Journal of Computers Vol. 35 No. 5, October 2024, pp. 15-33
doi: 10.53106/199115992024103505002

Multimodal Image Fusion Algorithm Based on Two-stage Multiscale Decomposition

Lei Zhang[1*], Yi-Yuan Cheng[1], Jing Li[1], Zhi-She Wang[2], Lin-Na Ji[3], and Wei Liu[1]

[1] Nan Yang Normal University, Nan Yang, 473061, China
[2] Taiyuan University of Science and Technology, Taiyuan 030051, China
[3] North University of China, Taiyuan 030051, China

zhanglei000223@163.com

Received 19 September 2023; Revised 22 September 2023; Accepted 22 September 2023

Abstract. Feature-intensity differences exist between different modal images; hence, complementary features are easily ignored or drowned in most multimodal image fusion algorithms. In this study, a novel two-stage fusion algorithm is proposed to reduce the loss of complementary features. The fusion algorithm is divided into two stages: the first stage adopts the multiscale transform based on the hybrid $\ell_0\ell_1$ layer decomposition and Gauss filter to decompose the source images into the structure-, large-scale-, and detail-layer images. Further, the previous fusion images were generated using a linear combination of one kind of modal image and the different feature layer images of the other type of modal image; this eliminated the differences between the images of the different sensors. The second stage enhances feature fusion and improves the fusion effect. The previous fusion images are decomposed by using non-down-sampling shear wave transform (NSST), and the low-frequency fusion images are respectively fused using principal component analysis (PCA) and local engineering texture; subsequently, the different fusion images are integrated by using local contrasting weighting to obtain the final fusion image of the low-frequency images. High-frequency images are respectively fused by using the local standard deviation matching degree of the exponential function and the local modified spatial frequency weighted average and then the different fusion images are integrated by using a weighted average to obtain the final high-frequency fusion images. Finally, the final fusion image is obtained by the inversion of the NSST. Results show that the fusion images quality have considerably improved and is sufficiently clear.

Keywords: multimodal images, two-stage fusion, multiscale decomposition

1 Introduction

The sensors of the imaging mechanism can capture different features of the detected scene. The visible image may have a favorable visual effect and rich, detailed information under suitable weather conditions. The infrared polarization intensity images have rich detail and local contrast features. The infrared images have a strong contrast between the target and background, and infrared imaging works in all weather conditions, but the infrared images lack detailed features of the scene [1]. Computed tomography (CT) images describe the hard tissue feature, primarily the structure feature and have a high spatial resolution; magnetic resonance imaging (MRI) describes the soft substance feature and describes the detail feature. A single modal image cannot fully describe the full features of the target. Image fusion can fuse the complementary features of the different modal images into a single image, and the fused image can completely describe the feature of the object. Thus, multimodal image fusion has important applications in state situation awareness, target detection, driverless driving, and accurate diagnosis.

Image fusion is classified into spatial domain fusion algorithms, transformation domain fusion algorithms, sparse representation (SP), deep learning fusion algorithms, and hybrid fusion algorithms. Spatial domain fusion algorithms, such as weight averaging and principal component analysis (PCA), are simple and rapid. Multimodal images have been integrated using PCA previously [2], indicating that the fusion image can preserve the main features of different modal medical images and operate effectively. However, in the spatial fusion, the large pixel values and the regions with large local energy occupy large fusion weights, such as the strong contrasting region. Fusion algorithms based on the spatial domain can easily lose both contrast and texture features [3, 4].

* Corresponding Author

Fusion algorithms based on the SP mainly extract the detail feature and fine texture feature [5]. This method [6] adopts joint sparse representation to fuse the medical images, and the reference [7] uses directional dictionaries to fuse the different frequency sub-band images. F. Fatemeh et al. [8] trained the dictionary basis through offline learning methods. These fusion methods based on the SP afford a suitable effect. However, during the fusion processing, the structure features and local contrast features are easily lost. The key to the SP fusion depends on the training sample and the construction approach. Thus, dictionaries do not always meet the requirements of multiple scenario fusion. Liu [9] analyzed different fusion algorithms based on SP, and the dictionary is the key of obtaining a suitable fusion effect.

Fusion algorithms based on the transform domain are mainly multiscale transforms. The source images are decomposed into low-frequency sub-band images and high-frequency sub-band images, and the different frequency sub-band images are fused by using the different fusion rules. The advantage of the multiscale transform is that the different types of features in images are separated. The multiscale transform fusion algorithms are mainly classified based on the wavelet transformation and the multiscale edge-preserving filter. The multiscale fusion algorithms based on the wavelet mainly include the discrete wave transform (DWT), the dual-tree complex wave (DTCWT), the nonsubsampled contourlet wave transform (NSCT), and NSST et al. [10-15]. These multiscale transforms can extract the high-frequency features suitably; however, the low-frequency features, including the contour and edge features are not extracted efficiently. The edge-preserving filters mainly include the guider filter, bilateral filter, and weighted least square optimization. These edge-preserving filters can smooth the image and retain the edge and contour features. Therefore, the combination of the edge-preserving filter and multiscale decomposition can extract the structure feature and high-frequency feature efficiently. The fusion methods based on the edge-preserving filter are mainly focused on [16-18]. However, the intensity and contrast features of the different modal images have large differences, and the transformation domain algorithms usually adopt the same transformation method. The effect of feature extraction has considerable differences for different mode images in the transform with the same transformation method, and because the feature extraction effect determines the fusion effect, guaranteeing the same transform function for different modal images is difficult. Thus, fusion images are prone to the loss of complementary features between different modal images.

The advantage of deep learning is the extraction of different features from the image. Thus, deep learning can considerably improve target classification and recognition, which is currently a research hot spot. Due to the powerful ability of deep learning for image feature extraction, it is introduced into image fusion. Deep learning is applied to image fusion in two main ways: first, the features of the image are extracted by using a neural network, and subsequently, the feature images are mapped into the fusion weight map to achieve the final fused image. A conventional neural network is adopted to extract the features of the source images [19, 20], and the feature images are used to get fusion weights. The generative adversarial network (GAN) can automatically generate pictures by training the generator and the discriminator; therefore, the GNA is often used in other fusion methods based on deep learning. However, fusion methods based on deep learning require numerous data to ensure the stability of the network; the size of the training set directly determines the reliability of the network and the effect of the fusion. Till date, no study has shown that a stable and adaptable network requires a large amount of data. Although the fusion algorithms based on the GAN do not need the fusion rule, they also require big data, while the input image of the discriminator should be the real images. Nonetheless, no standard fusion image exists.

Different transform methods and feature extraction methods have different advantages, and the hybridization of different methods with fusion can considerably improve the quality of fusion relative to traditional fusion algorithms. Therefore, hybrid fusion algorithms have received increasing attention [21]. Hybrid fusion methods are mainly included: First, the multiscale transform methods are combined with the other methods, and the fusion methods based on these combinations first decompose the source images by using the multiscale transform, and then the other methods such as PCNN, deep learning, and edge-preserving filters are adopted to process the low-frequency sub-band image or the high-frequency sub-band image. The hybrid of the multiscale transform and SP and the hybrid of the NSST and CNN have been explored in previous studies [22-25]. These fusion methods can better separate the images of different features, thereby reducing feature aliasing and information loss in the process of fusion and improving the quality of the fusion image over a single transformation. However, in fusion processing, the use of a single filter to extract features remains unchanged. Second, different multiscale transformations or different filters are combined, such as the hybrid of NSCT and NSST, LSWT and NSST, guider filter and Gauss filter (roll guider filter), or multiscale transform and guider filter [26-29]. These fusion methods can decompose image features into large-scale, medium scale, and fine-scale features and the different types of feature images adopt different fusion rules that can considerably improve the effect of fusion. Hybrid multiscale decomposition can comprehensively utilize the advantages of different multiscale transformation filters to effectively separate and fuse inter-image features. Moreover, it can reduce the integration loss of weak features.

In summary, fusion methods have achieved good fusion results. Because they directly extract the features of the image to obtain a fusion decision map and then fuse the source images, they ignore the differences among multimodal images, causing large intensity differences among the extracted features. Moreover, the features of multimodal images are easily lost during the fusion process, which is detrimental to subsequent identification, detection, and diagnosis. To improve the effect of fusion, this study attempts to reduce feature loss and the differences among multimodal images.

Based on the aforementioned analysis, multiscale decomposition can accurately extract different types of features from an image and subsequently fuse the images. Therefore, multiscale decomposition is most frequently adopted in fusion algorithms and also forms the basis of the proposed fusion method. The first stage reduces the difference between the multimodal images to obtain the previous fusion images by using the first type of multiscale decomposition, and then the second stage integrates the previous fusion images to get the final fusion image by using the second type of the previous fusion images, and the effect of the complementary feature fusion of multimodal images is further enhanced, effectively preventing feature loss.

Following are the contributions of the proposed method:

(1) Large difference between different modal images renders it difficult to extract the features of different modes with the same transformation method. Reducing the pixel difference between the images can ensure the feature extraction effect. This study introduces the serial structure fusion framework to reduce the difference between multimodal images and enhance the feature fusion effect.

