

Received May 25, 2021, accepted June 12, 2021, date of publication June 18, 2021, date of current version July 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3090436

Improving the Performance of Infrared and Visible Image Fusion Based on Latent Low-Rank Representation Nested With Rolling Guided Image Filtering

CE GAO¹, CONGCONG SONG, YANCHAO ZHANG, DONGHAO QI¹, AND YI YU¹

Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

Corresponding author: Ce Gao (50616636@qq.com)

This work was supported in part by the General Program of National Natural Science Foundation of China under Grant 51675506, and in part by the National Science and Technology Major Project under Grant 2017ZX10304403.

ABSTRACT The fusion quality of infrared and visible image is very important for subsequent human understanding of image information and target processing. The fusion quality of the existing infrared and visible image fusion methods still has room for improvement in terms of image contrast, sharpness and richness of detailed information. To obtain better fusion performance, an infrared and visible image fusion algorithm based on latent low-rank representation (LatLRR) nested with rolling guided image filtering (RGIF) is proposed that is a novel solution that integrates two-level decomposition and three-layer fusion. First, infrared and visible images are decomposed using LatLRR to obtain the low-rank sublayers, saliency sublayers, and sparse noise sublayers. Then, RGIF is used to perform further multiscale decomposition of the low-rank sublayers to extract multiple detail layers, which are fused using convolutional neural network (CNN)-based fusion rules to obtain the detail-enhanced layer. Next, an algorithm based on improved visual saliency mapping with weighted guided image filtering (IVSM-GIF) is used to fuse the low-rank sublayers, and an algorithm for adaptive weighting of regional energy features based on Laplacian pyramid decomposition is used to fuse the saliency sublayers. Finally, the fused low-rank sublayer, saliency sublayer, and detail-enhanced layer are used to reconstruct the final image. The experimental results show that the proposed method outperforms other state-of-the-art fusion methods in terms of visual quality and objective evaluation, achieving the highest average values in six objective evaluation metrics.

INDEX TERMS Image fusion, rolling guided image filtering, latent low-rank representation, detail-enhanced layer.

I. INTRODUCTION

Since richer, more comprehensive scene information cannot be obtained using a single image sensor, which leads to certain limitations, multiple sensors tend to be used to capture images for image fusion. An algorithm is employed to extract and integrate the effective information from multiple images captured from the same scene at the same moment for multidirectional, multiangle fusion to obtain good visual effects and rich detailed information. Therefore, image fusion research has received much attention in a variety of fields, such as military applications [1], [2], computer vision [3], [4],

remote sensing detection [5], [6], and medical imaging [7], with very broad and important application value. In particular, infrared and visible image fusion is a focus in multisensor image fusion research because it is the foundation of image fusion research.

The mechanism of infrared images is different from that of visible images, and the two types of images have different characteristics. Specifically, visible images have high spatial resolution and rich background information, and they are suitable for human visual perception. However, they are susceptible to poor lighting, smoke, and adverse weather conditions. Infrared images, depending on the detectors, are able to perceive thermal radiation at different wavelengths and have strong night vision and fog penetration capabilities;

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Yen Hsu¹.

however, they suffer from low resolution. The complementary information of these two types of images can be used to compensate for the deficiencies in the respective imaging techniques, making the fused images more informative and understandable. At present, image fusion algorithms are generally categorized into spatial domain-based fusion, multiscale transform (MST)-based fusion, edge preserving filter-based fusion, sparse representation (SR)-based fusion, neural network (NN)-based fusion, and other methods.

In recent years, to improve the fusion quality and obtain clearer scene information and better target features, researchers have devoted themselves to the research of new infrared and visible image fusion methods, and a variety of infrared and visible image fusion methods have been proposed [8]–[19]. To overcome the shortcomings of methods based on MST and SR, Li *et al.* [8] proposed an image fusion framework integrating nonsubsampling contour transformation (NSCT) into sparse representation. In this method, NSCT is used for source image decomposition. The high-pass coefficients are fused by using the sum modified Laplacian (SML). A principal component analysis dictionary learning algorithm is applied to the low-pass subband to reduce the dimensionality of the learning dictionary as well as the computational cost. The low-pass coefficients are fused using the SR-based method. The final fusion image is obtained by NSCT inverse transform. Huang *et al.* [9] proposed an infrared and visible image information fusion method based on phase congruency and image entropy. In this method, NSCT is used to decompose the source image to obtain the corresponding high- and low-frequency subbands. For the top decomposition layer of the high-frequency subband, the PCNN model [10] is used to measure the activity level of the high-frequency coefficients. For other high-frequency decomposition layers, the absolute maximum rule is used to measure the activity level of the absolute value of each high-frequency coefficient. For the low-frequency subband, according to activity measures, such as the phase consistency (PC), the local measure of sharpness change (LSCM), and the local signal strength (LSS), the low-frequency fusion rule is formulated. Nie *et al.* [11] proposed a novel total variation (TV) with joint norm-based infrared and visible images. A weighted fidelity term is employed to fuse both the infrared objects in the infrared image and the salient scenes in the visible image. In the process, a weight estimation method is developed based on the global luminance contrast-based saliency, and the $l_{2,1,rc}$ -norm is employed to produce structural group sparseness. To fuse more details from the visible image, a gradient sparseness constraint is further introduced, resorting to the $l_{1/2}$ -norm; however, the ability to retain detailed information is limited. Li *et al.* [12] presented a novel fusion framework based on a deep residual network (ResNet) and zero-phase component analysis (ZCA). The pretrained model ResNet50 is employed as a feature extraction tool to extract deep features from source images. Then, the ZCA and l_1 -norm are used to normalize the deep features, and finally, the fused image is reconstructed by obtaining the

weight coefficients through a softmax operation. The fusion performance of this method is, to some extent, better than that of several existing fusion methods, but the visual effect of the fused image is not adequate. The DenseFuse [13] network requires a longer training time, the extracted features are inadequately comprehensive, and the texture precision is low in the fusion result of the FusionGAN-based fusion method [14]. Yang *et al.* [15] proposed an infrared and visible image fusion method based on a texture conditional generative adversarial network (TC-GAN) and a multiple decision map fusion strategy. TC-GAN is constructed to generate a combined texture map from the visible and infrared images to capture the gradient changes in the fused image. Then, the combined texture map is utilized as a guidance image of the adaptive guided filter (AGF) to filter the source image, and the decision map based on the combined texture map is obtained. Finally, a multiple decision map fusion strategy is proposed to reconstruct the fused image. This method can preserve the texture details of the source image while highlighting the infrared target information.

Latent low-rank representation (LatLRR) [17], an improved version of low-rank representation (LRR) [16], has also been introduced into the image fusion field. Li *et al.* [18] proposed a LatLRR-based fusion method for infrared and visible images that uses simple fusion strategies (weighted-average strategy and sum strategy) to fuse the low-rank parts and the saliency parts, resulting in insufficient information from the source image and poor visual perception. Subsequently, they [19] proposed a new image fusion framework based on latent low-rank representation of multilevel image decomposition (MDLatLRR). A projection matrix L is learned by LatLRR. The matrix is used to extract the detail and base parts of the input image at several representation levels. The detailed parts are reconstructed using a fusion strategy based on the kernel norm and a reshape operator. The base parts are reconstructed by averaging strategy fusion. The fusion performance of this method overly depends on the number of decomposition levels, and as the decomposition levels increase, the time complexity increases.

The existing fusion methods have improved fusion performance; however, problems such as loss of fine details and degradation of contrast still exist [14], [30], [32], [34]. Consequently, the fusion quality of the existing infrared and visible image fusion methods still has room for improvement in terms of image contrast, sharpness and richness of the detailed information. In this regard, the present study proposes a new infrared and visible image fusion method. LatLRR with denoising and local structure representation capabilities is used as an image decomposition tool, nested with RGIF as a means of image enhancement. Two improved algorithms are employed as fusion rules for sublayers to reconstruct images.

The contributions of the proposed algorithm are as follows:

(1) A novel solution of two-level decomposition and three-layer fusion is proposed. Specifically, the solution includes the two-level decomposition method of LatLRR nested RGIF

and the three-layer fusion ideas of low-rank sublayer fusion, saliency sublayer fusion and detail layer fusion.

(2) The detail layer information is extracted by using RGIF. The detail layers are fused by using convolutional neural network (CNN)-based fusion rules to obtain a detail-enhanced layer to address the low contrast and low sharpness of fused images.

