


Research Article

Infrared and Visible Image Fusion via Fast Approximate Bilateral Filter and Local Energy Characteristics

Zongping Li,¹ Wenxin Lei,² Xudong Li,¹ Tingting Liao,¹ and Jianming Zhang^{1,2} 

¹National Engineering Research Center of Sintering and Pelletizing Equipment System, Zhongye Changtian International Engineering Co., Ltd., Changsha 410205, China

²Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China

Correspondence should be addressed to Jianming Zhang; jmzhang@csust.edu.cn

Received 30 September 2021; Revised 8 November 2021; Accepted 20 November 2021; Published 7 December 2021

Academic Editor: Zhu Xiao

Copyright © 2021 Zongping Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image fusion is to effectively enhance the accuracy, stability, and comprehensiveness of information. Generally, infrared images lack enough background details to provide an accurate description of the target scene, while visible images are difficult to detect radiation under adverse conditions, such as low light. People hoped that the richness of image details can be improved by using effective fusion algorithms. In this paper, we propose an infrared and visible image fusion algorithm, aiming to overcome some common defects in the process of image fusion. Firstly, we use fast approximate bilateral filter to decompose the infrared image and visible image to obtain the small-scale layers, large-scale layer, and base layer. Then, the fused base layer is obtained based on local energy characteristics, which avoid information loss of traditional fusion rules. The fused small-scale layers are acquired by selecting the absolute maximum, and the fused large-scale layer is obtained by summation rule. Finally, the fused small-scale layers, large-scale layer, and base layer are merged to reconstruct the final fused image. Experimental results show that our method retains more detailed appearance information of the fused image and achieves good results in both qualitative and quantitative evaluations.

1. Introduction

With the popularity of infrared image applied in military surveillance [1], remote sensing [2], medical imaging [3], space exploration [4], and other fields, people pay more and more attention to the fusion of infrared and visible images. There are some differences in the characteristics and imaging regulation of infrared and visible images. Infrared image (IR) that captures thermal radiation has strong anti-interference capability and can operate all day without being affected by lighting conditions, but its contrast is low and its ability to distinguish details is poor. In contrast, the visible image (VI) that captures reflected light has high spatial discrimination and legible detail texture information, whereas visible image is greatly affected by weather and has poor imaging effect in cloudy, rainy, and night environments. Image fusion technology can make up for their respective shortcomings, enrich the

information of single source image, obtain sufficient and accurate target expression, and make the image more abundant.

The research of infrared and visible image fusion has important theoretical significance in many application scenarios of real life. The existing fusion methods are effective, but there are also some common problems, such as poor contrast, block effect, information distortion, and so on. To address these issues and get better fusion performance, a novel fusion algorithm based on fast approximate bilateral filter and local energy characteristics is developed. Our contributions are as follows:

- (1) We introduce a fast approximate bilateral filter to decompose the original infrared or visible image. Decompose the image into small-scale layers with rich details, large-scale layer with obvious edge features, and base layer with low-frequency information.

- (2) We perform fast approximate bilateral filter on the infrared or visible input image five times to obtain the image sequence that is increasingly coarsened. Then, the two images separated by one image in the sequence are subtracted to obtain multiple, small-scale layers. The large-scale layer is obtained by the difference between the fourth image and itself after Gaussian filter.
- (3) We put forward the fusion of the base layer by using local energy characteristics, fusion of the small-scale layers by using the absolute maximum rule, and fusion of the large-scale layer by using the summation rule, so as to retain more local structures and salient features in the fused image. It transmits most of the information from the source image to the composite image, and fusion loss is negligible.

For the purpose of verifying the effectiveness of our fusion algorithm, we compared it with those of 10 other infrared and visible fusion methods. Experimental results show that the image quality fused by our method is high and superior to similar methods on different IR and VI data sets. The organizational structure of this paper is as follows: Section 2 summarizes the related works of image fusion. Section 3 describes the proposed fusion algorithm in detail. Section 4 analyzes the experimental results of our algorithm compared with several classical algorithms. Finally, Section 5 gives the conclusion.

2. Related Works

In recent years, the advantages of image fusion have been highly valued by many researchers. Image fusion technology has developed rapidly in theory. From the weighted average method to color space, and then to multiscale transformation, image fusion technology has been developing rapidly. Generally, image fusion methods can be divided into spatial domain-based methods and frequency domain-based methods according to different space of the fusion process.

The fusion algorithm, which is based on spatial domain, refers to the calculation and processing of pixel gray value in the space composed of image pixels. Such methods mainly include principal component analysis [5], color space mapping method and pseudocolor image fusion method [6], gray or contrast modulation [7], and artificial neural network [8]. The image fusion algorithm, which is based on frequency domain, is to analyze and process the information characteristics of the transform coefficients converted to frequency domain through multiscale transform tools. In the mid-1980s, Burt and Adelson first proposed the Laplacian pyramid algorithm [9], which can represent more prominent feature information of images, containing Laplacian pyramid (LAP) [10], gradient pyramid [11], and contrast pyramid [12]. However, this method will lose image information and the decomposition process does not have directionality. In order to overcome this problem, Mallat proposed a multiresolution algorithm of wavelet transform [13], which can obtain not only low-frequency information, but also horizontal, vertical, and diagonal information of

high-frequency part. Compared with the traditional algorithm based on tower decomposition, the overall fusion performance of wavelet transform is better. However, due to the limited direction information obtained by wavelet transform, some texture and contour features of the image cannot be accurately captured. Then, Candes and Donoho proposed curvelets transform (CVT) [14], and Cunha et al. introduced the nonsubsampling contourlet transform [15], which overcomes the disadvantage of translation invariance.

With the progress of deep learning in target tracking [16–19], target detection [20], and image restoration [21], some algorithms that are based on deep learning have also appeared in the field of image processing [22, 23]. Compared with traditional methods, deep learning allows computing models composed of multiple processing layers to learn data representation with multiple levels of abstraction. Deep learning uses backpropagation algorithm to direct how the machine works, so as to find the complex structure in large data sets and solve the problem of poor consideration of handcrafted. Applications in the field of image processing mainly include CNN [24], GAN [25], Siam network [26], and automatic encoder [27].

3. Proposed Method

3.1. Overview and Notations. We summarize the main symbols used in this paper in Table 1. Specifically, we use IR to represent infrared image, VI to represent visible image, and FU to represent the fused image. On a special note, B_t represents the base layer with image type t , where $t \in \{IR, VI, FU\}$, D_t^n is the small-scale layer with image type t , where $t \in \{IR, VI, FU\}$, and n is the filtering times, $n = 1, 2, 3, 4$. D_t^5 is the large-scale layer. If the subscript is not specifically defined, it can be either an infrared image or a visible image by default.

