Infrared Physics & Technology 92 (2018) 30–43

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Regular article

Adaptive fusion framework of infrared and visual image using saliency detection and improved dual-channel PCNN in the LNSST domain

Boyang Cheng [a,b,*], Longxu Jin [a], Guoning Li [a]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

HIGHLIGHTS

- LNSST, which is an upgrade based on the NSST, is used as a multi-scale analysis tool.
- The adaptive weighted fusion of the low-pass sub-image is guided by saliency detection method.
- An improved dual-channel PCNN model is proposed to process the high-pass sub-images.
- An adaptive fusion method of infrared and visual image is proposed.

ARTICLE INFO

Article history:
Received 22 November 2017
Revised 24 April 2018
Accepted 28 April 2018
Available online 7 May 2018

Keywords:
LNSST
Image fusion
Improved dual-channel PCNN
Frequency-tuned saliency detection

ABSTRACT

This paper presents an adaptive fusion framework of infrared and visual images using saliency detection and an improved dual-channel pulse-coupled neural network (ID-PCNN) in the local non-subsampled shearlet transform (LNSST) domain. The first step is to use the LNSST, an upgrade of the non-subsampled shearlet transform, for multi-scale analysis to separate the source images into low-pass and high-pass sub-images. The final fusion effect is determined by the fusion rule of the low-pass component. Thus, an improved algorithm based on frequency-tuned saliency extraction is adopted to guide the adaptive weighted fusion of the low-pass sub-image. An ID-PCNN model is used as the fusion rule for high-pass sub-images. A sum of directional gradients acts as the linking strength to characterize the texture details of an image. A modified spatial frequency that reflects the gradient features of images is used to motivate neurons. A series of images from diverse scenes is used for fusion experiments. Fusion results are evaluated subjectively and objectively. The results show that our algorithm exhibits superior fusion performance and is more effective than typical fusion techniques.

1. Introduction

Image fusion is an important branch of information fusion. Certain multi-source images from the same scene can be merged into a new image following certain rules such that the scene is described accurately and comprehensively [1]. Infrared (IR) and visual image fusion is the most widely used heterologous image fusion. An IR sensor generates images by detecting the thermal radiation of objects; this tool can reveal hidden thermal targets and is not strongly affected by external disturbances. A visible light imaging sensor generates images by capturing the reflection of objects with detailed information of high resolution. The combination of these two sensors effectively highlights a target and increases image resolution [2].

Many fusion approaches have been proposed recently, particularly for pixel level-based visible image (VI) and IR image fusion. Ma et al. [3] presented a fusion method via gradient transfer and total variation minimization. This method integrates the gradient texture information of visible light into an IR image, which provides a basis for the rapid identification of IR targets. However, the fused images have low contrast and lose considerable visual information. Bai et al. [4–7] proposed a novel fusion model based on multiscale new top-hat transform and obtained good results. However, the algorithm requires three parameters, making it difficult for inexperienced users to implement. Li et al. [8] established a weighted average fusion method using a guided filter. This method solved the problem of multi-focus image fusion perfectly, but the saliency maps generated may not be fully suitable for VI and IR image fusion. Liu et al. [9] proposed a general framework for image

* Corresponding author at: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China.
E-mail addresses: boyangwudi@163.com, chengboyang15@mails.ucas.edu.cn (B. Cheng).

https://doi.org/10.1016/j.infrared.2018.04.017
1350-4495/

fusion based on multi-scale transform and sparse representation. The sparse representation algorithm requires intense dictionary training; thus, the algorithm is highly complicated and time-consuming. Chen et al. [10] used deep learning to guide the fusion problem, but this application is not yet mature.

Compared with the abovementioned algorithms, multi-scale geometric analysis (MGA) is a relatively simple and easy fusion method for obtaining an image's features. This method decomposes the source image into a series of multi-direction and multi-scale sub-images, and appropriate fusion rules are applied to transform the corresponding sub-band images into fused images. The following MGA methods are currently used: the contourlet transform [11], non-subsampled contourlet transform (NSCT) [12], dual-tree complex wavelet transform [13], shearlet transform [14], and non-subsampled shearlet transform (NSST) [15]. The NSST is widely preferred by many researchers due to its rapid decomposition rate and efficient decomposition. A large shearlet filter causes a spectral aliasing phenomenon in the directional localization stage of the NSST. Thus, a local non-subsampled shearlet transform (LNSST) is proposed in this paper as a tool for the multi-scale decomposition of an image based on the NSST. This method uses a local small shearlet filter to avoid the block effect and weaken the Gibbs-ringing phenomenon, which improves time domain convolution calculation efficiency.

A weighted summation is usually treated as the fusion rule for the decomposition of low-pass sub-images. However, for IR and VI image fusion, this rule reduces the contrast of the fused image due to the spectral difference between the two low-pass components. The fused image must incorporate hidden or salient IR target information adaptively into humans' sensitive visual environment such that the fused image is not unnaturally reconstructed, and the information pertaining to salient objects can be well retained in the source image [16]. Achanta et al. [17] proposed a frequency-tuned approach to compute saliency in images. This method identifies salient objects and regions in an image precisely and forms a saliency map. The saliency map contains the weight information of the spatial distribution of the gray-scale value of the image. This map is used as a weighting function. The fusion rule of a low-pass sub-image can be switched from a weighted-mean to a weighted-adaptive rule, which effectively merges salient information into the fused image and improves the fusion effect.

High-pass sub-images usually adopt the conventional "max-absolute" [18] as the fusion rule, in which large absolute values correspond to highly salient texture features, but this rule easily loses redundant information from the source image. A pulse-coupled neural network (PCNN) [19] is a novel artificial neural network model that has been widely applied in image fusion because of its characteristics of global coupling and pulse synchronization of neurons. However, the linking strength is always constant, and the external stimuli usually adopt a single pixel value in traditional PCNN models. To remedy these defects, an improved dual-channel PCNN model (ID-PCNN) is proposed to process high-pass sub-images information, exploit local image information thoroughly and extract image details.

Given the preceding review, this paper presents an adaptive fusion framework for IR and visual images using saliency detection and an improved dual-channel pulse-coupled neural network (ID-PCNN) in the LNSST domain. First, the LNSST is used as a multi-scale analysis tool to decompose source images into a low-pass sub-image and a series of high-pass sub-images. Second, an improved algorithm based on frequency-tuned (FT) saliency extraction is adopted to guide the adaptive weighted fusion of the low-pass sub-image. The ID-PCNN model is proposed for high-pass sub-images. This model retains the spatial 2D information integrity of the input image after the fusing process. Finally, each sub-image is modeled, and the corresponding fusion coefficients are produced. The algorithm can effectively express the characteristics of the image and obtain a good fusion effect using IR and visible light images in experiments.

The remainder of this paper is organized as follows. Chapter 2 details the principles of the LNSST. Chapter 3 introduces the theory underlying FT and ID-PCNN. Chapter 4 elaborates on the algorithm based on the new fusion rule. Chapter 5 presents the results of five experiments and intuitively compares the proposed method with other methods. Chapter 6 summarizes the findings.

2. LNSST

When the dimension is $n = 2$, the shearlet system function with discrete parameters is as follows:

$$S_{AB}(\varphi) = \{\varphi_{j,l,k} = |detA|^{j/2}\varphi\left(B^l A^j x - k\right); j, l \in Z, k \in Z^2\} \tag{1}$$

where $\varphi \in L^2(R^2)$, A and B are 2×2 reversible matrices, $|\det B| = 1$, j is the scale parameter, l is the direction parameter, and k is the spatial position. For $j \geq 0$, $-2^j \leq l \leq 2^j - 1$, $k \in Z^2$, and $d = 0, 1$, the Fourier transform of the shearlet can be expressed based on the tight support frame:

$$\widehat{\varphi}_{j,l,k}^{(d)} = 2^{3j/2} V(2^{-2j}\xi) W_{j,l}^{(d)}(\xi) e^{-2\pi i \xi A_d^{-j} B_d^{-l} k} \tag{2}$$

where $V(2^{-2j}\xi)$ is the scale function, $W_{j,l}^{(d)}$ is the window function localized on the trapezoidal pair, A_d is the heterosexual expansion matrix, and B_d is the shear matrix. The shearlet transform of the function $f \in L^2(R^2)$ can be calculated by Eq. (3).

$$\left\langle f, \widehat{\varphi}_{j,l,k}^{(d)} \right\rangle = 2^{3j/2} \int_{R^2} \widehat{f}(\xi) \overline{V(2^{-2j}\xi) W_{j,l}^{(d)}(\xi)} e^{-2\pi i \xi A_d^{-j} B_d^{-l} k} d\xi \tag{3}$$

Eq. (3) shows that the shearlet transform is divided into two steps. The first step is a multi-scale decomposition ($\widehat{f}(\xi)\overline{V(2^{-2j}\xi)}$), and the second step is the calculation of the direction of localization ($\widehat{f}(\xi)\overline{V(2^{-2j}\xi) W_{j,l}^{(d)}(\xi)}$).

Multi-scale decomposition: The image is subjected to non-subsampled pyramid decomposition using a non-subsampled 2*D* filter bank of dual channels to generate a low-pass sub-image and multiple high-pass sub-images with perfect reconstruction.

Directional localization: Directional localization is achieved by small shearlet filters and high-pass sub-image convolution calculations. The traditional NSST does not specify the size and scale parameters of shearlet filters. Large-size shearlet filters are prone to cause the Gibbs-ringing phenomenon, which introduces false error information into the final fusion image and affects the fusion effect. To remedy the abovementioned defects, the LNSST is constructed based on the NSST by using a local small shearlet filter to improve the computational efficiency of the time domain convolution and weaken the block effect of the image. To fully combine the shearlet characteristics of parabolic scaling with its localization, a parabolic constraint function is proposed as follows:

$$\begin{cases} L \leqslant min(\sqrt{M}, \sqrt{N}) \\ L \geqslant 2^j + 1 \\ L = n \cdot (2^{j-1} + 1) \end{cases}$$

where the size of the shearlet filter is $L \times L$, j is the scale parameter, and n is any positive integer. In this manner, the size and scale parameters of the shearlet filters are combined and constrained relative to one another; thus, the contour information of the image can be well extracted. Through a large number of experiments, it is found that when j is generally 2 or 3, the local window is usually 15×15; therefore, a relatively good decomposition effect of the image can be obtained.