(3) The low-rank sublayers are fused using an algorithm based on improved visual saliency mapping with weighted guided image filtering (IVSM-GIF), and the saliency sublayers are fused using an algorithm of adaptive weighting of regional energy features based on Laplacian pyramid decomposition. This preserves the detailed texture information of the source image and highlights important targets. The two fusion rules ensure good fusion performance.

The rest of this paper is organized as follows. In Section II, the basic principles of the LatLRR model and RGIF are introduced. Section III describes in detail the fusion methods for low-rank sublayers and saliency sublayers, acquisition of a detail-enhanced layer, and image reconstruction. The experimental results are analyzed and discussed in Section IV. Finally, the conclusions of this study are presented in Section V.

II. TECHNICAL BACKGROUND

A. LatLRR

In 2010, LRR theory was first proposed by Liu *et al.* as a method for exploring the spatial structure of data [16]. This method uses the observed data matrix itself as a dictionary and considers only the global structure of the image. Hence, its application is limited. To address this problem, these authors put forward LatLRR as an improved scheme based on LRR theory [17]. LatLRR uses the observed data matrix itself and the unobserved latent data to construct a dictionary, fully considering the aspects of global structure, local structure, and sparse noise of image data. The presence of noise in the image may introduce visual artifacts [20], which will have some negative effects on the fusion results. Many fusion algorithms are devoted to denoising before fusion. LatLRR is not only able to robustly extract salient features from corrupted data but also has strong robustness to noise and outliers. Thus, LatLRR is used as an image decomposition tool in our research. Specifically, the LatLRR problem can be formulated as the following optimization problem:

$$\begin{aligned} \min_{Z, L, E} \|Z\|_* + \|L\|_* + \lambda \|E\|_1 \\ \text{s.t.}, X = XZ + LX + E \end{aligned} \quad (1)$$

where $\lambda > 0$ is a noise-related equilibrium parameter; $\|\cdot\|_1$ is the l_1 -norm; $\|\cdot\|_*$ is the nuclear norm; X is the data matrix of the source image; Z and L are the low-rank coefficient matrix and the saliency coefficient matrix, respectively; and E represents the sparse noise. Equation 1 represents a type of minimization problem with the nuclear norm and the l_1 -norm. This is a convex optimization problem and can be solved using the augmented Lagrangian multiplier (ALM) method.

LatLRR effectively integrates image segmentation and feature extraction to eventually obtain the solved low-rank sublayer (XZ), the saliency sublayer (LX), and the sparse noise sublayer (E). The sparse noise sublayer is directly discarded in the proposed method, which is equivalent to pre-denoising of the source image, which helps to improve the quality of image fusion [20]. Figure 1 shows the decomposition results by taking a set of images as an example. The decomposition process is time consuming.

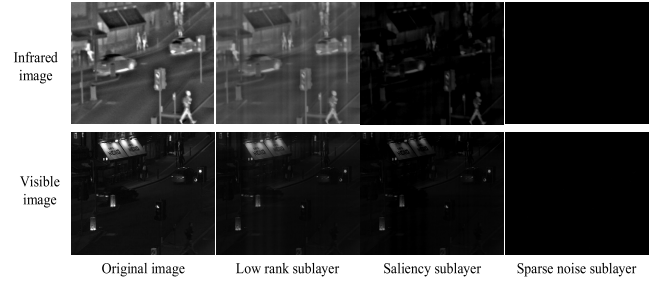


FIGURE 1. LatLRR decomposition operation.

B. RGIF

The concept of RGIF was first proposed by Zhang *et al.* [21]. RGIF performs iterative operations on the basis of an edge-preserving filter to control the smoothness of small structures under a scale measure and to recover edge structures through iterative processing. As a result, the dual effect of smoothing small structures and preserving edges is achieved. RGIF can preserve more image information and is more suitable for various types of image processing tasks. Therefore, in this study, the detail layers are extracted through multiscale decomposition by RGIF with scale-aware and edge-preserving features. They are used for the subsequent acquisition of the detail-enhanced layer. In general, RGIF can be divided into the following two main steps.

The first step is to smooth the image by using a Gaussian filter to remove small structures. The Gaussian filter for the input image with its center pixel at position n can be expressed as follows:

$$G_n = \frac{1}{K_m} \sum_{m \in S_n} \exp\left(-\frac{\|n - m\|}{2\sigma_s^2}\right) I_m \quad (2)$$

where $K_m = \sum_{m \in S_n} \exp(-\|n - m\|/2\sigma_s^2)$ is a normalization factor, m is a neighboring pixel of n , S_n is the set of neighboring pixels of pixel n , I_m is the grayscale value of pixel m in the input image, and σ_s is the standard deviation. The filter scale can be controlled by changing the σ_s value to achieve different smoothing results. The image output G from the Gaussian filter is used as the guidance image for the next step.

The second step is edge recovery through an iterative operation. In addition to the bilateral filter [22], which is an edge-preserving filter used in Reference [21], other well-known edge-preserving filters include guided image filtering (GIF) [23] and the weighted least squares filter [24].

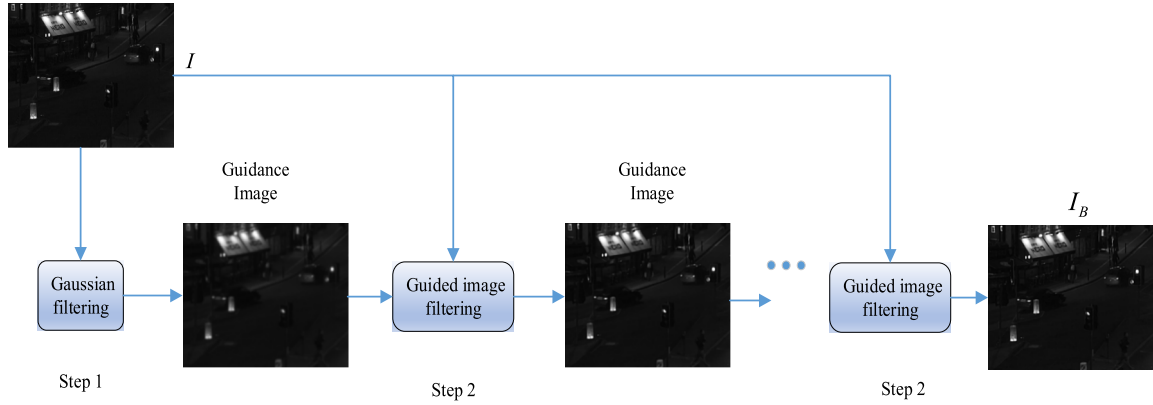


FIGURE 2. Illustration of the implementation process of RGIF.

Since GIF has a high computational efficiency, the second step of the present study is mainly based on GIF, which differs from the rolling guided filtering used in past studies.

The GIF process includes the guidance image, the input image, and the parameters that control the filtering performance. The i -th iteration can be expressed as follows:

$$G^i = GIF_{\sigma_s, \sigma_r^2}(G^{i-1}, I) \quad (i = 1, 2 \dots T) \quad (3)$$

where $GIF(\cdot)$ represents the GIF operator; G^i is the output image of the i -th iteration, which will be used as the guidance image of the next GIF; the initial guidance image G^0 is the output image of the Gaussian filter; I is the input image and remains unchanged during the iteration; σ_s is the radius of a square window, which is the same as Equation (2); σ_r is a regularization parameter, which controls the range weight; and T represents the number of iterations.

The first and second steps are combined to form the RGIF, which is defined as follows:

$$I_B = RGIF(I, \sigma_s, \sigma_r, T) \quad (4)$$

where $RGIF(\cdot)$ represents the RGIF operator and I_B is the base layer image filtered by RGIF. When $\sigma_s = 3$ and $\sigma_r = 0.001$, Figure 2 shows the entire iterative process taking a set of visible images as an example. After several iterations, the details of the image are smoothed, and the edges are restored. The detail layer image at this scale can be obtained by subtracting the image filtered by RGIF from the source image.

III. PROPOSED FUSION METHOD

A. FUSION FRAMEWORK

In this paper, we propose a new method for the fusion of infrared and visible images. The framework of the proposed method is shown in Figure 3. This method is based on LatLRR nested with RGIF. There are five steps, namely, image decomposition, acquisition of a detail-enhanced layer, fusion of low-rank sublayers, fusion of saliency sublayers, and image reconstruction.