The framework of the proposed algorithm is shown in Figure 1. At the beginning, we decompose the input image using fast approximate bilateral filter [28]. The specific decomposition process is introduced in the next sections. Then, we use the absolute maximum selecting rule, the summation rule, and local energy characteristics to fuse small-scale layers, large-scale layer, and base layer, respectively. Finally, the fused image is merged with the fused small-scale layers, the fused large-scale layer, and the fused base layer. Our method can better suppress the noise in source images, transfer the valuable detail texture information to fused image, and solve the problem of insufficient visual detail fusion in the fusion process.

3.2. Image Decomposition. We divide the image into three parts: the base layer containing most of the residual low-frequency information, the small-scale layer containing image detail texture information, and the large-scale layer containing image edge structure. Figure 2 shows the specific decomposition process. Firstly, the small structure information of the source infrared and visible image are repeatedly removed by the fast approximate bilateral filter. We perform fast approximate bilateral filter five times on input

TABLE 1: Notation usage in the paper.

Symbol	Meaning	Symbol	Meaning
\mathbb{S}	Spatial domain	\mathbb{R}	Intensity range domain
$A, W...$	2D function defined on \mathbb{S}	$a, w...$	3D function defined on $\mathbb{S} \times \mathbb{R}$
$p, C...$	Vectors	$p \in \mathbb{S}$	Pixel position in 2D vector
$\ x\ $	L_2 norm	$A_p \in \mathbb{R}$	Pixel intensity at point p
G_σ	1D Gaussian: $x \mapsto \exp(-x^2/2\sigma^2)$	G_{σ_s, σ_r}	3D Gaussian: $(x, \zeta) \mapsto \exp(-x \cdot x/2\sigma_s^2 - \zeta^2/2\sigma_r^2)$
W^b	Normalization factor	σ_s, σ_r	Standard deviation (space and range)
B_t	The base layer with image type t	D_t^n	The small-scale layers with image type t ($n = 1, 2, 3, 4$)
FU	The fused image	D_t^5	The large-scale layer with image type t

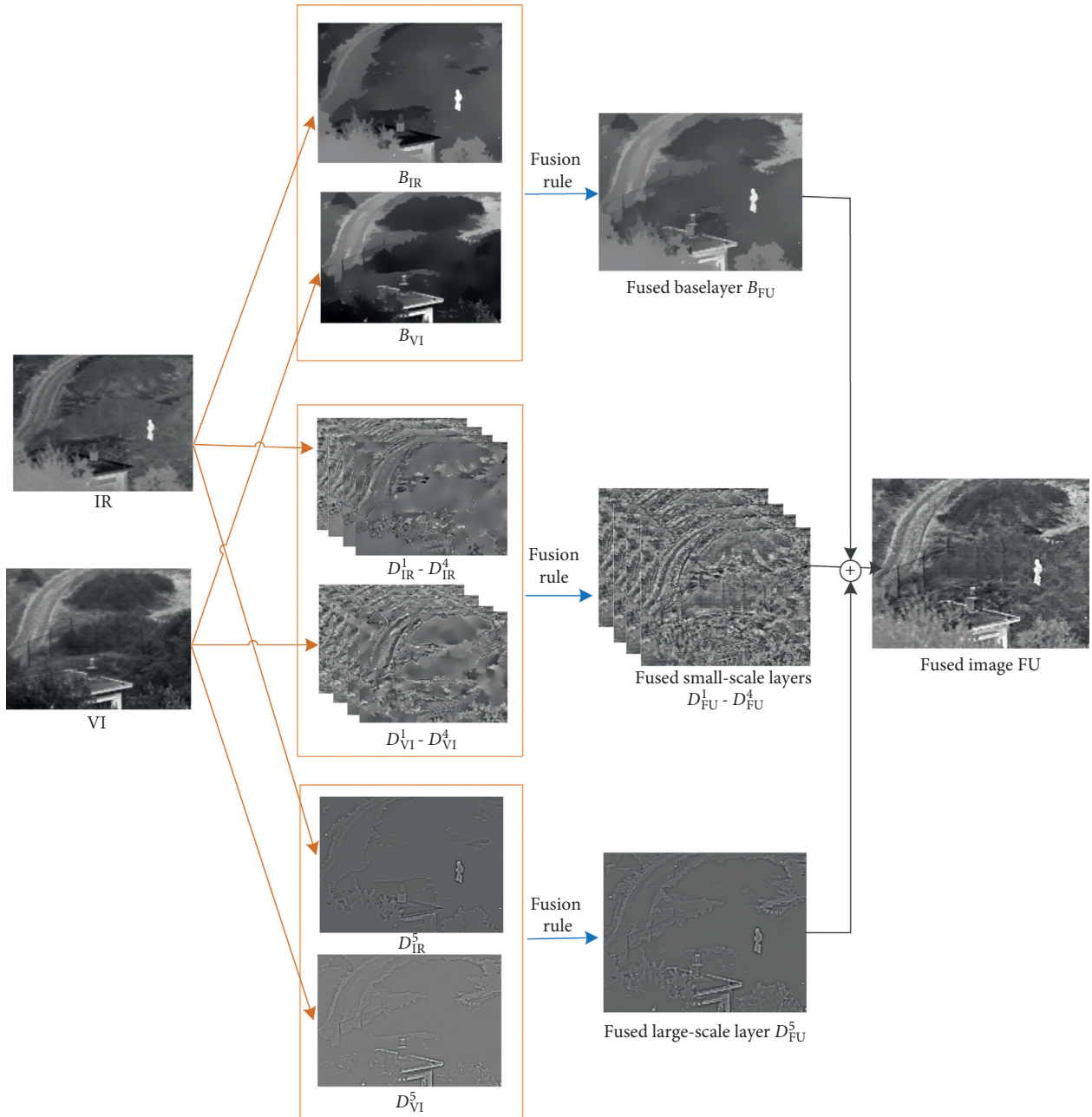


FIGURE 1: The framework of the proposed method.

