CrossMark

Infrared and visible images fusion using visual saliency and optimized spiking cortical model in non-subsampled shearlet transform domain

Ruichao Hou[1] · Rencan Nie[1,2] · Dongming Zhou[1] ·
Jinde Cao[2] · Dong Liu[1]

Abstract Aiming at some problems in existing infrared and visible image fusion methods such as edge blurring, low contrast, loss of details, a novel fusion scheme based on non-subsampled shearlet transform (NSST), visual saliency and multi-objective artificial bee colony (MOABC) optimizing spiking cortical mode (SCM) is proposed. NSST has many advantages such as multi-scale features and sparse representation. Moreover, the visual saliency map can improve the low frequency fusion strategy, and SCM has coupling and pulse synchronization properties. Firstly, NSST is utilized to decompose the source image into a low-frequency subband and a series of high-frequency subbands. Secondly, the low-frequency subband is fused by SCM, where SCM is motivated by the edge saliency map of the low-frequency subband of the source image, and then the high-frequency subbands are also fused by SCM, where the modified spatial frequency of the high-frequency subbands of the source image is adopted as the input stimulus of SCM, the parameters of SCM are optimized by the novel multi-objective artificial bee colony technique. Finally, the fused image is reconstructed by inverse NSST. Experimental results indicate that the proposed scheme performs well and has obvious superiorities over other current typical ones in both subjective visual performance and objective criteria.

Keywords Image fusion · Non-subsampled shearlet transform · Visual saliency map · Multi-objective artificial bee colony · Spiking cortical model

✉ Dongming Zhou
zhoudm@ynu.edu.cn

[1] Information College, Yunnan University, Kunming 650504, China

[2] Department of Mathematics, Southeast University, Nanjing 210096, China

🖄 Springer

1 Introduction

Image fusion plays an essential part in many applications such as computer vision, satellite cloud images, medical images, target detection, military, remote sensing and so on [12, 22]. The fusion of visible and infrared images is a significant research focus in the image fusion field. The infrared image (IR) consists of the thermal radiation characteristic, and can capture the target hidden under low-light conditions and recognize the objects in the camouflage. Although infrared imaging sensor is not affected by the various lightings or bad weather conditions, the obtained image lacks adequate background details. On the contrary, the visible light image is obtained by the spectral reflection of the object, and usually contains more textures and detail information of background, and also has higher spatial resolution, thus the visible image has a better visual quality than the infrared image [37]. Image fusion technique is to extract meaningful information from multiple images under the same scene, or different kinds of image sensors under diverse modes. The composite image synthesizes the advantages of the visible and infrared images and highlights localization of the target in the infrared image.

Currently, multi-scale geometric transform methods applied to image fusion have been studied extensively. Among the tools of multi-scale geometric transform, such as discrete wavelet transform (DWT) [20], Laplacian pyramid (LAP) [31], contourlet transform (CT) [5]. In order to lead to better frequency selectivity and regularity than CT, and remove pseudo-Gibbs phenomena along the edges to some extent, non-subsampled contourlet transform (NSCT) was proposed by Da Cunha et al. [2]. In comparison with other decomposition methods, NSCT requires a larger amount of computation. To reduce the computational complexity of NSCT, non-subsampled shearlet transform (NSST) was proposed by Easley et al. [7], NSST has the shift-invariance of non-subsampled processes and inherits the perfect properties from shearlet and wavelet, such as the characteristics of anisotropy, computing speed. Therefore, NSST has an advantage in obtaining more information for image fusion.

In addition, artificial neural network has become a research hotspot [24–27]. Pulse couple neuron network (PCNN) is a new generation of artificial neural network, which was developed by Johnson et al. [14], and owns some superior characters, such as coupling and pulse synchronization. It has been widely applied in image segmentation, image enhancement, pattern recognition, and so on [41]. Xin Jin et al. [11] proposed an images fusion based on NSST and PCNN. However, PCNN has a large number of parameters which are always set as constants by human experience leading to the lack of universality. In order to solve these problems, a modified neural network model called spiking cortical model (SCM) was proposed by Kong et al. [16], which devised a novel scheme based on SCM and NSST, and overcome the shortcoming of parameters setting and utilized the intensity distribution of pixels to optimize the iterative number. Meanwhile a large number of intelligent algorithms had been applied to parameters optimization of the neural network, such as genetic algorithm-PCNN (GA-PCNN) [43], particle swarm optimization-PCNN (PSO-PCNN) [13], and artificial bee colony-PCNN (ABC-PCNN) [3]. Commonly, these single objective optimization algorithms have only one fitness function, which ignore the influence of other factors so those algorithms do not achieve the best result in image fusion field.

Recently, the vision saliency detection and super resolution methods are also widely used for image processing [30, 38]. Jinlei Ma et al. [23] used visual saliency map to fuse the base

layers. Zhang et al. [44] presented a novel fusion method based on NSST and the visual saliency, although the performance was improved by the visual saliency map, the treating processes of background information was too simple so that the details were lost.

To alleviate the aforementioned problems and obtain better fusion performance, a novel image fusion scheme using the visual saliency detection and optimized SCM in NSST domain is proposed. At the beginning, NSST decomposes the source images into a low-frequency subband and a series of high-frequency subbands. Then the visual saliency map of low-frequency subband and the modified spatial frequency of high-frequency subbands are utilized to act as the SCM external stimulation, respectively. For the sake of overcoming the disadvantage of single objective optimization, we present to optimize parameters of SCM with multi-objective artificial bee colony algorithm, besides the iteration times is set by the time matrix. Finally, the fused image is obtained via optimization process. Experimental results show the proposed method does well in the fusion of infrared and visible image and can preserve not only the spectral information of the visible image but also the thermal target information of the infrared image, thus the fused result contains high contrast and rich background details.

The remaining sections of this paper are summarized as follows. Section 2 presents an overview of the proposed fusion scheme and reviews the theory of related algorithms. Section 3 describes the image fusion strategies and steps in detail. Experimental results and discussions are given in Section 4. Some conclusions are summarized in Section 5.

2 The proposed fusion scheme

Figure 1 sketches out the main scheme of the proposed fusion method. Firstly, the infrared image and the visible image are decomposed into a low-frequency subband
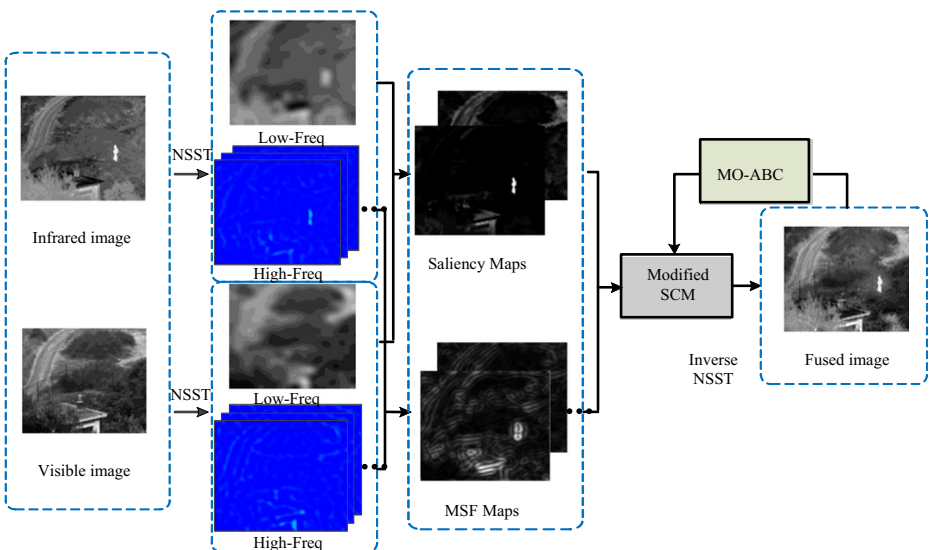

Fig. 1 Schematic diagram of the proposed image fusion framework

and a series of high-frequency subbands using NSST, respectively. Then, the modified frequency-tuned algorithm is used to extract the saliency map as external stimulation of SCM in the low-frequency subband, in the meanwhile, the modified spatial frequency (MSF) of the high-frequency subbands is used to stimulate the SCM. Next, the novel multi-objective artificial bee colony technique is utilized to optimize the parameters of SCM according to suitable fitness functions. Finally, the fused image is gained by taking inverse NSST.

