Infrared Weak-Small Targets Fusion Based on Latent Low-Rank Representation and DWT

XIAOZHU WANG[1], **JIANFEI YIN**[2], **KAI ZHANG**[1], **SHAOYI LI**[1], AND **JIE YAN**[1]

[1]School of Astronautics, Northwestern Polytechnical University, Xi'an710072, China
[2]Shanghai Institute of Spaceflight Control Technology, Shanghai 201109, China

Corresponding author: Xiaozhu Wang (wangxz@mail.nwpu.edu.cn)

ABSTRACT For the problem of anti-background interference of weak-small targets in infrared images, target extraction and texture detail processing are key tasks in the image fusion algorithm. The single-band infrared data can not fully reflect image details and contour information. There are texture differences in different bands of data, which makes it difficult to recognize targets. Therefore, it is necessary to fuse dual-band data to identify weak-small targets clearly. To solve these question, in this paper, we propose an effective image fusion framework using Latent Low-Rank Representation (LatLRR) and Discrete Wavelet Transform (DWT). Firstly, all source images are trained as L matrix by LatLRR which is used to extract salient features. And the original images are decomposed into high frequency and low frequency by DWT. Then high frequency parts are fused by maximum absolute value and low frequency parts are fused by weighted-average. On this basis, the training matrix L and high frequency fusion parts are used for contrast modulation fusion. Finally, the fused image is reconstructed by combining the contour parts and feature parts. The experimental results demonstrate that our proposed method achieves state-of-the-art performance in objective and subjective assessment.

INDEX TERMS Image fusion, infrared dual-band image, latent low-rank representation, discrete wavelet transform.

I. INTRODUCTION

In infrared dual-band sensor, the infrared dual-band image fusion is an important task. Infrared image reflects target and background radiation character. For single-band image, if the image has the weak-small target, it is easy to be submerged because of complex background. So the fusion of weak-small target needs comparison and analysis for the dual-band image.

At present, the fusion algorithms of infrared and visible imaging are more common. Because visible image can reflect the detail information of all scene, but visible light transmission is easily affected by the environment, resulting in limited detection distance. Infrared technique detects a long distance and uses thermal radiation to identify target. The disadvantage is that the image contrast is poor and the detail information is not rich. For infrared technology detection distance, the paper using infrared dual-band detector obtains medium wave image (3-5um) and long wave image (8-14um).

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

The image fusion method is used to enhance the detail and improve the recognition rate of the target.

As we all know, the extraction and processing of the features are key tasks in dual-band infrared image fusion, and the fusion performance is directly affected by the different features and processing methods undertaken.

A. ANALYSIS OF IMAGE FUSION ALGORITHM

Because the infrared dual-band image fusion algorithms are much fewer. So this paper learns from infrared and visible image fusion algorithm. At present, the image fusion methods are mainly divided into three categories which include multi-scale transformation method, sparse representation method and neural network method.

Multi-scale transform is a better algorithm for extracting image detail parts. It can decompose original images into components of different scales, where each component represents the sub-image at each scale and real-world objects typically comprise components at different scales. Many studies have shown that multi-scale transforms are consistent

with human visual characteristics, and this property can enable fused images to have good visual effect, such as Pyramid transform [1], Discrete wavelet transform (DWT) [2], Stationary wavelet transform (SWT) [3], Non-subsampled contour transform (NSCT) [4], [5] and Shift-invariant shearlet transform (NSST) [6] etc.

Because the multi-scale method can extract the edge and texture details of the image, so Huang *et al.* proposed an infrared and visible image fusion method that was based on curvelet transformation and visual attention mechanisms. Their model could elevate the signal-to-noise ratio of fused images and highlight dim targets [7]. Zhu *et al.* proposed an infrared and visible image fusion method that was based on an improved multi-scale top-hat transform model, this model could highlight the target of infrared images and preserve details of visible images [8].

Sparse representation is an effective tool for characterizing the human visual system and has been successfully applied in different fields, such as machine learning, image analysis, and pattern recognition. Sparse representation image fusion algorithms aim to learn an over-complete dictionary from a large number of high-quality natural images. Then, the source images can be sparsely represented by the learned dictionary, thereby potentially enhancing the representation of meaningful and stable images [9]. Meanwhile, sparse representation-based fusion methods divide source images into several overlapping patches by using a sliding window strategy, thereby potentially reducing visual artifacts and improving robustness to mis-registration [10].

There are many methods based on combining dictionary construction methods for image fusion, such as adaptive sparse representation [11], multi-scale dictionary learning [12], PCA [13], [14] and online dictionary learning [15]. Yang and Li took the first step to employ the SR theory for image fusion [16]. The main problems of SR-based fusion method are the sparse coding of the sparse coefficients and the dictionary construction algorithm. The fusion result is achieved by a linear combination of the sparse coefficients and the constructed dictionary. Yin *et al.* proposed a multi-scale dictionary learning method by integrating the wavelet into dictionary learning, thus potentially taking advantage of multi-scale representation and dictionary learning. Li and Wu [17] proposed a low-rank representation (LRR) based fusion method. They use LRR instead of SR to extract features, then $l_1 - norm$ and the max selection strategy are used to reconstruct the fused image, but this method does not go into the details. In [18], [19], Li *et al.* proposed a Latent Low-Rank Representation fusion algorithm. The source images decomposed into low-rank parts and saliency parts. The low-rank parts are fused by weighted-average strategy, and the saliency parts are simply fused by sum strategy. Han *et al.* [20] proposed an adaptive two-scale image fusion method. The algorithm decomposes infrared and visible images into a two-scale representation using LatLRR to generate low-rank parts and saliency parts, with respect to the fusion rule of the low-rank parts, then

construct adaptive weights by adopting fusion global-local-topology particle swarm optimization to obtain more useful information from the source images. However, in the two algorithms, the fusion image contains few details of thermal radiation. And Li *et al.* [21] proposed a novel deep decomposition method. The LatLRR algorithm is utilized to learn a project matrix which is used to extract salient features. The base part and multi-level detail parts are obtained by deep decomposition LatLRR. The fusion algorithm is useful for image detail extraction.

A neural network usually consists of many neurons, thus it can imitate the perception behavior mechanism of the human brain to deal with neuron information. The interactions among neurons characterize the transmission and processing of neuron information, and the neural network has the advantages of strong adaptability, fault tolerance and anti-noise capabilities.

