JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Fusion of Multifucus Image with Noise Based on Adaptive Sparse and Low-Rank Representations

Xin Feng[1,2,*], Haifeng Gong[1], Guohang Qiu[2], and Kaiqun Hu[2]

Abstract

Traditional multifucus image fusion often requires the inclusion of edge features, blurred details, and noise pollution when perturbed by noise. To address these problems, this study proposes a method for fusing noisy multifucus images using adaptive sparse and low-rank representations. The proposed method first decomposes the image into high- and low-frequency subband coefficients using a non-subsampled shearlet transform. Subsequently, the high-frequency energy components are fused and denoised using a low-rank representation. The corresponding fusion rules are then set using an adaptive sparse representation to fuse the low-frequency subband coefficients. The final fusion result is obtained by reconstructing the fused high- and low-frequency subband coefficients. Experimental results show that the proposed method outperforms traditional methods in terms of both subjective performance and objective indicators, making it a compelling fusion method for noisy multifucus images.

Keywords

Adaptive Sparse Representation, Image Fusion, Low-Rank Representation, Non-subsampled Shearlet Transform

1. Introduction

Obtaining clear panoramic images with a different focus is a significant challenge. The multifucus image fusion process can accurately extract multiple image targets with varying focal lengths in the same scene and combine the precise areas in these images into a high-quality main image. However, weather, imaging equipment limitations, and interference can cause noise during shooting. Therefore, it is essential to denoise the multifucus images during fusion. Multifucus image fusion is crucial for the subsequent image processing tasks and is widely used in microscopic and biomedical imaging, as well as industrial production. Currently, the classical method of multifucus image fusion is based on multiscale transformation (MST) [1].

Notably, the directional limitations inherent to the directional wavelet transform (DWT) present opportunities for improvement. To this end, the proposed multiscale geometric transformation enables the complex contours of an image to be accurately expressed. The contourlet transform (CT) and shearlet transform (ST) are widely utilized transformations; however, these two algorithms involve subsampling

during image decomposition, resulting in a noticeable Gibbs phenomenon in the final fused image. As a solution, the non-subsampled contourlet transform (NSCT) [2] and non-subsampled shearlet transform (NSST) [3] have been proposed, which utilize subbands of the same size as the source image. These approaches effectively resolve the Gibbs phenomenon caused by subsampling and facilitate the establishment of fusion rules.

The MST is a relatively refined NSST and not limited by the number of decomposition directions. Although this type of algorithm can complete image decomposition effectively, it cannot remove noise. Yang and Li [4] proposed using sparse representation (SR) to improve the image fusion effect. The core idea of the algorithm is to decompose images into several partially overlapping patches, thereby expressing these patches approximately using linear combinations of the vectors learned from the dictionary in the early stages. Unlike MST, SR uses a stepping sliding window to split the source image into partially overlapping patches during the decomposition process, to enable noise removal. However, because the dictionary matrix cannot completely cover all source image information, SR has a defect in extracting detailed image information. If the dimension of the dictionary matrix is blindly increased, computing efficiency is reduced, leading to redundant information in the fused image. In this case, noise is included in the approximate dictionary objects. Liu and Wang [5] proposed an advanced adaptive sparse representation (ASR) method. This method trains seven dictionaries in advance and divides them into six types based on the pixel gradient information of the training image. These patches are then expressed using the column vectors of the corresponding dictionaries. Compared with SR, ASR can improve the SR accuracy and denoising effect without significantly reducing computing efficiency. Liu et al. [6] proposed a method that uses images as dictionaries, called low-rank representation (LRR), which is superior to SR in image information extraction.

Furthermore, LRR has demonstrated strong anti-noise performance owing to its ability to separate the noise component from the image. However, the effectiveness of LRR is contingent upon two stringent conditions. First, the training data vector must be complete, which limits the ability of the LRR method to process images with a significant amount of information. Second, the LRR method is sensitive to noise, which diminishes its robustness when processing noisy images. To address these limitations, Li et al. [7] proposed a latent low-rank representation (LatLRR) method to overcome the problems of data volume and noise sensitivity. This approach augments the original LRR dictionary with hidden items, and decomposes the image into low-ranking, salient, and noisy components. Despite its potential, the salient component of LatLRR is incomplete compared with that of MST.

In conclusion, the MST technique can decompose source images into low- and high-frequency bands. However, MST lacks the ability to effectively remove noise, whereas SR can remove noise to some extent but suffers from the drawback of insufficient detail information extraction. Although LRR can effectively remove noise, it may not be suitable for processing images with a large amount of information. Furthermore, the LRR imposes restrictions on the degree of image noise pollution. Based on the aforementioned analysis, this paper proposes a novel method for noisy multifocus image fusion based on NSST decomposition combined with ASR and LRR regularization. The proposed method first employs NSST to decompose an image into high- and low-frequency bands. Subsequently, ASR is utilized to fuse low-frequency bands, thereby avoiding the extraction of information that is not sufficiently detailed. Considering that the information content of the high-frequency bands of the image is relatively small and the noise component is distributed in the high-frequency bands after image decomposition, LRR fusion

processing also satisfies the validity conditions. Finally, the processed high- and low-frequency bands are fused into an image through the inverse NSST. The comparative results of the experiments validate the advantages of the fusion algorithm proposed in this study.

