2	0.795 1
	Q^Y	0.925 8	0.929 0	0.944 6	0.910 6	0.927 5	0.931 6	0.923 4	0.965 9
	PSNR/dB	31.942 4	31.891 9	32.192 1	32.047 0	32.224 1	32.400 1	32.555 7	32.663 4
Multi-focus image (b)	MI	7.525 8	7.472 3	8.297 1	8.414 4	7.531 3	7.736 1	7.945 4	8.799 7
	$Q^{AB/F}$	0.733 1	0.730 0	0.756 1	0.752 2	0.744 2	0.705 7	0.690 1	0.750 2
	Q^Y	0.945 1	0.949 9	0.960 6	0.907 4	0.942 5	0.929 6	0.937 7	0.964 1
	PSNR/dB	29.621 5	29.640 5	29.631 8	29.612 3	29.526 0	29.647 8	29.841 5	29.672 1
Multi-focus image (c)	MI	6.161 2	6.159 1	7.670 9	7.031 1	6.523 9	5.913 3	5.225 0	8.476 4
	$Q^{AB/F}$	0.782 1	0.781 9	0.796 3	0.782 9	0.791 8	0.771 2	0.749 7	0.827 1
	Q^Y	0.958 4	0.957 9	0.961 7	0.955 4	0.962 2	0.954 8	0.942 0	0.980 5
	PSNR/dB	26.516 6	26.484 7	26.512 2	26.497 1	26.611 9	26.329 1	28.056 9	28.218 6
Multi-focus image (d)	MI	8.512 5	8.534 3	8.678 4	7.947 5	8.317 6	7.882 3	7.647 2	8.823 9
	$Q^{AB/F}$	0.741 1	0.740 4	0.745 9	0.728 8	0.743 5	0.723 0	0.709 8	0.731 4
	Q^Y	0.974 0	0.978 2	0.984 0	0.962 7	0.975 3	0.960 7	0.917 7	0.979 8
	PSNR/dB	26.340 3	26.355 5	26.621 5	26.343 5	26.489 5	26.481 2	26.623 0	26.630 3
Running time	T/s	23.854 1	23.070 5	12.507 1	21.357 5	39.695 4	41.642 95	39.267 5	32.178 9

6.3 Experiment and analysis of multi-modal medical image fusion

Four groups of computed tomography/magnetic resonance imaging (CT/MRI) multi-mode medical source images are used to verify the effectiveness of the proposed algorithm compared with NSCT, DTCWT, GFF, IM, SR, CVT-SR and NSCT-PCNN respectively in Fig. 2(b). Fig. 4 shows four sets of CT/MRI multi-modal medical source images and fused images of the above algorithms. The method based on the CVT-SR and NSCT-PCNN is superior to other methods in brightness and contrast, but local details are lost, as shown in Fig. 4(e6), Fig. 4(e7), Fig. 4(f6), Fig. 4(f7), Fig. 4(g6), Fig. 4(g7), Fig. 4(h6) and Fig. 4(h7). The local middle part of the image obtained by the DTCWT and IM method is ambiguous, as shown in Fig. 4(e2), Fig. 4(f2), Fig. 4(g2) and Fig. 4(h2). The fusion methods of GFF and NSCT not only lose a lot of details, but also cause serious artifacts, as shown in Fig. 4(e1), Fig. 4(e3), Fig. 4(f3) and Fig. 4(g3). Methods based on the SR lose local edge information (Fig. 4(e5)). From image fusion results in Fig. 4(h6) and Fig. 4(h7), the CVT-SR in Fig. 4(h6) introduces different degrees of artifacts, while the proposed algorithm result in Fig. 4(h7) does not. Through the comparison of the above fused images, the fused image obtained in this paper is superior to the fused image obtained by the above algorithms, and the obtained fused image can not only extract a large amount of detailed information of the source image, but also generate visible artifacts and brightness distortion.

