4-1-2020

Infrared and Visible Image Fusion Based on Oversampled Graph Filter Banks

Chunyan Song
Northeast Forestry University

Xueying Gao
Harbin Engineering University

Yu-Long Qiao
Harbin Engineering University

Kaige Zhang
Utah State University, zkghit@163.com

Recommended Citation

Chunyan Song, Xueying Gao, Yu-Long Qiao, and Kaige Zhang "Infrared and visible image fusion based on oversampled graph filter banks," Journal of Electronic Imaging 29(2), 023016 (1 April 2020). https://doi.org/10.1117/1.JEI.29.2.023016

Journal of Electronic Imaging

JElectronicImaging.org

Infrared and visible image fusion based on oversampled graph filter banks

Chunyan Song
Xueying Gao
Yulong Qiao
Kaige Zhang

SPIE. IS&T imaging.org

Chunyan Song, Xueying Gao, Yulong Qiao, Kaige Zhang, "Infrared and visible image fusion based on oversampled graph filter banks," *J. Electron. Imaging* **29**(2), 023016 (2020), doi: 10.1117/1.JEI.29.2.023016

Infrared and visible image fusion based on oversampled graph filter banks

Chunyan Song,[a] Xueying Gao,[b,*] Yulong Qiao,[b] and Kaige Zhang[c]
[a]Northeast Forestry University, College of Information and Computer Engineering, Dongli District, Harbin, Heilongjiang, China
[b]Harbin Engineering University, College of Information and Communications Engineering, Nangang District, Harbin, Heilongjiang, China
[c]Utah State University, Department of Computer Science, Logan, Utah, United States

Abstract. The infrared image (RI) and visible image (VI) fusion method merges complementary information from the infrared and visible imaging sensors to provide an effective way for understanding the scene. The graph filter bank-based graph wavelet transform possesses the advantages of the classic wavelet filter bank and graph representation of a signal. Therefore, we propose an RI and VI fusion method based on oversampled graph filter banks. Specifically, we consider the source images as signals on the regular graph and decompose them into the multiscale representations with M-channel oversampled graph filter banks. Then, the fusion rule for the low-frequency subband is constructed using the modified local coefficient of variation and the bilateral filter. The fusion maps of detail subbands are formed using the standard deviation-based local properties. Finally, the fusion image is obtained by applying the inverse transform on the fusion subband coefficients. The experimental results on benchmark images show the potential of the proposed method in the image fusion applications. [DOI: 10.1117/1.JEI.29.2.023016]

Keywords: image fusion; infrared image; graph filter banks; graph signal processing.

Paper 190978 received Oct. 22, 2019; accepted for publication Mar. 13, 2020; published online Apr. 1, 2020.

1 Introduction

The goal of the image fusion is to combine two or more images of the same scene into a fused image. The resulting image can provide more useful information about the scene for human visual perception or various applications based on image processing and analysis. In the image processing and computer vision, the multiview image fusion,[1] multisensor image fusion,[2] multitemporal image fusion,[3] multifocus image fusion,[4] and multiexposure image fusion[5] methods have been proposed in the literature. Among these methods, the multisensor image fusion technique merges complementary information from two or multiple imaging sensors to provide an effective way for understanding the scene under various conditions. Therefore, the multisensor image fusion methods have been applied in various fields, such as medical diagnosis, remote sensing, digital photography, and military surveillance.[6]

Infrared image (RI) and visible image (VI) fusion is an important part of the multisensor image fusion techniques. According to the imaging principle of infrared sensor, the RIs are captured by recording the thermal radiation of different objects, which reduce the influence of external environment, such as sunlight, smog, and other condition factors, to detect and identify the targets. However, the RIs are usually presented in low-spatial resolution and less detailed background information due to low-thermal contrast in the scene. Compared with the RI, the visible light images are captured by recording the visibly reflective properties of spectrum information of the scene, which has a strong contrast that can reflect more details. But in the situations of low-visual contrast (low illumination and camouflaged target), it can be hard or even impossible to locate the target in the visual band. Fortunately, the image fusion can effectively combine the

*Address all correspondence to Xueying Gao, E-mail: gaoxy@hrbeu.edu.cn

1017-9909/2020/$28.00

heat radiation characteristics of a scene in an RI and the details of the scene in a VI, which can provide an informative image for human visual perception, and improve the performance of object detection, target recognition, and scene analysis.[7,8]

In past years, various RI and VI fusion methods have been introduced. In general, these methods are divided into three levels: pixel level, feature level, and decision level. In the survey paper,[8] authors have classified the fusion methods into seven categories: multiscale transform,[9–15] sparse representation,[16] neural network,[17,18] subspace,[19] saliency-based methods,[15] hybrid models,[20] and fusion methods[21] based on another signal processing theory. The multiscale transform decomposes an image into multiscale components, in which each component (or coefficients) represents a part of information at a scale and can been processed separately. The multiscale mechanism is consistent with the human visual system, which may result in better fusion results. In order to achieve RI and VI fusion with multiscale transform, the source images are first decomposed into multiscale representation. Then, the fusion coefficients are obtained by merging multiscale coefficients with the corresponding fusion rules. Finally, the inverse transform of the fused coefficients is used to reconstruct the fusion image. In the image fusion field, the multiscale transforms, such as the pyramid representation,[9] wavelet transform,[10] graph wavelet transform,[11] curvelet transform,[12] nonsubsampled contourlet transform (NSCT),[13] and edge-preserving multiscale decomposition,[14,15] have been widely used for RI and VI fusion. The main advantages of these methods include preserving details of the source images and providing sensitive contrast information for the human visual system. Therefore, the high-quality fusion image can be obtained with the multiscale transform-based methods.

Graph signal processing is one of the emerging fields in signal processing. In past years, Crovella and Kolaczyk[22] designed wavelet-like functions on graphs that are localized in space and time. Jansen et al.[23] proposed the lifting-based wavelet transforms for the signal on the graph. Hammond et al.[24] constructed a class of wavelet operators in the graph spectral domain, which aroused the research on applying the concepts of filter banks from classical signal processing to graph signal. Because an image can be considered as a signal on the regular graph, Yan et al.[11] proposed the RI and VI fusion method based on graph spectral wavelet transform. They applied the unsharp filter to both RI and VI and then determined the fusion weight values. Finally, the weight values smoothed with the bilateral filter are used to fuse the graph spectral wavelet coefficients. This fusion method has been shown great potential in RI and VI fusion. However, the discrete spectral graph wavelet transform (SGWT) in Hammond et al.[24] is an overcomplete transform that can be considered as a natural discretization of the continue wavelet transform, in which the scaling functions and the spectral graph wavelets with sampled scales form a frame.

Recently, Narang and Ortega[25] showed that downsampling/upsampling operations in bipartite graphs lead to a spectral folding phenomenon, which is analogous to aliasing in regular signal domain, and constructed critically sampled wavelet filter banks on graph for analyzing graph signals defined on any arbitrary finite weighted graph. Then, the biorthogonal wavelet filter banks on graph[26] have been introduced. Tanaka and Sakiyama[27] proposed M-channel oversampled filter banks for graph signals. Tay and Zhang[28] introduced the polyphase representation for graph filter banks and constructed the biorthogonal graph filter banks with the ladder structure. Jiang et al.[29] designed biorthogonal graph filter banks with the lifting-based method. Teke and Vaidyanathan[30] extended the classical multirate signal processing theory to graphs. In the context of graph signal processing, the signal representation with a graph can explicitly consider the signal values and the structure and relationship among the signal samples. Therefore, it is reasonable that the graph wavelet transform combines the advantages of the classic wavelet filter bank and graph representation of a signal and can be used to form multiscale representation of an image, which may result in better image fusion.

The oversampled filter banks for regular signals have more freedom in their design, and it has been shown that they outperform critically sampled systems in the signal processing fields. It is reasonable that the oversampled graph filter banks would be useful for graph signals. Tanaka and Sakiyama[27] proposed M-channel oversampled graph filter bank that satisfies the perfect reconstruction condition and demonstrated its performance in graph signal denoising. Therefore, in this paper, we propose the M-channel oversampled graph filter banks-based RI and VI fusion method. The majority of this paper includes the following two aspects. First, we apply the image

decomposition with the M-channel oversampled graph filter banks to the image fusion. The source images are considered as signals on the regular graph and decomposed into low-frequency and high-frequency subbands with M-channel oversampled graph wavelet filter banks. Second, the fusion map of the approximation subband is constructed by introducing the modified local coefficient of variation (LCV).

The rest of this paper is organized as follows. In Sec. 2, we briefly review the graph signal and the principle of M-channel oversampled graph filter banks. The proposed RI and VI fusion method is introduced in Sec. 3. In Sec. 4, we conduct the fusion experiments on the benchmark images and present the detail discussion. Finally, our work is concluded in the last section.

