

A Noisy Infrared and Visible Light Image Fusion Algorithm

Yu Shen*, Keyun Xiang*, Xiaopeng Chen*, and Cheng Liu*

Abstract

To solve the problems of the low image contrast, fuzzy edge details and edge details missing in noisy image fusion, this study proposes a noisy infrared and visible light image fusion algorithm based on non-subsample contourlet transform (NSCT) and an improved bilateral filter, which uses NSCT to decompose an image into a low-frequency component and high-frequency component. High-frequency noise and edge information are mainly distributed in the high-frequency component, and the improved bilateral filtering method is used to process the high-frequency component of two images, filtering the noise of the images and calculating the image detail of the infrared image's high-frequency component. It can extract the edge details of the infrared image and visible image as much as possible by superimposing the high-frequency component of infrared image and visible image. At the same time, edge information is enhanced and the visual effect is clearer. For the fusion rule of low-frequency coefficient, the local area standard variance coefficient method is adopted. At last, we decompose the high- and low-frequency coefficient to obtain the fusion image according to the inverse transformation of NSCT. The fusion results show that the edge, contour, texture and other details are maintained and enhanced while the noise is filtered, and the fusion image with a clear edge is obtained. The algorithm could better filter noise and obtain clear fused images in noisy infrared and visible light image fusion.

Keywords

Bilateral Filter, Image Fusion, Local Area Standard Variance, Nonsubsample Contourlet Transform (NSCT)

1. Introduction

Wavelet transform is usually used in one-dimensional signal processing, and the two-dimensional wavelet transform cannot capture the multi-directional information of the image effectively. Do and Vetterli [1] proposed a Contourlet transform method that can conduct multi-dimensional and multi-directional decomposition in 2002. On the basis of contourlet transform, Da Cunha et al. [2] proposed the non-subsample contourlet transform (NSCT) theory in 2006. This theory not only has all possibilities of the contourlet transform, but also bears the characteristic of translation invariance. As the translation invariance can reduce the effect of registration error on the image fusion, the NSCT performance is suitable to process the image fusion. NSCT is characterized by multi-resolution, multi-direction and locality, which can effectively describe the high-dimensional singular feature of the image. At the same time, the decomposed image of by NSCT maintains the size of original image, so it is easy to find the

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received July 4, 2019; first revision March 23, 2020, second revision July 3, 2020; accepted July 11, 2020.

Corresponding Author: Xiaopeng Chen (18437927299@163.com)

* School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China (18609311366@163.com, 565878055@qq.com, 18437927299@163.com, 1290643579@qq.com)

corresponding relations between them, which is helpful for the subsequent image fusion processing.

At present, many researchers have proposed image fusion algorithms based on NSCT. Wang and Chen [3] proposed a synthetic aperture radar image and multi-spectral image fusion method; the pulse coupled neural network was used to fuse the coefficients and achieved a good fusion effect. Gomathi and Kalaavathi [4] proposed an algorithm of multi-channel medical image fusion based on NSCT, which uses the method of mean and variance to fuse the coefficients. Feng [5] decomposed the image on the basis of Tetrolet and formulated high- and low-frequency coefficient fusion rules with the ignition map of the pulse coupled neural network and the joint sparse coefficient active level map, which can fully retain the effective information of the source image. Deng et al. [6] decomposed the image with a non-subsampled dual-tree complex contourlet transform. The low-frequency fusion algorithm used an adaptive block method and the high-frequency fusion algorithm used the refined processing method of labeling images. This algorithm overcomes the block effect caused by block fusion. Yan and Xiang [7] proposed a new NSCT domain fusion method aimed at low infrared and visible light image contrast and the issue of incomplete preserved edges of the traditional fusion algorithm. The low-frequency coefficient adopts edge-based fusion rules. For the high-frequency fusion rule, the internal strength of the pulse-coupled neural network (PCNN) is adaptively adjusted according to the direction information for internal excitation, and the external excitation of PCNN is based on the improved spatial frequency characteristics; then, the high-frequency coefficients are fused using pulse ignition amplitude. Dai et al. [8] processed an adaptive medical image fusion method, as our eye to image contrast, edge, contour and texture information sensitivity is very high. So, the contrast between the fusion rules of PCNN energy, and regional Laplacian direction high-frequency coefficient fusion, the optimal weights were used as the objective function of adaptive weighted structural similarity selection of each sub-band coefficient, with the original information effectively based on the gray contrast enhanced image fusion, image fusion and better visual effect for the ordinary infrared and visual image fusion. Chen et al. [9] proposed an infrared and visible light image fusion method based on the wavelet transform and PCNN compensatory method in the NSCT domain. This method could effectively restrain the image distortion problem and bears better fusion precision and efficiency. Cai et al. [10] proposed an image fusion method combining the intuitionistic fuzzy set and region contrast together in the NSCT domain; this algorithm could enhance image contrast and preserve the edge and detail information in the original images. Zha and Guo [11] proposed an image fusion algorithm based on fast NSCT and the 4-directional sparse method. This fast algorithm could produce a 20-fold increase in efficiency. Wang and Cheng [12] proposed an image enhancement algorithm based on the MSSTO and NSCT methods. The original images were decomposed by the NSCT method, and the obtained low-frequency coefficients were further decomposed to obtain the brightness information and darkness information. The two kinds of information were processed to obtain better contrast and edges in the fused image. Zhang and Maldague [13] proposed a reconstruction algorithm based on the adaptive Gaussian fuzzy membership method, compressed sensing and total variation gradient descent method to fuse the infrared and visible light image. This method has better robustness and the fused image has better visual effect. Fu et al. [14] proposed an image-fused algorithm based on visual saliency and the translation invariance of NSCT. The results of this method reveal abundant visible light background with obvious infrared object. Wen et al. [15] proposed a NSCT fusion method according to the characteristics of the infrared and visible light image, and the fused image bears more detailed information and has a good effect. NSCT has a very strong sparsity, as high-frequency details can be extracted effectively. In the fusion process, we hope to preserve the details of edges and

contours as much as possible, so as to enrich the information of the fusion results. Chen et al. [16] decomposed infrared and visible light images through the Laplacian pyramid, and formulated fusion rules based on the pixel distribution information combined with the absolute value maximum criterion to achieve the goal of highlighting the salient target of the image and improving the fusion effect.

