CrossMark

Image fusion based on simultaneous empirical wavelet transform

Xiaoli Zhang[1,2] · Xiongfei Li[1,2] · Yuncong Feng[1,2]

Abstract In this paper, a new multi-scale image fusion algorithm for multi-sensor images is proposed based on Empirical Wavelet Transform (EWT). Different from traditional wavelet transform, the wavelets of EWT are not fixed, but the ones generated according to the processed signals themselves, which ensures that these wavelets are optimal for processed signals. In order to make EWT can be used in image fusion, Simultaneous Empirical Wavelet Transform (SEWT) for 1D and 2D signals are proposed, by which different signals can be projected into the same wavelet set generated according to all the signals. The fusion algorithm constructed on the 2D SEWT contains three steps: source images are decomposed into a coarse layer and a detail layer first; then, the algorithm fuses detail layers using maximum absolute values, and fuses coarse layers using the maximum global contrast selection; finally, coefficients in all the fused layers are combined to obtain the final fused image using 2D inverse SEWT. Experiments on various images are conducted to examine the performance of the proposed algorithm. The experimental results have shown that the fused images obtained by the proposed algorithm achieve satisfying visual perception; meanwhile, the algorithm is superior to other traditional algorithms in terms of objective measures.

Keywords Image fusion · Empirical wavelet transforms · Data driven · Simultaneous image decomposition

1 Introduction

Image fusion is an effective tool in many theoretical research and practical applications [1, 2, 9, 15]. Because of the limitation of imaging mechanism, any single imaging sensor can hardly

✉ Xiongfei Li
xiongfei@jlu.edu.cn

1   Key Laboratory of Symbol Computation and Knowledge Engineer of Ministry of Education, Jilin
     University, Changchun 130012, China

2   College of Computer Science and Technology, Jilin University, Changchun 130012, China

 Springer

provide a comprehensive presentation of a scene. Consequently, image fusion can be employed to combine images captured by several sensors from the same scene to generate an informative image. The attractive feature of image fusion is that the fused image is more valuable than the sum of images, because image fusion can abstract complementary information from each source image and eliminate redundant information.

Recent years have witnessed an increase in the number of image fusion algorithms, among which multi-scale image fusion accounts for a large percentage. Such kind of fusion algorithm decomposes source images into several layers with different morphological components first and then designs different fusion rules for different layers; finally fused coefficients are reconstructed to get the fused image. Until now, many multi-scale image presentation models have been utilized in image fusion. Pyramid family (including Laplacian, contrast, ratio pyramid and the like) are widely used in early stage. Then wavelets are brought into this field [9, 15]. As wavelets are well established in the scientific community with strict theoretical foundation and get more directional information, they perform better than Pyramid based algorithm. Recently, several geometric extensions of wavelets are introduced into image fusion, such as Ripplet transform [2], Ridgelet transform [1], Shearlet Transform [14], Curvelet transform [23] and Contourlet transform [10, 22].

However, multi-scale image representation models use fixed filters for any image, which makes it impossible to guarantee that the corresponding filters are the optimal ones to represent an image. A better alternative is adaptive image representation where the filters are generated according to the characteristics of images. In [7], Huang et al. proposed a new signal analysis method based on empirical mode decomposition (EMD) where the aim of the method is to find narrow band oscillatory components called intrinsic mode functions (IMFs). The EMD method is fully driven by signals, and it needs no prior assumptions. The method has been applied to image fusion [6, 13]. However, there are two basic difficulties in applying EMD on this field. The first one is its lack of mathematical theory; and the second one is that it does not ensure that decompositions of different source images are matched (the corresponding sub-bands of all the source images share the same morphological components), which makes it meaningless for multi-scale comparison [13]. In literature [3], the authors presented a new signal analysis method called empirical wavelet transform (EWT). It can separate input signals into amplitude modulated-frequency modulated (AM-FM) components by building an adaptive wavelet tight frame. Compared with EMD, EWT possesses a strict theoretical foundation. Hence, it performs better than EMD in theory. However, as EWT generates wavelets according to data to be processed, it cannot guarantee to generate the same wavelets for all the source images, and consequently different source images may be projected into different basis sets. As a result, it is nearly impossible to design proper fusion rules.

With the aim at solving the problems mentioned above, a straightforward extension of EWT is invented, which is named as Simultaneous Empirical Wavelet Transform (SEWT). With it, all the source images can be projected into the same basis set which is generated according to the source images themselves. Meanwhile, we propose a novel image fusion method based on SEWT, abbreviated as SEWTF.

The reminder of the paper is constructed as follows: in Section 2, the theory of SEWT is explained; in Section 3, the proposed image fusion algorithm is presented in detail; Section 4 presents the experimental results and discussions; finally, conclusions are drawn in the Section 5.

2 Simultaneous empirical wavelet transform

Wavelets are useful tools for signal (images) analysis. However, their wavelet base sets are fixed, which makes it impossible for us to design a filter that is optimal to represent any signal, and consequently it affects the results of signal analysis. So it is good to design different filter sets according to the signal to be processed, and decompose signals with their optimal filter sets. In literature [3], the authors have done a pioneering job where EWT is proposed with the aim at extracting amplitude modulated-frequency modulated components of an input signal. EWT can build a wavelet tight frame corresponding to an adaptive filter bank.

However, signal decomposition has been widely used in some cases where several signals need to be processed at the same time, for example, signal comparison, image fusion and the like. In this case, EWT cannot be used directly, because its filter bank is not fixed but determined by the processed signal. Different signals may produce quite different filter banks, and then these signals unavoidably be projected into different wavelet spaces. Adaptive filter bank design may be an attractive feature of EWT in some cases, but it brings us much trouble in the case of simultaneous signal decomposition.

In order to address the problem mentioned above, we present a EWT's straightforward revision named as Simultaneous Empirical Wavelet Transform (SEWT). The algorithm uses the same filter bank to process all the signals that need to be processed at the same time. In this way, signals can be projected into the same wavelet basis without the loss of adaptivity of filter bank. 1D SEWT and 2D SEWT are presented in subsection 2.1 and subsection 2.2, respectively.

2.1 1D SEWT

Traditional EWT algorithm contains three steps [4]: first the processed signal is transformed into Fourier domain; then find out the Fourier supports and build a tight frame set of wavelets based on them; finally the signal is decomposed with the assistance of the empirical wavelets. In order to get the same tight frame set of wavelets for several signals, we need to find Fourier supports based on all the signals rather than one single signal.