2.1 Non-subsampled shearlet transform

NSST, which was proposed by Easley [7], is an extension of the wavelet in multidimensional space and combines the non-subsampled pyramid (NSP) filter with shearlet transform to provide the multiscale decomposition. The shearlet transform (ST) is close to optimal sparse representation, the synthetic expansion of affine system is described as follows:

$$\Lambda_{AB}(\psi) = \left\{ \psi_{j,l,k}(x) = |\det A|^{j/2} \psi(B^l A^j x - k) : j, l \in Z, k \in Z^2 \right\}, \tag{1}$$

where $\psi_{j,\,l,\,k}$ is expressed as a composite wavelet, A denotes the anisotropy matrix for multi-scale decomposition, B is a shear matrix for directional analysis, j, l and k are scale, the direction of decomposition and shift parameter, respectively. When $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, the composite wavelet becomes shearlet, the structure of the frequency tiling by the shearlet is shown in Fig. 2.

The NSST decomposition is divided into two major steps: (I) Multi-scale decomposition. $(k + 1)$ subbands as same size as the source image can be obtained by using the k-class non-subsampled pyramid filter, including a low-frequency map and a series of high-frequency maps; (II) The direction of localization. In pseudo polarization grid coordinates, standard shearlet is calculated by Meyer window function, which requires the subsampled operation to obtain the shift-invariance. However, NSST direction of localization uses the modified shearlet filter, which can map from the pseudo polarization to
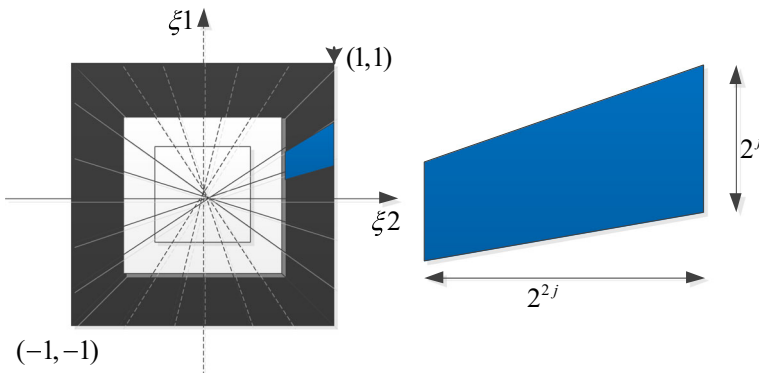

Fig. 2 The structure of the frequency tiling by the shearlet

the Cartesian coordinate system avoid the next sampling operation via Fourier inverse transform, so NSST has the characteristic of the shift-invariance.

2.2 Saliency detection of infrared image

Achanta et al. [29] introduced a frequency-tuned (FT) approach to estimate center-surround contrast using color and luminance features. This approach obtained the saliency map S for an image I of width W and height H pixels thus could be formulated as follows

$$S(x,y) = \left\| I_\mu - I_{\omega hc}(x,y) \right\|, \tag{2}$$

where I_μ is the arithmetic mean pixel value of the image, $I_{\omega hc}(x,y)$ is the pixel value of the source image in the Gaussian blurred version (5×5 separable binomial kernel), and ‖‖ is defined as Euclidean distance.

Guided filter was proposed by He et al. [8], which is a linear shift filter. The filtering output at a pixel i is expressed as a weighted average:

$$q_i = \sum_i W_{ij}(I)p_j, \tag{3}$$

where i and j are pixel indexes, W_{ij} is the filter kernel, I is guidance image, p is a filtering input image and q is an output image. The guidance image I is set according to different applications and can be taken as input image p directly.

The filter kernel weights are expressed by

$$W_{ij}(I) = \frac{1}{|\omega|^2} \sum_{k:(i,j)\in\omega_k} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \varepsilon} \right), \tag{4}$$

where $|\omega|$ is the number of pixels in the window, ω_k is the window of k kernel function, μ_k and σ_k^2 are the mean and variance of the guidance image I respectively, ε denotes the smoothing factor.

The conventional FT algorithm utilized a Gaussian blurry filter to process the input image. However, the guided filter kernel used the pixel mean and variance of the neighborhood as local estimation, and could adjust the output weight adaptively based on the content of the image, which had superior competence in retaining the edge information and performance of detail enhancement, so this paper makes an improvement on the FT approach by using the guided filter.

$$S(x,y) = \left\| I_\mu - I_G(x,y) \right\|, \tag{5}$$

where $I_G(x,y)$ is the guided filter output of the input image, the guidance image I is the same as input image p.

In the comparative study of well-known saliency detection methods, such as Itti model [19], saliency using natural statistics (SUN) [42], spectral residual approach (SR) [10]. Our modified method has the advantages in extracting target information of the infrared image, and keeping the edge details, and suppressing the background information of infrared image fully at the same time, as shown in Figs. 3 and 4. The X-axis and the Y-axis represent the position of the pixel, and the Z-axis represents the value of gray-scale in the three-dimensional diagram of gray-scale image.
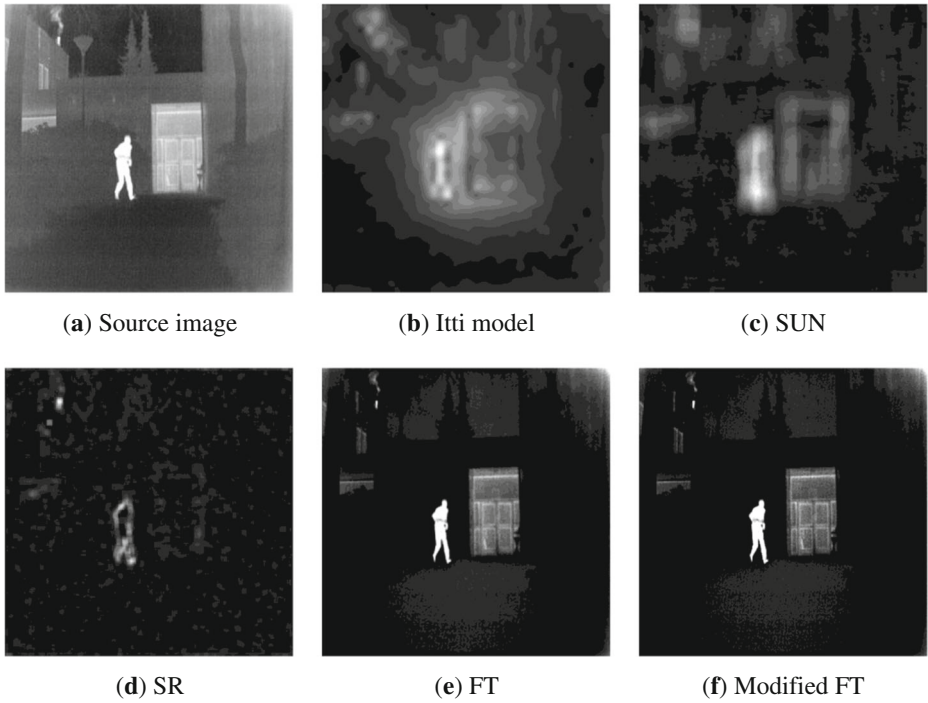
(**a**) Source image (**b**) Itti model (**c**) SUN

(**d**) SR (**e**) FT (**f**) Modified FT

Fig. 3 Saliency detection of the infrared image

2.3 Multi-objective ABC algorithm

Artificial bee colony (ABC) is a group intelligence optimization algorithm initially proposed by Karaboga [15] through imitating bee feeding behavior which uses various kinds of bees with a different division of labor to share information during the search process.
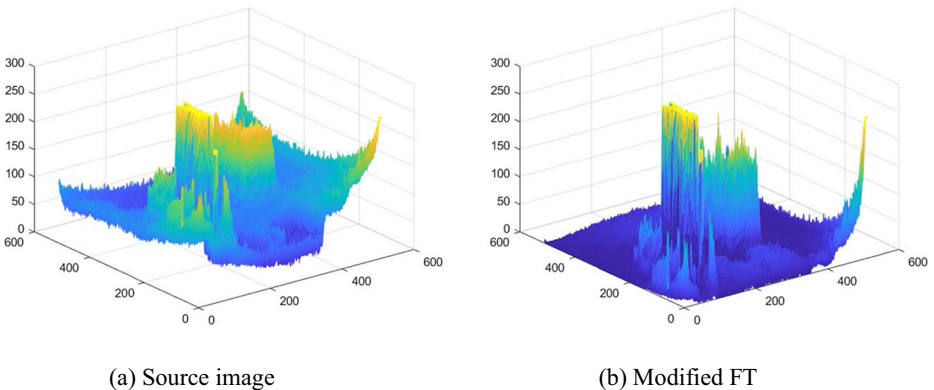

(a) Source image (b) Modified FT

Fig. 4 Three-dimensional diagram of the gray-scale image

The ABC algorithm consists of three groups: employed bees, onlooker bees, scout bees. Each nectar position represents a possible solution, and the degree of income of the nectar corresponds to the fitness of the solution.