2 M-Channel Oversampled Graph Filter Banks

2.1 *Graph Signals*

Due to the image fusion applications, we only consider a finite undirected graph with no-loops or multiple links. Formally, a graph G is represented as $G = \{V, \mathcal{E}\}$, where V is the set of nodes or vertices, and $\mathcal{E}$ is the set of edges (each edge is associated with two vertices). A graph signal is a real-value scalar function $f(\nu): V \to R$ defined on the graph such that $f(\nu)$ is a sample of function on $\nu \in V$. The adjacency matrix is one of the matrix representation methods for a graph. The adjacency matrix $\mathbf{A}$ is defined as follows:

$$\mathbf{A}(m,n) = \begin{cases} \omega_{mn} & \text{if nodes } m \text{ and } n \text{ are connected} \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where ω_{mn} is the weight of the edge between m and n. For an undirected graph, its adjacency matrix is symmetric. The degree matrix $\mathbf{D}$ is a diagonal matrix, and its m'th diagonal element is $\mathrm{d_{mm}} = \Sigma_{\mathrm{n}}\mathbf{A}(\mathrm{m},\mathrm{n})$. The graph Laplacian matrix and the symmetric normalized version are defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$ and $\mathbf{L}' = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$, respectively. The eigenvalues of the symmetric normalized graph Laplacian matrix lie in the interval [0, 2], thus the matrix $\mathbf{L}'$ is always used in graph filter banks design. It is assumed that λ_i is one of the eigenvalues of $\mathbf{L}'$, and $\mathbf{u}_{\lambda_i}$ is the corresponding eigenvector that satisfies $\mathbf{L}'\mathbf{u}_{\lambda_i} = \lambda_i\mathbf{u}_{\lambda_i}$. Then, the eigenspace projection matrix is defined as follows:

$$\mathbf{P}_{\lambda_i} = \sum_{\lambda=\lambda_i} \mathbf{u}_\lambda \mathbf{u}_\lambda^T. \tag{2}$$

According to the work,[24] the spectral-domain filter is defined as follows:

$$\mathbf{H} = \sum_{\lambda_i \in \sigma(G)} h(\lambda_i)\mathbf{P}_{\lambda_i}, \tag{3}$$

where $h(\lambda_i)$ is the spectral kernel of filter $\mathbf{H}$, and $\sigma(G)$ is the set of eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ and called the spectrum of the graph G.

2.2 *Critically Sampled Graph Filter Banks*

Narang[31] first proposed critically sampled wavelet filter banks for graph-structured signals. They defined the downsampling and upsampling operators for the graph signal and derived the perfect reconstruction condition. In case of bipartite graphs, the spectrum of the graph is symmetric and the deformed eigenvectors are also the eigenvectors of the same graph. This spectral folding phenomenon forms the basis of the two-channel critically sampled graph filter banks for the bipartite graph.

Let a bipartite graph is defined as $G = \{L, H, \mathcal{E}\}$, where L and H are the partitions of the vertex set such that all the links in the set $\mathcal{E}$ connect nodes of different partitions. For the sake of convenience, the nodes in L are called the lowpass channel and those in H are called the highpass channel. Similar to the case of regular signals, the downsampling-then-upsampling operation[31] is defined as follows:

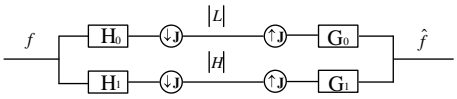

Fig. 1 Critically sampled graph filter bank.

$$\mathbf{D}_{\mathrm{du},L}=\frac{1}{2}(\mathbf{I}_N-\mathbf{J}),\quad \mathbf{D}_{\mathrm{du},H}=\frac{1}{2}(\mathbf{I}_N+\mathbf{J}), \tag{4}$$

where

$$\mathbf{J}_{mm}=\begin{cases}+1 & \text{if } f(m) \text{ belongs to } H,\\ -1 & \text{if } f(m) \text{ belongs to } L.\end{cases} \tag{5}$$

As shown in Fig. 1, under the restriction condition of perfect reconstruction, the whole transfer function should be equal to the identity matrix:

$$\mathbf{T}=\frac{1}{2}\mathbf{G}_0(\mathbf{I}-\mathbf{J})\mathbf{H}_0+\frac{1}{2}\mathbf{G}_1(\mathbf{I}+\mathbf{J})\mathbf{H}_1=\frac{1}{2}(\mathbf{G}_0\mathbf{H}_0+\mathbf{G}_1\mathbf{H}_1)+\frac{1}{2}(\mathbf{G}_1\mathbf{J}\mathbf{H}_1-\mathbf{G}_0\mathbf{J}\mathbf{H}_0)=\mathbf{I}_N, \tag{6}$$

where $\mathbf{H}_k=\Sigma_{\lambda_i\in\sigma(G)}h_k(\lambda_i)\mathbf{P}_{\lambda_i}$ and $\mathbf{G}_k=\Sigma_{\lambda_i\in\sigma(G)}g_k(\lambda_i)\mathbf{P}_{\lambda_i}$, $k=0,1$ are graph filters. Based on this condition, Narang and Ortega[25] obtained the critically sampled graph filter bank perfect reconstruction condition:

$$g_0(\lambda)h_0(\lambda)+g_1(\lambda)h_1(\lambda)=2,\quad g_0(\lambda)h_0(2-\lambda)-g_1(\lambda)h_1(2-\lambda)=0, \tag{7}$$

that is similar to the traditional wavelet filter bank.

2.3 *Oversampled Graph Filter Banks*

It is observed that the oversampled filter banks have more freedom in their design and have been shown that they are better than critically sampled filter banks in some applications. Tanaka and Sakiyama[27] proposed M-channel oversampled graph filter banks for graph signals and demonstrated the potential in the graph signal denoising. Therefore, we introduce M-channel oversampled graph filter banks-based RI and VI fusion method.

The graph signal decomposition and reconstruction with M-channel oversampled graph filter banks ($M=4$) is shown in Fig. 2. It is analogous to the traditional filter banks decomposition. The signal passes the first $M/2$ graph filters, and each channel contains $|L|$ signals after the downsampling operation. The other $M/2$ channels keep $|H|$ signals. For the signal reconstruction, the decomposed signals are upsampled and passed the synthesis graph filters. For example, $M=4$, the reconstructed signal $\hat{f}_k$ is represented as

$$\hat{f}_k=\begin{cases}\frac{1}{2}\mathbf{G}_k(\mathbf{I}-\mathbf{J})\mathbf{H}_k f & k=0,1\\ \frac{1}{2}\mathbf{G}_k(\mathbf{I}+\mathbf{J})\mathbf{H}_k f & k=2,3\end{cases}. \tag{8}$$

Therefore, the overall transfer function $\mathbf{T}$ is

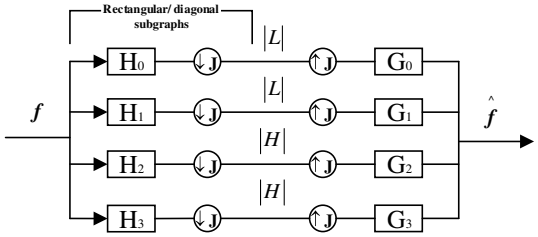

Fig. 2 Oversampled graph filter banks.

$$\mathbf{T} = \frac{1}{2}\sum_{\lambda_i \in \sigma(G)} \sum_{k=0}^{3} g_k(\lambda_i) h_k(\lambda_i) \mathbf{P}_{\lambda_i} + \frac{1}{2}\sum_{\lambda_i \in \sigma(G)} [g_2(\lambda_i) h_2(2-\lambda_i) + g_3(\lambda_i) h_3(2-\lambda_i) - g_0(\lambda_i) h_0(2-\lambda_i) - g_1(\lambda_i) h_1(2-\lambda_i)] \mathbf{P}_{\lambda_i} \mathbf{J}. \tag{9}$$

Then, the perfect reconstruction condition becomes

$$\sum_{k=0}^{3} g_k(\lambda) h_k(\lambda) = 2, \tag{10}$$

$$g_2(\lambda) h_2(2-\lambda) + g_3(\lambda) h_3(2-\lambda) - g_0(\lambda) h_0(2-\lambda) - g_1(\lambda) h_1(2-\lambda) = 0. \tag{11}$$

The M-channel oversampled graph filter banks can be designed with this perfect reconstruction condition.

3 Proposed Method

In this paper, we consider an image as a signal on the regular graph. As the traditional wavelet transform, after decomposing the graph signal (image) with M-channel oversampled graph filter banks, we can obtain the multiscale representation of an image in the graph spectral domain. Then, the fusion weight values are determined using the local property of oversampled graph transform coefficients of the RI and VI. Finally, the fusion image is obtained by applying the inverse transform on the fusion subband coefficients. The proposed RI and VI fusion scheme is shown in Fig. 3.