2. Non-subsampled Shearlet Transform

The shearlet and contourlet transforms are improved versions of DWT that address the basic limitations. The ST has two notable advantages over DWT. First, the number of decomposition directions in ST is not limited. Second, the calculation efficiency of ST is excellent. However, the subsampling process in ST can result in Gibbs phenomenon in the fusion image. To address this issue, Easley et al. [8] introduced a non-subsampled process called NSST, which decomposes source images into high- and low-frequency bands using a non-subsampled pyramid filter (NSPF). Fig. 1 shows the coefficient components of each layer of the two-layer NSST decomposition structure.
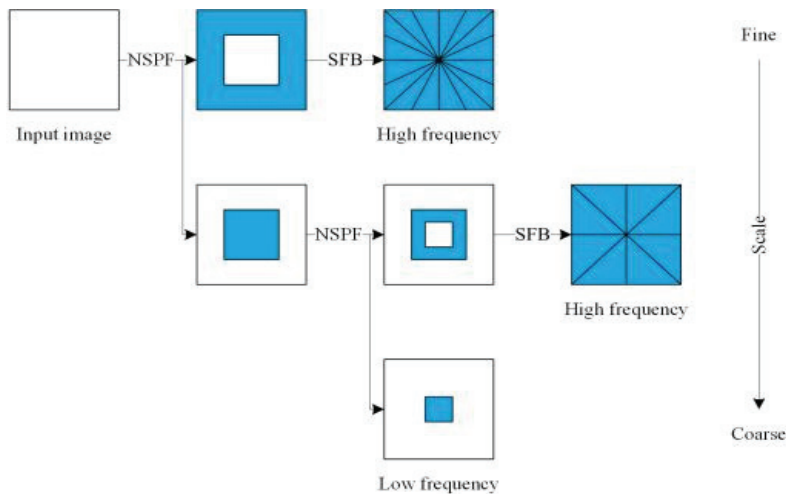

Fig. 1. Two-level NSST decomposition diagram.
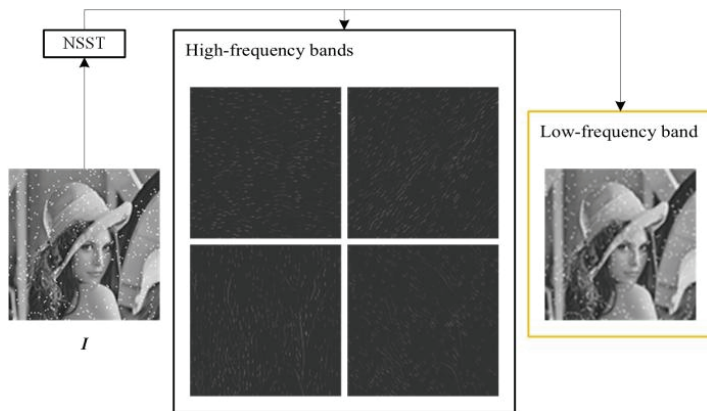

Fig. 2. NSST decomposition diagram of the image containing noise.

After decomposing the noisy image using NSST, the noise component is distributed across each subband. As depicted in Fig. 2, the original image I is decomposed using the 1-layer 2-direction NSST, resulting in two high-frequency bands and one low-frequency band. It is evident from Fig. 2 that the degree of noise pollution in the low-frequency band is significantly reduced compared with that in the original image, whereas noise is still present in the high-frequency bands. Increasing the number of decomposition layers and directions in the NSST results in a higher concentration of noise in the high-frequency bands, whereas the residual noise component in the low-frequency band decreases. Thus, denoising high-frequency bands is crucial for the effectiveness of the proposed algorithm.

3. Fusion Regulation

As illustrated in Fig. 2, the amount of information in the high-frequency bands is relatively low. With an increase in the number of decomposition layers and the direction in which they are applied, the high-frequency bands expand, resulting in a sparser distribution of the noise components within each band. This characteristic renders high-frequency bands suitable for LRR processing. Conversely, as the NSST decomposition becomes more thorough, the low-frequency band contains less edge feature information and fewer noise components. Therefore, utilizing ASR to process the low-frequency band can effectively prevent the loss of edge feature information and reduce the impact of noise on the dictionary.

3.1 Adaptive Sparse Representation

The core idea of SR is to approximate signals $x \in R_n$ by partial column vectors in the over-complete dictionary $D \in R_{n \times m}$, that is, $x \approx D\alpha$, where $\alpha \in R_m$ is the unknown sparse coefficient vector that can be solved by orthogonal matching pursuit (OMP) [9] of Eq. (1):

$$\max_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|x - D\alpha\|_2 < \varepsilon, \tag{1}$$

where $\varepsilon > 0$ is the error tolerance; $\|\cdot\|_0$ denotes l_0-norm; and $\|\cdot\|_2$ denotes the l_2-norm.

ASR is an improved version of SR that solves the time-consuming training process and completeness of the dictionary. The ASR dictionary training set is first constructed to sample high-quality images using a sliding window to obtain patches, and the pixel mean value of the sampled patches is set to zero. Subsequently, a preset threshold is used to exclude patches whose edge structures are not apparent. After collecting all of the M patches, they can be denoted as $P = \{p_1, p_2, \cdots, p_M\}$. For each $p_i \in P$, which can be classified according to the gradient of the pixel, the Sobel operator is used to calculate the horizontal gradient $G_x(x, y)$:

$$G_x(x, y) = p(x, y) * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

and vertical gradient $G_y(x, y)$:

$$G_y(x, y) = p(x, y) * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{3}$$

where $*$ denotes convolution. Then, the gradient magnitude $G_x(x, y)$ and orientation $\theta(x, y)$ can be calculated as follows:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{4}$$

and

$$\boldsymbol{\theta}(x, y) = \arctan\left[\frac{G_y(x, y)}{G_x(x, y)}\right] \tag{5}$$