Step 1: The infrared and visible images are decomposed using LatLRR, and the sparse noise sublayers are discarded. We obtain the low-rank sublayers (IR_{lrr} , VIS_{lrr}) and saliency sublayers (IR_{sal} , VIS_{sal}).

Step 2: Multiscale decomposition of IR_{lrr} , VIS_{lrr} is performed using the above RGIF to extract the information of multiple detail layers, the weight mapping matrix is obtained by extracting the deep features using the CNN-based model, and the detail-enhanced layer is obtained by fusing the detail layers through the weight mapping matrix.

Step 3: IR_{lrr} , VIS_{lrr} are fused using the IVSM-GIF algorithm. The visual saliency maps of IR_{lrr} , VIS_{lrr} are obtained and compared to generate the initial weight map values P_{ir} and P_{vis} . Then, optimizations of P_{ir} and P_{vis} are performed using the GIF to obtain the final weight map values w_{ir} and w_{vis} .

Step 4: IR_{sal} , VIS_{sal} are fused using an algorithm of adaptive weighting of regional energy features based on Laplacian pyramid decomposition. Pyramid decomposition of IR_{sal} , VIS_{sal} is carried out, and then, the fusion weight coefficients of the decomposed layers are assigned adaptively according to the local regional energy similarity.

Step 5: Reconstruction is performed with the obtained detail-enhanced layer and the fused low-rank sublayer and saliency sublayer.

B. ACQUISITION OF DETAIL-ENHANCED LAYER

The low-rank sublayers obtained from LatLRR decomposition still contain some details in addition to the global information and contrast information of the source image. To further ensure the integrity of the fused image information, the multiscale decomposition of the low-rank sublayers is performed using RGIF to extract more information prior to the fusion of low-rank sublayers. In the present study, we set the same regularization parameter σ_r , and images with different degrees of blurring are generated by changing the filter window radius value σ_s . That is, when $\sigma_s^j > \sigma_s^{j-1}$, the corresponding I_B^j is smoother and more blurred with less structural information than I_B^{j-1} , and the detail layer can

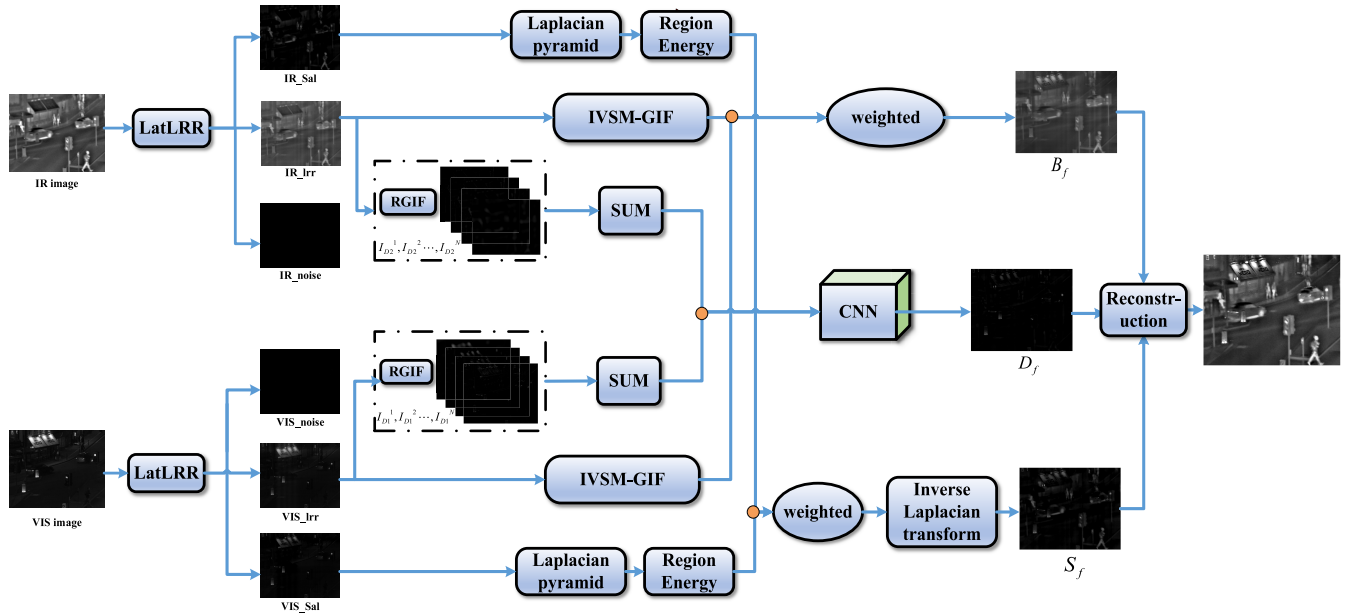


FIGURE 3. Schematic visualization of the proposed method.

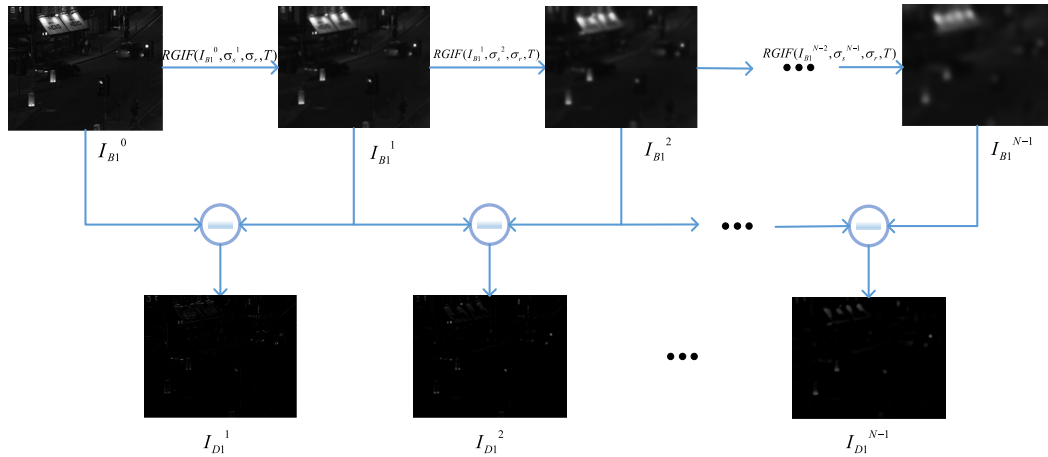


FIGURE 4. RGIF-based multiscale decomposition.

be obtained by the following equation. Assuming that the number of multiscale decomposition layers is N , the detail layers I_D^j at different scales are as follows:

$$I_D^j = I_B^{j-1} - I_B^j = I_B^{j-1} - RGIF(I_B^{j-1}, \sigma_s^j, \sigma_r, T) \quad (j = 1, 2, \dots, N) \quad (5)$$

where I_D^j is the detail layer image of the j th layer after decomposition and I_B^j is the base layer image of the j th layer after decomposition. In particular, I_B^0 is the source input image without filtering, namely, $I_{B1}^0 = VIS_lrr$, and $I_{B2}^0 = IR_lrr$. Thus, the detail layers $I_{D1(2)}^1, I_{D1(2)}^2, \dots, I_{D1(2)}^N$ are obtained through the above process of multiscale decomposition of VIS_lrr, IR_lrr . The process of RGIF-based multiscale decomposition is shown in Figure 4.

Subsequently, the extracted detail layers are summed separately as follows:

$$I_{D_VIS} = \sum_{i=1}^N I_{D1}^i \quad (6)$$

$$I_{D_IR} = \sum_{i=1}^N I_{D2}^i \quad (7)$$

As demonstrated in Reference [25], the CNN model can be used to achieve feature extraction and weight assignment in an “optimal” manner through network learning parameters, indicating the feasibility of the CNN for image fusion. Therefore, this network is employed in the present study to generate decision mapping containing the integrated pixel activity information from the input image. That is, the processed

images I_{D_IR} and I_{D_VIS} are input into the CNN model, the decision map values are obtained through deep feature extraction and used for image fusion. The network includes three convolutional layers (with a step size of 1 and a zero padding of 1 for the convolution process involved), one maximum pooling layer, and one fully connected layer [26]. The specific process for generating decision mapping D is described in Reference [26].