In general, details such as edges and contours in images are high frequency, but noise is also a high-frequency component. How to “edge-preserve” and “denoise” at the same time is rarely involved in the previous fusion process. To solve this problem, this paper uses bilateral filter to deal with the high-frequency coefficient. The two-side filtering (bilateral filtering) concept was first proposed by Tomasi and Manduchi [17], the gray value of each adjacent pixel or color information, not only taking into account the relationship between adjacent geometry, but also taking into account the similarity in brightness, through nonlinear combination of the two with an adaptive filter to obtain a smooth image. The image processed in this way can preserve the edge information well while filtering the noise.

The remaining sections of this paper are organized as follows: Sections 2 and 3 introduce the NSCT decomposition framework and the improved bilateral filter. Section 4 is the proposed image fusion algorithm based on the NSCT and improved bilateral filter. Section 5 is the experiment results and evaluation. Section 6 is the conclusion.

2. Non-subsampled Contourlet Transform

In image fusion, wavelet transform can capture a very limited number of directions [18]. The support interval of the wavelet basis is a small square, as shown in Fig. 1(a). The higher the resolution, the smaller the square, until it is close to the point singularity to approach the curve. Therefore, the wavelet transform has good surface singularity. It functions by using point singularities to approximate the curve.

The basic function of contourlet transform is processed using contour as the structure, the support interval as anisotropic, and its aspect ratio being adjusted arbitrarily. It is a “long bar” structure with variable size, as shown in Fig. 1(b).

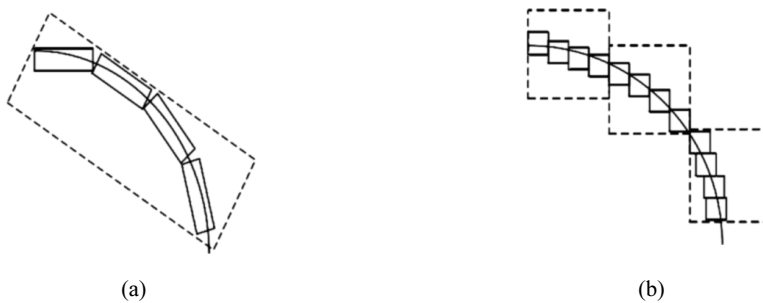


Fig. 1. (a) Wavelet base and (b) contourlet base.

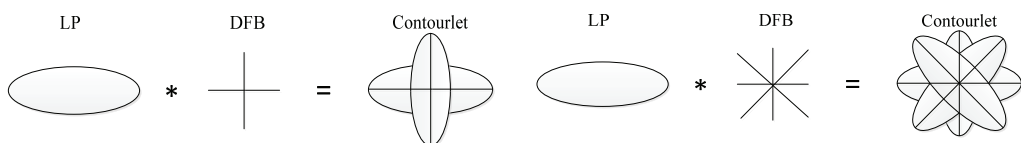


Fig. 2. Multidirectional transformation of contourlet.

Contourlet transform comprises two essential parts, one is Laplacian pyramid (LP) transform, to realize the multiscale geometric analysis function of contourlet transform [19]; the other is directional filter bank (DFB), to realize the direction analysis function of contourlet transform, as shown in Fig. 2. The specific process of contourlet transform—multiscale decomposition of image based on LP transform—mainly describes the singular points of the image. Then, with the high-frequency components of each level of LP decomposition of multi-direction decomposition by DFB, the purpose is to obtain the singular point in the same direction as the line. NSCT transformation also consists of two parts, which completes the multiscale decomposition process by the non-subsampled pyramid filter bank (NSPFB) and performs the multiple directional analysis process by the non-subsampled directional filter bank (NSDFB).

2.1 Non-subsampled Pyramid Decomposition Filter Bank

The multiscale decomposition of the NSCT transform is achieved by a set of dual channel NSPFB, which is shown in Fig. 3(a).

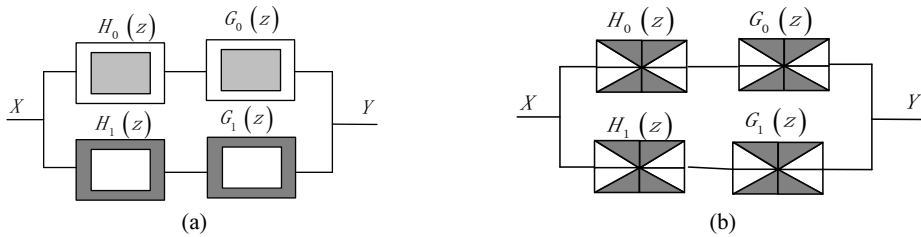


Fig. 3. NSCT two-channel: (a) NSPEB and (b) NSDFB.

In order to avoid the direct image sampling operations, NSCT performs translation invariance by performing upsampling in the decomposition filter and synthesis filter. In each layer of the NSPFB decomposition, the upper layer of the NSPFB decomposition is used by the sampling matrix for the decomposition of this layer to obtain a new filter. The low-frequency components decomposed from the previous layer are passed as input objects to the high-pass filter and low-pass filter to obtain the high-frequency components and low-frequency components of this layer. If the decomposition level is N , the original image is decomposed to $N+1$ sub-bands, and the size of each sub-band image is consistent with the original image. This completes the decomposition of the N layer of the image, or the N level multiscale decomposition. Fig. 4 is the decomposition process of three layers and the corresponding band allocation of the NSCT.

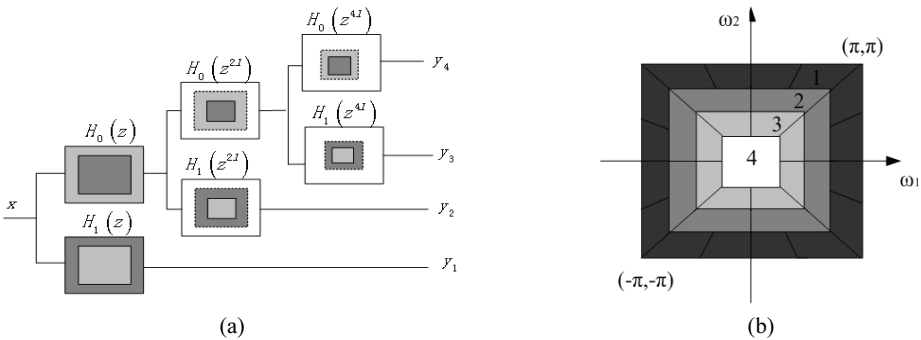


Fig. 4. (a) Three-layer NSPFB decomposition and (b) band allocation graph.

