Technique for infrared and visible image fusion based on non-subsampled shearlet transform and spiking cortical model

Weiwei Kong [a,*], Binghe Wang [a], Yang Lei [b]

[a] Department of Information Engineering, Engineering University of Armed Police Force, Xi'an 710086, China
[b] Department of Electronic Technology, Engineering University of Armed Police Force, Xi'an 710086, China

HIGHLIGHTS

- The combination between NSST and SCM is a successful attempt in both theoretical research and practical applications.
- NSST is an excellent tool for multi-scale geometric analysis of images.
- With great biological background, SCM is firstly utilized to complete the course of image fusion.
- The NSST–SCM-based algorithm is devised and given in detail.
- The experimentation results of NSST–SCM-based model indicate the great potential and benefits of combining NSST and SCM.

ARTICLE INFO

ABSTRACT

Fusion of infrared and visible images is an active research area in image processing, and a variety of relevant algorithms have been developed. However, the existing techniques commonly cannot gain good fusion performance and acceptable computational complexity simultaneously. This paper proposes a novel image fusion approach that integrates the non-subsampled shearlet transform (NSST) with spiking cortical model (SCM) to overcome the above drawbacks. On the one hand, using NSST to conduct the decompositions and reconstruction not only consists with human vision characteristics, but also effectively decreases the computational complexity compared with the current popular multi-resolution analysis tools such as non-subsampled contourlet transform (NSCT). On the other hand, SCM, which has been considered to be an optimal neuron network model recently, is responsible for the fusion of sub-images from different scales and directions. Experimental results indicate that the proposed method is promising, and it does significantly improve the fusion quality in both aspects of subjective visual performance and objective comparisons compared with other current popular ones.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The purpose of image fusion [1] is to integrate two or more images of the same scene obtained from different kinds of image sensors or the same one functioning in diverse modes, to create a single composite image which contains much more useful and valuable information. Obviously, the final fused image will thus be more informative and provide a better view for human and machine perception or further image processing tasks. Nowadays, despite a variety of sensors, a majority of them commonly places emphasis on the description of certain kind of images or certain partition of an image, like infrared image sensors, multi-focus image sensors, and so on. Hence, the fusion course is very

necessary and an effective way of solving the above mentioned problem. Currently, the image fusion has played an increasingly significant role in enhancing the performance of some civilian and military applications, such as remote sensing [2,3], optical microscopy [4,5] and medical imaging [6,7].

To date, a variety of algorithms for image fusion have been devised and proposed. According to the nature of the processing domain of the input images, these algorithms can be primarily categorized into two categories including spatial domain analysis (SDA) methods and frequency domain analysis (FDA) ones.

Generally, SDA methods are conducted in the spatial domain directly, and more typical image pixels or regions of source images are chosen to construct the final fused image. Due to the existing differences in theoretical basis, SDA can be further classified as linear weighting (LW) methods and artificial neuron network (ANN) ones. The former is much simpler compared with the latter.

* Corresponding author.
 E-mail address: kwwking@163.com (W. Kong).

Basically, the procedure of LW follows two steps. Firstly, certain fusion rule should be devised or chosen to set the weights of the pixels in each source image, aiming at setting the participation rate of each image to be fused. Then, the final fused image can be reconstructed as the scheme established above. For instance, Burt et al. [8] combined the ideas of averaging and choosing together, and then generated a matching matrix to describe the similarity degree between several source images. If the images are similar to a great extent, the weights of them should be equal, whereas the weight should be in direct proportion to the clarity extent of the image. Besides, as another important and improved branch of LW methods, principal component analysis (PCA) [9,10] has also been applied to image fusion. Although LW methods have many remarkable advantages, such as low computational costs and good characteristics of real-time, they also bring some side effects in the final fused image. Traditional LW methods often lead to the decline of the contrast level. Strong correlation between the original component and the substitute is required in PCA, but the reality commonly cannot satisfy the requirement. Unlike LW, ANN is a computational model inspired by animal central nervous systems (especially the brain) that are capable of machine learning and pattern recognition. ANN is usually presented as systems of interconnected "neurons" that can compute values from inputs by feeding information through the network. In the 2007 GRSS data fusion contest, the winning algorithm was based on an ANN approach which extracts land use/land cover maps in and around urban areas by exploiting multi-temporal and multisource coarse-resolution data sets [11]. Known as the typical member and the third generation of ANN, pulse coupled neural network (PCNN) [12] is characterized by global coupling and pulse synchronization of neurons. In recent years, a series of PCNN-based fusion methods and their modified versions have been widely developed. Generally, these methods can be divided into two groups. One is to fuse source images via PCNN theory or its improved versions independently. The other is to conduct the fusion course by using PCNN theory together with other models. For instance, Wang et al. [13] and El-taweel et al. [14] introduced the multi-channel mode and improved the traditional PCNN model to be a novel one, respectively. Kong et al. [15,16] proposed an improved PCNN model, and utilized the novel model and non-subsampled shearlet transform (NSST) to achieve the image fusion. Chai et al. [17] presented a novel algorithm based on lifting stationary wavelet transform and improved PCNN. Yang et al. [18,19] proposed two algorithms concerning PCNN and other models to fuse remote sensing images. According to the fusion performance of the above references, the methods based on PCNN often have obvious superiorities over other mainstream popular ones. However, the application extent of PCNN is still limited due to its complex function mechanism and multiple parameters to be set. In order to overcome the above problem, an improved neural network model called intersecting cortical model (ICM) with no linking neurons was proposed by Kinser [20]. Kong [21] devised a novel algorithm based on improved ICM and NSST, and presented it as a method to fuse multi-sensor images. Spiking cortical model (SCM) [22] is another improved version of PCNN and combines the advantages of both PCNN and ICM. However, only a few studies [23] address the area of image fusion based on SCM.

Unlike SDA, FDA methods are performed in the frequency domain, which coincide with the nature of human vision characteristics. Despite a variety of FDA methods, the fundamental distinction among them mainly lies in which category of the frequency domain transform selected. The common transforms include wavelet transform [24,25], lifting stationary wavelet transform [17], quaternion curvelet transform [26], contourlet transform [27], sparse representation [28], shearlet transform (ST) [29], and recently popularized NSCT [30,31]. Although ST and

NSCT have much better fusion performance compared with past categories of frequency domain transforms, their inherent drawbacks still cannot be ignored. Due to the lack of shift-invariance property, artifacts often appear in the final fused image based on ST. NSCT overcomes the above problem, but its computational costs are very high, so that it is not suitable for real-time applications. In comparison with ST and NSCT, the NSST [32] model has not only much better ability to extract the features and fuse them into the final fused result, but also more acceptable level of computational complexities. However, the relevant work [15,21] is very few.

Considering the SDA and FDA methods mentioned above, two points can be directly concluded. On the one hand, SDA methods commonly have low computational costs, but it does not place sufficient emphasis on the human vision effects. On the other hand, since FDA methods have less consideration of the spatial information during the courses of the initial decompositions and final reconstructions, the phenomena of artifacts and distortions may emerge in the fused image to some extent. Therefore, this paper attempts to combine the advantages of both SDA and FDA methods together, and proposes a novel image fusion technique based on NSST and improved SCM.

The contributions of this paper are summarized as follows:

- The combination between NSST and SCM is a successful attempt in both theoretical research and practical applications.
- NSST is an excellent tool for multi-scale geometric analysis of images, and has remarkable advantages in image processing over other current multi-resolution transform models.
- With great biological background, SCM is firstly utilized to complete the course of image fusion.
- The NSST–SCM-based algorithm is devised and given in detail.
- The experimentation results of NSST–SCM-based model indicate the great potential and benefits of combining NSST and SCM.

The remainder of this paper is organized as follows. In Section 2, the preliminary knowledge of NSST and the traditional SCM is briefly reviewed. Section 3 presents the improved SCM (ISCM) and concretely describes the proposed image fusion method. Some experiments will be conducted, and their results, together with relevant discussions, will be reported in Section 4. Conclusions and future work are summarized in the end.