Finally, the detail-enhanced layer is obtained by the following pixel weighting strategy:

$$D_f = D(i, j)I_{D_IR}(i, j) + (1 - D(i, j))I_{D_VIS}(i, j) \quad (8)$$

C. FUSION OF LOW-RANK SUBLAYERS

It is known that the low-rank sublayer contains global information, contrast information, and other background information of the source image. The fusion strategy of the low-rank sublayers affects the overall appearance and contrast of the fused image to some extent. Therefore, the fusion of the low-rank sublayers should not be neglected. Simple weighted fusion is the most basic, easy to implement, and widely used fusion strategy. However, since this strategy is not fully applicable to any imaging scene, certain information of some infrared and visible images is lost to varying extents, leading to the degradation of the quality of the fused images. Therefore, fully considering the overall visual effect of the imaging, the IVSM-GIF algorithm is proposed in the present study for the fusion of low-rank sublayers. The VSM method [27] defines the saliency of each pixel by calculating the intensity contrast information of that pixel relative to other pixels in the image. Hence, the obvious visual structures and regions in the source image can be well identified, and better fusion of low-rank sublayers can be achieved. The method includes the following steps.

Step 1: The saliency value of each pixel is determined, the visual saliency maps of IR_lrr , VIS_lrr are obtained.

Suppose that pixel p in image I has an intensity of I_p ; then, the saliency value $V(p)$ of pixel p is as follows:

$$V(p) = \sum_{\forall q \in I} |I_p - I_q| \\ = |I_p - I_1| + |I_p - I_2| + \dots + |I_p - I_W| \quad (9)$$

where W is the total number of pixels in image I . Considering the presence of multiple pixels with the same intensity, for the simplicity of the algorithm, the above equation can be rewritten as follows:

$$V(p) = \sum_{t=0}^{L-1} K_t |I_p - I_t| \quad (10)$$

where t is the pixel intensity, K_t is the number of pixels with the same intensity of t , and L represents the number of gray levels in the image. In this paper, all the images used for testing are 8-bit images with a total of 256 gray levels, and thus, $L = 256$. The saliency values of other pixels are calculated successively, the normalized visual saliency maps V_{ir} , V_{vis} are ultimately obtained.

Step 2: Compare the above obtained saliency maps V_{ir} , V_{vis} pixel by pixel to obtain the initial weight mapping, which is defined as follows:

$$P_{ir}(i, j) = \begin{cases} 1 & \text{if } V_{ir}(i, j) > V_{vis}(i, j) \\ 0 & \text{otherwise} \end{cases} \\ P_{vis}(i, j) = 1 - P_{ir}(i, j) \quad (11)$$

where $P_{ir}(i, j)$ is the initial weight mapping at the position (i, j) in IR_lrr . This weight mapping is essentially a binary mapping and is implemented using the maximum selection strategy.

Step 3: Each initial weight mapping is optimized by introducing the GIF, and its corresponding low-rank sublayer as the guidance image, is used to obtain the final weight mapping w_{ir} and w_{vis} .

$$w_{ir} = GIF_{r, \varepsilon}(IR_lrr, P_{ir}) \\ w_{vis} = GIF_{r, \varepsilon}(VIS_lrr, P_{vis}) \quad (12)$$

where r is the filter window radius, and ε is the regularization parameter.

Step 4: Pixel-by-pixel weighting is performed using the final weight mapping, and the fused low-rank sublayer is obtained as follows:

$$B_f = w_{ir}IR_lrr(i, j) + w_{vis}VIS_lrr(i, j) \quad (13)$$

D. FUSION OF SALIENCY SUBLAYERS

To accurately integrate the important targets and features of the source image, an adaptive weighted fusion method of regional energy features based on Laplacian pyramid decomposition is adopted [28] for the fusion of IR_sal and VIS_sal with salient structures. To flexibly accomplish the fusion task, the method adaptively adjusts the fusion coefficients by comparing the similarity of the local regional energy.

First, IR_sal and VIS_sal are decomposed by the Laplacian pyramid as follows:

$$L_A^l = \text{laplacian}(IR_Sal) \\ L_B^l = \text{laplacian}(VIS_Sal) \quad (14)$$

where L_A^l and L_B^l denote the Laplacian pyramid decomposition of IR_sal and VIS_sal , respectively, and l represents the decomposition level.

Next, the local energy features centered at pixel (m, n) are obtained for the decomposition layers L_A^l and L_B^l , respectively, using the corresponding equations as follows:

$$E_A^l(m, n) = \sum_p \sum_q w \cdot L_A^l(m + p, n + q)^2 \\ E_B^l(m, n) = \sum_p \sum_q w \cdot L_B^l(m + p, n + q)^2 \quad (15)$$

where w is a region window of size 3×3 and p and q are the offsets of the neighboring pixels relative to the center pixel in w . In this paper,

$$w = \frac{1}{16} \times \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

Similar to the covariance of two local windows to define M_{AB} , it characterizes the similarity of the local energy features of L_A^l and L_B^l , i.e.,

$$M_{AB}^l(m, n) = \frac{(\sum_p \sum_q w \cdot L_A^l(m+p, n+q) L_B^l(m+p, n+q))^2}{E_A^l(m, n) \cdot E_B^l(m, n)} \quad (16)$$

In addition, a threshold th is set to determine the fusion coefficient. If $M_{AB}^l(m, n) < th$, the similarity of the local energy features of the point is considered low, in which case the energy of image layers at this point are simply compared, and the image layer with a higher energy at this point is selected, while the rest are discarded. If $M_{AB}^l(m, n) > th$, the similarity of the point is considered high, so weights are assigned according to the energy ratio.

The fusion rule is as follows:

$$L_F^l(m, n) = w_1 L_A^l(m, n) + w_2 L_B^l(m, n) \quad (17)$$

where w_1 and w_2 are the adaptive fusion coefficients, which are obtained by the following equation:

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{cases} [1, 0]^T & \text{if } M_{AB}^l < th \text{ \& } E_A^l > E_B^l \\ \left[\frac{E_A^l}{E_A^l + E_B^l}, \frac{E_B^l}{E_A^l + E_B^l} \right]^T & \text{if } M_{AB}^l > th \\ [0, 1]^T & \text{if } M_{AB}^l < th \text{ \& } E_A^l < E_B^l \end{cases} \quad (18)$$

where $[\cdot]^T$ is the matrix transpose operator. This method adaptively adjusts the fusion coefficients by comparing the similarity of the local region energies to ensure the best contribution to the fusion results. The $L_F^l(m, n)$ of each layer is determined by the above equation, the fused image S_f of the saliency sublayers is reconstructed by inverse Laplacian transform.

E. RECONSTRUCTION PROCESS

Finally, the final fused image is reconstructed from Equation (19) using the acquired detail-enhanced layer D_f , the fused low-rank sublayer B_f , and the saliency sublayer S_f to obtain the final fused image as follows:

$$F = B_f + D_f + S_f \quad (19)$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENT SETTINGS AND IMPLEMENTATION DETAILS

To verify the feasibility and superiority of the proposed fusion method, ten infrared and visible images in different scenes are selected from the publicly available TNO Image Fusion Dataset for the fusion performance test. The contribution of the detail-enhanced layer to our method

is first verified, then the fusion performance is compared between our proposed method and other state-of-the-art methods, including (1) convolutional sparse representation fusion (ConvSR) [29]; (2) gradient transfer fusion (GTF) [30]; (3) a fusion method based on weighted least squares (WLS) optimization [31]; (4) a fusion method that uses infrared feature extraction and visual information preservation (FEIP) [32]; (5) a fusion method based on multi-level Gaussian curvature filtering (GCF) [33]; (6) a fusion framework based on ResNet50 and zero-phase component analysis (ResNet50) [12]; (7) Bayesian fusion (Bayesian) [34]; and (8) a fusion framework based on LatLRR of multi-level image decomposition (MDLatLRR) [19]. All algorithms used for comparison with the proposed method are implemented using the code provided by the original authors, and the default parameter settings of the above methods reported in the respective publications are adopted.