(2) The first stage of fusion is to obtain similar previous fusion images, which is the input for the next stage. The hybrid $\ell_0\ell_1$ layer decomposition with the Gauss filter is introduced to decompose the multimodal images into structure-layer images, detail images, and large-scale edge features, and these feature images are integrated to obtain the previous fusion images that exhibit similar structure, brightness, and detail features in preparation for further enhanced integration.

(3) The second stage of fusion is to obtain high-quality fusion images. The NSST is adopted to enhance the fusion effect. For different frequency sub-band images, the local texture contrast and local engineering are introduced to highlight the fine features and local contrast features of low-frequency sub-images, and the local standard deviation matching degree and local modified spatial frequency (SF) matching degree of the exponential function are introduced to highlight the detail features, edge features, and weak texture detail features.

To demonstrate the fusion effect of the proposed method, its performance is compared with those of seven established fusion methods on visible and infrared images, infrared and infrared polarization intensity images, and medical images. The evaluation of the object and subject show that the proposed method yields brighter, more detailed fusion images than the other methods, with more edge features and a superior fusion effect.

The rest of the paper is organized as follows: Section 2 presents the framework of the two-stage fusion. The detailed process of the fusion algorithm is presented in Section 3. In Section 4, performance evaluation is conducted based on the experimental results. Finally, Section 5 concludes the study with a summary of the main contributions.

2 Multiscale Decomposition Analysis

2.1 The Hybrid $\ell_0\ell_1$ Layer Decomposition

The hybrid $\ell_0\ell_1$ layer decomposition is expressed as follows [30]:

$$\min_B = \sum_p^N \left\{ (S_P - S_B)^2 + \lambda_1 \sum_{i=\{x,y\}} \left| \partial_i B_P \right| + \lambda_2 \sum_{\mathbf{i}=\{x,y\}} F(\partial_i (S_P - B_P)) \right\} \quad (1)$$

where P is the pixel coordinates and N is the number of pixels. The first term $(S_P - B_P)^2$ forces the base layer to be close to the original image. The spatial property of the base layer is formulated as an ℓ_1 gradient sparsity term $\left|\partial_i B_P\right|$, $i = \{x, y\}$, ∂_i is the partial derivative operation along the x or y direction. The spatial property of the detail layer is formulated as an ℓ_0 gradient sparsity term with an indicating function $F(x)$:

$$F(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{2}$$

The base and detail layers of an image can be obtained by solving Eq. (1). The decomposition modal is as follows:

$$B^i = mod_{\ell_0\ell_1}(B^{i-1}) \tag{3}$$

$$D^i = B^{i-1} - B^i \tag{4}$$

where B_i is the i level structure-layer image, D_i is the i level detail-layer image, $B_0 = I$, and I is the source image.

The hybrid $\ell_0\ell_1$ layer decomposition combines the advantages of ℓ_0 and ℓ_1 decomposition. To reduce the halo artifacts and preserve the edges, the ℓ_1 gradient term is applied to the base layer. The hybrid $\ell_0\ell_1$ layer decomposition separates the image structure, details, and edge features well, thereby reducing the loss during fusion.

The first stage is intended to reduce the differences among multimodal images. Accordingly, different features of an image are integrated with those of another image to obtain the previous fusion images, which are similar. The hybrid $\ell_0\ell_1$ layer decomposition better meets the needs of this study.

2.2 Non-down-sampling Shear Wave Transform

NSST performs shift-invariant, anisotropic, multidirectional, and quick operations and can better extract different features from an image than NSCT, DWT, and SWT. Because the previous fusion images are highly similar and complementary, the second stage selects only the strongest multiscale transformations. The images are then decomposed using NSST.

3 Multimodal Image Fusion Based on Two-stage Multiscale Decomposition

Fig. 1 illustrates the framework of the proposed fusion method. First, the hybrid $\ell_0\ell_1$ layer decomposition separates a multimodal image into a structure- and detail-layer images. The structure-layer image is passed through a Gauss filter to obtain a large-scale edge image and the final structure image. Second, the structure-layer, detail-layer, and large-scale edge images are added to the source image to obtain the previous images. Third, the previous images are decomposed via NSST and the low-frequency and high-frequency sub-images are fused using multifeatures. The final fusion image is obtained by inverse NSST transform.

3.1 First-stage Fusion Based on the Hybrid $\ell_0\ell_1$ Layer Decomposition

Here, the structure feature and detail feature are decomposed using hybrid $\ell_0\ell_1$ layer decomposition. Increasing the number of decomposition levels improves the extraction effect. We control the number of decomposition layers by calculating the degree of loss of image detail information. SF describes the clarity of the image; hence, we use the SF similarity to describe the loss of detail feature to control the number of decomposition layers. The formula is as follows:

$$S = \frac{2 * SF_{B^i} * SF_{B^{i-1}}}{SF_{B^i}^2 + SF_{B^{i-1}}^2} \tag{5}$$

where S is the SF similarity, SF_{B^i} is the SF of the structure-layer image of the previous hybrid $\ell_0\ell_1$ layer decom-

position, and $SF_{B^{i-1}}$ is the *SF* of the structure-layer image of the latter hybrid $\ell_0\ell_1$ layer decomposition. The formula for *SF* is expressed as follows:

$$SF_B = \sqrt{RF^2 + CF^2} \tag{6}$$

where RF and CF are defined as follows:

$$RF = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(B(x,y) - B(x,y-1))}{M \times N}} \tag{7}$$

$$CF = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(B(x,y) - B(x-1,y))}{M \times N}} \tag{8}$$

where SF_B is the *SF* of the structure-layer image of the hybrid $\ell_0\ell_1$ layer decomposition, *RF* is the row *SF* intensity of the SF_B, and *CF* is the column *SF* intensity of the SF_B.

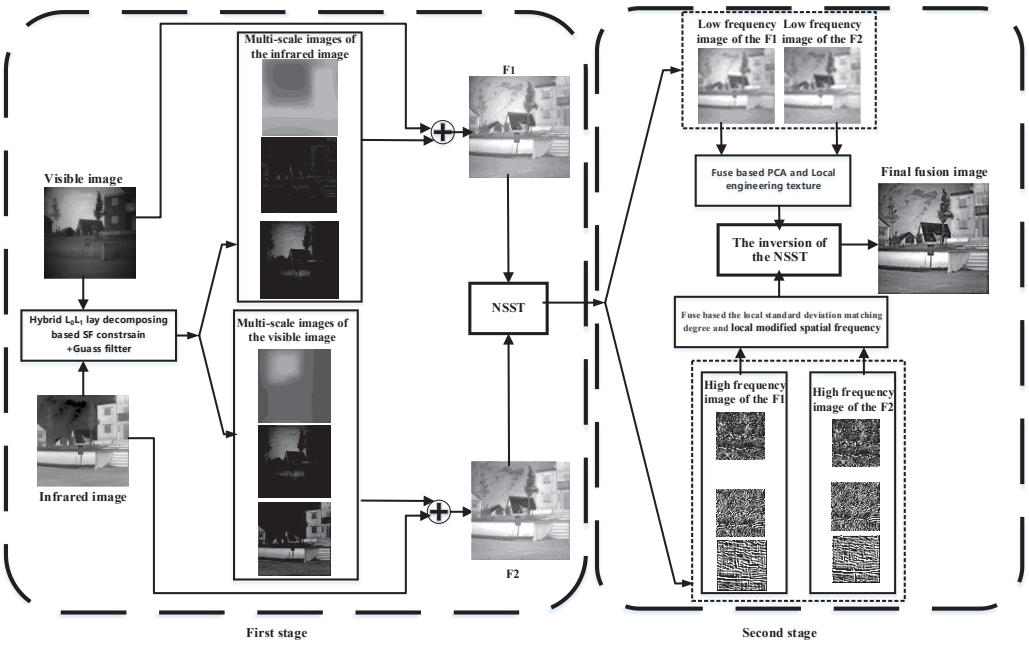


Fig. 1. The framework of the proposed fusion method

The hybrid $\ell_0\ell_1$ layer decomposition based on the *S* similarity constraint is expressed as follows:

$$\begin{cases} i+1 \ \ and \ \ B^{'} = \mathrm{mod}_{\ell_0\ell_1}(B^{i-1}) \ \ if \ \ S \geq th \\ stop\,\mathrm{mod}_{\ell_0\ell_1} \quad if \ \ S < th \end{cases} \tag{9}$$

where *i* is the number level of the decomposition and *th* is the threshold of the *S*.

Fig. 2 illustrates the final structure images of the hybrid $\ell_0\ell_1$ layer decomposition of the threshold of the different *S* and *SF* variance ratios. Fig. 3 shows that with decreasing *S*, the detail features of the source images are extracted. Further, when $S \leq 0.6$, the visible images and infrared polarization intensity images only leave the structure features, and when $S \leq 0.7$, the detail features of the infrared images, MRI images, and CT images are completely extracted. The figure of the *SF* variance ratio demonstrates that with decreasing *S*, *V* exhibits a greater drop in yield, particularly when $S \leq 0.7$ and *V* is negligible. Therefore, when the visible images and infrared polarization images are decomposed, the threshold of the *S* is 0.6 and when the infrared, MRI, and CT images are decomposed, the threshold of the *S* is 0.7 based on the experiment. The *SF* variance ratio is as follows:

$$V = \frac{SF_S}{SF_I} \tag{10}$$

where *V* is the *SF* variance ratio, SF_S is the *SF* of the structure image, and SF_I is the *SF* of the source image.