Partitioning $\theta(x, y)$ into six regions in the range of 360˚, as in Fig. 3, every pixel in $p_i \in P$ can be classified into one of six regions. Grouping all pixels of $p_i \in P$ into $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$ and calculating $G(x, y)$ as the magnitude for each pixel, $p_i \in P$ can be classified into p_{k_i} using Eq. (6):

$$k_i = \begin{cases} 0 & \sum_{k=1}^{\frac{\theta_{\max}}{6}} \theta_k < \frac{1}{3} \\ k^* & \text{otherwise} \end{cases} \tag{6}$$

where $\theta_{\max} = \max\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$ is the highest amplitude among the six regions, and $k^* = \arg\max_k\{\theta_k | k = 1, 2, 3, 4, 5, 6\}$ is the index of $\theta_{\max}$. If $k = 0$, the main direction of p_i is unclear. In this study, 50 high-quality images were sampled using an 8×8 sliding window to obtain approximately 100,000 patches. After setting the average pixel value to zero and excluding patches without apparent edges, approximately 80,000 patches remained for constructing the dictionary training set. Grouping these 80,000 patches using Eq. (6), seven training subsets were obtained, denoted as $\{P_0, P_1, P_2, P_3, P_4, P_5, P_6\}$. The patches in P_0 do not have an apparent main direction, whereas the other patches do.
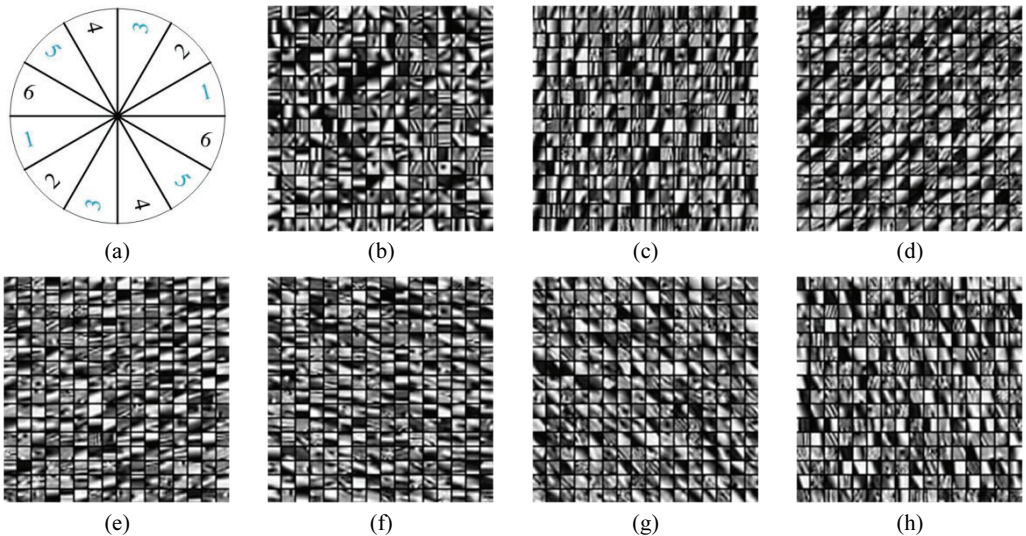

Fig. 3. Dictionary classification: (a) directional partition, (b) sub-dictionary 0, (c) sub-dictionary 1, (d) sub-dictionary 2, (e) sub-dictionary 3, (f) sub-dictionary 4, (g) sub-dictionary 5, and (h) sub-dictionary 6.

The training model is shown in Eq. (7), which can be solved using K-singular value decomposition (K-SVD) [9].

$$\min_{D,\{\alpha_i\}_{i=1}^{M}} \sum_{i=1}^{M} \|\alpha_i\|_0 \qquad \text{s.t.} \|y_i - D\alpha_i\|_2 < \varepsilon, i = 1, 2, \cdots, M \qquad (7)$$

where M is the number of patches; y_i is the target vector rearranged by p_i; α_i is a sparse vector; and D is the dictionary that needs to be trained. Because there are seven training subsets, there are seven sub-dictionaries, which are denoted as $\{D_0, D_1, D_2, D_3, D_4, D_5, D_6\}$. D_0 is trained using all 80,000 patches as shown in Fig. 3(b). $D_1 - D_6$ are trained using $p_1 - p_6$, respectively, as shown in Fig. 3(c)–3(h). The ASR fusion rule can be divided into five steps:

Step 1. Sampling the source images I_A and I_B using an 8×8 sliding window to obtain M patches, which can be denoted as $\{p_A^i\}_{i=1}^{M}$ and $\{p_B^i\}_{i=1}^{M}$;

Step 2. Rearranging p_A^i and p_B^i into 64×1 vectors $\hat{V}_A^i$ and $\hat{V}_B^i$, respectively, and subtracting the average values $\bar{v}_A^i$ and $\bar{v}_B^i$ to obtain zero mean value vectors V_A^i and V_B^i, as shown in Eq. (8):

$$\begin{cases} V_A^i = \hat{V}_A^i - \bar{v}_A^i \cdot 1 \\ V_B^i = \hat{V}_B^i - \bar{v}_B^i \cdot 1 \end{cases} \qquad (8)$$

where 1 denotes a 64×1 vector;

Step 3. Grouping p_A^i and p_B^i using Eq. (6) and training the dictionaries as in Eq. (7). Subsequently, sparse vectors α_A^i and α_A^i are calculated using Eq. (9):

$$\begin{cases} \alpha_A^i = arg \ \min_{\alpha}\|\alpha\|_0 & \text{s.t.} \left\|V_A^i - D\alpha\right\|_2 \\ \alpha_B^i = arg \ \min_{\alpha}\|\alpha\|_0 & \text{s.t.} \left\|V_B^i - D\alpha\right\|_2 \end{cases} \qquad (9)$$