First of all, the ABC algorithm generates initial populations randomly, N denotes the number of bees, and also the number of the nectar. At the initial time of algorithm flow, all bees are set to scout bees.

Secondly, each solution consists of D dimensional vectors, where D denotes the number of the neural network parameters which need to be optimized. And then the nectar position also expresses the solution of the corresponding problem, which is searched using the iteration of the three kinds of bees, employed bees search for and calculate the income of the new location, which also is known as the fitness of the practical problem in the neighborhood based on the local information in the memory. According to the greedy rule, if the income of the new location is high enough, we should record the new location to replace the original location.

Finally, the obtained information should pass to the waiting onlooker bees by unique dance like the shape of '8', while the search process is finished by the employed bees. Then onlooker bees start to search and choose a better solution by analyzing the obtained information, the rule is: the higher the probability of fitness is, the greater the probability of the choice of nectar position is.

The probability formula is as follow

$$p_i = \frac{fit_i}{\sum\limits_{n=1}^{N} fit_n}, \tag{6}$$

in which, fit_i denotes the value of the fitness function of the ith solution, n represents the number of the nectar or the number of employed bees.

The ith employed bees and onlooker bees search for new nectar position formula

$$V_{ij} = X_{ij} + \varphi_{ij}(X_{ij} - X_{kj}), \tag{7}$$

where $k \in \{1, 2 \cdots, N\}, j \in \{1, 2 \cdots, D\}, k \neq i, \varphi_{ij} \in rand(-1, 1)$ is used to limit the honey of the nectar location X_{ij}.

Equation (7) shows the situation, namely, the smaller the difference between X_{kj} and X_{ij} is, and the smaller the disturbance is. The optimal solution is achieved by the search area, and can shorten step size adaptively. Thus this algorithm has the advantage of adaptive convergence.

If the fitness still cannot be improved during a certain number of cycles, and the source nectar will be discarded. Scout bees will go to search by generating the random new nectar position.

Pareto domination is one of the effective methods for judging individual merits in low-dimensional multi-objective optimization [1, 28, 34]. Based on the concepts of Pareto non-inferior ranking and crowding distance in multi-objective evolutionary algorithm, we present the MOABC algorithm, the pseudo-code of algorithm flow is shown in Table 1, where N is the number of the employed bees; MCN is the maximum number of iterations; $Limit$ is the number of honey source update times and 'archive' represents the external population.

Table 1 The pseudo-code of MOABC algorithm

Framework MOABC
1: **Input**: *N, MCN, Limit*
Begin
2: Initialization and update the archive
3: **While** (iteration<=*MCN*) **do**
4: Send employed bees using above searching strategy
5: Make an optional Pareto local search
6: **if** (employed bees having no preferences) **then**
7: Form new nectar sources according to Eq. (6) Eq. (7)
8: **endif**
9: Evaluate the quality of the new nectar sources according to the fitness function
10: **if** (employed bees having no preferences) **then**
11: Send onlookers
12: **endif**
13: **if**(the fitness still can't be improved and nectar source will be discarded)**then**
14: Send Scout bees
15:**endif**
16: iteration++
17: **end while**
18: **Output** archive
End

2.4 Spiking cortical model

SCM was presented by Zhan K et al. [40], has the simple structure and fewer parameters, as shown in Fig. 5. It consists of multiple neurons, and each neuron contains three main function units: receptive field, modulation field, and pulse generator. Moreover, it does not need to learn or train, and can extract the useful information from the complex background. The mathematical expressions of the model are as follows

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

$$U_{ij}(n) = fU_{ij}(n-1) + S_{ij} \sum_{kl} W_{kl} Y_{kl}(n-1), \tag{9}$$

$$E_{ij}(n) = gE_{ij}(n-1) + V_\theta Y_{ij}(n-1), \tag{10}$$

$$X_{ij}(n) = \frac{1}{1 + e^{(E_{ij} - U_{ij})}}, \tag{11}$$

$$Y_{ij}(n) = \begin{cases} 1, & if \ X_{ij}(n) > 0.5 \\ 0, & otherwise \end{cases}, \tag{12}$$

where *n* denotes the iteration times, (i, j) is the location of the image pixel, $F_{ij}(n)$ describes the feedback input signal of the neuron, $S_{ij}(n)$ is the input excitation signal, $U_{ij}(n)$ refers to the
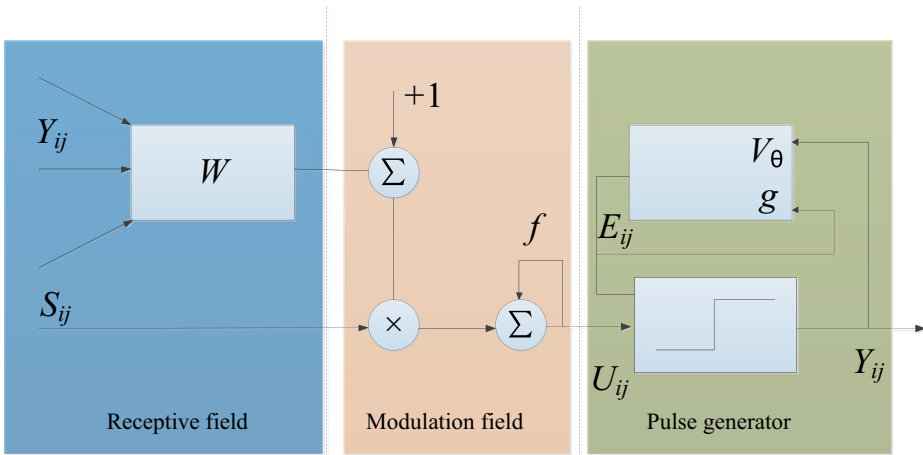
Fig. 5 The structure of the basic SCM neuron

internal active state of the neuron, W_{kl} is the weighted coefficient matrix of linking between neurons, $E_{ij}(n)$ is the dynamic threshold, V_θ is the threshold of amplification factor, $Y_{ij}(n)$ is the output signal of the neuron at nth iteration, f and g are the internal active and dynamic threshold signal decay coefficients, respectively (Fig. 5).

In order to show the difference within ignition range, the sigmoid function is used to improve the neuron output signal [39], as shown in (11), $X_{ij}(n)$ denotes the pixel pulse ignition output amplitude, as $X_{ij}(n) > 0.5$, the neuron produces a pulse, which is called one firing time, the signal is captured by the linking matrix W_{kl}, and the adjacent neurons achieve synchronization pulse release at the spatial position. $T_{ij}(n)$ expresses the neuron firing times matrix after nth iteration, the formula is described as follows

$$T_{ij}(n) = T_{ij}(n-1) + Y_{ij}(n). \tag{13}$$

2.5 Multi-objective artificial bee colony optimization SCM

Commonly, the quality of image fusion need to be evaluated by using various evaluation metrics comprehensively. However, these single objective optimization algorithms were presented by Jin Xin et al. [13] and Banharnsakun A [3], and have only one fitness function so ignore the influence of other factors in the image fusion field. To achieve better fused results, we introduce the multi-objective optimization algorithm.

The main task is to optimize the parameters of SCM, namely, it is equivalent to finding the optimal solution set of the two-dimensional equation and the bees corresponding to SCM's parameters f and g.

It is key point to select suitable fitness function, so we introduce several alternative objective evaluation metrics as the hybrid fitness function of MOABC algorithm. Those objective evaluation metrics include mutual information (MI) [9], mean structural similarity ($MSSIM$) [33], standard deviation (SD) [11], spatial frequency (SF) [11], image entropy (IE) [11] and margin information retention ($Q^{AB/F}$) [32].