The Previous Fusion Images. After the decomposition of the multimodal images, they are decomposed into multilayer detail and structure images. The structure image includes the local contrast feature and the large-scale edge feature. Thus, we use the Gauss filter to smooth the structure image to obtain the large-scale edge image and the final structure image. The formulas are as follows:

$$I_{Smooth} = B^k \otimes W \tag{11}$$

$$I_{edge} = B^k - I_{smooth} \tag{12}$$

$$W(x, y) = K * e^{-\frac{x^2+y^2}{\sigma^2}} \tag{13}$$

where I_{smooth} is the filtered image by the Gauss filter, B^k is the last level stricture image, *W* is the Gauss filter, and I_{edge} is the large edge image. In the study, we use the 3×3 Gauss filter.

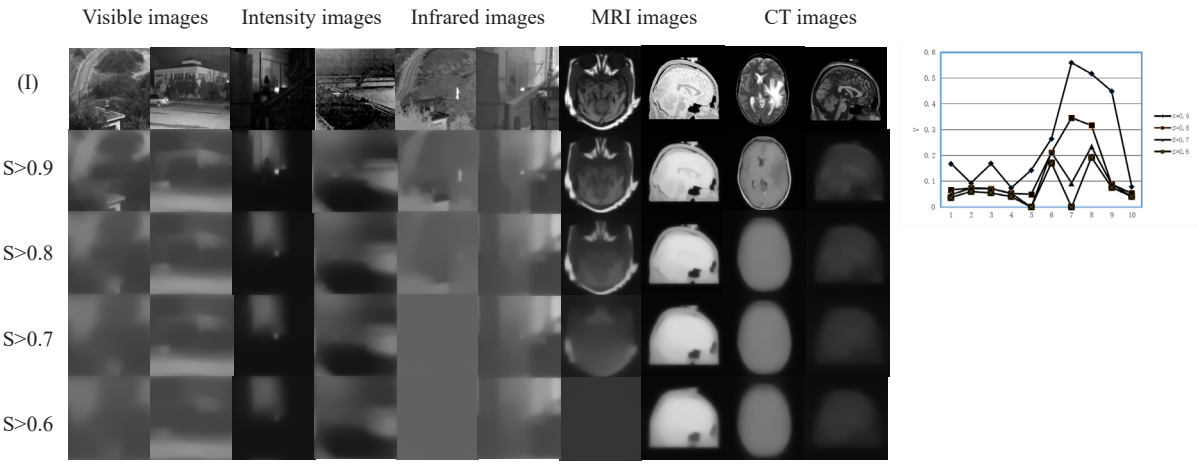


Fig. 2. (a) The final structure images of the $\ell_0\ell_1$ hybrid layer decomposing of the threshold of the different *S*, (I) source images (b) *SF* variance ratio

The multimodal images are decomposed into detail layers, large-scale edges, and structure images based on the hybrid $\ell_0\ell_1$ layer decomposition and the Gauss filter. The detail and large-scale edge images usually belong to the weak features, and the structure image belongs to the low-frequency feature and concentrates the main energy of the image and is the strong feature. Therefore, to reduce the loss of features and excessive fusion, the

detail and large-scale edge images are retained in the previous images and the structure image is fused by using the weighted average. The formulas are as follows:

$$F_1 = I_1 + I_{edge}^2 + \sum_{i=1}^{k} D_2^i + \frac{u_{smooth}^1}{u_{smooth}^1 + u_{smooth}^2} * I_{smooth}^2 \tag{14}$$

$$F_2 = I_2 + I_{edge}^1 + \sum_{i=1}^{k} D_1^i + \frac{u_{smooth}^2}{u_{smooth}^1 + u_{smooth}^2} * I_{smooth}^1 \tag{15}$$

$$u_{smooth}^l = \frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=}^{N} I_{smooth}^l (x, y) \tag{16}$$

where F_1 and F_2 are the previous fusion images, D_1^i and D_2^i are the detail layer images, i is the number of decomposition level, i = 1, 2, ..., k, u_{smooth}^1 and u_{smooth}^2 are the mean values of I_{smooth}^1 and I_{smooth}^2, respectively, and I_1 and I_2 are the source images.

3.2 Second-stage Fusion Based on the NSST

Fig. 3 shows the previous images F_1, F_2, the final fusion images (F_2) of the visible and infrared images, the final fusion images (F) of the infrared and infrared polarization intensity images, the evaluation index of F_1, F_2, and F, and the structure similarity between F_1 and F_2.

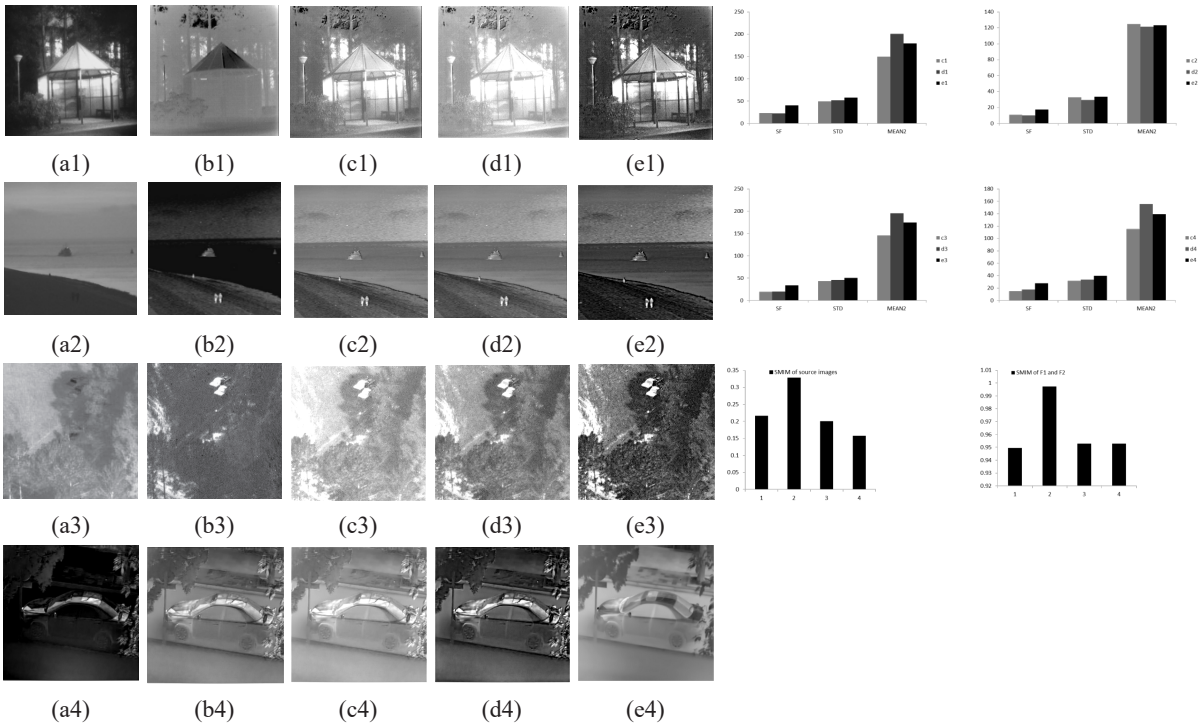


Fig. 3. (a1)-(a2) Visible images, (b1)-(b2) Infrared images, (a3)-(a4) infrared image, (b3)-(b4) Infrared polarization intensity image, (c1)-(c4) F_1, (d1)-(d4) F_2, (e1)-e4) F, (f) The evaluation index values of the F_1, F_2, F, and the structure similarity between the source images, and the structure similarity between F_1 and F_2

Fig. 3 demonstrates that F_1 and F_2 are similar and integrate the features of the multimodal images; however, F_1 and F_2 have complementary features; for example, F_1 exhibits good visual effects and F_2 exhibits good detail. However, the contrast in F_2 is too strong. After the fusion of F_1 and F_2, F exhibits good visual, detail, and contrasting features, fuses the advantages of F_1 and F_2, and considerably improves the fusion effect. The figures of

the evaluation index show that the structure similarity (SMIM) between F_1 and F_2 is more than 0.9, and the structure similarity between the source images is considerably small. The difference between F_1 and F_2 is small, and the *SF* and standard deviation of *F* are higher than F_1 and F_2, and the gray mean values of *F* are the middle value between F_1 and F_2. Furthermore, the fusion of F_1 and F_2 could considerably improve the fusion image.

Fusion Rule of the Low-frequency Sub-band Image. The current fusion algorithms mainly consider the structure feature; thus, the fusion rules usually use a weighted average based on the pixel value. However, the low-frequency sub-band image focuses on the main energy and includes the structure feature but also includes the texture feature; hence, we fuse the low-frequency sub-band images based on these two features.