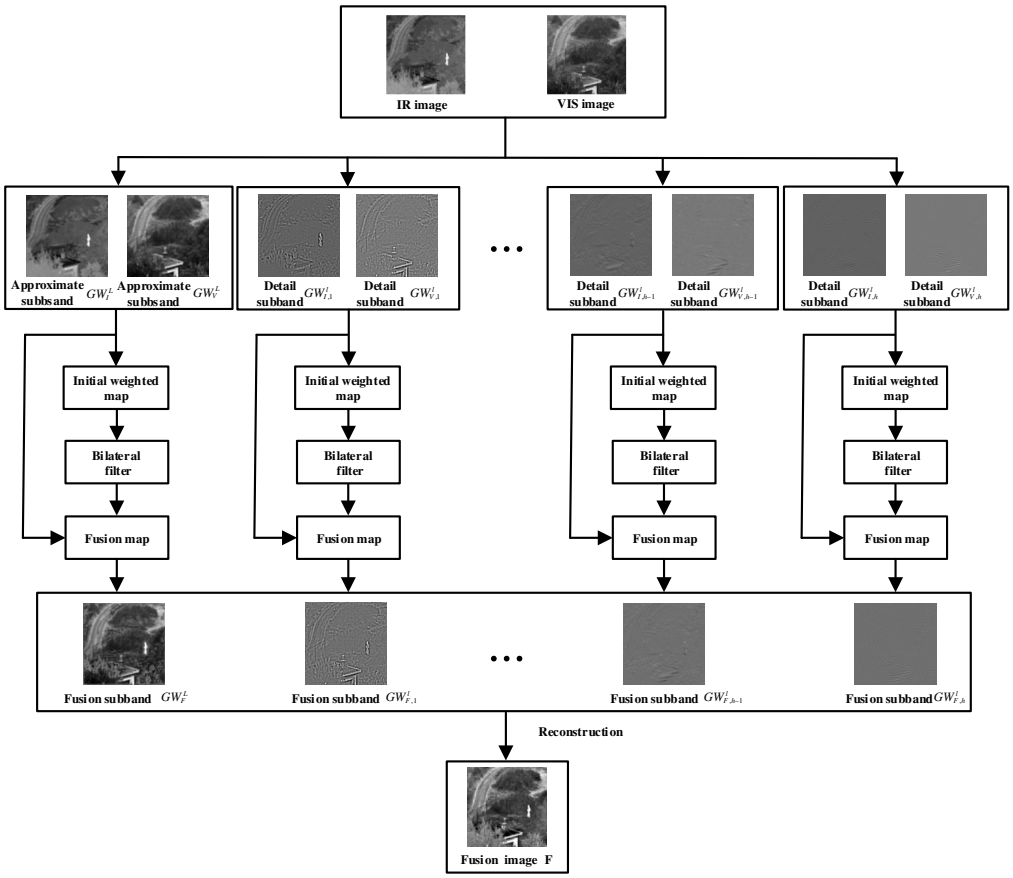

Fig. 3 Proposed RI and VI fusion method.

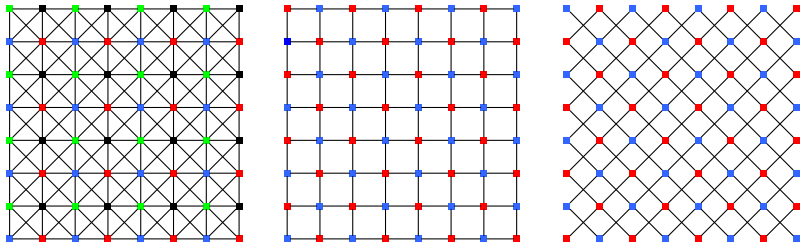

Fig. 4 (a) Image graph. (b) Rectangular bipartite subgraph. (c) Diagonal bipartite subgraph.

3.1 *Multiscale Decomposition with M-Channel Oversampled Graph Filter Banks*

An image is a regular two-dimensional signal. In the graph-based image processing, an image can be viewed as a graph signal by connecting each pixel with its neighboring pixels (edges) and interpreting a pixel values as the function value on a node. In Ref. 25, they use eight-connected graph representation of an image shown in Fig. 4(a). In order to do "separable" transform for an image with graph filter banks, a graph is decomposed into two bipartite subgraphs [Figs. 4(b) and 4(c)]. After filtering the graph signal along one bipartite subgraph with the graph filters, the results are stored in vertices, and filtering operation is applied to the resulting graph signals following the edges of the other bipartite subgraph. Finally, the multiscale decomposition is obtained by applying the graph filter banks in a "cascaded" manner.[27] In this paper, we use M-channel oversampled graph filter banks for the image transform as shown in Fig. 2, in which the filtering and downsampling operations are similar to those of the critically sampled graph wavelet transform. In the implementation of graph filter bank decomposition, we construct the adjacency matrix $\mathbf{A}$ for the bipartite graphs as shown in Fig. 4, whose elements are set as Refs. 25–27. The resulting normalized graph Laplacian matrix is used in the fast transform based on the Chebychev polynomials approximation.

For example, we set $M = 4$ since the image-graph representation is modeled as a four-color graph, the transform will result in 16 subband signals in the first-level decomposition, including one low-frequency subband and 15 high-frequency subbands. Then, in a "cascaded" manner, the low-frequency subband signal will be further decomposed with the oversampled graph filter banks. For the L-level decomposition, there are $15L$ high-frequency subbands and one low-frequency subband. In our image fusion application, the source RI and VI are transformed into multiscale representations with L-level M-channel oversampled graph filter bank decomposition. Their low-frequency subbands are denoted as GW_I^L and GW_V^L. The h'th high-frequency subband at the level l is denoted as $GW_{I,h}^l$ and $GW_{V,h}^l$. Here, I and V represent the RI and VI, respectively. In general, the subbands GW_I^L and GW_V^L contain the low-frequency contents of the images, whereas the subbands $GW_{I,h}^l$ and $GW_{V,h}^l$ contain middle- and high-frequency contents, such as edges and contours.

3.2 *Fusion Rules*

Yan et al.[11] applied the unsharp filter to both original RI and VI and then determined the initial weighed values by comparing the filtered RI and VI. The initial weight matrix is smoothed with the bilateral filter to construct the final weighted map. The fusion rule has been shown the potential in the image fusion. However, this rule is formed with unsharp filtered signals instead of the graph wavelet transform coefficients. In this paper, we will construct the fusion weighted map using the local properties of the transform coefficients after M-channel oversampled filter banks decomposition.

The low-frequency subband contains approximate coefficients that are the reflection of the energy distribution of an image. In order to determine the fusion weight values for the approximate subbands, we can use the choose-Max scheme based on the properties of the transform coefficients, such as the local standard deviation (SD) and the local mean. However, it is well

known that the imaging principle of the RI is different from that of the VI. The former is mainly based on the thermal radiation, and the scene information is reflected in the VI.[13] Because the gray distribution of the same scene is different in the RI and VI, the choose-Max scheme based on the SD or the mean cannot effectively fuse the two kinds of images. Fortunately, in the probability theory and statistics, the coefficient of variation is defined as the ratio of the SD to the mean, which is often used to measure the dispersion of the data. The advantage of the coefficient of variation is that it measures the variance of the data in the context of the mean. However, the coefficient of variation is sensitive to the change of the mean. Therefore, in this paper, we form the initial fusion weight values using the modified LCV as follows:

$$W_I^L(i,j) = \begin{cases} 1, & \mathrm{LCV}_I^L(i,j) \geq \mathrm{LCV}_V^L(i,j) \\ 0, & \text{otherwise} \end{cases}, \tag{12}$$

and

$$W_V^L(i,j) = \begin{cases} 1, & \mathrm{LCV}_I^L(i,j) < \mathrm{LCV}_V^L(i,j) \\ 0, & \text{otherwise} \end{cases}, \tag{13}$$

where $W_I^L(i,j)$ and $W_V^L(i,j)$ are the fusion weight values for the approximate coefficients $GW_I^L(i,j)$ and $GW_V^L(i,j)$, respectively, and $\mathrm{LCV}_I^L(i,j)$ and $\mathrm{LCV}_V^L(i,j)$ are the corresponding LCVs. They are defined as

$$\mathrm{LCV}_I^L(i,j) = \frac{\mathrm{LSD}_w[GW_I^L(i,j)]}{\mathrm{mean}(GW_I^L)}, \tag{14}$$

$$\mathrm{LCV}_V^L(i,j) = \frac{\mathrm{LSD}_w[GW_V^L(i,j)]}{\mathrm{mean}(GW_V^L)}, \tag{15}$$

where $\mathrm{LSD}_w[GW_I^L(i,j)]$ and $\mathrm{LSD}_w[GW_V^L(i,j)]$ are the local SDs of coefficients of the windowed regions (whose size is w) centered at the position (i,j) in the approximate subbands, respectively. The operator $\mathrm{mean}(\cdot)$ is the mean of that subband. Here, we use the mean of the approximate subband instead of the local mean in the windowed region. The reason is that the local mean in a windowed region may be zero, and its small change may result in larger change of the coefficient of variation. Therefore, we modified the general coefficient of variation with the global mean.