2.2 Non-subsampled Direction Filter Bank

Two-channel non-subsampled filter banks, the structure of which are flabellate, are used in directional decomposition of NSCT transform. Fig. 3(b) shows the decomposition filter.

Sampling is performed in the directional filter instead of the sampling operation on the image signal. In this way, it keeps the translation invariance. Each layer of high-frequency multiscale decomposition will be transmitted to the upsampling filter bank as the input object. When entering the next layer, first of all, the upper layer is decomposed using NSDFB to obtain the new NSDFB of the layer, and then filtering the high-frequency image. If the directional decomposition in a layer is 1, the high-frequency image of this layer is decomposed to 2^l directions. The directional decomposition of NSCT will not change the size of image. Fig. 5 shows the diagram of the four-direction decomposition.

As shown in Fig. 5, first-layer decomposition uses the horizontal component and the vertical component of flabellate filter $U_0(z)$ and $U_1(z)$ image separation; in the second-layer decomposition, first of all, to structure the new direction filter that this layer needs, upsampling is executed on the top of the flabellate filter $U_0(z)$ and $U_0(z)$ using quincunx matrix $Q = \begin{bmatrix} 1 & - \\ -1 & 1 \end{bmatrix}$. The directional decomposition of this layer is performed by a new directional filter, and the directional sub-band image of four channels is generated.

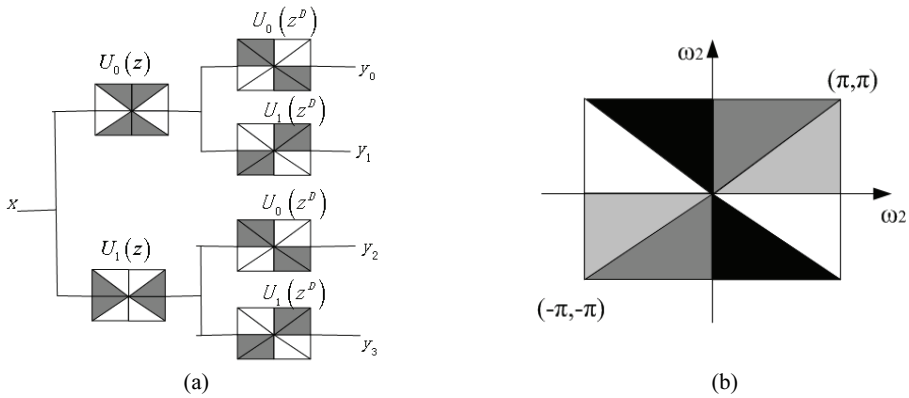


Fig. 5. Four-channel directional sub-band decomposition graph: (a) filter structure and (b) band allocation graph.

3. Bilateral Filter and its Improvement

3.1 Bilateral Filter

The bilateral filter is a kind of nonlinear filter [20], which is divided into two parts: spatial filter and range filter. The distance factor and gray difference between pixels are comprehensively considered.

The bilateral filter is defined as follows:

$$H_a = \frac{\sum_{b \in S} K(\|a - b\|, \sigma_s) K(|I_a - I_b|, \sigma_r) I_a}{\sum_{b \in S} K(\|a - b\|, \sigma_s) K(|I_a - I_b|, \sigma_r)} \quad (1)$$

where I_a is the gray value of the image I on the central pixel point $a(x, y)$; after passing through the two-side filter, the pixel of this point changes into H_a , the neighborhood pixel of $a(x, y)$ is $b(x, y)$. I_b is the gray value of neighborhood pixel $b(x, y)$. The neighborhood pixel point set S is made up by all of $b(x, y)$. K presents the filtered kernel function, which is in the form of Gauss functions:

$$K(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (2)$$

In Eq. (1), $K(\|a - b\|, \sigma_s)$ represents the spatial proximity factor, $K(|I_a - I_b|, \sigma_r)$ represents the gray similarity factor; the definition is

$$K(\|a - b\|, \sigma_s) = e^{-\frac{(x-u)^2 + (y-v)^2}{2\sigma_s^2}} \quad (3)$$

$$K(|I_a - I_b|, \sigma_r) = e^{-\frac{(I_a - I_b)^2}{2\sigma_r^2}} \quad (4)$$

where σ_s represents the distance standard deviation and σ_r represents the gray standard deviation. The two parameters represent the features of size and contrast of the image, which play a decisive role in the smoothness of the image; the two parameters also represent the filter width. By adjusting the width of σ_s and σ_r , we can make a balance between the image feature and the noise burst. The bigger the value of σ_s , the bigger the influence of pixel distance on the gray value after central point filtering. The bigger the value of σ_r , the bigger the gray value difference and the bigger the influence of the gray value of the central point.

Fig. 6 describes the characteristics of the bilateral filter in the vicinity of the edge. It can be seen that it has excellent edge protection abilities while denoising.

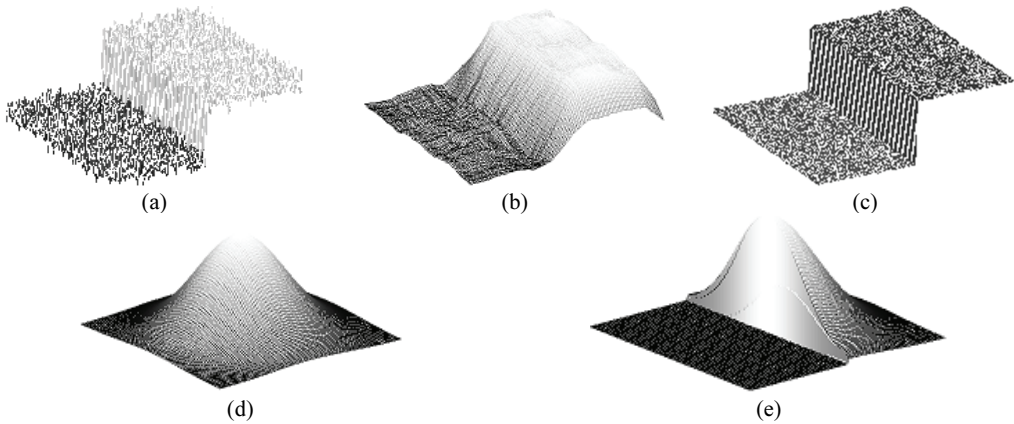


Fig. 6. Effect of the bilateral filter on image edge: (a) image region with noises, (b) effect of the domain filter, (c) effect of bilateral filter, (d) spatial domain filter function, and (e) composite function of spatial and value domain filter.