The abovementioned shearlet transformation is called the LNSST. The local non-subsampled shearlet removes the sampling operation in the decomposition stage, which presents translation invariance because the local small shearlet filter can avoid the aliasing of the spectrum and improves image decomposition and reconstruction.

A local non-subsampled shearlet exhibits good local properties, excellent direction selectivity, and parabolic edge characteristics. The frequency domain-implemented and time domain-implanted shearlet filters are shown in Fig. 1(a)–(d). The number of directional sub-bands in the first layer is 4 and 8 in the second layer, and the size of the shearlet filter is 15×15.

3. Relevant theory

3.1. Frequency-tuned salient region detection

Achanta et al. proposed a frequency-tuned approach to compute saliency in images by choosing the DoG filter as the bandpass filter for identifying salient features in the frequency domain. The DoG filter is defined as

$$DoG(x,y) = \frac{1}{2\pi}\left[\frac{1}{\sigma_1^2}e^{-\frac{(x^2+y^2)}{2\sigma_1^2}} - \frac{1}{\sigma_2^2}e^{-\frac{(x^2+y^2)}{2\sigma_2^2}}\right] = G(x,y,\sigma_1) - G(x,y,\sigma_2) \tag{5}$$

where σ_1 and σ_2 are the standard deviations of the Gaussian function.

The DoG filter is a simple bandpass filter. The low-frequency cut-off value *wlc* is determined by σ_1. The high-frequency cut-off value *whc* is controlled by σ_2. The extracted saliency map preserves the expected spatial frequency characteristics by selecting appropriate values for σ_1 and σ_2. The value of σ_1 is set to infinity to retain all the other frequencies. Small values of σ_2 are chosen to eliminate high-frequency noise and texture. For an input image, the saliency map calculated by the DoG filter is

$$S(x,y) = ||I_\mu - I_{whc}|| \tag{6}$$

This method uses the *Lab* color space, and each pixel location is an $[L,a,b]^T$ vector, where $\boldsymbol{I_\mu}$ is the average of the image feature vectors, $\boldsymbol{I_{whc}}$ is the Gaussian blur filtered value of the corresponding image pixel vector at the pixel location (x, y), and $||\cdot||$ is the Euclidean distance. The algorithm diagram is displayed in Fig. 2.

An example of saliency map extraction of IR and VI images for the same scene is illustrated in Fig. 3(a)–(d). The original IR and VI images are shown in Fig. 3(a)–(b), and the corresponding saliency maps obtained by the FT algorithm are shown in Fig. 3(c)–(d). The saliency map extracts relatively important regions or regions to which human vision is sensitive, which lays the foundation for the adaptive weighted fusion of low-pass components.

3.2. Improved dual-channel PCNN model

A PCNN is obtained by studying neuron stimulation in the mammalian visual system and conforms to the visual information system of the human brain. Because human eyes are highly sensitive to the high-frequency components that represent the edge features of an image, the PCNN model with a visual restoration mechanism can perform bionic processing of high-pass sub-images and perfectly restore detailed information in the fused images [20,21]. The traditional single-channel PCNN model only reflects the information in one source image, while other original images are not considered. In addition, the traditional PCNN model features numerous peripheral parameters that cannot be ignored. Many values must rely on experience value settings, which significantly reduce the integration of accuracy. Thus, an improved PCNN model with dual-channel (ID-PCNN) is considered in this paper. The sum of directional gradients (SDG) represents the image information changed in the horizontal, vertical, and diagonal directions, which can be used as the adaptive linking strength of the high-pass sub-image to characterize the textural details of an image. The ID-PCNN model is as follows.

The received part is

$$F_{ij}{}^1(n) = S_{ij}{}^1(n) \tag{7}$$

$$F_{ij}{}^2(n) = S_{ij}{}^2(n) \tag{8}$$

$$L_{ij}(n) = \begin{cases} 1 & if \sum_{N(i,j)} Y_{ij}(n-1) > 0 \\ 0 & otherwise \end{cases} \tag{9}$$

$S^1{}_{ij}$ and $S^2{}_{ij}$ represent the gray-scale values normalized at the (i, j) position of the two source images. Their values are regarded as the model's external stimuli. L_{ij} is the linking input, and $F^1{}_{ij}$ and $F^2{}_{ij}$ represent two symmetrical feedback inputs. Y_{ij} denotes the external output of the neurons after internal processing.

The modulation part is

$$U_{ij}(n) = max\left\{F^1{}_{ij}(n)(1+\beta_{ij}^1 L_{ij}(n)), F^2{}_{ij}(n)(1+\beta_{ij}^2 L_{ij}(n))\right\} \tag{10}$$

U_{ij} is treated as the internal activity item of the neurons. $\beta^1{}_{ij}$ and $\beta^2{}_{ij}$ represent the linking strength, which reflects the strength of the human visual system's response to different characteristic regions in the image.

The pulse-generated domain is

$$Y_{ij}(n) = \begin{cases} 1 & if\ U_{ij}(n) \geqslant \theta_{ij}(n-1) \\ 0 & otherwise \end{cases} \tag{11}$$

$$\theta_{ij}(n) = \theta_{ij}(n-1) - \Delta + V_\theta Y_{ij}(n) \tag{12}$$

Δ is the declining extent of the dynamic threshold, which can be regarded as 0.01 to ensure that the decay rate of the dynamic threshold is mild. θ_{ij} is the threshold function. V_θ determines the threshold of the fired neuron and should be set to a relatively large value (e.g., 10).

The firing times of each neuron are

$$T_{ij} = \begin{cases} n & if\ Y_{ij}(n) = 1 \text{ for the first time} \\ T_{ij}(n-1) & \text{otherwise} \end{cases} \tag{13}$$

n denotes the iteration number, and T_{ij} [22] is used to adaptively determine the number of reasonable iterations.

The adaptive linking strength is computed as follows:

$$SDG = \nabla D_h + \nabla D_v + \nabla D_{md} + \nabla D_{vd} \tag{14}$$

$$\nabla C_h = \sqrt{\left(D_l^k(i,j) - D_l^k(i-1,j)\right)^2 + \left(D_l^k(i,j) - D_l^k(i+1,j)\right)^2} \tag{15}$$

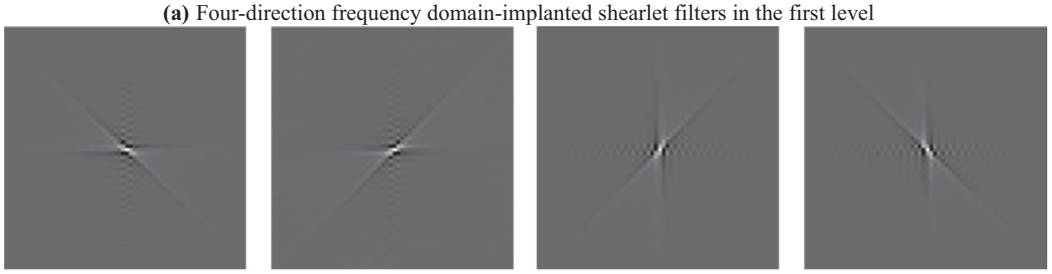
$$\nabla C_v = \sqrt{\left(D_l^k(i,j) - D_l^k(i,j-1)\right)^2 + \left(D_l^k(i,j) - D_l^k(i,j+1)\right)^2} \tag{16}$$

$$\nabla C_{md} = \sqrt{\left((D_l^k(i,j) - D_l^k(i-1,j-1))^2 + (D_l^k(i,j) - D_l^k(i+1,j+1))^2\right)/\sqrt{2}} \tag{17}$$

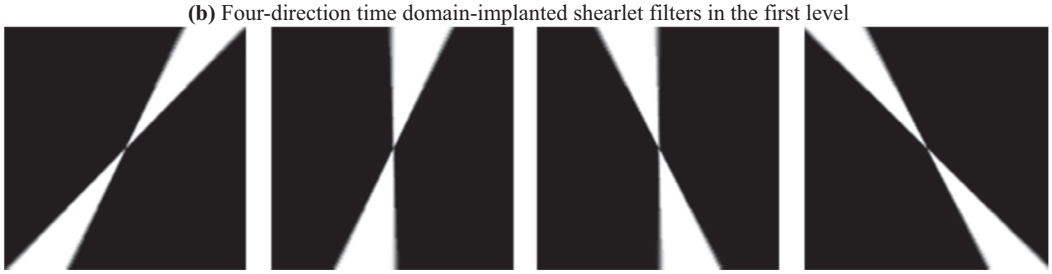
$$\nabla C_{vd} = \sqrt{\left((D_l^k(i,j) - D_l^k(i-1,j+1))^2 + (D_l^k(i,j) - D_l^k(i+1,j-1))^2\right)/\sqrt{2}} \tag{18}$$

$$\beta_2(i,j) = SDG_l^k(i,j) \tag{19}$$

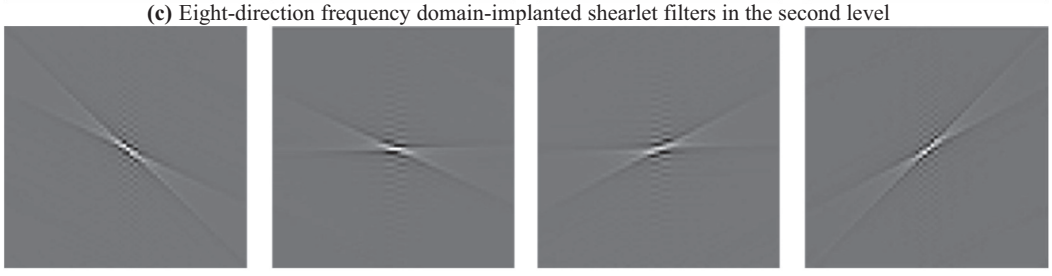
(a) Four-direction frequency domain-implanted shearlet filters in the first level



(b) Four-direction time domain-implanted shearlet filters in the first level



(c) Eight-direction frequency domain-implanted shearlet filters in the second level



(d) Eight-direction time domain-implanted shearlet filters in the second level

Fig. 1. Schematic of shearlet filters.

where $C_l^k(i,j)$ is the coefficient obtained by multi-scale analysis, l is the number of decomposed layers, and k is the number of decomposition directions of each layer. SDG represents the sum of the direction gradients at the pixel element. ∇C_h and ∇C_v represent changes in the horizontal and vertical directions, respectively. ∇C_{md} and ∇C_{vd} represent changes in the diagonal directions. Large direction gradient sums guarantee high local area clarity and fast neuronal activation.

