An Efficient Algorithm to Highlight Details in Infrared and Visible Image Fusion

PEIJIN LIU[ID], LICAI ZHANG[ID], MINGYANG LI[ID], AND XIANGRUI ZHANG[ID]

School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

Corresponding author: Licai Zhang (zhanglicai@xauat.edu.cn)

ABSTRACT To improve the fusion quality of infrared and visible images and highlight target and scene details, in this paper, a novel infrared and visible image fusion algorithm is proposed. First, a method for combining dynamic range compression and contrast restoration based on a guided filter is adopted to enhance the contrast of visible source images. Second, guided filter-based image multiscale decomposition is used to decompose images into base layers and detail layers. For base layer fusion, a fusion strategy based on the detail and energy measurements of the source image is proposed to determine the pixel value of the fused image base layer such that the energy loss of the fusion can be reduced and the texture detail features are highlighted to obtain more source image details. Finally, recursive separation and weighted histogram equalization methods are applied to optimize the fused image. Experimental results show that the fusion algorithm and fusion strategy proposed in this paper can effectively improve fusion image clarity, while more detailed target and scene information can still be retained.

INDEX TERMS Image fusion, highlighted details, multiscale decomposition, guided filter, fusion strategy.

I. INTRODUCTION

The fusion of infrared and visible images has been widely used in military, medical, remote sensing and many other fields in recent years [1]. Processing the source image from different sensors on the same target or scene according to a certain strategy, removing redundant information, and performing multidirectional and multiangle fusion can obtain a more accurate fusion image that reflects the target scene information [2]. With the development of image fusion technology, the cost of hardware equipment is reduced, and more comprehensive and reliable detailed information can be obtained [3]. Therefore, infrared and visible light image fusion technology has been extensively studied by scholars [4].

Although the imaging mechanisms are quite different in infrared sensors and visible sensors, they complement each other to some extent [5]. Infrared sensors can perceive heat radiation of different wavelengths, which can capture hidden target contour information, and they have great night vision and fog penetration capabilities, but they cannot obtain detailed information and have poor resolution [6]. Visible light sensors characterize objects through spectral reflection,

and the results have high resolution and rich background information, which are suitable for human visual perception, but the image quality is easily affected by the environment, especially at night and under low visibility conditions [3]. If the images from these two sensors are fused, i.e., the main target information is obtained from the infrared image, while the main detailed background information is obtained from the visible image, then the fused image can provide better target features and more detailed scene information [3]. Therefore, infrared and visible image fusion technology is a research hotspot, and it plays a key role in target tracking and detection, face recognition and other fields [7], [8].

According to the differences in image representation, image fusion technology can be divided into three levels: pixel-level fusion, feature-level fusion, and decision-level fusion [9]. Among them, the pixel-level image fusion method is currently the most common fusion method due to its relative source image information completeness [10]. Among the pixel-level image fusion methods, multiscale decomposition (MSD) has played a very important role and has proven to be extremely useful for image fusion [11].

For those methods based on MSD, generally, the visible source images are directly fused with the infrared images without being preprocessed. Therefore, under weak light or extreme weather conditions, the fusion effect of the visible

---

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

source images and the infrared images cannot be guaranteed, as the detailed scene information is severely lost. In this regard, we propose that contrast enhancement in visible images should also be an important part of image fusion. At present, the enhancement algorithms for the visible images include histogram equalization, automatic color enhancement (ACE), and multiscale retinex (MSR) [12]–[14], but these algorithms may produce artifacts or tend to be over-enhanced. Liu *et al.* [15] presented an algorithm based on an adaptive threshold to improve contrast, but it cannot fully display detailed information. Zhou *et al.* [16] presented a guided filter-based enhancement method, which can achieve some effects in dark areas of visible images, but over-enhancement occurs in bright areas. This means that visible images cannot be adaptively enhanced.

In image fusion schemes, fusion strategies play an important role. Averaging methods are adopted to fuse the base layer in most of the methods based on MSD [17]. However, because most of the energy is contained in the base layer, it is likely to cause energy loss during the averaging process, which eventually results in lower contrast [18]. In recent years, due to the research of many researchers, many infrared and visible image fusion methods have been proposed. Wei *et al.* [19] proposed a method based on robust principal component analysis (RPCA) and a nonsubsampled shear wave transform (NSST). They decomposed the source images by RPCA to obtain a sparse matrix, and then the source images were decomposed by the NSST transform to obtain the corresponding high-frequency and low-frequency components. The high-frequency part obtained was combined with a large absolute value and a sparse matrix, the low-frequency part was guided by a sparse matrix to the low-frequency part, and the fusion image was obtained through the NSST inverse transformation. However, this method easily loses visible detail information. Deng *et al.* [20] proposed a method based on a nonsubsampled shear wave transform. In this method, NSST is used for source image decomposition. For the fusion of high-frequency sub-band images, a fusion rule based on improved regional contrast is adopted by combining human visual characteristics. A fusion rule based on a saliency map is adopted for low-frequency sub-bands. After NSST inverse transformation, the fused image is obtained. However, this method may alter part of the image structure. Huang *et al.* [21] proposed an infrared and visible image information fusion method based on phase consistency and image entropy. In this method, the nonsubsampled contourlet transform (NSCT) is used for source image decomposition. The pulse coupled neural network (PCNN) fusion rule is used for the top layer of the high-frequency sub-band, and the absolute maximum rule is used for the remaining decomposition layers of the high-frequency sub-band. For the low-frequency sub-bands, the low-frequency fusion rules are designed according to activity measures such as phase consistency (PC), sharpness change (LSCM) and local signal strength (LSS). Li *et al.* [22] proposed an image fusion framework based on multilevel image decomposition of latent low-rank representation (MDLatLRR). This method uses MDLatLRR to decompose the source image into detailed and base parts. For the detailed parts, a fusion strategy based on the kernel norm and a reshape operator is used. The average fusion rule is adopted for the base parts. However, the fusion performance of this method is excessively dependent on the number of decomposition layers. When the number of decomposition layers increases, the algorithm complexity increases, and the time consumption is large. Nie *et al.* [23] proposed a fusion method based on total variation (TV). In this method, the weighted fidelity term is used to fuse the targets in the infrared image and the salient scenes in the visible image, and a weight estimation method based on the saliency of global brightness contrast was proposed. Additionally, $l_{2,1,rc}$ is introduced to provide structure group sparsity for fidelity items, $l_{1/2}$ provides better gradient sparsity for detail preservation items, and l_2 is used for brightness degradation prevention items. However, the ability of this method to retain detailed information is limited.

The performance of current fusion methods is continuously improving, but there are still some problems, such as loss of visible image detail information and contrast degradation. To address the above problems, a method for highlighting details is proposed in this paper. The main contributions of the proposed method are as follows:

1) By taking advantage of the combination of dynamic range compression and contrast restoration based on a guided filter, the contrast in the visible images can be increased adaptively to improve the quality of the fused input source images.

2) To fuse the detail layers, a multiscale gradient function is used to generate a saliency map, from which a weight map is derived after normalization such that the complexity of the algorithm can be greatly reduced, thereby reducing time consumption.

3) For base layer fusion, a fusion strategy is presented to determine the pixel value of the base layers of the fused image by calculating the detail and energy measures of the source image, which aims to prevent energy loss, highlight texture details, and obtain more source image details so that the fusion image information is richer and consistent with the human visual effect.

4) Recursive separation and weighted histogram equalization are applied to the fused image to optimize the image contrast as much as possible, maintain a higher level of brightness, and avoid artifacts and unnatural enhancement caused by excessive equalization, thereby improving the quality of the fused image and getting closer to the human visual effect.

The rest of this article is organized as follows. The second section introduces in detail the acquisition and fusion rules of the detail and base layers, image reconstruction and other related theories. In the third section, the experimental simulation is carried out, and the results of infrared and visible light image fusion are analyzed subjectively and objectively. The fourth section is the conclusion of this article.
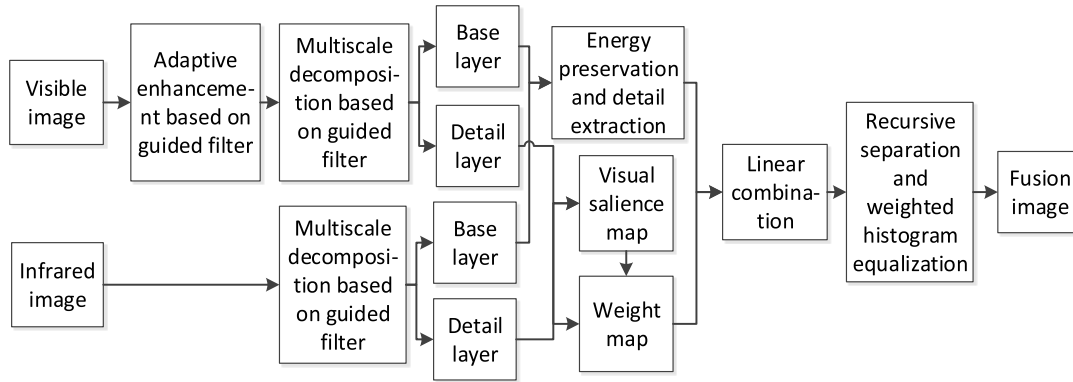
FIGURE 1. Overall framework of the fusion method.