2. Preliminaries

2.1. Non-subsampled shearlet transform [15,16,32]

According to Refs. [15,16,32], let dimension $n = 2$, the affine systems of ST can be expressed as follows:

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

where ψ is a collection of basis function and satisfies $\psi \in L^2(R^2)$; A denotes the anisotropy matrix for multi-scale partitions, S is a shear matrix for directional analysis. j, l, k are scale, direction and shift parameter respectively. A, S are both 2×2 invertible matrices and $|\det S| = 1$. For each $a > 0$ and $s \in \mathbf{R}$, the matrices of A and S are given as follows:

$$A = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \quad S = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \tag{2}$$

Assume $a = 4$, $s = 1$, (2) can be modified further.

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{3}$$

For any $\xi = (\xi_1, \xi_2) \in \hat{R}^2, \xi_1 \neq 0$, the mathematical expression of basic function $\hat{\psi}^{(0)}$ for ST can be given by [31]

$$\hat{\psi}^{(0)}(\xi) = \hat{\psi}^{(0)}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\hat{\psi}_2(\xi_2/\xi_1) \quad (4)$$

where $\hat{\psi}$ is the Fourier transform of ψ. $\hat{\psi}_1 \in C^\infty(R), \hat{\psi}_2 \in C^\infty(R)$ are both wavelet, and $\sup p\hat{\psi}_1 \subset [-1/2, -1/16] \cup [1/16, 1/2]$, $\sup p\hat{\psi}_2 \subset [-1, 1]$. It implies that $\hat{\psi}_0 \in C^\infty(R)$ and compactly supported with $\sup p\hat{\psi}_0 \subset [-1/2, 1/2]^2$. In addition, we assume that

$$\sum_{j \geq 0} |\hat{\psi}_1(2^{-2j}\omega)|^2 = 1, \ |\omega| \geq 1/8 \quad (5)$$

And for each $j \geq 0$, $\hat{\psi}_2$ satisfies that

$$\sum_{l=-2^j}^{2^j-1} |\hat{\psi}_2(2^j\omega - l)| = 1, \ |\omega| \leq 1 \quad (6)$$

From the conditions on the support of $\hat{\psi}_1, \hat{\psi}_2$ one can obtain that the function $\psi_{j,l,k}$ has the frequency support listed below:

$$\sup p\hat{\psi}^0_{j,l,k} \subset \{(\xi_1, \xi_2) | \xi_1$$
$$\in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}], |\xi_2/\xi_1 + l2^{-j}| \leq 2^{-j}\} \quad (7)$$
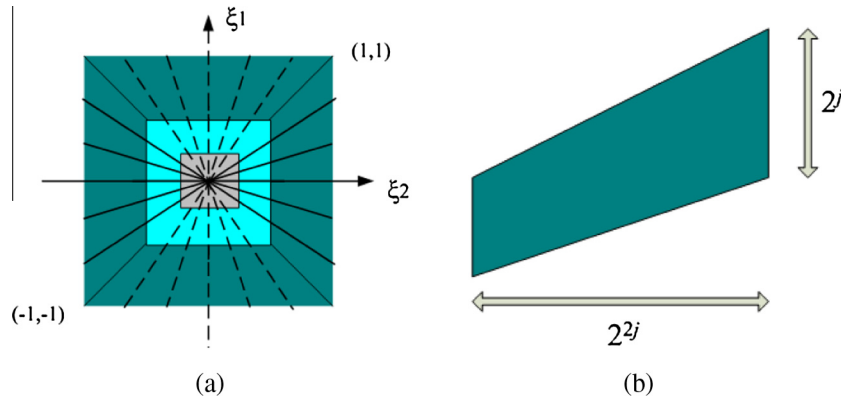

Fig. 1. The structure of the frequency tiling by the shearlet. (a) The tiling of the frequency plane R^2 induced by the shearlet. (b) The size of the frequency support of a shearlet $\psi_{j,l,k}$.
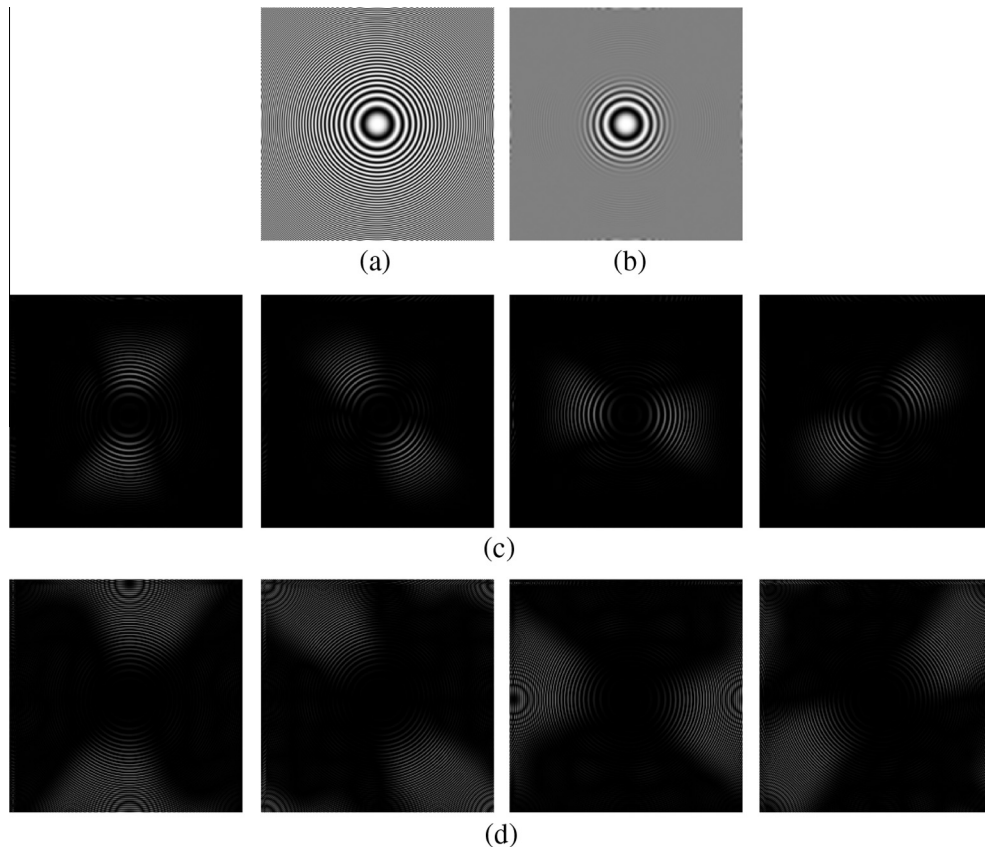

Fig. 2. An illustration of the NSST. (a) Zoneplate (256 * 256). (b) The approximate NSST coefficients. (c) Images of the detailed coefficients at level 1. (d) Images of the detailed coefficients at level 2.
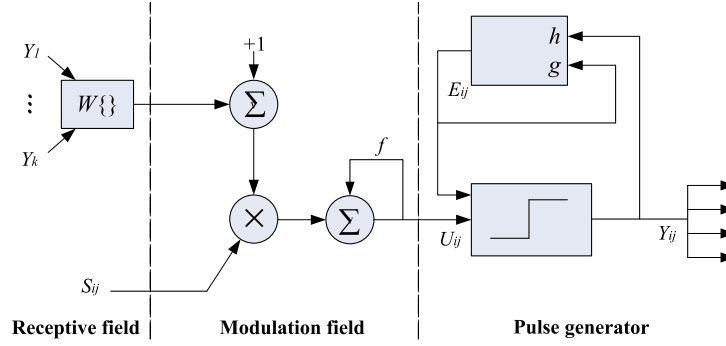
Fig. 3. The structure of the basic SCM neuron.

That is, each element $\hat{\psi}_{j,l,k}$ is supported on a pair of trapeziform zones, whose sizes all approximate to $2^{2j} \times 2^{j}$. The tiling of the frequency by shearlet and the size of the frequency support of $\psi_{j,l,k}$ are illustrated in Fig. 1. Note that Fig. 1(b) only shows the frequency support for $\xi_1 > 0$; the figure of the other support for $\xi_1 < 0$ is symmetrical.