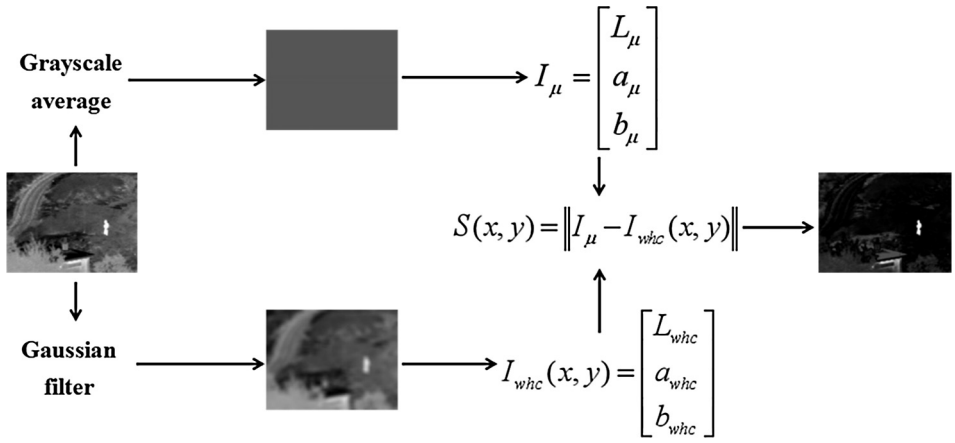


Fig. 2. Algorithm diagram of frequency-tuned salient region detection.

Fig. 3. Example of an extracted saliency map. (a) IR image, (b) VI image, and (c) and (d) saliency maps of IR image.

4. Fusion method

4.1. Fusion framework

The algorithm used in this paper aims to fuse heterologous images. The specific fusion framework is shown in Fig. 4. The proposed fusion steps can be summarized as follows.

(1) LNSST is used for the multi-scale decomposition of visible-light and IR images.
(2) Different fusion rules are obeyed to process the high-pass and low-pass sub-images given their differences. The low-pass sub-image adopts adaptive weight based on saliency detection as the fusion rule, whereas the high-pass sub-images are treated with the improved dual-channel PCNN model to output the high-pass fusion coefficients.
(3) The fused image is reintegrated based on an inverse NSST.

4.2. Fusion rule of low-pass sub-images

The low-pass component of the image represents the main energy of the image, which is an approximate part of the image. Thus, the fusion rules for the low-pass component determine the final fusion effect. To avoid incompatible spectral characteristics of heterologous images, an improved algorithm based on FT saliency extraction is adopted to guide the adaptive weighted fusion

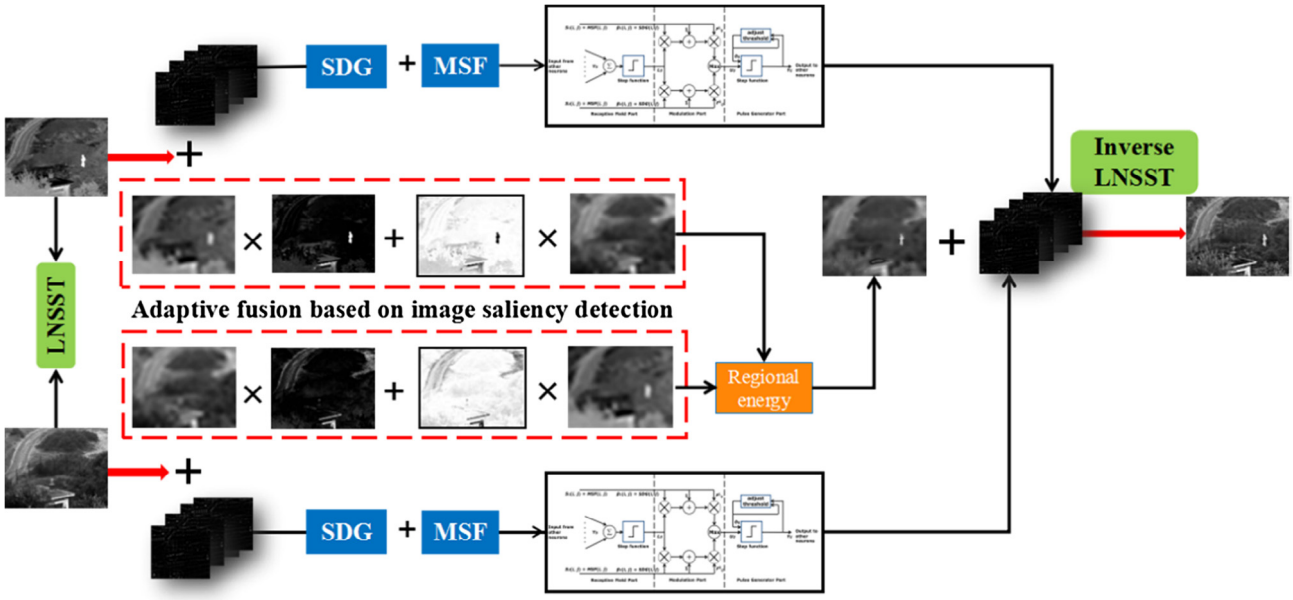


Fig. 4. Schematic of our proposed algorithm.

of the low-pass sub-image. The saliency map generated by the FT algorithm can intelligently identify the relatively important and salient parts of the image, thereby providing a basis for the adaptive weighting of the fusion rules. The fusion rules of the low-pass sub-image are as follows.

Step 1: FT algorithm is used for the saliency detection of IR and VI images, and the corresponding saliency maps SIR and SVI are generated. The gray values of the saliency maps are normalized to form the weighting coefficient matrices S1 and S_2.

$$S_1(i,j) = \frac{S_{IR}(i,j) - minS_{IR}(i,j)}{maxS_{IR}(i,j) - minS_{IR}(i,j)} \tag{20}$$

$$S_2(i,j) = \frac{S_{VI}(i,j) - minS_{VI}(i,j)}{maxS_{VI}(i,j) - minS_{VI}(i,j)} \tag{21}$$

Step 2: S_1 and S_2 are used to guide adaptive weighted fusion based on the saliency detection of IR and VI images, respectively [23,24]. The specific expression is as follows:

$$C^1_{l,k}(i,j) = S_1(i,j)\cdot C^{IR}_{l,k}(i,j) + (1 - S_1(i,j))\cdot C^{VI}_{l,k}(i,j)) \tag{22}$$

$$C^2_{l,k}(i,j) = S_2(i,j)\cdot C^{VI}_{l,k}(i,j) + (1 - S_2(i,j))\cdot C^{IR}_{l,k}(i,j)) \tag{23}$$

where $C^k_l\,(i, j)$ represents the multi-scale decomposition coefficients of the low-pass sub-image, l represents the decomposition number, k represents the direction of decomposition of each layer, and $C^1{}_{l,k}\,(i, j)$ and $C^2{}_{l,k}\,(i, j)$ represent the adaptive weighted fusion coefficients based on the saliency detection of IR and VI images, respectively.

Step 3: The two sets of fusion coefficients contain the results of the adaptive weighted fusion; thus, regional energy is used to combine the sets. This method can further combine the saliency information of IR and VI images based on the results of adaptive weight treatment. Thus, the effect of image fusion is robust. The process is as follows.

(1) The regional energy of the two groups of fusion coefficients is calculated.

$$E^1_{l,k}(i,j) = \sum_{m=-1}^{1}\sum_{n=-1}^{1} \omega(m+2,n+2)\left|C^1_{l,k}(m+i,n+j)\right| \tag{24}$$

$$E^2_{l,k}(i,j) = \sum_{m=-1}^{1}\sum_{n=-1}^{1} \omega(m+2,n+2)\left|C^2_{l,k}(m+i,n+j)\right| \tag{25}$$

$$\omega(m,n) = \frac{1}{15}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{26}$$

where $\omega(m, n)$ represents a window at position (i, j) in the 3×3 local region, and $E^1{}_{l,k}\,(i, j)$ and $E^2{}_{l,k}(i, j)$ represent the regional energy of the two groups of fusion coefficients in the window.

(2) The similarity between the two groups of fusion coefficients is calculated.

$$M^{1,2}_{l,k}(i,j) = \frac{2\sum_{m=-1}^{1}\sum_{n=-1}^{1}\omega(m+2,n+2)E^1_{l,k}(m+i,n+j)E^2_{l,k}(m+i,n+j)}{\left[E^1_{l,k}(m+i,n+j)\right]^2 + \left[E^2_{l,k}(m+i,n+j)\right]^2} \tag{27}$$

where a threshold T is set, which ranges between 0.3 and 0.4. When $M^{1,2}{}_{l,k}\,(i, j) < T$, the correlation between the two groups of coefficients is poor. A large regional energy value indicates that the region contains a considerable amount of salient information. Thus, the large coefficients are the final fusion coefficients. The consolidation rules for the fusion coefficients are as follows:

$$C^k_l(i,j) = \begin{cases} C^1_{l,k}(i,j) & if\ \mathrm{E}^1_{l,k}(i,j) > E^2_{l,k}(i,j) \\ C^2_{l,k}(i,j) & otherwise \end{cases} \tag{28}$$

When $M^{1,2}{}_{l,k}\,(i, j) > T$, both regions contain significant saliency information, and the two groups of fusion coefficients are combined into a weighted average. The two sets of fusion coefficients are obtained through adaptive weighted fusion. The weighted average strategy does not affect the fusion effect even when the coefficients are combined.

$$W(i,j) = \frac{1}{2} - \frac{1}{2}\left(\frac{1 - M^{l,k}_{1,2}(i,j)}{1 - T}\right) \tag{29}$$

$$C^k_l(i,j) = \begin{cases} W(i,j)\cdot C^1_{l,k}(i,j) + (1 - W(i,j))\cdot C^2_{l,k}(i,j) & E^1_{l,k}(i,j) > E^2_{l,k}(i,j) \\ (1 - W(i,j))\cdot C^1_{l,k}(i,j) + W(i,j)\times C^2_{l,k}(i,j) & E^1_{l,k}(i,j) \leqslant E^2_{l,k}(i,j) \end{cases} \tag{30}$$

where $W(i, j)$ represents the weight coefficient when the fusion coefficients are combined. The low-pass component is equivalent to twofold fusion, which solves the problem of low contrast and combines the salient information of the two source images. A schematic of the low-pass component fusion process is shown in Fig. 5.