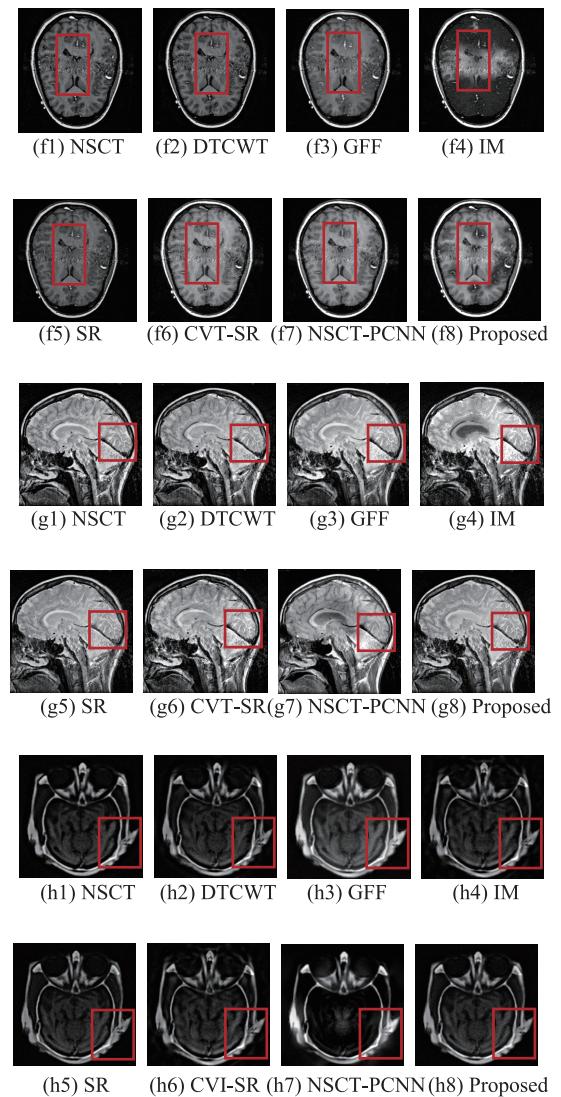
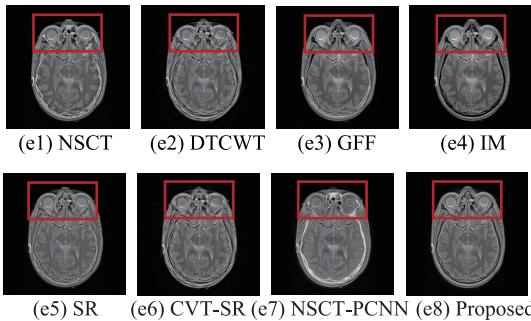


Fig. 4 Multi-modal medical source images and fusion results of different methods

The proposed method for four groups medical images is compared with other methods in Table 5. The proposed algorithm is superior to other fusion methods, because it can extract details from source image, and highlight the significant features.

Table 5 Objective indicators of different methods for multi-modal medical images

Source	Index	NSCT	DTCWT	GFF	IM	SR	CVT-SR	NSCT-PCNN	Proposed
Multi-modal medical image (e)	MI	3.594 2	3.187 0	4.280 7	3.710 2	3.933 4	3.753 0	3.301 1	4.386 1
	$Q^{AB/F}$	0.613 4	0.531 9	0.655 7	0.643 6	0.590 8	0.617 7	0.497 8	0.664 1
	Q^Y	0.850 8	0.743 3	0.885 8	0.830 5	0.873 3	0.800 3	0.595 8	0.929 8
	PSNR/dB	29.676 0	30.065 4	30.003 1	30.676 4	30.849 9	30.467 1	30.951 8	30.830 9
Multi-modal medical image (f)	MI	3.356 3	3.190 0	3.600 0	3.516 3	3.788 6	3.610 6	4.633 4	3.932 8
	$Q^{AB/F}$	0.614 4	0.546 1	0.623 0	0.590 4	0.606 2	0.614 9	0.618 7	0.645 1
	Q^Y	0.794 8	0.759 2	0.832 5	0.821 8	0.825 4	0.860 2	0.898 6	0.882 7
	PSNR/dB	29.127 9	29.189 7	29.385 4	29.004 9	29.794 4	29.481 0	29.405 1	29.803 9