With the rise of deep learning, deep features of source images are used to reconstruct the fused image. Liu and Wu [22] proposed a convolutional neural network (CNN) based fusion method. Image patches which contain different blur versions are used to train a network and obtain a decision map. The fused image is obtained by using the decision map and source images. But the fusion result has a poor contrast. The residual network (ResNet) and zero-phase component analysis (ZCA) were applied into image fusion tasks by Liu *et al.* [23]. The ResNet is used to extract deep features from source images, then ZCA is utilized to normalize the deep features and obtain initial weight maps. Finally the fused image is reconstructed using a weighted-averaging strategy. The result achieves less performance in visual quality. Li and Wu [24] proposed a densefuse to infrared and visible images. They propose a novel deep learning architecture which is constructed by encoding network and decoding network. Encoding network is used to extract image features and decoding network is used to obtain fused image. Because the encoding network and decoding network has more layers, so the algorithm operation is low. Ma *et al.* [25] proposed a generative adversarial network for infrared and visible image fusion. The method establishes an adversarial game between a generator and a discriminator, where the generator aims to generate a fused image combine major infrared intensities with additional visible gradients, and the discriminator aims to force the fused image to have more details existing in visible images. Though the algorithm keeps thermal radiation of infrared image, but many details are lost in the fusion of infrared dual-band images.

B. RESEARCH ON INFRARED DUAL-BAND IMAGE FUSION ALGORITHM

Although multi-scale transforms, sparse representation and deep learning methods obtain good fusion performance, but for infrared images fusion, the infrared image is different from the visible image. Visible light fully reflects the detailed information of the whole scene, but visible light transmission is difficult to penetrate fog, rain and dust, thus the detection

distance is limited. However infrared technology uses thermal radiation to convert infrared band information, that is more than the human eye to catch the visible information. Infrared technology can identify hot targets well, and the detection is far, but the image contrast is poor and the details are not rich.

At present, for infrared dual-band image fusion algorithm, Qu *et al.* [26] proposed dual-band infrared images fusion based on gradient pyramid decomposition. The method uses four Gradient operators, based on the Gaussian pyramid decomposition, to make a filtering in the horizontal, vertical, and two diagonal directions. By this way, they can extract the edge information of the source image well, and keep the details of the characteristics. The fused image has a better definition and contains enough effective messages. But the edges of the fused image are not clear. And for dim small targets fusion Method. Sun *et al.* [27] proposed dim small targets detection based on dual-band infrared image fusion. They use wavelet transformation to decompose the source images. In wavelet transformation domain high-frequency part, they calculate local fractal dimension and set up fusion rule to merge corresponding sub-images of two source images. In low frequency, they extract local maximum gray level to fuse them. Then reconstruct image by wavelet inverse transformation and obtain fused image. However, the background noise of the fused image is the same size as the target, so the target cannot be distinguished.

For a long distance target in complex background, a weak-small target is formed in the infrared dual-band image, which is mainly reflected in two aspects. On the one hand, weak-small targets have a small imaging area and only occupy one or a few pixels on the detector, lacking shape and structural information. Target detection often cannot directly utilize pattern, size and other features for pattern recognition. On the other hand, weak-small target in the image is usually the farthest distance imaging. After a long distance of atmospheric attenuation, the energy reaching the imaging system is usually weak, so the signal intensity of the target is weak. Since the background and noise occupy a large proportion in the field of view, the signal-to-noise ratio of the whole image is relatively low.

Infrared image fusion algorithms are applied to the weak-small target image fusion, those methods still have drawbacks. Fusion image background noise amplifies while detail information is lost. Therefore, it is necessary to propose a new method suitable for weak-small target image fusion. Eventually, the signal-to-noise ratio of the fused image is improved, and the intensity of the weak-small target is highlighted.

Based on the above questions, we propose a fusion algorithm based on LatLRR and DWT for infrared dual-band image fusion. The algorithm uses DWT to obtain the image saliency part, then uses LatLRR to extract the depth features of the saliency part. Firstly, the original image is decomposed into high frequency and low frequency by DWT. High frequency displays saliency parts and low frequency displays

contour parts. The LatLRR is utilized to learn a project matrix which is used to extract salient features. Secondly, high frequency parts are fused by maximum absolute value and low frequency parts are fused by weighted-average, The training matrix and high frequency fusion parts are used for contrast modulation fusion. Finally, the fused image is reconstructed by combining the contour parts and the detail parts. In fusion result, the contrast between target and background has obvious changes.

The rest of this paper is arranged as follows. Section 2 describes a brief introduction of related works. Section 3 introduces fusion framework based on LatLRR and DWT. Section 4 shows the experiments results. Section 5 draws the conclusions.

II. RELATED WORKS
A. FEATURE EXTRACTION ALGORITHM
At present, sparse representation and deep learning are commonly used as feature extraction methods. Liu and Yan [28] proposed Latent Low-Rank Representation for subspace segmentation and feature extraction. Li *et al.* [21] used the residual network (ResNet) to solve the image degradation problem, and extract the depth features through fast connections and residual representations. Liu *et al.* [29] used VGG19 to extract the depth features in the image and adopted multi-level fusion for the decomposed details. Ma *et al.* [25] used GAN to build a counter generator and discriminator to extract infrared radiation characteristics in the image.

We select different algorithms to do contrast experiments for infrared weak-small targets dual-band images. Through these experimental results, we can find that the LatLRR method is best than the other algorithms. Because LatLRR algorithm is an unsupervised feature extraction method, combining subspace segmentation with feature extraction can easily extract the salient features in the image. Given a source image, LatLRR decomposes it into the principal features: the salient features and the noise. The salient features are automatically extracted from original image so as to produce effective features for recognition.

B. IMAGE DECOMPOSITION THEORY
Image decomposition is a crucial part of image fusion processing. Multi-scale transforms can decompose original images into components of different scales. In this paper, we select various multi-scale decomposition methods for infrared images. Finally, the DWT is selected through comparative analysis. Because DWT can greatly remove the correlation between extracted different features by selecting appropriate filters. The high frequency wavelet coefficients in the image contain the overall details. The DWT in image processing is given as follows.

Given separable two-dimensional scaling and wavelet functions, the scaled and translated functions are:

$$\varphi_{j,m,n}(x, y) = 2^{j/2}\varphi(2^j x - m, 2^j y - n) \tag{1}$$

$$\psi^i_{j,m,n}(x, y) = 2^{j/2}\Psi(2^j x - m, 2^j y - n) i = \{H, V, D\} \tag{2}$$

where $\varphi_{j,m,n}(x,y)$ and $\psi^i_{j,m,n}(x,y)$ are, respectively, scale functions and wavelet functions at different scales and positions. j is the scale order. The larger the j is, the smaller the scale will be, which is equivalent to that the higher frequency makes it closer to the detail. m and n are the offsets of the position.