Step 4. Confirming sparse vector α_F^i of the fusion image by maximal l_1-norm, as in Eq. (10):

$$\alpha_F^i = \begin{cases} \alpha_A^i & if \ \left\|\alpha_A^i\right\|_1 > \left\|\alpha_B^i\right\|_1 \\ \alpha_B^i & \text{otherwise} \end{cases} \qquad (10)$$

then patching V_F^i of the fusion image, as in Eq. (11):

$$V_F^i = D\alpha_F^i + \bar{v}_F^i \cdot 1 \qquad (11)$$

where $\bar{v}_F^i$ is specified as in Eq. (12):

$$\bar{v}_F^i = \begin{cases} \bar{v}_A^i & \text{if } \alpha_F^i = \alpha_A^i \\ \bar{v}_B^i & \text{otherwise} \end{cases} \qquad (12)$$

Step 5. By processing all $\{p_A^i\}_{i=1}^{M}$ and $\{p_B^i\}_{i=1}^{M}$ following the above procedure, $\{V_F^i\}_{i=1}^{M}$ is obtained. Rearranging $\{V_F^i\}_{i=1}^{M}$ into 8×8 fusion image patches $\{p_F^i\}_{i=1}^{M}$, then placing them in the corresponding positions, and dividing repeatedly for every pixel, the fusion image I_F is obtained. The fusion processing rule proposed in this section is shown in Fig. 4.
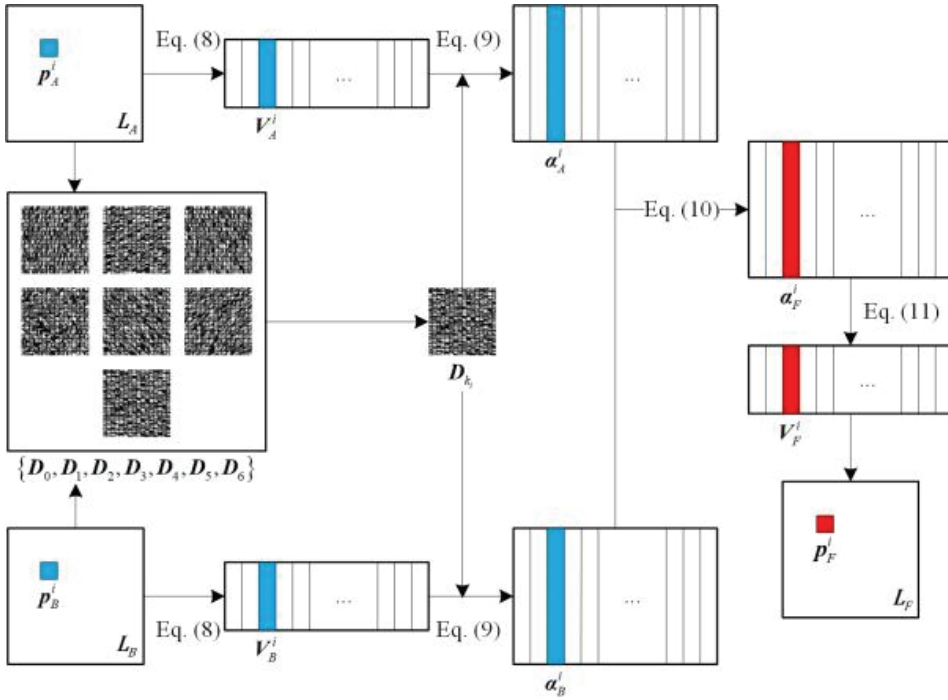
Fig. 4. Schematic diagram of ASR fusion.

3.2 Low-Rank Representation

The principle of LRR is similar to SR, which is based on the theory that vector $X = \{x_1, x_2, \cdots, x_M\}$ in space R_n can be represented by a linear combination of vectors in an over-complete dictionary $D \in R_{n \times m}(n < m)$, as in Eq. (13):

$$X = DZ \tag{13}$$

where $Z = \{z_1, z_2, \cdots, z_M\}$ is the coefficient matrix in space R_m. The difference in LRR is that it calculates the lowest rank, as expressed in Eq. (14).

$$\min_{Z}\|Z\|_* \quad \text{s.t. } X = DZ \tag{14}$$

where $\|\cdot\|_*$ denotes the nuclear norm, which is the sum of the singular values of the matrix. X is adopted as the dictionary to avoid studying the dictionary, as follows:

$$\min_{Z}\|Z\|_* \quad \text{s.t. } X = XZ \tag{15}$$

Here, Z is typically a block diagonal matrix. If X does not contain a sufficient number of data vectors, then $Z = I$ is probably the only feasible solution; thus, low-rank representation may fail. The most notable advantage of LRR is its ability to separate noise from images. Eq. (16) is derived from Eq. (15) by adding the noise component E.

$$\min_{Z,E}\|Z\|_* + \lambda\|E\|_{1,2} \quad \text{s.t. } X = XZ + E \tag{16}$$

where $\lambda > 0$ is the balance coefficient, $\|E\|_{1,2}$ denotes the $l_{1,2}$-norm of E, which can be calculated as follows:

$$\|\boldsymbol{E}\|_{1,2} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{n} \left([\boldsymbol{E}]_{ij}\right)^2} \tag{17}$$

The value of λ has a direct impact on the effect of image fusion, which is discussed explicitly in Section 5. Solving Eq. (16) using the augmented Lagrange multiplier (ALM), the image without noise is denoted as XZ or $X - E$. The LRR decomposition results are shown in Fig. 5, in which noise separation is observed.
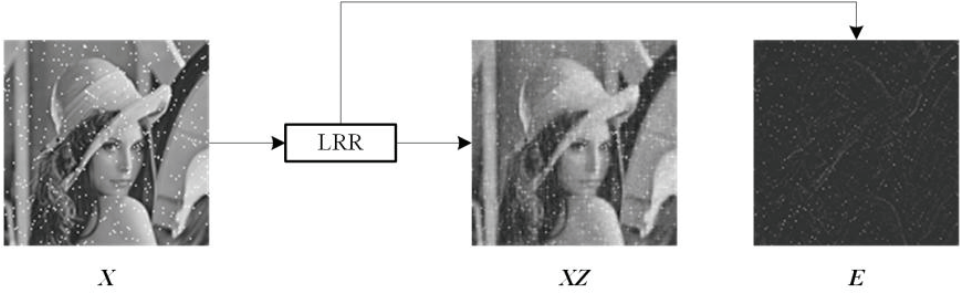

Fig. 5. LRR decomposition diagram.