Continued

Source	Index	NSCT	DTCWT	GFF	IM	SR	CVT-SR	NSCT-PCNN	Proposed
Multi-model medical image (g)	MI	3.300 0	3.184 3	3.463 5	3.437 7	3.497 8	3.501 6	3.211 1	3.959 4
	$Q^{AB/F}$	0.555 4	0.501 1	0.590 5	0.601 4	0.556 1	0.591 1	0.459 2	0.593 6
	Q^Y	0.836 0	0.810 4	0.870 9	0.890 6	0.861 2	0.831 0	0.762 2	0.895 0
	PSNR/dB	26.833 7	26.926 9	26.596 7	27.413 7	27.923 2	26.904 1	27.840 5	27.937 0
Multi-model medical image (h)	MI	2.2277	1.8767	3.4313	2.448 8	2.740 1	1.592 8	1.282 4	2.868 0
	$Q^{AB/F}$	0.706 2	0.584 3	0.778 9	0.555 0	0.701 4	0.544 4	0.576 4	0.726 7
	Q^Y	0.686 7	0.613 2	0.896 4	0.621 9	0.709 9	0.526 5	0.541 9	0.732 4
	PSNR/dB	36.928 9	36.988 7	37.302 5	32.098 6	37.320 3	37.302 5	37.812 7	37.660 8
Running time	T/s	28.949 9	29.322 8	12.3674 1	24.354 3	36.029 3	38.429 1	36.726 8	31.921 6

6.4 Experiment and analysis of infrared and visible image fusion

For the infrared and visible image fusion experiments, four sets of registered infrared and visible images (Fig. 2(c)) are selected to verify the correctness of the algorithm proposed in this chapter. It can be seen from Fig. 5 that the GFF-based fused image shows severe distortion, as shown in Fig. 5(i3) and Fig. 5(k3). The contrast of the fusion result image obtained by the DTCWT and NSCT algorithms is reduced, and the bright square panel becomes blurred, which does not reflect the partial texture information in the visible light image well, as shown in Fig. 5(i1) and Fig. 5(i2)). The edge of the fused image obtained by the fusion method based on the NSCT-PCNN is lost, as shown in Fig. 5(i7), Fig. 5(j7), Fig. 5(k7) and Fig. 5(l7). The details of the visible light appearing in the fused image based on the CVT-SR, SR and GFF are lost, and serious artifacts appear, as shown in Fig. 5(k5) and Fig. 5(k6). Although the fused image obtained by the NSCT-based fusion algorithm has been greatly improved visually, a small amount of detail information is lost, as shown in Fig. 5(i3), Fig. 5(j3), Fig. 5(k3) and Fig. 5(l3).