$f(x,y)$ is an intensity image of size $M \times N$, and then the discrete wavelet transform of the image is:

$$W_\varphi(j_0, m, n) = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\varphi_{j_0,m,n}(x,y)}{\sqrt{MN}} \quad (3)$$

$$W^i_\psi(j, m, n) = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\psi^i_{j,m,n}(x,y)}{\sqrt{MN}} \quad (4)$$

$W_\varphi(j_0, m, n)$ is an approximation of $f(x,y)$ at scale j_0. Its horizontal, vertical and diagonal details for scale $j \geq j_0$ are $W^i_\psi(j, m, n)$ coefficients. Given W_φ and W^i_ψ of Equations (3) and (4), $f(x,y)$ is obtained via the inverse discrete wavelet transform by the following formula:

$$f(x,y) = \frac{\sum_m \sum_n W_\varphi(j_0, m, n)\varphi_{j_0,m,n}(x,y)}{\sqrt{MN}}$$
$$+ \frac{\sum_{i=H,V,D}\sum_{j=j_0}^{\infty}\sum_m \sum_n W^i_\psi(j, m, n)\psi^i_{j,m,n}(x,y)}{\sqrt{MN}} \quad (5)$$

III. THE PROPOSED FUSION METHOD

In this section, the proposed algorithm based on LatLRR and DWT fusion is introduced in detail. The decomposition method and fusion strategies for contour parts and saliency parts will be presented in the next subsection.

The frame of image fusion algorithm is shown in Fig. 1. Input infrared images are medium wave image and long wave image. All process divided into two parts. In the first part, all source images are trained as L matrix by LatLRR. In the second part, medium wave image and long wave image are decomposed to low frequency (contour) and high frequency (saliency), respectively, by DWT. Contour feature uses Weighted-average strategy fusion, and saliency feature uses Maximum absolute value strategy fusion. Then L matrix is used to extract the detailed information. Gray modulate is used to extract deep detail feature by saliency feature and detail information. Finally, the fused image is reconstructed by adaptive strategies.

A. TRAINING THE PROJECT MATRIX L

LRR is an effective method for exploring the multiple subspace structures of data. Usually, the observed data matrix itself is chosen as the dictionary, which is a key aspect of LRR. But this representation method can not preserve the local structure information. In 2011, Liu and Yan [28] proposed LatLRR theory, LatLRR seamlessly integrates subspace segmentation and feature extraction into a unified framework. As a subspace segmentation algorithm, LatLRR is an enhanced version of LRR and outperforms the state-of-the-art algorithms. Being an unsupervised feature extraction algorithm, LatLRR is able to robustly extract salient features from corrupted data, and thus can work much better than the benchmark that utilizes the original data vectors as features for classification. So the low-rank structure and salient structure can be extracted by LatLRR from raw data. The LatLRR problem is reduced to solve the following optimization problem.

$$\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda\|E\|_1 \quad (6)$$

$$s.t. \ X = XZ + LX + E \quad (7)$$

where $\lambda > 0$ is the balance coefficient, $\|\cdot\|_*$ denotes the nuclear norm which is the sum of the singular values of matrix and $\|\cdot\|_1$ is $l_1 - norm$. X denotes observed data matrix, Z is low-rank coefficients, XZ denotes principal features, L is a project matrix which is named salient coefficients, LX denotes salient features, and E is sparse noisy matrix.

Equation (6) is solved by the inexact Augmented Lagrangian Multiplier (ALM) [29]. The salient component LX is obtained by (7).

In this paper, the source image is divided into M image patches, and the size of image patch is $n \times n$. X indicates the observed matrix and each column denotes an image patch. The size of project matrix L is just related to image patch size. Once the project matrix is learned by LatLRR, it can be used to process other images which are arbitrary size.

We use thermal imager (IRCAM Equus 327) to collect infrared medium wave and long wave images under different backgrounds. At the same time, we also select some infrared and visible images by downloading the TNO Image Fusion Dataset [30]. The project matrix L is learned by LatLRR, where the size of L is just related with image patch.

In the learning phase, all training images are divided into image patches by sliding windows technique without overlapping. We select $16*16$ image patch size. 1200 image patches are randomly chosen to generate an input matrix X in which each column indicates all pixels of one image patch. Then the size of X is $N*M$, where $N = n \times n$ and $M = 1200$.

The project matrix L could be learned by LatLRR. The $256*256$ size of L will be obtained to use extract features.

B. DISCRETE WAVELET TRANSFORM (DWT) DECOMPOSITION THEORY

Because the dual-band infrared images quantization digits are uniform, the radiation in the high-temperature scene is mainly concentrated on the middle band. Its corresponding radiation contrast of the long band is relatively small. So the quantized target cannot be distinguished from the scene and vice versa at low temperatures.

In order to enhance the contrast between weak-small target and background in the fused image, and the detector can clearly identify target. Based on this, we use the discrete
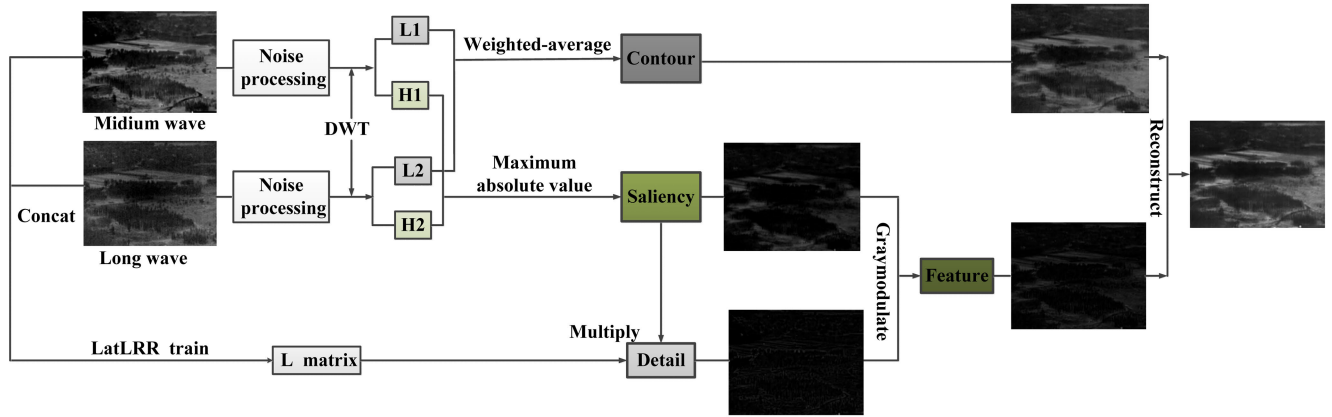
FIGURE 1. The framework of image fusion algorithm.

wavelet transform decomposition theory, because discrete wavelet transform has good localization properties in time domain and frequency domain [31]. High frequency components using subtle time domain sampling step, it can be focused to any details of an object, making a pace of time domain and frequency domain. It also can decompose a signal into sub-signals with different spatial resolution and frequency domain characteristics, without losing the information contained in the original signal, and find orthogonal basis to realize non-redundant signal decomposition. Therefore, we choose discrete wavelet transform to decompose the image.

Image decomposition method is shown in Fig. 1. In the medium and long wave thermal imager, random noise is generated duo to the influence of the detector effect which causes noise in the image. Therefore, it is necessary to remove the image surface noise. The paper uses the median filter function to filter out the noise in the image. Because median filtering is a nonlinear smoothing technique, it can effectively suppress noise and protect edge information [32]. Assume medium wave image and long wave image are A and B, respectively. The fusion steps are as follows:

The low frequency parts adopt the weighted-average fusion method based on local energy.