3.2 Improved Bilateral Filter

The traditional bilateral filter uses a Gaussian function as the kernel function. In the calculation process, there are many parameters, so calculation is complicated. Therefore, the bilateral filter is improved by

replacing the Gaussian spatial filtering with the mean space filter and Gaussian range filtering with the bilateral exponential range filter.

$$H_a = \frac{\sum_{b \in S} K(\|a-b\|)K(r)I_a}{\sum_{b \in S} K(\|a-b\|)K(r)} \quad (5)$$

Here, r represents the gray value difference of $a(x, y)$ and $b(x, y)$, shown as follows:

$$r = \begin{cases} I_a - I_b & I_a - I_b \neq 0 \\ 1 & I_a - I_b = 0 \end{cases} \quad (6)$$

$K(\|a-b\|)$ represents the spatial proximity factor, shown as follows:

$$K(\|a-b\|) = \begin{cases} \frac{1}{(\|a-b\|)^2} & \|a-b\| \neq 0 \\ 1 & \|a-b\| = 0 \end{cases} \quad (7)$$

$$K(r) = \frac{1}{r} \quad (8)$$

In this way, the calculation amount of local window is reduced from the original N^2 to $1+N$, which greatly accelerates the filtering process. $K(t)$ is the improved filter kernel function. The role of neighborhood pixels in filtering is decided by $K(\|a-b\|)$ and $K(r)$.

For the improved filter kernel function $K(t)$, $K(r)$ increases with the gray value similarity of $a(x, y)$ and $b(x, y)$ increasing. $K(r)$ decreases with the gray similarity decreasing; especially when $I_a = I_b$ and $K(r) = 1$. At this time, H_a is equivalent to a pure spatial filter. The longer the distance between $a(x, y)$ and $b(x, y)$, the smaller the distance similarity and $K(\|a-b\|)$ is; as the distance between $a(x, y)$ and $b(x, y)$ decreases, the value of $K(\|a-b\|)$ increases. Especially, $K(\|a-b\|) = 1$ when $a(x, y) = b(x, y)$. By comparison, we can see that the improved filter kernel function can achieve the same effect as the original Gaussian filter kernel function and reduce the computational complexity.

4. Image Fusion Method

NSCT has the characteristics of orientation, anisotropy, translation invariance and so on. The bilateral filter has a great advantage in image line feature detection and can better observe edge details while filtering. In this method, after NSCT decomposition of original images, an improved bilateral filter is adopted to filter the high-frequency (HF) coefficient, and linear information such as edges and contours are extracted to obtain the HF coefficient after fusion. For the fusion rules of low-frequency (LF) components, the local standard deviation coefficient method is used.

The flowchart of this method is shown in Fig. 7, and the steps are shown as follows:

- Step 1: Decompose the infrared image A and visible image B to obtain their LF coefficients $C_{R,A}^L(p, q)$ and $C_{R,B}^L(p, q)$ and their HF coefficients $E_{R,A}^H(p, q)$ and $E_{R,B}^H(p, q)$, respectively.
- Step 2: The LF coefficients $C_{R,A}^L(p, q)$ and $C_{R,B}^L(p, q)$ are fused according to the rule of local area standard deviation; the output is $C_R^F(p, q)$.

Step 3: The HF coefficients $E_{R,A}^H(p, q)$ and $E_{R,B}^H(p, q)$ are fused by improved bilateral filter, obtaining the fusion HF coefficients E_R^F .

Step 4: The fused coefficients C_R^F and E_R^F are inversely transformed by NSCT method into the fused image.

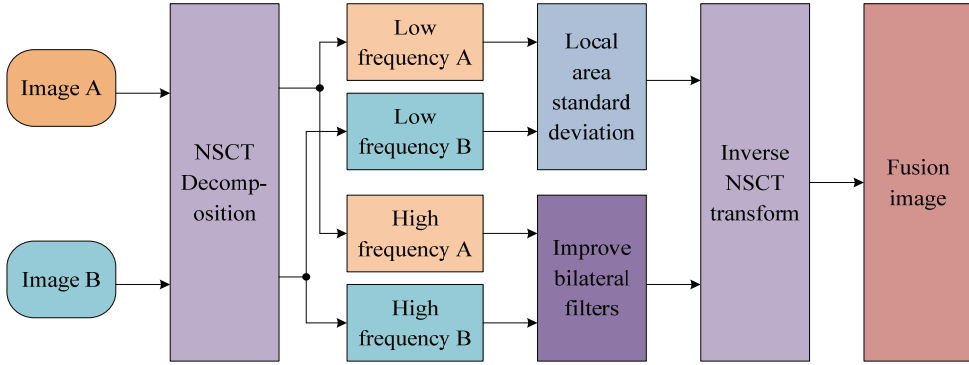


Fig. 7. Image fusion graph.

4.1 LF Coefficients Fusion

The LF and HF coefficients are the output of infrared original image and visible light original image after NSCT multiscale and multi-direction decomposition. The LF coefficients represent the gray smooth area, contain most of the energy of an image and usually represent the background information. The HF coefficients usually represent the HF information of the image, such as edge, contour and noise, and can reflect the details. To obtain the fusion image, it must first obtain the HF and LF coefficients of the fused image; thus, appropriate fusion rules must be adopted according to the characteristics of the coefficients.