II. FUSION METHOD

The overall process of the fusion method proposed in this paper is illustrated in Fig. 1. First, adaptive enhancement based on a guided filter is used before fusion. Second, using the guided filter-based image MSD method, the infrared and enhanced visible images are divided into base layers and detail layers, which can extract texture details while maintaining good edge features. Then, to fuse the detail layer, a multiscale gradient function is used to generate a saliency map, from which a weight map is derived after normalization such that the complexity of the algorithm can be greatly reduced. For base layer fusion, a fusion strategy is presented to determine the pixel value of the base layer of the fused image by calculating the measures of details and energy of the source image, which aims to prevent energy loss, highlight texture details, and obtain more source image details. Finally, recursive separation and weighted histogram equalization are applied to the fused image for further optimization.

A. GUIDED FILTER

The guided filter (GF) is a type of good edge-preserving filtering method [24] that utilizes a box filter implemented with integral image technology to achieve rapid implementation instead of directly performing convolution. When the input image p and the guided image I are identical, the guided filter can effectively maintain the edge features, which is equivalent to the bilateral filter at this time. When the ε value is large enough, the guided filter can replace the Gaussian filter. Currently, GF is widely used in saliency detection [25] and image fusion [26].

Assuming that GF is in the local window ω_k centered at pixel k, the guided image I can be linearly transformed by Eq. (1) to obtain the filtered output image q.

$$q_i = a_k I_i + b_k, \quad \forall i \in \omega_k \tag{1}$$

where ω_k is a local window with radius r, and a_k, b_k are constant coefficients over ω_k. By minimizing Eq. (2), q can be approximated the most to the input image p in ω_k.

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left((a_k I_i + b_k - p_i)^2 + \varepsilon a_k^2 \right) \tag{2}$$

Here, ε is a regularization parameter to prevent a_k from being too large, and it can be deduced that ε is less than 1, and its value range is between 0 and 1. In addition, the smaller ε is, the smaller the overlay smoothing factor, the larger a_k, and the clearer the image edge that is maintained; the larger ε is, the larger the overlay smoothing factor, the smaller a_k, the lower the edge retention of the image, and the image area is smoothed. Therefore, a_k and ε are used to jointly determine the edge retention and smoothness of the output image in the guided filter. Then, a_k, b_k can be solved as

$$a_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i p_i - \bar{I}_k \bar{p}_k}{\sigma_k^2 + \varepsilon}$$
$$b_k = \bar{p}_k - a_k \bar{I}_k \tag{3}$$

where $|\omega|$ is the number of pixels in ω_k, $\bar{I}_k$ and σ_k^2 are the average value and variance in the guide image I in ω_k, respectively, and $\bar{p}_k$ is the average value of the input image p in ω_k.

Regarding the pixels i in different local windows, the filtered output image q_i will change with respect to the window. This problem can be solved by calculating the average of all possible values through

$$q_i = \bar{a}_i I_i + \bar{b}_i \tag{4}$$

where $\bar{a}_i = \frac{1}{|\omega|} \sum_{k \in \omega_k} a_k$, $\bar{b}_i = \frac{1}{|\omega|} \sum_{k \in \omega_k} b_k$ are the average coefficients.

B. AN ADAPTIVE ENHANCEMENT ALGORITHM FOR VISIBLE IMAGE

Tao *et al.* [27] proposed an enhancement method composed of dynamic range compression and contrast enhancement to improve the visibility in visible images. However, Liu *et al.* [28] found that although this method can achieve some effects in the dark areas of visible images, over-enhancement occurs in the bright areas, which means

that the visible images cannot be adaptively enhanced. Therefore, a hybrid method to adaptively improve the contrast of visible images is used here.

First, the input visible image I is normalized to the range [0, 255] by

$$I_1 = \frac{I - \min(I)}{[\max(I) - \min(I)]} \quad (5)$$

I_1 is processed by using a guided filter. Then, we obtain filtered image I_b by $I_b = GF_{r,\varepsilon}(I_1)$, where r, ε are the GF size and edge-preserving degree. According to Durand *et al.* [29], it is transformed into the logarithmic domain to acquire the basic layer by

$$\hat{I}_b = \log(I_b + \xi) \quad (6)$$

and the detail layer by

$$\hat{I}_d = \log(I_1 + \xi) - \hat{I}_b \quad (7)$$

respectively, where log represents the natural logarithm algorithm. Here, let $\xi = 1$ to prevent the result from being negative and the unreasonable situation of $\log 0$.

Since the dynamic range of the base layer is relatively high, we set the scale factor β and the restoration factor γ and use them to perform dynamic range compression and contrast restoration on $\hat{I}_b$. They can be expressed as

$$\hat{u} = \beta\hat{I}_b + \hat{I}_d + \gamma \quad (8)$$

Generally, we can define a target contrast T to which the contrast of the base layer is reduced; then, the value of β can be expressed as

$$\beta = \frac{\log(T)}{\max\left(\hat{I}_b\right) - \min\left(\hat{I}_b\right)} \quad (9)$$

where $\max\left(\hat{I}_b\right)$ and $\min\left(\hat{I}_b\right)$ represent the maximum and minimum intensities of $\hat{I}_b$. The overall contrast is reduced because the dynamic range is compressed, and the contrast needs to be restored. γ is calculated as

$$\gamma = (1 - \beta)\max\left(\hat{I}_b\right) \quad (10)$$

Finally, the enhanced image can be obtained by

$$u = \exp(\hat{u}) \quad (11)$$

The size of the guided filter is $r = \lfloor \varepsilon T \max(\omega, h) \rfloor$, where ω, h are the width and height of the image, respectively. Since the target contrast T determines the effect of enhancing the image, different T values can be set through experiments. The experimental results are shown in Fig. 2. Under low illumination conditions, setting the parameters $T = 4$ and $\varepsilon = 0.01$ can effectively enhance the visible images. Additionally, due to the adaptive enhancement effect of this method, the visible images are prevented from being over-enhanced under better lighting conditions.
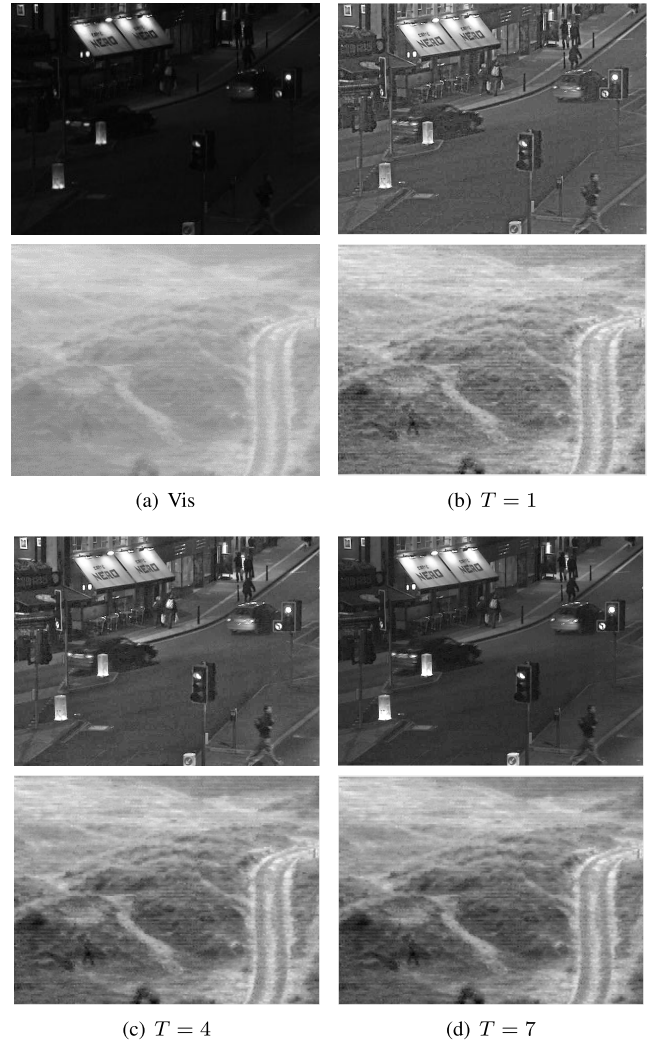

(a) Vis   (b) $T = 1$

(c) $T = 4$ (d) $T = 7$

FIGURE 2. Enhancement results at different *T* values.

C. MULTISCALE DECOMPOSITION BASED ON A GUIDED FILTER

Using GF, multiscale decomposition can be performed. The flowchart of the guided filter multiscale decomposition of image I is shown in Fig.3. According to the Li *et al.* [26], the guided filter operator is denoted as $GF_{r,\varepsilon}()$.