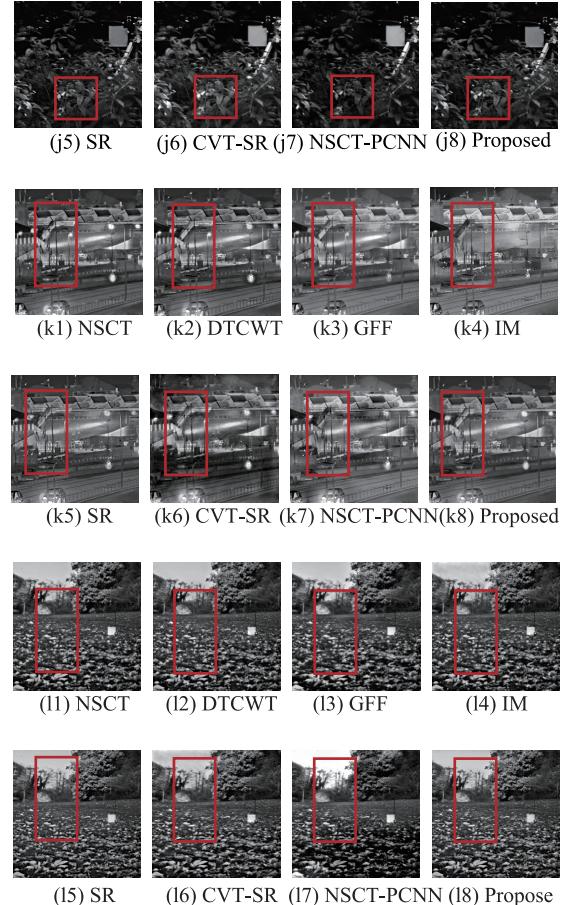
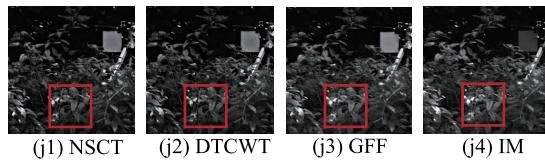
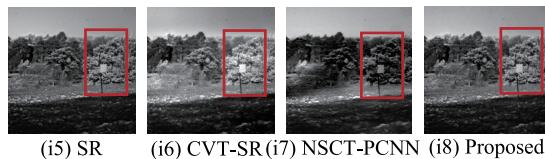
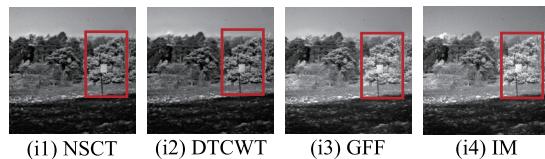


Fig. 5 Fusion results of infrared and visible source images with different methods

From the visible light image of the third group in Fig. 2, the railings and street lamps on the side of the road can be seen. In the infrared images, the moving people, the car and the house can be seen. However, a small amount of distortions occur in the localized image of the fused image obtained by the IM algorithm (Fig. 5(k4)). In Fig. 5(k8), the fused image can clearly distinguish the infrared targets, which is better than the above algorithms in the visual effect, and can better represent the texture details in the visible images. In addition, in the comparison of these four

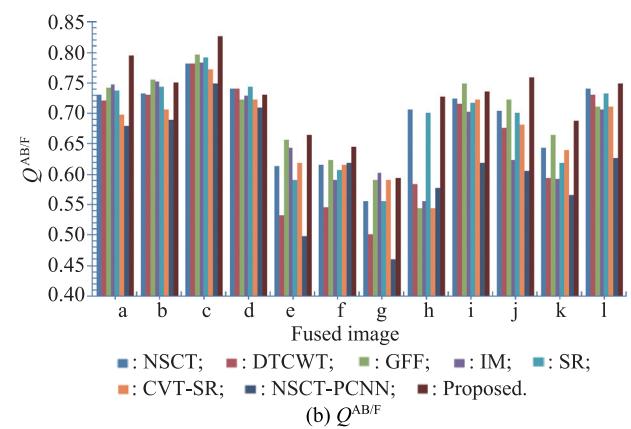
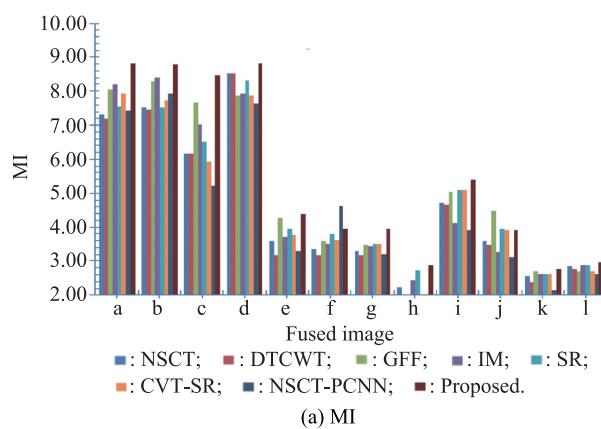
sets of objective data, as shown in Table 6, some indicators of the algorithm may be slightly lower than those of the comparison algorithm. However, integrating subjective visual and objective evaluation indicators, the improved fusion algorithm is better than other methods.