4.3. Fusion rule of high-pass sub-images

The high-pass sub-images always reflect the edge features and textural details of the image. Using the traditional "maximum absolute value" as the fusion rule results in a significant loss of detailed information. Therefore, an ID-PCNN model is adopted to guide the fusion of high-pass components. The human visual system is highly sensitive to edges, directions, and other features, yet the true brightness of an independent position is not. A single pixel value is insufficient, which is why image features are used for the PCNN external stimulus. This study uses the modified spatial frequency (MSF) [25] as the external stimulus, which represents the change in the gray value of an image and reflects changes in image edges and texture details. A schematic of the high-pass component fusion process is shown in Fig. 6. The fusion steps are as follows:

Step 1: The parameters of the ID-PCNN model are initialized, and the value of the SDG of each pixel is used as the linking strength (Eqs. (14)-(19)).
Step 2: The MSF of the high-pass sub-images coefficients is calculated and acts as the external stimulus for the ID-PCNN model. The formula for the MSF is as follows:

$$MSF = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} RF + CF + MDF + SDF \tag{31}$$

$$RF = \left[D^k_l(i,j) - D^k_l(i,j-1)\right]^2 \tag{32}$$

$$CF = \left[D^k_l(i,j) - D^k_l(i-1,j)\right]^2 \tag{33}$$

$$MDF = \left[D^k_l(i,j) - D^k_l(i-1,j-1)\right]^2/\sqrt{2} \tag{34}$$

$$SDF = \left[D^k_l(i,j) - D^k_l(i-1,j+1)\right]^2/\sqrt{2} \tag{35}$$

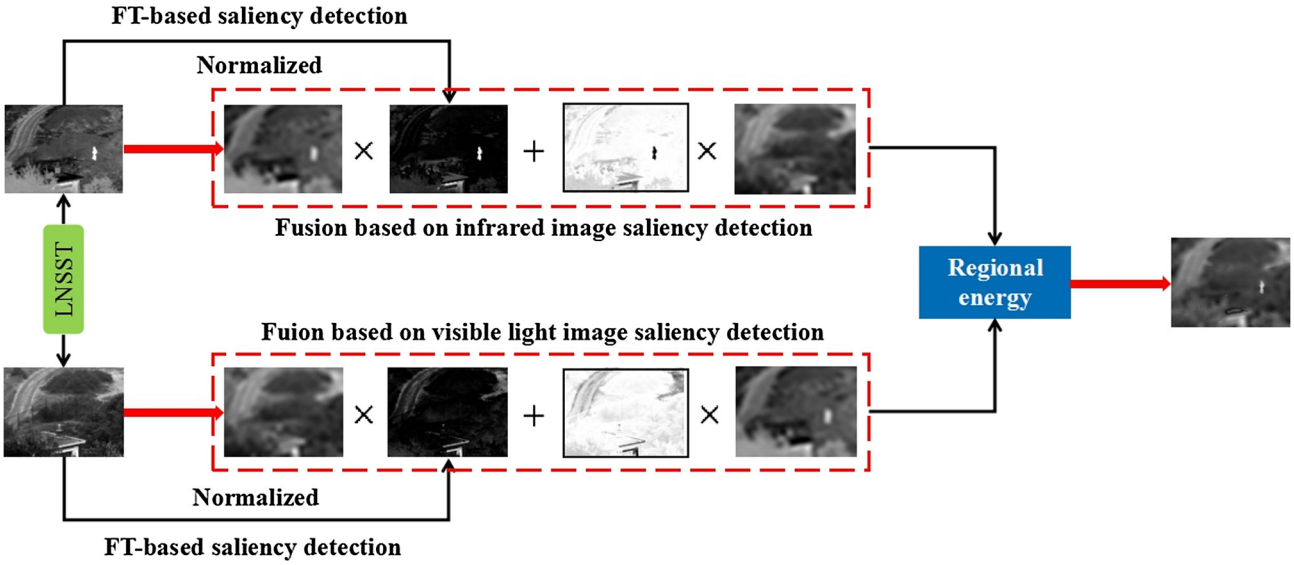


Fig. 5. Fusion diagram of low-pass component.

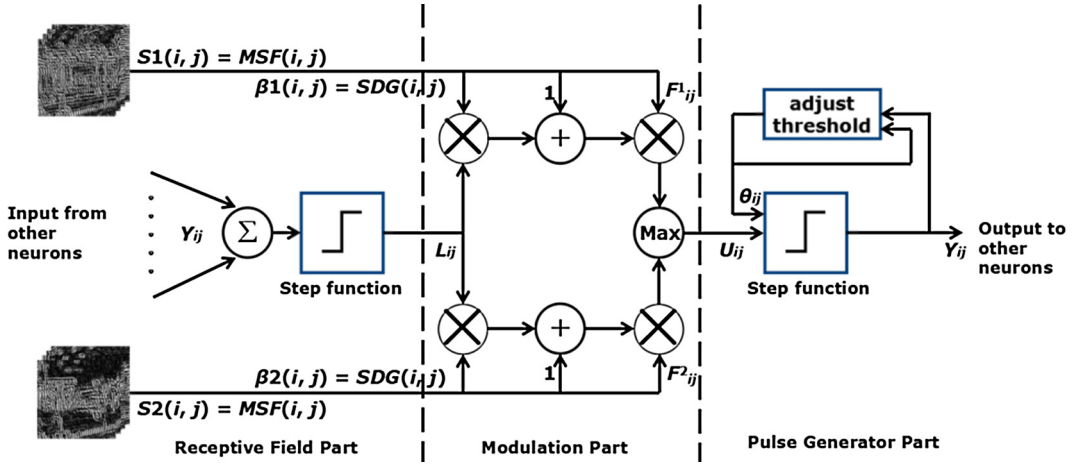


Fig. 6. Fusion diagram of high-pass sub-images.

where *RF*, *CF*, *MDF*, and *SDF* represent row frequency, column frequency, main diagonal frequency, and auxiliary diagonal frequency, respectively, and $M \times N$ is 3×3.

Step 3: Eqs. (7)–(13) are iterated until all the neurons are ignited, and $U_{ij}(n)$, $L_{ij}(n)$, $\theta_{ij}(n)$, $T_{ij}(n)$, and $Y_{ij}(n)$ are calculated. The fused coefficients are selected as follows:

$$C_l^k(i,j) = \begin{cases} C_L^{l,k}(i,j) & U_{ij}(n) = U_{ij}^L(n) \\ C_V^{l,k}(i,j) & U_{ij}(n) = U_{ij}^V(n) \end{cases} \tag{36}$$

$$U_{ij}^L(n) = F_{ij}^L(n)(1 + \beta_{ij}^L L_{ij}(n)) \tag{37}$$

$$U_{ij}^V(n) = F_{ij}^V(n)(1 + \beta_{ij}^V L_{ij}(n)) \tag{38}$$

5. Experimental results and analysis

To verify the superiority of the proposed method, our method is compared with the following methods: NSST-SF-PCNN [26], GFF [8], CNN [10], IFE-VIP [27], GTF [3], FT [28], LNSST-PCNN, and NSST-ID-PCNN. The LNSST-PCNN method combines the traditional single-channel PCNN model with the LNSST, and the NSST-ID-PCNN method combines the traditional NSST with the ID-PCNN model. Eight different environments are selected to demonstrate the versatility of the proposed algorithm. The first group shows a Jeep in front of a house, the second group depicts a man hidden in a forest, the third group shows a hidden tank, the fourth group depicts a man walking in a forest, the fifth group depicts an airport, the sixth group shows a ship on a lake, the seventh group depicts a coast, and the eighth group shows a woman standing in front of a car door. The original image sizes of the eight groups are 620×450, 505×510, 472×354, 360×270, 256×256, 505×510, 505×510, and 656×490, respectively. Our method adopts "maxflat" and [2,2,3,3] as the direction filter and pyramid filter, respectively. All experiments are conducted in MATLAB 2012a on a PC with an Intel Core i7/3.4 GHz/4G processor. The five sets of source images are presented in Fig. 7(a)–(h), from the first group to the fifth group.

5.1. Subjective evaluation

The first to fifth experiments are shown from top to bottom. The results of the NSST-SF-PCNN, GFF, CNN, IFE-VIP, GTF, FT, LNSST-PCNN, and NSST-ID-PCNN methods and the proposed method are

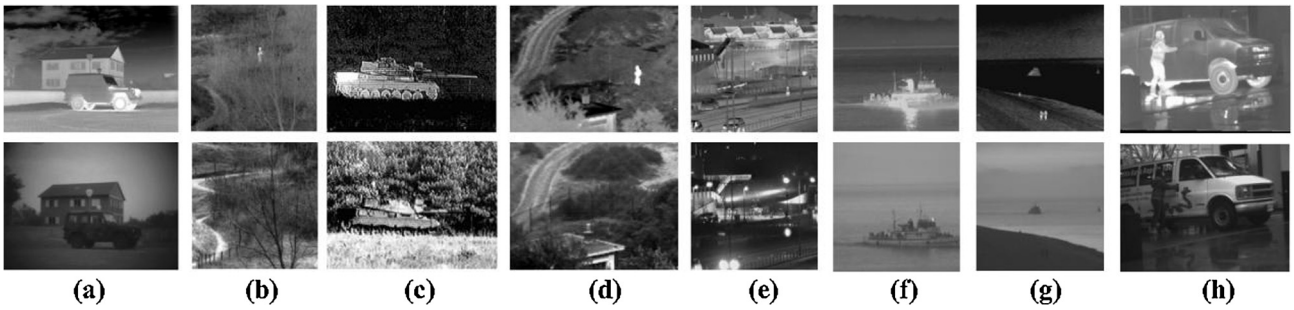


Fig. 7. Source images.