The settings of all parameters in our study are based on the magnitudes provided by the relevant references [18], [21], [31], [35], [36] and empirical analysis. Test experiments within a reasonable range are also carried out. According to the objective evaluation indicators, the corresponding results are compared and analyzed. To achieve a trade-off between performance and processing time, a more appropriate parameter value is selected. The parameters are set as follows: the regularization parameter for LatLRR $\lambda = 0.6$; the number of decomposition levels for RGIF $N = 4$; all regularization parameters σ_r are set to 0.02; and the parameter $\sigma_s = 3$ initially and then doubles successively (i.e., 6, 12, ...). As a result, we can obtain four detail layers. The parameters r and ε for the GIF in the IVSM-GIF algorithm are 45 and 0.1, respectively. The number of pyramid decomposition layers is set to 8, and the threshold th is set to 0.6 in the fusion method of the saliency sublayers. The experimental simulation is implemented on a platform with an Intel Core i7-10510U CPU, 8 GB RAM, and a 64-bit Windows 10 system. The programming environment is MATLAB 2014b.

B. SUBJECTIVE PERFORMANCE EVALUATION

It is straightforward to subjectively evaluate the image fusion quality. Therefore, the fusion performance with and without the detail-enhanced layer D_f are subjectively compared. Four images are selected for verification. The fusion results with and without D_f are shown in Figure 5.

By comparing the overall visual effects of the enlarged trash bin in Figures 5(b) and 5(B), the enlarged soldiers in Figures 5(c) and 5(C), and the enlarged windows in Figures 5(d) and 5(D), it can be seen that the fusion method with D_f highlights the edges and texture of the objects and significantly improves the overall contrast of the images, making the images more visible.

Next, we analyze the fusion results of the proposed method and the compared methods. A comparison of the fusion results from different methods is shown in Figure 6, where the infrared images are in row (a), the visible images are in row (b), the fused images by other methods are in rows

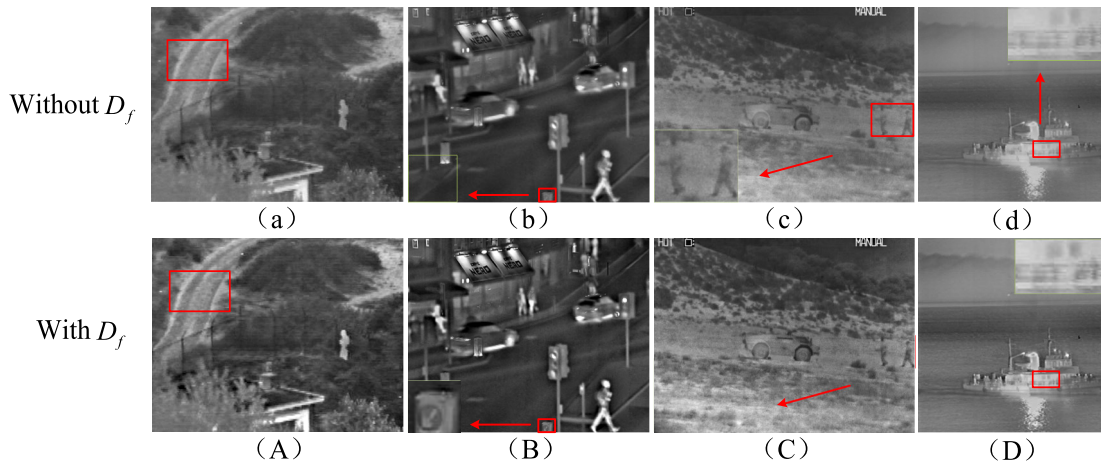


FIGURE 5. Fusion results with D_f and without D_f .

(c) to (m), and the fusion images by the proposed method are in row (n). An overall comparison of the fusion results reveals the following. The fused image from method (d) is blurred, and detailed information is lost, resulting in low visibility. The infrared target of the fused image from method (f) is significant, but some of the visible image details are lost; e.g., in the third group of images, the windows of the ship are missing and in the 10th group of images, the soldiers' eyes are missing. The contrast of the fusion image for method (h) is low, and the visual perception is not very good. The fused image of method (i) is visually closer to the visible image and has rich texture details, but the infrared target features are not prominent. Methods (j) to (m) correspond to the fusion results of the different decomposition levels (between 1 and 4) of the MDLatLRR method. The results show that with the increasing number of decomposition levels, the salient features are more enhanced and the detailed information becomes clearer; however, the artifacts become more obvious. In contrast, the fused images obtained by methods (e), (g) and the proposed method have better visual quality.

To fully verify the above statement, we selected representative images “Kaptein” and “marching” for detailed analysis. The fusion results for the “Kaptein” image are shown in Figure 7, where the canopy/shrubs and person are marked with red and blue boxes, respectively, and enlarged in the analysis. The data in Figure 7 show the following. There are some white shadows at the edges of the canopy contour fused by method (c). The fused target by method (d) is more obvious, but the other information is too blurred; e.g., the boundary of the image contour is not clear, and the texture details are not well preserved. In addition, the sky background of the fused image is dark, which affects the error in information acquisition. Method (e) has obvious noise around the tree canopy. In addition, the fused images of methods (h), (i), and (j) have significantly less contrast information than that of the proposed method and have poor visual effects. The fused images of methods (k), (l), and (m) have clear shrub texture and prominent infrared features, but obvious artifacts

appear on the edges of the canopy contours and the figure contours. However, the proposed method exhibits sharper shrubs and more distinct lighting at the glass door, showing good visibility.

The fusion results for the “marching” image are shown in Figure 8. The vehicle/soldiers and trees in the background are marked with red and blue boxes, respectively, and are enlarged. The data in Figure 8 show the following. Methods (e), (f), (g), (h) and (j) lead to the loss of the tire texture details to varying degrees. The contours of the soldiers by methods (f), (h), and (j) are blurred. In addition, a comparison of the fusion results of the trees on the hillside shows that the upper part of the trees fused by method (c) almost blends with the night sky background of the image, making it difficult to determine the edges of the trees. The trees fused by methods (d), (f), (h), (i) and (j) exhibit boundaries with the background of the image, whereas they do not exhibit an advantage in clearly extracting texture edges between trees. The trees fused by methods (e), (l), and (m) and the proposed method have a sharp texture, with the intertwining of tree branches clearly visible. In summary, the proposed method can effectively preserve texture detail information and has good performance in overall image sharpness and contrast.

C. OBJECTIVE PERFORMANCE EVALUATION

Due to differences in visual sensitivity, it is not possible to compare the advantages and disadvantages of fusion methods without bias. To evaluate the quality of fused images more comprehensively and convincingly, the fusion results of different methods are evaluated objectively and comprehensively. Different types of evaluation metrics are selected from the objective evaluation methods to quantitatively evaluate the fusion results of different methods. These metrics are based on information theory, structural similarity, image features, and human visual perception, and include information entropy (EN) [37], mutual information (MI) [38], multiscale structural similarity (MS-SSIM) [39], standard deviation (SD) [40], average gradient (AG) [41], edge-based