1) *MI* shows the correlation between two events, the *MI* of *U* and *V* can be defined as follows:

$$MI(U, V) = \sum_{v \in V} \sum_{u \in U} p(u, v) \log_2 \frac{p(u, v)}{p(u)p(v)}, \tag{14}$$

where $p(u,v)$ is the joint probability distribution of *U* and *V*, $p(u)$ and $p(v)$ are the marginal probability distribution of *U* and *V*, respectively. The sum of mutual information between the fused image and two source images can be calculated to denote the difference of fusion quality, and then the mutual information metric can be described as follows:

$$MI(A, B, F) = MI(A, F) + MI(B, F), \tag{15}$$

Eq. (15) reflects a total amount of information that fused image $F(i, j)$ contains about source image $A(i, j)$ and source image $B(i, j)$. The larger value of mutual information metric indicates that the fused image contains the more information and achieves the better the fusion effect.

2) *SD* is a measure of the dispersion degree of a set of image data averages. The standard deviation of an image is calculated as.

$$SD = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (F(i, j) - \mu)^2}, \tag{16}$$

where $F(i, j)$ is the pixel value of the fused image at the location (i, j), and μ is the mean value.

3) *SF* is composed of row frequency (RF) and column frequency (CF), and is described as follows

$$SF = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (RF + CF), \tag{17}$$

in which *M* is the row of the image, *N* is column of the image.

4) *IE* represents the amount of information in the fused image. It can be acquired by (18)

$$IE = - \sum_{l=0}^{L} P(l) \log_2 P(l), \tag{18}$$

where $P(l)$ expresses the probability density of *L*.

5) *MSSIM* is an effective measure of similarity of two images, which is calculated as follows

$$MSSIM = \frac{SSIM(A, F) + SSIM(B, F)}{2}, \tag{19}$$

where *SSIM(A, F)* and *SSIM(B, F)* are correlation coefficients between infrared image and fused image, visible image and fused image respectively. *SSIM (i, j)* is defined as follows

$$SSIM(i, j) = \frac{\left(2\mu_i \mu_j + C_1\right)\left(2\sigma_{ij} + C_2\right)}{\left(\mu_i^2 + \mu_j^2 + C_1\right)\left(\sigma_i^2 + \sigma_j^2 + C_2\right)}, \tag{20}$$

where μ_i, σ_j and σ_{ij} express the mean, standard deviation, and cross-correlation, respectively. *C1* and *C2* are used to ensure stability when the mean value and the variance are close to zero.

The rotationally symmetric Gaussian window with standard deviation 1.5 was selected in MSSIM.

6) $Q^{AB/F}$ represents the transformation degree of edge information of the fused image and the source image. It is defined as follows

$$Q^{AB/F} = \frac{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{M} \left(Q^{AF}(i,j) w^A(i,j) + Q^{BF}(i,j) w^B(i,j) \right)}{\sum_i^N \sum_j^M \left(w^A(i,j) + w^B(i,j) \right)}, \tag{21}$$

where $Q^{AF}(i,j) = Q_g^{AF}(i,j) Q_o^{AF}(i,j)$, $Q_g^{AF}(i,j)$ and $Q_o^{AF}(i,j)$ are the edge strength and orientation preservation value at the location (i,j), respectively. N and M are the size of the image, and $Q^{BF}(i,j)$ is similar to $Q^{AF}(i,j)$, $w^A(i,j)$ and $w^B(i,j)$ reflect the weight of $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$ respectively.

According to the principle, namely, the larger the value of objective evaluation metrics are, the better the performance of the fusion method is [13, 36], so we adopt two multi-criteria fitness functions, as shown as follows

$$fitness_1 = \max(MI + SD + IE), \tag{22}$$

$$fitness_2 = \max\left(Q^{AB/F} \right). \tag{23}$$

3 Fusion strategies and specific steps

3.1 Low-frequency subband fusion strategy

Commonly the low-frequency information is the main components of the source images. On the contrary, the high-frequency information contains the details of the image [7]. Most of the low-frequency coefficients are fused by the simple weighted averaging or maximum based strategies [35], which do not consider the relationship between pixels. In order to have a better fusion effect, we proposed a novel method that the improved edge saliency map is used as external excitation of SCM. We define edge saliency map as *Map*, which is described as follows:

$$Map_{IR}(i,j) = \max[S_{IR}(i,j), E_1(i,j)], \tag{24}$$

$$Map_V(i,j) = \max[S_V(i,j), E_2(i,j)], \tag{25}$$

$$E_1(i,j) = (L_{IR} * F)(i,j), \tag{26}$$

$$E_2(i,j) = (L_V * F)(i,j), \tag{27}$$

$$F = \begin{bmatrix} 0.2 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix}, \tag{28}$$

where $*$ denote convolution, $L_{IR}(i,j)$, $L_V(i,j)$ are the low-frequency coefficients. $E(i,j)$ represents the filtered image with convolution kernel F. $S_{IR}(i,j)$ and $S_V(i,j)$ represent the visual saliency map of the source images, which can be calculated using (5).

3.2 High frequency subband fusion strategy

The existing high-frequency fusion strategies contain the largest absolute value, regional energy, variance and gradient [7], but these strategies cannot extract detail information from the image adequately while only considering the individual pixels or regional characteristics. The gray value of a single pixel is used as the excitation of the neural network, this may lose image edges and texture features. Kong W et al. [17] introduced the modified spatial frequency which increases the gradient calculation of the diagonal direction, it can be utilized to extract more information in the infrared image sets.

Suppose $H(i,j)$ denotes the high-frequency coefficient at the location (i,j), and MSF is measured using slipping windows (the size is 3×3) of the coefficient, then MSF in each subband is used to motivate the neuron, and it is defined as follows:

$$MSF = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (RF + CF + MDF + SDF)^{1/2}, \tag{29}$$

$$RF = [H(i,j)-H(i,j-1)]^2, \tag{30}$$

$$CF = [H(i,j)-H(i-1,j)]^2, \tag{31}$$

$$MDF = [H(i,j)-H(i-1,j-1)]^2, \tag{32}$$

$$SDF = [H(i,j)-H(i-1,j+1)]^2, \tag{33}$$

where RF, CF, MDF, SDF denote the frequencies at rows, columns, main diagonal and auxiliary diagonal, respectively. N and M are the size of the slipping window.

3.3 Specific image fusion steps

Assume that the infrared and visible images have been matched and treated with uniform size accurately. The steps of the image fusion algorithm based on SCM as follows.

Step 1   Decompose the infrared and visible images using NSST to obtain their low-frequency subbands $\{L_{IR}^K, L_V^K\}$ and a series of high-frequency subbands $\{H_{IR}^{l,k}, H_V^{l,k}\}$ at each K-scale and l-direction, where $1 \leq k \leq K$.

Step 2   SCM is utilized to deal with the low-frequency subbands. Let the edge saliency maps
           be the feedback inputs of SCM.

(a)   Calculate the Map_{IR} and Map_V according to (24) and (25), and all coefficients are
        normalized.
(b)   Set the initial values as follows: $U_{ij}(0) = T_{ij}(0) = E_{ij}(0) = 0$. In the initial state, all the
        neurons are inactivated, so $Y_{ij}(0) = 0$.
(c)   Calculate $U_{ij}(n)$, $E_{ij}(n)$, $Y_{ij}(n)$ by (9), (10) and (12), respectively, and then compute the
        neuron's firing times $T_{ij}(n)$ according to (13). The fusion coefficients are selected
        according to $T_{ij}(n)$, N is the maximum number of iterations, and the rule is described as:

$$L_F^K(i,j) = \begin{cases} L_{IR}^K(i,j), & T_{ij}^{IR}(N) \geq T_{ij}^V(N) \\ L_V^K(i,j), & T_{ij}^{IR}(N) < T_{ij}^V(N) \end{cases}. \tag{34}$$

Step 3   Measure the MSF as the external excitation of SCM using (29). Referring to step 2,
           use SCM to fuse the high-frequency subbands $\{H_{IR}^{l,k}, H_V^{l,k}\}$. The fused coefficients
           can be determined as follows:

$$H_F^{l,k}(i,j) = \begin{cases} H_{IR}^{l,k}(i,j), & T_{ij}^{IR}(N) \geq T_{ij}^V(N) \\ H_V^{l,k}(i,j), & T_{ij}^{IR}(N) < T_{ij}^V(N) \end{cases}. \tag{35}$$

Step 4   Optimize the parameters of SCM using multi-objective artificial bee colony algo-
           rithm. First of all, initialize the bee populations and set maximum number of
           iterations. Then, find the optimal solution set according to the two fitness functions,
           as shown in (22) and (23). Finally, select the optimization solution based on the
           selection principle.
Step 5   Take the optimal parameters to set SCM and perform inverse NSST of the low-
           frequency and the high-frequency coefficients to obtain the fused image.