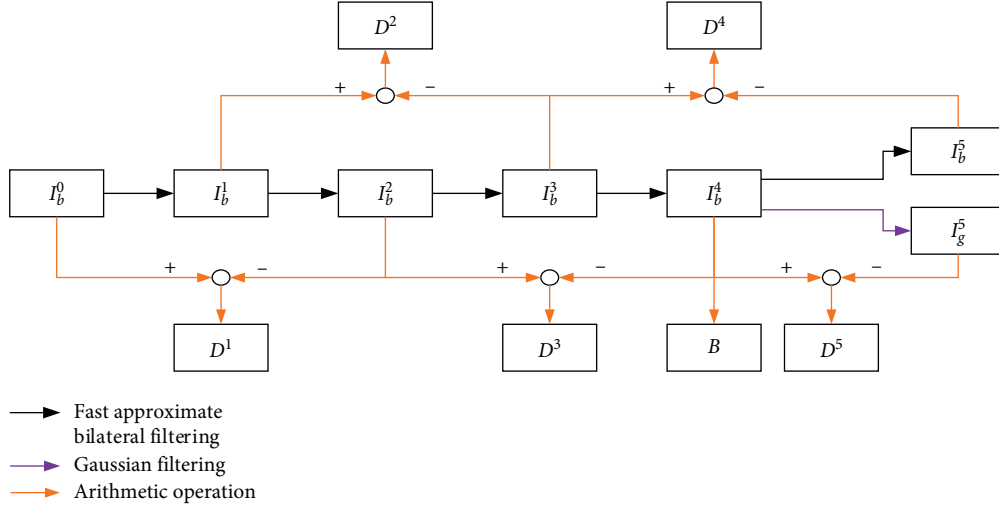


FIGURE 2: Decomposition process of IR and VI.

image to acquire the gradually coarsened image sequence $I_b^1, I_b^2, I_b^3, I_b^4, I_b^5$. Experiments show that the effect of five times filtering is better. Then, the two images separated by one image in the sequence are subtracted to obtain multiple small-scale layers D^1, D^2, D^3, D^4 . Finally, the large-scale layer D^5 is obtained by subtracting I_g^5 from I_b^4 . We can use this decomposition method to extract fine texture details from visible image, which is very important for the fusion of infrared and visible images.

The initial image I_b^0 is the original visible or infrared input image. We can think of this process as an iterative process. The n th iteration is calculated as follows:

$$\begin{aligned} I_b^n &= \text{FABF}(I_b^{n-1}, I_b^{n-1}, \sigma_r, \sigma_s), (n = 1, 2, 3, 4, 5) \\ D^n &= I_b^{n+1} - I_b^n, (n = 1, 2, 3, 4) \\ I_g^n &= \text{Gaussian}(I_b^{n-1}, \sigma_s), (n = 5) \\ D^n &= I_g^n - I_b^{n-1}, (n = 5), \end{aligned} \quad (1)$$

where $\text{FABF}()$ represents the fast approximate bilateral filter. I_g^n is the image obtained after Gaussian filter, σ_r, σ_s represent the standard deviations of the range and spatial Gaussians, parameter setting as [28]. σ_s is mainly determined by the image size, which is set as $\min(\text{size}(\text{size}(I_b^0))/30)$. σ_r is mainly determined by the pixel intensity difference of the image, which is set as maximum pixel difference of $I_b^0/30$. D^n represents the obtained small-scale layer or large-scale layer, Gaussian(I_b^{n-1}, σ_s) is the Gaussian filter, and scales smaller than σ_s structure can be removed by the filter.

Generally, the base layer does not need to retain the details and edge information of the source image and is the coarsest layer of the input image. The base layer is mainly used to control the appearance and global contrast of image. When too much detail and edge information is retained in the base layer, the image will lose some useful information. In our method, we choose I_b^4 as the base layer.

Figure 3 shows the decomposed layers. (a) and (b) are infrared and visible images, (c) is the fused base layer, and (d) is the fused image. The second row shows the infrared

image, and the third row shows the visible image. (a1)-(d1), (a2)-(d2) are small-scale layers from infrared and visible image decomposition. Different small-scale layers contain different features. It can be seen that the small-scale layer corresponding to (a1), (a2) mainly contains fine-detail texture information. From (b1), (b2) to (d1), (d2), the small-scale layers contain more and more abundant target feature information, and the hillside, trees, and road information contained in (d1) and (d2) are the most obvious. (e1), (e2) are the large-scale layers and (g1), (g2) are the base layers. We can see that each scale contains a corresponding proportion of content. In short, due to the characteristics of edge preservation and scale perception, this decomposition method can reduce the halo and preserve the edge features well, which is beneficial to the further image fusion.

3.3. Fast Approximate Bilateral Filter. Bilateral filter (BF) [29] is a filter that can denoise and maintain edges. It is a nonlinear filter. Because the filter consists of two functions, one of which determines the filter coefficient through pixel difference, and the other function determines the filter coefficient through geometric space distance, the denoising effect can be achieved. The advantage of bilateral filter over ordinary Gaussian low-pass filter [30] is that it considers both the influence of image position on the center pixel and the radiation difference in the pixel range. The filtered output formula of image A at pixel position p is as follows:

$$A_p^{bf} = \frac{1}{W_p^{bf}} \sum_{q \in \mathbb{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - A_q|) A_q, \quad (2)$$

$$\text{with } W_p^{bf} = \sum_{q \in \mathbb{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - A_q|), \quad (3)$$

where A_p^{bf} is the filtered image, W_p^{bf} represents a normalization factor, σ_s, σ_r are the standard deviation of the spatial Gaussian function and the distance Gaussian function, p, q

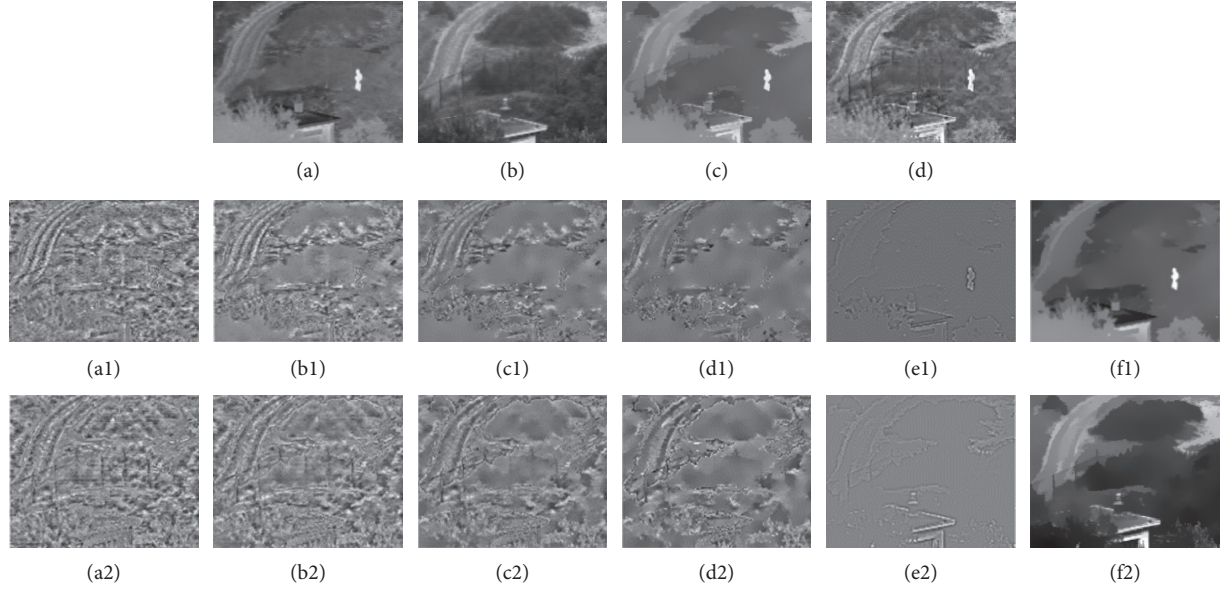


FIGURE 3: Decomposition of IR and VI. (a) IR. (b) VI. (c) Fused base layer. (d) Fused image: small-scale layers, large-scale layers, and base layers.

represent the pixel coordinates, A_p, A_q represent the pixel intensity values of pixels p, q , $\|p - q\|$ calculates the Euclidean distance of pixels p, q , $|A_p - A_q|$ calculates the absolute value of A_p, A_q , and \mathbb{S} is spatial domain.