1) The wavelet coefficients of the low frequency parts are extracted from the results of A and B wavelet transform.

2) For the two images A and B, the energy of the local region within the low frequency band is calculated respectively.

$$E(A, p) = \sum_{q \in Q} w(q) C_J^2(A, q) \quad (8)$$

$$\sum_{q \in Q} w(q) = 1 \quad (9)$$

where $w(q)$ denotes weight. When the point q is closer to the point p, the weight is larger. Q is a neighborhood of p. $E(B, p)$ can be obtained by the same principle.

3) Define a match matrix:

$$M(p) = \frac{2}{E(A, p) + E(B, p)} \cdot \sum_{q \in Q} w(q) C_J(A, q) C_J(B, q) \quad (10)$$

The value of each point in the matching matrix changes between 0 and 1, and the closer to 1 means the higher the degree of correlation.

4) Define thresholds TH, if $M(p) < TH$, so:

$$C_J(f, p) = \begin{cases} C_J(A, p), & E(A, p). \geq E(B, p) \\ C_J(B, p), & E(B, p). < E(A, p) \end{cases} \quad (11)$$

Otherwise:

$$C_J(F, p) = \begin{cases} if\ E(A, p). \geq E(B, p) \\ W_{max} C_J(A, p), \qquad W_{min} C_J(b, p) \\ else\ if\ E(B, p). \geq E(A, p) \\ W_{max} C_J(B, p), \qquad W_{min} C_J(A, p) \end{cases} \quad (12)$$

Among:

$$W_{min} = 0.5 - 0.5 \left(\frac{1 - M(p)}{1 - TH} \right) \quad (13)$$

$$W_{max} = 1 - W_{min} \quad (14)$$

The high frequency parts adopt the fusion method of select max absolute value:

$$D_F = \max(D_A, D_B) \quad (15)$$

D_A, D_B, D_F denote image A, image B, fusion image F.

C. THE ANALYSIS PRODUCE OF SALIENCY PARTS

In order to make weak-small targets clear and prevent the background noise amplification in the fused image, we process the image after high frequency fusion. Firstly, we use Matrix L to learn by LatLRR from training data (medium wave and long wave images). Secondly, Image details information is obtained by Matrix L and high frequency fusion result. On this basis, image details information and high frequency fusion result are used to obtain image features by grayscale contrast modulation. The process of saliency parts are shown in Fig. 2.

The detail parts are calculated by (16):
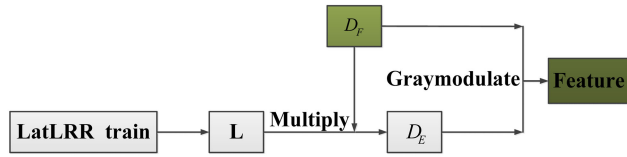
$$D_E = L \times D_F \quad (16)$$

FIGURE 2. The process of saliency parts.

where D_F and L denote, respectively, the high frequency fusion result and the project matrix which is learned by LatLRR. D_E is the extracted detail information.

At the same time, the high frequency fusion result D_F is established as Gaussian decomposition, then it is filtered in two dimensions. The filter window function uses the template $5 * 5$. According to symmetry, normalized type and other relevant constraints, the template can be obtained as follows:

$$w = \frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (17)$$

$D_F(i, j)$ is filtered by a two dimensional F filter in the matrix w, then $D_{F1}(i, j)$ is obtained by two dimensional cross-correlation. Namely:

$$D_{F1}(i, j) = filter2(w, D_F(i, j)) \quad (18)$$

Gray contrast formula is used to process the image to obtain:

$$C(i, j) = \frac{D_F(i, j) - D_{F1}(i, j)}{D_{F1}(i, j)} \quad (19)$$

The contrast $C(i, j)$ is extracted by gradient operator and discretized by gray level. New grayscale $I(x, y)$:

$$I(x, y) = \frac{C(i, j) - C(i, j)_{\min}}{C(i, j)_{\max} - C(i, j)_{\min}} \quad (20)$$

Gray maximum is $C(i, j)_{\max} = \max_{x,y} C(i, j)$; gray minimum is $C(i, j)_{\min} = \min_{x,y} C(i, j)$.

The grayscale image $I_1(x, y)$ is convolved by $I(x, y)$ and detail information D_E.

$$I_1(x, y) = I(x, y) * D_E(x, y) \quad (21)$$

The grayscale discretization of $I_1(x, y)$ is performed by the gradient operator, the gray value is enhanced by contrast, and the fused image based on the gray contrast is obtained $I_2(x, y)$:

$$I_2(x, y) = 255 \cdot \frac{I_1(i, j) - I_1(i, j)_{\min}}{I_1(i, j)_{\max} - I_1(i, j)_{\min}} \quad (22)$$

D. RECONSTRUCTION

Once the fused detail feature $I_2(x, y)$ is obtained, we use the fused detail feature $I_2(x, y)$ and low frequency fused part $D_{L(x,y)}$ to reconstruct the final fused image, as shown in (23):

$$F(x, y) = I_2(x, y) + D_{L(x,y)} \quad (23)$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The aim of the experiment is to validate the proposed method using objective criteria and to carry out a comparison with existing methods.

A. EXPERIMENTAL SETTINGS

In our experiment, as show in Fig. 3 and Fig. 4, we choose infrared medium wave and long wave images based on different backgrounds, and we also choose some infrared and visible images for comparison.

For comparison, we select several recent and classical fusion methods to perform the same experiment, including: principal component analysis (PCA) [33], non-subsampled contourlet transform (NSCT) [4], visual saliency map and weighted least square optimization (VS-WLSO) [34], ResNet and zero-phase component analysis (ResNet-ZPCA) [21], Latent Low-Rank Representation (LatLRR) [28], novel deep decomposition method (NDD) [22], Deep Learning Framework (DLF) [19], generative adversarial network (GAN) [25], DenseFuse (DF) [24].

For all fusion algorithms. GAN and DF are implemented with Tensorflow and GTX 1080Ti, 16GB RAM. ResNet-ZPCA and DLF are implemented in MATLAB R2018a and Visual Studio 2015. Other fusion algorithms are implemented in MATLAB R2018a.

B. SUBJECTIVE EVALUATION

The main purpose of the subjective visual effect evaluation is to discriminate the sharpness of the target in the image, the contrast of the target and the background. The general subjective quality assessment method is to judge the image quality by the observer's score normalization. The advantage is that it can truly reflect the visual quality of the image and the evaluation result is reliable. The disadvantage is that it cannot be described by applying a mathematical model, and it is difficult to achieve real-time quality evaluation. In this paper, nine kinds of contrast experiments are selected. Some algorithms are classic traditional fusion methods, and some are the latest image fusion algorithms proposed at this stage. Combining the algorithm proposed in this paper with the other nine algorithms, we select some observers to evaluate the results of ten algorithms. The evaluation is divided into three levels: good, general and poor.