The structure feature is the main feature of the low-frequency sub-band image, and PCA can extract the main feature of the image; therefore, we use the PCA to integrate the structure feature. The formulas are as follows:

$$L = \left[L_{F_1}(:) \quad L_{F_2}(:) \right] \tag{17}$$

$$G = COV(L', L) \tag{18}$$

$$[D \quad V] = eig(G) \tag{19}$$

where L_{F_1} and L_{F_2} are the low frequency sub-band images after the NSST transformation; $L_{F_1}(:)$ and $L_{F_2}(:)$ are the $MN\times1$ vector of the L_{F_1} and L_{F_2}, respectively; *L* is $MN\times2$ matrix L' is the transposition of the *L*, *G* is the covariance matrix of *L*, *D* is the eigenvalues of the covariance matrix, and *V* is the characteristic vector.

The eigenvalues of the covariance matrix reflect the importance of the image features, and the sine function exhibits the characteristics of the low-frequency filter and could remove the noise, so we use the eigenvalues and sine function to integrate the structure feature. The formulas are expressed as follows:

$$F_S^{LF} = \begin{cases} D(1,1) \geq D(2,2) \sin(\sqrt{D(1,1)/(D(1,1)+D(2,1))} * L_{F_1} + \sin(\sqrt{D(2,2)/(D(1,1)+D(2,1))} * L_{F_2} \\ D(1,1) \leq D(2,2) \sin(\sqrt{D(2,2)/(D(1,1)+D(2,1))} * L_{F_1} + \sin(\sqrt{D(1,1)/(D(1,1)+D(2,1))} * L_{F_2} \end{cases} \tag{20}$$

where F_S^{LF} is the fusion image of the structure feature of L_{F_1} and L_{F_2}.

The texture feature reflects the change between the gray values. The gradient can describe the change in direction of the image, and the larger the difference between the pixel values, the higher the gradient is. However, the local energy could reflect the large edge change of the low-frequency sub-band image. Herein, we adopt the gradient and local energy to describe the detail features of low-frequency sub-band images. The formulas are expressed as follows:

$$dx_{F_i}^2(x,y) = \sum_{x=1}^{M} \sum_{y=1}^{N} (L_{F_i}(x+1,y) - L_{F_i}(x-1,y))^2 \tag{21}$$

$$dx_{F_i}^2(x,y) = \sum_{x=1}^{M} \sum_{y=1}^{N} (L_{F_i}(x,y+1) - L_{F_i}(x,y-1))^2 \tag{22}$$

$$LE_{F_1}(x,y) = \sqrt{\frac{1}{9} \sum_{k=-1}^{1} \sum_{l=-1}^{1} L_{F_1}(x+k,y+k)^2} \tag{23}$$

$$LE_{F_2}(x,y) = \sqrt{\frac{1}{9} \sum_{k=-1}^{1} \sum_{l=-1}^{1} L_{F_2}(x+k,y+k)^2} \tag{24}$$

$$C_i^T(x,y) = LE_{F_i} + \sqrt{dx_i^2 + dy_i^2} \tag{25}$$

$$F_T^{LF}(x,y)=\sum_{x=1}^{M}\sum_{y=1}^{N} a\tan\left(\frac{C_{F_1}^{T}(x,y)}{C_{F_1}^{LF}(x,y)+C_{F_2}^{LF}(x,y)}\right)*L_{F_1}(x,y)+\sum_{x=1}^{M}\sum_{y=1}^{N} a\tan\left(\frac{C_{F_2}^{T}(x,y)}{C_{F_1}^{LF}(x,y)+C_{F_2}^{LF}(x,y)}\right)*L_{F_2}(x,y) \quad (26)$$

where dx_i^2 and dy_i^2 are the x and y direction gradients of L_{F_1} and L_{F_2}, respectively; C_i^T is the detail feature of the L_{F_1} and L_{F_2}, LE_i is the 3×3 local energy of L_{F_1} and L_{F_2}, $i = F_1$ or F_2, atan is the arctan function into smooth weight, and F_T^{LF} is the fusion image of the texture feature.

The contrasting feature is used to integrate F_T^{LF} and F_S^{LF}. The formulas are expressed as follows:

$$Fused\{1\}(x,y)=a\tan\left(\frac{Lu_{F_S^{LF}}}{u_1}\right)*F_S^{LF}(x,y)+a\tan\left(\frac{Lu_{F_T^{LF}}}{u_2}\right)*F_T^{LF}(x,y) \quad (27)$$

$$u_1=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}F_S^{LF}(x,y) \quad (28)$$

$$u_2=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}F_T^{LF}(x,y) \quad (29)$$

$$Lu_S^{LF}=\frac{1}{9}\sum_{k=-1}^{1}\sum_{l=-1}^{1}F_S^{LF}(x+k,y+l) \quad (30)$$

$$Lu_T^{LF}=\frac{1}{9}\sum_{k=-1}^{1}\sum_{l=-1}^{1}F_T^{LF}(x+k,y+l) \quad (31)$$

where u_1 and u_2 are the mean values of F_T^{LF} and F_S^{LF}, respectively; Lu_S^{LF} and Lu_T^{LF} are the local mean values of F_T^{LF} and F_S^{LF}, respectively; and $Fused\{1\}$ is the final fusion image of the low-frequency sub-band images.

Fusion Rule of the High-frequency sub-band Images Based on Different Active Measures. The high-frequency sub-band images of the NSST reflect the different types of detail features, including the large, middle, and small detail features. Therefore, the different active measures need to be used to describe the different types of detail features. Herein, the standard deviation reflects the contrast, which is used to measure the large and middle texture features, and the modified *SF* reflects the different direction change of the pixel value, which is used to measure the small texture feature. The modified spatial frequencies are expressed as follows:

$$MSF=\sqrt{FRF+FCF+FMD+FSD} \quad (32)$$

$$FRF(x,y)=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}(c_i\{d\}(x,y)-c_i\{d\}(x,y-1))^2 \quad (33)$$

$$FCF(x,y)=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}(c_i\{d\}(x,y)-c_i\{d\}(x-1,y))^2 \quad (34)$$

$$FMD(x,y)=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}(c_i\{d\}(x,y)-c_i\{d\}(x-1,y-1)-c_i\{d\}(x+1,y+1))^2 \quad (35)$$

$$FSD(x,y)=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}(c_i\{d\}(x,y)-c_i\{d\}(x-1,y+1)-c_i\{d\}(x+1,y-1))^2 \quad (36)$$

where *MSF* is the modified spatial frequency, *FRF* is the row *SF* intensity of high-frequency sub-band images, *FCF* is the column *SF* intensity of the high-frequency sub-band images, *FMD* and *FSD* are the diagonal line *SF* intensity of the high-frequency sub-band images, $c_i\{d\}$ is the high-frequency sub-band images of F_1 or F_2, and d is the high frequency decomposition number of layers of NSST.

The fusion based on the MSF is expressed as follows:

$$Fused_1\{d\}(x,y)=\sum_{x=1}^{M}\sum_{y=1}^{N}\frac{MSF_{F_1}\{d\}(x,y)}{MSF_{F_1}\{d\}(x,y)+MSF_{F_2}\{d\}(x,y)}c_{F_1}\{d\}(x,y)+\frac{MSF_{F_2}\{d\}(x,y)}{MSF_{F_1}\{d\}(x,y)+MSF_{F_2}\{d\}(x,y)}c_{F_2}\{d\}(x,y) \quad (37)$$

where $Fused_1\{d\}$ is the fusion image based on *MSF*; and $MSF_{F_1}\{d\}$ and $MSF_{F_2}\{d\}$ are *MSF* of the high-frequency sub-band images.

The fusion based on the *STD* is expressed as follows:

$$T\{d\}(x,y)=\sum_{x=1}^{M}\sum_{y=1}^{N}\frac{2*S_1*S_2}{S_1^2+S_2^2} \quad (38)$$

$$S_i\{d\}(x,y)=\sqrt{\frac{1}{9}*\sum_{k=-1}^{1}\sum_{l=-1}^{1}(c_i\{d\}(x+k,y+l)-u_i\{d\})^2} \quad (39)$$

$$u_i\{d\}=\frac{1}{9}*\sum_{k=-1}^{1}\sum_{l=-1}^{1}c_i\{d\}(x+k,y+k) \quad (40)$$

$$Fused_2\{d\}(x,y)=\sum_{x=1}^{M}\sum_{y=1}^{N}\exp(-(1-T\{d\}(x,y)))*c_{F_1}\{d\}(x,y)+\exp(-T\{d\}(x,y))*c_{F_2}\{d\}(x,y) \quad (41)$$

where $T\{d\}(x,y)$ is the local standard deviation match degree, $S_i\{d\}$ is the local standard deviation of the high-frequency sub-band images $u_i\{d\}$ is the local meaning value of the high-frequency sub-band images, and $Fused_2\{d\}$ is the fusion image based on the *STD*.