GF-based multiscale decomposition of I_1 and I_2 is performed to obtain base layers B_1^1, B_2^1 and detail layers D_1^1, D_2^1 as follows:

$$B_1^1 = GF\left(I_1, I_2, r^1, \varepsilon^1\right), \ B_2^1 = GF\left(I_2, I_1, r^1, \varepsilon^1\right) \quad (12)$$
$$D_1^1 = I_1 - B_1^1, D_2^1 = I_2 - B_2^1 \quad (13)$$

The continuous base layers B_1^k, B_2^k are expressed as:

$$B_1^k = GF\left(B_1^{k-1}, B_2^{k-1}, r^k, \varepsilon^k\right),$$
$$B_2^k = GF\left(B_2^{k-1}, B_1^{k-1}, r^k, \varepsilon^k\right),$$
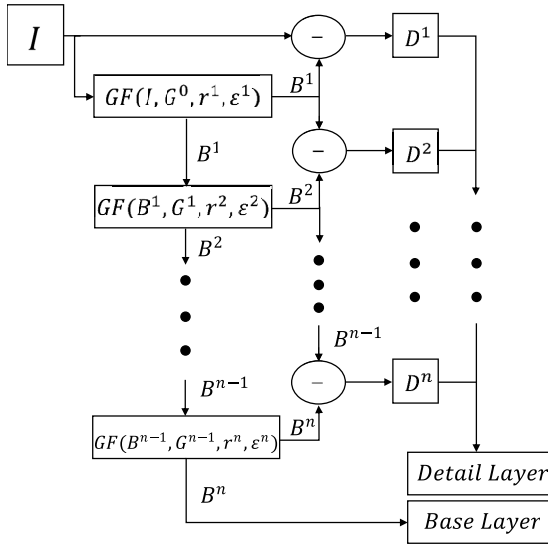$$k = 1, \ldots, n \quad (14)$$

FIGURE 3. GF-based multiscale decomposition of image I.

where B_1^{k-1}, B_2^{k-1} are base layers of I_1 and I_2 at the $k-1$ level and depend on the preceding level B_1^{k-2}, B_2^{k-2}, respectively. In addition, $r^k = 2 * r^{k-1}$, $\varepsilon^k = \varepsilon^{k-1}/4$. It can be noted that B_1^0, B_2^0 represent two source images I_1 and I_2.

With the same Eq. (13), the consecutive detail layers D_1^k, D_2^k are generated as:

$$D_1^k = B_1^{k-1} - B_1^k, \quad D_2^k = B_2^{k-1} - B_2^k \tag{15}$$

D. DETAIL LAYER FUSION

1) VISUAL SALIENCY DETECTION (VSD)

A multiscale VSD method is used for image fusion. The brief steps are described as follows. First, use Eq. (12) and Eq. (13) to obtain base layers B_1^k, B_2^k and detail layers D_1^k, D_2^k. Second, visual maps S_1^k, S_2^k are obtained by taking the absolute value of the magnitude of detail layers D_1^k, D_2^k. Third, S_1^k, S_2^k are averaged to find the visual maps S_1, S_2 corresponding to the source images. Finally, the final visual saliency S of the entire scene can be determined by taking the maximum value of S_1, S_2. It can be summarized as

$$S_1^k = \left| D_1^k \right|, \quad S_2^k = \left| D_2^k \right|$$

$$S_1 = \frac{1}{k} \sum_{k=1}^{n} S_1^k, \quad S_2 = \frac{1}{k} \sum_{k=1}^{n} S_2^k$$

$$S = \max(S_1, S_2) \tag{16}$$

2) WEIGHT MAP CALCULATION

Unlike existing fusion methods based on guided filters, a novel weight map structure operation that obtains saliency maps directly from the detail layers is used here, which greatly reduces the intricacy of the algorithm. Weight maps W_1^k, W_2^k can be obtained by normalizing S_1^k, S_2^k, which can

be expressed as

$$W_1^k = \frac{S_1^k}{\sum_{i=1}^{2} S_i^k}, \quad W_2^k = \frac{S_2^k}{\sum_{i=1}^{2} S_i^k}, \quad \forall k = 1, \ldots, n \tag{17}$$

Detailed layer information of each pixel can be integrated by saliency information-based weight map structure operation.

3) FUSION STRATEGY

The detail layers are integrated at each scale k with the help of weight maps W_1^k and W_2^k as follows:

$$D_F^k = W_1^k D_1^k + W_2^k D_2^k \tag{18}$$

The final detail layer D_F can be obtained by combining each integrated detail layer:

$$D_F = \sum_{k=1}^{n} D_F^k \tag{19}$$

E. BASE LAYER FUSION

Regarding image fusion, the intensities of different source images may vary significantly even at the same location [18]. Thus, the averaging-based base layer fusion strategy can often cause the loss of energy in fused images [17] and distortion of the fusion image, which results in inferior visual perception. To obtain a better fusion effect, a method to prevent energy loss and highlight the features of texture details is proposed. First, we define the energy level of the base layer, named EB, where the energy comes from Parseval's theorem (also known as the Rayleigh energy theorem). Specifically, the energy of a frequency band is defined as the sum of the squares of all frequency band coefficients. The formula is as follows:

$$EB_S(i,j)$$
$$= \sum_{m=-L}^{L} \sum_{n=-L}^{L} W(m+L+1, n+L+1) B_S(i,j)^2 \tag{20}$$

where $S \in \{1, 2\}$, W is a $(L-2) \times (L-2)$ weighting matrix with radius L, where L is the layer number of guided filter decomposition. Each element's value in W is set to 2^{L-2d}, where d is the distance from its surrounding elements to the center. $B_S(i,j)$ is the gray value at the pixel point (i,j) of the base layers of I_1 and I_2 at the L level.

Second, due to computational efficiency, there are limitations in the GF-based decomposition levels. Accordingly, detailed information exists in the base layer. To obtain more details, a measurement named DB is defined as

$$DB_S(i,j)$$
$$= \sum_{m=-L}^{L} \sum_{n=-L}^{L} W(m+L+1, n+L+1) AB_S(i,j) \tag{21}$$

where $S \in \{1, 2\}$, W has the same definition as that in Eq. (20). Compared with the original modified Laplace (ML), AB considers the influence of the diagonal coefficient to

make full use of neighborhood information. In addition, the value of AB is as follows:

$$
\begin{aligned}
AB_S\,(i,j) = \; & |2B_S(i,j) - 2B_S\,(i-1,j)| \\
& + |2B_S\,(i,j) - 2B_S\,(i,j-1)| \\
& + \frac{2}{\sqrt{2}}\,|2B_S\,(i,j) - 2B_S\,(i-1,j-1)|
\end{aligned} \quad (22)
$$

With the help of the multiplication of *EB* and *DB*, the final base layer is determined by the following strategy:

$$
B_F\,(i,j) = \begin{cases} B_1\,(i,j), & if \; EB_1(i,j) \cdot DB_1(i,j) \\ & \gg EB_2(i,j) \cdot DB_2(i,j) \\ B_2\,(i,j), & otherwise \end{cases} \quad (23)
$$

F. FUSED IMAGE RECONSTRUCTION AND FURTHER OPTIMIZATION

Because the detail layers are obtained by subtracting the base layers from the source images, to reduce the computational complexity, the final fused image can be obtained by linearly combining the fusion results of the detail layers and the base layers. The final fused images are integrated by the linear superposition of the final detail layer D_F and the final base layer B_F.

$$
F = D_F + B_F \quad (24)
$$

To allow a higher level of brightness maintenance, avoid artifacts and unnatural enhancement due to over-equalization, and enhance the contrast in the image as much as possible, in this paper, recursive separation and weighted histogram equalization methods are used to further optimize the final fused images to achieve an effect as close as possible to human visual perception.

The main steps of the proposed method are summarized in Algorithm 1.

III. EXPERIMENTS AND RESULTS ANALYSIS

In the experiments, we use the dataset collected from the website http://www.imagefusion.org/, which has been widely used in image fusion testing [30]. We select seven pairs of image samples, including the low visibility visible images. Other classic or newly released fusion algorithms, including the fusion method based on NSST [31], the fusion method based on GF multiscale decomposition (MGFF) [32], the fusion method based on infrared image structure extraction and visible image information retention (IFEVIP) [33],the multiscale fusion method through Gaussian and bilateral filters (HMSD) [34], the fusion method based on the convolutional neural network method (CNN) [35], the fusion method based on target-enhanced multiscale transform decomposition (TE-MSD) [36], and the fusion method based on multiscale transformation and norm optimization (MST-NO) [37], are compared with the proposed method on some image samples. The experimental parameters of the proposed method in this paper are set

---

Algorithm 1 Proposed Infrared and Visible Image Fusion Algorithm

---

Input: the source images: I_1 and I_2.
Parameters: GF size r, edge-preserving degree ε and number multiscale decomposition levels k.
Part 1: Adaptive enhancement for the visible image
01: Obtain new visible image I_1 by using Eq. (5)-(11).
Part 2: Multiscale decomposition
02: For each image $I_S = [I_1, \; I_2]$
03: Perform GF-based multiscale decomposition on S to obtain D_S^k, B_S^k;
04: End
Part 3: Detail layer fusion
05: For each level $l = 1 : k$
06:    For each detail layer $D_S^k = [D_1^k, D_2^k]$
07:        Obtain visual saliency S_1^k, S_2^k by using Eq. (16);
08:        Obtain Weight maps W_1^k, W_2^k by using Eq. (17);
09:        Merge D_1^k and D_2^k by using Eq. (18) to obtain D_F^k;
10:    End
11: End
12: Obtain final detail layer D_F by using Eq. (19);
Part 4: Base layer fusion
13: For base layer $B_S = [B_1, B_2]$
14:    Calculate the EB_S of B_S by using Eq. (20);
15:    Calculate the DB_S of B_S by using Eq. (20);
16: End
17: Merge B_1 and B_2 by using Eq. (23) to obtain final detail layer B_F;
Part 5: Fused image reconstruction and further optimization
18: Obtain fused image F by using Eq. (24);
19: Obtain final fused image F by using recursive separation and weighted histogram equalization methods;
Output: the fused image F.