shown in Fig. (a) to Fig. (i) in each group, respectively. Globally, Fig. (i) for each experiment consistently looks clearer than Fig. (h). The main reason is that the LNSST is more effective than the NSST in decomposing the image; thus, the texture details of the image can be well restored into the fusion image. The enlarged details of each group well support the suitable effect of the proposed algorithm. Similarly, the effect of Fig. (g) appears much worse than that of Fig. (i) in each group; the discrepancy arises mainly because the ID-PCNN model can handle the high-frequency details of the image better than the traditional single-channel PCNN model can. The first group of experiments are shown in Figs. 8(a)–(i). The Fig. 8(i) that obtained by the proposed method has the highest contrast, the most background information retained, and the best view. The Figs. 8(a)–(g) are all close to the IR image and do not integrate the VI information into the fused image. The second group of experiments shows that the textural details of the trees are not reconstructed well in Fig. 9(a) and (c), and the contrast of Fig. 9(d) is low. Fig. 9(b) shows a relatively good appearance, but compared with Fig. 9 (i), the edge details of the whole image and textures of the trees are not as good as those produced by the method used in this paper. The third experiment shows that the brightness of Fig. 10(d) (see Fig. 11) is high, which results in ambiguous details for the tank. The background information of Fig. 10(e) is almost completely lost compared with Fig. 10(i). The contrast of the fused image obtained in this paper is high, and the textural information of the tank wheel is largely restored. The salient information of the IR image and the textural details of the VI image are combined perfectly in the background. The fourth group of experiments shows that the fusion effect of the algorithm introduced is better than the effects of the other five contrast algorithms, in which the image contrast is the highest, and the appearance is the clearest. The other five groups of fused images appear more like IR source images. Thus, many edge details of the VI source images are lost. Regarding the fifth group of experiments, Fig. 12(e) is ambiguous and has a poor view, and the lamp information of the VI image is not reconstructed well in Fig. 12(a)–(c). The image is excessively bright, which results in low contrast. The fused image obtained in this paper exhibits the best appearance. The combination of light and body contour is perfect under the lighting conditions, and the overall brightness is moderate. Regarding the sixth group of experiments, the backgrounds of Fig. 13(c) and Fig. 13(g) do not fuse the information pertaining to the mountain profile in the infrared image very well, and the hulls of Fig. 13 (d) and Fig. 13(f) are closer to the IR images, thus losing the edge information of the VI images. The contrast of Fig. 13(e) is not as good as that produced by the algorithm proposed in this paper,

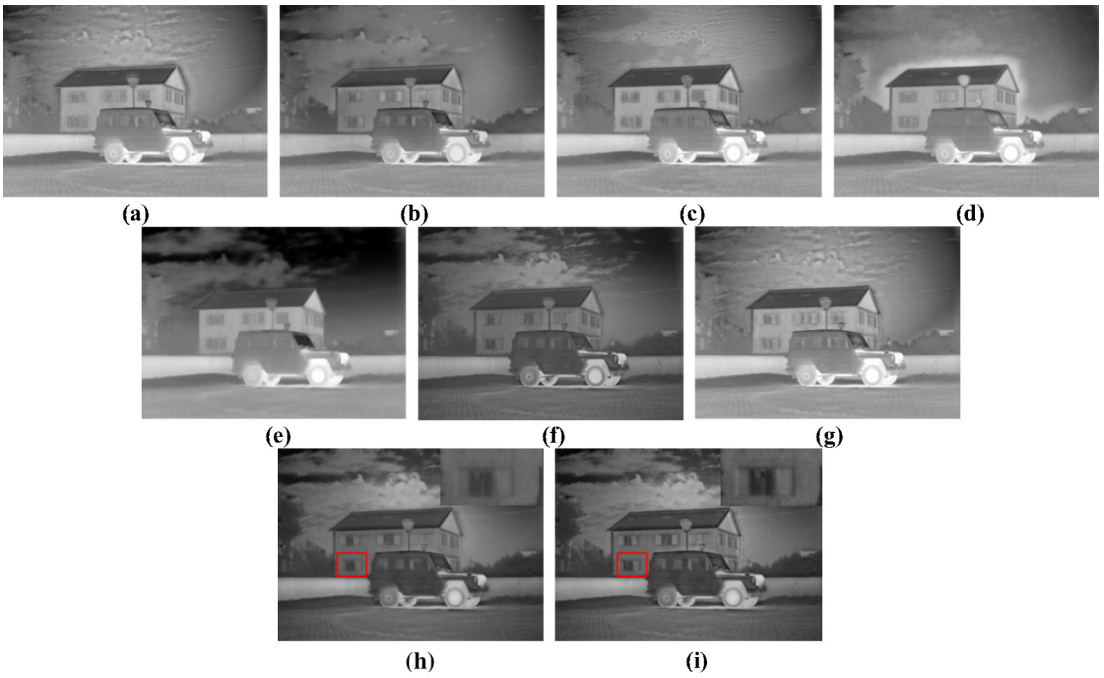


Fig. 8. First group of fusion experiments.

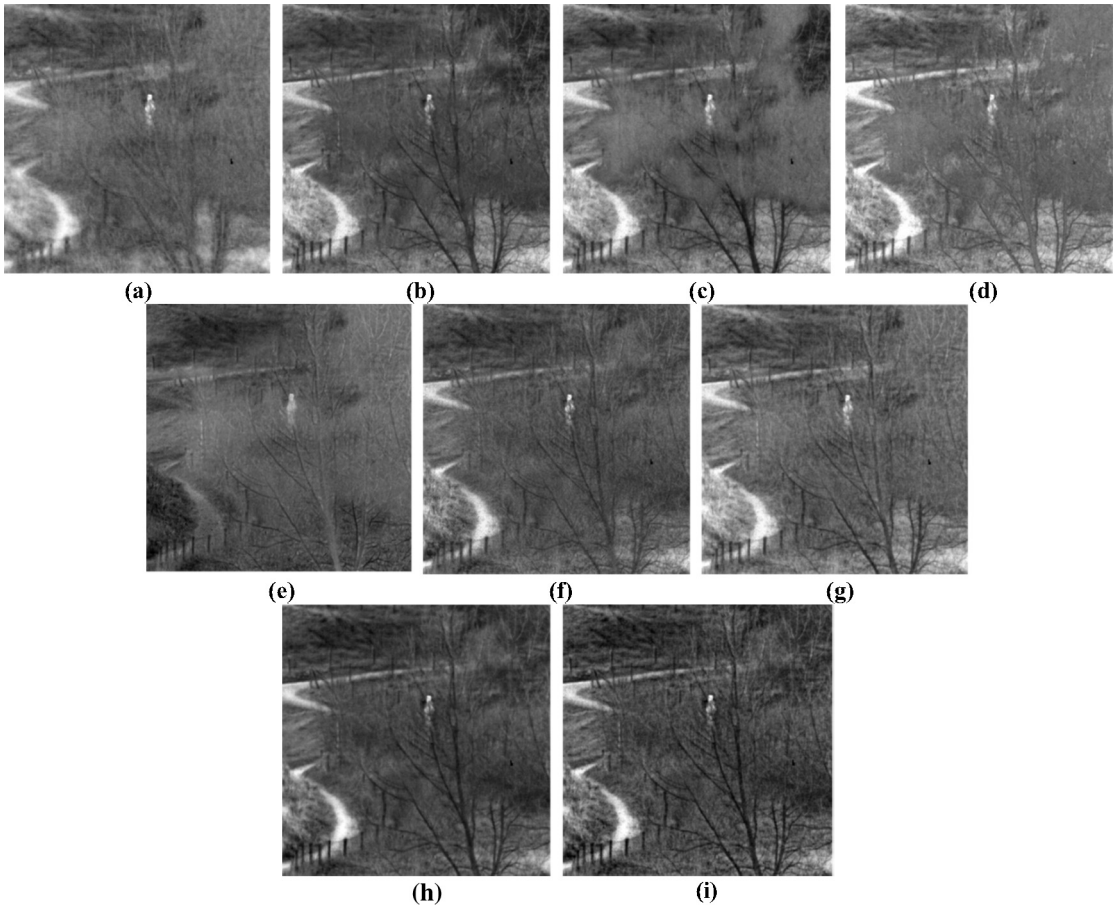


Fig. 9. Second group of fusion experiments.

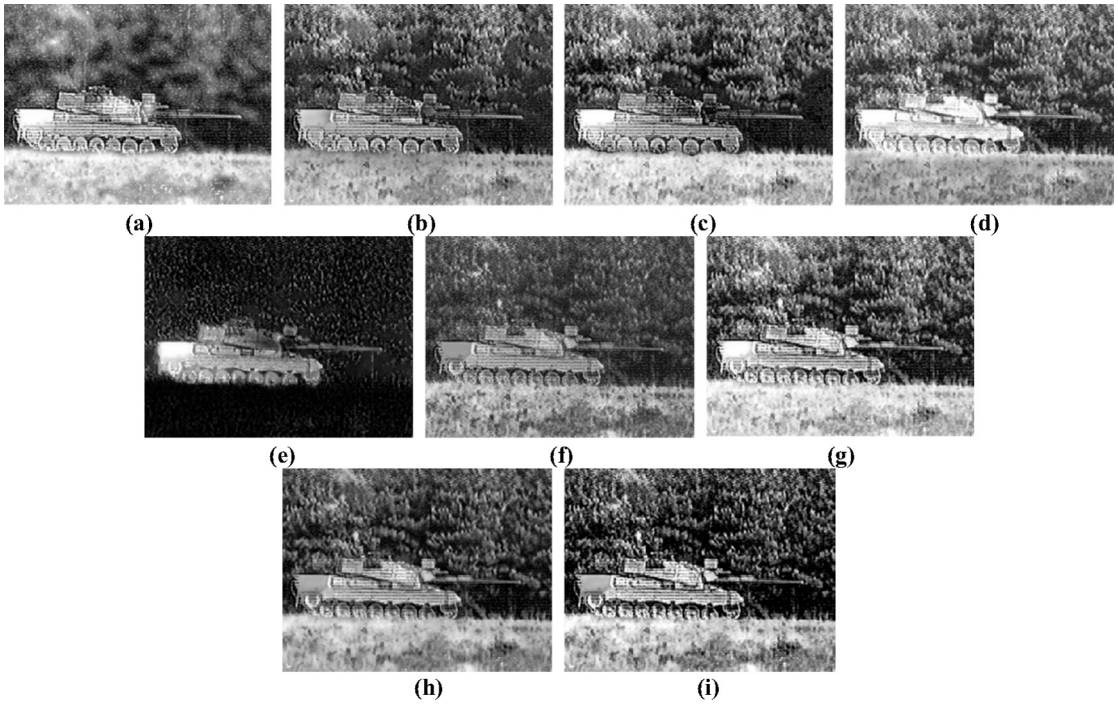


Fig. 10. Third group of fusion experiments.