The fusion results which are obtained by the nine existing methods and the proposed method are shown in Fig. 5. The first two columns, the fusion results on two typical infrared and visible image pairs from TNO database. After three columns, three typical infrared medium wave and long wave image pairs from collect data. For different test images, we evaluate the relative performance of the fusion methods using visual standard. As can be seen from the Fig. 5. The fused images contain many fuzzy phenomenon around the detail information by NSCT, ResNet-ZPCA, LatLRR and DLF methods. Although NDD and DF algorithms increased image grayscale, but the overall image details are not clear.
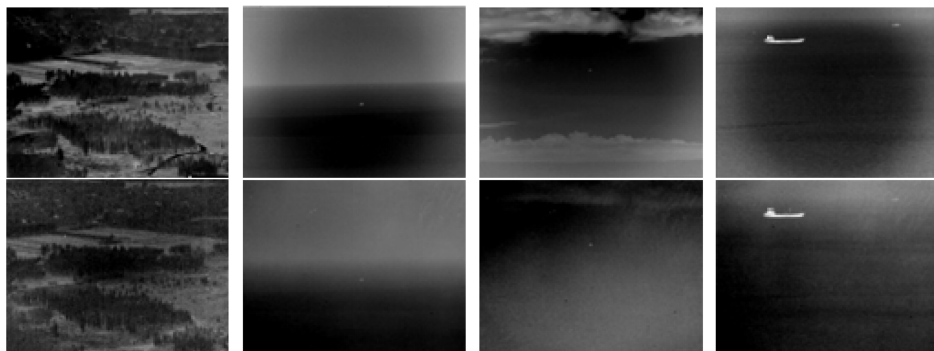
FIGURE 3. Four pairs of collect images. The top row contains infrared medium wave images, and the second row contains visible images.

FIGURE 4. Four pairs of source images. The top contains infrared images, and the second row contains visible images.

TABLE 1. Subjective visual evaluation result.

Evaluation index	Definition	Algorithm
Good	The fused image quality is good and the target is clear	Proposed algorithm LatLRR
General	The fused image quality is unchanged and the target is generally clear	PCA NSCT VS-WLSO NDD ResNet-ZPCA DLF
Poor	The fused image quality becomes worse and the target is blurred	GAN DF

Finally, through the normalization of the observer's evaluation, the subjective visual evaluation results are as shown in Table. 1.

In contrast, the best visual effect is obtained by our proposed method. The proposed method not only works better in terms of highlighting the target, but also enhance the image contrast and brightness. In summary, the proposed method is superior to other nine methods in both inheriting the features of source images and preserving important details. Meanwhile, the overall brightness of the fused image is improved by the proposed method due to adopting both the contour and saliency information.

C. OBJECTIVE EVALUATION

For infrared weak-small target images under different backgrounds, the following problems will appear in infrared fusion.

1) The infrared target is small and unclear, occupying only a few pixels in the image.

2) The remote contour of infrared image collected is not clear.

3) After fusion, the background noise is enlarged and the information is distorted.

So we put forward four evaluation indicators for specific problems.

1) For the infrared image target is small, only a few pixels in the image, the target is not clear. We propose information entropy (EN), which is an index to measure the richness of information contained in an image.

2) For the fused image which is easily distorted, we propose structural similarity ($SSIM$), which is used to evaluate the probability of image information being distorted. The larger value of $SSIM$ indicates the better image quality.

3) For the low signal-to-noise ratio of the fused image, we propose the peak signal to noise ratio ($PSNR$) index. The higher value of $PSNR$ indicates the better quality and effect of the fused image.
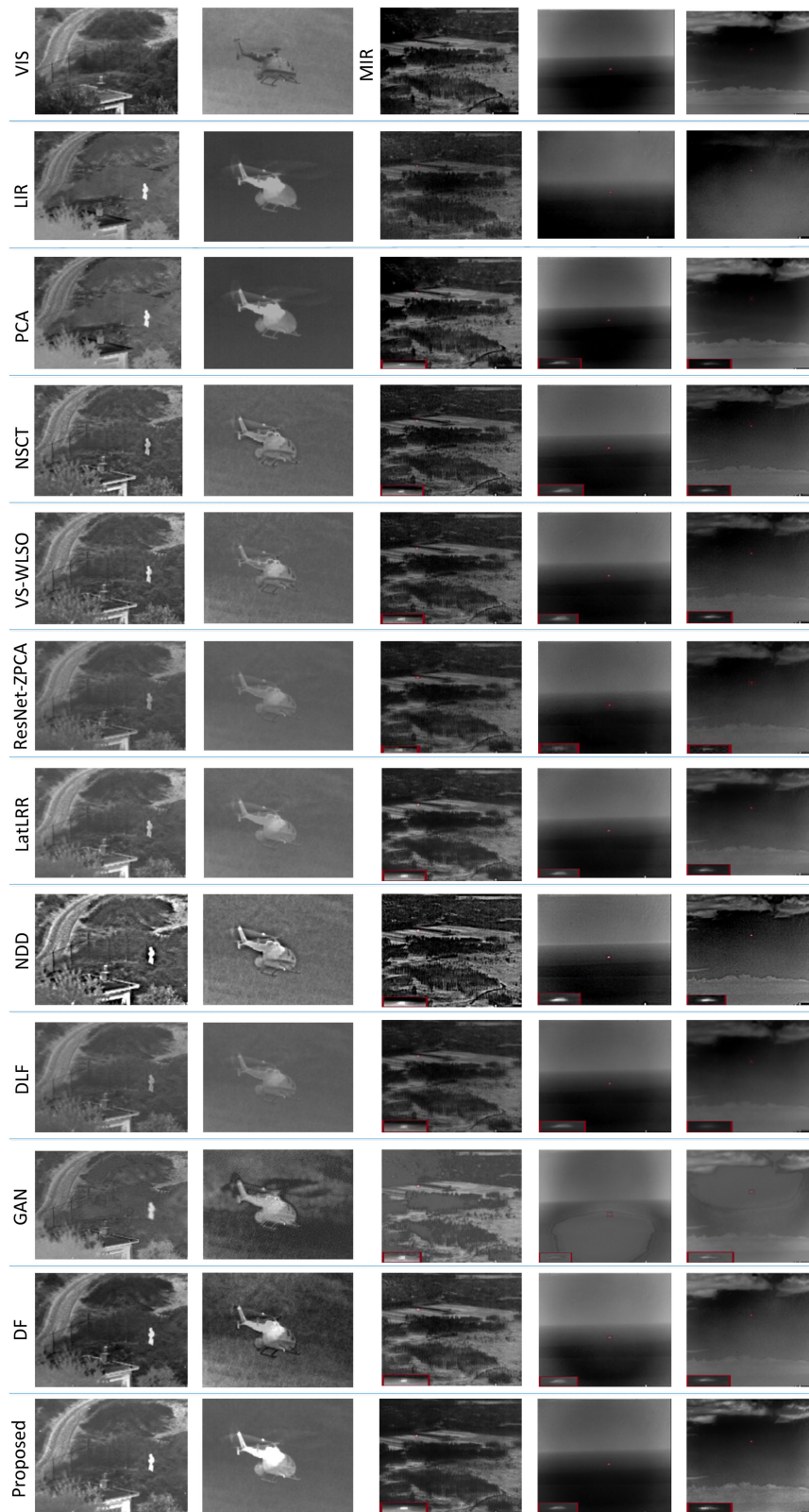
FIGURE 5. Comparison of the fusion results from different algorithms.
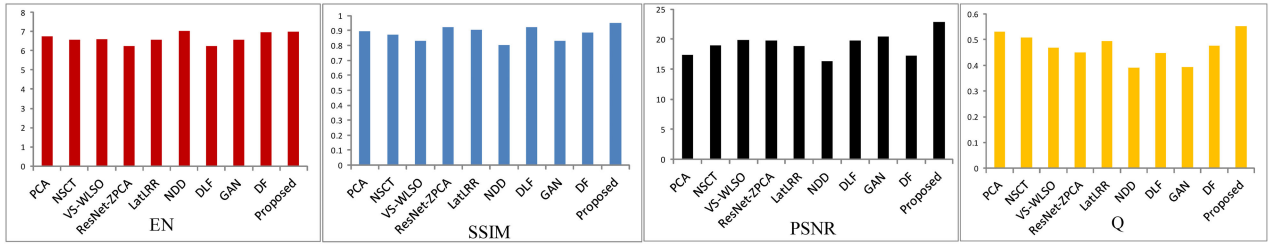
FIGURE 6. Quantitative comparisons of the first column image.