NSST combines the non-subsampled Laplacian pyramid (NSLP) transform with several different combinations of the shearing filters (SF). It is commonly acknowledged that NSST is the shift-invariant version of ST essentially. In order to eliminate the courses of up-sampling and sub-sampling, NSST utilizes NSLP filters as a substitute for the Laplacian pyramid filters used in the ST mechanism, so that it has superior performance in terms of shift-invariance, multi-scale and multi-directional properties. The discretization process of NSST is composed of two phases including multi-scale factorization and multi-directional factorization. NSLP is utilized to complete multi-scale factorization. The first phase ensures the multi-scale property by using two-channel non-sub-sampled filter bank, and one low frequency image and one high frequency image can be produced at each NSLP decomposition level. The subsequent NSLP decompositions are implemented to decompose the low frequency component available iteratively to capture the singularities in the image. The multi-directional factorization in NSST is realized via improved SF. These filters are formed by avoiding the sub-sampling to satisfy the property of shift-invariance. SF allows the direction decomposition with l stages in high frequency images from NSLP at each level and produces 2^{l} directional sub-images with the same size as the source image. Fig. 2 illustrates the two-level NSST decomposition of an image. The number of shearing directions is chosen to be 4 and 4 from finer to coarser scale.

2.2. Spiking cortical model

As a novel member of ANN-based models, spiking cortical model (SCM) combines the advantages of both PCNN and ICM. A SCM neuron commonly denoted by N_{ij} consists of three units: the receptive field, the modulation field, and the pulse generator. The structure of a basic SCM neuron is shown in Fig. 3.

The discrete mathematical expressions of SCM model can be described as follows [23]:

$$F_{ij}[n] = S_{ij} \tag{8}$$

$$L_{ij}[n] = V_L \sum_{kl} W_{ijkl} Y_{kl}[n-1] \tag{9}$$

$$U_{ij}[n] = fU_{ij}[n-1] + S_{ij}\sum_{kl} W_{ijkl} Y_{kl}[n-1] + S_{ij} \tag{10}$$

$$E_{ij}[n] = gE_{ij}[n-1] + hY_{ij}[n-1] \tag{11}$$

$$Y_{ij}[n] = \begin{cases} 1, & \text{if } 1/(1+\exp(-\gamma(U_{ij}[n]-E_{ij}[n]))) \geqslant 0.5 \\ 0, & \text{else} \end{cases} \tag{12}$$

As shown in Fig. 3, N_{ij} receives input signals via other neurons and external sources by two channels in the receptive field. One channel is the feeding input S_{ij} and the other one is the linking input L_{ij}, which correspond to (8) and (9), respectively. U_{ij} is the internal activity and combines the information of the above two channels in a second order mode to form the total internal activity, as shown in (10). According to (11), E_{ij}, the dynamic threshold, will influence the firing frequency of the neurons by comparing with the output Y_{ij}. Eq. (12) indicates that U_{ij} is then compared with E_{ij} to decide the value of the output Y_{ij}. If the requirement mentioned in (12) is satisfied, then the neuron N_{ij} will be activated and generate a pulse, which is characterized by $Y_{ij} = 1$, else $Y_{ij} = 0$. The result is an auto wave expanding from an active neuron to the whole region.

Apart from the parameters discussed above, there are still six other ones including V_L, W_{ijkl}, f, g, h and γ needing explaining. V_L is the magnitude scaling term; W_{ijkl} is the linking matrix; f and g denote the decay constants; h is the threshold magnitude coefficient; γ denotes a parameter of sigmoid function. Obviously, SCM has a more compact structure than tradition PCNN models in terms of the amount of parameters and operation mechanism.

3. Proposed method

After reviewing and discussing the preliminary basic theories, we propose a novel method that is based on the combination and enhancement of SCM and NSST to deal with the problem of the fusion of infrared and visible images. The schematic diagram of our proposed image fusion method is shown in Fig. 4. For simplicity, we take the fusion process of two multi-focus images for instance. Suppose that the two source images respectively denoted by A and B have been already accurately registered. The proposed method is composed of three main phases detailed as follows:

3.1. Improved SCM (ISCM) and its parameters setting

Although SCM combines the advantages of both PCNN and ICM models, its several inherent drawbacks still cannot be ignored by us, which are listed in detail as follows:

(1) The number of the parameters needing setting is large. On the one hand, by referring to the references on ANN, we can easily find that it is almost impossible to set all parameters to be the very optimal values to achieve the best fusion performance. On the other hand, in most conventional ANN based fusion methods, the parameters are often chosen manually on the basis of personal experience or numerous unnecessary, repeated and time consuming experiments.
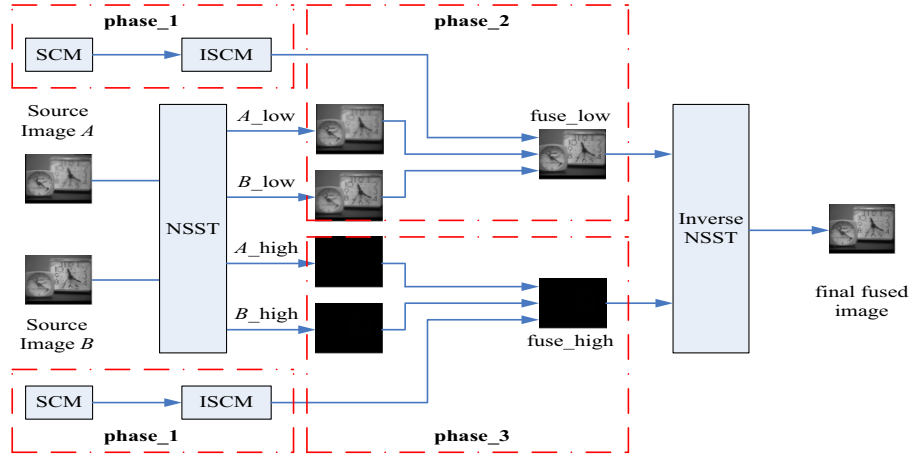
Fig. 4. The schematic diagram of our proposed image fusion method.

What is worse, the values you adjust are merely applicable in certain experiments or applications, but their performance on other occasions may be bad.

(2) The determination of the iterative number is also an onerous task for us. If n is too small, the neurons will not be activated sufficiently so that the advantage of the synchronous impulse bursting cannot be reflected enough. Of course, it is also irrational to set n as large as possible, the reason for which is that there commonly exists a comparatively desirable value or interval for the iterative number, an exceedingly large value will not only result in an adverse influence on the visual effects, but also increase the computational costs a lot. As a result, we have to develop an efficient scheme of iterative number setting.

In order to overcome the above two problems, an improved SCM (ISCM) is proposed whose concrete measures are given as follows:

(1) Except for the iterative number n, there are six parameters in the traditional SCM altogether.

(a) V_L: As known to us, V_L denotes the magnitude scaling term in the linking input unit. According to (9) and (10), if we omit the part of the former internal activity namely $fU_{ij}[n-1]$, (10) can be rewritten as follows.

Comparing (13) with the internal activity in PCNN model, we can conclude that the expression $1/V_L$ corresponds to the linking strength β in PCNN. Of course, here we suppose that $V_L \neq 0$. Since the parameter V_L is counteracted in the internal activity in (10), we do not have to consider setting its value.

$$
\begin{aligned}
U_{ij}[n] &= S_{ij}\sum_{kl}W_{ijkl}Y_{kl}[n-1] + S_{ij} \\
&= S_{ij}\left(1 + \sum_{kl}W_{ijkl}Y_{kl}[n-1]\right) \\
&= S_{ij}\left(1 + \frac{1}{V_L}\cdot V_L\sum_{kl}W_{ijkl}Y_{kl}[n-1]\right) \\
&= S_{ij}\left(1 + \frac{1}{V_L}\cdot L_{ij}[n]\right)
\end{aligned}
\tag{13}
$$

(b) W_{ijkl}: This matrix serves as a bridge to transfer the output information of the surrounding neurons into the central one. For simplicity, the spatial distance between the central neuron and its surrounding neighbors can be regarded as the elements in the matrix W_{ijkl}. Let the size of the square area is 3×3 and the radius is 1, so the distance covering the certral neuron with its neighbors equals to $2^{1/2}$. Based on the above, the expression of W_{ijkl} can be written as (14).

$$
W_{ijkl} = \begin{bmatrix} 0.707 & 1.000 & 0.707 \\ 1.000 & 0 & 1.000 \\ 0.707 & 1.000 & 0.707 \end{bmatrix}
\tag{14}
$$