As for the high-frequency subbands, because the local or global mean always approaches to zero, we make use of the local SD-based choose-Max scheme for the initial weight values as follows:

$$W_{I,h}^l(i,j) = \begin{cases} 1.4, & \mathrm{LSD}_w[GW_{I,h}^l(i,j)] \geq \mathrm{LSD}_w[GW_{V,h}^l(i,j)] \\ 0, & \text{otherwise} \end{cases}, \tag{16}$$

and

$$W_{V,h}^l(i,j) = \begin{cases} 1.4, & \mathrm{LSD}_w[GW_{I,h}^l(i,j)] < \mathrm{LSD}_w[GW_{V,h}^l(i,j)] \\ 0, & \text{otherwise} \end{cases}. \tag{17}$$

Here, the weight value $W_{I,h}^l(i,j)$ [or $W_{V,h}^l(i,j)$] at the position (i,j) for the h'th subband at the l'th level is determined by comparing the local SDs of the window regions with size of w. The weight values of the low-frequency subband and the high-frequency subband are different. This is because the high-frequency subbands contain different details. We set the fusion weight value to be 1.4 to highlight these details.

The choose-Max rule selects the maximum value of transform coefficients of RI and VI to determine the transform coefficients of the fusion image, which easily produces dramatic enhancement and causes artifacts in the fusion image.[11] Thus, as the smoothing operation in Ref. 11, we smooth the initial fusion maps with the bilateral filters to form the final fusion map matrices. The details of the bilateral filter can be found in Ref. 11. For the sake of simplicity,

we also use $W_I^L(i,j)$ and $W_V^L(i,j)$ to denote the smoothed fusion maps for the low-frequency subbands and $W_{I,h}^l(i,j)$ and $W_{V,h}^l(i,j)$ for the high-frequency subbands.

3.3 *Image Fusion*

Finally, the low-frequency and high-frequency subbands of the source RI and VI are fused as follows:

$$GW_F^L = W_I^L \times GW_I^L + W_V^L \times GW_V^L, \tag{18}$$

$$GW_{F,h}^l = W_{I,h}^l \times GW_{I,h}^l + W_{V,h}^l \times GW_{V,h}^l. \tag{19}$$

Then, the final fusion image is obtained by applying the inverse transform on the fusion subbands GW_F^L and $GW_{F,h}^l$.

4 Experiments and Discussions

In this section, we conduct the image fusion experiments on eight sets of IR and VR:[32] "Nato camp," "Kaptein_1123," "bench," "lake," "Kaptein_1654," "2_men in front of house," "soldier behind smoke," and "bunker," which are shown in Fig. 5. In order to simplify the oversampled filter bank decomposition, we resize the image so that its width and height are equal. In the following experiments, we make use of the four-channel oversampled graph filter banks designed for image transform by Tanaka and Sakiyama.[27] The decomposition level is determined so that the size of the approximate subband is not smaller than 16×16. For example, for an image with the size of 256×256, the corresponding decomposition level is 4. The other parameter is the window size for the local standard deviation (or LCV), which is set to be 5×5 in all experiments.

It is observed that the target is clear in the RI, and the scene information is reflected in the VI. We will subjectively evaluate whether the proposed method can fuse the two kinds of image contents. Meanwhile, seven fusion quality measures are used to objectively evaluate the fusion performance, which includes SD,[33] average gradient (AG),[33] entropy (EN),[33] spatial frequency (SF),[33] gradient-based fusion performance (Q_G),[34] image structural similarity-based measure (Q_Y),[34] and information theory-based measurement (Q_{MI}).[34] The measure SD is used to evaluate the distribution and contrast of the fusion image. The metric AG measures the characteristics of small details, that is, the contrast and texture changes of the fusion image. The EN represents the amount of information in the fusion image. The metric SF can measure the gradient distribution and reflect the clarity of the fusion image. The measure Q_G evaluates the amount of edge information transmitted from the source image to the fusion image. The metric Q_{MI} is usually used to reflect the preservation of original information in the fusion image in the literature.

In order to demonstrate the effectiveness of the introduced method, we compare it with other RI and VI fusion methods based on cross bilateral filter (CBF),[33] curvelet transform (CT),[12] dual-tree complex wavelet transform (DT-CWT),[35] visual saliency map and weighted least square

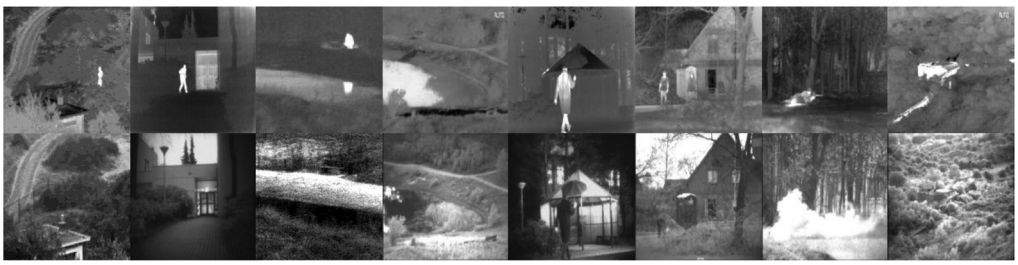

Fig. 5 Eight sets of infrared and visible source images. (a) Nato camp, (b) Kaptein_1123, (c) bench, (d) lake, (e) Kaptein_1654, (f) 2_men in front of house, (g) soldier behind smoke, and (h) bunker.

optimization (VSM-WLS),[36] NSCT,[13] ratio of low-pass pyramid (RP),[37] multiresolution SVD (MSVD),[38] local edge-preserving multiscale decomposition (LEP),[15] multiscale directional non-local means filter (MDNLM),[14] discrete cosine transform and local SF in discrete stationary wavelet transform-domain (SWT-DCT),[39] and SGWT.[11] For those methods, the parameters are in accordance with the requirements of the corresponding references. The fusion results are shown in Figs. 6–13.

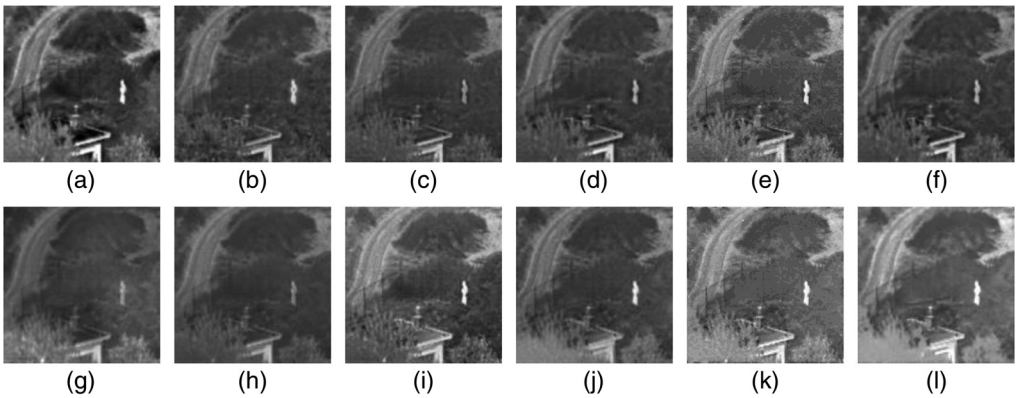

Fig. 6 Fusion results of images "Nato camp." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

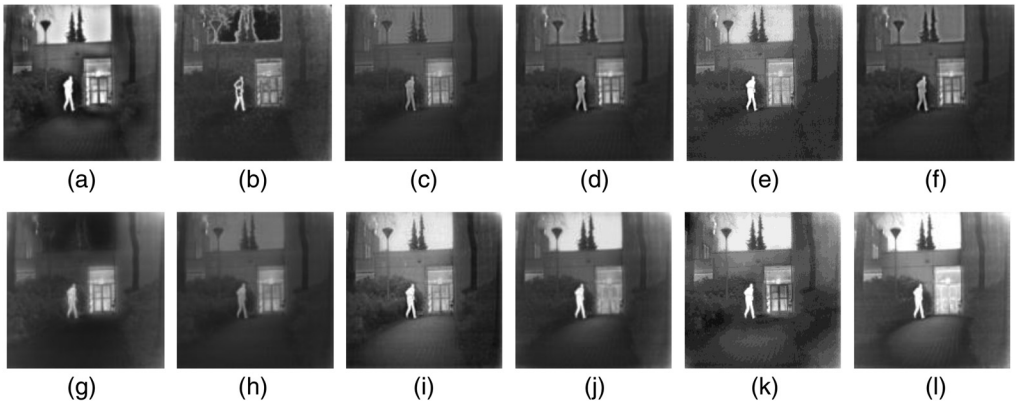

Fig. 7 Fusion results of images "Kaptein_1123." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

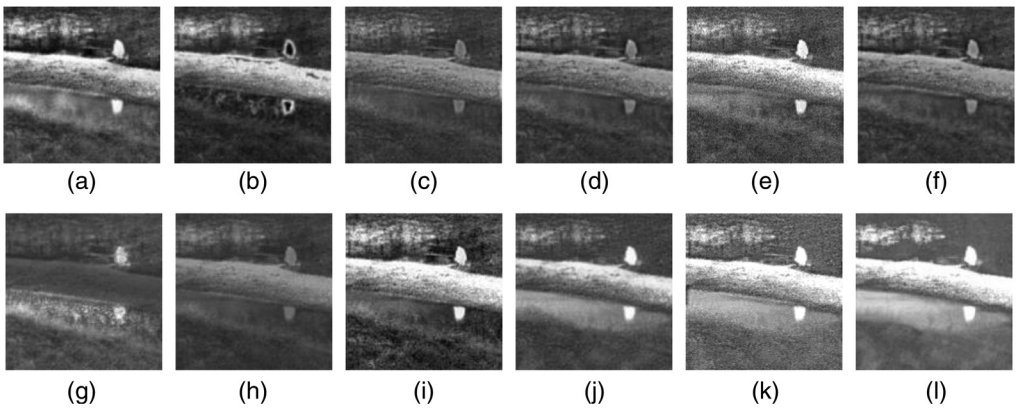

Fig. 8 Fusion results of images "bench." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