4) Aiming at the lost edge information of the fused image and the blurred image, we propose edge retention index $Q^{AB/F}$.

EN is defined based on information theory, which measures the amount of information the fused image contains [35]. Mathematically, *EN* is defined as follows:

$$EN = -\sum_{i=0}^{L-1} p_i \log_2 p_i \qquad (24)$$

where L denotes the number of gray level. We set it to 256 in our experiments. p_i is the normalized histogram of corresponding gray level in the fused image. The larger value of *EN* indicates the better performance fusion method.

SSIM is used to model the image loss and distortion, which measures the structural similarity between source images and fused images [36]. *SSIM* mainly consists of three components: loss of correlation, luminance distortion, and contrast distortion. The product of three components is the assessment result of the fused image, and *SSIM* is defined as follows:

$$SSIM_{X,F} = \sum_{x,f} \frac{2u_x u_f + C_1}{u_x^2 + u_f^2 + C_1}$$
$$\cdot \frac{2\sigma_x \sigma_f + C_2}{\sigma_x^2 \sigma_f^2 + C_2} \cdot \frac{\sigma_{xf} + C_3}{\sigma_x \sigma_f + C_3} \qquad (25)$$
$$SSIM = \left(SSIM_{A,F} + SSIM_{B,F}\right)/2 \qquad (26)$$

where $SSIM_{X,F}$ denotes the structural similarity between source image X and fused image F, x and f represents the image patch of the source image and fused image in a local window of size $M * N$, σ_x and σ_y denote the standard deviation, σ_{xf} is the standard covariance correlation of source and fused image, μ_x and μ_f denote the mean value of source image and fused image, respectively. C_1, C_2 and C_3 are parameters to make the algorithm stable. $SSIM_{A,F}$ and $SSIM_{B,F}$ denote the structural similarities between infrared/visible images and fused image.

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise [37]. The *PSNR* of the fusion result is defined as follows:

$$PSNR = 10\log_{10}\left(\frac{(f_{\max})^2 MN}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[R(m,n) - F(m,n)]^2}\right) \qquad (27)$$

where $R(m,n)$ and $F(m,n)$ are the reference and fused images, respectively. M and N are image dimensions. $f_{\max}$ is the maximum gray scale value of the pixels in the fused image.

$Q^{AB/F}$ is to evaluate the fusion performance by measuring how much edge information the fusion image obtains from the source image [38]. $Q^{AB/F}$ evaluation process is as follows. Firstly, the source images and fused images are edge-extracted. And then the edge retention between each source image and the fused image is calculated. Finally, the weighted edge retention is used as the final quantitative evaluation index. $Q^{AB/F}$ is defined as follows:

$$Q^{AB/F} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j))}{\sum_{i=1}^{N}\sum_{j=1}^{M}(w^A(i,j) + w^B(i,j))} \qquad (28)$$

$w^A(i,j)$ and $w^B(i,j)$ are the weights of the corresponding pixels. $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$ are similarity measures of input image and fused image. The larger index of $Q^{AB/F}$ indicates the more image edge information retained by the fused image.

The following figures show the evaluation indexes of ten algorithms for five images. The ten algorithms are principal component analysis (PCA), non-subsampled contourlet transform (NSCT), visual saliency map and weighted least square optimization (VS-WLSO), Latent Low-Rank Representation (LatLRR), novel deep decomposition method (NDD), ResNet and zero-phase component analysis (ResNet-ZPCA), Deep Learning Framework (DLF), generative adversarial network (GAN), DenseFuse (DF) and the proposed algorithm in order.

Fig. 6 and Fig. 7 show the comparison of evaluation indexes after the fusion of visible images and infrared images. It can be seen from the data that the algorithm proposed in this paper is better in all four evaluation indicators. Compared to other infrared and visible image algorithms, the algorithm highlights the infrared information of the target, retaining the details of the visible light.

Fig. 8 to Fig. 10 show the comparison of evaluation indexes after the fusion of medium wave images and long wave images in the different backgrounds. For weak-small targets, the proposed fusion algorithm obtains better results in *EN*, *SSIM*, *PSNR* and $Q^{AB/F}$. Analyze each fusion algorithm.
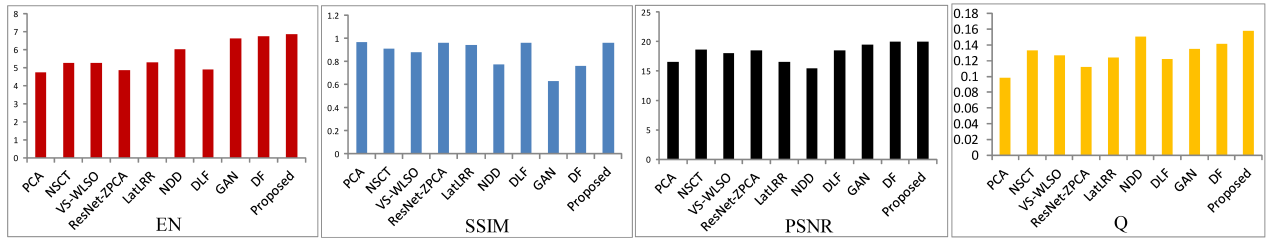
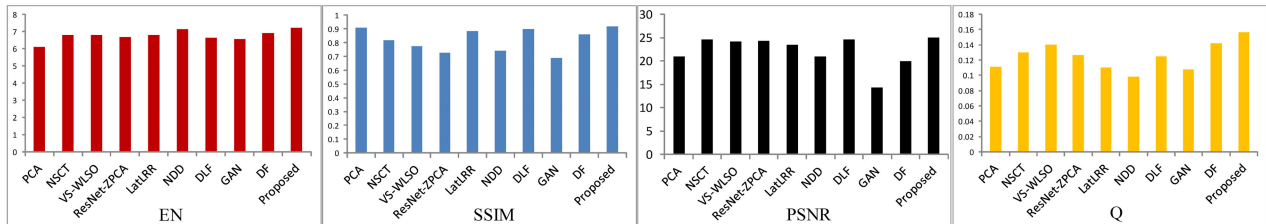**FIGURE 7.** Quantitative comparisons of the second column image.