Let $f_i(t)$ $(i = 1, \ldots, N_i)$ denote the processed signals, and $\mathfrak{F}_t(f_i)(\omega)$ $(i = 1, \ldots, N_i)$ denote the Fourier transform of $f_i(t)$, namely $\mathfrak{F}_t(f_i)(\omega) = \int_{-\infty}^{+\infty} f_i(t) e^{-iwt} dt$. Based on $|\mathfrak{F}_t(f_i)|(\omega)$, we can detect Fourier supports which can cut the signal into adjacent intervals. Fourier supports for each signal are detected by finding the local lowest minima. For the Fourier spectrum $\mathfrak{F}_t(f_i)(\omega)$, assume that it contains N local maximum $\nabla_i^n (n = 1, \ldots, N)$, and its location is indexed by $\hat{\omega}_i^n (n = 1, \ldots, N)$. Then we can get at least one local minima in the interval $[\hat{\omega}_i^{n-1}, \hat{\omega}_i^n]$ which is denoted as Δ_i^n. The location of these local minima $\omega_i^n (n = 1, \ldots, N)$ is given by

$$\omega_i^n = \begin{cases} 0, & i = 0 \\ \underset{\omega}{\arg\min} \Delta_i^n, & 0 < i < N_i \\ \pi, & i = N_i \end{cases} \tag{1}$$

From the procedure discussed above, it is straightforward to see that different signals may generate quite different Fourier supports. However, in order to generate the same tight frame set, we need to get a unique one, here we set $\omega = \overset{K}{\underset{i=1}{\cup}} \omega_i$. Based on the set ω, we can define a wavelet tight frame, $B = \{\phi_1(t), \{\varphi_m(t)\}_{m=1}^{M-1}\}$, where $\phi_1(\omega)$ refers to empirical wavelets, and

$\{\varphi_m(t)\}_{m=1}^{M-1}$ denotes the scaling function. Following the setting in [4], we can get the Fourier transforms of B using Meyer's and Littlewood-Paley, which are denoted as $\left\{ \mathfrak{F}_t(\phi_1)(\omega), \{\mathfrak{F}_t(\varphi_m)(\omega)\}_{m=1}^{M-1} \right\}$.

Once the wavelet tight frame is determined, we can get the approximation layer for each of signal

$$\mathcal{W}(f_i)(0,t) = \mathfrak{F}_\omega^{-1}\left(\mathfrak{F}_t(f_i)(\omega)\overline{\mathfrak{F}_t(\phi_1)(\omega)} \right) \tag{2}$$

and the detail layers for each signal

$$\mathcal{W}(f_i)(m,t) = \mathfrak{F}_\omega^{-1}\left(\mathfrak{F}_t(f_i)(\omega)\overline{\mathfrak{F}_t(\varphi_m)(\omega)} \right). \tag{3}$$

where $\mathfrak{F}_\omega^{-1}$ denotes the inverse Fourier Transform.

The inverse 1D Simultaneous Empirical Wavelet Transform can be defined by

$$f_i(t) = \mathcal{W}(f_i)(0,t) \circ (\phi_1)(t) + \sum_{m=1}^{M-1} \mathcal{W}(f_i)(m,t) \circ \varphi_m(t) \tag{4}$$

where $\circ$ is a convolution operator.

2.2 2D SEWT

In the real world, we need to process 2D signals. So it is of great significance to extend the newly proposed SEWT to 2D space. One straightforward alternative is to reshape the signal into 1D row by row or column by column; then performs 1D SEWT on the reshaped signal; and finally 1D approximation layer and detail layers are mapped into 2D space. However, such process may lose some vital spatial information among neighboring pixels. Here, 2D Littlewood-Paley wavelet transform is used. The transform can be regarded as a filter bank obtained in the Fourier domain on annuli supports for filtering images.

Let $I_i(\mathbf{p})$ ($i = 1,\ldots,N_i$) denote the processed images (2D signals) with $\mathbf{p}$ representing pixel locations, and $\mathfrak{F}_P(I_i)(\theta, |\omega|)$ denotes Pseudo-Polar Fourier Transform where θ refers to angle of Fourier Transform. For any image I_i, each θ produces a $\mathfrak{F}_P(I_i)(\theta, |\omega|)$. Before detecting Fourier boundary for each image, we need to compute the average spectrum with respect to θ. Namely,

$$\mathfrak{F}(I_i)(|\omega|) = \frac{1}{N_\theta} \sum_{k=1}^{N_\theta} \mathfrak{F}_P(I_i)(\theta_k, |\omega|). \tag{5}$$

The Fourier spectrum will indirectly influence the empirical Littlewood-Paley wavelets, so it should take into account all the images that need to be decomposed. Here, we define an average-average spectrum with respect to i:

$$\mathfrak{F}(|\omega|) = \frac{1}{N_i} \sum_{i=1}^{N_i} \mathfrak{F}(I_i)(|\omega|) = \frac{1}{N_i N_\theta} \sum_{i=1}^{N_i} \sum_{l=1}^{N_\theta} \mathfrak{F}_P(I_i)(\theta_l, |\omega|). \tag{6}$$

Fourier boundary detection can be conducted on the average-average spectrum in Eq. (6), and the detection results are kept in $\omega_i^m (m = 1, \ldots, M)$. Then build a set of 2D Simultaneous Empirical Wavelets $B = \{\phi_1(\mathbf{p}), \{\varphi_m(\mathbf{p})\}_{m=1}^{M-1}\}$, where $\phi_1(\mathbf{p})$ denotes 2D empirical wavelets, and $\{\varphi_m(\mathbf{p})\}_{m=1}^{M-1}$ denotes the 2D scaling function. The approximation layer of image $I_i(\mathbf{p})$ is defined as

$$\mathcal{W}(I_i)(0, \mathbf{p}) = \mathfrak{F}_\omega^{-1}\left(\mathfrak{F}_P(f_i)(\omega)\overline{\mathfrak{F}_\mathbf{p}(\phi_1)(\omega)} \right), \tag{7}$$

and the detail layers of image $I_i(\mathbf{p})$ are defined as

$$\mathcal{W}(I_i)(m, \mathbf{p}) = \mathfrak{F}_P^{-1}\left(\mathfrak{F}_\mathbf{p}(I_i)(\omega)\overline{\mathfrak{F}_\mathbf{p}(\varphi_m)(\omega)} \right). \tag{8}$$

The inverse 2D Simultaneous Empirical Wavelet Transform can be defined by

$$I_i(\mathbf{p}) = \mathcal{W}(I_i)(0, \mathbf{p}) \circ (\phi_1)(\mathbf{p}) + \sum_{m=1}^{M-1} \mathcal{W}(I_i)(m, \mathbf{p}) \circ \varphi_m(\mathbf{p}). \tag{9}$$

where $\circ$ is the 2D convolution operator.

EWT in [4] and SEWT presented in this paper can be regarded as extensions of traditional Wavelet Transform (WT). In addition, EWT and SEWT share the same mathematical theory of WT. The difference of EWT and SEWT with WT is that their wavelet functions are not fixed for all the images but generated according to each signal to be processed. Hence, EWT and SEWT also contain all the excellent properties of WT, including linear addition characteristic, time-shift invariance, multi-scale analysis, tight frame, and so on.