As a parameter of measuring the image clarity degree, the local area standard deviation can be expressed by the gray variation intensity in the image area. The region with obvious gray variation is also the region with concentrated image features, so it can effectively extract the features of the image with the help of the local standard deviation method. Therefore, the local area standard variance method is used as the fusion rule in the fusion of LF coefficients. The steps are as follows:

Firstly, calculate the local area standard deviation $\sigma^A(p, q)$ and $\sigma^B(p, q)$ of LF coefficients $C_{R,A}^L(p, q)$ and $C_{R,B}^L(p, q)$ with the size of neighborhood $H_1 \times W_1$ as 3×3 or 5×5 ; the formula is shown as follows:

$$\sigma^A = \sqrt{\frac{\sum_{i=-(H_1-1)/2}^{(H_1-1)/2} \sum_{j=-(W_1-1)/2}^{(W_1-1)/2} [C_{R,A}^L(p+i, q+j) - \overline{C_{R,A}^L}(p, q)]^2}{H_1 \times W_1}} \quad (9)$$

$$\sigma^B = \sqrt{\frac{\sum_{i=-(H_1-1)/2}^{(H_1-1)/2} \sum_{j=-(W_1-1)/2}^{(W_1-1)/2} [C_{R,B}^L(p+i, q+j) - \overline{C_{R,B}^L}(p, q)]^2}{H_1 \times W_1}} \quad (10)$$

Then, select LF coefficients after fusion based on Eq. (11). The fusion rule is used to compare the threshold value and difference between the standard deviation of the local region of the original images. If the former is large, the coefficient of the image will be larger; if the latter is large, the average value of

the coefficients of the two images will be taken. So, the threshold th is important, and we usually select it by experience; the range of th value is $[0.1, 0.3]$ [21].

$$C_R^F(p, q) = \begin{cases} C_{R,A}^L(p, q) & \sigma_A - \sigma_B > th \\ [C_{R,A}^L(p, q) + C_{R,B}^L(p, q)] / 2 & |\sigma_A - \sigma_B| < th \\ C_{R,B}^L(p, q) & \sigma_A - \sigma_B < -th \end{cases} \quad (11)$$

According to the above processes, we obtain the fusion LF coefficient $C_F^{m,p}$.

4.2 HF Coefficients Fusion

Fusing the infrared and visible light image HF coefficient by the improved bilateral filter method, the main idea is to compare the gray value difference between the two pixels in the same location in the HF components of source image HF components and use the larger difference as the similarity indicator. This is equivalent to using a subimage with clear edges as a gray similarity indicator, so that the fused HF components contains more edge, contour and texture details. At the same time, to further enhance the effect of the fused image, the detailed image of the HF component of the infrared image extracted by the bilateral filter is added to the HF component of the visible image with a certain weight. The reason why the detail image is superimposed on the HF component of the visible light image is due to the high resolution of the visible light image and the richer information such as edges, contours and textures in the scene. The detailed image extracted from the HF component of the infrared and visible light image is further superimposed on the HF component of the visible light image, which can obtain richer details of edges, contours and textures, and can achieve the effect of image enhancement. The steps are shown as follows: Step 1: Calculate the gray similarity value, using the larger difference between of the infrared HF components and the visible light HF similarity on $a(x, y)$ and $b(x, y)$ as the similarity indicator, the equation is as follows:

$$r(a, b) = \begin{cases} I_{Aa} - I_{Ab} & I_{Aa} - I_{Ab} \geq I_{Ba} - I_{Bb} \\ I_{Ba} - I_{Bb} & other \end{cases} \quad (12)$$

where r_A represents the gray value difference of the infrared image HF components on $a(x, y)$ and $b(x, y)$, and r_B represents the visible light image HF components on $a(x, y)$ and $b(x, y)$.

Step 2: According to Eq. (5), we use the improved bilateral filter to filter the HF coefficient $E_{R,A}^H$ of the infrared image and the HF coefficient $E_{R,B}^H$ of the visible light, and obtain the HF components $F_{R,A}^H$ and $F_{R,B}^H$.

Step 3: Calculate the detailed image of the infrared image HF components $E_{R,A}^{D,H}$; the detail is shown as follows:

$$E_{R,A}^{D,H} = E_{R,A}^H - F_{R,A}^H \quad (13)$$

Step 4: Multiply a weight with the detailed image $E_{R,A}^{D,H}$ and add it to the HF component of a visible image that has been bilateral filtered, we then obtain the HF coefficient after fusion:

$$E_R^H = F_{R,B}^H + \lambda E_{R,A}^{D,H} \quad (14)$$

where λ is the test parameter; here, it is 1.5. The detailed image obtained by bilateral filter contains abundant image details such as edge and texture; if it is properly enhanced and then overlapped with the HF component of the visible light image, it can effectively improve the clarity of the fusion image, so the value of λ should be greater than 1. However, if the value of λ is too large, it will cause partial blur of the fused image and, consequently, affect its visual effect. After repeated experiments, here, we take λ as 1.5.

4.3 Image Reconstruction

The LF and HF coefficients after NSCT decomposition are processed by fusion rules, then obtain the fused LF coefficient C_R^F and the fused HF coefficient E_R^F , NSCT inverse transform is performed on them, obtaining the fused image.

5. Experimental and Discussion

5.1 Data Sources

To verify the algorithm, this study uses three groups of original images, which are the infrared and visible images that have been strictly registered. The fusion results of this algorithm are compared with other five methods to verify the effectiveness of this algorithm. These five methods are discrete wavelet transform (DWT) fusion method, dual-tree complex wavelet transform (DTCWT), LP, contourlet transform method, and the NSCT fusion method.

5.2 Objective Evaluation Criteria

For the objective evaluation criteria, we use the standards of entropy (EN), mutual information (MI), average gradient (AvG), correlation coefficient (CC), and SSIM. Entropy mainly reflects the amount of information contained in the fused image. Mutual information reflects how much information the fused image obtains from the original image. The average gradient reflects the ability of the image to express the contrast of small details, indicating the sharpness of the image. SSIM is different to other indicators; the general objective evaluation indicators are based on the error sensitivity to measure image quality without considering the image correlation. CC is used to measure the linear correlation degree between the fused image and the original image. The larger the CC, the better the fused image. SSIM is calculated by comprehensively considering the image brightness, contrast and structure. It is a kind of evaluation standard that conforms to the human real perception, the value range is $[0, 1]$. Figs. 8–10 show the original images and the fused images of these methods, Tables 1–3 show the objective evaluation values of the fused image by these methods.