---

as follows. Set the magnitude of GF in the enhancement before fusion as $r = \lfloor 0.04 \max(\omega, h) \rfloor$, where ω, h are the image width and height respectively. The performance of the proposed method depends on free parameters: GF size r, edge-preserving degree ε and number multiscale decomposition levels k. Through a large number of experiments, as in the early stage, these parameters are tuned to $r = [2, 4, 8, 16]$, $\varepsilon = [10^{-2}, 2.5 \times 10^{-3}, 6.25 \times 10^{-4}, 1.5625 \times 10^{-4}]$, and $k = 4$. The remaining parameters of the comparison methods are consistent with the literature [31]–[37]. The fusion results of eight methods for seven pairs of image samples are shown in Figs. 4–10. To better observe, two local windows in each image are labeled and enlarged.

Overall, the seven fused images obtained by the proposed method have better visual quality and are more suitable for human visual perception than the fused images obtained by the other seven methods. Specifically, NSST-based fusion images are not capable of highlighting details, and the background is too blurry. In MGFF-based fusion images,
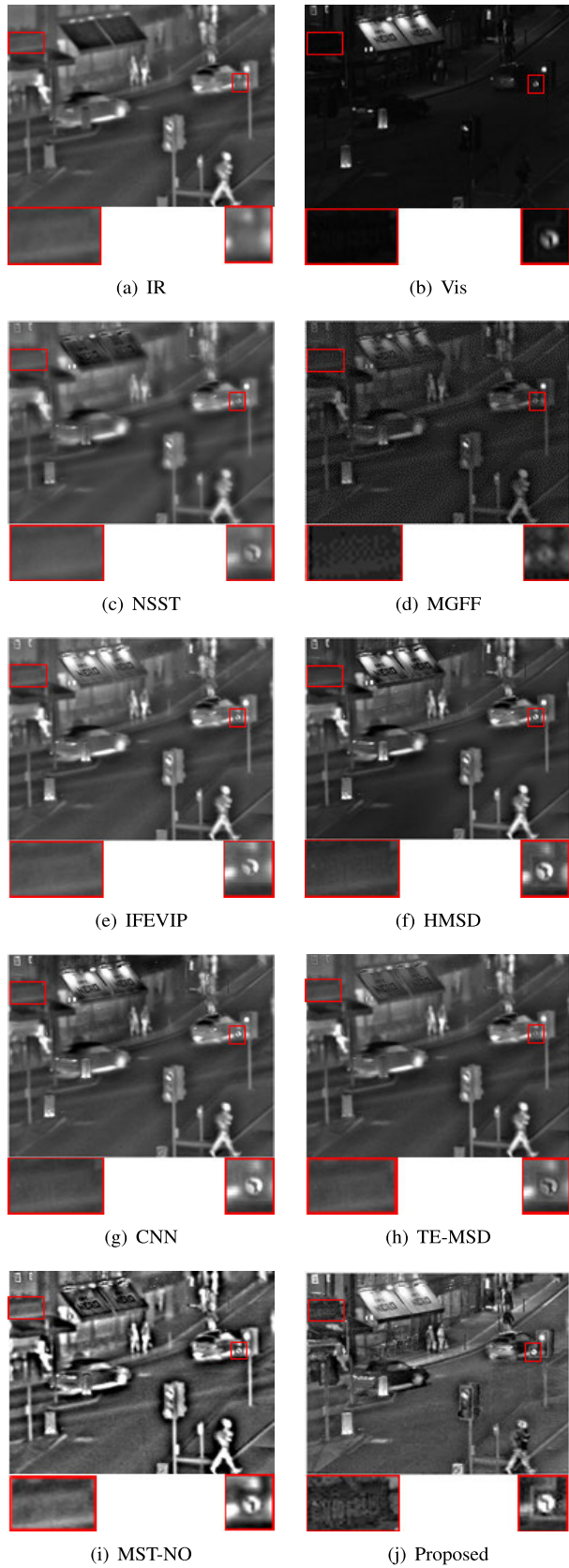
(a) IR

(b) Vis

(c) NSST

(d) MGFF

(e) IFEVIP

(f) HMSD

(g) CNN

(h) TE-MSD

(i) MST-NO

(j) Proposed

FIGURE 4. The results of the first group of comparative experiments.

(a) IR

(b) Vis

(c) NSST

(d) MGFF

(e) IFEVIP

(f) HMSD

(g) CNN

(h) TE-MSD

(i) MST-NO

(j) Proposed

FIGURE 5. The results of the second group of comparative experiments.

(a) IR (b) Vis

(c) NSST (d) MGFF

(e) IFEVIP (f) HMSD

(g) CNN (h) TE-MSD

(i) MST-NO (j) Proposed

FIGURE 6. The results of the third group of comparative experiments.

(a) IR (b) Vis

(c) NSST (d) MGFF

(e) IFEVIP (f) HMSD
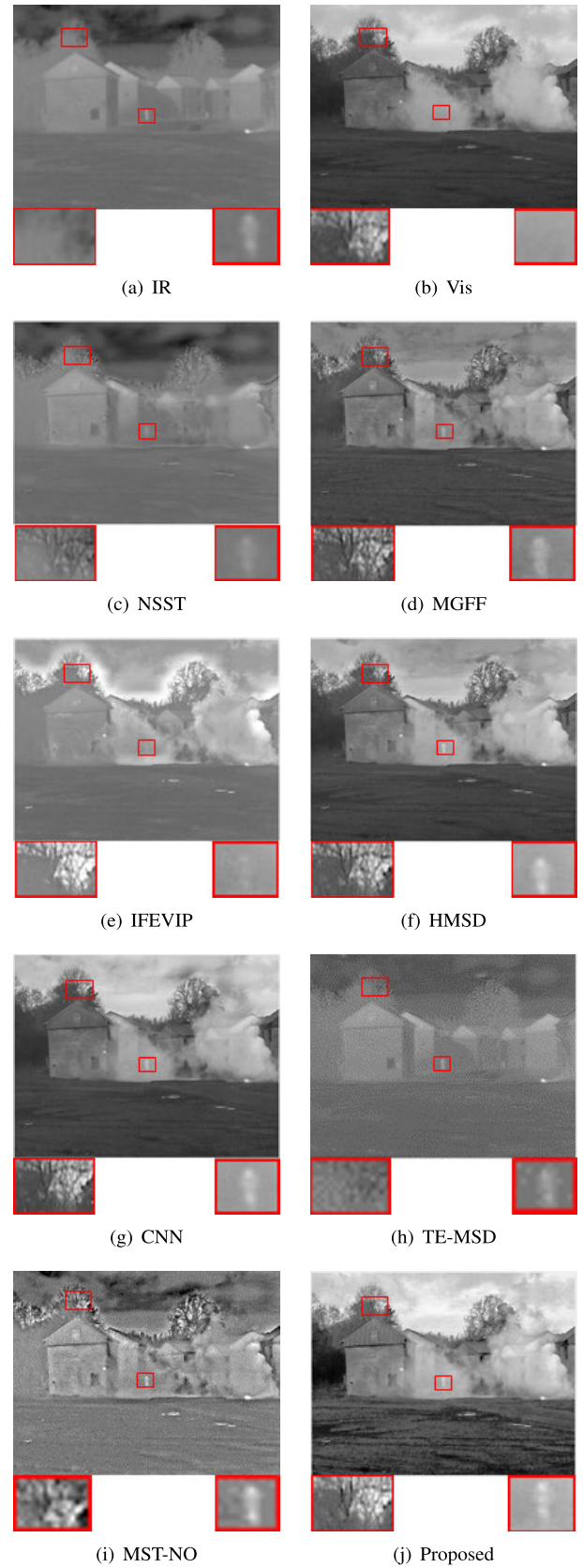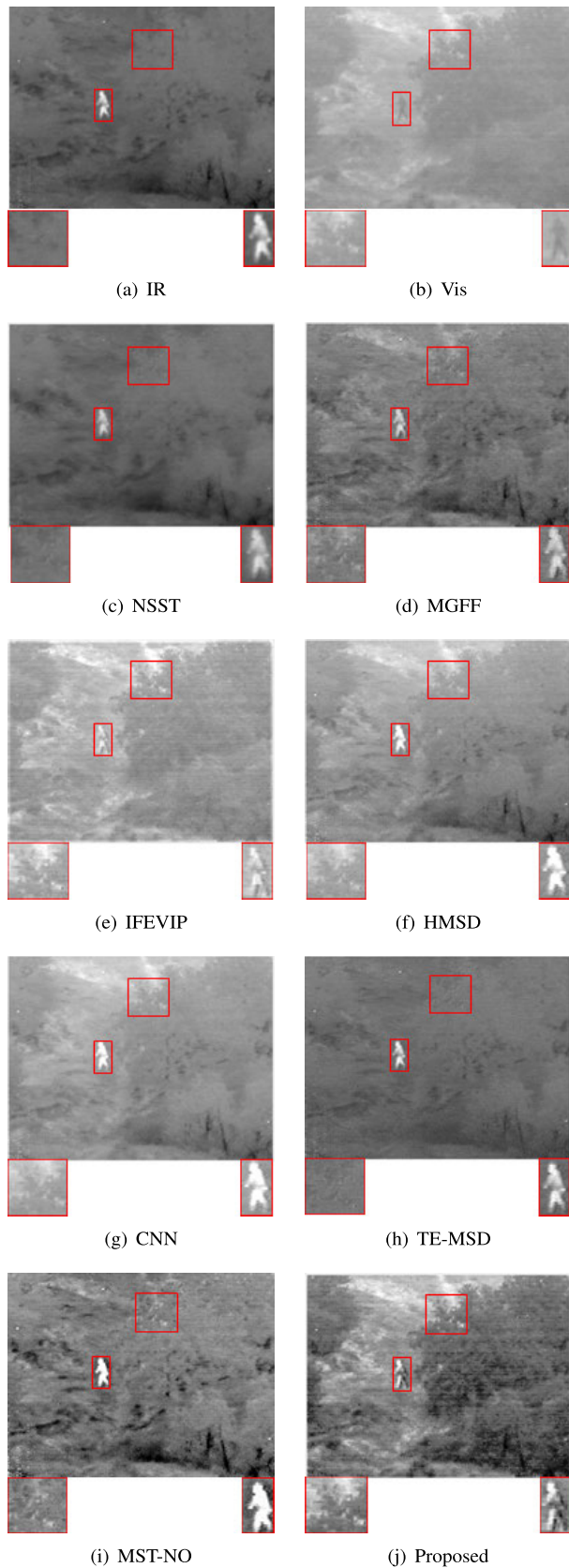
(g) CNN (h) TE-MSD