FIGURE 8. Quantitative comparisons of the third column image.
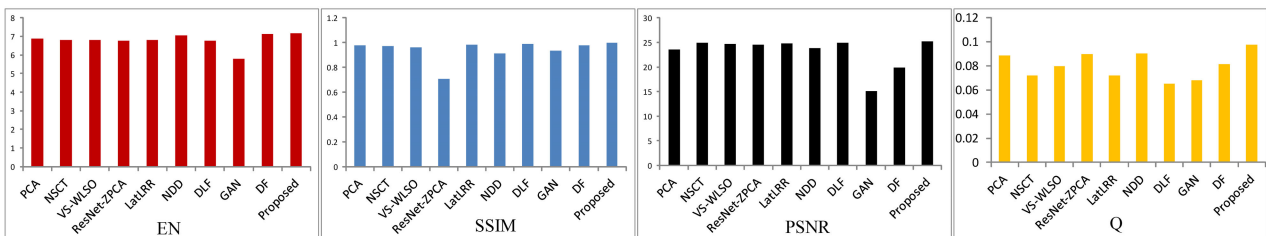

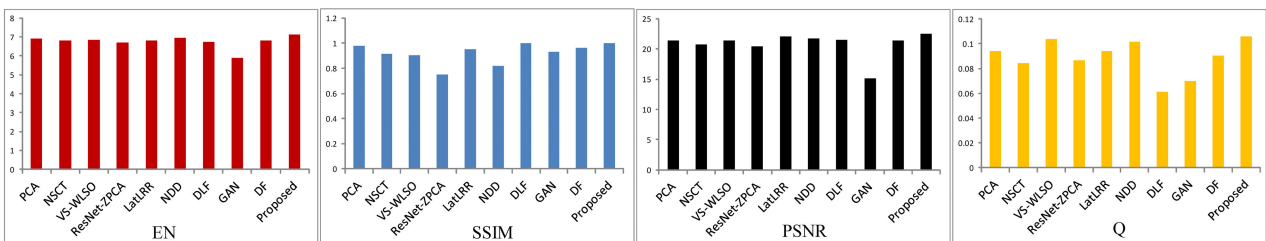**FIGURE 9.** Quantitative comparisons of the fourth column image.

FIGURE 10. Quantitative comparisons of the fifth column image.

The PCA method uses down-sampling processing to obtain an approximate image, and up-samples the approximate image to obtain image detail information. The final fused result has good indicators in *EN* and *SSIM*. The NSCT method is used to decompose image into a low frequency sub-band and a high frequency sub-band. The low frequency sub-band parts are fused by the weighted averaging method, and the high frequency sub-band parts are fused by the absolute value of the pixel. The final fused image has better indicators in *PSNR* and *EN*. The VS-WLSO method uses the rolling guidance filter and Gaussian filter to decompose input images into base and detail layers, the detail layers extract

multiple features. The fused image has better indicators in *EN*, *SSIM* and *PSNR*. The LatLRR method uses LatLRR to learn a project matrix which is used to extract salient features. The fused image has good indicators in *EN*, *SSIM* and *PSNR*. The NDD method uses deep decomposition LatLRR to obtain the base part and multi-level detail parts. With adaptive fusion strategies, the fused result has good indicators in *EN* and *PSNR*. The DLF method is used to decompose the image into base parts and detail content, the detail content use a deep learning network to extract multi-layer features. The fused image has better indicators in *SSIM* and *PSNR*. The ResNet is used to extract deep features from Source images. The ZCA

and $l_1 - norm$ are utilized to normalize the deep features and obtain initial weight maps. The fused result has a poor indicator in *SSIM*. The DF and the GAN methods are not adaptive to the network architecture, resulting in unsatisfactory image fusion results and poor evaluation indicators.

In particular, all algorithms are compared. The proposed method has rich image information, no distortion and good image quality. Overall, the proposed fusion algorithm has obvious advantages for the fusion of weak-small targets in infrared dual-band images.

V. CONCLUSION

In this paper, we have investigated the problem of weak-small target fusion in infrared dual-band image. At the same time, the multi-scale transforms, sparse representation and deep learning fusion algorithms have been proposed based on infrared image and visible image. We have proposed an infrared weak-small targets image fusion method based on LatLRR and DWT. Firstly, the source images are decomposed into high frequency and low frequency by DWT. High frequency displays saliency parts and low frequency displays contour parts. At the same time, the LatLRR is utilized to learn a project matrix which is used to extract salient features. Secondly, high frequency parts are fused by maximum absolute value and low frequency parts are fused by weighted-average. The training matrix and high frequency fusion parts are used for contrast modulation fusion. Finally, the fused image is reconstructed by combining the outline parts and detail parts. We use both subjective and objective methods to evaluate the proposed method. The experimental results show that the proposed method exhibits better performance than other compared methods. Especially, This algorithm can improve the detection distance of dual-band detectors, and provide a technical support for accurate target recognition and detection.

ACKNOWLEDGMENT

REFERENCES

[1] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COMM-31, no. 4, pp. 532–540, Apr. 1983.

[2] A. B. Hamza, Y. He, H. Krim, and A. Willsky, "A multiscale approach to pixel-level image fusion," *Integr. Comput.-Aided Eng.*, vol. 12, no. 2, pp. 135–146, 2005.

[3] Q. Jiang, X. Jin, S.-J. Lee, and S. Yao, "A novel multi-focus image fusion method based on stationary wavelet transform and local features of fuzzy sets," *IEEE Access*, vol. 5, pp. 20286–20302, 2017.

[4] X. Luo, X. Li, P. Wang, S. Qi, J. Guan, and Z. Zhang, "Infrared and visible image fusion based on NSCT and stacked sparse autoencoders," *Multimed. Tools Appl.*, vol. 77, no. 17, pp. 22407–22431, Sep. 2018.