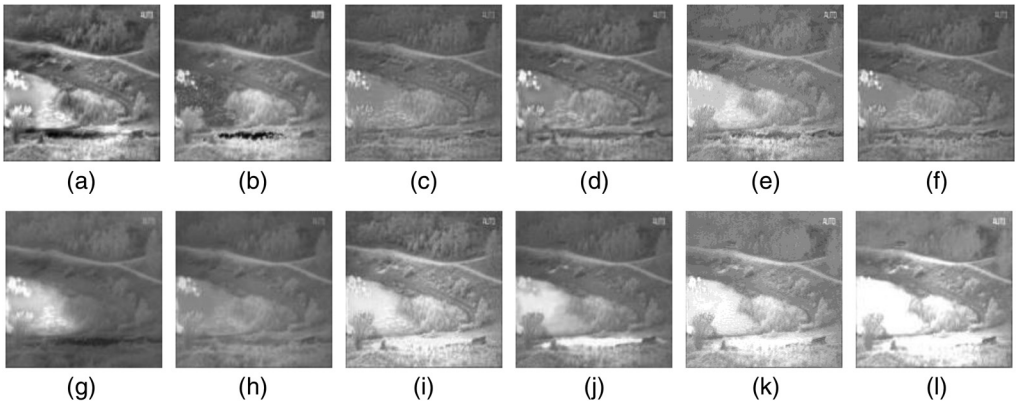

Fig. 9 Fusion results of images "lake." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

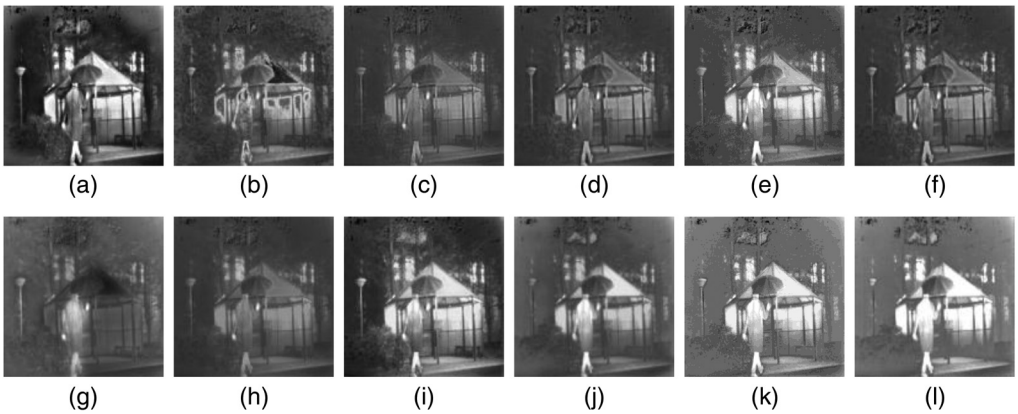

Fig. 10 Fusion results of images "Kaptein_1654." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

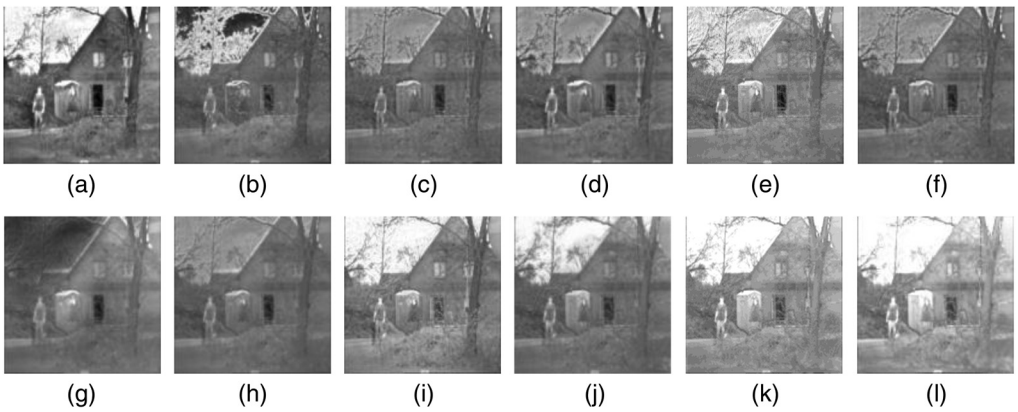

Fig. 11 Fusion results of images "2_men in front of house." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

One aim of the image fusion is to provide an image, from which the human visual perception system can perceive the information captured by the infrared and visible light sensors. Therefore, we subjectively evaluate the fusion results from the following two aspects: (1) whether the interesting object is prominent in the fusion image and (2) whether the detail background information

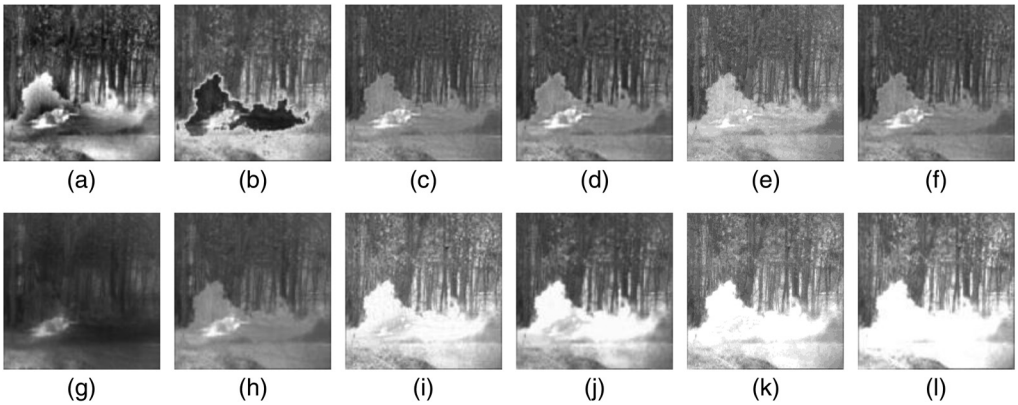

Fig. 12 Fusion results of images "soldier behind smoke." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

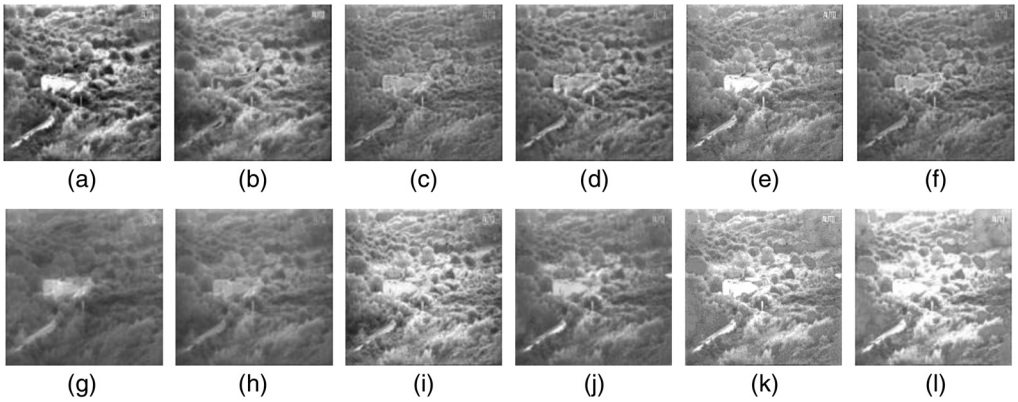

Fig. 13 Fusion results of images "bunker." (a) Proposed method, (b) CBF, (c) CT, (d) DT-CWT, (e) VSM-WLS, (f) NSCT, (g) RP, (h) MSVD, (i) LEP, (j) MDNLM, (k) SWT-DCT, and (l) SGWT.

of the VI has been transferred to the fusion image. Due to the limitations of the subjective evaluation, we also evaluate the performance with the objective measures that are listed in Tables 1–8.

From Fig. 6, it can be seen that the target in Figs. 6(a), 6(e), and 6(i)–6(l) are well highlighted. However, focusing on the background region, we can find that the fusion results in Figs. 6(a) and 6(i) are better than those in other figures. As the objective evaluations shown in Table 1, our proposed method has the best fusion quality indicators SD, AG, EN, and SF. The Q_G and Q_Y values are larger than those of the method CBF, CT, VSM-WLS, RP, MSVD, MDNLM, and SWT-DCT, but the Q_{MI} value is smaller than that of RP, LEP, MDNLM, SWT-DCT, and SGWT.