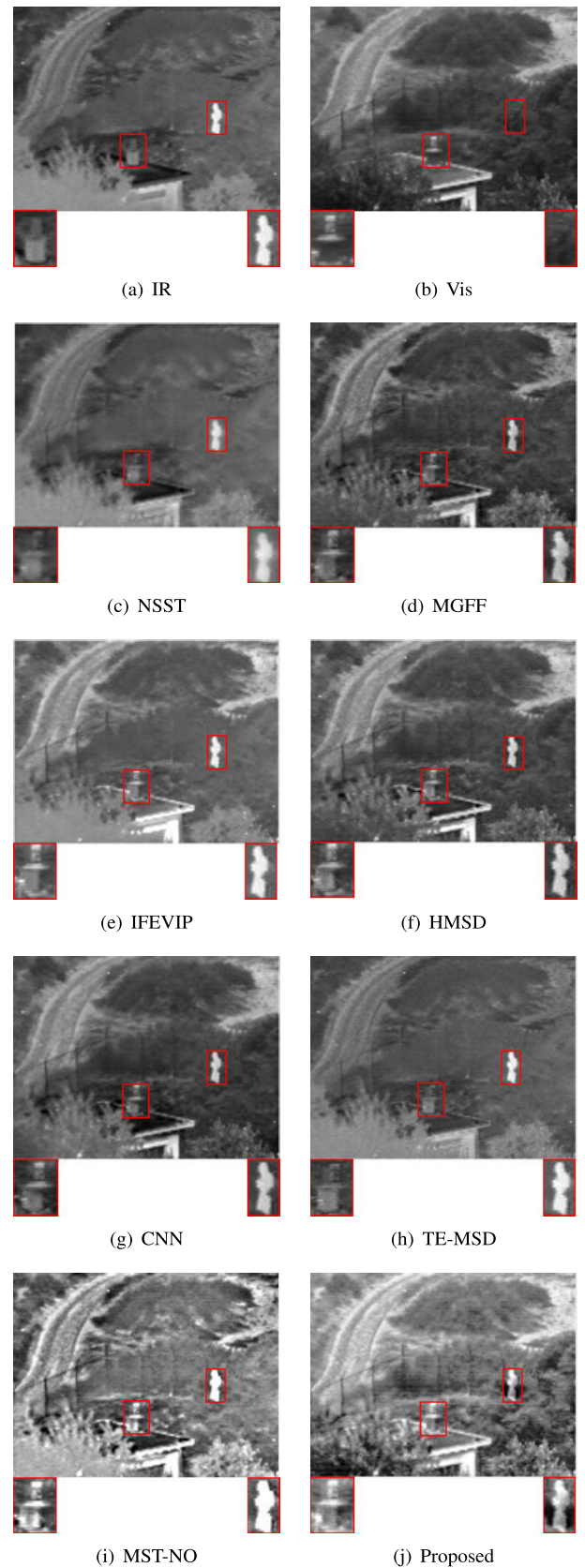(i) MST-NO (j) Proposed

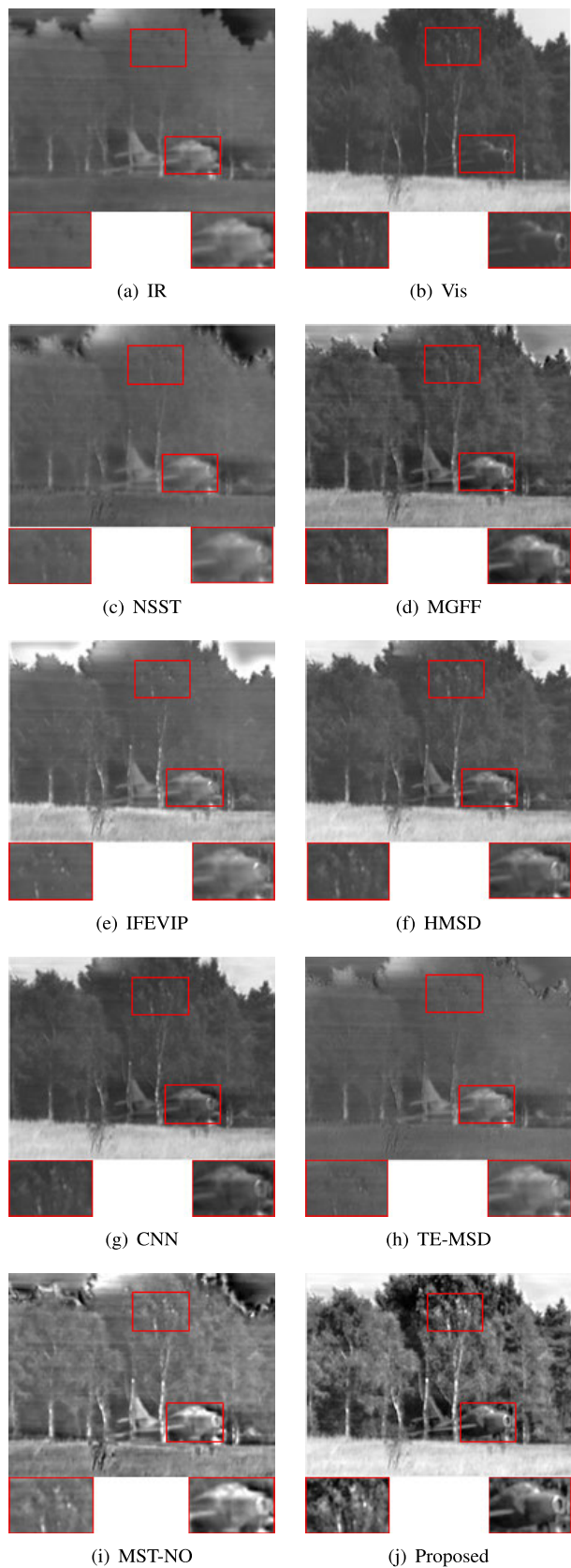**FIGURE 7.** The results of the fourth group of comparative experiments.