(c) f, g, h: Despite their different locations, the main functions of them are similar f and g denote the decay constants, which decide the participation radio of the internal activity U_{ij} and the dynamic threshold E_{ij} last time, respectively. Obviously, the values of U_{ij} and E_{ij} grow with f and g increasing. Besides, h denotes the threshold magnitude coefficient to determine the transformation extent of the last output $Y_{ij}[n-1]$ into $E_{ij}[n]$. Accordingly, the value of h directly concerns E_{ij} to influence the firing threshold of each neuron. Based on the analysis and discussions above, we can easily find that the setting of these parameters is very important because of their inherent adjustment effects. On the other hand, the task of setting them is very difficult. Bad final visual performance will inevitably appear if we fail to set them to be the optimal values. In order to make the parameter setting more objective and protect the final fused result from being interfered by human factor to a great extent, meanwhile, considering the correlation between $U_{ij}[n]$ and $U_{ij}[n-1]$ is not very conspicuous, we can rewrite (10) and (11) as follows:

$$
U_{ij}[n] = S_{ij}\sum_{kl}W_{ijkl}Y_{kl}[n-1] + S_{ij}
\tag{15}
$$

$$
E_{ij}[n] = E_{ij}[n-1] - \Delta + hY_{ij}[n-1]
\tag{16}
$$

In (15), the influence of the last internal activity $U_{ij}[n-1]$ is omitted. Step Δ is a positive constant. It guarantees that the dynamic threshold E_{ij} decreases linearly with the iterative number n increasing. In (16), h is usually set to be a relatively large value to ensure that the firing times of each neuron will not exceed one at most. Here we may as well set Δ and h to be 0.1 and 20, respectively.

(d) γ: γ denotes a parameter of sigmoid function. Let is review the requirement when $Y_{ij} = 1$.

$$1/(1 + \exp(-\gamma(U_{ij}[n] - E_{ij}[n]))) \geqslant 0.5$$
$$\Longleftrightarrow 1 + \exp(-\gamma(U_{ij}[n] - E_{ij}[n])) \leqslant 2$$
$$\Longleftrightarrow \exp(-\gamma(U_{ij}[n] - E_{ij}[n])) \leqslant 1 \qquad (17)$$
$$\Longleftrightarrow -\gamma(U_{ij}[n] - E_{ij}[n]) \leqslant 0$$
$$\Longleftrightarrow \gamma(U_{ij}[n] - E_{ij}[n]) \geqslant 0$$

Here we compare (17) with the pulse bursting condition in PCNN model, so we can set γ to be 1 for simplicity. Under such conditions, if $U_{ij} \geqslant E_{ij}$, then the neuron N_{ij} will be activated and generate a pulse, which is characterized by $Y_{ij} = 1$, whereas else $Y_{ij} = 0$.

(2) The setting of the iterative number n has great impacts on the computational costs and the final fusion effects. Motivated by Ref. [33], the time matrix model [33] with the same size of the source image is utilized to record the firing information of neurons. The expression of the time matrix is given as follows:

$$T_{ij}[n] = \begin{cases} n, & \text{if } Y_{ij} = 1 \text{ for the first time} \\ T_{ij}[n-1], & \text{else} \end{cases} \qquad (18)$$

where T is the time matrix, the iterative number n can be determined adaptively according to the intensity distribution of pixels in images. There are several aspects required to be noted: (i) T_{ij} will keep invariable if N_{ij} does not fire all the time; (ii) if N_{ij} fires for the first time, T_{ij} will be set to be the ordinal value of corresponding iteration; (iii) once N_{ij} has already fired, T_{ij} will not alter again. Its value relies on the ordinal number of the iteration, during which N_{ij} fired for the first time. Once all pixels have fired, the whole iteration process is over, and the value of the largest element in T indicates the total iteration times [16].

In conclusion, ISCM has overcome the drawbacks of the traditional SCM to a great extent, which can be listed as follows.

(1) The number of the parameters in ISCM declines a lot compared with SCM. The traditional SCM has seven variables needing setting in all, but ISCM has only four ones including W_{ijkl}, h, Δ and γ.
(2) The original function Y in ISCM is replaced by the time matrix T, which provides rich temporal and spatial information simultaneously for subsequent image processing.

3.2. Fusion course of the low-frequency sub images

By using NSST, each source image can be decomposed into a low-frequency sub image and a series of high-frequency ones. The coefficients in the low-frequency sub image are the approximate representation of the original source image, and inherit its characteristics, like the whole contrast level, the environmental information, and so on. In order to emphasize and preserve the main information of the source images well, ISCM proposed in Section 3.1 is utilized to complete the fusion course of the low-frequency sub images. Here we assume that two source images respectively denoted by A and B have been already accurately registered. A_low, B_low, and F_low are the low-frequency components of A, B and the final fused image, respectively. The concrete steps of the fusion algorithm can be described below:

Step (a): The coefficients of A_low and B_low are both normalized between 0 and 1. Take them respectively as the feeding input to stimulate the corresponding ISCM.
Step (b): Initialize $U_{ij}[0] = T_{ij}[0] = Y_{ij}[0] = 0$, $E_{ij}[0] = 1$.

Note that all neurons do not fire at first. The reason for $E_{ij}[0]$ setting as given above is to make neurons be activated as soon as possible, which can prevent unnecessary "void" iterations and save computational costs.
Step (c): Compute $U_{ij}[n]$, $T_{ij}[n]$, $Y_{ij}[n]$ and $E_{ij}[n]$ according to (15), (18), (12) and (16), respectively.
Step (d): Implement step (c) iteratively until all neurons are activated, namely the elements in $T_{ij}[n]$ are equal to one. Then, the coefficients of F_low can be chosen as follows.

$$F\_low_{ij} = \begin{cases} A\_low_{ij}, & \text{if } T_{ij,A\_low} \leqslant T_{ij,B\_low} \\ B\_low_{ij}, & \text{if } T_{ij,A\_low} > T_{ij,B\_low} \end{cases} \qquad (19)$$

In (19), if T_{ij,A_low} is less than T_{ij,B_low}, it means that the pixel located at (i,j) in A_low has much stronger feeding input than the corresponding pixel in the same place of B_low. Therefore, we choose A_low_{ij} as the corresponding pixel in F_low. Conversely, B_low_{ij} will be chosen.

3.3. Fusion course of the high-frequency sub images

Unlike the low-frequency component, the high-frequency sub images place great emphasis upon the description of the information of edges and details in original source images. As a result, we adopt ISCM to construct the maximum selection rule for the sake of capturing and preserving the details information as much as possible. Suppose that $A\_high_{k,l}$, $B\_high_{k,l}$, and $F\_high_{k,l}$ are the high-frequency components of A, B and the final fused image, respectively. k denotes the level number of multi-scale decompositions, and l is the stage number of multi-directional decompositions at the kth level. The concrete steps of the fusion algorithm can be described below:

Step (a): The coefficients of $A\_high_{k,l}$ and $B\_high_{k,l}$ are both taken to be the feeding input to stimulate the corresponding ISCM.
Step (b): Initialize $U_{ij}[0] = T_{ij}[0] = Y_{ij}[0] = 0$, $E_{ij}[0] = \max (|A\_high_{k,l}|, |B\_high_{k,l}|)$.

Note that all neurons do not fire at first. The reason for $E_{ij}[0]$ setting as given above is as follows. (i) Setting the largest coefficient in a pair of corresponding sub images to be the upper limit is helpful to guarantee that there must be certain neurons which will be activated during the first iteration. (ii) Unlike the low-frequency coefficients, negative coefficients in high-frequency sub images are permissible. Since positive and negative coefficients are responsible for representing and describing the edges and contours information from different angles respectively, it is not reasonable to normalize all coefficients within the interval [0,1]. On this condition, we could estimate the capacity to describe the edge information by comparing their corresponding absolute values.