As for the test images Kaptein_1123, the fusion results (Fig. 7) show that the sky regions in Figs. 7(a), 7(e), and 7(i)–7(l) are more similar with that in the original VI. However, in the top-left part of Figs. 7(j) and 7(l), the smoke region is unnoticeable, which is clear in the RI. It can be found from Table 2 that the AG and SF values of our method, SD and EN values of the method LEP, Q_G, Q_Y, and Q_{MI} values of SGWT are the largest in the corresponding measures. For our method, the SD value is smaller than those of LEP and SGWT, the EN value is slightly smaller than that of LEP, and the Q_G, Q_Y, and Q_{MI} values lie in the medium level.

It is observed from Fig. 8 that the target regions of Figs. 8(a), 8(e), and 8(i)–8(l) have been successfully fused into the resulting images. However, the texture regions of Figs. 8(e) and 8(j)–8(l) are not clear enough. Therefore, the fusion images of our method and LEP are better than other methods using the subjective evaluation. In Table 3, the measures Q_G, SD, AG, EN, and SF of our method are higher than those of LEP, whereas the measure Q_Y is lower than that of LEP. The Q_G value of our method is 0.6532 that is slightly smaller than that of DT-CWT and NSCT.

Table 1 Performance of different fusion methods on images "Nato camp."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.4655	0.7781	**40.7815**	**11.9400**	**7.2381**	0.2584	**16.3335**
CBF	0.4127	0.6535	26.7411	9.1421	6.4668	0.2365	12.1136
CT	0.4109	0.7294	25.0080	8.4913	6.4095	0.2307	11.2633
DT-CWT	0.4710	0.7706	26.6405	8.2878	6.5006	0.2374	11.5577
VSM-WLS	0.4437	0.7573	28.3654	9.6169	6.6112	0.2479	12.6808
NSCT	**0.5111**	0.8152	27.5181	8.5579	6.5678	0.2388	11.8229
RP	0.3688	0.7012	25.0935	6.3788	6.5422	0.2833	8.3920
MSVD	0.3933	0.6947	23.1825	8.2080	6.2940	0.2426	10.6994
LEP	0.4734	0.7789	36.0358	10.5094	6.9907	0.3269	13.9698
MDNLM	0.4244	0.7593	28.7867	8.6895	6.6365	0.3030	11.2437
SWT-DCT	0.4512	0.7612	30.6780	9.2804	6.7878	0.3266	12.1712
SGWT	0.5060	**0.9233**	34.9510	8.6948	6.9560	**0.6403**	11.7717

Note: The bold characters are used to show the best method under different evaluation measures.

Table 2 Performance of different fusion methods on images "Kaptein_1123."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.5137	0.7938	56.4023	**9.8101**	7.2554	0.2725	**13.4831**
CBF	0.4295	0.6943	34.1209	7.9993	6.8122	0.3108	11.5944
CT	0.4603	0.7723	32.8501	7.0809	6.6919	0.2683	9.3300
DT-CWT	0.5144	0.8060	34.0834	6.8952	6.7264	0.2629	9.6723
VSM-WLS	0.4944	0.7773	49.1963	8.3616	6.9621	0.3018	11.3519
NSCT	0.5479	0.8244	34.4099	7.0033	6.7327	0.2661	9.7881
RP	0.3335	0.7000	35.3903	4.8129	6.8651	0.3896	6.4829
MSVD	0.2322	0.5220	31.4118	4.3951	6.4997	0.3157	6.3684
LEP	0.5179	0.7323	**57.9370**	8.8251	**7.3660**	0.3989	11.8429
MDNLM	0.5083	0.8256	51.8953	7.4062	7.1441	0.4325	9.8044
SWT-DCT	0.5429	0.8262	51.3894	7.4353	7.1977	0.5163	10.1113
SGWT	**0.5523**	**0.8808**	56.9271	7.3426	7.0043	**0.7465**	10.3917

Note: The bold characters are used to show the best method under different evaluation measures.

As for the measure Q_{MI}, our method is only better than the method CT, DT-CWT, NSCT, and RP and worse than other methods.

As for the experimental results in Fig. 9, it is visually observed that our method, VSM-WLS, LEP, MDNLM, and SWT-DCT result in good fusion images. However, from the original images in Fig. 5(d), it is known that the water surface region and the ground region without grass are captured as the high gray level (white) and the low gray level (black) in the RI, respectively, whereas those regions are captured as opposite gray levels in the VI. This requires these two regions should exhibit as two different regions with opposite gray levels in the fused image. Among the above-mentioned methods, our method and VSM-WLS have the desired results. In Table 4, the measures Q_G, Q_Y, SD, AG, EN, and SF of our method are higher than those

Table 3 Performance of different fusion methods on images "bench."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.6532	0.8126	**61.2348**	**23.1286**	**7.4972**	0.2788	**28.4684**
CBF	0.6245	0.8359	55.0879	17.8081	7.3396	0.5358	21.5537
CT	0.6243	0.8408	32.2942	16.9260	6.8232	0.2700	20.3698
DT-CWT	0.6677	0.8568	34.0147	16.9518	6.9093	0.2634	20.6400
VSM-WLS	0.5883	0.8242	47.8256	19.4471	7.1366	0.3414	23.6843
NSCT	**0.6748**	0.8582	34.8013	16.9894	6.9360	0.2627	20.6819
RP	0.3517	0.6326	28.4179	14.8498	6.5059	0.1788	21.4595
MSVD	0.4509	0.7586	29.9328	14.7981	6.6529	0.2990	18.0920
LEP	0.5179	**0.8810**	57.9370	8.8251	7.3660	0.3989	11.8429
MDNLM	0.5775	0.7973	46.7913	16.8127	7.1893	0.3638	20.3705
SWT-DCT	0.6283	0.8013	48.1163	17.4978	7.1861	0.4379	21.2311
SGWT	0.4304	0.8625	49.4514	12.0254	6.6329	**0.6562**	17.8347

Note: The bold characters are used to show the best method under different evaluation measures.

Table 4 Performance of different fusion methods on images "lake."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.5638	0.8604	45.8461	**10.3952**	7.3756	0.2008	**15.3864**
CBF	0.5459	0.8141	38.8605	8.5613	7.1599	0.4566	13.9942
CT	0.5510	0.8692	27.1897	7.6428	6.6777	0.2267	12.3409
DT-CWT	**0.5943**	**0.8946**	28.9053	7.7254	6.7363	0.2142	12.1445
VSM-WLS	0.4895	0.8117	35.8123	8.5897	7.0036	0.2946	12.7529
NSCT	0.5909	0.8848	27.6341	7.7161	6.6831	0.2357	12.3026
RP	0.2927	0.6659	34.0238	4.1508	6.5450	0.2680	6.2892
MSVD	0.4104	0.7529	24.8118	6.7510	6.5662	0.2731	10.7757
LEP	0.5859	0.8766	**47.7822**	9.9929	**7.4634**	0.4463	14.0586
MDNLM	0.4531	0.8202	38.3313	7.1563	6.9675	0.3688	9.6027
SWT-DCT	0.5335	0.8380	42.3720	7.6761	7.0669	0.4753	12.1724
SGWT	0.4951	0.7968	43.2152	6.2915	6.3294	**0.6778**	11.1757

Note: The bold characters are used to show the best method under different evaluation measures.

of the method VSM-WLS. However, for our method, there are only two measures AG and SF that are better than other methods, and the SD and EN values are smaller than those of LEP. The measure Q_{MI} value is the smallest one.

From the fusion results in Fig. 10, we observed that the fused images in Figs. 10(a), 10(e), and 10(i) are the desired results. Although the measures AG, EN, and SF of our method are better than those of the method LEP (in Table 5), the regions in the top left and top right corners of our fused image are worse than those of the method LEP. Meanwhile, all measures of our method are better than those of the method VSM-WLS. In the objective evaluation Table 5, we can found that the AG, EN, and SF of our method are the best among those methods. Except the measure Q_{MI} and Q_Y, the values of Q_G and SD of the proposed method are close to the best one.

Table 5 Performance of different fusion methods on images "Kaptein_1654."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.5603	0.8140	56.1405	**11.4458**	**7.3978**	0.2532	**16.4752**
CBF	0.4678	0.5526	36.9554	10.3881	6.9058	0.2802	13.8107
CT	0.4794	0.7605	28.9725	7.9342	6.4299	0.2568	11.1371
DT-CWT	0.5460	0.8008	31.2598	8.0248	6.5029	0.2416	11.6469
VSM-WLS	0.5227	0.7764	41.9062	9.2361	6.7328	0.3033	12.6820
NSCT	**0.5789**	0.8229	31.9674	8.1839	6.5161	0.2434	11.8240
RP	0.3123	0.6388	24.8571	6.3056	6.4613	0.2126	8.4488
MSVD	0.4110	0.7108	27.5918	7.4549	6.3507	0.2964	10.4817
LEP	0.5619	0.8076	**56.7933**	10.3504	7.2529	0.4297	14.4504
MDNLM	0.5194	0.7991	38.2242	8.1894	6.8188	0.3760	11.2160
SWT-DCT	0.5585	0.7970	40.3506	8.5293	6.7529	0.4197	12.2101
SGWT	0.5449	**0.8841**	43.1867	7.8864	6.5246	**0.6885**	12.3793

Note: The bold characters are used to show the best method under different evaluation measures.