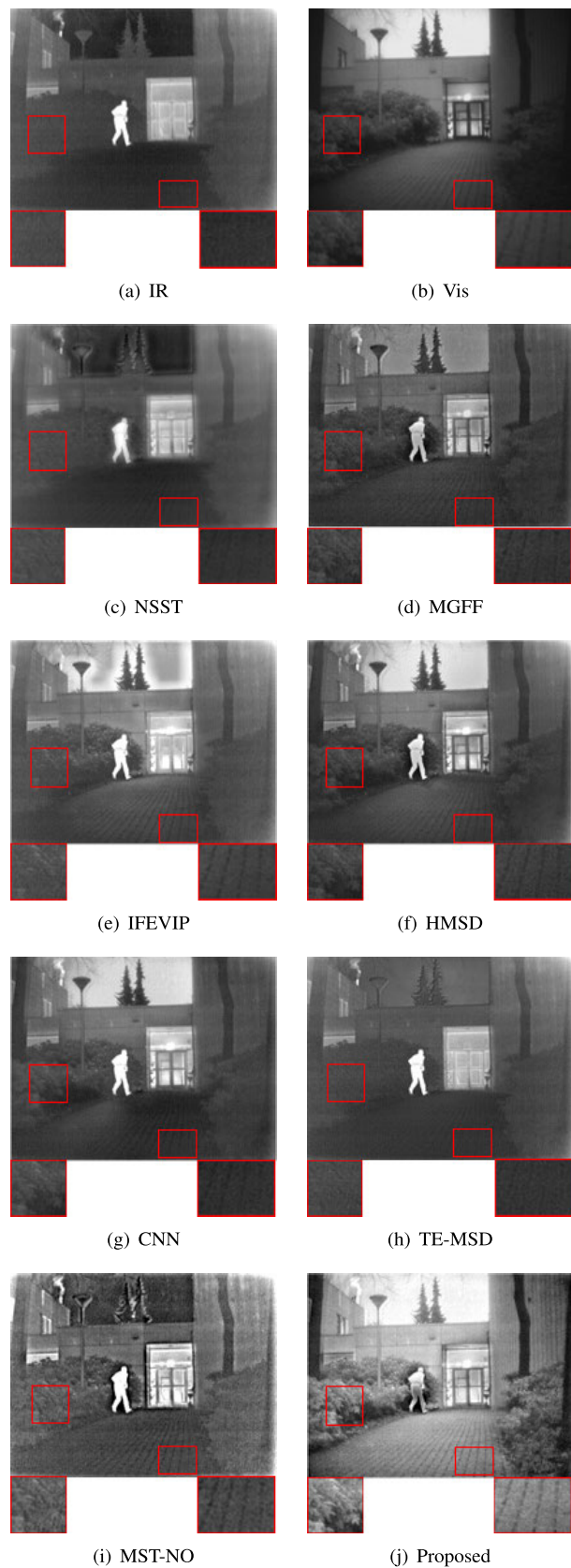**FIGURE 8.** The results of the fifth group of comparative experiments.

FIGURE 9. The results of the sixth group of comparative experiments.
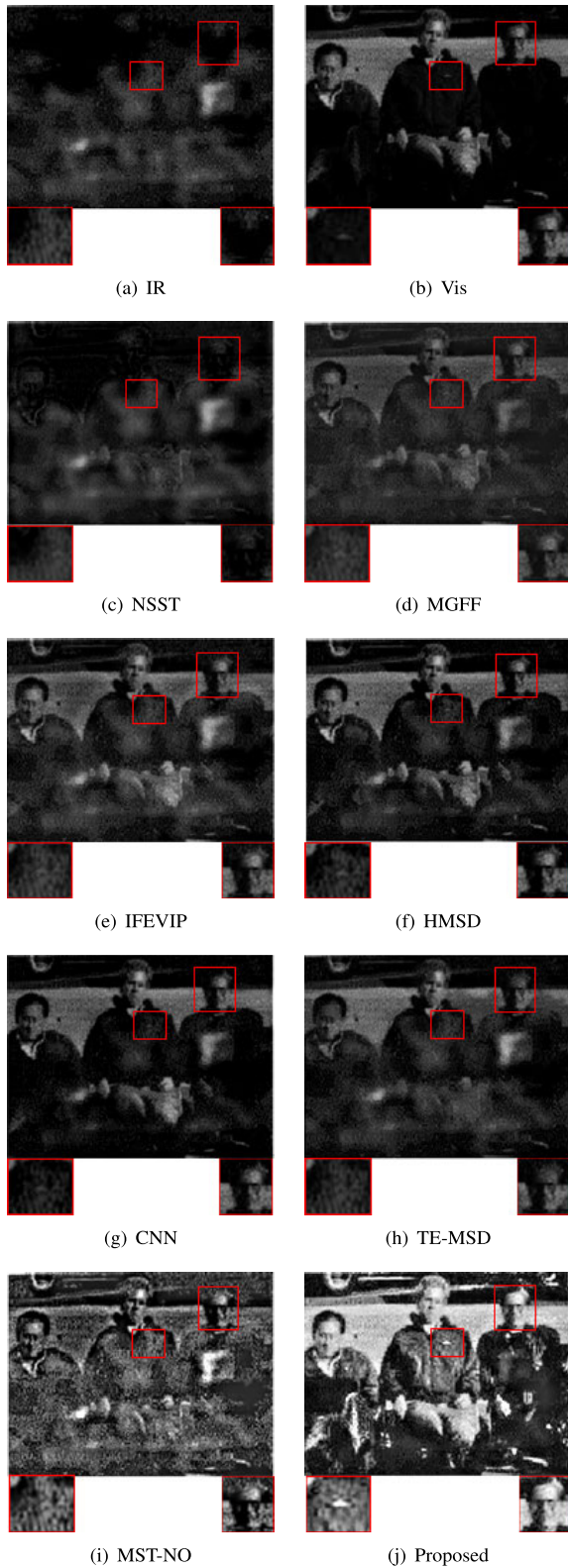
FIGURE 10. The results of the seventh group of comparative experiments.

FIGURE 11. Quantitative evaluation of different methods for seven image pairs.

the brightness of the infrared salient target is low, and insufficient detailed information is retained. There are clear halos in most IFEVIP-based fusion images, which lead to
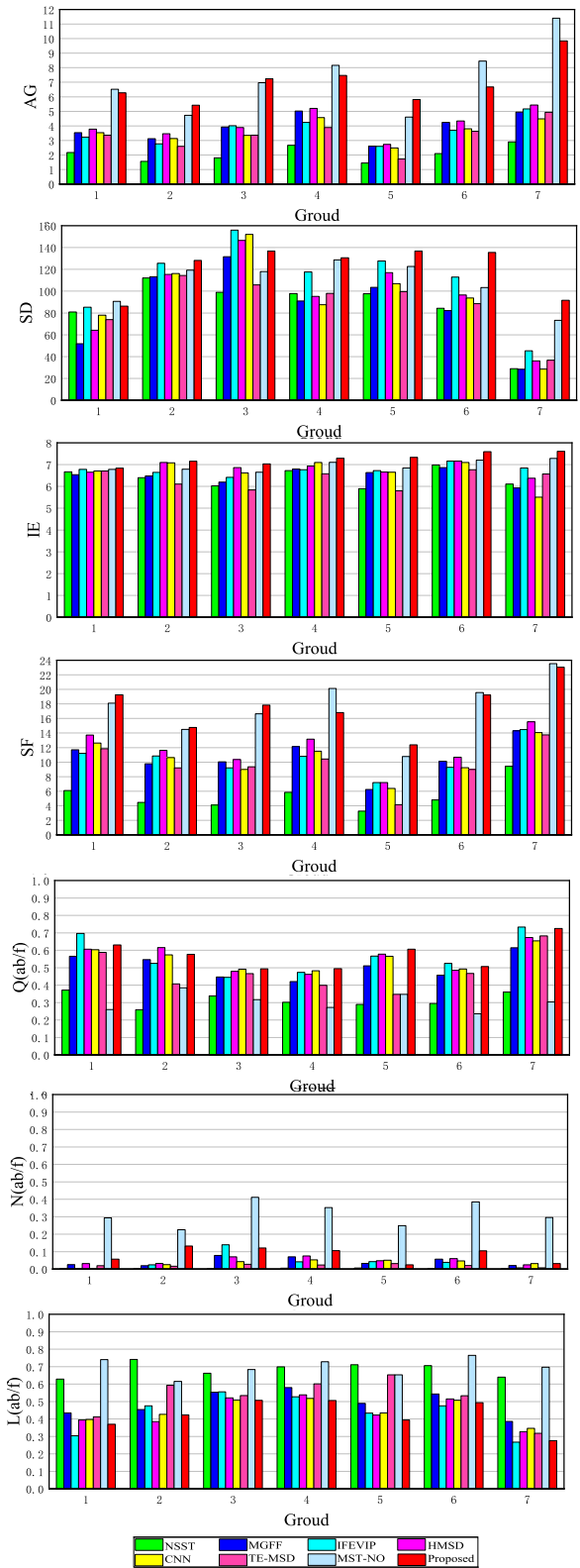
obvious discontinuity. In the fusion images based on HSMD and CNN, the details in dark areas of the visible images are lost. The TE-MSD-based fusion images have the same effects

TABLE 1. Quantitative evaluation of different methods for seven image pairs.