3.3 Reconstruction

The algorithms mentioned above cannot directly process color images. Therefore, the original color images I_A and I_B must first be split into three subbands of red, green, and blue channels, as shown in Eq. (18):

$$\begin{cases} I_A = I_{A,r} + I_{A,g} + I_{A,b} \\ I_B = I_{B,r} + I_{B,g} + I_{B,b} \end{cases} \tag{18}$$

where $I_{A,r}$, $I_{A,g}$, and $I_{A,b}$ are the red, green, and blue subbands of I_A, respectively, and similarly $I_{B,r}$, $I_{B,g}$, and $I_{B,b}$ are subbands of I_B. Therefore, the three subbands can be processed separately using the proposed method.

First, $I_{A,i}$ and $I_{B,i}$ are decomposed using NSST to obtain $\{H_{A,i}^{l,k}, L_{A,i}\}$ and $\{H_{B,i}^{l,k}, L_{B,i}\}$, where $H_{A,i}^{l,k}$ and $L_{A,i}$ are the high- and low-frequency bands of $I_{A,i}$, respectively, and similarly $H_{B,i}^{l,k}$ and $L_{B,i}$ are the high- and low-frequency bands of $I_{B,i}$, respectively; $i \in [r, g, b]$ denotes the three color subbands of the source image; $l \in [1, N]$ denotes the decomposition levels of NSST; and $k \in [1, K(l)]$ denotes the directions of each level of l.

Second, $H_{A,i}^{l,k}$ and $H_{B,i}^{l,k}$ are processed using LRR to obtain the high-frequency bands without noise $\widehat{H}_A^{l,k}$, $\widehat{H}_B^{l,k}$, and noise components $E_A^{l,k}$, $E_B^{l,k}$, respectively. The maximum rule is used to confirm the fused high-frequency bands $H_{F,i}^{l,k}$, as shown in Eq. (19):

$$H_{F,i}^{l,k} = \max\left(\widehat{H}_{A,i}^{l,k}, \widehat{H}_{B,i}^{l,k}\right) \tag{19}$$

Then, $L_{A,i}$ and $L_{B,i}$ are processed by ASR to obtain the fused low-frequency band $L_{F,i}$. Finally, the fused subband $I_{F,i}$ can be obtained by applying the inverse NSST to $H_{F,i}^{l,k}$ and $L_{F,i}$. Finally, the three fused subbands are superimposed to obtain the final fusion result I_F, as shown in Eq. (20):

$$I_F = I_{F,r} + I_{F,g} + I_{F,b} \tag{20}$$

4. Experimental Results

The proposed method involves processing a pair of registered multifucus images, I_A and I_B, that are intentionally corrupted by noise. A flowchart outlining the steps of this method is shown in Fig. 6.

The main steps of the algorithm proposed in this paper are as follows:

Algorithm 1. Proposed image fusion algorithm

  Input images: I_A and I_B.

  Output images: Result image I_F

/* Section1: Color format conversion.   */

---

1 for every [I_A, I_B] **do**

 2 Decomposition of each original image into red, green, and blue channels to obtain the corresponding subbands $I_A = \{I_{A,r}, I_{A,g}, I_{A,b}\}$ and $I_B = \{I_{B,r}, I_{B,g}, I_{B,b}\}$.

3 end

4 for $i = [r, g, b]$ **do**

 Part 2: NSST decomposition.   */

 5 Applying NSST on I_A and I_B to obtain $\{\boldsymbol{H}_{A,i}^{l,k}, L_{A,i}\}$ and $\{\boldsymbol{H}_{B,i}^{l,k}, L_{B,i}\}$, $l \in [1, N]$, $k \in [1, K(l)]$

 /* Part 3: Fusion of high frequency bands.   */

 6 for $l = [1{:}N]$ **do**

  /* N denotes the decomposition level of the NSST;   */

  7 for $k = [1{:}K(l)]$ **do**

   /* $K(l)$ denotes directions of each level of l;   */

   8 Perform LRR on $\{\boldsymbol{H}_{A,i}^{l,k}, \boldsymbol{H}_{B,i}^{l,k}\}$ to obtain, $\{\widehat{\boldsymbol{H}}_{A,i}^{l,k}, \widehat{\boldsymbol{H}}_{B,i}^{l,k}\}$ and $\{\boldsymbol{E}_{A,i}^{l,k}, \boldsymbol{E}_{B,i}^{l,k}\}$ by Eq. (16);

   9 Perform Max-rule on $\{\widehat{\boldsymbol{H}}_{A,i}^{l,k}, \widehat{\boldsymbol{H}}_{B,i}^{l,k}\}$ to obtain $\boldsymbol{H}_{F,i}^{l,k}$ as $\boldsymbol{H}_{F,i}^{l,k} = \max(\widehat{\boldsymbol{H}}_{A,i}^{l,k}, \widehat{\boldsymbol{H}}_{B,i}^{l,k})$.