5.3 Experimental Results Discussion

Fig. 8 shows the storing base of UN. In the scene, a man is walking outside the fence; the fence surrounds important military supplies. Fig. 8(a) is the scene of the infrared image; the person is clear but the background is extremely vague. Fig. 8(b) is the visible image of the same scene, not the target, but the background is relatively clear; buildings, trees, roads, fences and other details are clear. Fig. 8(c)–(h)

are the results corresponding to the six methods. In the fusion result of DWT and contourlet transform methods, the overall clarity is not high; the point features are relatively clear; and the edge, contour and texture are fuzzy. Much of the information is lost through artifacts and blocking effects. In the fusion results of DTCWT and LP methods, the images are clearer than the results of DWT method, and the outline of the target is clear; however, they lose most edge details of the background and display the Gibbs phenomenon. In the fusion result of NSCT, it is clearer and eliminates the Gibbs phenomenon because of its translation invariance, and the overall details produced the effect shown in Fig. 8(h). In Fig. 8(h), the fusion result of this algorithm is the clearest image of the six results. The image is of high gray contrast and its edge, contour texture and lines are rich and clear.

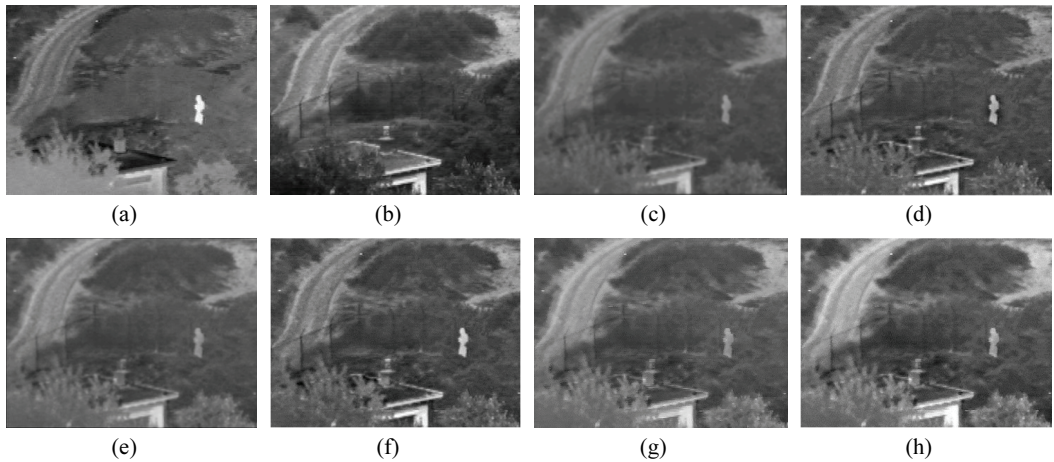


Fig. 8. Camp fusion results: (a) infrared image-Camp, (b) visible light image-Camp, (c) DWT, (d) DTCWT, (e) contourlet, (f) LP, (g) NSCT, and (h) this algorithm.

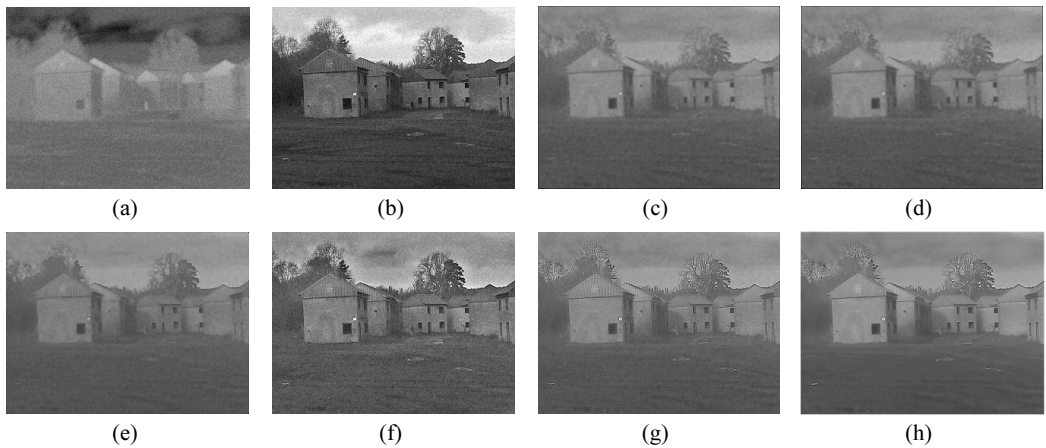


Fig. 9. Smoke fusion results (variance 0.002): (a) infrared image-Smoke, (b) visible light image-Smoke, (c) DWT, (d) DTCWT, (e) contourlet, (f) LP, (g) NSCT, and (h) this algorithm.

Fig. 9 shows the image of the building, with the sky and trees as the background and the buildings in the foreground, with people moving in the middle around the buildings. Fig. 9(a) is an infrared image. The figure is clear, but the background only shows a rough outline. Fig. 9(b) is a visible image. The sky

and trees in the scene are clear and the buildings show clear outline details, but the targets are not easy to recognize. To verify the filtering performance of this algorithm for noisy images, this group of experiments adds Gaussian white noise to the standard images; its mean value is 0 and variance is 0.001. Fig. 9(c)–(h) are the fused images obtained by the six methods. DWT and contourlet method fusion results lose most details of line, edge, contour and noise. In the fusion results of the DTCWT and LP methods, there are many details missing in the edges of buildings and the contours of human targets are clear with noise. The fusion result of NSCT method can capture most of the contours and edges in the original visible image, but it is not clear enough and the noise is obvious. In Fig. 9(h), the fusion result of this algorithm adopts bilateral filter for line detection and denoising of the image, so the edges and contours in the results are very clear, the effect is best and noise pollution is effectively reduced.