Experimental results	Index	Methods							
		NSST	MGFF	IFEVIP	HMSD	CNN	TE-MSD	MST-NO	Proposed
first group	AG	2.1691	3.5294	3.2284	3.7809	3.5297	3.3601	6.5238	6.2707
	SD	80.911	51.861	85.238	64.237	77.977	73.872	90.579	86.267
	IE	6.6664	6.5395	6.7825	6.6489	6.7072	6.7091	6.7853	6.8423
	SF	6.0813	11.703	11.209	13.726	12.619	11.835	18.104	19.251
	$Q_{AB/F}$	0.3712	0.5653	0.6964	0.6057	0.6027	0.5882	0.2600	0.6302
	$N_{AB/F}$	0.0033	0.0251	0.0029	0.0316	0.0038	0.0197	0.2931	0.0567
	$L_{AB/F}$	0.6288	0.4347	0.3036	0.3943	0.3973	0.4118	0.7400	0.3698
second group	AG	1.5552	3.1206	2.7639	3.4658	3.1321	2.5950	4.7245	5.4164
	SD	112.12	113.26	125.53	115.34	116.09	114.33	119.21	128.18
	IE	6.4003	6.4782	6.6452	7.0995	7.0827	6.1069	6.7892	7.1555
	SF	4.455	9.7798	10.832	11.615	10.612	9.2044	14.490	14.761
	$Q_{AB/F}$	0.2586	0.5457	0.5252	0.6152	0.5738	0.4076	0.3849	0.5766
	$N_{AB/F}$	0.0018	0.0195	0.0245	0.0327	0.0258	0.0163	0.2251	0.1324
	$L_{AB/F}$	0.7414	0.4543	0.4748	0.3848	0.4262	0.5924	0.6151	0.4234
third group	AG	1.7993	3.9282	4.0081	3.8900	3.3532	3.3625	6.9637	7.2414
	SD	98.872	131.52	156.01	146.44	152.05	105.80	117.92	136.71
	IE	6.0248	6.1950	6.4185	6.8656	6.6159	5.8396	6.6478	7.0351
	SF	4.0961	10.045	9.2050	10.358	9.0020	9.3488	16.652	17.833
	$Q_{AB/F}$	0.3382	0.4472	0.4448	0.4789	0.4911	0.4657	0.3172	0.4937
	$N_{AB/F}$	0.0014	0.0780	0.1389	0.0697	0.0426	0.0279	0.411	0.1211
	$L_{AB/F}$	0.6618	0.5528	0.5552	0.5211	0.5089	0.5343	0.6828	0.5063
fourth group	AG	2.6693	5.0151	4.2439	5.2041	4.5691	3.8913	8.1583	7.4706
	SD	97.796	91.035	117.62	95.248	87.476	97.991	128.72	130.61
	IE	6.7246	6.8007	6.7682	6.9346	7.1018	6.5754	7.1104	7.2974
	SF	5.8531	12.132	10.805	13.164	11.505	10.439	20.129	16.824
	$Q_{AB/F}$	0.3013	0.4201	0.4729	0.4617	0.4819	0.3991	0.2721	0.4940
	$N_{AB/F}$	0.0049	0.0704	0.0415	0.0749	0.0524	0.0229	0.3522	0.1053
	$L_{AB/F}$	0.6987	0.5799	0.5271	0.5383	0.5181	0.6009	0.7279	0.5060
fifth group	AG	1.4492	2.6116	2.5924	2.7317	2.4843	1.7305	4.6062	5.8117
	SD	97.682	103.44	127.66	117.01	106.85	99.568	122.62	136.82
	IE	5.8986	6.6381	6.7232	6.6589	6.6542	5.8011	6.8461	7.3380
	SF	3.2659	6.2285	7.1919	7.1961	6.3732	4.1499	10.763	12.382
	$Q_{AB/F}$	0.2894	0.5106	0.5661	0.5773	0.5651	0.3471	0.3469	0.6057
	$N_{AB/F}$	0.0061	0.0328	0.0431	0.0477	0.0512	0.0311	0.2485	0.0247
	$L_{AB/F}$	0.7106	0.4894	0.4339	0.4227	0.4349	0.6529	0.6531	0.3943
sixth group	AG	2.0973	4.2341	3.6989	4.3325	3.804	3.6288	8.4526	6.6725
	SD	84.353	82.231	112.85	96.548	93.784	88.506	103.23	135.49
	IE	6.9851	6.8663	7.1651	7.1628	7.1000	6.7660	7.2019	7.5957
	SF	4.7941	10.101	9.2726	10.683	9.2395	8.9828	19.564	16.228
	$Q_{AB/F}$	0.2943	0.4564	0.5246	0.4855	0.4923	0.4675	0.2355	0.5065
	$N_{AB/F}$	0.0034	0.0567	0.0383	0.0599	0.0463	0.0201	0.3844	0.1045
	$L_{AB/F}$	0.7057	0.5436	0.4754	0.5145	0.5077	0.5325	0.7645	0.4935
seventh group	AG	2.8944	4.9479	5.1674	5.4276	4.4779	4.9296	11.403	9.8463
	SD	28.922	28.409	45.359	36.116	28.542	36.663	73.269	91.692
	IE	6.1083	5.9275	6.8504	6.3784	5.5099	6.5718	7.2853	7.6092
	SF	9.4389	14.302	14.479	15.535	14.050	13.742	23.541	23.052
	$Q_{AB/F}$	0.3599	0.6141	0.7334	0.6729	0.6534	0.6825	0.3032	0.7246
	$N_{AB/F}$	0.0007	0.0198	0.0075	0.0242	0.0312	0.0076	0.2959	0.0313
	$L_{AB/F}$	0.6401	0.3859	0.2666	0.3271	0.3466	0.3175	0.6968	0.2754

as the NSST under the condition of high brightness of the source images, which are in low contrast. Under low illumination conditions, the effects of preserving detailed information are not good. The fusion images based on MST-NO have a better ability to highlight infrared targets, but the ability to retain visible source image information is poor. Additionally, many artifacts are added in the fusion process, which results in poor visual quality. In contrast, the background information and details are better retained in the proposed fused images, such as the billboard information in Fig. 3 (j) and the facial expressions and clothing logos in Fig. 9. Most of our fusion images exhibit good continuity without any halos. Therefore, the method in this paper demonstrated outstanding advantages in the fusion of infrared images and visible images, such as clearer fused images, more details and background information of the visible images retained, and higher overall visual quality.

To further assess the proposed method, we introduced seven well-known objective evaluation indicators based on the statistical characteristics of the fusion image,

TABLE 2. Running Time with the different methods on 360 × 270 source images.

Methods	NSST	MGFF	IFEVIP	HMSD	CNN	TE-MSD	MST-NO	Proposed
Time (s)	6.86	0.15	0.064	2.08	18.54	1.89	15.17	2.51

i.e., average gradient (AG) [38], standard deviation (SD) [39], information entropy (IE) [40], spatial frequency (SF) [41], total fusion performance parameter ($Q_{AB/F}$) [42], fusion artifacts ($N_{AB/F}$) [42], and fusion loss ($L_{AB/F}$) [42], in addition to subjective human visual effect analysis. The experimental evaluation indicators are shown in Table 1, where the value marked in red indicates the best performance of all methods, and the value marked in blue indicates the second-ranked performance. The proposed method occupies the top two positions in almost all metrics scores, and most of them rank first. To obtain a more intuitive grasp of the quantitative evaluation of various methods, the values in Table 1 are visualized in Fig. 11. In AG, the proposed method and MST-NO-based method are significantly ahead of the other six methods, but MST-NO is 1.04% higher than the proposed method on average. The proposed method has an average improvement of 41%, 41%, 10%, 26%, 28%, 37% and 12%, in SD. In IE, the proposed method has an average improvement of 14%, 12%, 8%, 7%, 9%, 15% and 5%. In SF, the proposed method and MST-NO are significantly ahead of the other six methods, but compared with MST-NO, the proposed method has an average increase of 0.07%. In $Q_{AB/F}$, except for the relatively small values of NSST and MST-NO, the other methods are not much different from the proposed method. The proposed method has an average increase of 1.7% over the second-ranked IFEVIP. In $N_{AB/F}$, NSST has a substantial advantage, and the other methods are not much different. Therefore, it can be concluded that the proposed method has obvious advantages over the other five selected methods in fusing infrared images and visible images, including retaining more information, better image quality, and higher definition.

Finally, the time spent with the different methods on 360 × 270 source images is provided in Table 2. All the experiments are performed using MATLAB R2019(a) on a laptop with a 2.67 GHz Intel Core CPU and 4 GB RAM. IFEVIP has an absolute advantage in time consumption, followed by MGFF. In addition, the method in this paper is slightly slower than HMSD and TE-MSD. From the overall point of view of the proposed method, GF is a simple local linear filtering method. The algorithm process does not need to be iterated, and the local sliding window strategy inside the algorithm can be efficiently implemented by a box filter, so the algorithm is very efficient and can reach the O(N) level. In the fusion process of the detail layers, the weight coefficient is directly obtained from the detail layer, which also shortens the time. Therefore, the reason the proposed method consumes considerable time is mainly due to the need to judge each pixel value in the fusion of the base layers. The improvement of the base layer fusion rules will be one of the main works in the future.