Bilateral filter (BF) has been widely used in many fields and proved to be very effective. However, bilateral filter needs the gray information of each central point neighborhood to determine its coefficients, which leads to its long running time. In this paper, we suggest decomposing the original image using fast approximate bilateral filter [28]; we use a signal processing method, mainly adding the one-dimensional signal strength to the original domain to form a high-dimensional space, and downsampling in the high-dimensional space. This scheme produces equivalent running time but significantly improves the accuracy. Next, we describe the implementation of this method in detail.

Firstly, we rewrite (2) using a two-dimensional vector:

$$\begin{pmatrix} W_p^{bf} A_p^{bf} \\ W_p^{bf} \end{pmatrix} = \sum_{q \in \mathbb{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - A_q|) \begin{pmatrix} A_q \\ 1 \end{pmatrix}. \quad (4)$$

We specify the weight $W = 1$ to keep the characteristics of the weighted average value of the bilateral filter:

$$\begin{pmatrix} W_p^{bf} A_p^{bf} \\ W_p^{bf} \end{pmatrix} = \sum_{q \in \mathbb{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - A_q|) \begin{pmatrix} W_q A_q \\ W_q \end{pmatrix}. \quad (5)$$

We further promote the interpretation of the above equation in 3D space. In order to more conveniently define the sum of the whole three-dimensional space, we added a new dimension ζ and Kronecker's symbol $\sigma(\zeta)$ (1 if $\zeta = 0, 0$ otherwise), define each point (x, y, ζ) intensity of A , and rewrite equation (5) using $\sigma(\zeta - A_q) = 1$; when $\zeta \neq A_q$, the terms are cancelled:

$$\begin{pmatrix} W_p^{bf} A_p^{bf} \\ W_p^{bf} \end{pmatrix} = \sum_{q \in \mathbb{S}} \sum_{\zeta \in \mathbb{R}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - \zeta|) \cdot \sigma(\zeta - A_q) \begin{pmatrix} W_q A_q \\ W_q \end{pmatrix}. \quad (6)$$

Equation (6) is a sum on a product space $\mathbb{S} \times \mathbb{R}$. The product $G_{\sigma_s} G_{\sigma_r}$ represents a separable Gaussian kernel $g_{\sigma_s} g_{\sigma_r}$ on $\mathbb{S} \times \mathbb{R}$:

$$g_{\sigma_s, \sigma_r}: (x \in \mathbb{S}, \zeta \in \mathbb{R}) \mapsto G_{\sigma_s}(\|x\|) G_{\sigma_r}(|\zeta|). \quad (7)$$

Then, we introduce two new functions A and w :

$$A: (x \in \mathbb{S}, \zeta \in \mathbb{R}) \mapsto A_x, \quad (8)$$

$$\begin{aligned} w: (x \in \mathbb{S}, \zeta \in \mathbb{R}) &\mapsto \sigma(\zeta - A_x) \\ &= \sigma(\zeta - A_x) W_x \text{ Since } W_x = 1. \end{aligned} \quad (9)$$

We rewrite the right part of (6) according to (8) and (9):

$$\begin{aligned} \sigma(\zeta - A_q) \begin{pmatrix} W_q A_q \\ W_q \end{pmatrix} &= \begin{pmatrix} \sigma(\zeta - A_q) W_q A_q \\ \sigma(\zeta - A_q) W_q \end{pmatrix} \\ &= \begin{pmatrix} w(q, \zeta) a(q, \zeta) \\ w(q, \zeta) \end{pmatrix}, \end{aligned} \quad (10)$$

with definition (10), we get

$$\begin{pmatrix} W_p^{bf} A_p^{bf} \\ W_p^{bf} \end{pmatrix} = \sum_{(q, \zeta) \in \mathbb{S} \times \mathbb{R}} g_{\sigma_s, \sigma_r}(p - q, A_p - \zeta) \cdot \begin{pmatrix} w(q, \zeta) a(q, \zeta) \\ w(q, \zeta) \end{pmatrix}. \quad (11)$$

The value of the above formula at point (wa, w) is as follows:

$$\begin{pmatrix} W_p^{bf} A_p^{bf} \\ W_p^{bf} \end{pmatrix} = \left[g_{\sigma_s, \sigma_r} \otimes \begin{pmatrix} wa \\ w \end{pmatrix} \right] (p, A_p), \quad (12)$$

where \otimes represents convolution operation. On the basis of the above formula, we introduce the functions a^{bf} and w^{bf} :

$$(w^{bf} a^{bf}, w^{bf}) = [g_{\sigma_s, \sigma_r} \otimes (wa, w)]. \quad (13)$$

Therefore, the bilateral filter is represented as convolution operation, and then we perform nonlinear operation:

$$(w^{bf} a^{bf}, w^{bf}) = g_{\sigma_s, \sigma_r} \otimes (wa, w), \quad (14)$$

$$A_p^{bf} = \frac{w^{bf}(p, A_p) a^{bf}(p, A_p)}{w^{bf}(p, A_p)}. \quad (15)$$

In fact, the nonlinear part consists of two operations. The first operation is slicing, that is, evaluating the function of $w^{bf} a^{bf}$ and w^{bf} at point (p, A_p) . The second is division. In our case, the slice and segmentation results are independent of their corresponding position order.

