Infrared Physics & Technology 72 (2015) 266–275

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

A fast fusion scheme for infrared and visible light images in NSCT domain

Chunhui Zhao *, Yunting Guo, Yulei Wang

Information and Communication Engineering, Harbin Engineering University, Nantong Street, Harbin City, China

HIGHLIGHTS

- We propose a fast realization of NSCT instead of traditional NSCT for image decomposition.
- The computational efficiency is improved through combining the filters in each channel of NSCT.
- Once-for-all calculation of the global attribute is used to obtain higher efficiency.
- The fusion weights are determined by the difference between pixel and global image attribute.

ARTICLE INFO

Article history:
Received 27 May 2015
Available online 20 August 2015

Keywords:
Infrared images
Fast non-subsampled contourlet transform
Pixel information evaluation
Image fusion

ABSTRACT

Fusion of infrared and visible light images is an effective way to obtain a simultaneous visualization of details of background provided by visible light image and hiding target information provided by infrared image, which is more suitable for browsing and further processing. Two crucial components for infrared and visual light image fusion are improving its fusion performance as well as reducing its computational burden. In this paper, a novel fusion algorithm named pixel information estimation is proposed, which determines the weights by evaluating the information of pixel and is well applied in visible light and infrared image fusion with better fusion quality and lower time-consumption. Besides, a fast realization of non-subsampled contourlet transform is also proposed in this paper to improve the computational efficiency. To verify the advantage of the proposed method, this paper compares it with several popular ones in six evaluation metrics over four different image groups. Experimental results show that the proposed algorithm gets a more effective result with much less time consuming and performs well in both subjective evaluation and objective indicators.

1. Introduction

Due to different imaging mechanism and technical conditions, information contained in a single image or obtained by a single sensor cannot fully reflect the features of object and scene. Usually, it's difficult for human to observe two or more source images directly without losing any important information, or to locate target accurately when target and background are displayed in different source images. However, these tasks are widely appeared in military surveillance, industrial applications and medical imaging applications, etc. Therefore, image fusion technology is developed to integrate source images obtained by different sensors into a single one, which contains more complete description of the observation object and is more suitable for observation and further processing. In order to facilitate the tasks described above, an ideal fusion algorithm should preserve as much useful information as possible in source images within permissive time, which is our research goal.

To date, a variety of algorithms have been proposed for image fusion, which can be categorized into spatial domain analysis (SDA) and multi-scale analysis (MSA). Generally speaking, SDA methods determine the pixels' weights of source images directly according to a certain participation rate calculation rule, such as linear weighing based methods, artificial neural networks based methods [1] and guided filtering based weighted average method [2]. While MSA decomposes the source images into multi-scale components, then integrates these components to form a new multi-scale representation, finally, an inverse multi-scale transform is used to reconstruct the fused image. The classical multi-scale transforms include Laplacian pyramid (LP) [3] and discrete wavelet transform (DWT) [4]. To better capture the intrinsic

* Corresponding author. Tel.: +86 18845870228.
E-mail address: zhaochunhui@hrbeu.edu.cn (C. Zhao).

http://dx.doi.org/10.1016/j.infrared.2015.07.026
1350-4495/

geometrical structure of images, various multi-scale geometrical analysis methods have been developed, including ridgelet transform [5], curvelet transform [6], and contourlet transform (CT) [7]. These theories construct a framework of multi-scale and multi-direction analysis. Thereafter, non-subsampled contourlet transform (NSCT) [8], a shift-invariant version of contourlet transform, is presented to solve the pseudo-Gibbs phenomena around singularities, and provide an asymptotic optimal representation of images compared to other transform methods. However, due to its inefficient implementation, time-consuming of the NSCT is unacceptable, which limits its development. Since then, several new methods are presented to improve the fusion effect and reduce the time consumption of image fusion, such as multi-scale directional bilateral filter (MDBF) [9], sparse representation (SR) based fusion methods [10,11] and non-subsampled shearlet transform (NSST) [12,13]. In MDBF-based method, the authors use a bilateral filter instead of Gaussian filter for pyramid decomposition. It seems to be effective in preserving more edge information, but actually, it just transfers some of the high-frequency information into the low-frequency sub-band. Refs. [10,11] combine the SR theory and multi-scale transform (MST), adopt the SR-based fusion rule for the low-pass bands and obtained a better fusion result, but the problem of timeliness is still under resolution. Refs. [12,13] present a new algorithm called non-subsampled shearlet transform, but its decomposition result is similar to NSCT, and it costs more processing time when compared with the fast NSCT presented in this paper.

To solve the problem discussed above, this paper presents a novel fusion algorithm based on a fast NSCT realization. The outline of this paper is presented as follows. In Section 2, the principle of NSCT is reviewed. Section 3 analyzes the features of the NSCT, and puts forward its fast realization. The proposed algorithm for visual light and infrared image fusion is described in details in Section 4. Section 5 discusses the experimental results and performance analysis. Conclusion and further work are given in Section 6.

2. Non-subsampled contourlet transform

Non-subsampled contourlet transform (NSCT), based on contourlet transform theory, utilizes contour segment to approximate the source image [14]. It can be divided into non-subsampled pyramid (NSP) and non-subsampled directional filter bank (NSDFB). The former step ensures the multi-scale feature of NSCT, and the latter step ensures its multi-direction feature. Fig. 1 shows the structure of the NSCT.

2.1. Non-subsampled pyramid

Non-subsampled pyramid (NSP), completed by a two-channel non-subsampled filter bank, produces high- and low-frequency sub-bands at each decomposition level, and the low-frequency sub-band is filtered iteratively to get the subsequent stages of decomposition. As a result, a K-level NSP produces $K+1$ sub-bands with the same size as source image, consisting of K high-frequency and one low-frequency sub-bands. Fig. 2 shows the three level NSP decomposition, where $H_0\left(z^{2^k I}\right)$ $(k=0,1,2)$ denotes low-pass filter bank and $H_1(z^{2^k I})$ $(k=0,1,2)$ denotes high-pass filter bank, k is the decomposition level, and the gray region means the frequency pass-band of each NSP decomposition stage.