3 The proposed fusion algorithm

3.1 Framework

The image fusion method based on 2D Simultaneous Empirical Wavelet Transform (SEWTF) can be summarized in Fig. 1. The proposed SEWTF algorithm contains no registration steps,
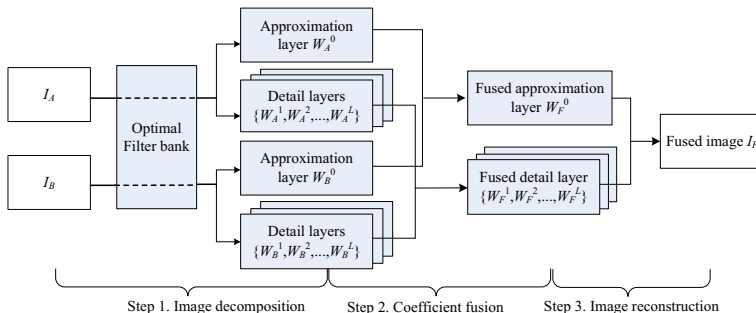
Fig. 1 Framework of the image fusion method based on 2D Simultaneous Empirical Wavelet Transform

so it is assumed that the source images are well registered. If the source images are not geometrically aligned in practice, users need to pre-process them using the registration methods such as [12, 18] before running this fusion algorithm.

In this paper, let I_A and I_B denote the well registered source images with the size of $M \times N$, and I_F denotes the fused image. The fusion framework takes the following three steps,

(1) Based on the two source images I_A and I_B, build a set of 2D Simultaneous Empirical Wavelets $B = \{\phi_1, \{\varphi_i\}_{i=1}^{L-1}\}$. Decompose the source images I_A and I_B by 2D SEWT into subbands $\mathcal{W}_A = \{\mathcal{W}_A^0, \mathcal{W}_A^1, \cdots, \mathcal{W}_A^L\}$ and $\mathcal{W}_B = \{\mathcal{W}_B^0, \mathcal{W}_B^1, \cdots, \mathcal{W}_B^L\}$ using Eqs. (7) and (8) respectively. Here, $\mathcal{W}_*^0$ refers to the approximation layer and $\mathcal{W}_*^1, \cdots, \mathcal{W}_*^L$ represent the detail layers.

(2) With the assistance of the designed fusion rules, the approximation layer and detail layers are combined to get the fused layers $\mathcal{W}_F = \{\mathcal{W}_F^0, \mathcal{W}_F^1, \cdots, \mathcal{W}_F^L\}$. Let $r = \{r_0, \{r_i\}_{i=1}^L\}$ denote the fusion map set where r_0 is the fusion map for approximation layer, and $\{r_i\}_{i=1}^L$ are the ones for detail layers. Then we have

$$\mathcal{W}_F^0 = r_0\mathcal{W}_A^0 + (1-r_0)\mathcal{W}_B^0 \tag{10}$$

and

$$\mathcal{W}_F^i = r_i\mathcal{W}_A^i + (1-r_i)\mathcal{W}_B^i \quad i = 1, \ldots, L \tag{11}$$

(3) Conduct inverse 2D SEWT on $\mathcal{W}_F$ to obtain the final fused image I_F using Eq. (9).

In the proposed fusion algorithm, 2D Simultaneous Empirical Wavelet Transform plays a vital rule. It can decompose each source image into several subbands with different morphological components. In addition, the 2D wavelets are generated adaptively according to the source images themselves, which guarantees that the wavelets are the optimal ones for decomposing these images. Note that each group of source images will generate a tight frame set of wavelets.

Figure 2 displays an example of 2D SEWT: Fig. 2a shows a pair of source images (a CT image and an MRI image), their decomposed layers by 2D SEWT are listed in Fig. 2b. The tight frame set of wavelets is determined adaptively according to the two source images shown in Fig. 2a.

(a)   (b)
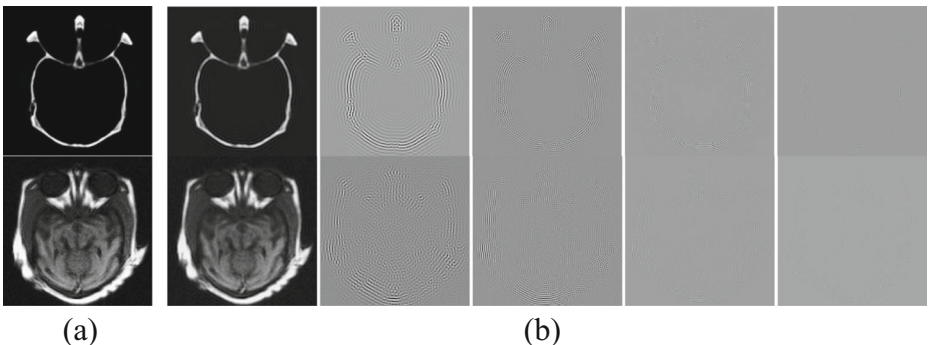
Fig. 2 An illustration of 2D SEWT: (a) source images; (b) decomposed layers

The proposed SEWTF algorithm can be easily extended to fuse an image sequence which contains more than two source images. First, two source images are randomly selected for fusion using SEWTF algorithm, and then the fused image is combined with another source image using SEWTF; repeat such operation until all the source images are fused. Finally, we can get the fused image of the image sequence.

3.2 Fusion rules

2D SEWT can decompose each source image into one approximation layer and several detail layers. The approximation layer preserves the basic visual information of the source images, such as intensity level and structures; while the detail layers capture high frequency information, such as strong edges and textures. So it is necessary to design different fusion rules for the two kinds of layers in a bid to abstract salient visual information instead of implicit information from each source image.

Theory on biological vision shows that human vision system is sensitive to contrast in visual signal, so we need to preserve the visual information in all the layers of each source images, though the approximation layer contains little such information. Here, a global contrast measure is used to predict the contrast of pixels in approximation layer of each source image

$$S\big(\mathcal{W}_*^0(\mathbf{p})\big) = \frac{1}{M \times N} \sum_{w(\mathbf{q}) \in \mathcal{W}_*^0} D\big(\mathcal{W}_*^0(\mathbf{p}), \mathcal{W}_*^0(\mathbf{q})\big), \tag{12}$$

where $D(W_*^0(\mathbf{p}), W_*^0(\mathbf{q}))$ refers to the intensity distance between coefficients at $\mathbf{p}$ and $\mathbf{q}$, namely, $D(W_*^0(\mathbf{p}), W_*^0(\mathbf{q})) = |W_*^0(\mathbf{p}) - W_*^0(\mathbf{q})|$. The fusion rule for approximation coefficients is defined as such a sigmoid function

$$r_0 = \frac{1}{1 + e^{-a\big(S(\mathcal{W}_A^0) - S(\mathcal{W}_B^0)\big)}}, \tag{13}$$

where a is a positive constant that determines how threshold-like the sigmoid function is. Namely, if the parameter is assigned to be a high value, r_0 shall be very close to the following threshold function

$$r_0 = \begin{cases} 1, & S\big(\mathcal{W}_A^0\big) > S\big(\mathcal{W}_B^0\big) \\ 0, & otherwise \end{cases}. \tag{14}$$

However, if a is close to zero, r_0 is close to 0.5. It is straightforward to see that sigmoid function is a tradeoff of averaging and maximum selection. Figure 3 displays a sigmoid function and a threshold function. In this algorithm, we choose the sigmoid function instead of the threshold function because the former one is smooth and continuous, which makes the fused images smoother and avoids blocking artifacts effectively.