Step (c): Compute $U_{ij}[n]$, $T_{ij}[n]$, $Y_{ij}[n]$ and $E_{ij}[n]$ according to (15), (18), (12) and (16), respectively.
Step (d): Implement step (c) iteratively until all neurons are activated, namely the elements in $T_{ij}[n]$ are equal to one. Then, the coefficients of $F\_high_{k,l}$ can be chosen as follows:

$$F\_high_{ij,k,l} = \begin{cases} A\_high_{ij,k,l}, & \text{if } T_{ij,A\_high_{k,l}} \leqslant T_{ij,B\_high_{k,l}} \\ B\_high_{ij,k,l}, & \text{if } T_{ij,A\_high_{k,l}} > T_{ij,B\_high_{k,l}} \end{cases} \qquad (20)$$

According to (19) and (20), by using ISCM, the courses of low-frequency sub images fusion and high-frequency sub image fusion can be realized, which correspond to phase_2 and phase_3, respectively. Then, the final fused result can be obtained by using inverse NSST.
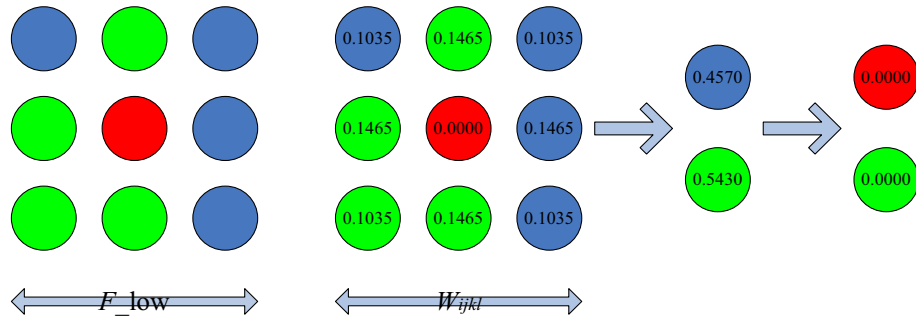
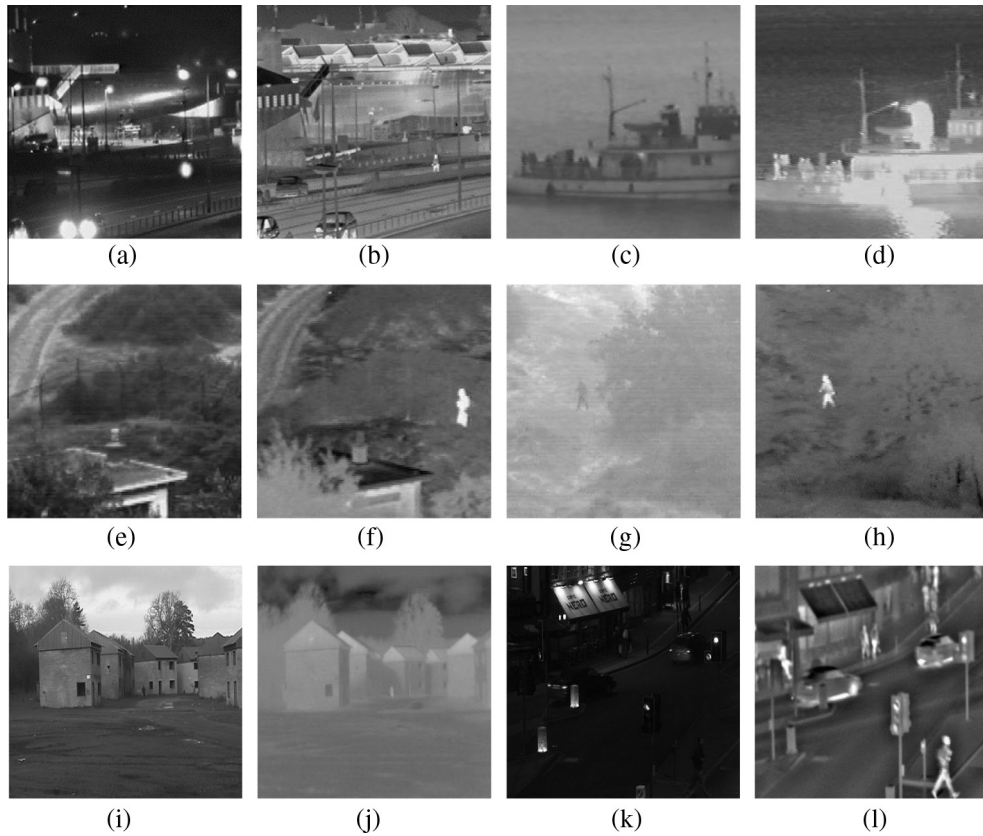**Fig. 5.** An example for "Smooth" rule.

Fig. 6. Source images. (a) Visible night image (256 ∗ 256). (b) Infrared night image (256 ∗ 256). (c) Visible boat image (256 ∗ 256). (d) Infrared boat image (256 ∗ 256). (e) Visible uncamp image (256 ∗ 256). (f) Infrared uncamp image (256 ∗ 256). (g) Visible forest image (256 ∗ 256). (h) Infrared forest image (256 ∗ 256). (i) Visible octec image (512 ∗ 512). (j) Infrared octec image (512 ∗ 512). (k) Visible quad image (512 ∗ 512). (l) Infrared quad image (512 ∗ 512).

However, because the final fused image is composed of a number of pixels, artifacts may appear in the border area between the pixels which come from different source images. In order to make the fused image look better and more natural, another step called "Smooth" is devised to optimize the F_low and $F\_high_{k,l}$. Take the fusion course of the low-frequency sub images in phase_2 for example, and the framework of "Smooth" is given as follows.

Suppose that the size of the square area centered in $f(i,j)$ is 3×3, other eight pixels are chosen from the source image A or B. Besides, W_{ijkl} is the linking matrix which directly decides the participation extent of the surrounding neurons in a square area centered in another certain neuron. For this reason, we can give a comprehensive consideration of the locations of the eight pixels and the linking matrix W_{ijkl} to smooth the final fused low-frequency image. The smooth rule can be expressed as follows:

$$F\_low_{ij} = \begin{cases} A\_low_{ij}, & \text{if} \quad \text{surrounding weight sum based on source image } A \\ \geqslant \text{surrounding weight sum based on source image } BB\_low_{ij}, & \text{if} \quad \text{surrounding weight sum based on source image } A \\ < \text{surrounding weight sum based on source image } B \end{cases}$$
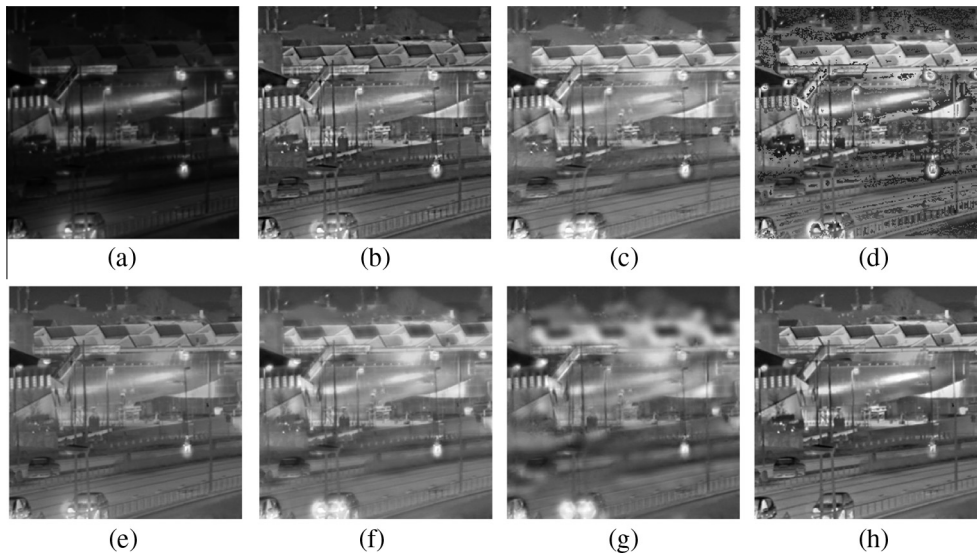
(21)

Fig. 7. Fused results of 'night'. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.