To further verify the filtering effect, we add the *Street* images with Gaussian white noise with mean of zero and variance of 0.002. The Gaussian white noise in Fig. 10(a) and (b) are obvious. The fusion result of this algorithm is weakly affected by the interference of the noise. The fusion results of DTCWT and LP are highly affected by the noise, as its sensitivity with the point singularity, edges and contours are almost destroyed in the results. The results of DWT, contourlet and NSCT transform are also affected by the noise.

It can be seen from the fusion results of the three groups of images that the proposed algorithm has higher fusion image clarity. Not only are the prominent targets in the scene more prominent, but also information such as edge details and textures are preserved, which is more in line with the observation habits of human vision and improves the recognition of the scene.

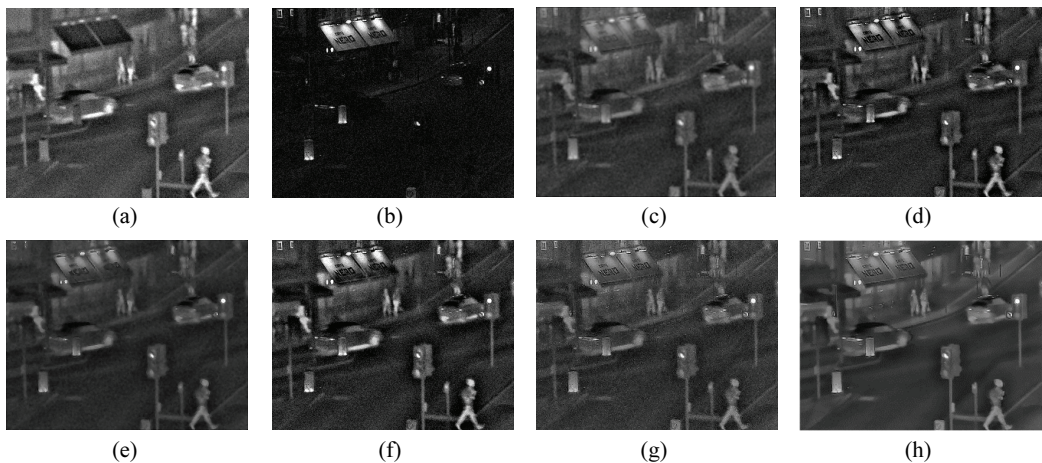


Fig. 10. Street fusion results (variance 0.002): (a) infrared image-Street, (b) visible light image-Street, (c) DWT, (d) DTCWT, (e) contourlet, (f) LP, (g) NSCT, and (h) this algorithm.

The quantitative results are shown in Tables 1–3. According to the comprehensive comparison of objective indicators, the method in this study is better than the other three methods on MI, AvG, CC, and SSIM. For the processing of images with noise, this algorithm has a strong ability for suppressing noise. Filtering the noise actually reduces the information of the image, so the result of this algorithm has the lowest EN, which does not mean that fusion quality is bad for the noise is filtered and the useful information is not reduced. So, image quality is guaranteed and the image becomes clearer. Therefore, the algorithm in this paper is the best of the four methods.

Table 1. Quantitative comparison of Fig. 8

Method	MI	EN	AvG	CC	SSIM
Contourlet	0.6358	6.5222	4.7309	0.9549	0.7314
DTCWT	0.6734	6.4845	7.3444	0.9624	0.7760
DWT	0.6839	6.2469	5.2610	0.9605	0.8027
LP	0.6753	6.6548	7.9342	0.9609	0.7596
NSCT	0.6538	6.7637	11.1187	0.9455	0.6364
This algorithm	0.7474	2.6866	19.9408	0.9666	0.8028

Table 2. Quantitative comparison of Fig. 9

Method	MI	EN	AvG	CC	SSIM
Contourlet	0.7372	2.5142	29.0072	0.5011	0.6733
DTCWT	1.1382	6.8605	22.9270	0.9508	0.7955
DWT	0.7024	2.4244	21.7486	0.5059	0.6195
LP	1.0552	6.9962	24.6681	0.9563	0.7537
NSCT	0.7099	2.5963	24.0375	0.5137	0.6176
This algorithm	1.1712	2.4075	28.9951	0.9816	0.7354

Table 3. Quantitative comparison of Fig. 10

Method	MI	EN	AvG	CC	SSIM
Contourlet	0.1236	2.3959	16.2314	0.6410	0.1748
DTCWT	0.2811	6.6887	24.7739	0.6694	0.3426
DWT	0.1136	2.0837	26.5228	0.5173	0.1259
LP	0.2855	6.8617	26.9704	0.6389	0.3062
NSCT	0.2712	2.5714	31.0152	0.4246	0.3569
This algorithm	0.3391	2.4474	33.9296	0.6987	0.4349

As can be seen from Tables 1–3, compared to the comparison algorithm, this algorithm has different degrees of leading in five evaluation indexes, which shows that the fusion image obtained by this algorithm can effectively retain the effective information of the source image, and the fusion image and source image have high structural similarity and linear correlation, consistent with the subjective evaluation results. In summary, this algorithm combines NSCT and bilateral filtering methods for processing, while maintaining and enhancing linear details (such as edges, contours, textures and other details); it has good noise filtering ability, and can obtain a fusion image with clear edges, improve the comprehensibility of the scene and meet the observation needs of the human eye on the scene.

6. Conclusion

With the new ideas of using the bilateral filter in the NSCT domain, this study proposes a new infrared and visible light fusion method. NSCT features such as multiresolution, local localization and multi-directionality can make it more effective for capturing the high-dimensional directional parity of images. The image size after NSCT decomposition is the same as that of the original image, which can easily find the corresponding relationship with each other and be better for the subsequent fusion processing. The

characteristics of NSCT result in its innate advantages in image fusion. In addition, to solve the contradiction between denoising and edge preservation in image filtering and avoid the problem of blurred edges while filtering, this study introduced the bilateral filter to denoise the HF components of the original images and calculate the detailed image of the infrared image HF component. By superimposing the detail image and the visible image, we can extract as much detail information as possible from the fusion image, such as the edge, contour and texture of the source image. At the same time, the edge information is also enhanced, which makes the image fusion, edge, contour and texture information clearer. The fusion results show that this algorithm has a good ability to filter the noise and extract the edge of the image.