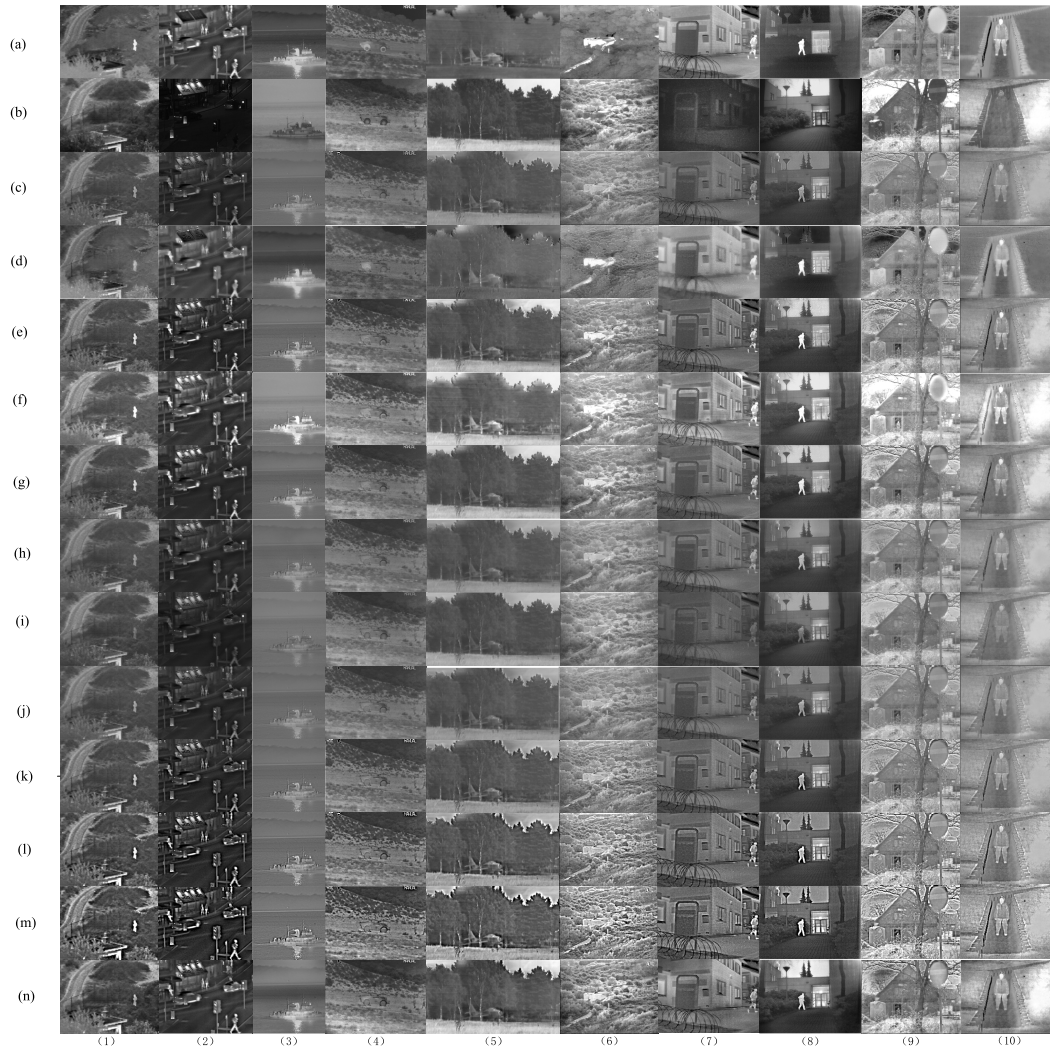


FIGURE 6. Comparison of fusion results from different methods on ten infrared and visible image pairs. (a) Infrared images; (b) Visible images; (c) ConvSR; (d) GTF; (e) WLS; (f) FEIP; (g) GCF; (h) ResNet50; (i) Bayesian; (j) MDLatLRR_level1; (k) MDLatLRR_level2; (l) MDLatLRR_level3; (m) MDLatLRR_level4 and (n) the proposed method.

similarity (Q_{abf}) [42], sum of the correlations of differences (SCD) [43], and visual information fidelity (VIF) [44].

The changes in each metric for the cases with and without the detail-enhanced layer D_f are analyzed in detail using objective evaluation methods. The results of the objective evaluation of Figure 5 are shown in Table 1. As shown in Table 1, most of the objective evaluation metrics for the fusion methods with detail-enhanced layer D_f tend to be higher than those without D_f . To measure the changes, we calculate the average values of the differences between the metric values of the cases with and without D_f , as listed in the last column of Table 1. These average values show that the metrics of AG, MI, SD, and VIF increase significantly, indicating that the superposition of the detail-enhanced layer makes the fused image have rich gradient information (e.g., detail and texture) and remarkable contrast information, improving the visual effect. This is consistent with the

evaluation by the subjective method, thereby confirming the contribution of the detail-enhanced layer.

Next, the proposed method and other methods are objectively evaluated and analyzed. Quantitative comparisons of the eight objective evaluation metrics using ten different fusion methods on the ten infrared and visible image pairs are shown in Figure 9.

The higher the values of these eight metrics are, the better the quality of the fused images. As seen from the summary of the various metrics in Figure 9, the proposed method is one of the best in terms of metrics EN, MI, MS_SSIM, Q_{abf} , SCD, and SD, indicating that the proposed method can preserve rich and effective information, provides high contrast, and produces a good overall structural similarity between the fused image and the source image. The proposed method is above average in terms of metrics AG and VIF, indicating that the proposed method can preserve most of the

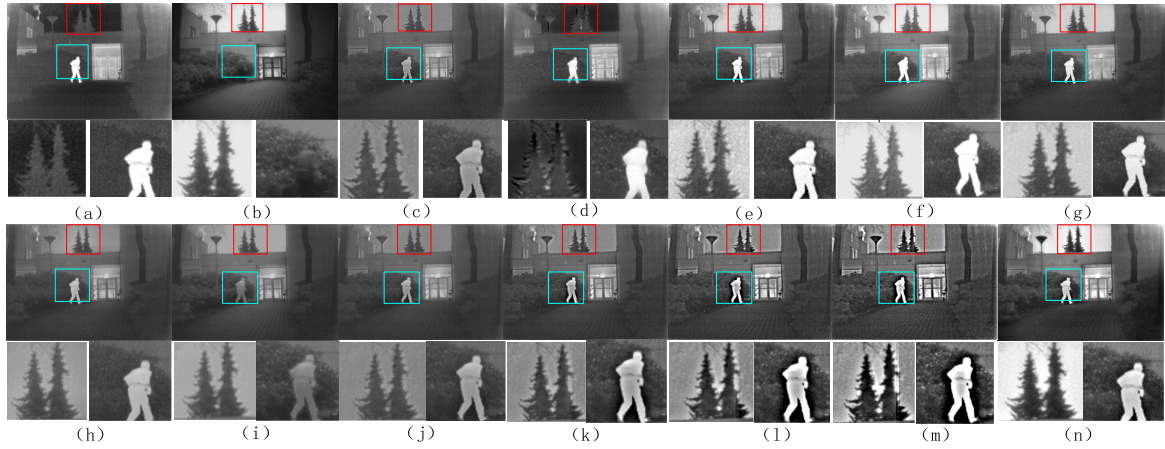


FIGURE 7. Comparison of fusion results for the source image “Kaptein” using different methods: (a) Infrared images; (b) Visible images; (c) ConvSR; (d) GTF; (e) WLS; (f) FEIP; (g) GCF; (h) ResNet50; (i) Bayesian; (j) MDLatLRR_level1; (k) MDLatLRR_level2; (l) MDLatLRR_level3; (m) MDLatLRR_level4 and (n) the proposed method.

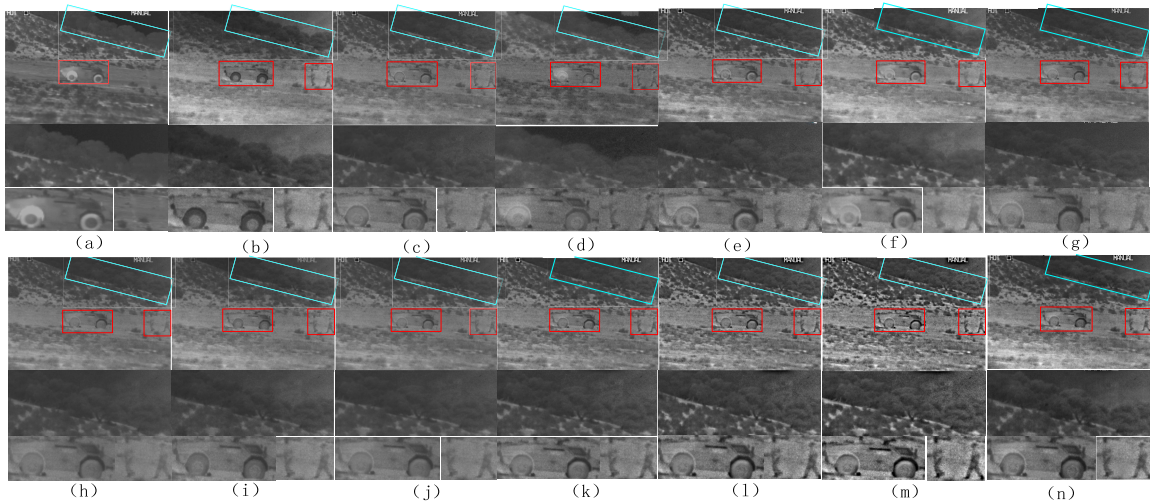


FIGURE 8. Comparison of fusion results of the “marching” source image using different methods: (a) Infrared images; (b) Visible images; (c) ConvSR; (d) GTF; (e) WLS; (f) FEIP; (g) GCF; (h) ResNet50; (i) Bayesian; (j) MDLatLRR_level1; (k) MDLatLRR_level2; (l) MDLatLRR_level3; (m) MDLatLRR_level4 and (n) the proposed method.

TABLE 1. Results of the objective evaluation of cases with and without D_f , respectively.