Fig. 8. Fused results of 'boat'. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.
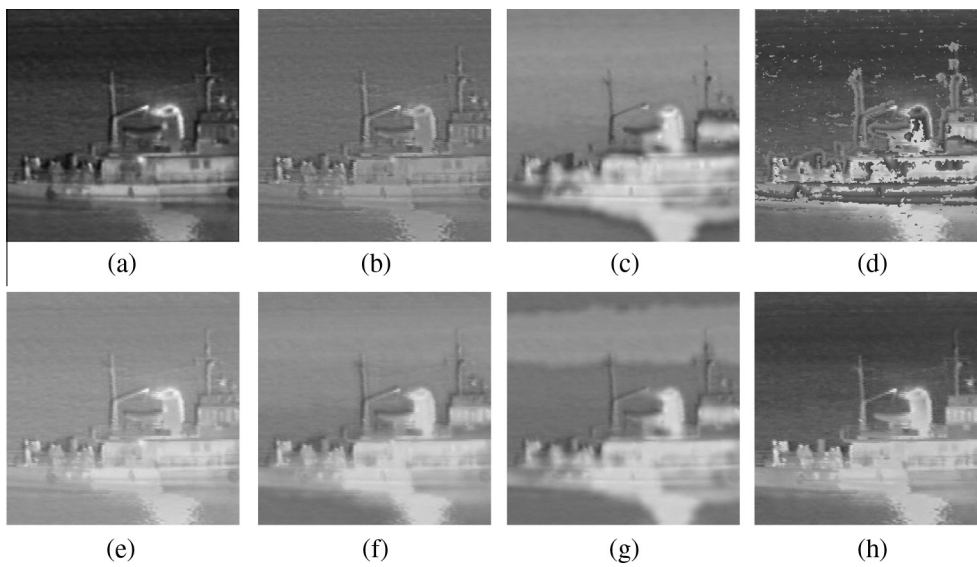
Fig. 5 gives an example for the above "Smooth" rule.

In Fig. 5, the red[1] circle denotes the central pixel in a square area; the green and blue circles indicate that they are from different source images. By computing the weight sum of the eight surrounding pixels, we can easily observe that the weight sum in green is less than that in blue. According to (21), the central pixel should be replaced by the corresponding one in the source image for which the green color stands.

4. Experimental results and related analysis

In order to verify the effectiveness of the proposed method, a series of simulation experiments are conducted in this section.

The section mainly consists of three parts as follows: (a) Experimental introduction: this part includes PC hardware conditions, the source images used, and a brief list of the current popular methods to compare with in the following experiments. (b) Performance evaluation: this part is divided into two aspects including subjective qualitative evaluation and objective quantitative evaluation. Of course, the evaluation indexes are given in this part. (c) Related analysis and discussions: in addition to part (b), further analysis and discussions on the simulation results are carried out in this part.

4.1. Experimental introduction

The simulation experiments are conducted by MATLAB 2010a software on a PC with Intel Core i5/2.3 GHz/2G. Six sets of infrared and visible images which are shown in Fig. 6 are utilized to verify
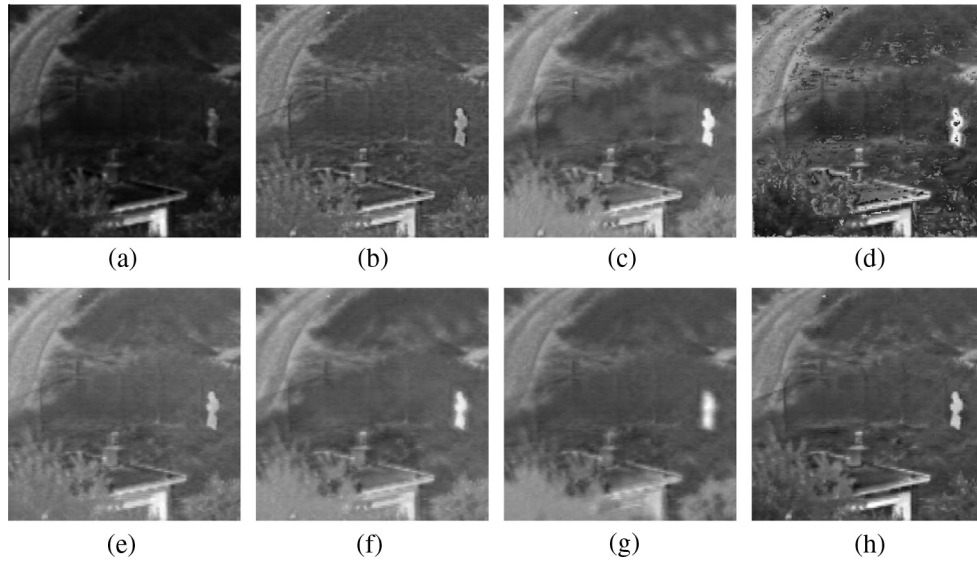
---

[1] For interpretation of color in Fig. 5, the reader is referred to the web version of this article.

Fig. 9. Fused results of 'uncamp'. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.

Fig. 10. Fused results of 'forest'. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.
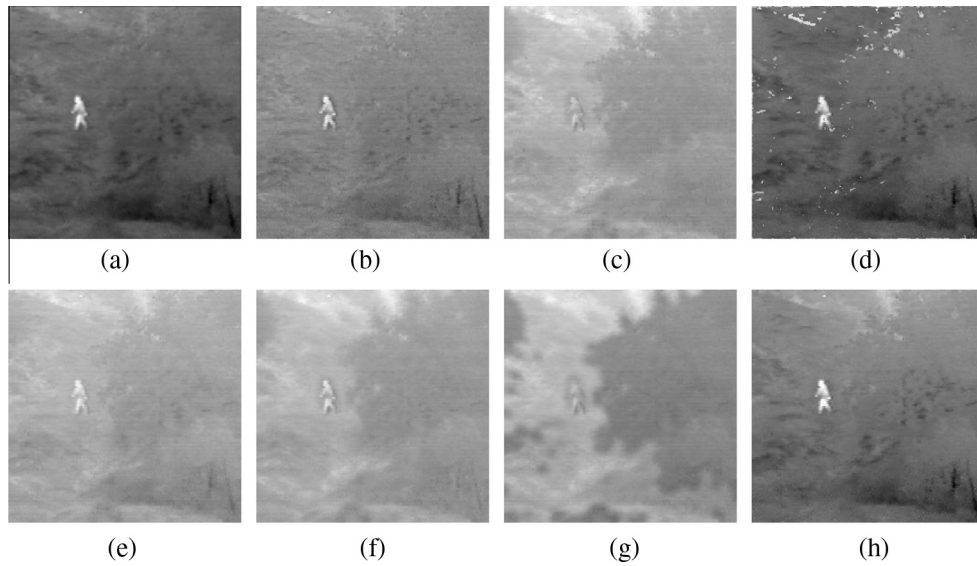
the superiority of the proposed method. All of them cover 256 gray levels.

For simplicity, we call the proposed method M8. Apart from M8, seven other current popular ones are adopted including m-PCNN based method [13] (M1), Image fusion based on NSST-improved-PCNN based method [15] (M2), SW-NSCT-PCNN based method [18] (M3), NSST-I^{2}CM based method [21] (M4), SCM based method [23] (M5), ST based method [29] (M6), and NSCT based method [30] (M7). The above seven methods are still implemented as the Refs. [13,15,17,20,22,28,29]. If you want to obtain more details, please see the related contents. To M8, the level of multi-scale decompositions is set to be 3, and the number of direction from coarser to finer scale is set to be 8, 16 and 16, respectively. The size of the shearing window is set to be 3. The size of the neighborhood is 3×3. W_{ijkl} is [0.707,1.000,0.707; 1.000,0.000, 1.000; 0.707,1.000,0.707]. $\varDelta$ and h are set to be 0.1 and 20, respectively.

4.2. Performance evaluation

This section consists of two aspects including subjective qualitative evaluation and objective quantitative evaluation. The final fused results on the six sets of source images based on M1–M8 are illustrated in Figs. 7–12.

As shown in Figs. 7–12, it can be observed that the final fused images based on M3, M5, M6, M7 and M8 are much better than those based on M1, M2 and M4 in terms of contours definition and the whole contrast level from the visual performance point of view. For instance, in Fig. 12, the final fused results based on M1 and M2 mainly inherit the characteristics of original visible images, but they do not preserve the corresponding main and details information well in the source infrared image, so that we can scarcely notice the person in the left region of the source image. There are so much noise-like artifacts in the result based on M4. By comparison, the fused images based on M3, M5, M6

Fig. 11. Fused results of 'octec'. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.