For detail layers, as they contain high frequency information of source images, we adopt absolute maximum selection as their fusion rules. In images, high frequency always corresponds to salient information, such as edges or textures. This fusion rule, on the one hand, is
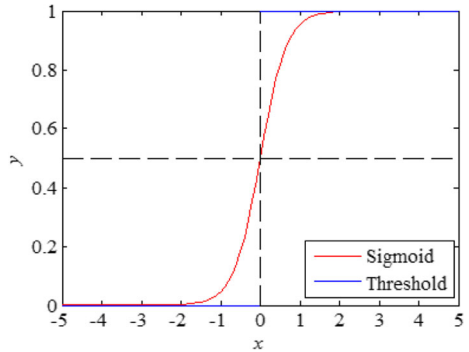
conducive to preserve salient information in final fused images; and on the other hand, performs well in terms of computation complexity. The fusion rule for detail layers is formally defined as,

$$r_i = \begin{cases} 1, & |\mathcal{W}_A^i| > |\mathcal{W}_B^i| \\ 0, & otherwise \end{cases} \quad (i = 1, \ldots, L) \tag{15}$$

Once getting the fusion map set $r = \{r_0, \{r_i\}_{i=1}^L\}$, we can get a fused approximation layer and L detail layers using Eqs. (10) and (11).

4 Experimental results and discussions

4.1 Experimental settings

In order to examine the performance of the proposed fusion algorithm, experiments on two types of images are performed, namely multi-modal medical images and Visible-IR images. In addition, medical images cover CT (Computed Tomography), MRI (Magnetic Resonance Imaging), CBF (Cerebral Blood Flow), and SPECT (Single-Photon Emission Computed Tomography). In the experiments, medical images come from two websites (http://www.imagefusion.org) and (http://www.med.harvard.edu/AANLIB/home.html). Visible-IR images are downloaded from Dr. Liu's Homepage (http://home.ustc.edu.cn/~liuyu1/).

The proposed fusion algorithm is compared with five popular methods: Laplacian Pyramid based algorithm (LP), Discrete wavelet transform based algorithm (DWT), Nonsubsampled contourlet transform based algorithm (NSCT), Guided filtering based fusion algorithm (GFF) [8], and the algorithm proposed by Liu et al. (LIU) [11]. In the fusion algorithms LP, DWT, NSCT, source images are decomposed into four layers. The coefficients in approximation layers are combined by averaging, and the ones in detail layers are combined by selecting maximum absolute values. The settings of the algorithms GFF and LIU are the default ones in the provided code.

To quantitatively compare the fusion performance of the proposed algorithm with others, five quality measures are adopted in the experiments, including Q_{AB}^F [16], two Piella measures (Q_0, Q_E) [17], and VIFF [5]. Q_{AB}^F evaluates how much edge information are transferred from source images to their fused one; Both Q_0 and Q_E, constructed on the Structural Similarity

theory [20, 21], measure similarity of fused images and source images from the aspects of correlation, luminance and contrast. *VIFF* measures the quality of fused images using visual information fidelity. For all the measures, a higher value corresponds to a better fusion quality.
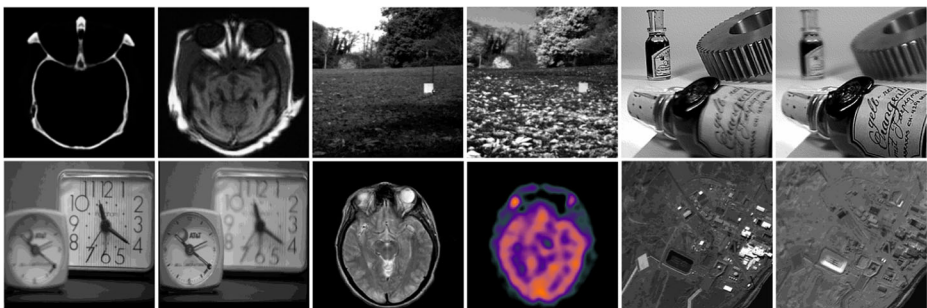
4.2 Parameter selections

There is only one parameter that should be determined, namely a in Eq. (13). It controls the sharpness of the sigmoid function. As discussed above, $a = 0$ corresponds to $r^0 = 0.5$ which means that approximation layer of the fused image is obtained by averaging; $a = +\infty$ corresponds to an indicator function which means that approximation layer of the fused image is obtained by selecting maximum global contrast. It is clear that neither 0 nor $+\infty$ is the optimal value for the parameter a. In order to tackle this problem, an empirical method is used.

Six pairs of source images shown in Fig. 4 are adopted here to empirically investigate the influence of the parameter a on the performance of the proposed SEWTF algorithm. First, these source images are fused by the proposed method with different parameter setting. Then, the objective evaluation measures are calculated. The results with respect to the different parameter setting are displayed in Fig. 5. In the figure, x-axis corresponds to a's value, and y-axis corresponds to the scores predicted by objective evaluation measures, including Q^F_{AB}, two Piella measures (Q_0, Q_E), and *VIFF*. For all the measures, larger scores indicate better fused results. Based on the curves in Fig. 5, it can be seen that the performance of proposed method is the worst, when the parameter a is set to be zero. It suggests that averaging coefficients in approximation layers is not a proper fusion rule. Meanwhile, we also note that there are nearly no difference in terms of the performance of proposed method when a varies from 1 to 5. We also check the performance of the proposed algorithm when the parameter a is distributed in the interval of [10, 100], the performance is the same with that in the case of a varying from 1 to 5.

Figure 6 shows the relationship of the metric with the parameter a with the step of 0.2 in the interval of [0.1, 0.9]. Here, we note that Q^F_{AB} and Q_0 suggest that a larger a may be conducive to produce a better fusion, while, Q_E and *VIFF* make the contrary conclusion. So in our experiments, we set a as 0.5. In this case, the proposed algorithm can get satisfactory fusion performance.

Note that in the proposed fusion algorithm, there is only one time decomposition and reconstruction processing (DRP). Here, we test whether more times DRP are needed. The six groups of source images shown in Fig. 4 are also adopted to check the influence of the time of DRP on the performance of the proposed SEWTF algorithm. The objective evaluation

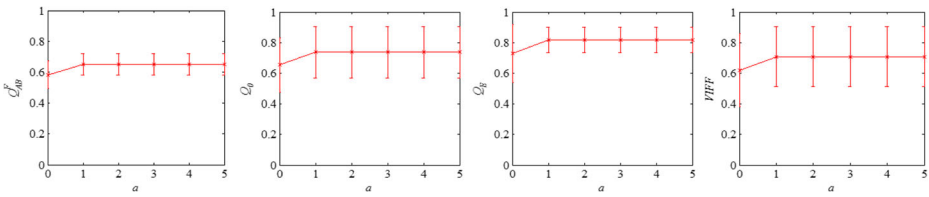**Fig. 4** Source images for parameter selections

Fig. 5 The relationship of the metric with the parameter a within the interval of [0, 5]

measures are employed to compare the fused images. The results with respect to the different time of DRP are displayed in Fig. 7. The figure reveals that the scores given by evaluation measures Q_0, Q_E and *VIFF* do not increase with time of DRP, which means that more times DRP does not result in better fusion. In addition, more times DRP need more computation time. Hence, in the following experiments, there is only one time DRP.