  10 end

  /* Part 4: Fusion of low-frequency bands.   */

  11 Perform ASR on $\{\boldsymbol{L}_{A,i}, \boldsymbol{L}_{B,i}\}$ to obtain $\boldsymbol{L}_{F,i}$.

 12 end

 Part 5: NSST restraint traction.   */

 13 Perform inverse NSST on $\{\boldsymbol{H}_{F,i}^{l,k}, L_{F,i}\}$ to obtain $\boldsymbol{I}_{F,i}$.

14 end

 Part 6: Image reconstruction.   */

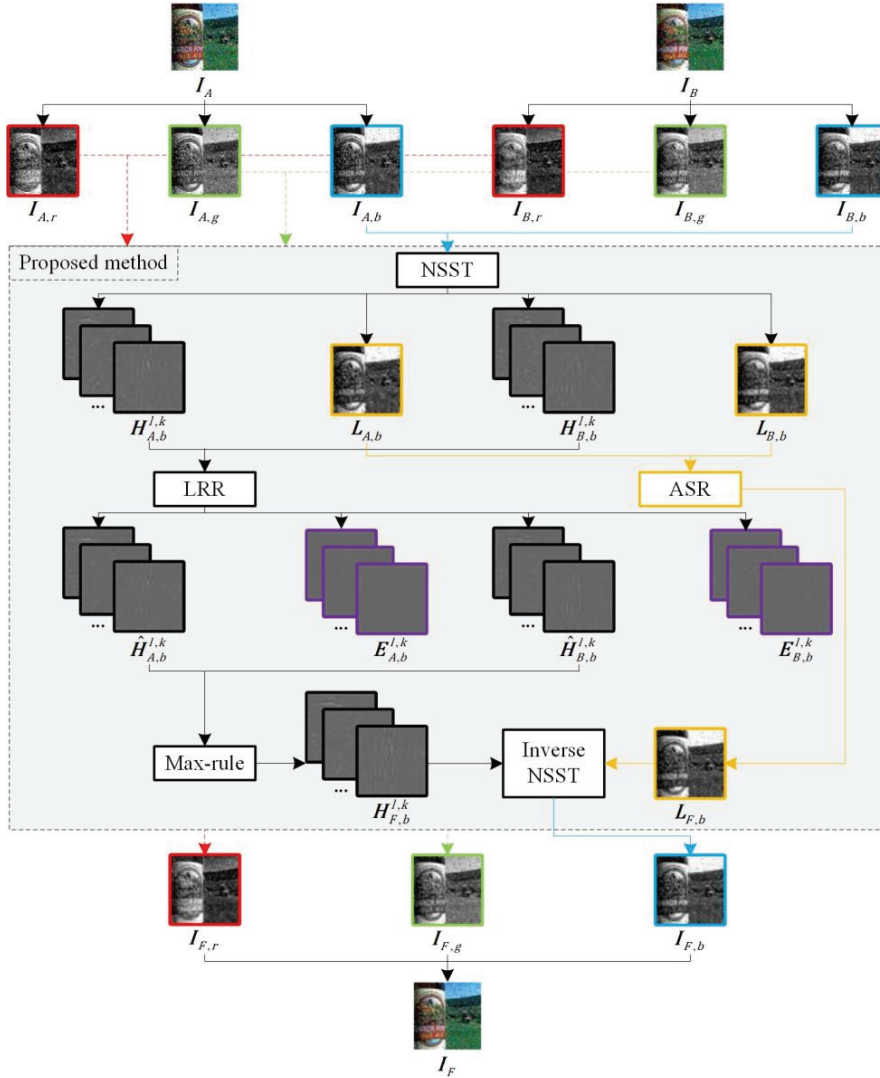**15** $I_F = I_{F,r} + I_{F,g} + I_{F,b}$

Fig. 6. Algorithm flowchart of the proposed method.

5. Experiment

Two sets of partially focused images collected from the Lytro multifucus dataset [10] and their corresponding noisy images are used to verify the effectiveness of the proposed algorithm. The validity of the experiments was confirmed using four objective assessment indexes: normalized mutual information (Q_{MI}), phase congruency (Q_{G}), image structural similarity metrics (Q_{Y}), and human perception-inspired fusion metrics (Q_{CB}). These four indicators measure the fusion results from different perspectives. Q_{MI} is used to measure the information transfer from the source image to the fusion result; Q_G is used to analyze the gradient information that the source image retains in the fusion result; Q_Y is used to analyze the structural information in the fusion results; and Q_{CB} is used to measure the global quality. By combining these four indicators, the objective performance of the fusion results can be evaluated from local and

global perspectives. More information on the four indices can be found in [11]. In this study, the number of decomposition levels of the NSST transformation was set to 3. The optimal values of these indices were obtained for $\lambda = 0.05$, so the value of λ was set to 0.05. The experimental platform hardware for the algorithm was an AMD CPU Ryzen 5 5700X (3.59 GHz) with 64 GB DDR4 3200 MHz memory. The software platform used was MATLAB R2020a on a Windows 10 64-bit operating system.