Fig. 5 shows the average evaluation criteria values of the different fusion methods on five pairs of test infrared visible images. It clearly shows that the proposed algorithm significantly outperforms other methods for the tested images with constantly higher scores in terms of fusion metrics. These three largest criteria values confirm the objective assessment, which means the images obtained by the proposed method generally incorporate more information from the visible image together with important targets from the infrared image.

To further compare the performance of the algorithms, Fig. 6 shows the objective evaluation criteria values of different image fusion methods on four groups of test images. It clearly shows that the proposed algorithm significantly outperforms other methods for the tested images with constantly higher scores in terms of fusion metrics, because the image fusion method in the spatial domain can realize the fast multi-scale decomposition and reconstruction of the image through the simple subtraction and addition of the image pixel value. Compared with the spatial domain, the pixel level multi-scale method in the frequency domain realizes the multi-scale decomposition and reconstruction of the image through the high time complexity Fourier transform and inverse Fourier transform of image pixel values.

Table 6 Objective indicators of infrared and visible image with different methods

Source	Index	NSCT	DTCWT	GFF	IM	SR	CVT-SR	NSCT-PCNN	Proposed
Visible-infrared image (i)	MI	4.720 6	4.645 9	5.028 3	4.121 6	5.092 0	5.088 2	3.902 1	5.391 4
	$Q^{AB/F}$	0.723 9	0.716 7	0.749 0	0.701 9	0.717 1	0.722 0	0.618 3	0.735 3
	Q^Y	0.862 7	0.855 1	0.914 2	0.872 6	0.835 1	0.798 4	0.795 9	0.920 6
	PSNR	31.546 9	31.569 8	31.417 2	31.071 1	31.694 7	31.694 5	32.571 8	33.014 2
Visible-infrared image (j)	MI	3.596 5	3.467 4	4.470 2	3.279 5	3.944 1	3.904 3	3.124 3	3.907 9
	$Q^{AB/F}$	0.705 0	0.675 9	0.722 8	0.623 0	0.701 5	0.681 1	0.605 2	0.759 2
	Q^Y	0.878 7	0.865 3	0.943 3	0.841 7	0.886 0	0.906 0	0.847 0	0.909 2
	PSNR	30.342 1	30.436 4	30.501 9	30.209 5	31.035 9	31.359 8	31.488 2	31.603 5
Visible-infrared image (k)	MI	2.553 8	2.373 2	2.711 0	2.612 2	2.610 8	2.627 6	2.136 5	2.757 9
	$Q^{AB/F}$	0.643 8	0.593 2	0.664 8	0.591 4	0.618 5	0.640 5	0.565 8	0.688 1
	Q^Y	0.857 7	0.821 0	0.931 1	0.932 7	0.849 8	0.861 3	0.868 8	0.983 0
	PSNR	29.849 8	29.904 6	30.109 7	30.207 8	30.366 8	30.422 7	30.492 9	30.326 3
Visible-infrared image (l)	MI	2.852 6	2.772 8	2.834 4	2.893 4	2.879 0	2.704 4	2.605 7	2.981 1
	$Q^{AB/F}$	0.740 2	0.730 7	0.759 9	0.705 9	0.732 2	0.711 3	0.627 0	0.749 5
	Q^Y	0.919 2	0.913 1	0.909 4	0.887 6	0.923 6	0.880 6	0.767 5	0.914 1
	PSNR	26.738 9	26.747 7	26.800 2	26.537 5	26.958 9	26.800 2	27.840 7	26.963 0
Running time	T/s	32.942 9	25.325 8	12.547 7	23.862 4	32.188 4	34.146 8	35.534 4	34.128 9