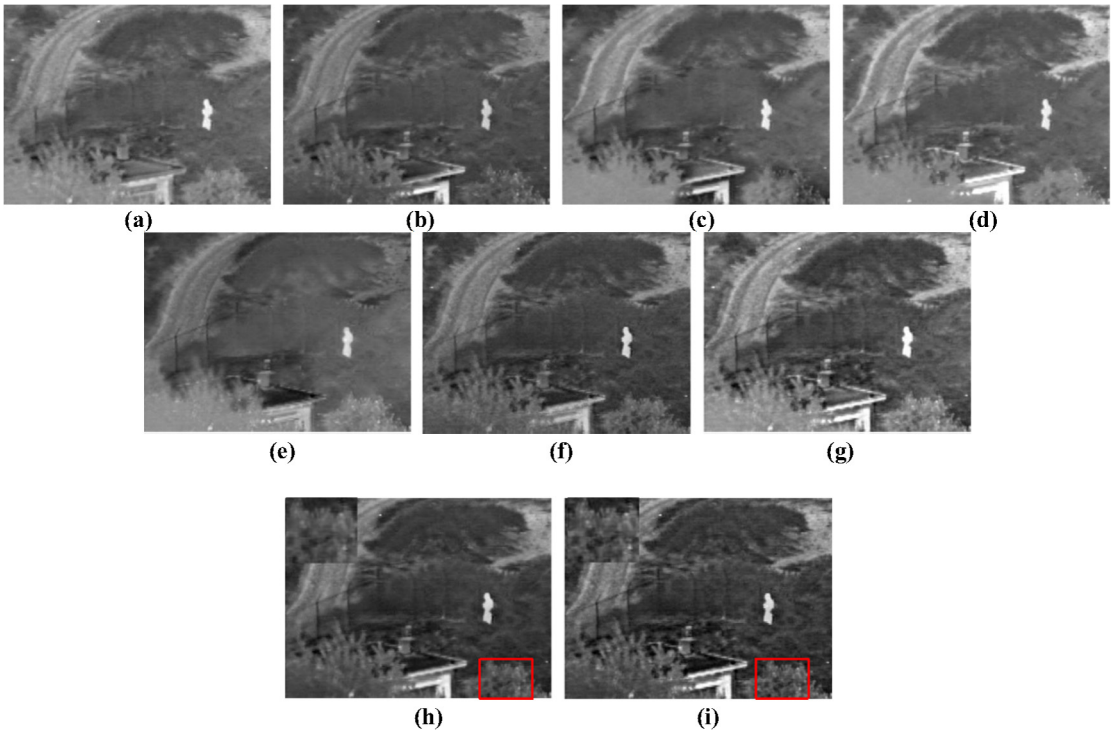


Fig. 11. Fourth group of fusion experiments.

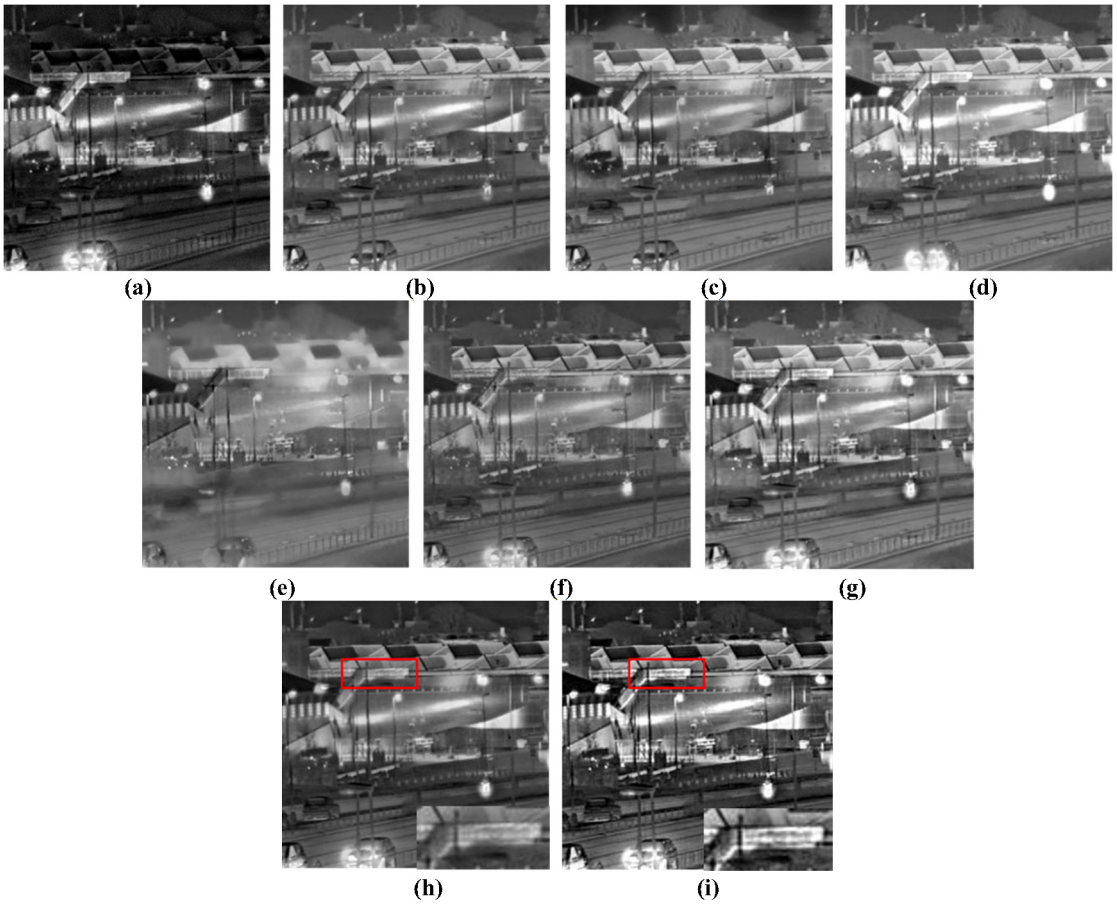


Fig. 12. Fifth group of fusion experiments.

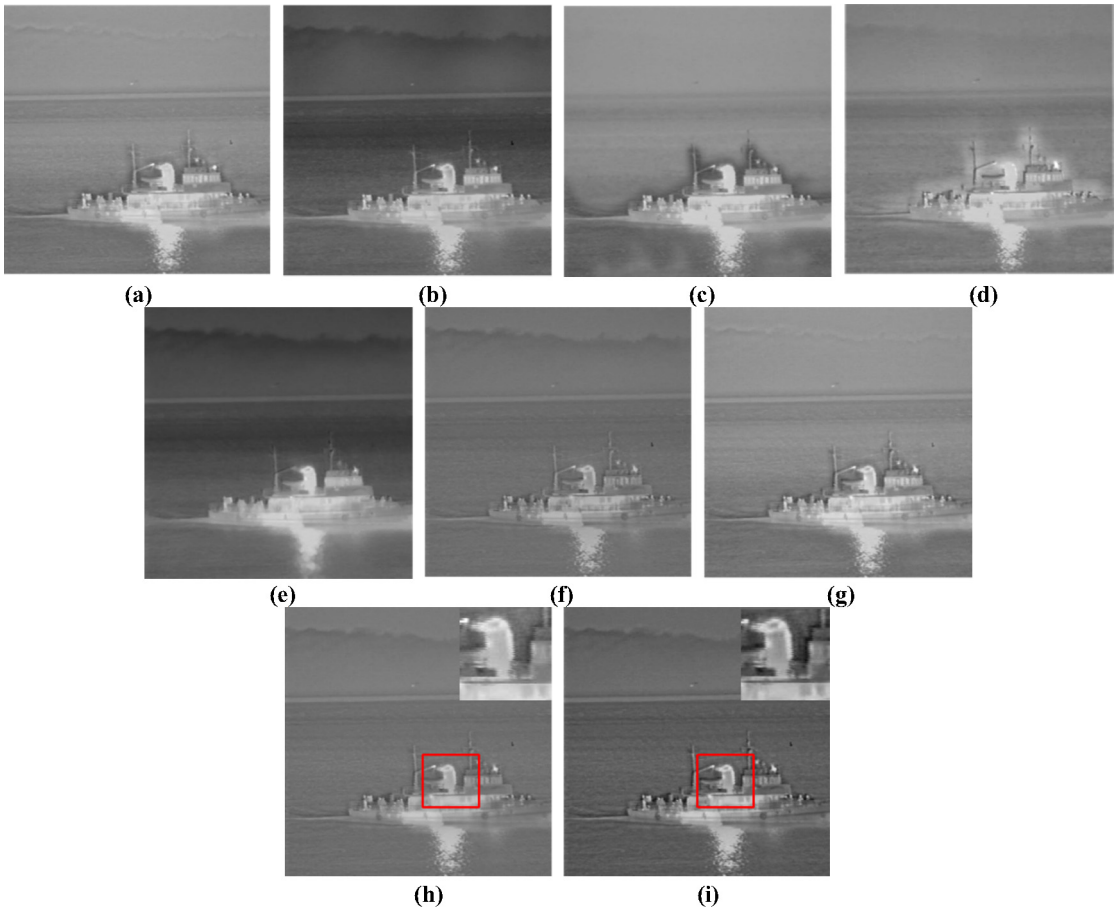


Fig. 13. Sixth group of fusion experiments.

and the water wave texture on the sea surface is not as good as that shown in Fig. 13(i). Regarding the seventh group of experiments, the appearance of Fig. 14(g) is similar to that of the IR image. Fig. 14(d) has the disadvantage that the transition of the gray value is not uniform in the background. Fig. 14(e) is too dark, and the contrast is low. Fig. 14(c) has white pseudo-noise along the coastline. Regarding the eighth group of experiments, Fig. 15(d) and Fig. 15(g) are close to the IR images, and Fig. 15(c) and Fig. 15(f) are close to the VI images. Therefore, none of the four abovementioned fusion algorithms fully combines the characteristics of the two source images. The appearance of Fig. 15(e) is close to that of the proposed algorithm, but the details of the texture are as good as those shown in Fig. 15(i). Compared with those yielded by the abovementioned algorithms, the fused images generated by our algorithm have the highest contrast ratio and can fuse the IR target and background information in the human visual field. Furthermore, the proposed algorithm does not lose edge detail information in the fusion process; thus, its performance is the best.