Acknowledgement

The work has been supported by National Natural Science Foundation of China (Nos. 61861025, 61562057, 51969011); Ministry of Education (No. KFKT2018-9); Longyuan Youth Innovative and Entrepreneurial Talents (Team) Project in 2021; Gansu Provincial Department of Education: Young Doctor Fund Project (No. 2021QB-049); College Students Employment and Entrepreneurship Ability Improvement Project of Gansu Province (No. 2021-C-123); Intelligent Tunnel Supervision Robot Research Project (China Railway Research Institute [Research Institute], No. 2020-KJ016-Z016-A2); Youth Foundation of Lanzhou Jiaotong University (No. 2015005); Scientific Research Project of Higher Education Institutions of Gansu Province (No. 2016A-018); and Open project of Gansu Provincial Research Center for Conservation of Dunhuang Cultural Heritage (No. GDW2021YB15).

References

- [1] M. N. Do and M. Vetterli, "Contourlets: a new directional multiresolution image representation," in *Conference Record of the 36th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2002, pp. 497-501.
- [2] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsamped contourlet transform: theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089-3101, 2006.
- [3] X. L. Wang and C. X. Chen, "Image fusion for synthetic aperture radar and multispectral images based on sub-band-modulated non-subsampled contourlet transform and pulse coupled neural network methods," *The Imaging Science Journal*, vol. 64, no. 2, pp. 87-93, 2016.
- [4] P. S. Gomathi and B. Kalaavathi, "Multimodal medical image fusion in non-subsampled contourlet transform domain," *Circuits and Systems*, vol. 7, no. 8, pp. 1598-1610, 2016.
- [5] X. Feng, "Fusion of infrared and visible images based on Tetrolet framework," *Acta Photonica Sinica*, vol. 48, no. 2, pp. 76-84, 2019.
- [6] H. Deng, C. Wang, Y. Hu, and Y. Zhang, "Fusion of infrared and visible images based on non-subsampled dualtree complex contourlet and adaptive block," *Acta Photonica Sinica*, vol. 48, no. 7, pp. 136-146, 2019.
- [7] L. Yan and T. Z. Xiang, "Fusion of infrared and visible images based on edge feature and adaptive PCNN in NSCT domain," *Acta Electronica Sinica*, vol. 44, no. 4, pp. 761-766, 2019.
- [8] W. Z. Dai, X. L. Jiang, J. F. Li, "Adaptive medical image fusion based on human visual features," *Acta Electronica Sinica*, vol. 44, no. 8, pp. 1932-1939, 2016.

- [9] Z. Chen, X. Yang, and C. Zhang, "Infrared and visible image fusion based on the compensation mechanism in NSCT domain," *Chinese Journal of Scientific Instrument*, vol. 37, no. 4, pp. 860-870, 2016.
- [10] H. Y. Cai, L. R. Zhuo, P. Zhu, Z. H. Huang, and X. Y. Wu, "Fusion of infrared and visible images based on non-subsampled contourlet transform and intuitionistic fuzzy set," *Acta Photonica Sinica*, vol. 47, no. 6, pp. 225-234, 2018.
- [11] C. H. Zha and Y. T. Guo, "Fast image fusion algorithm based on sparse representation and non-subsampled contourlet transform," *Journal of Electronics & Information Technology*, vol. 38, no. 7, pp. 1773-1780, 2016.
- [12] F. Wang and Y. M. Cheng, "Visible and infrared image enhanced fusion based on MSSTO and NSCT transform," *Control and Decision*, vol. 32, no. 2, pp. 269-274, 2017.
- [13] Q. Zhang and X. Maldague, "An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing," *Infrared Physics & Technology*, vol. 74, pp. 11-20, 2016.
- [14] Z. Fu, X. Wang, X. Li, and J. Xu, "Infrared and visible image fusion based on visual saliency and NSCT," *Journal of University of Electronic Science and Technology of China*, vol. 46, no. 2, pp. 357-362, 2017.
- [15] G. Wen, J. Pengchong, and Z. Tianchen, "Infrared image and visual image fusion algorithm based on NSCT and improved weight average," in *Proceedings of 2015 6th International Conference on Intelligent Systems Design and Engineering Applications (ISDEA)*, Guiyang, China, 2015, pp. 456-459.
- [16] J. Chen, X. Li, L. Luo, X. Mei, and J. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Information Sciences*, vol. 508, pp. 64-78, 2020.
- [17] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the 6th International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, Bombay, India, 1998, pp. 839-846.
- [18] E. Lallier and M. Farooq, "A real time pixel-level based image fusion via adaptive weight averaging," in *Proceedings of the 3rd International Conference on Information Fusion*, Paris, France, 2000.
- [19] C. Q. Ye, "Research on multi-sensor image fusion algorithm based on multiscale decomposition," Xidian University, Xi'an, China, 2009.
- [20] H. G. Jia, Z. P. Wu, M. C. Zhu, M. Xuan, and H. Liu, "Infrared image enhancement based on generalized linear operation and bilateral filter," *Optics and Precision Engineering*, vol. 21, no. 12, pp. 3272-3282, 2013.
- [21] Y. Shen, J. W. Dang, Y. P. Wang, X. Feng, and W. W. Luo, "A novel medical image fusion method based on the multi-scale geometric analysis tool," *Journal of Optoelectronics-Laser*, vol. 24, no. 12, pp. 2446-2451, 2013.



Yu Shen <https://orcid.org/0000-0001-5396-9657>

She received M.S. and Ph.D. degrees in School of Electronic and Information Engineering from Lanzhou Jiaotong University in 2008 and 2017, respectively. Her current research interests includes image processing and image fusion.



Keyun Xiang <https://orcid.org/0000-0003-0106-659X>

He received B.S. degree in School of Electronic and Information Engineering from Lanzhou Jiaotong University in 2020. His current research interests includes image and video processing.



Xiaopeng Chen <https://orcid.org/0000-0002-4249-0381>

He received M.S. degree in School of Electronic and Information Engineering from Lanzhou Jiaotong University in 2021. His current research interests includes image and video processing.



Cheng Liu <https://orcid.org/0000-0001-7424-7557>

He received M.S. degree in School of Electronic and Information Engineering from Lanzhou Jiaotong University in 2019. His current research interests includes image and video processing.