Table 6 Performance of different fusion methods on images "2_men in front of house."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.5289	0.8166	**54.2621**	**11.1017**	**7.1707**	0.2404	**17.0527**
CBF	0.4928	0.6919	40.0314	10.8112	7.0327	0.3765	16.9928
CT	0.5104	0.8053	28.3749	8.4204	6.6902	0.2068	13.0977
DT-CWT	0.5470	0.8344	30.3315	8.4048	6.8017	0.2039	13.2038
VSM-WLS	0.4807	0.7748	38.9201	9.9665	6.9906	0.2680	14.5957
NSCT	**0.5680**	**0.8455**	28.4700	8.5607	6.7157	0.2140	13.2749
RP	0.3049	0.7070	28.2526	4.7685	6.7850	0.2702	7.0767
MSVD	0.4135	0.7284	24.0750	7.4540	6.4788	0.2635	11.5649
LEP	0.5591	0.8333	51.3861	10.7834	7.0652	0.3767	15.5299
MDNLM	0.4646	0.7910	43.4567	7.8148	6.8795	0.3465	10.7854
SWT-DCT	0.5198	0.7983	46.0848	8.5616	6.8834	0.4362	13.4468
SGWT	0.4612	0.8325	42.0006	6.5424	6.3515	**0.6793**	11.4151

Note: The bold characters are used to show the best method under different evaluation measures.

By checking the sky region and the person target, it can be seen that the fusion images in Figs. 11(a), 11(e), and 11(i)–11(l) are better than those of other methods. However, the window regions near the tree of the fusion images of the methods VSM-WLS, MDNLM, SWT-DCT, and SGWT are not clear. Therefore, the fusion results of our method and LEP are competitive. However, as listed in Table 6, the SD, AG, EN, and SF of our method are better than those of other methods. As for the measure Q_G and Q_Y, NSCT is the best method among these fusion techniques, and LEP is better than our introduced method.

For the set of images "soldier behind smoke," the fusion results are shown in Fig. 12. It is demonstrated from Fig. 12(a) that the target (the person with the gun behind the smoke) is clear in our fusion image, and the background, such as the smoke and the texture of the trees, is well

Table 7 Performance of different fusion methods on images "soldier behind smoke."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.5773	0.8146	55.7168	**13.3780**	**7.6546**	0.1792	**18.2730**
CBF	0.5053	0.7702	51.6231	11.3968	7.5529	0.4117	17.0591
CT	0.5604	0.8337	37.4487	9.6787	7.0688	0.2470	14.2475
DT-CWT	**0.6104**	**0.8596**	38.8829	9.7190	7.1289	0.2361	14.0594
VSM-WLS	0.4833	0.7685	42.0497	10.1970	7.1773	0.2818	14.0243
NSCT	0.5971	0.8475	37.9667	9.7993	7.0976	0.2484	14.2437
RP	0.2154	0.6552	24.9352	3.7946	6.4484	0.2509	5.7607
MSVD	0.4161	0.7236	34.8551	8.5153	6.9543	0.2951	12.5497
LEP	0.5865	0.8351	56.8321	12.5389	7.4159	0.4013	17.0721
MDNLM	0.5107	0.8141	**59.7291**	9.3762	7.2878	0.3952	12.9669
SWT-DCT	0.5648	0.8332	59.0421	9.7144	7.1830	0.5266	14.6920
SGWT	0.5364	0.8009	56.6775	8.6970	6.0439	**0.7337**	14.3637

Note: The bold characters are used to show the best method under different evaluation measures.

Table 8 Performance of different fusion methods on images "bunker."

Metrics	Q_G	Q_Y	SD	AG	EN	Q_{MI}	SF
Proposed	0.6289	0.9021	**48.1053**	**15.7203**	7.5168	0.1762	**19.1694**
CBF	0.5872	0.9041	40.4572	11.2998	7.3211	0.4573	14.4357
CT	0.5680	0.9066	30.7614	11.3814	6.9330	0.1845	14.1507
DT-CWT	0.6301	0.9360	34.2209	11.6106	7.0867	0.1904	14.4819
VSM-WLS	0.5040	0.8661	35.9489	12.2201	7.1098	0.2473	15.3172
NSCT	0.6202	0.9301	31.6366	11.5364	6.9732	0.1967	14.4787
RP	0.2322	0.6407	24.6557	5.5882	6.5538	0.1473	7.2019
MSVD	0.2795	0.6521	25.7477	8.1649	6.7115	0.2036	10.4611
LEP	**0.6338**	**0.9327**	46.3427	15.1125	**7.5198**	0.3727	18.2443
MDNLM	0.4670	0.8482	31.2678	10.5562	6.9148	0.2669	12.4671
SWT-DCT	0.5954	0.9004	38.5976	11.6275	7.1578	0.4580	14.6297
SGWT	0.5324	0.8663	40.1099	10.1809	6.8083	**0.5633**	13.7700

Note: The bold characters are used to show the best method under different evaluation measures.

fused into the final result. The results with VSM-WLS, LEP, MDNLM, and SWT-DCT [Figs. 12(e) and 12(i)–12(k)] have good background information received from the VI. However, the targets in the two fusion images are hard to be identified from the background. Therefore, our method is the best one among all algorithms under the subjectively evaluation. From the objective evaluation shown in Table 7, it can be seen that AG, EN, and SF of our method are better than those of other methods. The ranks of both Q_G and SD of our method are the fourth. The measure Q_Y value lies in the medium level.

In our fusion result [Fig. 13(a)] for the set of images "bunker," the target "bunker" can be clearly identified from a complex background, and the jungle around the target bunker is also clearly visible, which are also achieved in the methods VSM-WLS and SWT-DCT. In other

methods, either the target is not distinct from the background or the texture background is not well fused into the final image. From Table 8, it has been shown that the measures SD, AG, and SF of the proposed method are better than those of other approaches. The Q_G, Q_Y, and EN values are close to the best measures that are achieved by the method LEP. However, as for the measure Q_{MI}, our method is still worse than other method except for the method RP.

From Tables 1–8, it has been shown that the measures Q_G, Q_Y, and Q_{MI} of our method are not the best, especially for the evaluation measure Q_{MI}. In order to further evaluate our method, we conduct the experiments on images "bunker" using different decomposition levels (level = 1, level = 3, and level = 5) and changing the initial weight values of Eqs. (16) and (17) to 1.0. The fusion results are given in Fig. 14. Meanwhile, for comparison and explanation, we assume a special case that the infrared (visible) image is not fused into the visible (infrared) image. Then, the fusion image is the original visible (infrared) image. We also calculate their evaluation measure values listed in Table 9. The measure value Q_{MI} of VI is larger than those of all methods in Tables 8 and 9. Therefore, it may be reasonable that this measure should be improved for evaluating the fusion methods, which is out of the scope of this paper. In addition, from Fig. 14, we find the fusion image in Fig. 14(c) is not better than those in Figs. 14(d) and 14(e). However, its corresponding Q_{MI} value is the largest among the three fusion results. The reason is that the change (fusion operation) of each transform coefficient may affect a region of the fusion image, because the reconstruction is performed by applying the synthesis graph filter bank, which is also demonstrated by other multiscale transform-based fusion methods, such as DT-CWT. Meanwhile, for other measures SD, AG, EN, and SF, their values also decrease with the increase of the decomposition level. Therefore, initial weight values of Eqs. (16) and (17) are set to be the empirical value 1.4 that results in good fusion performance with the subjective and objective evaluation.

In order to show the computational complexity of the proposed method, the average running time is listed in Table 10 for the experimental images shown in Fig. 5. In general, the computational complexity is higher than other methods. The reason is that the image decomposition with oversampled graph filter bank results in more subbands than other multiscale transform.

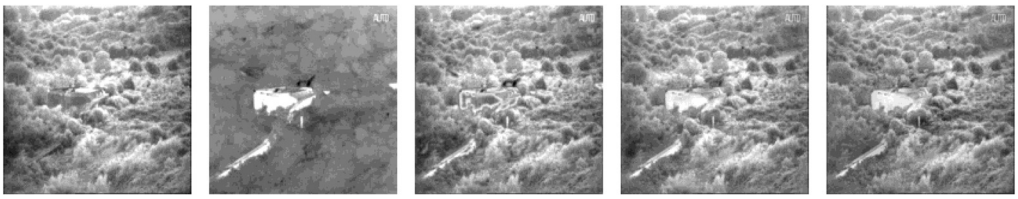

Fig. 14 (a) Original VI, (b) original RI, (c) fusion image with level = 1, (d) fusion image with level = 3, and (e) fusion image with level = 5.

Table 9 Performance analysis of the evaluation measures of our method.