In NSP decomposition, the ideal frequency support of low-pass filter of k stage is $\left[-(\pi/2^k),(\pi/2^k)\right]^2$, and the ideal support of high-pass filter in the same stage is $\left[-(\pi/2^{(k-1)}),(\pi/2^{(k-1)})\right]^2/\left[-(\pi/2^k),(\pi/2^k)\right]^2$.. The filters of subsequent NSP stages are

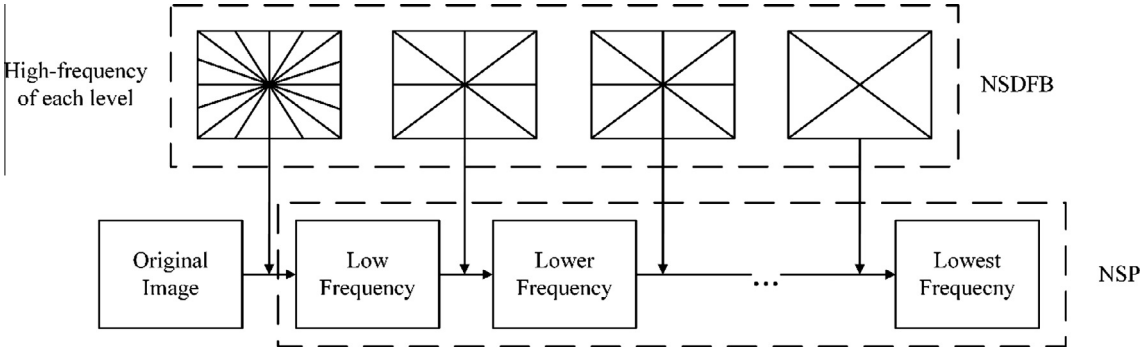

Fig. 1. Block diagram of NSCT.

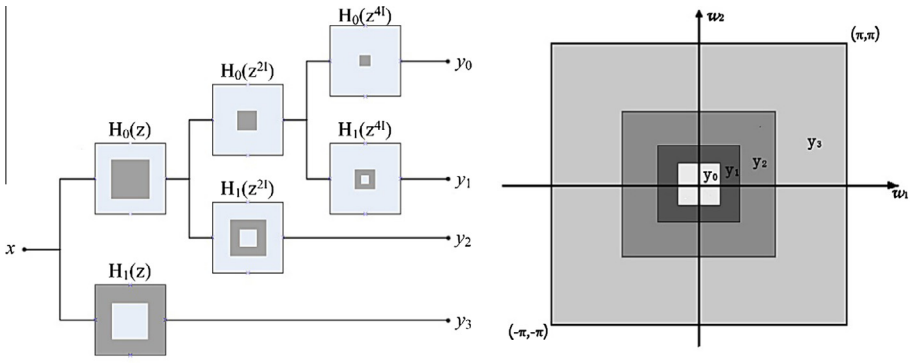

Fig. 2. Three-stage NSP and its frequency domain support.

obtained by up-sampling the filters of the previous stage. The filters of the K-stage NSP are defined as:

$$H_n^{\mathrm{eq}}(z)\begin{cases} H_1\left(z^{2^{n-1}}\right)\prod_{k=0}^{n-2}H_0\left(z^{2^k}\right), & 1 \leqslant n \leqslant 2^K \\ \prod_{k=0}^{n-2}H_0\left(z^{2^k}\right), & n = 2^K \end{cases} \tag{1}$$

where z^k denotes $[z_1^k, z_2^k]$.

2.2. Non-subsampled directional filter bank

Non-subsampled directional filter bank (NSDFB) is a tree-structure non-subsampled filter bank which is composed of a series of directional fan filters. It provides us a wealth of directional details information by producing 2^l directional sub-bands with the same size as source image, where l is the number of NSDFB decomposition stage. A 2-level NSDFB and its frequency decomposition is illustrated in Fig. 3, where $U_0(z)$ and $U_1(z)$ are fan filters, and $U_0(z^Q)$ and $U_1(z^Q)$ are quadrant filters.

In NSDFB, all the filters can be produced by rotating a series of fan filters with the same computational complexity as the original filter, i.e. all filters are produced by up-sampling the original filter by a quincunx matrix given by $Q = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

3. Fast realization of NSCT

The acknowledged disadvantage of NSCT is its huge computation cost, resulting from the inefficient iterative filtering in NSCT. As described in part 2, traditional NSCT is divided into two parts, and the source image which will be decomposed need to be processed through NSP and NSDFB one by one. Thus, in order to generate a directional decomposition sub-band, the source image needs to be convoluted with a set of band-pass filters in NSP and a set of fan filters in NSDFB. However, if a directional decomposition sub-band can be obtained by once filter, it will be much more effective than the traditional one. Therefore, we illustrate the realization of our fast NSCT (FNSCT) in a channel of NSCT filter bank.

As shown in Fig. 4, $H_0(z)$ and $H_1(z^2)$ denote the low-pass filter and high-pass filter in NSP respectively, while $U_0(z)$ and $U_0(z^Q)$ denote the fan filter and quadrant filter in NSDFB respectively. One of the final decomposed image is obtained by filtering the source image with these four filters one by one. However, due to the size difference between source image and digital filters, convolution of the filters is more efficient than multiple image filtering. Therefore, filters in each channel of NSCT filter bank ($H_0(z), H_1(z^2), U_0(z)$ and $U_0(z^Q)$) are combined into a single decomposition filter $F(z)$ of FNSCT.

$$F(z) = H_0(z) * H_1(z^2) * U_0(z) * U_0(z^Q) \tag{2}$$

Thus, the tree composed filter bank of NSCT is reconstructed into a multi-channel filter bank, and the number of filters is reduced from $2^{K+2} - 4$ to $2^{K+1} - 1$ (assume that the structure of NSCT filter bank is a binary tree), where K denotes the decomposition level. Fig. 5 shows a comparison of the two-level decomposition filter banks of traditional NSCT and FNSCT, where the light gray region denotes the frequency pass-band of filters in each step, and the dark gray region denotes the frequency band of each decomposition sub-band. Since the filters of FNSCT are constituted

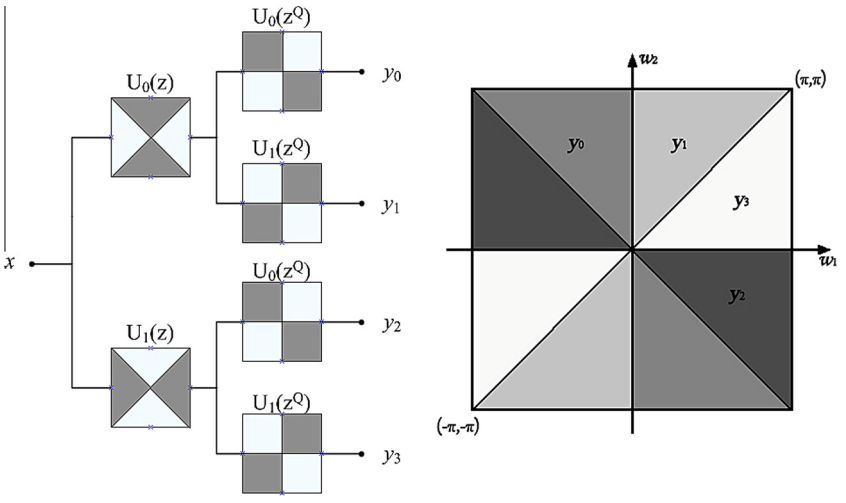

Fig. 3. Four-channel NSDFB and its frequency decomposition.