Images Metrics	(a)	(A)	(b)	(B)	(c)	(C)	(d)	(D)	Average change
AG	3.563	5.178	3.030	3.907	3.474	4.728	1.577	2.172	1.085
EN	6.868	7.061	6.610	6.862	7.135	7.239	6.443	6.535	0.160
MI	13.737	14.122	13.221	13.724	14.271	14.478	12.886	13.069	0.320
MS_SSIM	0.924	0.928	0.973	0.972	0.942	0.945	0.917	0.948	0.009
$\bar{Q}_{abf}$	0.461	0.473	0.636	0.633	0.545	0.545	0.495	0.531	0.011
SCD	1.672	1.769	1.666	1.711	1.448	1.621	1.786	1.798	0.082
SD	33.003	37.778	34.359	41.113	37.688	40.810	23.827	25.931	4.189
VIF	0.469	0.757	0.609	0.969	0.6053	1.1502	0.413	0.6656	0.361

gradient information (e.g., detail and texture) well and has good performance in fidelity of visual information.

To accurately give the metric values of the proposed method, we obtain the average metric values of the ten image

pairs for each fusion method to comprehensively evaluate the advantages and disadvantages of the proposed method. The average values of the objective evaluation metrics for different methods are shown in Table 2. In Table 2, each

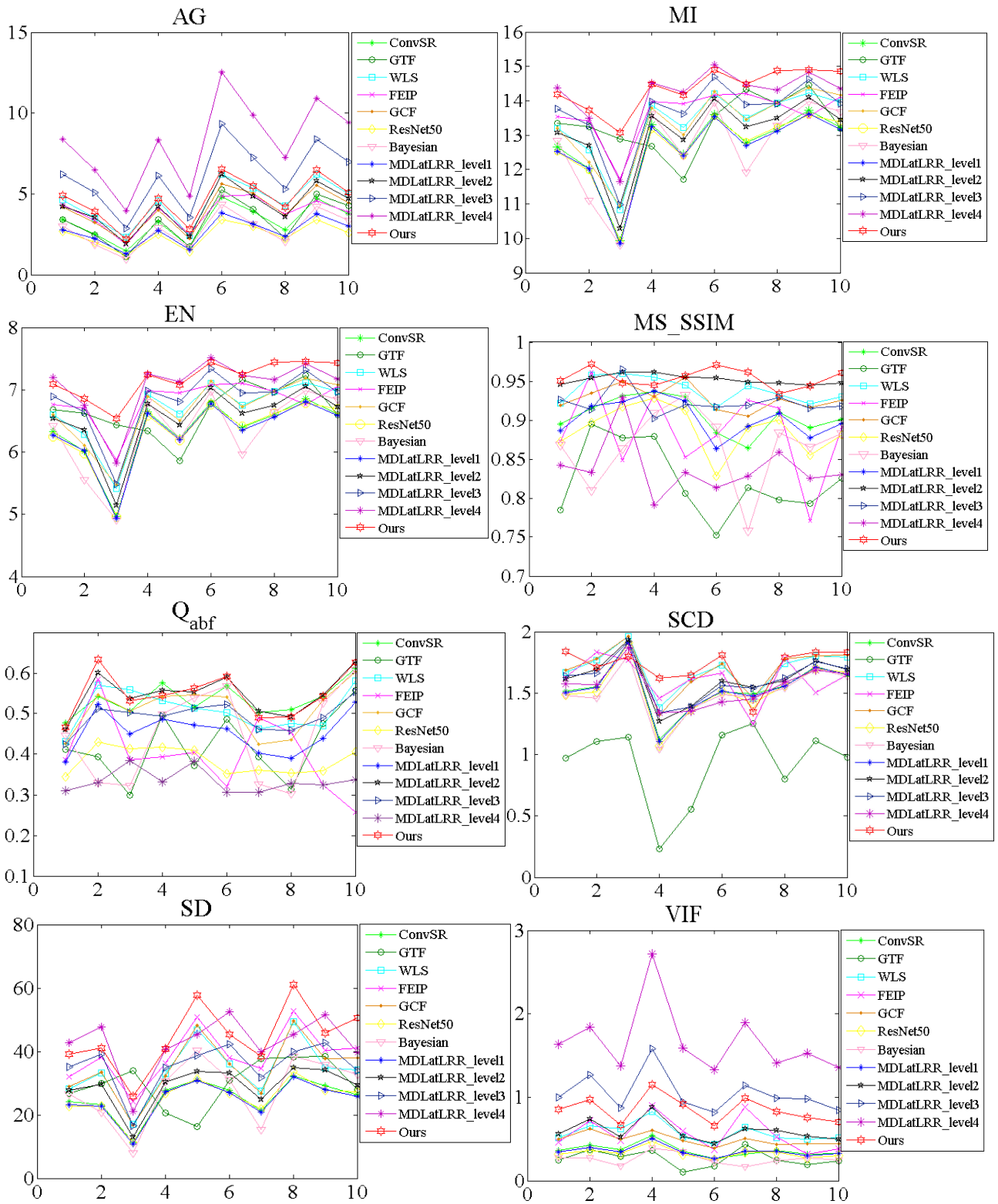


FIGURE 9. Quantitative comparison of eight objective evaluation metrics using ten different fusion methods on the ten infrared and visible image pairs.

column corresponds to an objective evaluation metric, and the values in red and in green are the highest values and the second highest values of the average values of that metric,

respectively. The average values for the proposed method are the highest with respect to six metrics (EN, MI, MS_SSIM, Q_{abf} , SCD, and SD) and the third highest with respect to

TABLE 2. Average values of objective evaluation metrics for different methods.

Metrics Methods	AG	EN	MI	MS_SSIM	Q_{abf}	SCD	SD	VIF
ConvSR	3.245	6.344	12.688	0.906	0.535	1.54	25.441	0.364
GTF	3.292	6.667	13.333	0.822	0.42	0.929	30.079	0.266
WLS	4.448	6.675	13.351	0.938	0.51	1.693	34.01	0.575
FEIP	3.812	6.826	13.652	0.885	0.4	1.616	38.865	0.564
GCF	4.002	6.668	13.336	0.928	0.511	1.693	34.791	0.49
RestNet50	2.454	6.326	12.652	0.888	0.385	1.52	25.244	0.331
Bayesian	2.761	6.331	12.661	0.866	0.447	1.499	27.961	0.26
MDLatLRR_level1	2.673	6.311	12.623	0.903	0.454	1.536	24.861	0.352
MDLatLRR_level2	4.142	6.543	13.087	0.952	0.547	1.612	29.132	0.594
MDLatLRR_level3	6.115	6.833	13.666	0.923	0.493	1.613	35.517	1.043
MDLatLRR_level4	8.198	7.065	14.129	0.838	0.334	1.55	42.634	1.668
Ours	4.633	7.182	14.363	0.953	0.548	1.720	44.537	0.844

TABLE 3. Average time consumption of different methods when fusing two images of size 256×256 pixels.

Methods	ConvSR	GTF	WLS	FEIP	GCF	RestNet50	Bayesian	MDLatLRR_level1	MDLatLRR_level2	MDLatLRR_level3	MDLatLRR_level4	Ours
Time(s)	40.156	0.495	0.493	0.052	116.514	1.071	0.204	7.697	18.758	29.371	40.131	36.094

two metrics (AG and VIF). Among them, the highest value and the second highest value of AG and VIF appear in the decomposition three layers and the decomposition four layers in the MDLatLRR method, respectively. As we know from the earlier analysis, an increase in the number of decomposition levels results in enhanced salient features. The detailed information becomes clearer, but the artifacts become more obvious. As the number of decomposition levels increases to three levels and four levels, the MS_SSIM, Q_{abf} , and SCD indicators decrease, among which MS_SSIM and Q_{abf} have reached poor values. The proposed method performs remarkably well in terms of the average values of the eight objective metrics.

In summary, the objective evaluation results are generally consistent with the subjective evaluation results, and the proposed method has advantages in the richness of information, sharpness and fidelity of images, showing good fusion performance.

D. COMPUTATIONAL EFFICIENCY ANALYSIS

The computational efficiency of the algorithm is generally evaluated by the time complexity. Therefore, we calculated the average running time of different methods when fusing image pairs of size 256×256 pixels with 10 fusion operations. All experiments were performed under the same conditions. The results of the running time are shown in Table 3.