$Fused_1\{d\}$ and $Fused_2\{d\}$ are the different feature fusion images. Herein, we use the average weighted in to fuse the different feature fusion images. The formulas are expressed as follows:

$$Fused\{d\}(x,y)=\begin{cases}Fused_1\{d\}(x,y)+\dfrac{Fused_2\{d\}(x,y)}{Fused_2\{d\}(x,y)+Fused_1\{d\}(x,y)}*Fused_2\{d\}(x,y) & if\ Fused_1\{d\}(x,y)\geq Fused_1\{d\}(x,y)\\ Fused_2\{d\}(x,y)+\dfrac{Fused_1\{d\}(x,y)}{Fused_2\{d\}(x,y)+Fused_1\{d\}(x,y)}*Fused_1\{d\}(x,y) & if\ Fused_1\{d\}(x,y)\leq Fused_1\{d\}(x,y)\end{cases} \quad (42)$$

where $Fused\{d\}$ is the final fusion image of the high-frequency sub-band images.

4 Analysis of the Experimental Result

To illustrate the validity of the proposed fusion algorithm, the low illumination visual and infrared images, the high illumination visual and infrared images, the infrared polarization and infrared intensity images, and the multimodal medical images are used as the fusion data.

4.1 Fusion Images of the Visual and Infrared Images

For the visual and infrared images as well as infrared polarization and infrared intensity images, seven fusion algorithms are used to compare with the proposed method. The proposed fusion algorithm was compared with RGF_MDFB [31], NSCT-SR [9], Hybrid-MSD [32], CNN [33], GFF [34], VILS [35], and GTF [36] for multimodal infrared image fusion, infrared and visual image fusion.

Fig. 4 shows the infrared and low illumination visual images and the fusion images of the different fusion algorithms. Fig. 4 demonstrates that the obtained fusion images better fuse the complementary features between the infrared and low-light visual images and better integrate the detail features of the visual images with contrasting and brightness features of the infrared images, such as the detail features of the tree and the sky, contrasting features of the people, and structure feature of the building. The fusion images obtained by this paper have the better visual effect than the other fusion algorithms, and reduce the loss of the bright feature, and the quality of the fusion images is salient improved.

Fig. 5 shows the infrared and high illumination visual images and the fusion images of the different fusion algorithms. The figure shows that the obtained fusion images better preserve the contrasting feature of the infrared images and the detail feature of the visual images, are not supersaturated, do not introduce noise, and the fusion effect is better than the other fusion algorithms, such as the structure feature of the electric wire and the cloud, texture feature of the tree, brightness of the light, feature of the person, and detail and structure features of the car. The other fusion images introduce the bright distortion, and lose the detail features, and do not fuse the complementary features between the infrared and visible images. While the proposed fusion algorithm effectively reduces the halo effect of the headlights.

Although the other fusion algorithms also fuse the images well, the other fusion images have loss, and the other fusion algorithms have distortion, such as the RGF-GDBF, Hybrid-MSD, CNN, and GFF, which lose the contrast feature of the infrared image, there is a loss of the detail feature of the visual image in the VILS and GTF, and the fusion images of the NSCT-SR, which produce distortion.

4.2 Fusion Images of the Infrared Polarization and Intensity Images

Fig. 6 shows the fusion images of the infrared polarization and intensity images. The figure shows that the obtained fusion images have better clarity and better preserve the detail feature, the edge feature, and the local contrasting feature of the infrared polarization image; for example, the texture feature of the tree, the car window, the surface of the water, and the top of the building have strong polarization. While the proposed algorithm fused the structural features of the infrared intensity images, such as the car, building, and person, and does not introduce the distortion. However the GFF exhibited loss of the features of the infrared intensity images, and the NSCT-SR exhibited distortion, and the other fusion methods do not fuse the complementary features between the infrared polarization and intensity images. Hence, the proposed algorithm could integrate the detail, structure, contrasting, and brightness features of the infrared polarization and intensity images, and the fusion algorithm proposed by this paper salient improves the fusion effect.

4.3 Fusion Images of the CT and MRI

For multimodal medical image fusion, the proposed fusion algorithm was compared with RGF_MDFB [31], NSCT–PCNN [37], as well as models based on Guide filter (GFS) [38], deep learning (CNN) [39], morphological component analysis (MCA) [40], and DWT–PCA (DWP) [41].

Fig. 7 demonstrates that the fusion effect of fusion images obtained by using the proposed algorithm is better than the other fusion images. The fusion images proposed by this paper have the best edge, detail, and contrast features. For example, the fusion images of the MAC did not preserve the detail feature well; the edge feature of the secondary column images is not preserved well in the other fusion images; and the DWT–PCA loses the contrast feature.

4.4 Objective Evaluation

Herein, the four evaluation indexes are used to evaluate the different fusion algorithms. The SF is adopted to evaluate the clarity of the fusion image, the standard deviation (*STD*) is adopted to evaluate the contrast of the fusion image, and the edge strength (*QE*) estimates the edge feature. For evaluating the transfer ability of the proposed fusion algorithm, the difference seminary index ($R_{ab/f}$) [42] is used, which evaluates the similarity between the fusion image and the source image.

The formulas of the *QE* are expressed as follows:

$$QE = \sum_{i=1}^{M}\sum_{j=1}^{N} GF(i,j) \tag{43}$$

The gradient value of the pixel (GF) is defined as follows.

$$GF(i,j) = \sqrt{S_F^x(i,j)^2 + S_F^y(i,j)^2} \tag{44}$$

Here, $S_F^x(i,j)^2$ and $S_F^y(i,j)^2$ indicate the sobel operator convoluting image F horizontally and vertically, respectively

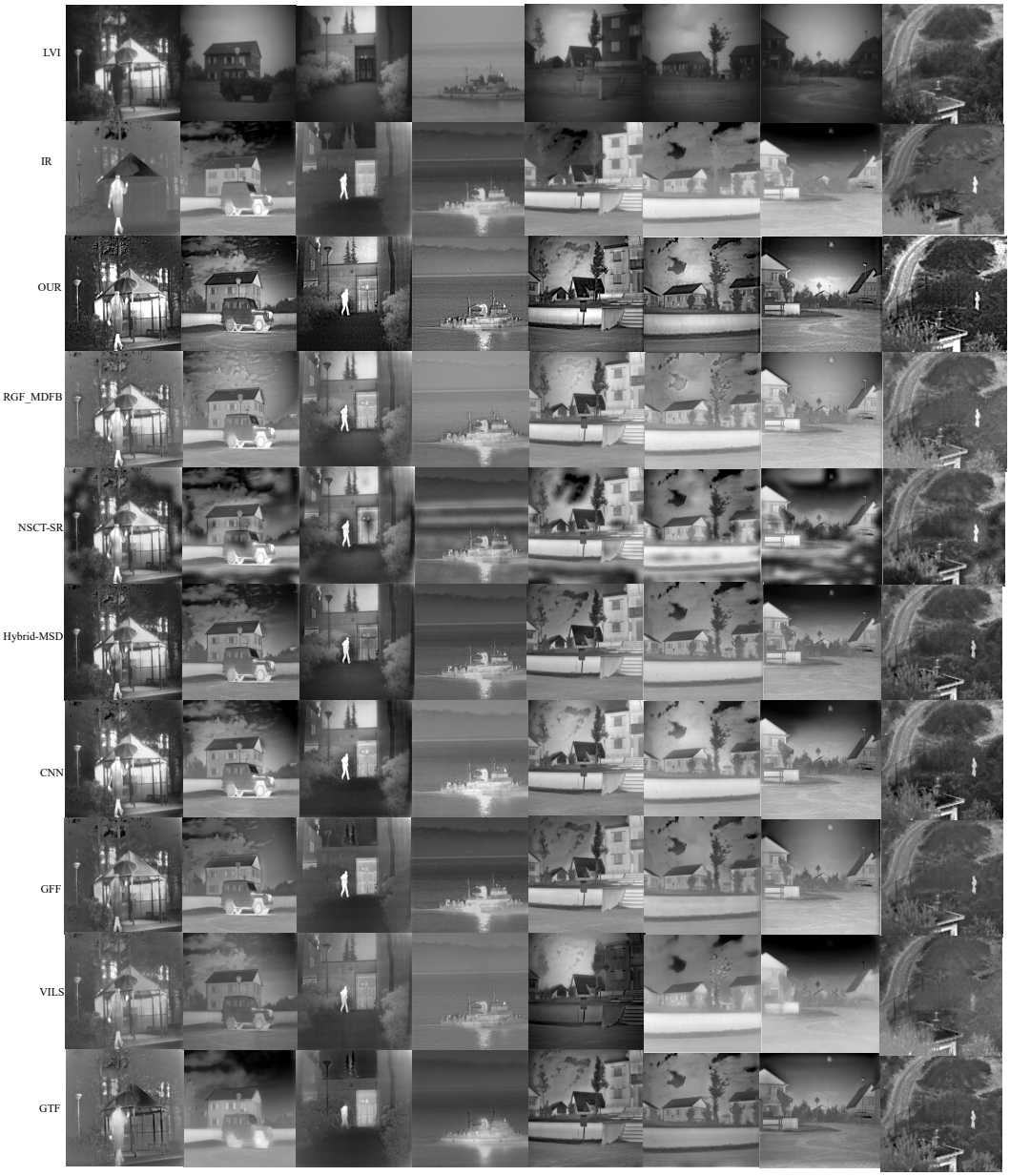


Fig. 4. The infrared and low illumination visual image and the infrared and visual image fusion of the different fusion algorithms