4.3 Experiments on multi-modal medical images

Nowadays, medical images play an increasingly important role in clinical diagnosis and treatment. For example, CT (Computed Tomography) images are widely used in orthopaedics, because they can capture status of bones clearly; MRI (Magnetic Resonance Imaging) images can reflect soft tissue information clearly, so they are usually adopted for diagnosing intracerebral hematoma, brain tumors and the like. CBF (cerebral blood flow) can objectively reflect the changes of tension and elasticity of cerebral blood vessels. SPECT (Single-Photon Emission Computed Tomography) can measure the biology activities of cellular and molecular. However, these images contain their own limitations, which may restrict their applications in real applications. For example, SPECT images are not clear enough and they cannot capture the structure of tissues, so SPECT images are always used along with MRI or CT in clinical diagnosis.

In this subsection, the proposed algorithm is compared with others on various types of medical images. The source images, as shown in Fig. 8, contain CT, MRI, CBF, and SPECT images. In each row of the figure, two pairs of source images are displayed. Each pair contains two images captured by different medical imaging equipments.

Figure 9 shows the fused images obtained by different methods. If taking a close look at the figure, we can note the difference of the proposed algorithm with other schemes. As shown in Fig. 9-(f1), the final image fused by the proposed SEWTF algorithm can well preserve the complementary visual information of different source images, namely the fused image preserves the soft tissue of human brain from the MRI image and the bone from CT image simultaneously. However, in the brain tissue part of the fused images obtained by LP and DWT, the brightness is decreased. In the fused image obtained by NSCT, the brain tissue is neglected. LIU algorithm does not work well for this example as well, because it produces some visible artifacts around the bone in the fused image.
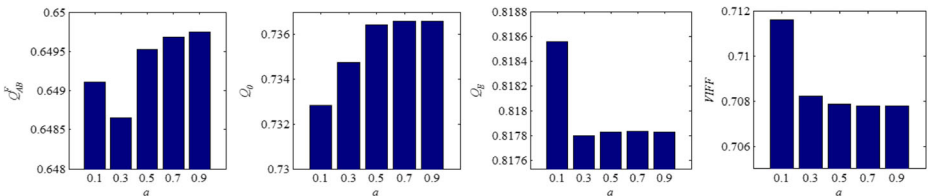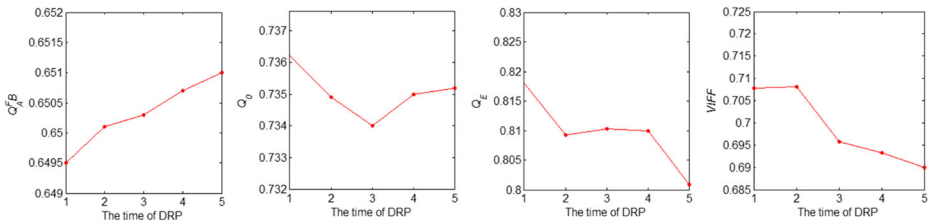

Fig. 6 The relationship of the metric with the parameter a within the interval of [0.1, 0.9]
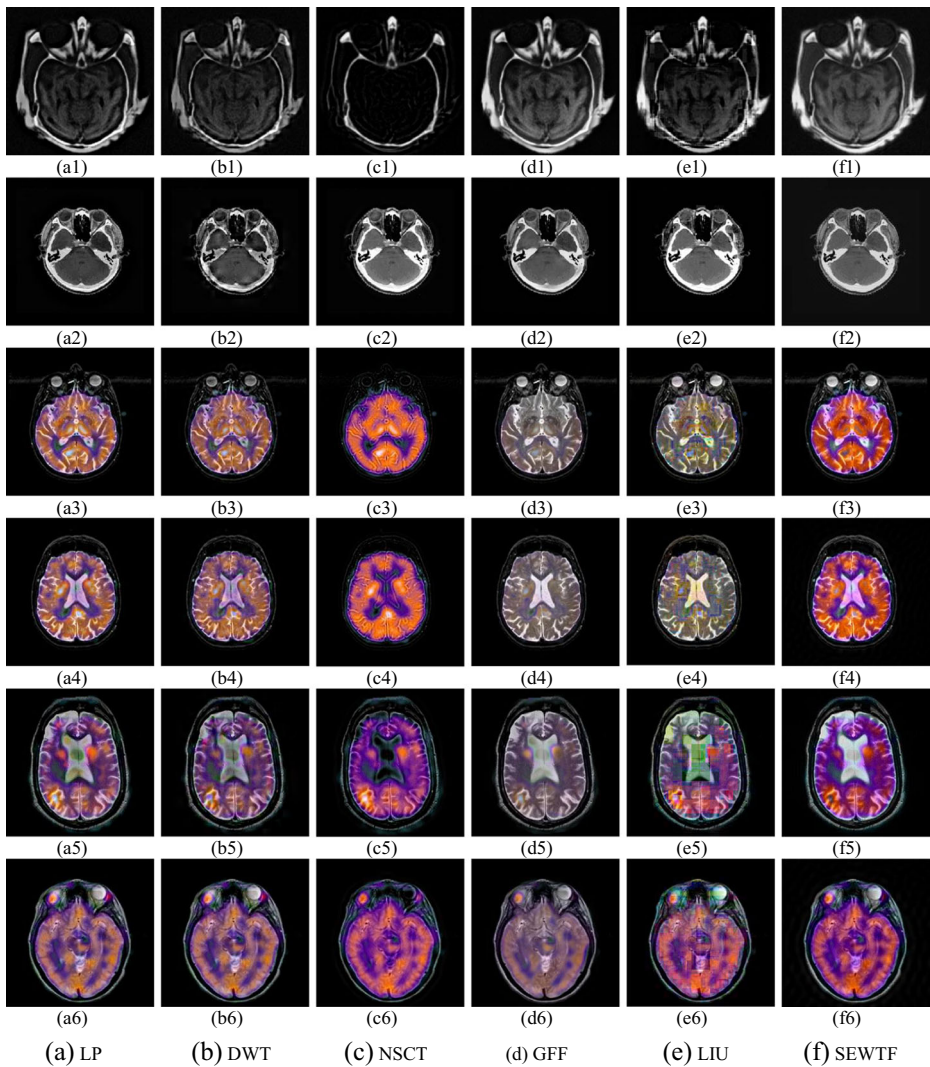
Fig. 7 The relationship of the metric with the time of DRP

From Fig. 9-(a3) to (f6), we list 4 groups of color fused images. Different from the case of gray image fusion, color image fusion needs to preserve salient visual information of each source image; meanwhile, it should avoid color distortion. Figure 9-(f3) shows that the proposed SEWTF algorithm can preserve the color from CBF image and the soft tissue of human brain from MRI. From Fig. 9(a3)–(e3), we can note that the fused images generated by the LP, DWT, GFF, and LIU method lose the color of CBF image, and consequently makes some details invisible. Figure 9-(c3) shows that the NSCT based algorithm fails in producing satisfactory fused image, because it tends to loss texture parts.