Fig. 12. Fused results of 'quad'. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.
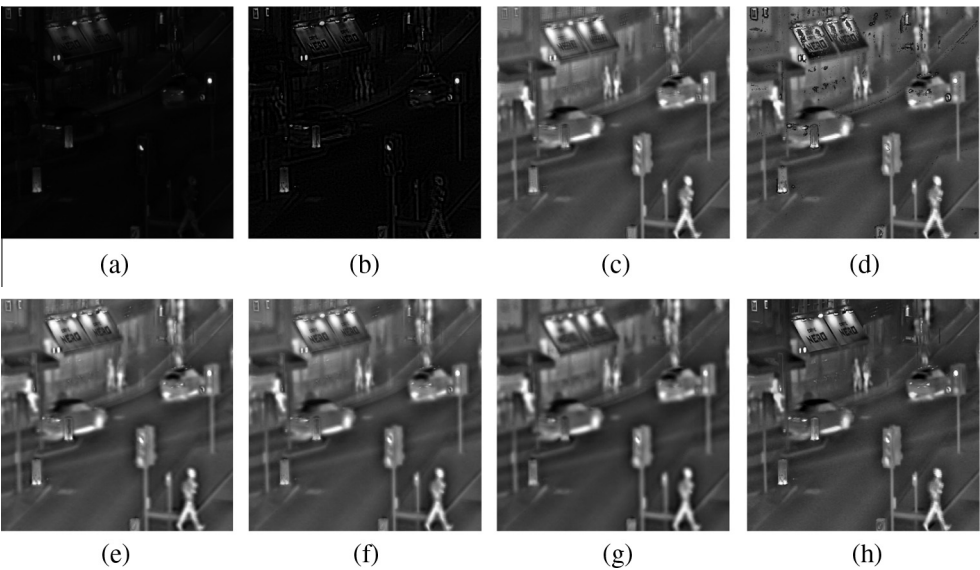
Table 1
Objective evaluation results based on M1–M8 about 'night' and 'boat'.

	'Night'			'Boat'		
	SF	*MI*	*IE*	*SF*	*MI*	*IE*
M1	9.9144	5.2185	6.2846	9.6301	4.9708	6.7938
M2	20.3511	4.9976	7.1397	9.4278	3.8424	5.8035
M3	18.8833	5.1901	7.2942	5.5861	5.4469	6.5695
M4	9.6509	**6.9795**	7.2266	9.4038	6.4787	5.8578
M5	17.5105	4.9201	7.3135	7.8137	6.0555	6.0556
M6	9.5483	5.0400	7.2179	6.2969	4.1280	6.3347
M7	10.7534	4.9544	7.2551	4.8526	4.1826	6.3463
M8	**21.1382**	6.1139	**7.3403**	**10.3739**	**6.6250**	**6.9713**

Table 2
Objective evaluation results based on M1–M8 about 'uncamp' and 'forest'.

	'Uncamp'			'Forest'		
	SF	*MI*	*IE*	*SF*	*MI*	*IE*
M1	7.2600	4.8157	3.5782	6.5667	4.4808	6.9815
M2	11.2556	4.2082	4.3955	**7.1702**	3.5205	7.2992
M3	9.6115	4.3169	6.7317	4.9049	5.6270	6.4930
M4	10.9755	4.6140	6.7472	6.5623	6.1507	6.5522
M5	7.9240	4.4503	6.8156	4.6514	5.2084	6.9927
M6	7.1665	4.3356	6.7297	4.3388	3.5346	6.5131
M7	6.3448	4.4550	6.7694	3.9757	3.9806	6.4507
M8	**11.4604**	**4.9102**	**6.9104**	6.8563	**6.3763**	**7.3341**

and M7 combine the advantages of the both source images together, what is more, they also share reasonable contrast levels. Of course, after careful comparison, we can easily find that M8 is superior to the above four ones in both preserving important details information and inheriting the characteristics of source images.

In addition to subjective evaluations, the objective quality metrics including space frequency (SF) [34], mutual information (MI) [35], and information entropy (IE) have also been used for evaluating the fused images based on different methods from the perspective of the quantity.

Tables 1–3 report the objective evaluation results based on M1–M8. Bold values indicate the best result in the same index among the mentioned methods. Obviously, the results of quantitative comparisons concerning M1–M8 are also consistent with the above subjective visual effects.

4.3. Further analysis and discussions

The above six simulation experiments indicate that the proposed method (M8) in this paper has a notably superiority in both subjective visual effects and objective evaluation criteria over other seven current popular methods. In this section, further analysis and discussions on the experimental results will be carried out in terms of several other aspects.

4.3.1. Computational efficiency

Take the experimental results of the 'quad' source images for example. Table 4 reports the actual ART of M1–M8. Due to the absence of the multi-resolution analysis mechanism, M1 and M5 are time-saving which only cost 3.286 s and 4.375 s respectively. On the contrary, since NSCT is very time-consuming, the computational complexities of M3 and M7 involving NSCT are rather high than other six methods. Similar to M1 and M5, M6 which is based on ST with a lower computational cost does not require much computer resource. M2, M4 and M8 are based on the

theories of both NSST and ANN, so they can be considered as the same type of fusion methods in fact. Compared with the IPCNN, I^2CM is more efficient which directly leads to a much lower complexity. SCM with the both superiorities of PCNN and ICM is one of the simplified PCNN models, so the time M8 consumes should range from M2 and M4.

4.3.2. Enlarged details of regions

In addition to the real-time performance of eight different methods, the subjective visual effects of the final fused results remain a heated topic that attracts much attention in the research field of image fusion. As a result, in this section, we will evaluate the definition level of the fused images from another angle.

Still take the 'quad' source images for instance. The enlarged details of tarpaulins in the eight pictures in Fig. 12 are given in Fig. 13. Obviously, the contrast levels of the fused images based on M1 and M2 are unreasonable. Fig. 13(c) and (g) is so blurred that we can hardly distinguish the characters. In Fig. 13(d), although the characters namely 'NERO' can be noticed and obtained, a great deal of grid-like artificial information appears. As for M5 and M6, the definition levels of the fused results are not very satisfactory either. On the contrary, the superiorities of the fused image based on M8 are very clear. Firstly, the subjective visual effect of Fig. 13(h) is very good. Secondly, the overall contrast level is very reasonable. Thirdly, there is hardly any artificial information in the final fused image.

5. Conclusions

In this paper, a novel fusion technique for infrared and visible images based on NSST and ISCM is proposed. The novel method combines multi-resolution analysis theory and artificial neural network, and sufficiently takes the superiorities of the above both areas. Firstly, multi-scale and multi-directional sub images can be obtained by using NSST. Compared with current popular multi-resolution analysis tools such as NSCT, NSST owns much lower computational costs and better sparse representations. Then, traditional SCM is improved to be ISCM to fuse the sub images according to the fusion rules devised in this paper. Finally, by using the inverse NSST, the final fused image can be obtained. In order to verify the effectiveness of the proposed method, six pairs of source images involving four categories are utilized. Besides, seven other current popular methods are also adopted to conduct the simulation experiments. Experimental results indicate that the proposed method has obvious superiorities over other seven ones in terms of both subjective visual performance and objective quantitative criteria. How to further optimize and enhance the performance of the new method will attract our attention and be the emphasis of our future work.

Table 3
Objective evaluation results based on M1–M8 about 'octec' and 'quad'.

	'Octec'			'Quad'		
	SF	*MI*	*IE*	*SF*	*MI*	*IE*
M1	10.0049	4.7886	6.9815	3.7534	2.9510	3.5782
M2	14.0450	4.7332	**7.2992**	10.1856	1.3769	4.3955
M3	6.7434	3.7299	6.4930	9.0044	3.9542	6.7317
M4	7.9722	3.6728	6.5522	8.5516	3.6134	6.7472
M5	14.1109	4.9432	6.9927	10.7528	3.6936	6.8156
M6	7.4041	3.7386	6.5131	8.1347	3.7686	6.7297
M7	3.1061	4.2615	6.4507	5.9708	3.3950	6.7694
M8	**14.4272**	**5.3502**	7.1834	**12.2939**	**4.0552**	**6.8774**

Table 4
Average running time (ART) based on M1–M8 about 'quad'.