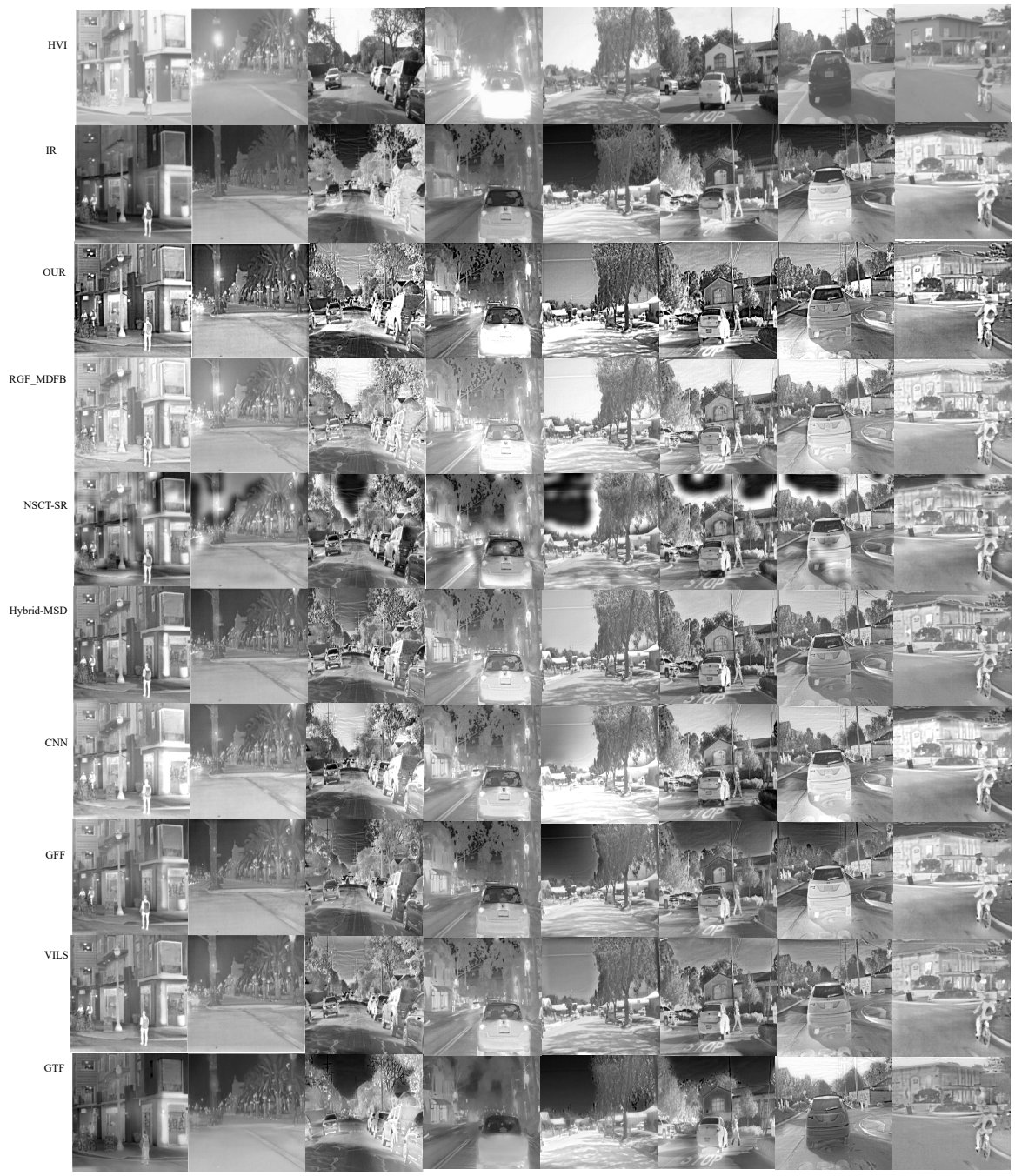


Fig. 5. The infrared and high illumination visual image, and the infrared and visual image fusion of the different fusion algorithms

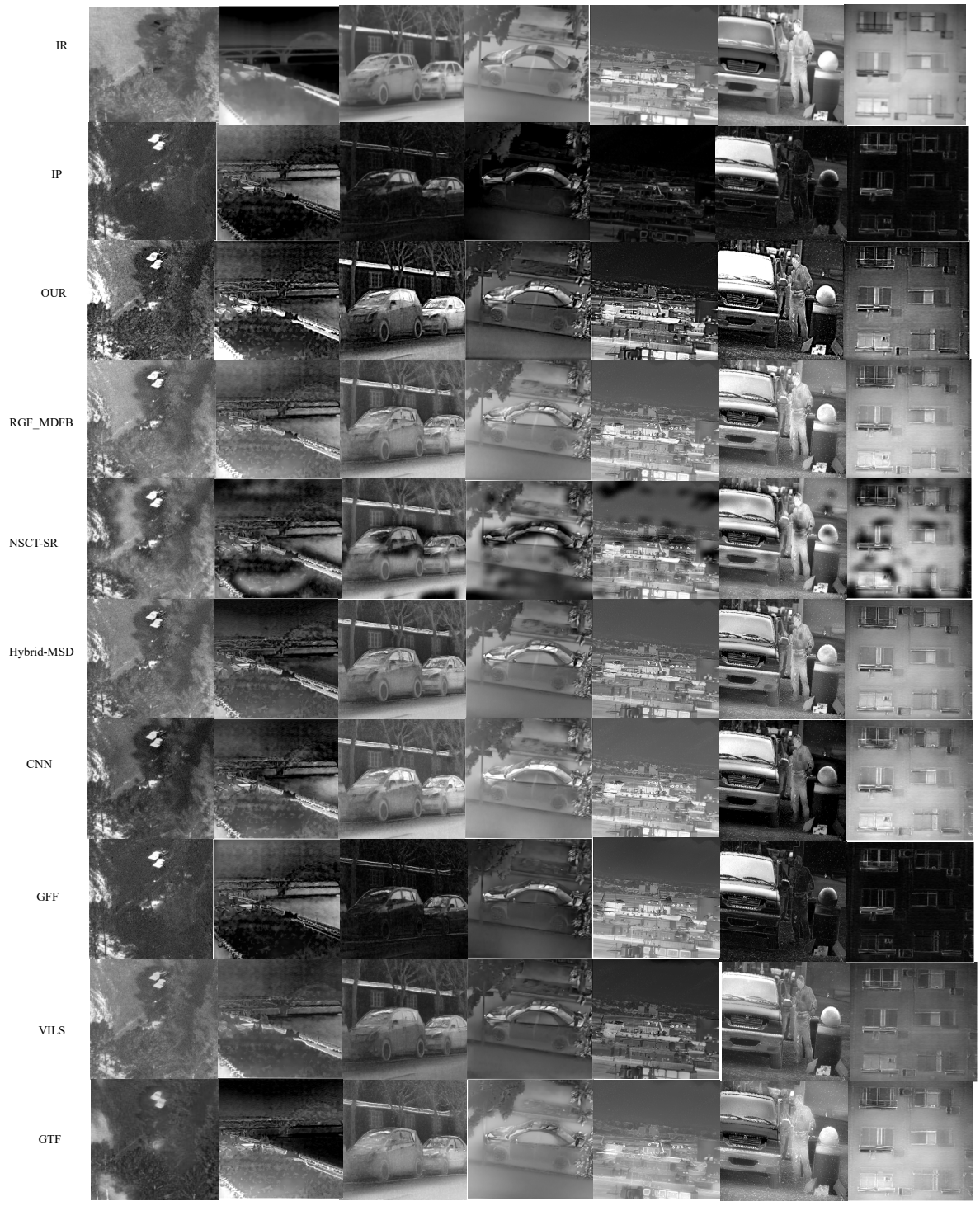


Fig. 6. The infrared intensity and polarization images, and the image fusion of the different fusion algorithms