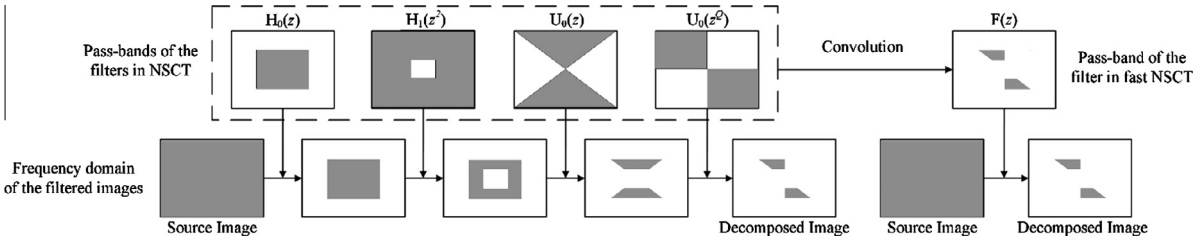

Fig. 4. Illustration of a channel of the traditional NSCT and fast NSCT.

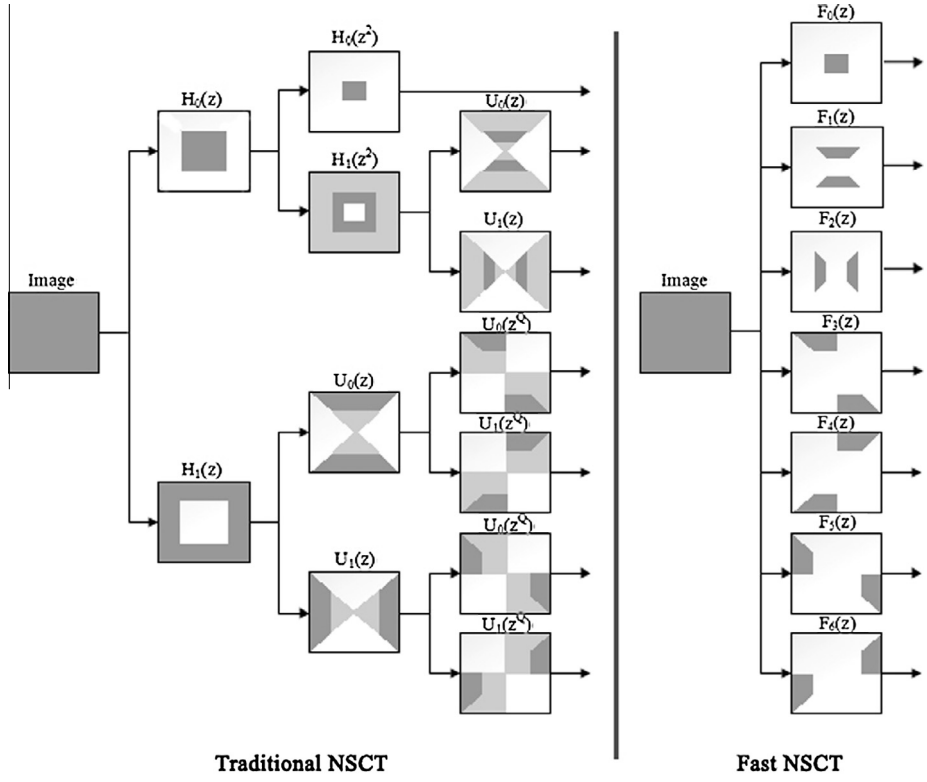

Fig. 5. Comparison of traditional NSCT and FNSCT in a two-level decomposition.

by filters of NSCT, their frequency support are intersections of the filters' frequency support in each channel of traditional NSCT, let ψ_i be the pass-band of filter $F_i(z)$, it satisfies that

$$\sum_{i=1}^{2^{K+1}-1} |\psi_i| = 1 \tag{3}$$

and

$$\psi_i \cap \psi_j = 0, \quad i \neq j \tag{4}$$

According to Figs. 4 and 5, the steps of FNSCT are summarized as follows:

(1) Initialization: Design the low-pass filter $H_0(z)$ and high-pass filter $H_1(z)$ in NSP, and fan filters $U_0(z)$ and $U_1(z)$ in NSDFB. Determine the NSP stage n and direction decomposition level l_n in each NSP decomposition stage.
(2) Adjust the filters: According to n and l_n, adjust filters $H_0(z)$ and $H_1(z)$ into multi-scale filters and adjust filters $U_0(z)$ and $U_1(z)$ into multi-direction filters by up-sampling and quincunx resampling matrix descripted in Section 2.
(3) Obtain the decomposition filter bank: Combine the filters in each processing flow into a final filter with different scales and directions.
(4) Obtain the decomposition sub-bands: Filter the source image with the decomposition filters to obtain the final decomposition sub-bands.

It should be noted in step (3) that, size of the combining digital filters will increase because of the convolution operation, which will influence the computation performance of FNSCT. Fortunately, all the filters in NSCT have wave attenuation property. Thus, though size of the combining digital filter is increasing, most weights of the extended region are closed to zero, wiping off the weights less than a small threshold, which is set to 10^{-6} in this paper, will ensure the efficiency without influencing the decomposition performance. In order to verify the decomposition and reconstruction accuracy of our proposed FNSCT, we give the mean error (ME), which defined as formula (5), between the sub-bands of FNSCT and traditional NSCT over a 3-level decomposition in Table 1. The statistical data shows that the cutoff of filter with a small threshold will not affect the decomposition and reconstruction performance.

$$\mathrm{ME} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|\mathrm{FNSCT}(i,j) - \mathrm{TNSCT}(i,j)| \tag{5}$$

Similar steps are used in the inverse transform to obtain the reconstruction filter bank, then add the convolution result of the decomposition sub-bands and reconstruction filters to obtain the final reconstructed image.

In this way, each directional decomposition sub-band can be obtained by once filter, moreover, it is noteworthy that the image data is much huger than decomposition filters, time consumption of combining the filters can be ignored when compared with image processing. In order to verify the effectiveness of FNSCT, Table 2 gives the comparison of the processing time (including decomposition processing and reconstruction processing) of traditional NSCT and FNSCT. The comparison result shows that the proposed FNSCT is able to save about half processing time, and is robustness for different image sizes and decomposition levels.

4. Proposed fusion algorithm

Pixel level image fusion integrates the pixels of source images into one fused pixel directly. The calculation of the fused pixel is accordance with the significance of the source pixels (it may include pixel's neighbor information). Thus, no matter what the fusion rule is, it can be summarized as a calculation of the weights

Table 1
ME between the sub-bands of FNSCT and traditional NSCT.