In order to compare the proposed SEWTF algorithm with others objectively, Q^F_{AB} [16], two Piella measures (Q_0, Q_E) [17], and *VIFF* [5] are employed to predict the quality of the fused images obtained by different algorithm. The objective evaluations for the medical test images are shown in Table 1. The objective evaluations indicate that SEWTF performs better than other methods, which justifies the subjective conclusions drawn above.
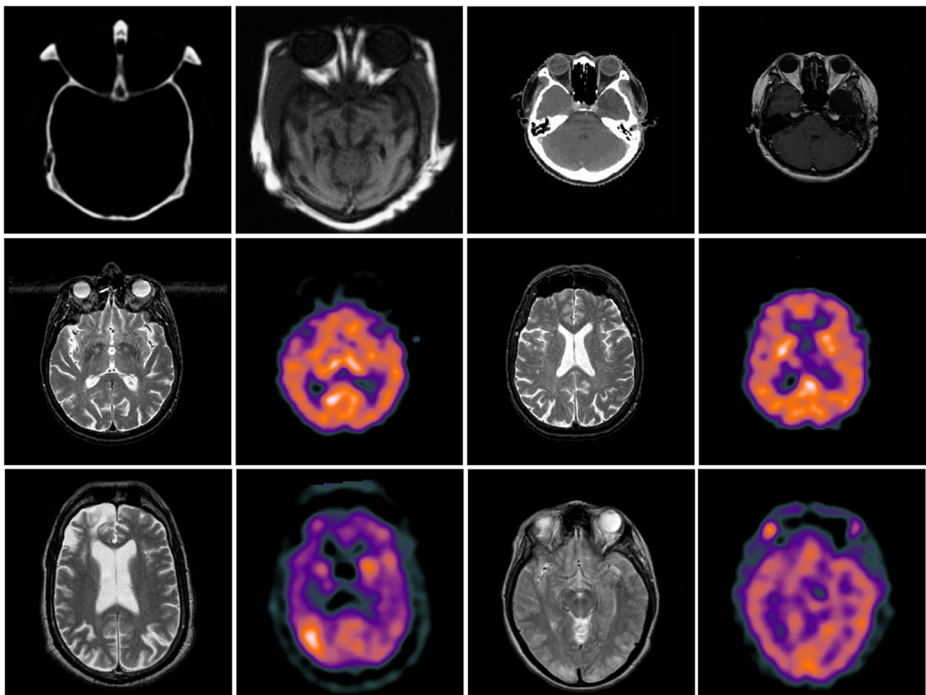
Fig. 8 Medical source images: each row contains two pairs of source images. The top row lists CT-MRI images; the middle rule lists MRI-CBF images; and the bottom row lists MRI-SPECT images

(a1)	(b1)	(c1)	(d1)	(e1)	(f1)
(a2)	(b2)	(c2)	(d2)	(e2)	(f2)
(a3)	(b3)	(c3)	(d3)	(e3)	(f3)
(a4)	(b4)	(c4)	(d4)	(e4)	(f4)
(a5)	(b5)	(c5)	(d5)	(e5)	(f5)
(a6)	(b6)	(c6)	(d6)	(e6)	(f6)
(a) LP	(b) DWT	(c) NSCT	(d) GFF	(e) LIU	(f) SEWTF

Fig. 9 Fused medical images obtained by different algorithms

4.4 Experiments on visible-ir images

In the real world, visible images and IR(infrared) images are always fused into one image. Visible image can only capture the surface information, while, the IR image can reflect the inner information based on their temperature. The fusion of Visible-IR images allows improved detection and directly recognizable localization of a target in the IR image with respect to its background in the visible image [19].

In this subsection, the proposed algorithm is compared with others on Visible-IR images. In each row of the figure, two pairs of source images are displayed. Each pair contains two images captured by different scanners from the same scene (Fig. 10).

Table 1 Objective evaluation results of various algorithms on medical images

Source image		Q^F_{AB}	Q_0	Q_E	VIFF
#1	(a1)	0.7184	0.6197	0.7316	0.5256
	(b1)	0.6073	0.5857	0.6728	0.3743
	(c1)	0.4122	0.1958	0.5129	0.1244
	(d1)	0.7119	0.7510	0.6985	0.4847
	(e1)	0.5645	0.6540	0.6529	0.3574
	(f1)	0.7454	0.8659	0.7080	0.5068
#2	(a2)	0.5892	0.5166	0.7753	0.5972
	(b2)	0.5238	0.5032	0.7386	0.4705
	(c2)	0.5919	0.5739	0.8054	0.5569
	(d2)	0.5763	0.3934	0.7811	0.5176
	(e2)	0.5785	0.6542	0.8014	0.5241
	(f2)	0.6109	0.6633	0.7924	0.6340
#3	(a3)	0.6820	0.4057	0.8687	0.5761
	(b3)	0.6413	0.4010	0.8453	0.5006
	(c3)	0.4140	0.2228	0.5025	0.1955
	(d3)	0.6941	0.4111	0.8929	0.4947
	(e3)	0.7399	0.4634	0.8863	0.4796
	(f3)	0.7638	0.4705	0.9021	0.5939
#4	(a4)	0.6643	0.3361	0.8707	0.5898
	(b4)	0.6185	0.3205	0.8419	0.4998
	(c4)	0.4196	0.1826	0.5052	0.2062
	(d4)	0.6613	0.3335	0.8868	0.4902
	(e4)	0.7071	0.3611	0.8862	0.4770
	(f4)	0.7210	0.3849	0.8248	0.5984
#5	(a5)	0.6418	0.5303	0.8820	0.6545
	(b5)	0.6148	0.4938	0.8712	0.5646
	(c5)	0.4799	0.3572	0.6529	0.2762
	(d5)	0.6367	0.4656	0.8786	0.5673
	(e5)	0.6441	0.5747	0.8711	0.5616
	(f5)	0.6843	0.5836	0.9021	0.5680
#6	(a6)	0.5953	0.4298	0.8474	0.6692
	(b6)	0.5618	0.4051	0.8299	0.5706
	(c6)	0.4768	0.3358	0.6781	0.3900
	(d6)	0.5967	0.4112	0.8325	0.4848
	(e6)	0.5329	0.4215	0.8035	0.5481
	(f6)	0.5988	0.4325	0.8546	0.6864