3.4. Image Reconstruction

3.4.1. Base Layer Fusion. The traditional average fusion rules cannot effectively fuse the information of base layer of source images, which leads to the loss of some low-frequency information. We preserve more image information by fusing local energy characteristics to obtain better visual effect. The gray value of the image is what we call energy. In this paper, we

define energy as the weighted average gray value of a region. The greater the gray value, the higher the energy of the image. For image fusion, each image has good parts and bad parts, what we need to do is to extract the good parts of each image. As can be seen in Figure 3 (f1), the human information is more prominent, so the energy is greater. In Figure 3 (f2), there is strong-detail texture information in the left corner of the image, and the energy in this area is greater. When fusing, we choose the part with large energy in the local region for fusion, the fused image contains most of the useful information of the original image. The fused base layer B_{FU} is calculated as follows:

$$B_{FU}(p, q) = \max_{t \in \{IR, VI\}} \left\{ \sum_{k=1}^m \sum_{l=1}^n W(k, l) * |B_t(p+k, q+l)| \right\}, \quad (16)$$

where k and l define the size of the region, B_t is the base layer of IR or VI. (p, q) is the central point of the neighborhood. In this paper, we set $m = n = 3$, and W is the weight template, expressed as follows:

$$W = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (17)$$

3.4.2. Small-Scale Layers Fusion. At the small-scale level, we select all decomposition levels from $n = 1$ to $n = 4$. The absolute maximum selection rule is used to integrate all important texture features and edge intensity into the fused small-scale layers. The formula is given by

$$D_{FU}^n(p, q) = \begin{cases} D_{IR}^n(p, q), & |D_{IR}^n(p, q)| > |D_{VI}^n(p, q)| \\ D_{VI}^n(p, q), & \text{otherwise} \end{cases} \quad (n = 1, 2, 3, 4), \quad (18)$$

where D_{FU}^n is the fused small-scale layer image, (p, q) represents the corresponding position of D_{IR}^n , D_{VI}^n , and D_{FU}^n , and n denotes the number of decomposition levels.

3.4.3. Large-Scale Layer Fusion. At the large-scale level, we only select the bottom layer of decomposition ($n = 5$). Most of the feature information contained in them does not overlap due to the different imaging principles of infrared and visible images. It can be seen from Figure 3 (e1), (e2) that the two layers have a large amount of supplementary information. Obviously, using the weighted average method will lose large amounts of important information. Therefore, we use the summation rule to fuse the large-scale layers to prevent information loss. The formula is as follows:

$$D_{FU}^n(p, q) = w_1 D_{IR}^n(p, q) + w_2 D_{VI}^n(p, q) \quad (n = 5), \quad (19)$$

where w_1 , w_2 are the weight values of D_{IR}^n and D_{VI}^n , respectively. We choose $w_1 = 1$ and $w_2 = 1$ to save more image

information. In order to show the superiority of using summation rule to fuse large-scale layer, we compare it with absolute maximum selection rule and weighted average rule and use API [31], SD [32], AG [32], and EN [33] to measure the quality of these three methods. It can be easily seen from Table 2 that except that the API is slightly lower than the absolute maximum selection rule, the fusion result obtained by the summation rule is better in performance indicators.

3.4.4. Image Composition. Finally, the decomposition information at all levels is fused to obtain the fused image FU and is given by

$$FU = B_{FU} + \sum_{n=1}^N D_{FU}^n \quad (n = 5). \quad (20)$$

This process is briefly introduced in Algorithm 1.

TABLE 2: The qualitative evaluation results of different fusion methods in large-scale layer.

Method	API	SD	AG	EN
Absolute maximum selection rule	112.10	35.79	14.13	7.06
Weighted average rule	112.09	34.48	13.10	7.03
Summation rule	112.09	35.95	14.19	7.07

```

Input:  $I_{IR}, I_{VI}$ 
Output: FU
1)  $I_b^0 = I_{IR} \text{ OR } I_{VI}$ 
2) Image decomposition
   for  $n = 1, 2, 3, 4, 5$  do
      $I_b^n = FABF(I_b^{n-1}, I_b^{n-1}, \sigma_s, \sigma_r)$ 
   end for
    $I_g^5 = \text{Gaussian}(I_b^4, \sigma_s)$ 
   for  $n = 1, 2, 3, 4$  do
      $D^n = I_b^{n+1} - I_b^{n-1}$ 
   end for
    $D^5 = I_g^5 - I_b^4$ 
3) Image fusion
   The fused base layer is obtained by (16)
   for  $n = 1, 2, 3, 4$  do
     The fused small-scale layers are obtained by (18)
   end for
   The fused large-scale layer is obtained by (19)
4) The fused image is obtained by (20)

```

ALGORITHM 1: Image decomposition and reconstruction.

4. Results and Discussion

4.1. Experimental Setup. For the purpose of verifying the effectiveness of our fusion algorithm, we compared it with those of 10 other latest infrared and visible fusion methods, containing lp [9], CVT [14], NSCT [34], DTCWT [35], MSVD [36], TIF [37], VSM-WLS [38], FPDE [39], MGFF [40], and GTF [41] methods. All the experimental parameters of these ten algorithms are set on the basis of the original paper.

In our experiments, the test images are obtained from the TNO_Image_Fusion_Dataset. Five pairs of infrared and visible images of different scenes are used. All the experiments in this paper are implemented in MATLAB (2017b), using an Intel® Core™ I5-10210U CPU.

4.2. Qualitative Evaluation. The visual perception quality comparison of Camp, Road, Keptein, Meting, and Steamboat data sets are shown from Figures 4 to 8. Among them, (a) is infrared image, (b) is visible image, (c)-(l) are the results of compared methods, and (m) is the proposed method. To see the difference more distinctly, we mark a rectangle in the fused image; what we can see is that our method transfers most of the important background and details to the fused image, the important infrared structure information is properly retained, and the image noise is also reduced. Figure 4 is a fusion example on the “Camp” image. In the red rectangle, we can observe that the fence information presented by MSVD, FPDE, and GTF methods are fuzzy and difficult to identify. In

contrast, our method has clearer fence texture details and reduces the interference of noise to a certain extent.

In Figure 5, we can find that our method has more advantages when we obtain some detail features from the source image under weak light conditions because we improve the overall brightness of the image, while other methods obviously do not achieve good visual results. Among them, FPDE and NSCT have fusion artifacts, and the characters of NSCT, MSVD, and TIF are not clearly displayed.

In Figure 6, our method makes the edge structure of tree branches vivid and the ground texture information clear. In contrast, DTCWT, CVT, and TIF have distorted artifacts due to the influence of incorrect introduction of visible light information; the overall images of lp, MSVD, and GTF are dark and spectral features are not rich enough.

In Figure 7, our method can accurately fuse the obvious person information of the infrared image and the brick hole structure detail on both sides of the visible image. It can be seen from the figure that MSVD, VSM-WLS, FPDE, and GTF tend to produce some unnatural artifact information. Because the brightness difference between source images is large, and these methods cannot effectively suppress noise, they are difficult to show a clear hole structure. Overall, our method can produce better fusion results. Similarly, as shown in Figure 8, for the “steamboat” image, our method has better contour and contrast and improves the clarity of the image.