[5] S. Wang and M. Zhen, "A new visible and infrared image fusion algorithm based on NSCT," in *Proc. Int. Conf. Sensor Netw. Signal Process. (SNSP)*, Oct. 2018, pp. 181–184.

[6] L. Wang, B. Li, and L.-F. Tian, "EGGDD: An explicit dependency model for multi-modal medical image fusion in shift-invariant shearlet transform domain," *Inf. Fusion*, vol. 19, pp. 29–37, Sep. 2014.

[7] Y. Huang, K. Gao, C. Gong, L. Han, and Y. Guo, "Infrared and visible image fusion with the target marked based on multi-resolution visual attention mechanisms," *Proc. SPIE*, vol. 10255, Mar. 2016, Art. no. 102552V.

[8] P. Zhu, X. Ma, and Z. Huang, "Fusion of infrared-visible images using improved multi-scale top-hat transform and suitable fusion rules," *Infr. Phys. Technol.*, vol. 81, pp. 282–295, Mar. 2017.

[9] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Inf. Fusion*, vol. 40, pp. 57–75, Mar. 2018.

[10] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.

[11] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, May 2015.

[12] Y. Haitao, "Sparse representation with learned multiscale dictionary for image fusion," *Neurocomputing*, vol. 148, pp. 600–610, Jan. 2015.

[13] R. Bashir, R. Junejo, N. N. Qadri, M. Fleury, and M. Y. Qadri, "SWT and PCA image fusion methods for multi-modal imagery," *Multimed. Tools Appl.*, vol. 78, no. 2, pp. 1235–1263, Jan. 2019.

[14] Y. Yang, M. Ding, S. Huang, Y. Que, W. Wan, M. Yang, and J. Sun, "Multi-focus image fusion via clustering PCA based joint dictionary learning," *IEEE Access*, vol. 5, pp. 16985–16997, 2017.

[15] Y. Yao, P. Guo, X. Xin, and Z. Jiang, "Image fusion by hierarchical joint sparse representation," *Cognit. Comput.*, vol. 6, no. 3, pp. 281–292, Sep. 2014.

[16] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.

[17] H. Li and X.-J. Wu, "Multi-focus image fusion using dictionary learning and low-rank representation," in *Proc. ICIG*, Dec. 2017, pp. 675–686.

[18] H. Li and X. J. Wu, "Infrared and visible image fusion using latent low-rank representation," *J. Vis. Commun. Image*, vol. 3, pp. 1–20, Apr. 2018.

[19] H. Li, X. J. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," 2018, *arXiv:1806.07119*. [Online]. Available: https://arxiv.org/abs/1806.07119

[20] X. Y. Han, T. Lv, X. Song, T. Nie, H. Liang, B. He, and A. Kuijper, "An adaptive two-scale image fusion of visible and infrared images," *IEEE Access*, vol. 7, pp. 56341–56352, 2019.

[21] H. Li, X.-J. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," in *Proc. IEEE CVPR*, Dec. 2018, pp. 1–21.

[22] H. Li and X. J. Wu, "Infrared and visible image fusion using a novel deep decomposition method," 2018, *arXiv:1811.02291*. [Online]. Available: https://arxiv.org/abs/1811.02291

[23] Y. Liu, X. Chen, H. Peng, and Z. F. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[24] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2018.

[25] J. Y. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2018.

[26] X. Qu, F. Zhang, Y. Zhang, and L. Tang, "A method of dual-band infrared images fusion based on Gradeint pyramid decomposition," in *Proc. IET Int. Conf. Inf. Sci. Control Eng.*, Dec. 2012, pp. 1–4.

[27] Y.-Q. Sun, J.-W. Tian, and J. Liu, "Dim small targets detection based on dualband infrared image fusion," in *Proc. IEEE ICIT.*, 2006, pp. 3003–3007.

[28] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. IEEE ICCV*, Nov. 2011, pp. 1615–1621.

[29] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. ICML*, 2010, pp. 663–670.

[30] Alexander Toet. (2014). *TNO Image Fusion Dataset*. [Online]. Available: https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029

[31] R. Singh, M. Vatsa, and A. Noore, "Multimodal medical image fusion using redundant discrete wavelet transform," in *Proc. ICAPR*, Feb. 2009, pp. 4–6.

[32] G. George, R. M. Oommen, S. Shelly, S. S. Philipose, and A. M. Varghese, "A survey on various median filtering techniques for removal of impulse noise from digital image," *IEEE Access*, Mar. 2018, pp. 23–238.

[33] Z. Z. Wang, J. R. Deller, Jr., and B. D. Fleet, "Pixel-level multisensor image fusion based on matrix completion and robust principal component analysis," *Proc. SPIE*, vol. 25, no. 1, Jan. 2016, Art. no. 013007.

[34] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infr. Phys. Technol.*, vol. 82, pp. 8–17, May 2017.

[35] J. W. Roberts, J. van Aardt, and F. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, Jan. 2008, Art. no. 023522.

[36] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[37] P. Sreeja and S. Hariharan, "An improved feature based image fusion technique for enhancement of liver lesions," *Biocybern. Biomed. Eng.*, vol. 38, no. 3, pp. 611–623, 2018.

[38] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

XIAOZHU WANG received the B.S. degree from the Changchun University of Science and Technology, in 2013, and the M.S. degree from the School of Electronics and Information, Northwestern Polytechnical University, in 2016, where he is currently pursuing the Ph.D. degree with the School of Astronautics. He has presented papers in national and international journals and conferences. His research interests include infrared target characteristics and multi-band fusion.
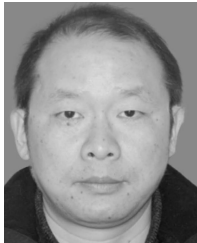
JIANFEI YIN received the B.S. degree from the School of Astronautics, Northwestern Polytechnical University, in 2001. He is currently an Associate Professor with the Research and Development Center of Infrared Detection Technology CASC, Shanghai Institute of Spaceflight Control Technology. His research interests include image processing and infrared multi-band seeker design.

KAI ZHANG received the Ph.D. degree from the School of Astronautics, Northwestern Polytechnical University, in 2009, where he is currently an Associate Professor. His research interests include guidance, navigation, and control.
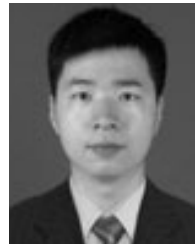
SHAOYI LI received the bachelor's degree from Southwest Jiaotong University, in 2008, and the master's and Ph.D. degrees from Northwestern Polytechnical University, in 2011 and 2015, respectively, where he is currently an Assistant Researcher. His current research interests include artificial intelligence and image processing.

JIE YAN was born in 1960. He received the Ph.D. degree from the School of Astronautics, Northwestern Polytechnical University, in 1988, where he is currently a Professor and a Ph.D. Candidate Supervisor. His research interests include flight control, guidance, system simulation, and aircraft design.

• • •