Metrics	VI	RI	Level = 1	Level = 3	Level = 5
Q_G	0.7037	0.3606	0.6484	0.6576	0.6585
Q_Y	1.0000	0.9999	0.9336	0.9547	0.9489
SD	45.5734	29.8983	42.4140	41.8312	38.5161
AG	11.5237	4.1519	11.8201	11.5725	11.5494
EN	7.4796	6.5308	7.4135	7.3673	7.2486
Q_{MI}	1.0213	1.0213	0.5927	0.4366	0.2137
SF	14.3546	6.2188	14.6698	14.4286	14.3290

Table 10 Running time of the proposed method.

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Size	256×256	448×448	256×256	576×576	448×448	576×576	576×576	576×576
Time(s)	51.6	162.2	52.3	266.9	160.5	261.7	261.5	260.4

In this paper, an image is considered as a signal on graph by connecting each pixel (vertex) in the image to its adjacent pixels (vertices) and defining the pixel value as the value of the graph signal. The oversampled graph wavelet filter bank is used to form the multiscale components of a graph signal (image), in which the pixel values and the relationship among pixels are explicitly explored. Therefore, this multiscale representation can describe the image very well. In addition, we also introduce the modified LCV to determine the fusion map. It is reasonable that the proposed method performs better than some transform-based algorithms. In the experiments, we evaluate the performance of the fusion methods by first determining the good fusion images with the visual perception (subjective evaluation) and then evaluating the acceptable fusion images with the objective measures. Using the subjective evaluation on the fusion results of eight sets of RI and VI, it is observed that the proposed method can well balance the intensity between the RI and VR and retain the information in the source images as much as possible. As can be seen from fusion quality indicators in Tables 1–8, the proposed method provides the best AG and SF and relatively good SD and EN for all test images. This means that the fusion image with our method has higher contrast, sharpness, and texture information. Therefore, our method has good performance and has been shown potential in the image fusion field.

5 Conclusions

Graph signal processing is one of the emerging fields in signal processing. By representing an image as the signal on a graph, in the context of graph signal processing, we can explicitly explore the pixel values and the relationship among the pixels. Therefore, the graph wavelet filter banks combine the advantages of the classic wavelet filter banks and graph representation of a signal. Based on this observation, this paper proposes an improved image fusion method based on M-channel oversampled graph filter banks, in which the fusion rule is determined using the local properties based on the SD in each subband. We conduct the experiments on eight sets of RI and VI to evaluate the performance. The results demonstrate that our method provides a much better fusion image for human visual perception. In addition, the objective measures also show that the proposed method can effectively combine different information in RI and VI.

Acknowledgments

The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the paper. This research was funded by the National Natural Science Foundation of China (Grant No. 61871142) and the Fundamental Research Funds for the Central Universities (Grant No. 2572018BF13). There is no conflict of interest.

References

1. J. Swoger et al., "Multi-view image fusion improves resolution in three-dimensional microscopy," *Opt. Express* **15**(13), 8029–8042 (2007).
2. K. Zhan et al., "Multimodal image seamless fusion," *J. Electron. Imaging* **28**(2), 023027 (2019).
3. C. Berger et al., "Multi-modal and multi-temporal data fusion: outcome of the 2012 GRSS data fusion contest," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**(3), 1324–1340 (2013).

4. K. Zhan et al., "Multifocus image fusion using phase congruency," *J. Electron. Imaging* **24**(3), 033014 (2015).
5. K. Ma et al., "Robust multi-exposure image fusion: a structural patch decomposition approach," *IEEE Trans. Image Process.* **26**(5), 2519–2532 (2017).
6. A. A. Dogra, B. Goyal, and S. Agrawal, "From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications," *IEEE Access* **5**, 16040–16067 (2017).
7. X. Jin et al., "A survey of infrared and visual image fusion methods," *Infrared Phys. Technol.* **85**, 478–501 (2017).
8. J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: a survey," *Inf. Fusion* **45**, 153–178 (2019).
9. H. Xu et al., "Infrared and multi-type images fusion algorithm based on contrast pyramid transform," *Infrared Phys. Technol.* **78**, 133–146 (2016).
10. Y. Ben-Shoshan and Y. Yitzhaky, "Improvements of image fusion methods," *J. Electron. Imaging* **23**(2), 023021 (2014).
11. X. Yan et al., "Infrared and visible image fusion with spectral graph wavelet transform," *J. Opt. Soc. Am. A* **32**, 1643–1652 (2015).
12. Z. Shao, J. Liu, and Q. Cheng, "Fusion of infrared and visible images based on focus measure operators in the curvelet domain," *Appl. Opt.* **51**, 1910–1921 (2012).
13. J. Adu et al., "Image fusion based on nonsubsampled contourlet transform for infrared and visible light image," *Infrared Phys. Technol.* **61**, 94–100 (2013).
14. X. Yan et al., "Infrared and visible image fusion using multiscale directional nonlocal means filter," *Appl. Opt.* **54**, 4299–4308 (2015).
15. X. Zhang et al., "Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition," *J. Opt. Soc. Am. A* **34**, 1400–1410 (2017).
16. J. Wang et al., "Fusion method for infrared and visible images by using non-negative sparse representation," *Infrared Phys. Technol.* **67**, 477–489 (2014).
17. H. Li and X. Wu, "Densefuse: a fusion approach to infrared and visible images," *IEEE Trans. Image Process.* **28**(5), 2614–2623 (2019).
18. D. Liu et al., "Infrared and visible image fusion based on convolutional neural network model and saliency detection via hybrid l0-l1 layer decomposition," *J. Electron. Imaging* **27**(6), 063036 (2018).
19. W. Kong, Y. Lei, and H. Zhao, "Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization," *Infrared Phys. Technol.* **67**, 161–172 (2014).
20. Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion* **24**, 147–164 (2015).
21. J. Ma et al., "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion* **31**, 100–109 (2016).
22. M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Twenty-Second Annu. Joint Conf. IEEE Comput. and Commun. Soc.*, San Francisco, pp. 1848–1857 (2003).
23. M. Jansen, G. P. Nason, and B. W. Silverman, "Multiscale methods for data on graphs and irregular multidimensional situations," *J. R. Stat. Soc. Ser. B* **71**, 97–125 (2009).
24. D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.* **30**, 129–150 (2011).
25. S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.* **60**, 2786–2799 (2012).
26. S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.* **61**, 4673–4685 (2013).
27. Y. Tanaka and A. Sakiyama, "M-channel oversampled perfect reconstruction filter banks for graph signals," in *IEEE Int. Conf. Acoust., Speech and Signal Process.*, Florence, Italy, pp. 2623–2627 (2014).
28. D. B. H. Tay and J. Zhang, "Techniques for constructing biorthogonal bipartite graph filter banks," *IEEE Trans. Signal Process.* **63**, 5772–5783 (2015).
29. J. Z. Jiang, F. Zhou, and P. L. Shui, "Lifting-based design of two-channel biorthogonal graph filter bank," *IET Signal Process.* **10**, 670–675 (2016).

30. O. Teke and P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs part I: fundamentals," *IEEE Trans. Signal Process.* **65**, 409–422 (2017).
31. S. K. Narang, "Critically sampled wavelet filterbanks on graphs," PhD Dissertations, University of Southern California (2012).
32. T. Alexander, "The TNO multiband image collection," figshare. Collection (2017).
33. B. K. Shreyamsha Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal Image Video Process.* **9**, 1193–1204 (2015).
34. Z. Liu et al., "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 94–109 (2012).
35. J. X. Huang, "Image fusion algorithm of adaptive simplified PCNN based on dual-tree complex wavelet transform," in *Int. Conf. Comput. Sci. and Environ. Eng.*, Science and Engineering Research Center (2015).
36. J. Ma et al., "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Phys. Technol.* **82**, 8–17 (2017).
37. T. Alexander, "Image fusion by a ratio of low-pass pyramid," *Pattern Recognit. Lett.* **9**, 245–253 (1989).
38. V. P. S. Naidu, "Image fusion technique using multi-resolution singular value decomposition," *Def. Sci. J.* **61**, 479–484 (2011).
39. X. Jin et al., "Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain," *Infrared Phys. Technol.* **88**, 1–12 (2018).

Chunyan Song received her MS and PhD degrees from Harbin Engineering University, Harbin, China, in 2004 and 2010, respectively. Currently, she is an associate professor at Northeast Forestry University. Her current research interests include image processing, chaos theory, and communication technique.

Xueying Gao received her BEng degree in information and communication engineering from Northeast Forestry University. Currently, she is an MS degree candidate at Harbin Engineering University, China. Her current research interests include graph signal processing, image processing, and image fusion.

Yulong Qiao received his BS and MS degrees in applied mathematics and his PhD from Harbin Institute of Technology, Harbin, China. He is a professor at Harbin Engineering University. His current research interests include texture analysis, image and video processing, graph signal processing, and pattern recognition.

Kaige Zhang received his BEng degree from Harbin Institute of Technology and his MS degree from Harbin Engineering University, Harbin, China. Currently, he is a PhD candidate at Utah State University, Logan, Utah, USA. His current research interests include image processing and computer vision.