4.3. Quantitative Evaluation. It is difficult for human vision to distinguish small differences in images. Thus, it is unreliable to evaluate the effect of fused image only from the qualitative evaluation. The quality of image fusion also needs quantitative analysis. For quantitative evaluation, we use four indicators to evaluate the fusion accuracy. The formula of these indicators is defined as follows:

- (1) Average pixel intensity (API) or mean \bar{u} is the arithmetic mean of the gray values of all pixels in the image and is calculated by $\bar{u} = \frac{\sum_{i=1}^M \sum_{j=1}^N H(i, j)}{M \times N}$, where $M \times N$ is the size of the image and $H(i, j)$ is pixel intensity.
- (2) Standard deviation (SD) is the dispersion of image gray relative to the mean, which is used to evaluate the size of image contrast and is calculated by $SD = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N [H(i, j) - \bar{u}]^2}{M \times N}}$.
- (3) Average gradient (AG) can sensitively reflect the expression level of image contrast to small details and is calculated by

$$AG = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\frac{(H(i+1, j) - H(i, j))^2 + (H(i, j+1) - H(i, j))^2}{2}} \quad (21)$$

- (4) Entropy (EN) is mainly an objective evaluation index to measure the amount of information contained in image

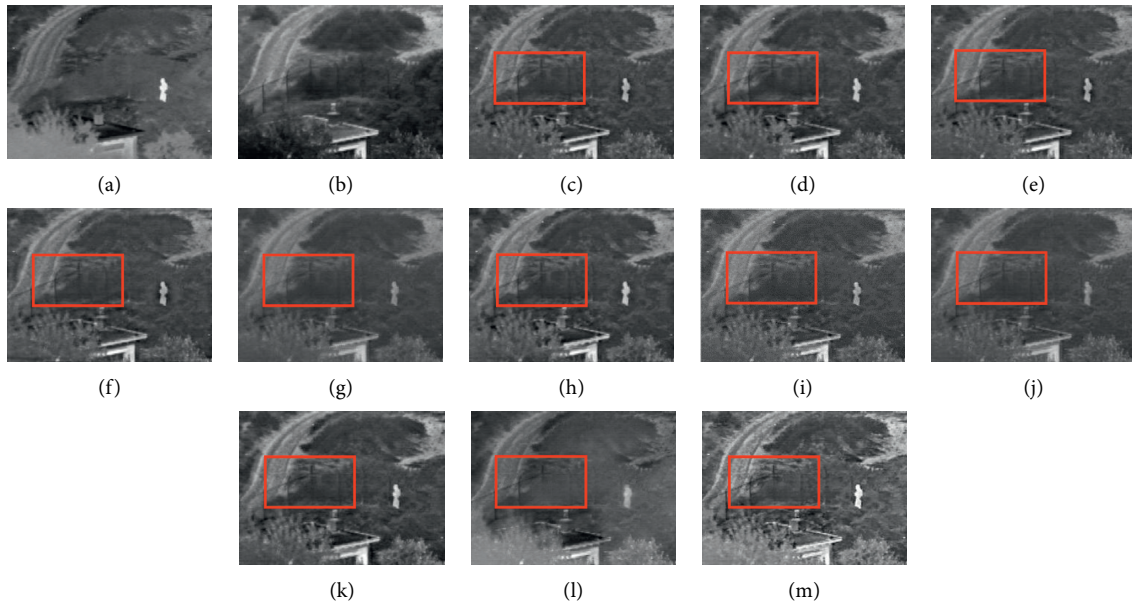


FIGURE 4: Experiment on “Camp” images. (a) Infrared image. (b) Visible image. (c) NSCT. (d) lp. (e) DTCWT. (f) CVT. (g) MSVD. (h) TIF. (i) VSM-WLS. (j) FPDE. (k) MGFF. (l) GTF. (m) Proposed.

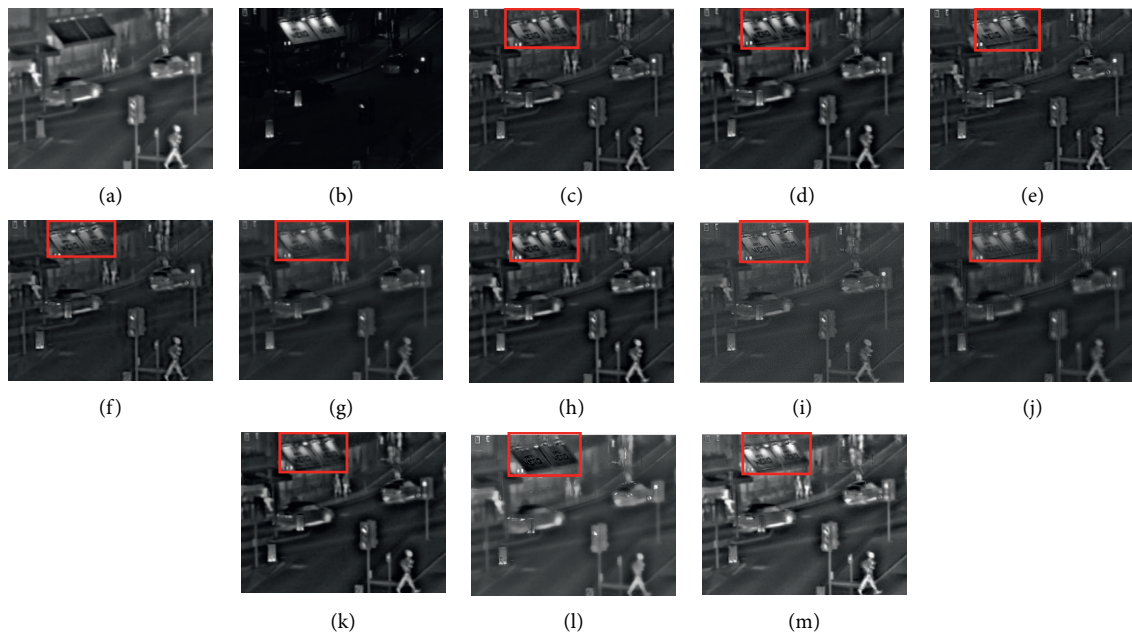


FIGURE 5: Experiment on “Road” images. (a) Infrared image. (b) Visible image. (c) NSCT. (d) lp. (e) DTCWT. (f) CVT. (g) MSVD. (h) TIF. (i) VSM-WLS. (j) FPDE. (k) MGFF. (l) GTF. (m) Proposed.

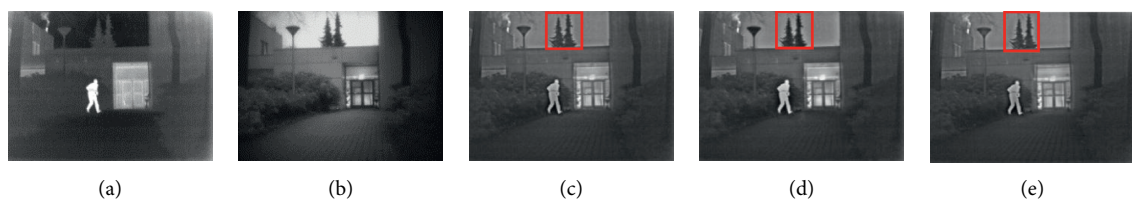


FIGURE 6: Continued.