According to Table 3, the GTF, FEIP, RestNet50 and Bayesian methods show better calculation efficiency than

other fusion methods, but the fusion performance is not the most satisfactory. The running times of GCF, ConvSR, MDLatLRR and the proposed method are relatively long. Among them, the GCF method is the most time consuming. The ConvSR method is based on the improvement in the sparse representation algorithm; thus, the algorithm efficiency is relatively low. The running time of the MDLatLRR method is closely related to the decomposition level. As the number of decomposition levels increases, the running time of the algorithm becomes increasingly longer. The proposed method is based on latent low-rank decomposition. It is acceptable that the proposed method produces superior fused images at the cost of more running time than other fusion methods.

V. CONCLUSION

In this paper, an infrared and visible image fusion method based on LatLRR nested with RGIF is proposed. The method uses LatLRR with a denoising capability as an image decomposition tool to obtain low-rank sublayers and saliency sublayers. The low-rank sublayers are decomposed at multiple scales using RGIF to acquire the detail-enhanced layer. The IVSM-GIF algorithm and an algorithm for adaptive weighting of regional energy features based on Laplacian pyramid decomposition are used to fuse the low-rank sublayers and the saliency sublayers. The fusion performance of the proposed method is comprehensively evaluated using both a subjective evaluation method and an objective evaluation method containing eight evaluation metrics. The contribution

of the detail-enhanced layer to the proposed method is verified first, followed by a comparative analysis with other state-of-the-art methods. The experimental results show that the proposed method outperforms other representative fusion methods in terms of visual quality and objective evaluation, as demonstrated by achieving the highest average values in six objective evaluation metrics (EN, MI, MS_SSIM, Q_{abf}, SCD, SD). The fused image obtained by the proposed method not only preserves the important information of the source image but also has high sharpness.

REFERENCES

- [1] A. C. Müller and S. Narayanan, "Cognitively-engineered multisensor image fusion for military applications," *Inf. Fusion*, vol. 10, no. 2, pp. 137–149, Apr. 2009.
- [2] Z. J. Feng, X. L. Zhang, L. Y. Yuan, and J. N. Wang, "Infrared target detection and location for visual surveillance using fusion scheme of visible and infrared images," *Math. Problems Eng.*, vol. 3, pp. 831–842, Aug. 2013.
- [3] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, Mar. 2015.
- [4] R. Singh, M. Vatsa, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition," *Pattern Recognit.*, vol. 41, no. 3, pp. 880–893, Mar. 2008.
- [5] L. Han, B. Wulie, Y. Yang, and H. Wang, "Direct fusion of geostationary meteorological satellite visible and infrared images based on thermal physical properties," *Sensors*, vol. 15, no. 1, pp. 703–714, Jan. 2015.
- [6] X. Chang, L. Jiao, F. Liu, and F. Xin, "Multicontourlet-based adaptive fusion of infrared and visible remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 549–553, Jul. 2010.
- [7] X. Liu, W. Mei, and H. Du, "Structure tensor and nonsubsampling shearlet transform based algorithm for CT and MRI image fusion," *Neurocomputing*, vol. 235, pp. 131–139, Apr. 2017.
- [8] Y. Li, Y. Sun, X. Huang, G. Qi, M. Zheng, and Z. Zhu, "An image fusion method based on sparse representation and sum modified-Laplacian in NSCT domain," *Entropy*, vol. 20, no. 7, p. 522, Jul. 2018.
- [9] X. Huang, G. Qi, H. Wei, Y. Chai, and J. Sim, "A novel infrared and visible image information fusion method based on phase congruency and image entropy," *Entropy*, vol. 21, no. 12, p. 1135, Nov. 2019.
- [10] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampling shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.
- [11] R. Nie, C. Ma, J. Cao, H. Ding, and D. Zhou, "A total variation with joint norms for infrared and visible image fusion," *IEEE Trans. Multimedia*, early access, 12, 2021, doi: [10.1109/TMM.2021.3065496](https://doi.org/10.1109/TMM.2021.3065496).
- [12] H. Li, X.-J. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infr. Phys. Technol.*, vol. 102, Nov. 2019, Art. no. 103039.
- [13] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019.
- [14] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.
- [15] Y. Yang, J. Liu, S. Huang, W. Wan, W. Wen, and J. Guan, "Infrared and visible image fusion via texture conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jan. 26, 2021, doi: [10.1109/TCSVT.2021.3054584](https://doi.org/10.1109/TCSVT.2021.3054584).
- [16] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. ICML*, vol. 1, Aug. 2010, p. 8.
- [17] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1615–1622.
- [18] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, *arXiv:1804.08992*. [Online]. Available: <http://arxiv.org/abs/1804.08992>
- [19] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: A novel decomposition method for infrared and visible image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4733–4746, Nov. 2020.
- [20] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jan. 2017.
- [21] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2014, pp. 815–830.
- [22] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, Feb. 1998, pp. 839–846.
- [23] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [24] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.
- [25] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 3, May 2018, Art. no. 1850018.
- [26] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.
- [27] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th Annu. ACM Int. Conf. Multimedia (MULTIMEDIA)*, 2006, pp. 815–824.
- [28] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *Readings Comput. Vis.*, vol. 31, no. 4, pp. 671–679, 1987.
- [29] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.
- [30] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, Sep. 2016.
- [31] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infr. Phys. Technol.*, vol. 82, pp. 8–17, May 2017.
- [32] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visible image fusion through infrared feature extraction and visual information preservation," *Infr. Phys. Technol.*, vol. 83, pp. 227–237, Jun. 2017.
- [33] W. Tan, H. Zhou, J. Song, H. Li, Y. Yu, and J. Du, "Infrared and visible image perceptual fusion through multi-level Gaussian curvature filtering image decomposition," *Appl. Opt.*, vol. 58, no. 12, pp. 3064–3073, Apr. 2019.
- [34] Z. Zhao, S. Xu, C. Zhang, J. Liu, and J. Zhang, "Bayesian fusion for infrared and visible images," *Signal Process.*, vol. 177, Dec. 2020, Art. no. 107734.
- [35] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Inf. Fusion*, vol. 12, no. 2, pp. 74–84, Apr. 2011.
- [36] L. Jian, X. Yang, Z. Zhou, K. Zhou, and K. Liu, "Multi-scale image fusion through rolling guidance filter," *Future Gener. Comput. Syst.*, vol. 83, pp. 310–325, Jun. 2018.
- [37] J. Van Aardt, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, May 2008, Art. no. 023522.
- [38] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'information measure for performance of image fusion,'" *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.
- [39] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [40] Y.-J. Rao, "In-fibre Bragg grating sensors," *Meas. Sci. Technol.*, vol. 8, no. 4, p. 355, 1997.
- [41] H. Jin and Y. Wang, "A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization," *Infr. Phys. Technol.*, vol. 64, pp. 134–142, May 2014.
- [42] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.
- [43] V. Aslantas and E. Bendes, "A new image quality metric for image fusion: The sum of the correlations of differences," *AEU Int. J. Electron. Commun.*, vol. 69, no. 12, pp. 1890–1896, Dec. 2015.
- [44] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion*, vol. 14, no. 2, pp. 127–135, Apr. 2013.



CE GAO was born in 1981. He received the bachelor's degree in computer science and technology from Tianjin University, in 2004, and the master's degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2011. He is currently an Associate Research Fellow. His research interests include computer applications, image processing, and optoelectronic measurement and control technology.



DONGHAO QI was born in 1994. He received the master's degree in mechanical and electronic engineering from Xidian University, in 2020. He is currently a Research Intern. His research interest includes technology related to image processing.



CONGCONG SONG was born in 1994. She received the bachelor's degree in communication engineering from the Changchun University of Science and Technology, in 2017, and the master's degree in electronics and communication engineering from Jilin University, in 2020. She is currently working with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. Her current research interests include image fusion and deep learning.



YANCHAO ZHANG was born in 1985. She received the bachelor's and master's degrees in engineering from the School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, in 2007 and 2009, respectively, and the Ph.D. degree in optical engineering from the University of Chinese Academy of Sciences, in 2015. She is currently an Associate Research Fellow. Her research interests include image processing and photoelectronic measurement and control technology.



YI YU was born in 1979. He received the Ph.D. degree in optical engineering from the Changchun Institute of Technology, Changchun, China, in 2016. He is currently a Professor with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun. His research interests include optoelectronic measurement, image/video processing, computer vision, and machine learning.

...