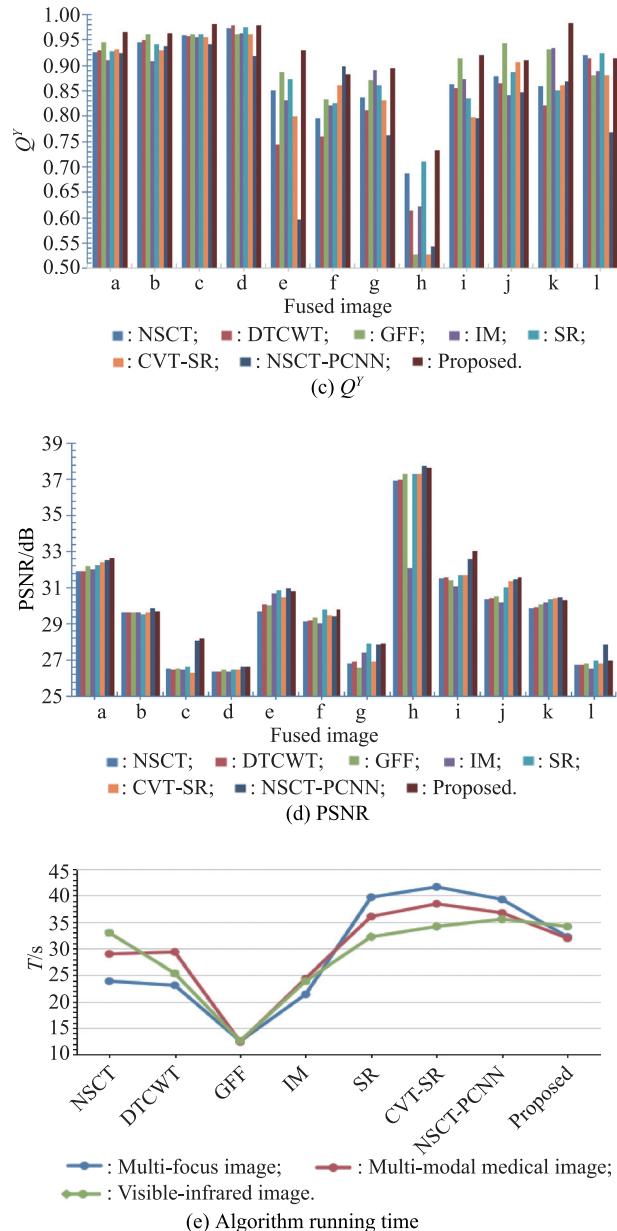


Fig. 6 Objective indicators of fused images obtained based on different fusion algorithms

Therefore, it can be seen that the proposed algorithm's complexity is higher than that of the NSCT, GFF, IM and DTCWT, but lower than that of the SR, CVT-SR and NSCT-PCNN.

7. Conclusions

The fusion results of the proposed method are compared with seven mainstream fusion algorithms. The gradient regularized CSR is introduced to the multi-source image fusion algorithm, which greatly compensates for the shortcomings of the multi-source fusion algorithm based on the SR in image detail preservation. The experimental results

show that the proposed algorithm is superior to the traditional multi-source image fusion algorithms, as shown in Fig. 6. The regularization parameters of the gradient regularization CSR and the size of the window are studied. However, the setting of these parameters is not unique. The proposed algorithm needs to be further optimized to improve the performance of the fusion. In the future development, convolution sparseness will have greater potentials in the field of image fusion.