Sub-bands	Low-frequency	1-level decomposition	2-level decomposition	3-level decomposition	Reconstruction image
Average ME	2.18×10^{-16}	2.14×10^{-5}	5.62×10^{-5}	7.12×10^{-5}	2.04×10^{-4}

Table 2
Comparison of the processing time(s) of traditional NSCT and FNSCT.

Image size	Decomposition level	Traditional NSCT	Fast NSCT	Percentage reduction (%)
$256 * 256$	2	0.4360	0.1946	55.4
	3	1.1276	0.5554	50.8
	4	3.7780	1.9717	47.8
	5	9.9746	5.4761	45.1
$632 * 496$	2	1.6003	0.6259	60.9
	3	5.1478	2.2559	56.2
	4	16.7599	8.1675	51.3
	5	42.7214	22.5421	47.2

of source pixels, and these weights reflect the significances of the source pixels.

The basic requirement for image fusion is to retain the information present in source images as much as possible. Hence, pixels with larger information content require higher weights in fusion step to keep more useful information, while pixels with less information content are regarded as useless information and need to be removed. Therefore, we propose a novel criterion based on Shannon information theory to evaluate the information content of the pixels. The proposed criterion is based on the following two postulates:

(1) The less pixels with similar value in the source image, the more information these pixels have.
(2) The greater the difference between the pixels' gray values and the image's background brightness, the greater the amount of information these pixels have.

For postulate (1), less pixels means lower appearing probability and higher information entropy, which means these pixels contain more information. It is also worth noting in postulate (2) that background and target are distinguished according to the gray contrast in infrared image, larger contrast between pixels and background means the pixels are more likely to be targets, they contain more information and are more important. Similar conclusions are obtained in visual light image.

Mathematically, aforementioned postulates are given as Eqs. (6) and (7):

$$\mathrm{Inf}[I(i,j)] \propto 1/\mathrm{num}\left[R_{I(i,j)}\right] \tag{6}$$

where $I(i, j)$ denotes the pixel at location (i, j), $\mathrm{Inf}[x]$ denotes the information content of x, $R_{I(i,j)}$ denotes the pixels with approximate value to $I(i, j)$, $\mathrm{num}[X]$ denotes the number of the pixels in region X.

$$\mathrm{Inf}[I(i,j)] \propto \mathrm{diff}[I(i,j), G] \tag{7}$$

where G denotes global feature of image I, $\mathrm{diff}[x, y]$ denotes the difference between x and y.

In the following part, we will describe the proposed Pixel Information Estimation (PIE) fusion rule in detail.

4.1. Fusion rule of high frequency

In NSCT, the high-frequency sub-bands reflect the texture or edges information, they are small in smooth region of source image, but will increase dramatically in edges. Different coefficient values reflect different object features, and need different rules to fuse them. Thus, $c_{vis}(m, n)$ and $c_{inf}(m, n)$, which are high-frequency coefficients at row m and column n of the visual light and infrared images respectively, with approximation value are considered to be redundant data. It means that they both come from smooth regions or edges, weighted summation is applicative to obtain the fused coefficient. On the other hand, $c_{vis}(m, n)$ and $c_{inf}(m, n)$ with different characteristics are considered to be complementary information, which means that one of them comes from edges and the other comes from smooth region, and the larger one will be preserved. Therefore, a threshold should be found to distinguish the smooth regions and the edges, and different rules are used in different parts. The proposed fusion rule PIE for high-frequency (PIE-H) is presented as follows:

(1) Calculate the threshold T for each high-frequency bands to distinguish the edges and smooth regions.

$$T = (\mu_h + M_h) \times (2^k - 1)/2^{k+1} \tag{8}$$

where k is the decomposition level, μ_h and M_h denote the mean and median of high-frequency coefficients respectively. It should be noted in formula (8) that it consists of two parts, $(\mu_h + M_h)/2$ and $(2^k - 1)/2^k$. Since the high-frequency coefficients of edge area are much larger than smooth area, threshold simply obtained by mean operation may be too larger to enhance the coefficients of textures, which usually is just a bit larger than the average value. So the average of mean and median of the coefficients will lead to a better performance. In addition, coefficients of different decomposition level have different statistical property. Lower decomposition level sub-bands usually contains more prominent edges information and leads to a too-high threshold, the weight factor $(2^k - 1)/2^k$ is used to inhibitory this phenomenon.

(2) Calculate the weights according to the coefficients' characteristics. The weights of coefficients larger than threshold T have an exponential increase to ensure their significances, and weights of coefficients less than threshold T are set to the coefficients' value to achieve a better weighted mean effect.

$$\omega_{vis} = \begin{cases} e^{(\gamma_H \cdot |I_{vis}(i,j)| - T_{vis}|)} & |I_{vis}(i,j)| > T_{vis}, |I_{inf}(i,j)| > T_{inf} \\ |I_{vis}(i,j)| & |I_{vis}(i,j)| < T_{vis}, |I_{inf}(i,j)| < T_{inf} \\ 1 & |I_{vis}(i,j)| > T_{vis}, |I_{inf}(i,j)| < T_{inf} \\ 0 & |I_{vis}(i,j)| < T_{vis}, |I_{inf}(i,j)| > T_{inf} \end{cases} \tag{9}$$

$$\omega_{inf} = \begin{cases} e^{(\gamma_H \cdot |I_{inf}(i,j)| - T_{inf}|)} & |I_{vis}(i,j)| > T_{vis}, |I_{inf}(i,j)| > T_{inf} \\ |I_{inf}(i,j)| & |I_{vis}(i,j)| < T_{vis}, |I_{inf}(i,j)| < T_{inf} \\ 0 & |I_{vis}(i,j)| > T_{vis}, |I_{inf}(i,j)| < T_{inf} \\ 1 & |I_{vis}(i,j)| < T_{vis}, |I_{inf}(i,j)| > T_{inf} \end{cases} \tag{10}$$

where subscript *vis* and *inf* denotes visible light image and infrared image respectively, γ is a parameter to control weight, and in this paper, γ_H is set to 12.

(3) Obtain the fused coefficients according to the weights obtained in step (2).

$$F(i,j) = \frac{\omega_{vis} \times I_{vis}(i,j) + \omega_{inf} \times I_{inf}(i,j)}{\omega_{vis} + \omega_{inf}} \tag{11}$$

where $F(i,j)$ denotes the fusion pixel in location (i,j).

4.2. Fusion rule of low frequency

The low-frequency sub-band is the overview of source image, which contains the essence interpretations of the source images, such as gray-level distribution, main gray scale contrast and main object of observation and so on. Fusion result of the low-frequency sub-band is directly related to the quality of the image fusion, i.e. an effective method is crucial important.