	M1	M2	M3	M4	M5	M6	M7	M8
ART	**3.286**	5.574	390.096	4.708	4.375	4.882	309.135	5.178

The bold statistics indicate the best result among the eight methods.
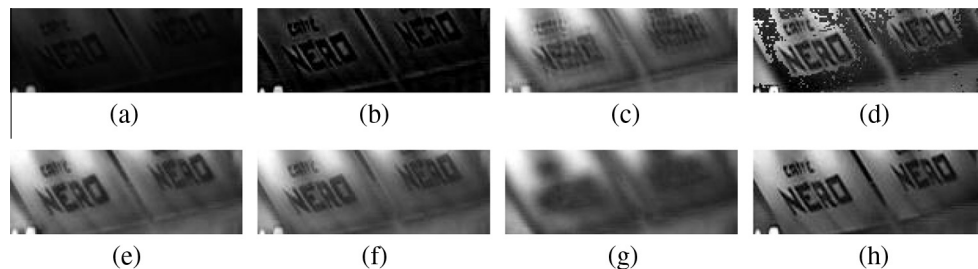

Fig. 13. Enlarged details of tarpaulins in the fused results based on M1–M8 for 'quad' source images. (a) Result based on M1. (b) Result based on M2. (c) Result based on M3. (d) Result based on M4. (e) Result based on M5. (f) Result based on M6. (g) Result based on M7. (h) Result based on M8.

Conflict of interest

Acknowledgements

References

[1] M. Ehlers, Multi-sensor image fusion techniques in remote sensing, Isprs. J. Photogramm. 46 (1) (1991) 19–30.

[2] L. Bin, M.M. Khan, T. Bienvenu, J. Chanussot, Z. Liangpei, Decision-based fusion for pansharpening of remote sensing images, IEEE Trans. Geosci. Remote. S. 10 (1) (2013) 19–23.

[3] S.T. Li, H.T. Yin, L.Y. Fang, Remote sensing image fusion via sparse representations over learned dictionaries, IEEE Trans. Geosci. Rem. 51 (9) (2013) 4779–4789.

[4] J. Swoger, P. Verveer, K. Greger, J. Huisken, E.H.K. Stelzer, Multi-view image fusion improves resolution in three-dimensional microscopy, Opt. Express 15 (13) (2007) 8029–8042.

[5] Y. Chen, L. Wang, Z.B. Sun, Y.D. Jiang, G.J. Zhai, Fusion of color microscopic images based on bidimensional empirical mode decomposition, Opt. Express 18 (21) (2010) 21757–21769.

[6] G.H. Qu, D.L. Zhang, P.F. Yan, Medical image fusion by wavelet transform modulus maxima, Opt. Express 9 (4) (2001) 184–190.

[7] G. Bhatnagar, Q.M.J. Wu, Z. Liu, Human visual system inspired multi-modal medical image fusion framework, Expert Syst. Appl. 40 (5) (2013) 1708–1720.

[8] P.J. Burt, R.J. Kolcznski, Enhanced image capture through fusion, Proc. Conf. Comp. Vis. (1993) 173–182.

[9] S.Y. Yang, M. Wang, L.C. Jiao, Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis, Inform. Fusion 13 (3) (2012) 177–184.

[10] T. Wan, C.C. Zhu, Z.C. Qin, Multifocus image fusion based on robust principal component analysis, Pattern Recogn. Lett. 34 (9) (2013) 1001–1008.

[11] F. Pacifici, F.D. Frate, W.J. Emery, P. Gamba, J. Chanussot, Urban mapping using coarse SAR and optical data: outcome of the 2007 GRSS data fusion contest, IEEE Trans. Geosci. Rem. S. 5 (3) (2008) 331–335.

[12] R.P. Broussard, S.K. Rogers, M.E. Oxley, G.L. Tarr, Physiologically motivated image fusion for object detection using a pulse coupled neural network, IEEE Trans. Neur. Net. 10 (3) (1999) 554–563.

[13] Z.B. Wang, Y.D. Ma, Medical image fusion using m-PCNN, Inform. Fusion 9 (2) (2008) 176–185.

[14] G.S. El-taweel, A.K. Helmy, Image fusion scheme based on modified dual pulse coupled neural network, IET. Image. Process. 7 (5) (2013) 407–414.

[15] W.W. Kong, J.P. Liu, Technique for image fusion based on nonsubsampled shearlet transform and improved pulse-coupled neural network, Opt. Eng. 52 (1) (2013) 017001-1–017001-12.

[16] W.W. Kong, L.J. Zhang, Y. Lei, Novel fusion method for visible light and infrared images based on NSST-SF-PCNN, Infrared. Phys. Techn. 65 (7) (2014) 103–112.

[17] Y. Chai, H.F. Li, M.Y. Guo, Multifocus image fusion scheme based on features of multiscale products and PCNN in lifting stationary wavelet domain, Opt. Commun. 284 (5) (2011) 1146–1158.

[18] S.Y. Yang, M. Wang, Y.X. Lu, W.D. Qi, L.C. Jiao, Fusion of multiparametric SAR images based on SW-nonsampled contourlet and PCNN, Signal Process. 89 (12) (2009) 2596–2608.

[19] S.Y. Yang, M. Wang, L.C. Jiao, Contourlet hidden Markov tree and clarity-saliency driven pcnn based remote sensing images fusion, Appl. Soft. Comput. 12 (1) (2012) 228–237.

[20] J.M. Kinser, Simplified pulse-coupled neural network, Proc. Conf. Appl. Arti. Neur. Net. (1996) 563–567.

[21] W.W. Kong, Multi-sensor image fusion based on NSST domain I²CM, Electron. Lett. 49 (13) (2013) 802–803.

[22] K. Zhan, H. Zhang, Y.D. Ma, New spiking cortical model for invariant texture retrieval and image processing, IEEE Trans. Neur. Net. 20 (12) (2009) 1980–1986.

[23] N.Y. Wang, Y.D. Ma, K. Zhan, Spiking cortical model for multifocus image fusion, Neurocomputer 3 (2013). Available online 3 August.

[24] H.M. Lu, L.F. Zhang, S. Serikawa, Maximum local energy: an effective approach for multisensory image fusion in beyond wavelet transform domain, Comput. Math. Appl. 64 (5) (2012) 996–1003.

[25] A. Ellmauthaler, C.L. Pagliari, E.A.B. Dasilva, Multiscale image fusion using the undecimated wavelet transform with spectral factorization and nonorthogonal filter, IEEE Trans. Image Process. 22 (3) (2013) 1005–1017.

[26] L. Guo, M. Dai, M. Zhu, Multifocus color image fusion based on quaternion curvelet transform, Opt. Express 20 (17) (2012) 18846–18860.

[27] H.J. Wang, Q.K. Yang, R. Li, Tunable-Q contourlet-based multi-sensor image fusion, Signal Process 93 (7) (2013) 1879–1891.

[28] L. Chen, J.B. Li, C.L. Philip Chen, Regional multifocus image fusion using sparse representation, Opt. Express 21 (4) (2013) 5182–5197.

[29] Q.G. Miao, C. Shi, P.F. Xu, M. Yang, Y.B. Shi, A novel algorithm of image fusion using shearlets, Opt. Commun. 284 (6) (2011) 1540–1547.

[30] J.H. Adu, J.H. Gan, Y. Wang, J. Huang, Image fusion based on nonsubsampled contourlet transform for infrared and visible light image, Infrared. Phys. Techn. 61 (1) (2013) 94–100.

[31] G. Bhatnagar, Q.M.J. Wu, L. Zheng, Directive contrast based multimodal medical image fusion in NSCT domain, IEEE Trans. Multimedia 15 (5) (2013) 1014–1024.

[32] G. Easley, D. Labate, W.Q. Lim, Sparse directional image representation using the discrete shearlet transforms, Appl. Comput. Harmon. A 25 (1) (2008) 25–46.

[33] Y.D. Ma, D.M. Lin, B.D. Zhang, A novel algorithm of image enhancement based on pulse coupled neural network time matrix and rough set, Proc. IEEE. Conf. Fuzzy Sys. Knowl. Disc. (2007) 86–90.

[34] Y. Zheng, E.A. Essock, B.C. Hansen, A.M. Haun, A new metric based on extended spatial frequency and its application to dwt based fusion algorithms, Inform. Fusion 8 (2) (2007) 177–192.

[35] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315.