References

- [1] MA J, MA Y, LI C, et al. Infrared and visible image fusion methods and applications: a survey. *Information Fusion*, 2019, 45: 153 – 178. DOI:10.1016/j.inffus.2018.02.004.
- [2] WANG J, YANG K, REN P, et al. Multi-source image fusion algorithm based on fast weighted guided filter. *Journal of Systems Engineering and Electronics*, 2019, 30(5): 831 – 840.
- [3] YANG D, HU S, LIU S Q, et al. Multi-focus image fusion based on block matching in 3D transform domain. *Journal of Systems Engineering and Electronics*, 2018, 29(2): 415 – 428.
- [4] MA J, YU W, LIANG P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion. *Information Fusion*, 2019, 48(8): 11 – 26.
- [5] BLANC P, WALD L, RANCHIN T, et al. Importance and effect of coregistration quality in an example of pixel to pixel fusion process. *Proc. of the 2nd International Conference on Fusion Earth Data*, 1998: 67 – 74.
- [6] ZHANG Z, BLUM R S. A hybrid image registration technique for a digital camera image fusion application. *Information Fusion*, 2001, 2(2): 135 – 149.
- [7] LI S, KANG X, FANG L, et al. Pixel-level image fusion: a survey of the state of the art. *Information Fusion*, 2017, 33(1): 100 – 112.
- [8] JIN X, JIANG Q, YAO S. A survey of infrared and visual image fusion methods. *Infrared Physics & Technology*, 2017, 85(9): 478 – 501.
- [9] ZHANG Q, LIU Y, RICK S B, et al. Sparse representation based multi-sensor image fusion for multi-focus and multi-modal images. *Information Fusion*, 2018, 40(3): 57 – 75.
- [10] ABAVISANI M, PATEL V M. Deep sparse representation-based classification. *IEEE Signal Processing Letters*, 2019, 26(6): 948 – 952.
- [11] LIU Y, CHEN X, RABAB K, et al. Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*, 2016, 23(12): 1882 – 1886.
- [12] YANG B, LI S. Multifocus image fusion and restoration with sparse representation. *IEEE Trans. on Instrumentation and Measurement*, 2010, 59(4): 884 – 892.
- [13] LI S, YIN H, FANG L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. on Biomedical Engineering*, 2012, 59(12): 3450 – 3459.
- [14] CHEN C, LI Y, LIU W, et al. Image fusion with local spectral consistency and dynamic gradient sparsity. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 2760 – 2765.
- [15] YANG B, LI S. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Information Fusion*, 2012, 13(1): 10 – 19.
- [16] ZHANG Q H, FU Y, LI H F, et al. Dictionary learning method for joint sparse representation-based image fusion. *Optical Engineering*, 2013, 52(5): 057006. DOI:10.1117/1.OE.52.5.057006.
- [17] YIN H, LI S, FANG L, et al. Simultaneous image fusion and