In NSCT, low-frequency sub-band is severely blurred by NSP, which means that a small neighborhood of a pixel does not provide more information than pixel itself, wasting time in calculating the neighborhood information of pixel is unnecessary. In addition, the popular region energy based fusion rules overlook the dark targets, and region variance based fusion rules only enhance the edge pixels instead of target pixels. Therefore, in order to solve the problem above, PIE for low-frequency is proposed, which reduce calculation amount by calculating the global attribute once-for-all instead of calculating each pixel's neighborhood information, and determines the pixels' weights by utilizing the differences between pixels and image's background brightness. PIE-L is departed into three steps as follows:

(1) Calculate the global attribute G, which reflects the general background brightness of the source image.

$$G = \frac{\mu_l + M_l}{2} \tag{12}$$

where μ_l and M_l denote the mean and median of the gray values of low-frequency sub-band respectively.

(2) Calculate the weight of each pixel form source images according to the differences between pixel's value and image's global attribute G.

$$\omega_{vis} = e^{\gamma_L \cdot |I_{vis}(i,j) - G_{vis}|} \tag{13}$$

$$\omega_{inf} = e^{\gamma_L \cdot |I_{inf}(i,j) - G_{inf}|} \tag{14}$$

where γ_L is set to 4, and the natural exponential function is used to wipe off the error points (such as zero and null point) and highlight the object.

(3) Calculate the fusion coefficients.

$$F(i,j) = \frac{\omega_{vis} \times I_{vis}(i,j) + \omega_{inf} \times I_{inf}(i,j)}{\omega_{vis} + \omega_{inf}} \tag{15}$$

4.3. Fusion scheme based on FNSCT and PIE

According to above-mentioned parts, the schematic diagram of proposed fusion algorithm is illustrated in Fig. 6. The source images are firstly decomposed by FNSCT to obtain their multi-scale and multi-direction decompositions. Then these decompositions are combined by PIE-L and PIE-H to obtain the fused sub-bands. Finally, inverse FNSCT is applied to the fused sub-bands to reconstruct the fused image. More formally, the steps of the proposed fusion scheme are given as follows:

Step 1: image decomposition. For each source image, a K level FNSCT is used to obtain one low-frequency sub-band and $2^{K+1} - 2$ high-frequency sub-bands.

Step 2: coefficient fusion. For coefficients of low-frequency sub-bands coming from different source images, PIE-L discussed in Section 4.2 is used to fuse them into a single one; and the $2^{K+1} - 2$ fused high-frequency sub-bands are obtained by PIE-H discussed in Section 4.1.

Step 3: image reconstruction. The inverse FNSCT of fused one low- and $2^{K+1} - 2$ high-frequency sub-bands generates the final fusion image.

5. Experimental results and discussions

In this section, four different infrared and visible image datasets are provided to demonstrate the performance of the proposed algorithm. For each group, the proposed method is compared with five current excellent algorithms which are NSCT and PCNN based fusion method [15], multi-scale directional bilateral filter (MDFB) based method [9], SR based method [10], MST–SR based method [11] and NSST based method [16], they are all novel and representative in non-linear fusion, image decomposition representation and neighborhood information based fusion respectively. The decomposition level is set to 3 (if needed), and all the parameters settings in these methods are still implemented according to the content in references. Besides, all of the experiments are conducted in MATLAB 2010b on a PC with Intel Core i7/2.1 GHz/4G.

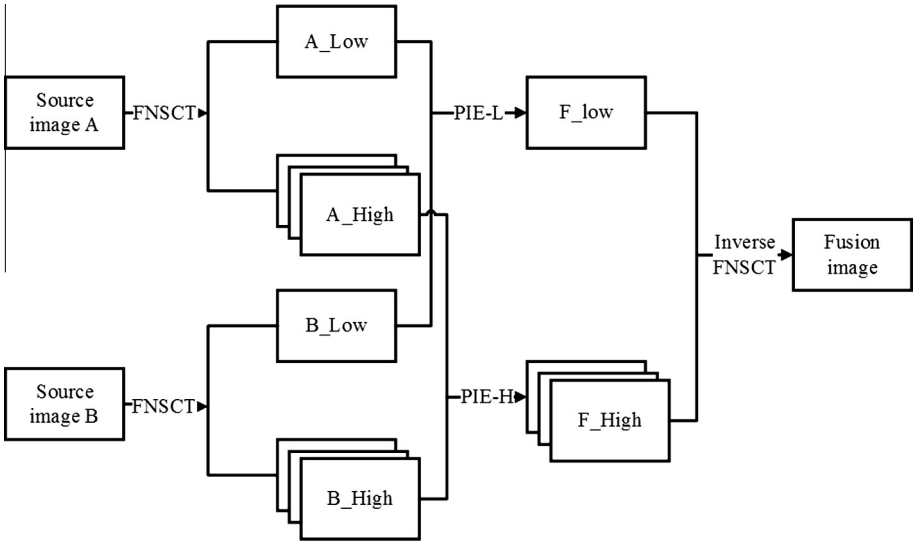

Fig. 6. Schematic diagram of the proposed image fusion algorithm.

These four image datasets used in fusion of visible light and infrared images come from "UN Camp", "Bristol Queens Road", "Octec" and "Trees" image set. They cover 256 gray levels, and have been widely used in many related Refs. [10,11,16], etc. For more accurate results, watermark in image set "Octec" is cropped.

5.1. Subjective quality assessment

Figs. 7–10 list the comparisons of the fusion results from the four image groups. In these figures, (a) and (b) are the original visual light and infrared images, (c)–(f) are the fusion results based on the four different fusion algorithms mentioned above, including the proposed algorithm.

Fig. 7 is a comparison of fusion results of "UN camp" image set. Obviously, the visible light image well describes the environmental information, and infrared image emphasizes the target information, a person hiding in trees. As we can see, all the methods preserve the main information and characteristics of source images, but significant differences are still exist in the details. Sub-image (c) has a good visual effect and layering, but there are severely blurred in most regions, especially the fence in the left side of the image, that is because PCNN emphasizes the visual contrast instead of texture details. Contrast of sub-image (d) is poor because the averaging fusion rule used for low-frequency sub-band narrows the range of gray-level. Sub-image (e) blurs most of the texture information and sub-image (f) has inappropriate gray level over the shrub region. Sub-image (g) is too light to reproduce the real gray level in source images, it implies that NSST based method is weak in keeping the real gray-scale of source images. And sub-image (h), has the best performance in all respects including contrast, sharpness, target discrimination and edge details, it implies that PIE is suitable for infrared and visible light images fusion.