In this study, we evaluated the performance of the proposed method by comparing it with seven commonly used noisy image fusion algorithms: LP combined with SR (LP_SR) [4], NSCT combined with SR (NSCT_SR) [12], dual-tree complex wavelet transform combined with SR (DTCWT_SR) [13], ASR [5], CSR [6], LatLRR [14], and MDLatLRR [7]. For all the methods, we set the multiscale decomposition level to four. The original images are shown in Fig. 7(a) (front-focused) and Fig. 7(b) (back-focused). Fig. 7(c) and 7(d) show the noisy images with speckle noise (green box, mean $\mu=0$ and variance $\sigma = 0.05$) added to Fig. 7(a) and 7(b), respectively. Fig. 7(e)–7(l) show the fused images of the seven comparison methods and our proposed method. As shown in the green box in Fig. 7(l), the proposed method is the best in removing speckle noise, and the multifucus fusion target is within the recognizable range. We also conducted a comparison experiment for the fifth set of images, as shown in Fig. 8. Fig. 8(a) and 8(b) show the top- and bottom-focused original images, respectively. In contrast, Fig. 8(c) and 8(d) show noisy images with the same noise (green box) added to Fig. 8(a) and 8(b), respectively, and Fig. 8(e)–8(l) show the fused images of the seven comparison methods and our proposed method. As shown in the green box of Fig. 8(l), the best noise removal is achieved by the proposed method in such an extreme case, and the multifucus fusion target is within the recognizable range. Note that if a higher level of target discrimination is required for multifucus fusion, the value of λ can be increased accordingly. In this manner, more noise and image are retained in the fused image, making the target more explicit despite the additional noise. In further experiments, the noise was effectively eliminated by setting λ to 0.05. However, decreasing the λ value further does not significantly reduce the noise component in the fused image, instead, resulting in a blurred multifocus-fused target.
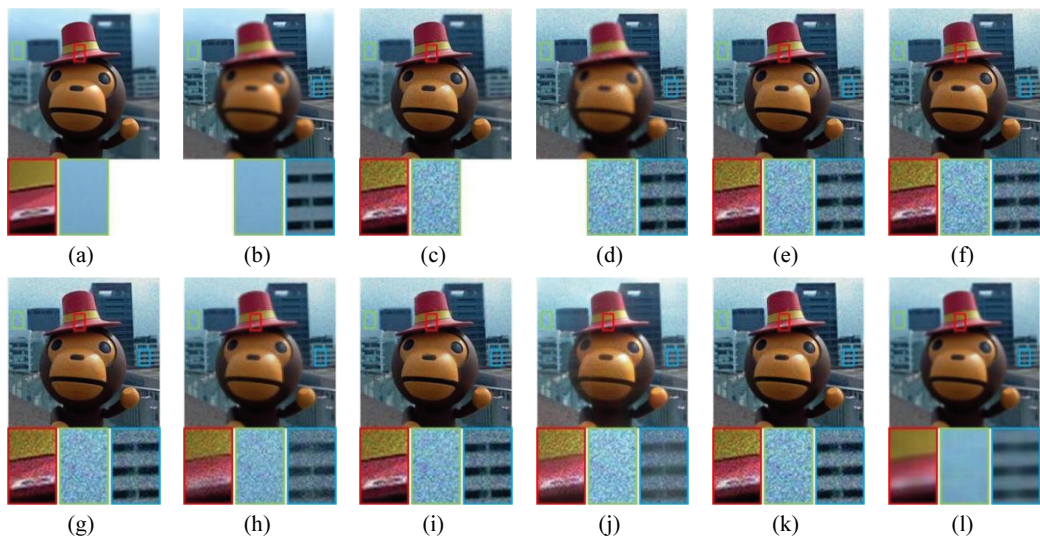

Fig. 7. Comparison results of the first group: (a) front-focused, (b) back-focused, (c) noise1 added, (d) noise2 added, (e) LP_SR, (f) NSCT_SR, (g) DTCWT, (h) ASR, (i) CSR, (j) LatLRR, (k) MDLatLRR, and (l) proposed method.

Fig. 8. Comparison results of the second group: (a) top-focused, (b) bottom-focused, (c) noise1 added, (d) noise2 added, (e) LP_SR, (f) NSCT_SR, (g) DTCWT, (h) ASR, (i) CSR, (j) LatLRR, (k) MDLatLRR, and (l) proposed method.
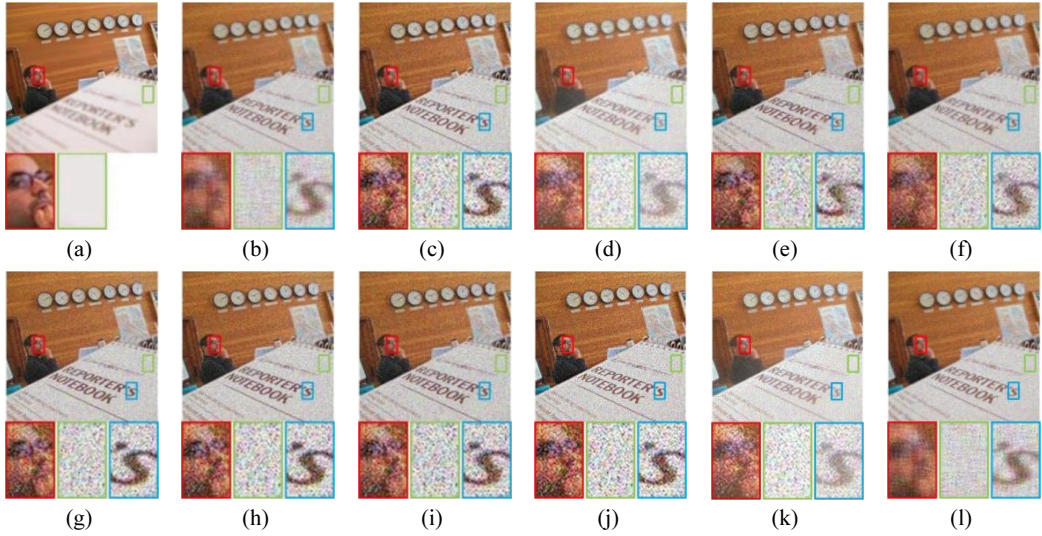
Table 1 presents the objective assessment indices for the various fusion methods applied to the two sets of images. Based on the objective data in Table 1, it is evident that the proposed method outperformed most indicators. Notably, in the second group of experimental results, the proposed method surpassed Q_{MI} and Q_Y, which indicates its superior ability to preserve the information and structure of the source image. In addition, the proposed method exhibited relative advantages in terms of the Q_Y index, indicating its ability to extract edge information from the source image. However, the proposed method lagged in the Q_{CB} index, suggesting that it reduces the contrast of the fusion results after effective denoising. This analysis demonstrates that the proposed method is an effective noisy multifucus image fusion method that can handle noise.