4 Experimental results and analysis

The simulation experiments were conducted by MATLAB2014a software on PC with Intel E5
2670 2.6 GHz, 16 GB RAM. We take several groups of accurate matching of infrared image
and visible light image to test. All of them cover 256 or 512 Gy levels. The source infrared and
visible images were collected from http://www.imagefusion.org/ and https://figshare.
com/articles/TNO_Image_Fusion_Dataset/1008029.

4.1 Experiment parameters setting

According to Ref. [1], we initialize the bee populations as follows: feasible solutions number is
2, the sum of bees is 20 (the number of employed bees and onlooker bees is 10 respectively),
the largest number of search limit is set to 10, the maximum number of iterations is 50. The 2-
D initial random values are $f \in [0, 1]$ and $g \in [0, 1]$.

At the same time, so as to show the optimization effect of this method, the un-optimized
SCM fusion method is used to compare and analyze. The high frequency coefficient adopts the
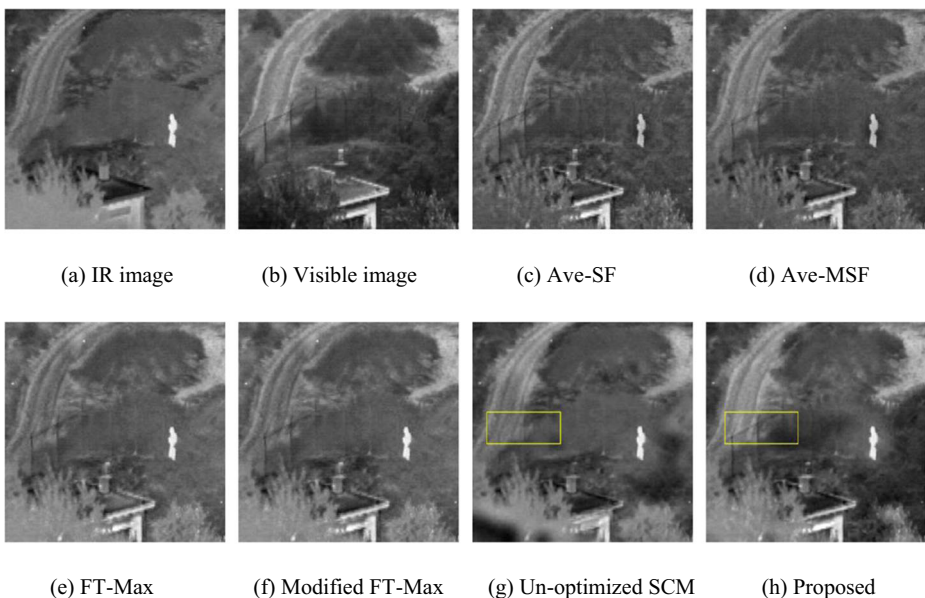
Table 2 The optimal solution sets of *UN Camp*

Number	f	g
1	0.2045	0.2395
2	0.1996	0.6288
3	0.0704	0.4101
4	0.1959	0.4116
5	0.2169	0.5088
6	0.2209	0.5805
7	0.0903	0.5456
8	0.2054	0.5275
9	0.3403	0.5216
10	0.8728	0.4592

modified spatial frequency as the fusion strategy, the low-frequency coefficient fusion strategy selects the saliency map of the image. According to Ref. [8] and the parameters of the conventional SCM are set as follows:

$$f = 0.2, \ g = 0.6, \ V_\theta = 20, \ W = \begin{bmatrix} 0.1091 & 0.1409 & 0.1091 \\ 0.1409 & 0 & 0.1409 \\ 0.1091 & 0.1409 & 0.1091 \end{bmatrix}, \ n = 20, \ V_\theta = 20.$$

In addition, the parameters of *f, g,* and iteration *n* are set by the optimized SCM adaptively and the remaining parameters are the same as that of the conventional SCM in our method. In our implementation, the proposed fusion method is compared with three representative conventional fusion methods and two state-of-the-art fusion methods, such as wavelet-based method (DWT) [20], Laplacian pyramid (LAP) [31], multiscale transform-based method (NSST-SCM) [18], multiscale transform and sparse representation (MST-SR) [21], guide the

(a) IR image   (b) Visible image   (c) Ave-SF   (d) Ave-MSF

(e) FT-Max   (f) Modified FT-Max   (g) Un-optimized SCM   (h) Proposed

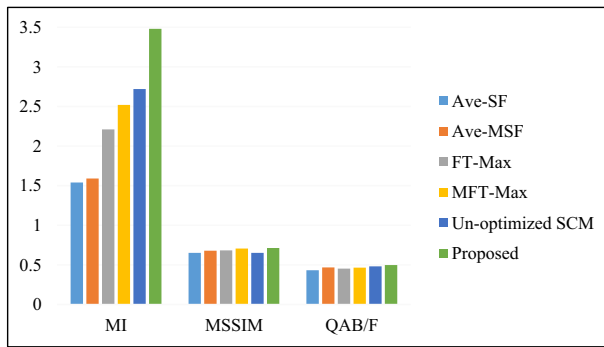**Fig. 6** Comparison of different fusion strategies

Fig. 7 The chart of evaluation indexes

filtering-based method (GFF) [6]. The 'db2' wavelet adopts discrete wavelet decomposition; NSST uses a non-subsampling pyramid 'maxflat' filter and its decomposition directions are set as [12, 20, 37].

4.2 Parameters optimization

In order to verify the rationality of the parameters and the fusion strategies in our proposed method, several experiments were conducted on the image sets, we selected "*UN Camp*" for specific analysis. At the beginning, four groups of different fusion strategies are compared, Group 1: the low-frequency subbands are fused by a simple weighted averaging strategy, and
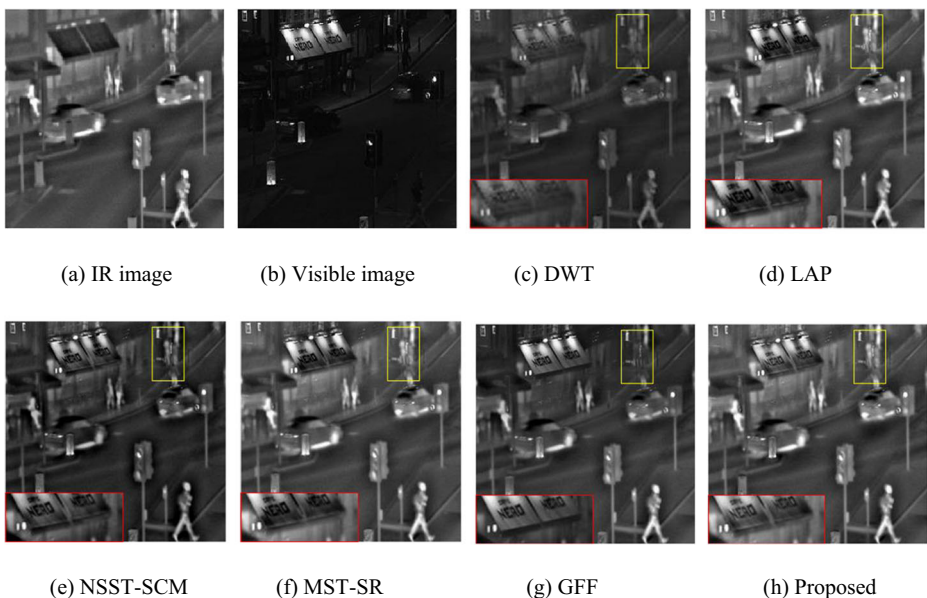

(a) IR image   (b) Visible image   (c) DWT   (d) LAP

(e) NSST-SCM   (f) MST-SR   (g) GFF   (h) Proposed

Fig. 8 Image fusion results of "Bristol Queen's Road"