Fig. 8 shows the street night of Bristol Queens Road, in the scene of week light, the visible light image can only show lights and lighted billboards, but in infrared image, much more details are displayed. As for the fusion result, sub-image (c) preserves strong edge well, but lost most of the information in weak edge. Edges and texture in sub-image (d) are clear, but its contrast is unacceptable. Both sub-image (e) and sub-image (f) have poor fusion performance on the poster boards at the up-left corner of

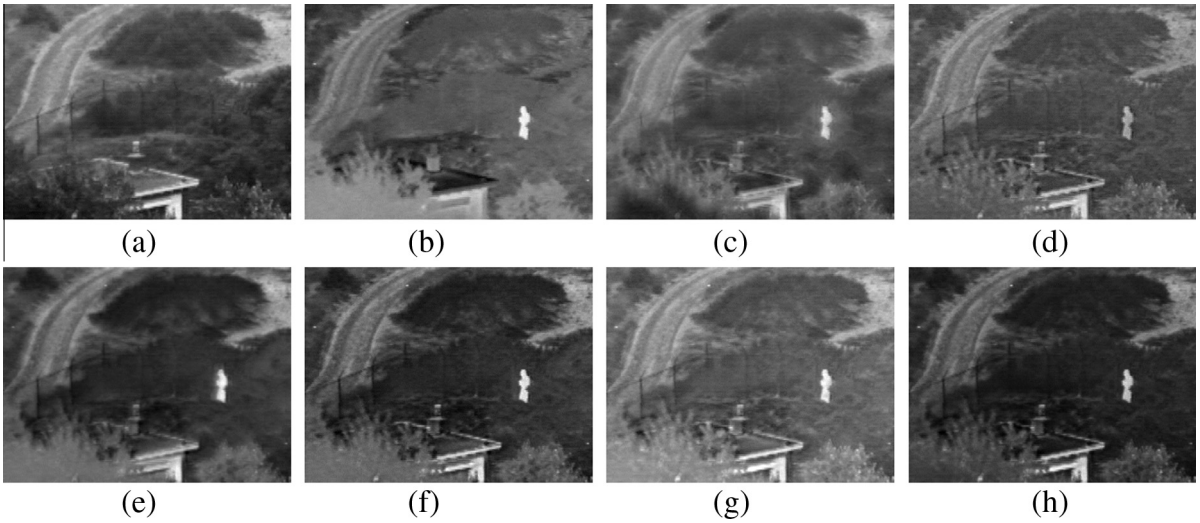

Fig. 7. Subjective fusion results of "UN camp". (a) Visible light image; (b) infrared image; (c) NSCT–PCNN based fusion result; (d) MDBF based fusion result; (e) SR based fusion result; (f) MST–SR based fusion result; (g) NSST based fusion result; (h) fusion result of PM.

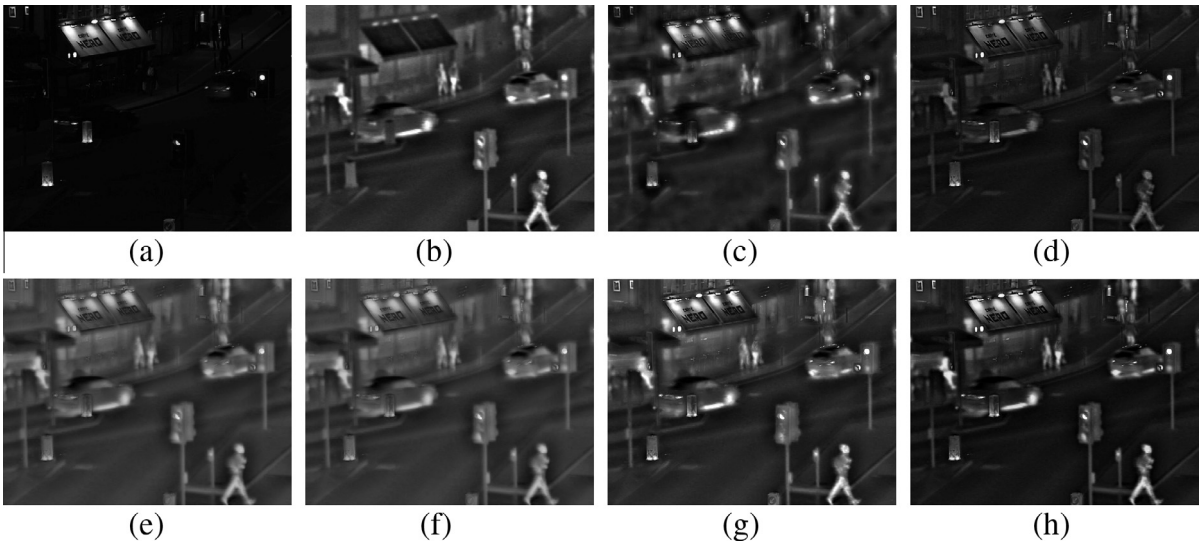

Fig. 8. Subjective fusion results of "Bristol Queens Road". (a) Visible light image; (b) infrared image; (c) NSCT–PCNN based fusion result; (d) MDBF based fusion result; (e) SR based fusion result; (f) MST–SR based fusion result; (g) NSST based fusion result; (h) fusion result of PM.

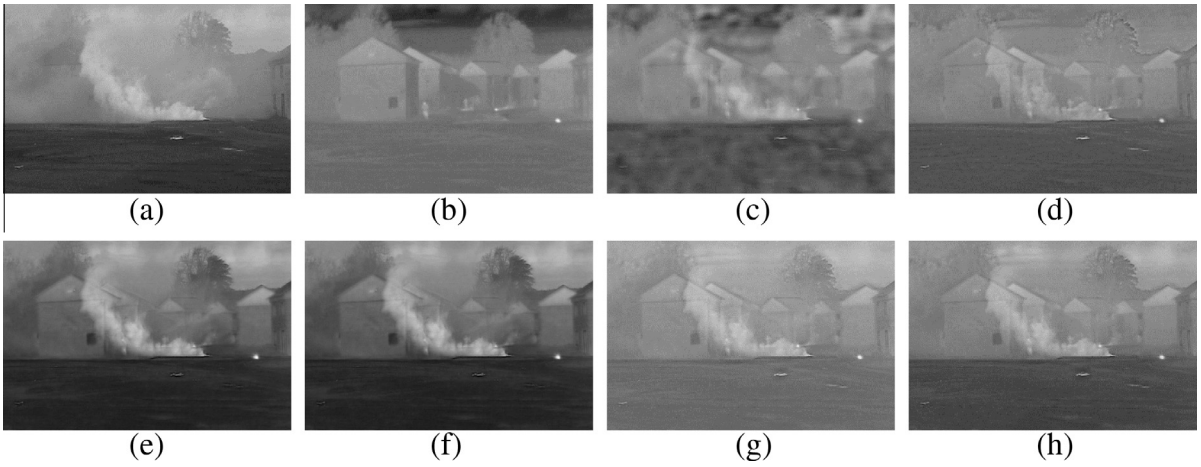

Fig. 9. Subjective fusion result of "Octec". (a) Visible light image; (b) infrared image; (c) NSCT–PCNN based fusion result; (d) MDBF based fusion result; (e) SR based fusion result; (f) MST–SR based fusion result; (g) NSST based fusion result; (h) fusion result of PM.