Figure 11 shows the fused images obtained by different methods. By taking a careful look at the figure, we note the final image fused by the proposed SEWT can well preserve the complementary visual information of different source images. Take the first group of images as an example, Fig. 11-(f1) suggests that SEWTF produces a more informative and natural fused image. It can be noted that the images fused by NSCT based algorithm nearly neglect the information from the IR image. The reason is that only the major principal components are preserved and minor components are neglected in this fusion algorithm.
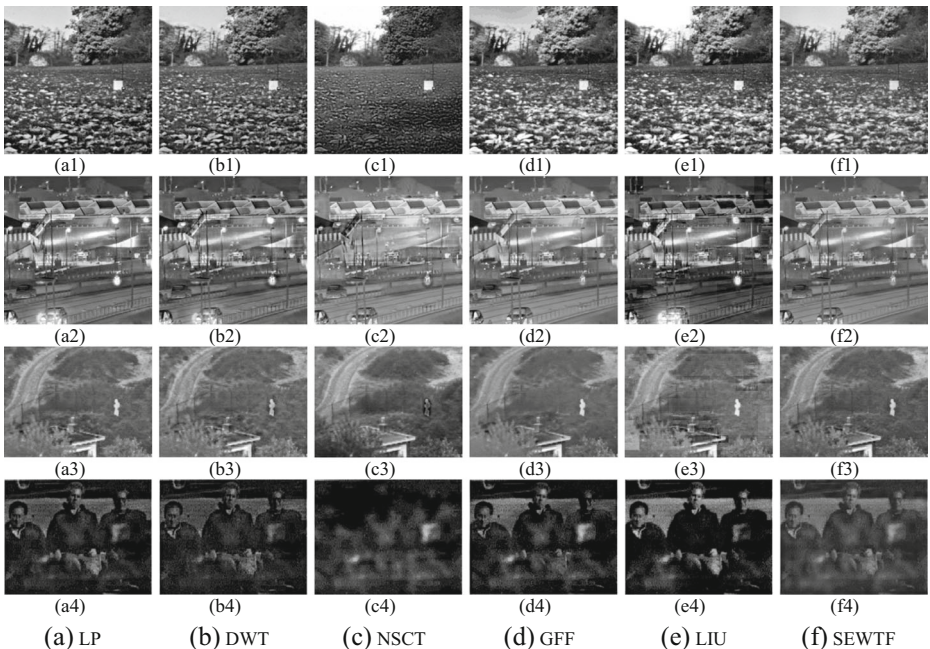
Fig. 10 Visible-IR source images

The objective evaluations for the Visible-IR test images are shown in Table 2. The objective evaluations indicate that SEWTF performs better than other methods, which justifies the subjective conclusions drawn above.

4.5 Image sequence fusion

Figure 12 displays an image sequence which is constructed by blurring an image's different regions. Two source images are selected for fusion using the proposed SEWTF algorithm, and then the fused image is combined with another source image using SEWTF; repeat such

| (a1) | (b1) | (c1) | (d1) | (e1) | (f1) |
(a2)	(b2)	(c2)	(d2)	(e2)	(f2)
(a3)	(b3)	(c3)	(d3)	(e3)	(f3)
(a4)	(b4)	(c4)	(d4)	(e4)	(f4)
(a) LP	(b) DWT	(c) NSCT	(d) GFF	(e) LIU	(f) SEWTF

Fig. 11 Fused Visible-IR images obtained by different algorithms

Table 2 Objective evaluation results of various algorithms on Visible-IR images

Source images		Q^F_{AB}	Q_0	Q_E	VIFF
#1	(a1)	0.6826	0.8296	0.8596	0.8735
	(b1)	0.6403	0.8032	0.8414	0.7550
	(c1)	0.4990	0.5242	0.5658	0.3738
	(d1)	0.6774	0.8585	0.8908	0.8394
	(e1)	0.7150	0.8637	0.8966	0.8582
	(f1)	0.6875	0.8658	0.8986	0.8738
#2	(a2)	0.6333	0.7625	0.7788	0.6531
	(b2)	0.5837	0.7196	0.7589	0.5197
	(c2)	0.5646	0.6746	0.7121	0.2956
	(d2)	0.6084	0.7402	0.7691	0.4642
	(e2)	0.5518	0.6947	0.6961	0.5084
	(f2)	0.6546	0.7669	0.7824	0.5250
#3	(a3)	0.4828	0.6373	0.7254	0.5241
	(b3)	0.4426	0.5802	0.7114	0.4197
	(c3)	0.4499	0.6140	0.6500	0.2993
	(d3)	0.4467	0.4901	0.6911	0.2648
	(e3)	0.4532	0.5938	0.6823	0.3301
	(f3)	0.4923	0.6238	0.7367	0.5632
#4	(a4)	0.6811	0.7090	0.9362	0.6390
	(b4)	0.6219	0.6813	0.9186	0.4950
	(c4)	0.6310	0.6161	0.6162	0.1268
	(d4)	0.6566	0.7929	0.8483	0.4782
	(e4)	0.7151	0.7117	0.9434	0.6618
	(f4)	0.6519	0.7986	0.9437	0.6924

operation until all the source images are fused. Finally, we can get the fused image of the image sequence, which is shown in Fig. 13. It can be observed that all the coins in the fused image are clear.
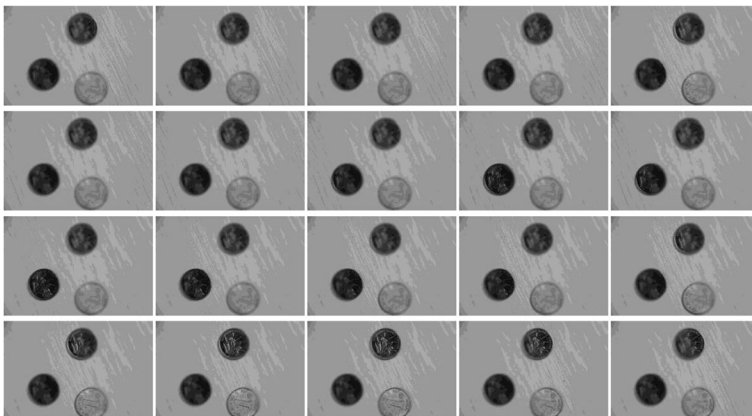
Fig. 12 An image sequence

Fig. 13 Fused image

4.6 Computation time comparison

Image fusion algorithms have been widely employed in various fields, such as remote sensing, optical imaging systems, and clinical diagnosis. Nearly all the applications demand that the fusion algorithms run in real time. Hence, it is necessary to analyze the computation consumption of the proposed algorithm. The algorithm contains three parts with different computation time. First, source images are decomposed into several layers using SEWT; then, the approximation layer and detail layers are fused, respectively; finally, the fused approximation layer and detail layers are combined to get the final fused image.