5.2. Objective evaluation

Image fusion results are usually evaluated subjectively and objectively. Fusion results often present limited differences. Correctly and subjectively evaluating fusion results can be difficult. Thus, the fusion effect is frequently determined via objective quality evaluations. The following five objective quality indexes are selected as the evaluation criteria: (1) AVG [29], (2) information entropy (IE) [30], (3) edge retentiveness ($Q^{AB/F}$) [31], and (4) standard deviation (SD) [32].

The results of detailed quantitative evaluation of the five groups of IR and VI images are shown in Tables 1–3. The values in bold font represent the best results.

The AVG and IE values of the proposed algorithm are higher than those of the other algorithms for each group of experiments. Thus, the textural features and edge details of the proposed algorithm are well restored, and the content of the fused image is abundant. In the first group of experiments, the parameters of the proposed algorithm are all higher than those of the other algorithms, which show that the fusion effect of this algorithm is the best and agrees with actual perception. In the second experiment, the $Q^{AB/F}$ value of the proposed algorithm is slightly lower than that of the CNN algorithm, whereas the values of other parameters are high. In the third experiment, the $Q^{AB/F}$ value is also slightly lower than that of the IFE-VIP algorithm. In the fourth, sixth, and eighth experiments, the values of the evaluation parameters of the algorithm are higher than those of other algorithms, in good agreement with the previous subjective evaluation. In the fifth and seventh experiments, the SD of the proposed algorithm is lower than that of the NSST-PCNN because the NSST-PCNN algorithm produces many black artifacts in the fusion. These artifacts lead to an abnormal increase in the gray value. Thus, the SD is unrealistic. Based on the objective evaluation parameters, the final result is the same as that of the subjective evaluation. The proposed algorithm is superior to the other algorithms in terms of the image gray value distribution, edge detail, and clarity.

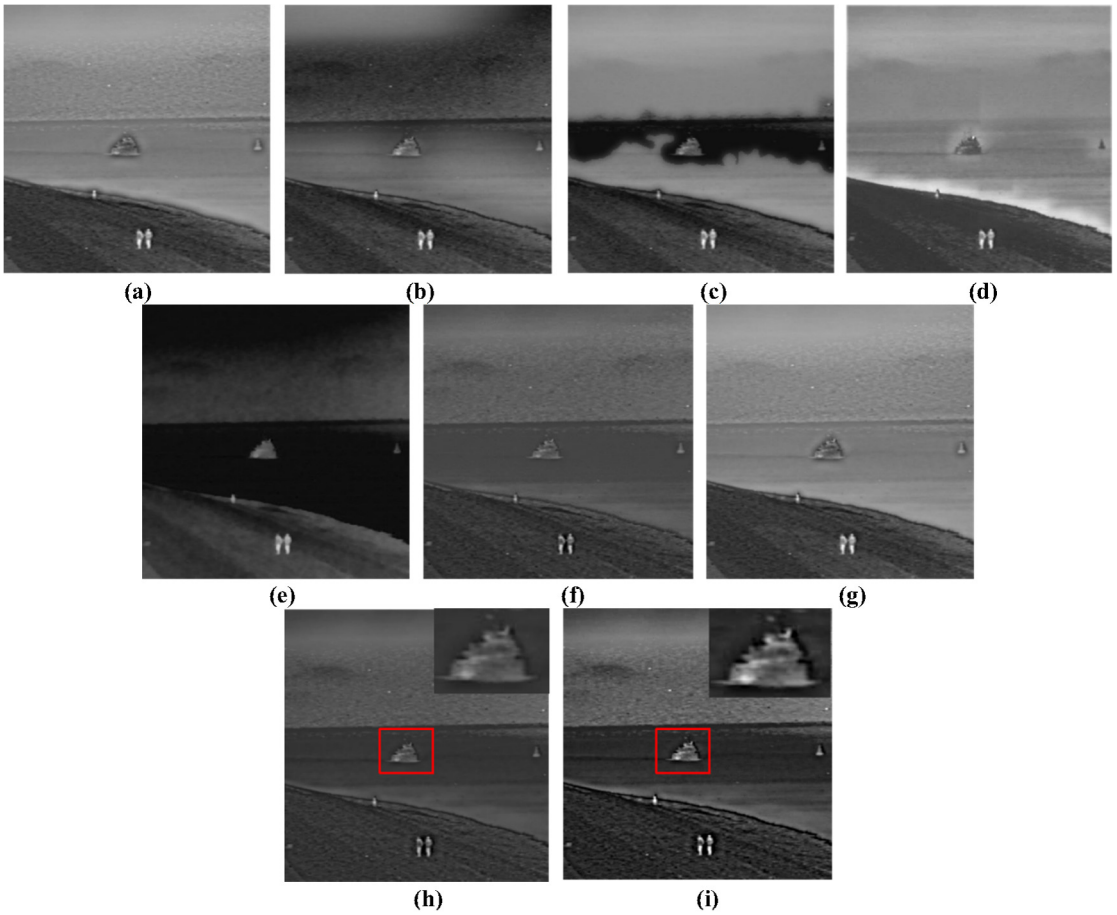


Fig. 14. Seventh group of fusion experiments.

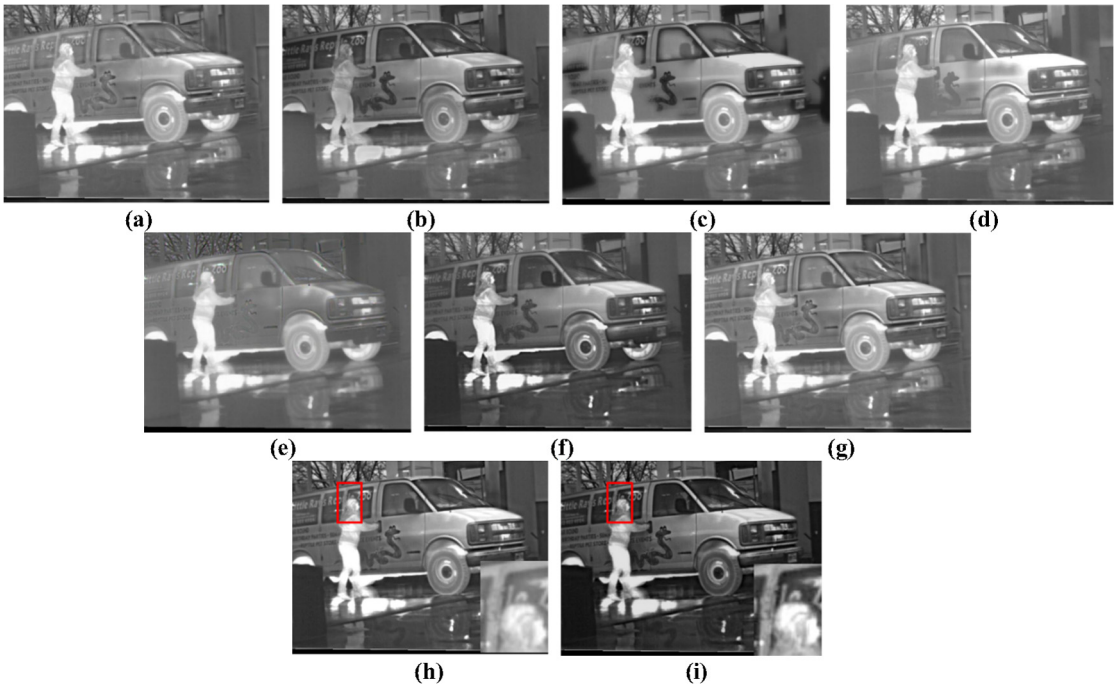


Fig. 15. Eighth group of fusion experiments.

Table 1
Objective evaluation results of the first three groups of fused images.

Group	Fusion methods	Evaluation index			
		AVG	IE	$Q^{AB/F}$	SD
1	NSST-PCNN	3.3474	7.2022	0.5686	6.9689
	GFF	2.8348	7.1193	0.5831	6.1894
	CNN	3.3754	7.1753	0.5005	7.0172
	IFE-VIP	3.2863	7.1574	0.5659	6.8684
	GTF	3.4752	7.1932	0.4952	7.0242
	FT	4.5490	7.1456	0.4832	7.0065
	LNSST-PCNN	3.8907	7.1456	0.4755	6.9543
	NSST-ID-PCNN	5.0531	7.3145	0.5987	8.3477
	Proposed	**5.1420**	**7.3522**	**0.6193**	**8.5655**
2	NSST-PCNN	5.1420	6.6522	0.5193	8.5655
	GFF	5.6612	6.8422	0.5514	9.7923
	CNN	5.7914	6.9653	**0.5598**	10.2403
	IFE-VIP	6.7687	6.7255	0.4969	11.6890
	GTF	5.9627	6.6752	0.4216	11.5228
	FT	7.0831	6.9876	0.5314	12.7657
	LNSST-PCNN	5.4731	6.7580	0.5076	9.8901
	NSST-ID-PCNN	7.3214	7.0145	0.5231	12.1156
	Proposed	**7.3764**	**7.1396**	0.5484	**13.2142**
3	NSST-PCNN	9.3850	7.2819	0.5355	41.1616
	GFF	9.5599	7.2261	0.5605	40.8538
	CNN	9.5616	7.2835	0.6018	44.2468
	IFE-VIP	9.9535	7.3017	**0.6188**	42.1068
	GTF	9.4752	7.1932	0.4952	41.9325
	FT	9.7890	7.1145	0.5781	43.2310
	LNSST-PCNN	9.3670	7.1765	0.5123	42.1234
	NSST-ID-PCNN	9.9671	7.2090	0.5965	44.1783
	Proposed	**10.0521**	**7.3159**	0.6081	**46.1198**

Table 2
Objective evaluation results of the middle three groups of fused images.