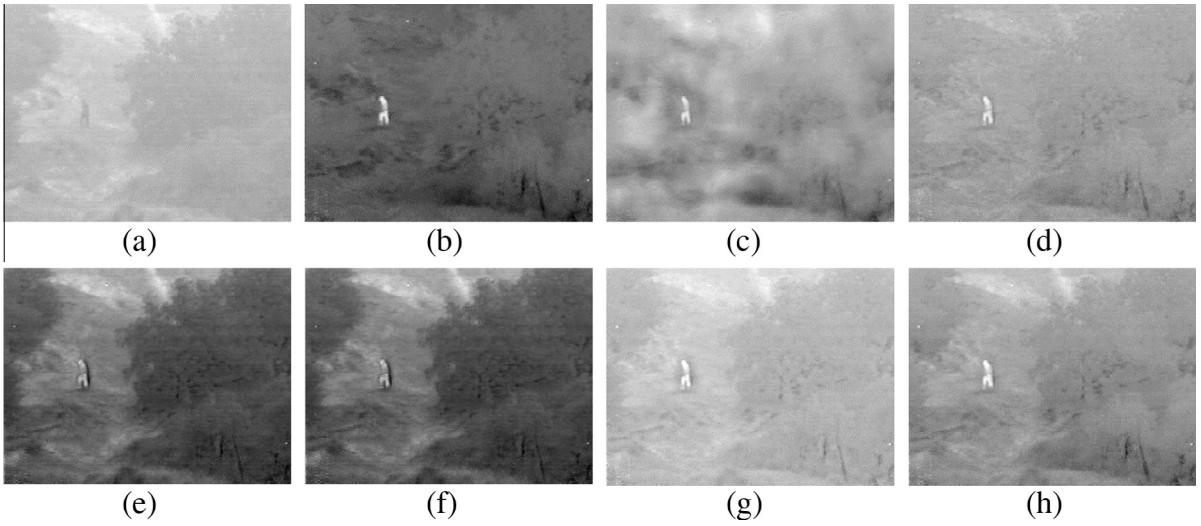

Fig. 10. Subjective fusion result of "Trees". (a) Visible light image; (b) infrared image; (c) NSCT–PCNN based fusion result; (d) MDBF based fusion result; (e) SR based fusion result; (f) MST–SR based fusion result; (g) NSST based fusion result; (h) fusion result of PM.

Table 3
Comparison of the fusion rules for "UN Camp" image.

Method	MI	Q	Q_E	Q_{ABF}	VIF	SSIM
NSCT–PCNN	2.7279	0.7839	0.4898	0.6911	0.4683	0.7293
MDBF	2.5679	**0.8359**	0.5198	0.7145	0.5270	0.7343
SR	**3.0088**	0.7321	0.4604	0.7051	0.6729	0.7264
MST–SR	3.0052	0.7735	0.4594	0.7202	0.8071	0.7118
NSST	2.2695	0.7863	0.4942	0.7136	**0.8433**	0.6807
PM	2.8809	0.8252	**0.5249**	**0.7209**	0.7336	**0.7354**

Table 4
Comparison of the fusion rules for "Bristol Queens Road" image.

Method	MI	Q	Q_E	Q_{ABF}	VIF	SSIM
NSCT–PCNN	**1.8739**	0.5809	0.2645	0.6990	0.7110	0.6298
MDBF	1.3696	0.6280	0.3004	0.7359	0.6209	0.6664
SR	1.3058	0.6102	0.2939	0.7287	0.4635	0.6686
MST–SR	1.3296	0.6450	0.3275	**0.7710**	0.4724	0.6716
NSST	1.3566	0.6280	0.3045	0.7013	0.8398	0.6211
PM	1.2999	**0.6475**	**0.3283**	0.7407	**0.9653**	**0.6719**

Table 5
Comparison of the fusion rules for "Octec" image.

Method	MI	Q	Q_E	Q_{ABF}	VIF	SSIM
NSCT–PCNN	4.0606	0.6852	0.4221	0.6647	0.7154	0.8287
MDBF	1.9860	0.6410	0.3876	0.6435	0.7888	0.8488
SR	**4.4977**	0.6493	0.4093	0.6414	0.8374	0.8274
MST–SR	4.4454	0.6575	0.3638	0.6608	**0.8729**	0.8137
NSST	2.2819	0.7033	0.4397	0.6771	0.6810	0.8621
PM	2.5318	**0.7114**	**0.4490**	**0.6786**	0.7917	**0.8803**

Table 6
Comparison of the fusion rules for "Trees" image.

Method	MI	Q	Q_E	Q_{ABF}	VIF	SSIM
NSCT–PCNN	1.6797	0.6903	0.4198	0.6731	0.7771	0.8148
MDBF	1.5589	0.7500	0.4532	0.7076	0.5543	0.8364
SR	**3.0210**	0.7028	0.3961	0.6835	0.7364	0.7483
MST–SR	2.8028	0.7041	0.3806	0.6984	0.7680	0.7396
NSST	2.3440	0.7267	0.4358	0.7119	0.6679	0.8020
PM	2.0703	**0.7617**	**0.4668**	**0.7294**	**0.8109**	**0.8486**

Table 7
Comparison of processing time (s).

Image set	Size	NSCT–PCNN	MDBF	SR	MST–SR	NSST	PM
UN Camp	320 * 240	13.9831	10.2916	37.5816	38.7776	3.0317	**1.3272**
Bristol Queens Road	632 * 496	54.8938	36.5106	157.0087	158.5251	8.4827	**5.2127**
Octec	640 * 421	51.9926	32.1598	135.9638	137.4814	8.0112	**4.8921**
Trees	360 * 270	18.4469	13.4674	44.7191	47.3456	3.7058	**1.5050**

the image. Sub-image (g) and sub-image (h) inherited the useful information in the source image perfectly, but the contrast of sub-image (h) is much better.