- super-resolution using sparse representation. *Information Fusion*, 2013, 14(3): 229–240.
- [18] LI S T, YIN H T, FANG L Y, et al. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Trans. on Geoscience and Remote Sensing*, 2013, 51(9): 4779–4789.
- [19] NEJATI M, SAMAVI S. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*, 2015, 25(1): 72–84.
- [20] KIM M, HAN D K, KO H, et al. Joint patch clustering-based dictionary learning for multi-modal image fusion. *Information Fusion*, 2016, 27(1): 198–214.
- [21] WANG W, JIAO L, YANG S. Fusion of multispectral and panchromatic images via sparse representation and local autoregressive model. *Information Fusion*, 2014, 20(1): 73–87.
- [22] ZHANG Q, LEVINE M D. Robust multi-focus image fusion using multi-task sparse representation and spatial context. *IEEE Trans. on Image Processing*, 2016, 25(5): 2045–2058.
- [23] AISHWARYA N, ABIRAMI S, AMUTHA R. Multifocus image fusion using discrete wavelet transform and sparse representation. Proc. of the IEEE International Conference on Wireless Communications, Signal Processing and Networking, 2016: 2377–2382.
- [24] RONG C, JIA Y, YANG Y, et al. Fusion of infrared and visible images through a hybrid image decomposition and sparse representation. Proc. of the International Conference on Intelligent Human-Machine Systems and Cybernetics, 2018: 15–26.
- [25] RONG C, JIA Y, YANG Y, et al. Fusion of infrared and visible images based on multi-scale edge-preserving decomposition and sparse representation. Proc. of the International Congress on Image and Signal Processing, 2018: 1–9.
- [26] GAI D, SHEN X, CHENG H, et al. Medical image fusion via PCNN based on edge preservation and improved sparse representation in NSST domain. *IEEE Access*, 2019, 7: 85413–85429.
- [27] WOHLBERG B. Efficient algorithms for convolutional sparse representations. *IEEE Trans. on Image Processing*, 2015, 25(1): 301–315.
- [28] WOHLBERG B. Convolutional sparse representations as an image model for impulse noise restoration. Proc. of the Image, Video, and Multidimensional Signal Processing Workshop, 2016: 1–5.
- [29] LI S, KA X D, HU J W, et al. Image fusion with guided filtering. *IEEE Trans. on Image Processing*, 2013, 22(7): 2864–2875.
- [30] WOHLBERG B. Convolutional sparse representations with gradient penalties. Proc. of the IEEE International Conference on Acoustics, 2017: 15–20.
- [31] ZHANG Q, GUO B L. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Processing*, 2009, 89(7): 1334–1346.
- [32] YANG W, CHAI Q, WANG L M, et al. Multi-focus image fusion method based on dual-tree complex wavelet transform. *Computer Engineering & Applications*, 2007, 43(28): 12–14.
- [33] HUANG H Y, YANG H. MR image reconstruction via guided filter. *Medical & Biological Engineering & Computing*, 2018, 56(4): 635–648.
- [34] YANG Y, QUE Y, HUANG S, et al. Multiple visual features measurement with gradient domain guided filtering for multi-sensor image fusion. *IEEE Trans. on Instrumentation and Measurement*, 2017, 66(4): 691–703.
- [35] LI S, KANG X, HU J, et al. Image matting for fusion of multi-focus images in dynamic scenes. *Information Fusion*, 2013, 14(2): 147–162.
- [36] LIU Y, WANG Z. Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Processing*, 2014, 9(5): 347–357.
- [37] LIU Y, LIU S, WANG Z F, et al. A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*, 2015, 24(7): 147–164.
- [38] NENCINI F, GARZELLI A, BARONTI S, et al. Remote sensing image fusion using the curvelet transform. *Information Fusion*, 2007, 8(2): 143–156.
- [39] QU X, YAN J, XIAO H Z, et al. Image fusion method based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain. *Acta Automatica Sinica*, 2008, 34(12): 1508–1514.

Biographies



WANG Jian was born in 1972. He received his Ph.D. degree in signal and information processing from Northwestern Polytechnical University in 2005. Now he is an assistant professor at the School of Electronics and Information Engineering, Northwestern Polytechnical University, Xi'an, China. His current research interests include UAV intelligent processing technology, UAV ground observation video signal processing technology, and multi-source information intelligent processing technology.



QIN Chunxia was born in 1995. She received her B.S. degree in electronic and information engineering from Northwestern Normal University, Lanzhou, China in 2017. She is currently pursing her M.S. degree in the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. Her research interests include signal processing, target detection and recognition, etc.

E-mail: chunxia_qin@163.com



ZHANG Xiufei was born in 1989. He received his M.S. degree from the School of Electronics and Information, Northwestern Polytechnical University in 2017. He is now pursing his Ph.D. degree in the School of Automation, Northwestern Polytechnical University, Xi'an, China. His research interests include signal processing, image fusion, etc.

E-mail: 921391314@qq.com



YANG Ke was born in 1995. She received her B.S. degree in electronic and information engineering from Taiyuan University of Technology, Taiyuan, China in 2017. She is currently pursing her M.S. degree in the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. Her research interests include signal processing, image fusion, etc.

E-mail: xgdms_yk@mail.nwpu.edu.cn



REN Ping was born in 1993. She received her B.S. degree in electronic and information engineering from Northwestern University in 2016. She is now pursing her M.S. degree in the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. Her research interests include signal processing, target recognition, etc.

E-mail: 1403147639@mail.nwpu.edu.cn