The comparison of computation time among the proposed algorithm and other methods is presented. Fusion methods include (LP, DWT, NSCT, GFF, LIU). All experiments are performed on a PC with an Intel Xeon 2.66 GHz CPU and 3.25 GB RAM using Matlab 7.9.0. The elapsed time of fusion algorithms that process 1024 pixels are listed in Table 3.

From Table 3, it can be seen that the consumption time of GFF is the lowest, followed by the DWT, LP, and LIU. As NSCT based fusion algorithm needs complicated image decomposition and reconstruction, its computation time is the highest among these fusion algorithms. The consumption time of the proposed algorithms is longer than other four algorithms (except NSCT based fusion algorithm). In the proposed fusion algorithm, time is mainly consumed in the process of image decomposition, because the algorithm needs to find optimal wavelets rather than design fixed wavelets.

Table 3 Comparison on time consumption

Image fusion algorithm	LP	DWT	NSCT	GFF	LIU	SEWTF
Time(s/1 K pixels)	0.0195	0.0174	2.4620	0.0059	0.0340	0.0489

5 Conclusions

In this paper, we presented a novel multi-scale image fusion algorithm for multi-sensor images based on simultaneous empirical wavelet transform. Like other multi-scale fusion algorithms, it contains three steps: image decomposition, coefficient combination, and image reconstruction. But compared with traditional ones, an attractive feature of the proposed algorithm is that it can simultaneously decompose the source images into the same wavelet set which is the optimal one for these images. The global contrast based salience detection model is utilized to fuse smooth parts, and maximum absolute value selection is employed to fuse the detail layers. The proposed algorithm is tested on various types of image infusion, including multimodal medical image fusion and Visible-IR fusion. The experimental results show that it can preserve salient and complementary information of the source images. In addition, the results show good performance on color image fusion.

In the real applications, the algorithm can be used in various fields, such as clinical diagnosis, target detection, and digital camera. For example, in clinical diagnosis, MRI can capture structural and anatomical information with high resolution, and SPECT images reveal the metabolic change which is significant to clinical diagnosis. The proposed algorithm can combine the two images into one image which contains more information. Hence, the algorithm can help doctors to make correct diagnosis. In target detection, visible image can only capture the surface information, while, the IR image can reflect the inner information based on their temperature. The fusion of Visible-IR images allows improved detection and directly recognizable localization of a target in the IR image with respect to its background in the visible image.

However, there are still some limitations in the algorithm. First, the time consumption of the algorithm is relatively high, which restricts its application in real time systems. Time is mainly consumed in the process of image decomposition, because it needs to find out the optimal wavelets for signals. Second, we do not test the proposed algorithm on noised images. Considering that noises unavoidably affect the results in the process of generating optimal wavelets, it is necessary to conduct denoising operations before running the proposed fusion algorithm.

References

1. Chen T, Zhang J, Zhang Y (2005) Remote sensing image fusion based on ridgelet transform. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS'05). IEEE 1150–1153
2. Geng P, Huang M, Liu S, Feng J, Bao P. Multifocus image fusion method of Ripplet transform based on cycle spinning. Multimed Tools Appl 1–11
3. Gilles J (2013) Empirical wavelet transform. IEEE Trans Signal Process 61(16):3999–4010
4. Gilles J, Tran G, Osher S (2014) 2D empirical transforms. wavelets, ridgelets, and curvelets revisited. SIAM J Imaging Sci 7(1):157–186
5. Han Y, Cai Y, Cao Y, Xu X (2013) A new image fusion performance metric based on visual information fidelity. Inf Fusion 14(2):127–135

6. Hariharan H, Gribok A, Abidi MA, Koschan A (2006) Image fusion and enhancement via empirical mode decomposition. J Pattern Recogn Res 1(1):16–32
7. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Ro Soc Lond A Math Phys Eng Sci R Soc 903–995
8. Li S, Kang X, Hu J (2013) Image fusion with guided filtering. IEEE Trans Image Process 22(7):2864–2875
9. Li H, Manjunath B, Mitra SK (1995) Multisensor image fusion using the wavelet transform. Graph Model Image Process 57(3):235–245
10. Li TJ, Wang YY (2011) Biological image fusion using a NSCT based variable-weight method. Inf Fusion 12(2):85–92
11. Liu Y, Liu S, Wang Z (2015) A general framework for image fusion based on multi-scale transform and sparse representation. Inf Fusion 24:147–164
12. Loeckx D, Slagmolen P, Maes F, Vandermeulen D, Suetens P (2010) Nonrigid image registration using conditional mutual information. IEEE Trans Med Imaging 29(1):19–29
13. Looney D, Mandic DP (2009) Multiscale image fusion using complex extensions of EMD. IEEE Trans Signal Process 57(4):1626–1630
14. Miao QG, Shi C, Xu PF, Yang M, Shi YB (2011) A novel algorithm of image fusion using shearlets. Opt Commun 284(6):1540–1547
15. Pajares G, De La Cruz JM (2004) A wavelet-based image fusion tutorial. Pattern Recogn 37(9): 1855–1872
16. Petrovic V, Xydeas C (2005) Objective evaluation of signal-level image fusion performance. Opt Eng 44(8)
17. Piella G, Heijmans H (2003) A new quality metric for image fusion. Proc Int Conf Image Process 3:173–176
18. Siu AMK, Lau RWH (2005) Image registration for image-based rendering. IEEE Trans Image Process 14(2): 241–252
19. Toet A, Franken EM (2003) Perceptual evaluation of different image fusion schemes. Displays 24(1):25–37
20. Wang Z, Bovik AC (2002) A universal image quality index. IEEE Signal Process Lett 9(3):81–84
21. Wang Z, Bovik AC (2006) Modern image quality assessment. Synth Lect Image Video Multimed Process 2(1):1–156
22. Yang SY, Wang M, Jiao LC, Wu RX, Wang ZX (2010) Image fusion based on a new contourlet packet. Inf Fusion 11(2):78–84
23. Zhenfeng S, Jun L, Qimin C (2012) Fusion of infrared and visible images based on focus measure operators in the curvelet domain. Appl Optics 51(12):1910–1921

Xiaoli Zhang received his MSc degree in computer science and technology from Jilin University, China, in 2012. He is currently a PhD candidate in college of computer science and technology, Jilin University. His research interests include information fusion, algorithm evaluation, and data mining. He has published more than 20 papers in journals and conferences.

Xiongfei Li received the BS degree in computer software in 1985 from Nanjing University, the MSc degree in computer software in 1988 from the Chinese academy of sciences, the PhD degree in communication and information system in 2002 from Jilin University. Since 1988, he has been a member of the faculty of the computer science and technology at Jilin University, Changchun, China. He is a professor of computer software and theory at Jilin University. He has authored more than 100 research papers. His research interests include data mining, intelligent network, image processing and analysis. Prof. Li is a member of the IEEE.

Yuncong Feng received her MSc degree in computer science and technology from Jilin University, China in 2014. She is currently a PhD candidate in college of computer science and technology, Jilin University. Her research interests include information fusion and data mining.