More comparisons are shown in Figs. 9 and 10, and the advantage of our proposed method is clearer. Fusion results of NSCT–PCNN displayed in Figs. 9(c) and 10(c) have serious vaporific blur and lose almost all of the background details. Contrasts of Figs. 9(d) and 10(d) are too low to distinguish the target and background clearly, and the edges in Fig. 9(d) are not natural enough. The SR and MST–SR based method performance well in contrast of fused image, but all of the targets (humans) are unclear. As for NSST based method, loss of the dark information is more obvious in these two image sets, intensity distortion of the ground in Fig. 9(g) and the unapparent trunk in Fig. 10(g) show its weakness in dark information preserving. And compared with all these methods, the proposed method provides more satisfactory results in all these aspects.

5.2. Objective quality assessment

The subjective visual perception gives us direct comparisons, but in some cases it may be difficult to judge which one is the best. Whereas, an objective evaluation can provide us with quantitative analysis of the fusion quality. In this part, six quality metrics, including Q, Q_E, Q_{ABF}, MI (mutual information), VIF (visual information) and SSIM (structural similarity), are used to evaluate the performance of above algorithms. The metric Q gives an overall distortion evaluation of loss of correlation, luminance distortion and contrast distortion [17]. Both metrics Q_E [18] and Q_{ABF} [19] evaluate the edge information in the fused image. And MI evaluates the amount of information transferred from source images to fused image [20]. The VIF [21] quantifies the distortions including additive noises, blurs, and global or local contrast changes. And SSIM [22], which combines the luminance distortion, contrast distortion and structure distortion between fused image and source images, can accurately measure the fusion quality. It is noteworthy that the larger the values of these metrics mentioned above, the better the fusion quality is. Finally, comparisons of each image group are showed in Tables 3–6 respectively, and the best metrics are highlight in bold.

In Table 3, the evaluation values of PM are much better than others except Q and VIF index, which means the fusion result of PM has richer information, more obvious contrast and less structure distortion. As for Q index, though MDBF is higher than PM, it should be noted that the difference is very slight, and the superiorities of PM in other indexes are obvious. Similar conclusion can be obtained in VIF metric, though NSST based method outperforms our proposed method in VIF, it's obviously that our method has a better overall performance. In Tables 4–6, advantages of the proposed method are clearer, it achieves the highest score in nearly all of the metrics. Besides, though SR and MST–SR based method performance well in MI index over three image sets, they cost much more time and are slightly inferior to PM in the comprehensive evaluation.

Based on the statistics in Tables 3–6 and the above analysis, it is obvious that the proposed method is effective in preserving useful information, reducing structural distortion and keeping a reasonable contrast, and it is superior to current popular image fusion methods from the visual analysis and quantitative evaluation.

Finally, comparison of the average processing time of these four methods working in 3-stage decomposition is shown in Table 7 to prove the efficiency of the proposed method.

In Table 7, NSCT–PCNN adopts traditional NSCT to decompose the source images and PCNN to fuse the coefficients, its time consumption is unacceptable. Fusion rule of MDBF is the fastest one over these four methods, but it cost too much time in multi-scale directional bilateral filter, which affects its efficiency. The sliding window and OMP algorithm using in SR and MST–SR method lead to extremely high computational complexity, and cost even more time than NSCT–PCNN based method. NSST is faster than the first two methods, but compared with the proposed one, it cost more time in both images decomposition stage and coefficients fusion stage. Statistics shown in Table 7 prove the efficiency of the proposed method.

6. Conclusions and future work

Different types of sensors have different explanations for the same scene, which can be integrated to construct a more complete expression of the scene. So it is necessary to fuse the complementary information into the final fused image, and it is significant in both civilian and military applications. In order to effectively conduct the fusion image, a novel fusion algorithm as well as a fast realization of NSCT is proposed in this paper, which improves the efficiency of tradition NSCT and provides an effective fusion rule named PIE for coefficient fusion, and achieves a better fusion result with less time consumption when compared with several current popular fusion algorithms.

Future work will involve finding a better way to express image and optimizing the fusion quality. The former is helpful to extract the image feature, the latter is directly related to the fusion effect and time consumption. Consequently, the above two facets will be the emphasis of future work.

Conflict of interest

There is no conflict of interest.

Acknowledgements

This study is partially supported by the National Science Foundation of China (No. 61405041), the China Postdoctoral Science Foundation (Grant No. 2014M551221), the Heilongjiang Postdoctoral Science Found (Grant No. LBHZ13057), the Key Program of Heilongjiang Natural Science Foundation (No. ZD201216), Program Excellent Academic Leaders of Harbin (No. RC2013XK009003) and the Fundamental Research Funds for the Central Universities (No. HEUCF1508).

References

[1] A. Malek, M. Yashtini, Image fusion algorithms for color and gray level images based on LCLS method and novel artificial neural network, Neurocomputing 73 (4–6) (2010) 937–943.
[2] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013) 2864–2875.

[3] P.J. Burt, E.H. Andelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[4] H. Li, B. Manjunath, S. Mitra, Multisensor image fusion using the wavelet transform, Graph. Models Image Process. 57 (3) (1995) 235–245.
[5] N.D. Minh, V. Martin, The finite ridgelet transform for image representation, IEEE Trans. Image Process. 12 (1) (2003) 16–28.
[6] F. Nencini, A. Garzelli, et al., Remote sensing image fusion using the curvelet transform, Inform. Fusion 8 (2) (2007) 143–156.
[7] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
[8] A.L. Cunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.
[9] J. Hu, S. Li, The multiscale directional bilateral filter and its application to multisensor image fusion, Inform. Fusion 13 (2012) 196–206.
[10] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Inform. Fusion 13 (1) (2012) 10–19.
[11] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inform. Fusion 24 (1) (2015) 147–164.
[12] G. Easley, D. Labate, W.Q. Lim, Sparse directional image representation using the discrete shearlet transforms, Appl. Comput. Harmon. Anal. 25 (1) (2008) 25–46.
[13] W. Kong, J. Liu, Technique for image fusion based on NSST domain improved fast non-classical RF, Infrared Phys. Technol. 61 (2013) 27–36.
[14] J.S. Shan, W. Yang, X.M. Zhang, Infrared Image Processing Analysis and Fusion, Science Press, Beijing, 2009, pp. 169–179.
[15] T. Xiang, Y. Li, R. Gao, A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain, Infrared Phys. Technol. 69 (2015) 53–61.
[16] W. Kong, Technique for gray-scale visual light and infrared image fusion based on non-subsampled shearlet transform, Infrared Phys. Technol. 63 (2014) 110–118.
[17] Z. Wang, A. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.
[18] G. Piella, H. Heijmans, A new quality metric for image fusion, in: Proceedings of the Tenth International Conference on Image Processing, Barcelona, Spain, 2003.
[19] C.S. Xydeas, V. Petrović, Objective image fusion performance measure, Electron. Lett. 36 (4) (2000) 308–309.
[20] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315.
[21] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[22] SSIM index code webpage, <http://www.ece.uwaterloo.ca/~z70wang/research/ssim/>.