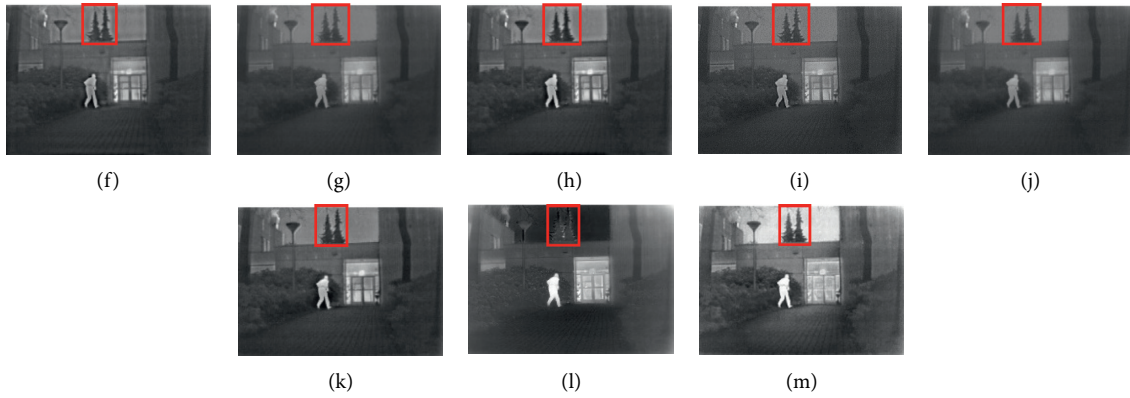


FIGURE 6: Experiment on “Keptein_1123” images. (a) Infrared image. (b) Visible image. (c) NSCT. (d) lp. (e) DTCWT. (f) CVT. (g) MSVD. (h) TIF. (i) VSM-WLS. (j) FPDE. (k) MGFF. (l) GTF. (m) Proposed.

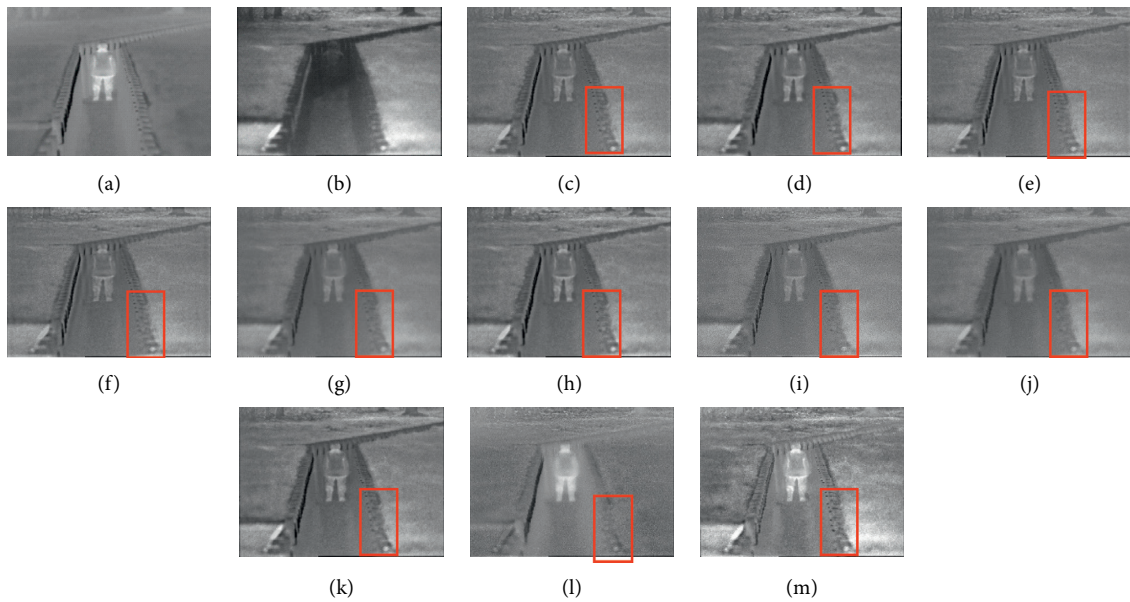


FIGURE 7: Experiment on “Meting055” images. (a) Infrared image. (b) Visible image. (c) NSCT. (d) lp. (e) DTCWT. (f) CVT. (g) MSVD. (h) TIF. (i) VSM-WLS. (j) FPDE. (k) MGFF. (l) GTF. (m) Proposed.

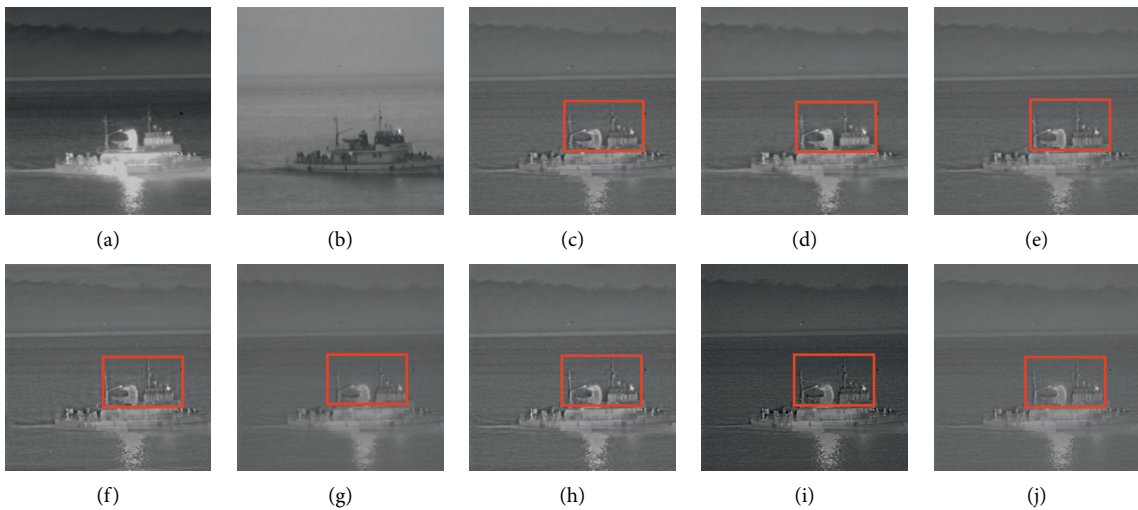


FIGURE 8: Continued.