(a) IR image   (b) Visible image   (c) DWT   (d) LAP

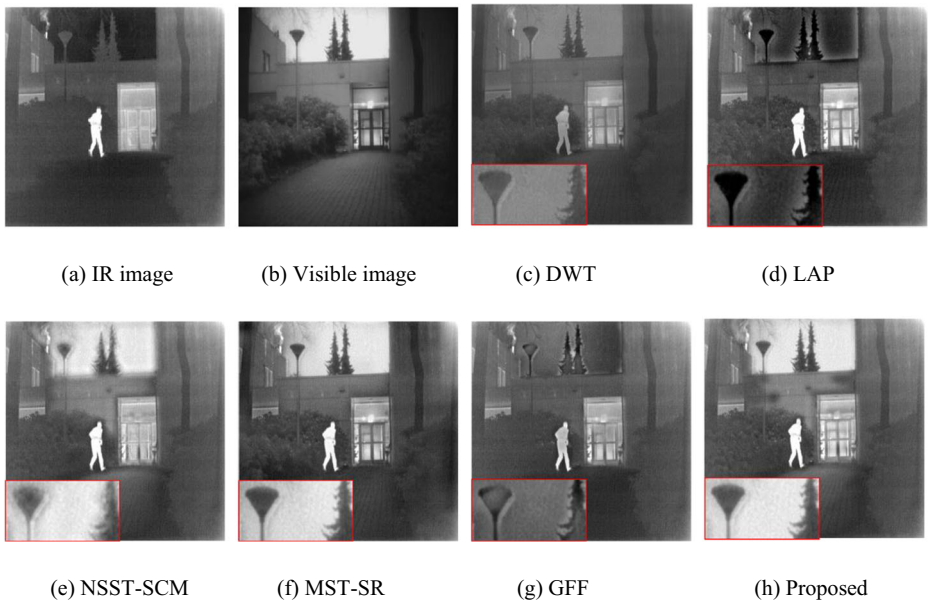(e) NSST-SCM   (f) MST-SR   (g) GFF   (h) Proposed

Fig. 9 Image fusion results of "Kaptein"

the high-frequency subbands adopt SF as external excitation of SCM. Group 2: the low-frequency subbands are also fused by a simple weighted averaging strategy, and the high-frequency subbands adopt MSF as external excitation of SCM. Group 3: the saliency map is used as the external stimulus of SCM in the low-frequency subbands, and the high- frequency subbands are fused by the largest absolute value strategy. Group 4:
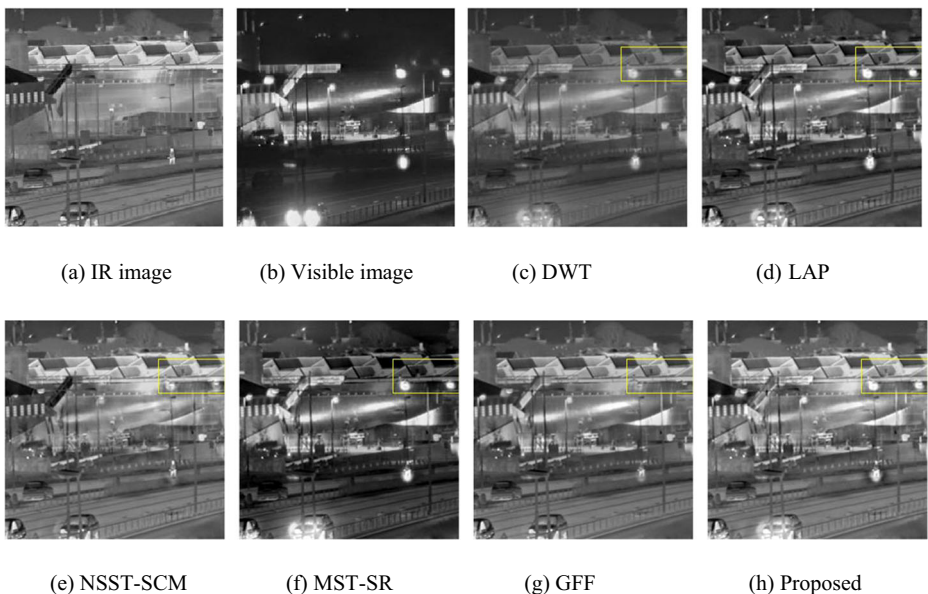

(a) IR image   (b) Visible image   (c) DWT   (d) LAP

(e) NSST-SCM   (f) MST-SR   (g) GFF   (h) Proposed

Fig. 10 Image fusion results of "Street"

(a) IR image   (b) Visible image   (c) DWT   (d) LAP

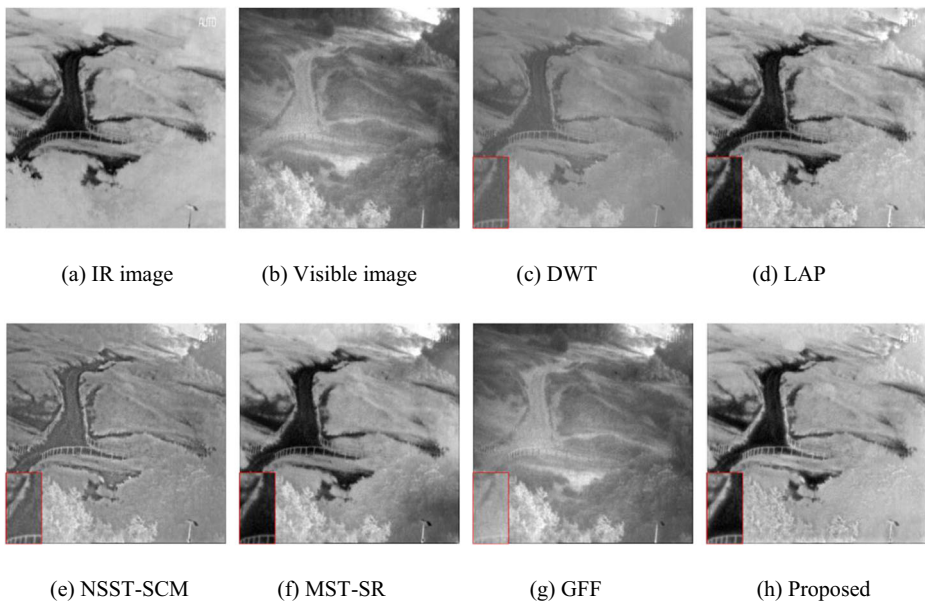(e) NSST-SCM   (f) MST-SR   (g) GFF   (h) Proposed

Fig. 11 Image fusion results of "Heather"

the low frequency subbands are fused using modified saliency map as external excitation of SCM, and the high-frequency subbands are fused by the largest absolute value strategy. Next, the fifth experiment uses the conventional SCM, so the number of iterations cannot be optimized. The last is our fusion method, the ten sets of solutions about SCM parameters were listed in Table 2.

However, it is difficult to select which set of optimal solutions to be the final parameters of SCM, so we introduce the concept of the best compromise solution [4]. Generally, $Q^{AB/F}$ can better reflect the object edge information in the fused image, so the solution of the maximum value of this index is $f = 0.2209$ and $g = 0.5805$, which is selected as the final solution, the selection of parameters is realized adaptively based on this criterion.

It can be seen from Fig. 6 that Fig. 6c to f correspond to four groups of comparative experiments. Figure 6g and h show the fused results of the un-optimized SCM and the proposed method, respectively. First of all, in terms of visual effects, these methods take a simple weighted averaging strategy, and have bad fusion effect, the modified FT method can preserve more edge information than original FT method, such as the details of eaves in the fused image. Obviously, the un-optimized SCM lacks the details of the fence in the regions marked by the yellow rectangle. Then we utilize objective evaluation indexes [9, 32, 33] to measure the fused results, and the data show that the modified strategies improve MI and $Q^{AB/F}$ to a certain extent, as shown in Fig. 7.

4.3 Subjective evaluations

The fused results based on the different methods above are illustrated in Fig. 8, 9, 10, and 11, and the red rectangle and yellow rectangle region represent the enlarged details

of the region and the contrast region, respectively. For the case of Fig. 8, the key point of image fusion is to fuse the information of pedestrians and vehicles into the final image fully, and maintain environmental information as much as possible. In terms of visual effects, although these algorithms can fuse most information of the source images, both DWT method and GFF method have lost infrared character details, as shown in the yellow rectangular marked region. It can be seen that our method can well retain the details of the tarpaulins in the red rectangular marked area because of the uniform gray-scale distribution. However, the fusion effects of NSST-SCM and MST-SR methods are inferior to our method. The next set of the image is "Kaptein", as shown in Fig. 9. Among DWT, LAP and GFF methods do not to fuse the sky area properly due to contaminated with the dark IR spectral information, while the sky in our result is brighter and less noisy. Moreover, as we can see in marked region, the edge of the street lamp and the trees have some shadow in the result by NSST-SCM method. Both MST-SR and the proposed method can achieve good visual effects compared with other methods.