Table 1. Objective assessment of the two sets of images

Fusion method	The first group				The second group			
	Q_{MI}	Q_G	Q_Y	Q_{CB}	Q_{MI}	Q_G	Q_Y	Q_{CB}
LP_SR	0.489	0.173	0.318	0.540	0.220	0.080	0.130	0.437
NSCT_SR	0.501	0.178	0.331	0.529	0.235	0.088	0.146	0.445
DTCWT_SR	0.507	0.185	0.338	**0.552**	0.242	0.092	0.150	**0.456**
ASR	0.545	0.172	0.324	0.468	0.284	0.084	0.133	0.400
CSR	0.534	0.186	0.352	0.531	0.258	0.091	0.147	0.430
LatLRR	0.502	0.150	0.301	0.438	0.298	0.085	0.139	0.410
MDLatLRR	0.494	0.171	0.325	0.514	0.236	0.083	0.140	0.435
Proposed method	**0.532**	**0.217**	**0.414**	0.432	**0.475**	**0.219**	**0.244**	0.397

The bold font indicates the best performance in each test.

Overall, in a subjective comparison, the method proposed in this paper exhibited significantly superior denoising effects compared to the other methods. Subjective visual perception revealed almost no

perceptible graininess, and the cleanliness of the image was highly similar to that of the original image without noise contamination. Furthermore, the proposed method effectively preserved the overall contour of the image, resulting in the highest target recognition ability. Regarding the objective comparison, the proposed method outperformed the other Q_{MI}, Q_G, and Q_Y metrics in both datasets, with a slight disadvantage in the Q_{CB} metric.

6. Conclusion

This paper proposes a noisy multifucus image fusion method using a NSST framework. Compared to other classical image fusion methods, our proposed method can effectively remove noise while maintaining a certain degree of target discrimination in multifucus image fusion. Unlike traditional filter denoising, which targets only specific types of noise, such as Salt-and-Pepper noise using a median filter, the proposed method is generalizable and demonstrates a better effect on all types of noise, improving scene recognition accuracy and playing a unique and essential role in image recognition and decision-making applications. However, although the proposed method effectively removes noise, it also removes detailed information from the image, resulting in a blurred display of the target. The focus of the subsequent research stage will be to retain the detailed information of the image while removing noise.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

References

[1] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: a survey of the state of the art," *Information Fusion*, vol. 64, pp. 71-91, 2020. https://doi.org/10.1016/j.inffus.2020.06.013

[2] B. Li, H. Peng, and J. Wang, "A novel fusion method based on dynamic threshold neural P systems and nonsubsampled contourlet transform for multi-modality medical images," *Signal Processing*, vol. 178, article no. 107793, 2021. https://doi.org/10.1016/j.sigpro.2020.107793

[3] B. Li, H. Peng, X. Luo, J. Wang, X. Song, M. J. Perez-Jimenez, and A. Riscos-Nunez, "Medical image fusion method based on coupled neural P systems in nonsubsampled shearlet transform domain," *International Journal of Neural Systems*, vol. 31, no. 1, article no. 2050050, 2021. https://doi.org/10.1142/S0129065720500501

[4] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884-892, 2010. https://doi.org/10.1109/TIM.2009.2026612

[5] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Processing*, vol. 9, no. 5, pp. 347-357, 2015. https://doi.org/10.1049/iet-ipr.2014.0311

[6] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882-1886, 2016. https://doi.org/10.1109/LSP.2016.2618776

[7] H. Li, X. J. Wu, and J. Kittler, "MDLatLRR: a novel decomposition method for infrared and visible image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4733-4746, 2020. https://doi.org/10.1109/TIP.2020.2975984

[8] G. Easley, D. Labate, and W. Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25-46, 2008. https://doi.org/10.1016/j.acha.2007.09.003

[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, 2006. https://doi.org/10.1109/TSP.2006.881199

[10] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Information Fusion*, vol. 25, pp. 72-84, 2015. https://doi.org/10.1016/j.inffus.2014.10.004

[11] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 94-109, 2012. https://doi.org/10.1109/TPAMI.2011.109

[12] S. Kollem, K. R. Reddy, and D. S. Rao, "Improved partial differential equation-based total variation approach to non-subsampled contourlet transform for medical image denoising," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2663-2689, 2021. https://doi.org/10.1007/s11042-020-09745-1

[13] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147-164, 2015. https://doi.org/10.1016/j.inffus.2014.09.004

[14] H. Li and X. J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018 [Online]. Available: https://arxiv.org/abs/1804.08992.

Xin Feng https://orcid.org/0000-0001-8793-3775

He received his Ph.D. in Computer Science and Engineering from Lanzhou University of Technology in 2012. Since March 2013, he has been teaching at the School of Chongqing Technology and Business University. His main research is in the areas of technology, computer control, and image processing.

Haifeng Gong https://orcid.org/0000-0002-2589-5483

He received his Ph.D. in Oil and Gas Storage and Transportation Engineering from the School of Logistics Engineering in 2010. He is currently a Professor at the Engineering Research Centre for Waste Oil Recovery Technology and Equipment of the Ministry of Education, Chongqing. His main research interests include intelligent multiphase flow separation equipment and technology.

Guohan Qui https://orcid.org/0000-0002-3273-5090

He received his B.S. from the School of Smart Grid Information Engineering, Zhengzhou University of Light Industry in 2021. Currently, he is studying towards his M.S. at Chongqing Technology and Business University. His main research interests include image fusion and image processing.

Kaiqun Hu https://orcid.org/0000-0002-5590-2584
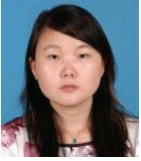
She received her Ph.D. from China Agriculture University in 2011. Currently, she is a lecturer, teaching at the School of Chongqing Technology and Business University. Her main research interests include computer control technology and image processing.