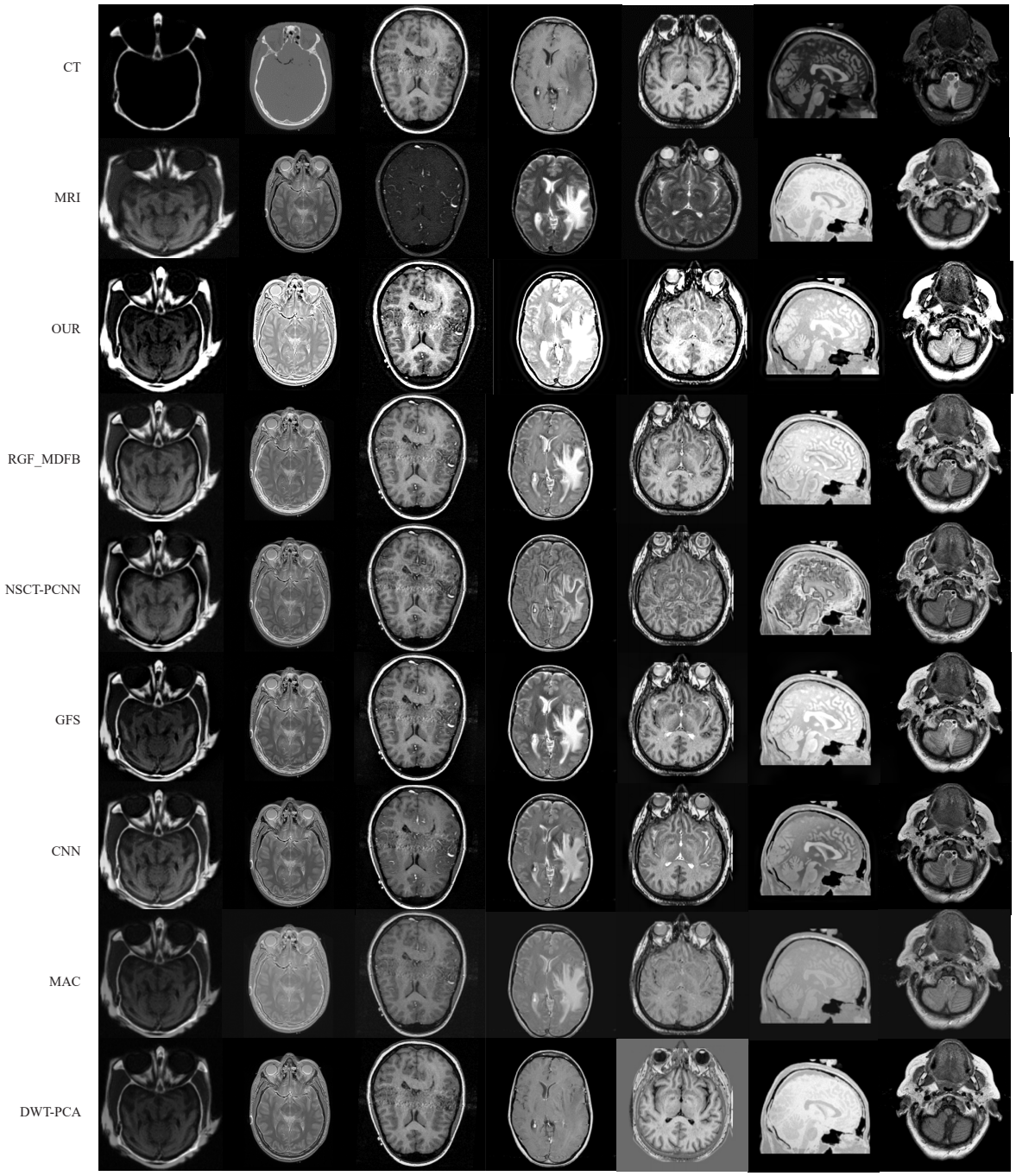


Fig. 7. CT and MRI image, and the CT and MRI image fusion of the different fusion algorithms

The formulas for $R_{ab/f}$ are expressed as follows:

$$r(D,S) = \frac{\sum_i \sum_j (D(i,j) - \overline{D}) * (S(i,j) - \overline{S})}{\sqrt{\sum_i \sum_j (D(i,j) - \overline{D})^2 + \sum_i \sum_j S(i,j) - \overline{S})^2}} \quad (45)$$

$$R_{ab/f} = r(D_I, S_I) + r(D_k, S_k) \tag{46}$$

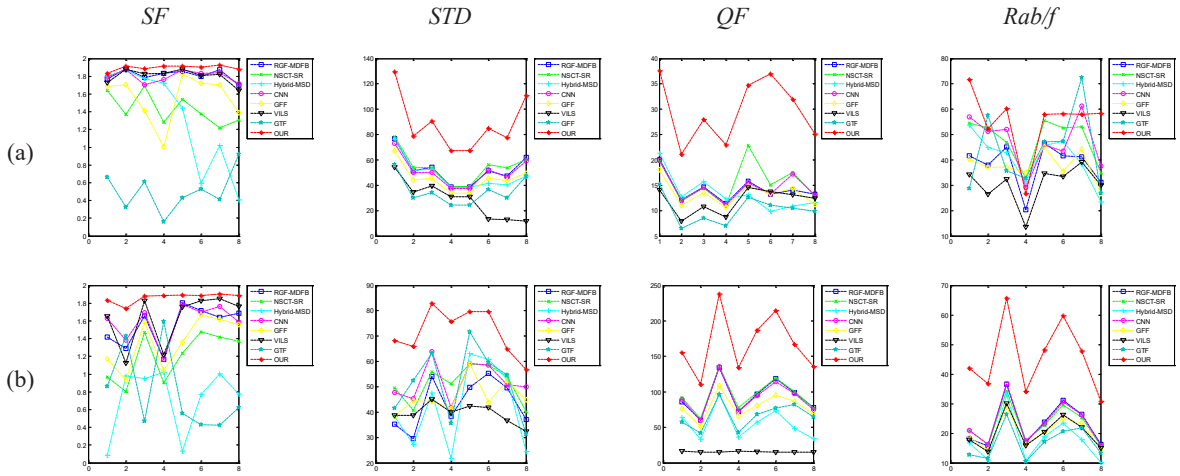
where r is the correlation of the differences, D is the difference image between the fused image and source images, S represents the source images, D_I is the difference between F and S_k, and D_k is the difference between F and S_I. In addition, $\overline{D}$ and $\overline{S}$ are the mean values of D and S respectively, and $R_{ab/f}$ is the sum of the correlation of the differences.

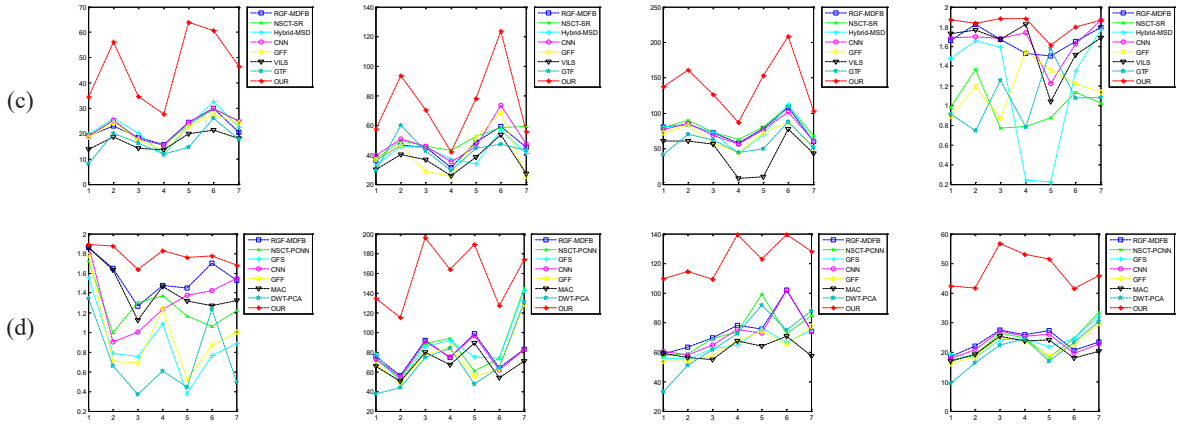
Fig. 8 demonstrates the objective evaluation results of the different fusion algorithms, which indicate that *SF* and *QF* of the method of this paper are the best; hence, the fusion images have the best clarity and edge features. Further, because the *STD* of the fusion algorithm is overall high, the fusion images have contrasting features. In addition, $R_{ab/f}$ is larger than the other fusion algorithms. Thus, the method of this paper can transfer the complementary features between the source images well, which shows that the method of this paper can effectively reduce the information loss of the source images, and integrates the structure, detail, brightness, and the contrasting features very well, and salient improves the quality of multimodal images. Based on the analysis of the objective evaluation, the proposed method can fuse the detail feature, edge feature, and structure feature of the modal image and silently improve the fusion effect.

In summary, the clarity, edge, and contrasting feature values of the fusion images considerably improve, while the fusion images cannot introduce distortion and have a good visible effect, and the proposed method can integrate the complementary features between the modal images very well and can availably reduce the feature loss.

5 Conclusion

In this study, multimodal image fusion with a two-stage fusion algorithm based on hybrid multiscale decomposition is proposed. This method combines the advantages of hybrid $\ell_0\ell_1$ layer decomposition and NSST. In the first stage, the hybrid $\ell_0\ell_1$ layer decomposition based on the SF constraint is used to obtain the previous fusion images to reduce the difference between the source images, and the previous fusion images are used in the input images of the next stage. In the second stage, the NSST based on different fusion rules is used to fuse the previous fusion images. The experimental results show the proposed method could availably fuse the detail, edge, contrast, structure, and weak features of the different types of multimodal images and observably improves the quality of the fusion image of the different types of modal images. In the future, the research focus is to improve the efficiency of multiscale decomposition to better extract different types of image features, while reducing the operation time and apply it to practice.





(a) The evaluation values of low illumination visual and infrared fusion images; (b) The evaluation values of high illumination visual and infrared fusion images; (c) The evaluation values of infrared polarization and intensity fusion images; (d) The evaluation values of CT and MRI fusion images.

Fig. 8. The evaluation values of the different fusion algorithms

6 Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61672472), and the Doctoral Program of Nanyang normal University (19202), and the open fund project in 2020 of Key Laboratory of Marine Environmental Information Technology (The research on remote sensing image sea ice classification based on generated adversarial network), and the Key Scientific and Technological Project of Henan Province (Grant No. 232102211047).