Figure 10 shows a scene that contains multiple targets and complex source of lights, which is similar to Fig. 8. Compared with the proposed method, both the bulb luminance and the contrast of the fused results obtained by NSST-SCM and GFF methods are a little lower, and the details of two lamps in the upper right are not fused; the background scenery in MST-SR result contains more IR noise, and the contrast of the DWT result is low. From Fig. 11, we can easily find out that the result based on GFF looks like the visible image which lost the infrared information, whereas the result of the proposed method contains more details of the natural scene and obvious target.

In summary, the proposed method is superior to other methods in both inheriting the characteristics of the source images and preserving background details on the visual level.

4.4 Quantitative comparison

Another essential evaluation criterion is quantitative comparison, so the image fusion effect is measured by some above objective evaluation indexes [9, 11, 32, 33]. From Table 3, 4, 5, and 6 report the objective evaluation results based on six methods. Moreover, the two important relative evaluation metrics that MI and $Q^{AB/F}$ will be represented as graph intuitively, as shown in Fig. 12. It can be seen that these two indexes are superior to other methods, this indicates the fused image generated by the proposed method contains more significant information from the source image, and the details of the two source images are reflected more accurately, the remaining metrics are slightly better than other comparison methods, this proves that the image fusion quality of the proposed method is better objectively.

Table 3 Quantitative results of experiment on"Bristol Queen's Road"

Methods	MI	EI	SF	SD	MSSIM	$Q^{AB/F}$
DWT	1.7641	5.9502	10.8103	26.8452	0.6084	0.4105
LAP	2.7743	6.7614	17.6038	36.0527	0.6004	0.6719
NSST-SCM	2.7080	6.4549	17.3505	36.1998	0.5901	0.6407
MST-SR	3.5980	6.7462	17.6871	35.6520	0.5994	0.6750
GFF	1.8704	6.5639	16.6055	32.7254	0.6052	0.6666
Proposed	**3.6116**	**6.8114**	**17.6940**	**36.6063**	**0.6209**	**0.6783**

Table 4 Quantitative results of experiment on "Kaptein"

Methods	MI	EI	SF	SD	MSSIM	$Q^{AB/F}$
DWT	1.9752	6.5779	7.5664	31.6990	0.7036	0.3605
LAP	2.4530	7.0261	**9.9516**	40.4013	0.7164	0.4939
NSST-SCM	2.3740	7.3433	9.7017	51.4140	0.7041	0.4763
MST-SR	2.5068	7.1490	9.6775	**54.5026**	0.7239	0.5279
GFF	2.5989	6.9828	8.6265	34.1017	0.7125	0.5224
Proposed	**2.6214**	**7.4392**	9.6068	54.0522	**0.7253**	**0.5332**

Table 5 Quantitative results of experiment on "Street"

| Methods	MI	EI	SF	SD	MSSIM	$Q^{AB/F}$
DWT	2.7036	6.9454	12.1065	34.7115	0.6037	0.4612
LAP	2.4962	7.2680	**22.1301**	42.9834	0.6129	0.6192
NSST-SCM	2.7553	7.2680	19.8159	40.3688	0.6008	0.5630
MST-SR	2.6138	7.3779	22.0925	40.7620	0.5733	0.6326
GFF	2.7105	7.2098	20.1805	40.5091	**0.6155**	0.6248
Proposed	**3.1125**	**7.4204**	21.4458	**43.5274**	0.6043	**0.6335**

In conclusion, our proposed method retains the effective information of the source images and plays a significant role in the fusion of infrared and visible images.

5 Conclusions

In this paper, a novel infrared and visible light image fusion scheme is proposed, in which visual saliency map improves the low-frequency fusion strategy, the spatial frequency is utilized as the external incentive of SCM in NSST domain. Among them, the soft limiting function improves the output of SCM; at the meanwhile, the parameters of SCM are further optimized by multi-objective the artificial bee colony. Compared with other methods, the experimental results show that the modified SCM structure is simple which has fewer parameters to set, low computational costs, and objective is outstanding in the fused image, the outline is clear, rich background details in the fused image, the

Table 6 Quantitative results of experiment on "Heather"

Methods	MI	EI	SF	SD	MSSIM	$Q^{AB/F}$
DWT	1.7641	6.6894	10.8103	26.8452	0.6094	0.3447
LAP	2.4129	7.2744	13.8266	48.9163	0.6158	0.4721
NSST-SCM	2.2577	7.1917	13.2201	47.1087	0.6021	0.4588
MST-SR	2.3631	7.4123	13.3848	49.1882	0.6072	**0.5257**
GFF	2.5389	7.3006	12.4872	39.9741	0.6038	0.4701
Proposed	**2.7359**	**7.5715**	**13.8749**	**49.2002**	**0.6203**	0.5138
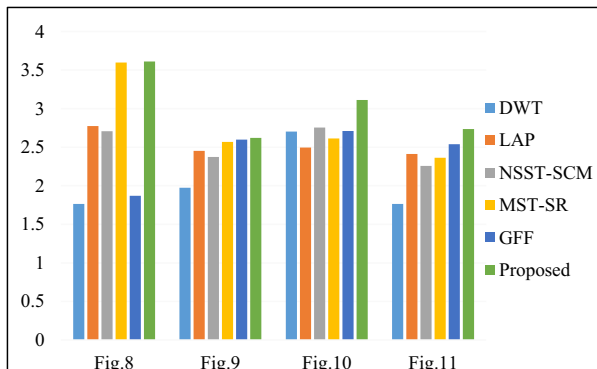
(a) The comparison bar chart for MI
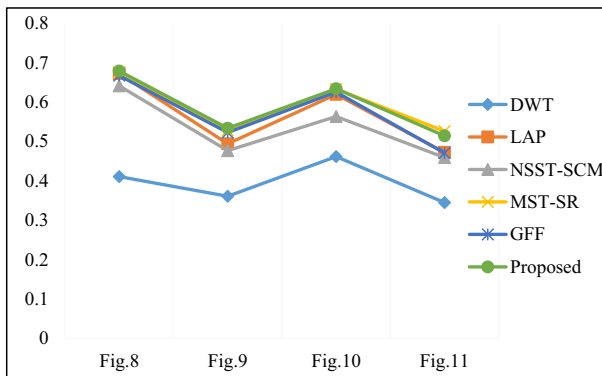
(b) The comparison line chart for $Q^{AB/F}$

Fig. 12 Objective evaluation results based on Figs. 8, 9, 10 and 11

fusion performance is better than the other state-of-the-art methods both the subjective and objective evaluation. Our next research goal is to use parallel computing to reduce computational costs and extend the application domain of this scheme.

**Compliance with ethical standards**

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this paper.

References

1. Akbari R, Hedayatzadeh R, Ziarati K, Hassanizadeh B (2012) A multi-objective artificial bee colony algorithm. Swarm Evol Comput 2(1):39–52
2. Al Da C, Zhou J, Do MN (2006) The nonsubsampled contourlet transform: theory, design, and applications. IEEE Trans Image Process 15(10):3089–3101
3. Banharnsakun A Multi-focus image fusion using best-so-far ABC strategies. Neural Comput & Applic:1–16, https://doi.org/10.1007/s00521-015-2061-2(2015)

4. Current John R, Revelle CS, Cohon JL (1990) Interactive approach to identify the best compromise solution for two objective shortest path problems. Comput Oper Res 17(2):187–198

5. Do Minh N, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans Image Process 14(12):2091–2106

6. Gan W, Wu X, Wu W, Yang X, Ren C, He X, Liu K (2015) Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter. Infrared Phys Technol 72:37–51

7. Glenn E, Labate D, Lim WQ (2008) Sparse directional image representations using the discrete shearlet transform. Appl Comput Harmon Anal 25(1):25–46

8. He K, Sun J, Tang X (2013) Guided image filtering. IEEE Trans Pattern Anal Mach Intell 35(6):1397–1409

9. Hossny M, Nahavandi S, Creighton D (2008) Comments on 'information measure for performance of image fusion. Electron Lett 44(18):1066–1067

10. Hou X, Zhang L. (2007) Saliency detection: a spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, Computer Vision and Pattern Recognition, CVPR '07, DOI: https://doi.org/10.1109/CVPR.2007.383267, 1–8: 2280