Group	Fusion methods	Evaluation index			
		AVG	IE	$Q^{AB/F}$	SD
4	NSST-PCNN	4.3398	6.9941	0.4344	39.4377
	GFF	4.1216	6.8889	0.6072	34.3422
	CNN	4.8633	7.1839	0.5074	51.8397
	IFE-VIP	4.2545	7.1547	0.5324	49.3352
	GTF	4.3243	6.8742	0.4634	50.8255
	FT	4.8467	7.0981	0.5891	52.8045
	LNSST-PCNN	4.5671	7.1145	0.5435	50.8931
	NSST-ID-PCNN	4.8901	7.3045	0.6076	55.4367
	Proposed	**4.9107**	**7.3250**	**0.6268**	**57.2323**
5	NSST-PCNN	15.6851	7.8985	0.4039	**64.0592**
	GFF	16.6480	7.7788	0.4803	60.0111
	CNN	17.0325	7.9565	0.5435	62.8997
	IFE-VIP	16.4775	7.7707	0.4761	60.5765
	GTF	15.5789	7.5675	0.4189	60.7524
	FT	16.8760	7.8876	0.5281	60.7340
	LNSST-PCNN	15.9231	7.8765	0.5087	60.6541
	NSST-ID-PCNN	16.8901	7.9760	0.5323	61.1090
	Proposed	**17.0062**	**8.0003**	**0.5521**	62.7661
6	NSST-PCNN	2.7707	5.9209	0.5654	19.8795
	GFF	2.4447	5.4152	0.5568	20.4683
	CNN	1.9172	5.9175	0.5704	19.4387
	IFE-VIP	2.4104	5.7929	0.4641	19.8898
	GTF	2.0647	5.4326	0.5089	17.7524
	FT	3.1300	5.4133	0.5699	16.9300
	LNSST-PCNN	2.0432	5.3456	0.5255	17.4567
	NSST-ID-PCNN	2.8789	5.9345	0.6022	21.4590
	Proposed	**2.9062**	**6.0003**	**0.6221**	**22.7661**

6. Conclusion

This paper presents an adaptive fusion framework for IR and VI images based on saliency detection and the improved dual-channel PCNN in the LNSST domain. Our method uses the LNSST as a multi-scale decomposition tool for images based on the NSST. To improve the effect of fused images, an improved algorithm based on FT saliency extraction is adopted to guide the adaptive weighted fusion of low-pass sub-images, and an ID-PCNN model is used as the fusion rule for high-pass sub-images. Five different scenarios are used to verify fusion performance. The results show that our algorithm can fuse visible light and IR images effectively with high con-

Table 3
Objective evaluation results of the last two groups of fused images.

Group	Fusion methods	Evaluation index			
		AVG	IE	$Q^{AB/F}$	SD
7	NSST-PCNN	5.0978	6.2414	0.4732	**25.0885**
	GFF	4.8765	6.1547	0.4948	19.3435
	CNN	4.8633	6.0839	0.4574	18.8397
	IFE-VIP	5.2898	6.4205	0.4446	21.6111
	GTF	5.2327	6.3752	0.4216	21.8340
	FT	5.9468	6.3167	0.4981	21.2330
	LNSST-PCNN	4.5671	6.2145	0.5235	18.8931
	NSST-ID-PCNN	5.8901	6.4035	0.5076	55.4367
	Proposed	**6.0764**	**6.5396**	**0.5584**	22.6234
8	NSST-PCNN	4.4945	7.0940	0.5900	45.7099
	GFF	4.3337	7.2006	0.5991	40.9522
	CNN	4.0325	7.0565	0.5435	42.8997
	IFE-VIP	4.2179	7.0110	0.6361	48.5090
	GTF	4.2327	7.0752	0.4216	42.8340
	FT	4.8720	7.0651	0.6034	48.4107
	LNSST-PCNN	4.5231	7.0765	0.6087	40.6541
	NSST-ID-PCNN	4.8901	7.1760	0.6323	48.1090
	Proposed	**4.9976**	**7.2736**	**0.6483**	**49.3437**

trast while retaining a wealth of textural and detail information without any artifacts. The experimental results of the subjective and objective evaluation indicate that our algorithm exhibits better fusion performance than typical fusion techniques do.

Conflict of interest

No conflict of interest.

Acknowledgements

The authors would like to thank the anonymous reviewers and editors for their invaluable suggestions. The paper is jointly supported by the National High-tech R&D Program of China (NO. 863-2-5-1-13B).

References

[1] Jufeng Zhao, Guangmang Cui, Xiaoli Gong, Yue Zang, Shuyin Tao, Daodang Wang, Fusion of visible and infrared images using global entropy and gradient constrained regularization, Infrared Phys. Technol. 81 (2017) 201–209.
[2] Xiangzhi Bai, Morphological center operator based infrared and visible image fusion through correlation coefficient, Infrared Phys. Technol. 76 (2016) 546–554.
[3] Jiayi Ma, Chen Chen, Chang Li, Jun Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, Inf. Fusion 31 (2016) 100–109.
[4] Xiangzhi Bai, Fugen Zhou, Bindang Xue, Infrared image enhancement through contrast enhancement by using multiscale new top-hat transform, Infrared Phys. Technol. 54 (2011) 61–69.
[5] Xiangzhi Bai, Xiaowu Chen, Fugen Zhou, Zhaoying Liu, Bindang Xue, Multiscale top-hat selection transform based infrared and visual image fusion with emphasis on extracting regions of interest, Infrared Phys. Technol. 60 (2013) 81–93.
[6] Xiangzhi Bai, Gu. Shuhang, Fugen Zhou, Bindang Xue, Weighted image fusion based on multi-scale top-hat transform: algorithms and a comparison study, Optik 124 (2013) 1660–1668.
[7] Pan Zhu, Xiaoqing Ma, Zhanhua Huang, Fusion of infrared-visible images using improved multi-scale top-hat transform and suitable fusion rules, Infrared Phys. Technol. 81 (2017) 282–295.
[8] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (2013) 2864–2875.
[9] Yu Liu, Shuping Liu, Zengfu Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inf. Fusion 24 (2015) 147–164.
[10] Yu Liu, Xun Chen, Hu Peng, Zengfu Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (2017) 191–207.
[11] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
[12] A.L. Da Cunha, Zhou Jianping, M.N. Do, The nonsubsampled contourlet transform: theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.
[13] I. Selesnick, R. Baraniuk, N. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Process. Mag. 22 (6) (2005) 123–151.
[14] G. Easley, D. Labate, W. Lim, Sparse directional image representations using the discrete shearlet transform, Appl. Comput. Harmon. Anal. 25 (1) (2008) 25–46.
[15] W.Q. Lim, The discrete shearlet transform: a new directional image representation and compactly supported shearlet frames, IEEE Trans. Image Process. 19 (5) (2010) 1166–1180.
[16] Jufeng Zhao, Qiang Zhou, Yueting Chen, Huajun Feng, Xu. Zhihai, Qi Li, Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition, Infrared Phys. Technol. 56 (2013) 93–99.
[17] R. Achanta, S. Hemami, F. Estrada, et al., Frequency-tuned Salient Region Detection, Computer Vision and Pattern Recognition, 2009, in: Proceedings of the IEEE Conference on CVPR 2009, 2009, pp. 1597–1604.
[18] Fu. Zhizhong, Xue Wang, Xu. Jin, Ning Zhou, Yufei Zhao, Infrared and visible images fusion based on RPCA and NSCT, Infrared Phys. Technol. 77 (2016) 114–123.
[19] Zhanwen Liu, Yan Feng, Yifan Zhang, Xu Li, A fusion algorithm for infrared and visible images based on RDU-PCNN and ICA-bases in NSST domain. Infrared Phys. Technol. 79 (2016) 183–190.
[20] Xu. Xinzheng, Dong Shan, Guanying Wang, Xiangying Jiang, Multimodal medical image fusion using PCNN optimized by the QPSO algorithm, Appl. Soft Comput. 46 (2016) 588–595.
[21] Ashraf K. Helmy, Gh. S. El-Taweel, Image segmentation scheme based on SOM–PCNN in frequency domain, Appl. Soft Comput. 40 (2016) 405–415.
[22] Liu Qiong, Ma Yide, A new algorithm for noise reducing of image based on PCNN time matrix, Electron. Inform. Technol. 8 (2008) 1869–1873.
[23] Jungong Han, Eric J. Pauwels, Paul de Zeeuw, Fast saliency-aware multi-modality image fusion, Neurocomputing 111 (2013) 70–80.
[24] Jinlei Ma, Zhiqiang Zhou, Bo Wang, Hu.a. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, Infrared Phys. Technol. 82 (2017) 8–17.
[25] Tianzhu Xiang, Li Yan, Rongrong Gao, A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain, Infrared Phys. Technol. 69 (2015) 53–61.
[26] Weiwei Kong, Longjun Zhang, Yang Lei, Novel fusion method for visible light and infrared images based on NSST–SF–PCNN, Infrared Phys. Technol. 65 (2014) 103–112.
[27] Yu Zhang, Lijia Zhang, Xiangzhi Bai, Li Zhang, Infrared and visual image fusion through infrared feature extraction and visual information preservation, Infrared Phys. Technol. 83 (2017) 227–237.
[28] Guangmang Cui, Huajun Feng, Xu Zhihai, Qi Li, Yueting Chen, Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition, Opt. Commun. 341 (2015) 199–209.
[29] Xiangzhi Bai, Infrared and visual image fusion through feature extraction by morphological sequential toggle operator, Infrared Phys. Technol. 71 (2015) 77–86.
[30] Yong Ma, Jun Chen, Chen Chen, Fan Fan, Jiayi Ma, Infrared and visible image fusion using total variation model, Neurocomputing 202 (2016) 12–19.
[31] Xiangzhi Bai, Yu Zhang, Fugen Zhou, Bindang Xue, Quadtree-based multi-focus image fusion using a weighted focus-measure. Information Fusion 22 (2015) 105–118. Infrared Physics & Technology 54 (2011) 61–69.
[32] Ming Yin, Puhong Duan, Wei Liu, Xiangyu Liang, A novel infrared and visible image fusion algorithm based on shift-invariant dual-tree complex shearlet transform and sparse representation, Neurocomputing 226 (2017) 182–191.