References

[1] L.H. Jian, R. Rayhana, L. Ma, S.W. Wu, Z. Liu, H.Q. Jiang, Infrared and Visible Image Fusion Based on Deep Decomposition Network and Saliency Analysis, IEEE Transactions on Multimedia 24(2022) 3314-3326.

[2] Y.L. Zhu, R.L. Liu, X.W. Li, J. Wang, Research on Medical Image Fusion Based on PCA, in: Proc. 2018 International Conference on Computational, modeling, Simulation and Mathematical Statistics, 2018.

[3] X. Jin, Q. Jiang, X. Chu, X. Lang, S.W. Yao, K.Q. Li, W. Zhou, Brain Medical Image Fusion Using L2-Norm-Based Features and Fuzzy-Weighted Measurements in 2-D Littlewood-Paley EWT Domain, IEEE Transactions on Instrumentation and Measurement 69(8)(2020) 5900-5913.

[4] H. Hermessi, O. Mourali, E. Zagrouba, Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain, Neural computing & applications 30(7)(2018) 2029-2045.

[5] X.S. Li, W.J. Wan, F.Q. Zhou, X.Q. Cheng, Y.C. Jie, H.S. Tan, Medical image fusion based on sparse representation and neighbor energy activity, Biomedical signal processing and control 80(2)(2023) 104353-1-104353-17.

[6] K. Kittusamy, L.S.V.S. Kumar, Sub-Sampled Contourlet with Joint Sparse Representation Based Medical Image Fusion, Computer Systems Science & Engineering 44(3)(2023) 1989-2005.

[7] B.Z. Wei, X.C. Feng, W.W. Wang, 3M: A Multi-scale and Multi-directional method for Multi-focus image fusion, IEEE Access 9(2021) 48531-48543.

[8] F. Fakhari, M.R. Mosavi, M.M. Lajvardi, Image fusion based on multi-scale transform and sparse representation: an image energy approach, Iet Image Processing 11(11)(2017) 1041-1049.

[9] Y. Liu, S.P. Liu, Z.F. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Information Fusion 24(2015) 147-164.

[10] S.K. Panguluri, L. Mohan, A DWT Based Novel Multimodal Image Fusion Method, Traitement du Signal 38(3)(2021) 607-617.

[11] W.W. Kong, Q.G. Miao, Y. Lei, C. Ren, Guided filter random walk and improved spiking cortical model based image fusion method in NSST domain, Neurocomputing 488(2022) 509-527.

[12] Q.Y. Xie, L. Ma, Z.Q. Guo, Q. Fu, Z.L. Shen, X.L. Wang, Infrared and visible image fusion based on NSST and phase consistency adaptive DUAL channel PCNN, Infrared Physics & Technology 131(2023) 104659-1-104659-14.
[13] N.A. Lawrance, T.S.S. Angel, Image Fusion Based on NSCT and Sparse Representation for Remote Sensing Data, Computer Systems Science and Engineering 46(3)(2023) 3439-3455.
[14] A. Selvaraj, P. Ganesan, Infrared and visible image fusion using Multi-scale NSCT and rolling-guidance filter, IET Image Processing 14(16)(2020) 4210-4219.
[15] B.S. Babu, M.V. Narayana, Two stage multi-modal medical image fusion with marine predator algorithm-based cascaded optimal DTCWT and NSST with deep learning, Biomedical Signal Processing and Control 85(2023) 104921-1-104921-16.
[16] W. Tan, W. Thitøn, P. Xiang, H.X. Zhou, Multi-modal brain image fusion based on multi-level edge-preserving filtering, Biomedical Signal Processing and Control 64(2021) 102280-1-102280-13.
[17] D.C. Lepch, A. Dogra, B. Goyal, J.S. Chohan, D. Koundal, A. Zaguia, H. Hamam, Multimodal Medical Image Fusion Based on Pixel Significance Using Anisotropic Diffusion and Cross Bilateral Filter, Human-centric Computing and Information Sciences 12(2022) 1-19.
[18] L. Ren, Z.B. Pan, J.Z. Cao, H. Zhang, H. Wang, Infrared and visible image fusion based on edge-preserving guided filter and infrared feature decomposition, Signal Processing 186(2021) 108108-1-108108-14.
[19] H.C. Park, R. Ghimire, S. Poudel, S.W. Lee, Deep Learning for Joint Classification and Segmentation of Histopathology Image, Journal of Internet Technology 23(4)(2022) 903-910.
[20] C.R. Dhivyaa, K. Nithya, K. Karthika, S. Mythili, Multi-Feature Integrated Concurrent Neural Network for Human Facial Expression Recognition, Journal of Internet Technology 23(6)(2022) 1263-1274.
[21] D.P. Zou, B. Yang, Infrared and low-light visible image fusion based on hybrid multiscale decomposition and adaptive light adjustment, Optics and Lasers in Engineering 160(2023) 107268-1-107268-18.
[22] S. Maqsoo, U. Javed, Multi-modal Medical Image Fusion based on Two-scale Image Decomposition and Sparse Representation, Biomedical Signal Processing and Control 57(2020) 101810-1-101810-8.
[23] Y. Yang, Y.M. Zhang, S.Y. Huang, Y.F. Zou, J.C. Sun, Infrared and Visible Image Fusion Using Visual Saliency Sparse Representation and Detail Injection model, IEEE Transactions on Instrumentation and Measurement 70(2020) 5001715-1- 5001715-15.
[24] M.G. Zhao, Y.P. Peng, A Multi-module Medical Image Fusion Method Based on Non-subsampled Shear Wave Transformation and Convolutional Neural Network, Sensing and Imaging 22(1)(2021) 1-16.
[25] A.I. Abas, H.E. Koçer, N.A. Baykan, Medical image fusion with convolutional neural network in multiscale transform domain, Turkish Journal of Electrical Engineering and Computer Sciences 29(SI-1)(2021) 2780-2794.
[26] A.U. Moonon, J.W. Hu, Multi-Focus Image Fusion Based on NSCT and NSST, Sensing and imaging 16(1)(2015) 1-16.
[27] J.W. Li, B.H. Li, Y.X. Jiang, An Infrared and Visible Image Fusion Algorithm Based on LSWT-NSST, IEEE Access 8(2020) 179857-179880.
[28] L.H. Jian, X.M. Yang, Z.L. Zhou, K. Zhou, K. Liu, Multi-scale image fusion through rolling guidance filter, Future Generations Computer Systems 83(2018) 310-325.
[29] C. Zhao, Y.D. Huang, Infrared and visible image fusion method based on rolling guidance filter and NSST, International Journal of Wavelets, Multiresolution and Information Processing 17(6)(2019) 1950045-1-1950045-31.
[30] Z. Liang, J. Xu, D. Zhang, Z. Cao, L. Zhang, A Hybrid l_0l_1 Layer Decomposition model for Tone Mapping, in: Proc. 2018 Conference on Computer Vision and Pattern Recognition, 2018.
[31] L.H. Chen, X.M. Yang, L. Lu, K. Liu, G. Jeon, W. Wu, An image fusion algorithm of infrared and visible imaging sensors for cyber-physical systems, Journal of intelligent & fuzzy systems: Applications in Engineering and Technology 36(5)(2019) 4277-4291.
[32] Z.Q. Zhou, B. Wang, S. Li, M.J. Dong, Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters, Information fusion 30(2016) 15-26.
[33] Y. Liu, X. Chen, J. Cheng, H. Peng, Z.F. Wang, Infrared and visible image fusion with convolutional neural networks, International Journal of Wavelets, Multiresolution and Information Processing 16(3)(2018) 1850018-1-1850018--20.
[34] S.T. Li, X.D. Kang, J.W. Hu, Image fusion with guided filtering, IEEE Transactions on Image Processing 22(7) (2013) 2864-2875.
[35] J.L. Ma, Z.Q. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, Infrared Physics & Technology 82(2017) 8-17.
[36] J.Y. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, Information Fusion 31(2016) 100-109.
[37] Z.Q. Zhu, M.G. Zheng, G.Q. Qi, D. Wang, Y. Xiang, A Phase Congruency and Local Laplacian Energy Based Multi-Modality Medical Image Fusion Method in NSCT Domain, IEEE Access 7(2019) 20811-20824 .
[38] D.P. Bavirisetti, V. Kollu, X. Gang, R. Dhuli, Fusion of MRI and CT images using guided image filter and image statistics, International Journal of Imaging Systems and Technology 27(3)(2017) 227-237.
[39] Y. Liu, X. Chen, H. Peng, Z.F. Wang, Multi-focus image fusion with a deep convolutional neural network, Information Fusion 36(2017) 191-207.

[40] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Medical Image Fusion via Convolutional Sparsity based Morphological Component Analysis, IEEE Signal Processing Letters 26(3)(2019) 485-489.
[41] R. Vijayarajan, S. Muttan, Discrete wavelet transform based principal component averaging fusion for medical images, AEU- International Journal of Electronics & Communications 69(6)(2015) 896-902.
[42] V. Aslantas, E. Bendes, A new image quality metric for image fusion: The sum of the correlations of differences, International Journal of Electronics and Communications 69(12)(2015) 1890-1896.