IV. CONCLUSION

In this paper, an efficient infrared and visible image fusion algorithm for highlighting details was proposed. First, a method combining dynamic range compression and contrast restoration based on a guided filter was introduced to adaptively increase the contrast of visible source images. Second, the infrared and enhanced visible images were divided into base layers and detail layers by using a GF-based image multiscale method. Then, to fuse the detail layer, by taking advantage of a multiscale gradient function, the saliency map was generated and then normalized to formulate the weight map, which greatly reduces the complexity of the algorithm. For base layer fusion, a fusion strategy was proposed in which the measures of details and energy of the source image are calculated to determine the pixel value of the base layer of the fused image to prevent energy loss, highlight texture details, and obtain more source image details. Finally, recursive separation and weighted histogram equalization were applied to the fused image for further optimization. The experimental results showed that the proposed method is not only subjectively consistent with the human visual effects but also highly evaluated in terms of the excellent scores on the quantitative evaluation index, which further proves its advantages and effectiveness over other methods.

The performance enhancement of the time-consuming procedures in the fusion will be pursued in future work.

REFERENCES

[1] D. D. Sworder, J. E. Boyd, and G. A. Clapp, "Image fusion for tracking manoeuvring targets," *Int. J. Syst. Sci.*, vol. 28, no. 1, pp. 1–14, Jan. 1997.

[2] X. Gang, Y. Bo, and J. Zhongliang, "Infrared and visible dynamic image sequence fusion based on region target detection," in *Proc. 10th Int. Conf. Inf. Fusion*, 2007, pp. 1–5.

[3] L. Junwu, B. Li, and Y. Jiang, "An infrared and visible image fusion algorithm based on LSWT-NSST," *IEEE Access*, vol. 8, pp. 179857–179880, 2020.

[4] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jun. 2017.

[5] W. Gan, X. Wu, W. Wu, X. Yang, C. Ren, X. He, and K. Liu, "Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter," *Infr. Phys. Technol.*, vol. 72, pp. 37–51, Sep. 2015.

[6] C. Gao, C. Song, Y. Zhang, D. Qi, and Y. Yu, "Improving the performance of infrared and visible image fusion based on latent low-rank representation nested with rolling guided image filtering," *IEEE Access*, vol. 9, pp. 91462–91475, 2021.

[7] X. Qian, L. Han, and Y. Cheng, "An object tracking method based on local matting for night fusion image," *Infr. Phys. Technol.*, vol. 67, pp. 455–461, Nov. 2014.

[8] L. Li, H. Ge, Y. Tong, and Y. Zhang, "Face recognition using Gabor-based feature extraction and feature space transformation fusion method for single image per person problem," *Neural Process. Lett.*, vol. 47, no. 3, pp. 1197–1217, 2018.

[9] Z. Zhang and R. Blum, "Region-based image fusion scheme for concealed weapon detection," in *Proc. 31st Annu. Conf. Inf. Sci. Syst.*, 1997, pp. 168–173.

[10] M. Li and Y. Dong, "Review on technology of pixel-level image fusion," in *Proc. 2nd Int. Conf. Meas., Inf. Control*, vol. 1, 2013, pp. 341–344.

[11] A. Dogra, B. Goyal, and S. Agrawal, "From multi-scale decomposition to non-multi-scale decomposition methods: A comprehensive survey of image fusion techniques and its applications," *IEEE Access*, vol. 5, pp. 16040–16067, 2017.

[12] Q. Zhang and X. Maldague, "An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing," *Infr. Phys. Technol.*, vol. 74, pp. 11–20, Jan. 2016.

[13] A. Rizzi, C. Gatta, and D. Marini, "A new algorithm for unsupervised global and local color correction," *Pattern Recognit. Lett.*, vol. 24, no. 11, pp. 1663–1677, Jul. 2003.

[14] F. Wang, B. Zhang, C. Zhang, W. Yan, Z. Zhao, and M. Wang, "Low-light image joint enhancement optimization algorithm based on frame accumulation and multi-scale retinex," *Ad Hoc Netw.*, vol. 113, Mar. 2021, Art. no. 102398.

[15] S. Liu, Y. Piao, and M. Tahir, "Research on fusion technology based on low-light visible image and infrared image," *Opt. Eng.*, vol. 55, no. 12, pp. 1–9, 2016.

[16] Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," *Appl. Opt.*, vol. 55, no. 23, pp. 6480–6490, Aug. 2016.

[17] 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.

[18] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shear-let transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.

[19] R. Wei, D. Zhu, W. Zhan, and Z. Hao, "Infrared and visible image fusion based on RPCA and NSST," in *Proc. IEEE Int. Conf. Power, Intell. Comput. Syst. (ICPICS)*, Jul. 2019, pp. 236–240.

[20] Y. X.-F. DENG Li-nuan, "Research on the fusion algorithm of infrared and visible images based on non-subsampled shearlet transform," (in Chinese), *Acta Electonica Sinica*, vol. 45, no. 12, pp. 2965–2970, 2017.

[21] X. Huang, G. Qi, H. Wei, Y. Chai, and J. Sim, "A novel infrared and visible image information fusion method based on phase congruency and image entropy," *Entropy*, vol. 21, no. 12, p. 1135, Nov. 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/12/1135

[22] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: A novel decomposition method for infrared and visible image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4733–4746, 2020.

[23] R. Nie, C. Ma, J. Cao, H. Ding, and D. Zhou, "A total variation with joint norms for infrared and visible image fusion," *IEEE Trans. Multimedia*, early access, Mar. 12, 2021, doi: 10.1109/TMM.2021.3065496.

[24] K. He and J. Sun, "Fast guided filter," *CoRR*, vol. abs/1505.00996, May 2015. [Online]. Available: http://arxiv.org/abs/1505.00996

[25] Y. Ding, J. Xiao, and J. Yu, "Importance filtering for image retargeting," in *Proc. CVPR*, 2011, pp. 89–96.

[26] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[27] H. N. Li Tao, "A multi-sensor image fusion and enhancement system for assisting drivers in poor lighting conditions," in *Proc. 34th Appl. Imag. Pattern Recognit. Workshop (AIPR)*, 2005, pp. 106–113.

[28] Z. Liu and R. Laganière, "Context enhancement through infrared vision: A modified fusion scheme," *Signal, Image Video Process.*, vol. 1, no. 4, pp. 293–301, 2007, doi: 10.1007/s11760-007-0025-4.

[29] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proc. 29th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*. New York, NY, USA: Association for Computing Machinery, 2002, pp. 257–266.

[30] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.

[31] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 25–46, Jul. 2008.

[32] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, "Multi-scale guided image and video fusion: A fast and efficient approach," *Circuits, Syst., Signal Process.*, vol. 38, no. 12, pp. 5576–5605, Dec. 2019.

[33] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visual image fusion through infrared feature extraction and visual information preservation," *Infr. Phys. Technol.*, vol. 83, pp. 227–237, Jun. 2017.

[34] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Inf. Fusion*, vol. 30, pp. 15–26, Jul. 2016.

[35] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 3, 2018, Art. no. 1850018.

[36] J. Chen, X. Li, L. Luo, X. Mei, and J. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Inf. Sci.*, vol. 508, pp. 64–78, Jan. 2020.

[37] G. Li, Y. Lin, and X. Qu, "An infrared and visible image fusion method based on multi-scale transformation and norm optimization," *Inf. Fusion*, vol. 71, pp. 109–129, Jul. 2021.

[38] 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.

[39] J. M. Román, J. V. Noguera, H. Legal-Ayala, D. Pinto-Roa, S. Gomez-Guerrero, and M. G. Torres, "Entropy and contrast enhancement of infrared thermal images using the multiscale top-hat transform," *Entropy*, vol. 21, no. 3, p. 244, Mar. 2019.

[40] P. Zhu, L. Ding, X. Ma, and Z. Huang, "Fusion of infrared polarization and intensity images based on improved toggle operator," *Opt. Laser Technol.*, vol. 98, pp. 139–151, Jan. 2018.

[41] X. Bai and Y. Zhang, "Detail preserved fusion of infrared and visual images by using opening and closing based toggle operator," *Opt. Laser Technol.*, vol. 63, pp. 105–113, Nov. 2014.

[42] K. C. Rajini and S. Roopa, "A review on recent improved image fusion techniques," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2017, pp. 149–153.

PEIJIN LIU received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1994 and 2000, respectively, and the Ph.D. degree in mechanical engineering from Xi'an University of Architecture and Technology, Xi'an, in 2012.

She is currently an Associate Professor with Xi'an University of Architecture and Technology. Her research interests include theory and technology in the field of status monitoring and recognition of electromechanical systems.
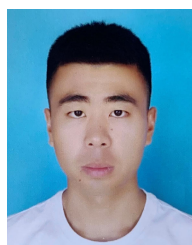
LICAI ZHANG received the B.S. degree in electrical engineering from Xi'an University of Architecture and Technology, Xi'an, China, in 2019, where he is currently pursuing the M.S. degree in electrical engineering.

His research interests include fault diagnosis and image fusion.

MINGYANG LI received the B.S. degree in electrical engineering from Xihua University, Chengdu, China, in 2019. He is currently pursuing the M.S. degree in electrical engineering with Xi'an University of Architecture and Technology, Xi'an, China.

His research interests include infrared image simulation and processing.

XIANGRUI ZHANG received the B.S. degree in building electrical and intelligence from Anhui University of Science and Technology, Anhui, China, in 2020. He is currently pursuing the M.S. degree in electrical engineering with Xi'an University of Architecture and Technology, Xi'an, China.

His research interest includes image processing.

● ● ●