11. Jin X, Nie RC, Zhou DM et al (2016) Multifocus color image fusion based on NSST and PCNN. J Sens 8359602. https://doi.org/10.1155/2016/8359602

12. Jin X, Jiang Q, Yao SW et al (2017) A survey of infrared and visual image fusion methods. Infrared Phys Technol 85:478–501

13. Jin X, Zhou DM, Yao SW et al (2017) Multi-focus image fusion method using S-PCNN optimized by particle swarm optimization. Soft Comput 1298:1–13. https://doi.org/10.1007/s00500-017-2694-4

14. Johnson JL, Padgett ML (1999) PCNN models and applications. IEEE Trans Neural Netw 17(3):480–498

15. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. Appl Soft Comput 8(1):687–697

16. Kong WW (2013) Multi-sensor image fusion based on NSST domain (ICM)-C-2. Electron Lett 49(13):802

17. Kong W, Zhang L, Lei Y (2014) Novel fusion method for visible light and infrared images based on NSST-SF-PCNN. *Infrared Phys Technol* 65:103–112

18. Kong W, Wang B, Lei Y (2015) Technique for infrared and visible image fusion based on non-subsampled shearlet transform and spiking cortical model. *Infrared Phys Technol* 71:87–98

19. Laurent I, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259

20. Li H, Manjunath BS, Mitra SK (1995) Multisensor image fusion using the wavelet transform. Graph Model Image Process 57(3):235–245

21. Liu Y, Liu S, Wang Z (2015) A general framework for image fusion based on multi-scale transform and sparse representation. Inf Fusion 24:147–164

22. Liu Z, Erik B, Gaurav B et al (2018) Fusing synergistic information from multi-sensor images: an overview from implementation to performance assessment. Inf Fusion 42:127–145

23. Ma JL, Zhou ZQ, Wang B et al (2017) Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Phys Technol 8-17:82

24. Manivannan R, Samidurai R, Cao JD et al (2017) Global exponential stability and dissipativity of generalized neural networks with time-varying delay signals. Neural Netw 87:149–159

25. Manivannan R, Mahendralcumar G, Samidurai R et al (2017) Exponential stability and extended dissipativity criteria for generalized neural networks with interval time-varying delay signals. J Frankl Inst Eng Appl Math 354(11):4353–4376

26. Manivannan R, Samidurai R, Cao JD et al (2017) Design of extended dissipativity state estimation for generalized neural networks with mixed time-varying delay signals. Inf Sci 424:175–203

27. Manivannan R, Samidurai R, Zhu QX (2017) Further improved results on stability and dissipativity analysis of static impulsive neural networks with interval time-varying delays. J Frankl Inst Eng Appl Math 354(14):6312–6340

28. Nasiraghdam H, Jadid S (2012) Optimal hybrid PV/WT/FC sizing and distribution system reconfiguration using multi-objective artificial bee colony (MOABC) algorithm. Sol Energy 86(10):3057–3071

29. Radhakrishna A, Sheila H, Francisco E, et al (2009) Frequency-tuned salient region detection. IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009:1597-1604

30. Song YD, Wu W, Liu Z et al (2016) An adaptive pansharpening method by using weighted least squares filter. IEEE Geosci Remote SensLett 13(1):18–22

31. Toet A (1989) Image fusion by a ratio of low-pass pyramid. Pattern Recogn Lett 9(4):245–253

32. Vladimir P, Costas, X (2004) Evaluation of image fusion performance with visible differences. 8th European Conference on Computer Vision, ECCV 2004, Lect Notes Comput Sci, 3023: 380–391

33. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

34. Wang Q, Zhou D, Nie R, et al (2016) Medical image fusion using pulse coupled neural network and multi-objective particle swarm optimization. Eighth International Conference on Digital Image Processing, ICDIP 2016, DOI: https://doi.org/10.1117/12.2245043, 10033: 100334K
35. Wei C, Blum RS (2010) Theoretical analysis of correlation-based quality measures for weighted averaging image fusion. Information Fusion 11(4):301–310
36. Xu X, Shan D, Wang G, Jiang X (2016) Multimodal medical image fusion using PCNN optimized by the QPSO algorithm. Appl Soft Comput 46:588–595
37. Yang X, Wu W, Yan B et al (2018) Infrared image super-resolution with parallel random Forest. Int J Parallel Prog 4:1–21
38. Yang XM, Wu W, Liu K et al (2018) Multi-semi-couple super-resolution method for edge computing. IEEE Access 6:5511–5520
39. Yoshifusa I (1991) Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. Neural Netw 4(3):385–394
40. Zhan K, Zhang H, Ma Y (2009) New spiking cortical model for invariant texture retrieval and image processing. IEEE Trans Neural Netw 20(12):1980–1986
41. Zhan K, Shi J, Wang H, Xie Y, Li Q (2017) Computational mechanisms of pulse-coupled neural networks: a comprehensive review. Arch Comput Meth Eng 24(3):573–588
42. Zhang L, Tong MH, Marks TK et al (2008) SUN: a Bayesian framework for saliency using natural statistics. J Vis 8(7):32.1-20
43. Zhang D, Mabu S, Hirasawa K (2010) Noise reduction using genetic algorithm based PCNN method. IEEE International Conference on Systems Man and Cybernetics IEEE: 2627–2633
44. Zhang BH, Lu XQ, Pei HQ et al (2015) A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled shearlet transform. Infrared Phys Technol 73:286–297
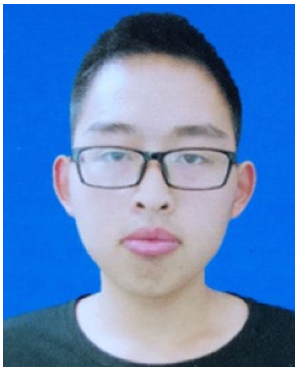
Ruichao Hou is a master's degree candidate at Yunnan University. He received his BE degree in electrical engineering and automation in 2016. His current research interests include neural networks theory and its application, image processing, computer vision, and deep learning.
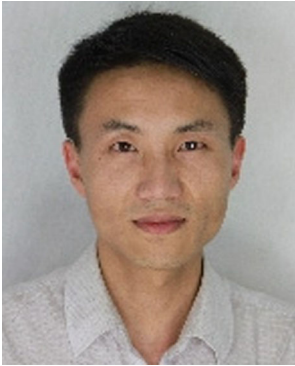
Rencan Nie is an associate professor at the Yunnan University. He received his BS and MS degrees in telecommunication engineering from the Yunnan University in 2004 and 2007, respectively, and his PhD in telecommunication engineering from Yunnan University in 2014.His current research interests include pulse coupled neural networks theory and its application, image processing, and machine learning.

Dongming Zhou is a professor at Yunnan University. He received his BS and MS degrees in automatic control engineering from Huazhong University of Science and Technology (HUST) in1985 and 1988, respectively, and his PhD in electronic circuit and system from Fudan University in 2004. His current research interests include neural network theory and applications, artificial intelligent control theory, and image processing.

Jinde Cao is a professor at Southeast University. He received the B.S. degree from Anhui Normal University, Wuhu, China, the M.S. degree from Yunnan University, Kunming, China, and the Ph.D. degree from Sichuan University, Chengdu, China, all in mathematics/applied mathematics, in 1986, 1989, and 1998, respectively. From March 1989 to May 2000, he was with Yunnan University. In May 2000, he joined the Department of Mathematics, Southeast University, Nanjing, China. From July 2001 to June 2002, he was a Postdoctoral Research Fellow in the Department of Automation and Computer-aided Engineering, Chinese University of Hong Kong, Hong Kong. From August 2002 to October 2002, he was a Senior Visiting Scholar at the Institute of Mathematics, Fudan University, Shanghai, China. From February 2003 to May 2003, he was a Senior Research Associate in the Department of Mathematics, City University of Hong Kong, Hong Kong. From July 2003 to September 2003, he was a Senior Visiting Scholar in the Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China. From January 2004 to April 2004, he was a Research Fellow in the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong. He is currently a Professor and Doctoral Advisor at the Southeast University. Prior to this, he was a Professor at Yunnan University from 1996 to 2000. His research interests include nonlinear systems, neural networks, complex systems and complex networks, stability theory, and applied mathematics.

Dong Liu is a master's degree candidate at Yunnan University. He received his BE degree in electrical automation from Tianjin Polytechnic University in 2016. His current research interests include image processing and image fusion.