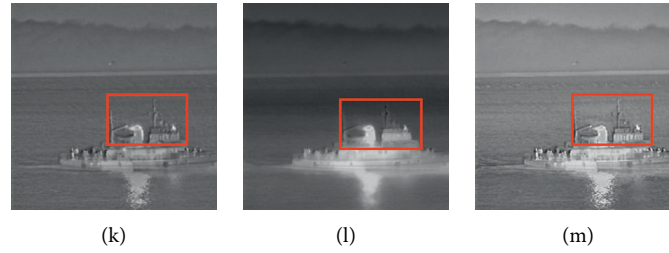


FIGURE 8: Experiment on “Steamboat” images. (a) Infrared image. (b) Visible image. (c) NSCT. (d) lp. (e) DTCWT. (f) CVT. (g) MSVD. (h) TIF. (i) VSM-WLS. (j) FPDE. (k) MGFF. (l) GTF. (m) Proposed.

TABLE 3: The qualitative evaluation results of different fusion methods.

Source images	Index	NSCT	lp	CVT	DTCWT	MSVD	TIF	FPDE	VSM-WLS	MGFF	GTF	Proposed
Camp	API	91.09	90.82	91.09	91.08	91.09	90.92	91.59	91.07	91.04	98.06	112.09
	SD	27.11	29.13	26.96	26.28	22.87	29.25	23.13	24.10	32.01	26.86	35.95
	AG	7.80	8.08	7.79	7.47	6.59	7.61	6.15	7.14	8.89	6.08	14.19
	EN	6.55	6.66	6.53	6.48	6.26	6.64	6.29	6.35	6.80	6.67	7.07
Road	API	51.90	52.48	51.90	51.90	51.90	52.34	52.39	51.88	51.86	80.19	84.28
	SD	26.49	31.89	25.40	25.49	21.50	30.35	22.20	22.91	32.49	29.78	41.20
	AG	6.12	6.36	6.10	6.05	5.23	5.82	7.89	5.54	6.45	4.12	10.26
	EN	6.19	6.47	6.17	6.16	5.95	6.41	6.07	6.01	4.94	6.62	7.00
Kaptein_1123	API	82.18	82.18	82.18	82.18	82.10	82.15	82.61	82.10	82.23	84.30	111.02
	SD	36.28	36.28	36.28	36.28	31.62	36.61	31.73	32.24	36.72	38.17	53.51
	AG	6.95	6.95	6.95	6.95	5.83	6.33	5.86	6.51	7.86	4.58	12.04
	EN	6.78	6.78	6.78	6.78	6.55	6.84	6.61	6.63	6.87	6.96	7.34
Meting055	API	127.80	127.64	127.80	127.80	127.80	127.81	128.28	127.82	127.74	130.90	147.91
	SD	24.48	26.39	24.28	24.29	20.30	25.84	21.07	22.53	25.74	20.94	33.53
	AG	10.83	10.90	10.83	10.84	6.88	9.15	9.90	10.49	10.29	10.30	17.66
	EN	6.49	6.57	6.49	6.48	6.32	6.57	6.39	6.42	6.63	6.14	6.92
Steamboat	API	110.92	110.92	110.92	110.91	110.92	110.94	111.42	110.91	110.96	91.73	140.22
	SD	12.65	14.07	12.47	12.32	10.69	13.17	10.78	11.27	14.17	33.95	22.10
	AG	3.85	3.90	3.88	3.78	3.19	3.53	3.10	3.49	4.18	1.89	6.73
	EN	5.16	5.31	5.21	5.14	4.93	5.23	4.97	5.01	5.33	6.44	6.27

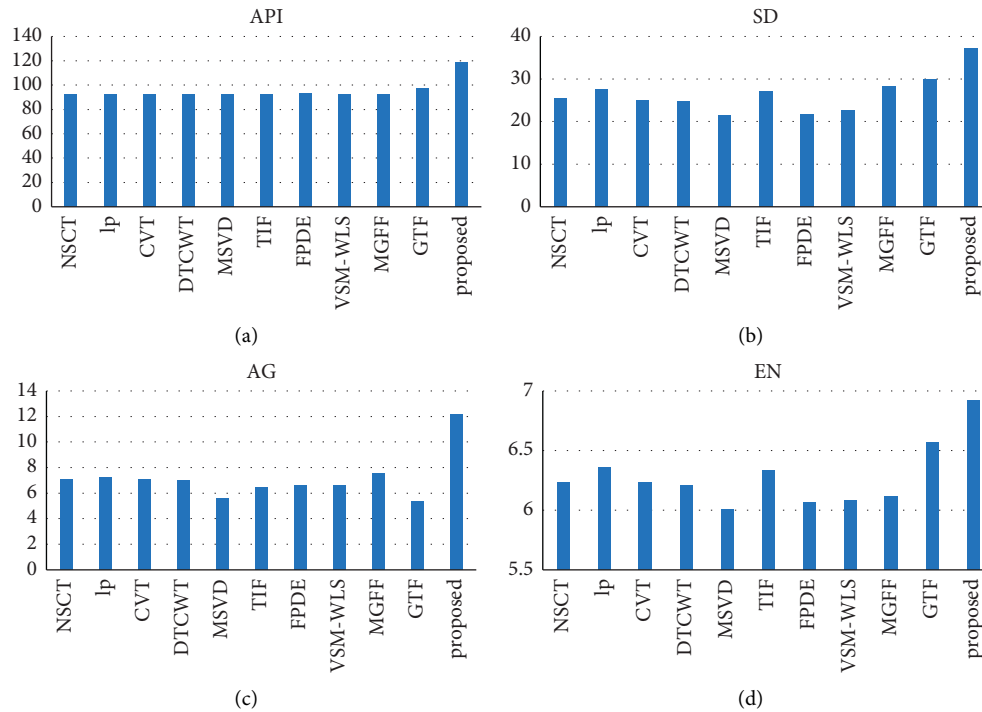


FIGURE 9: Bar comparison of average API, SD, AG, and EN values of five IR and VI image datasets of various fusion methods.

TABLE 4: Run time comparison of different fusion methods.

Times	Camp	Road	Kaptein_1123	Meting055	Steamboat	Average
NSCT	1.66	6.73	4.56	6.43	4.17	4.71
lp	0.01	0.02	0.01	0.02	0.01	0.01
CVT	0.9	1.87	1.45	2.27	1.47	1.59
DTCWT	0.3	0.47	0.35	0.53	0.4	0.41
MSVD	0.66	1.24	0.96	1.82	0.94	1.12
TIF	0.27	0.30	0.33	1.11	0.39	0.48
FPDE	1.24	3.12	2.68	4.13	2.77	2.79
VSM-WLS	2.02	3.74	3.35	4.81	3.06	3.40
MGFF	0.11	0.46	0.01	0.60	0.38	0.31
GTF	1.64	6.22	4.82	7.14	4.82	4.93
Ours	1.93	4.73	4.29	6.40	4.12	4.29

and is calculated by $EN = -\sum_{i=0}^{255} p_i \log_2(p_i)$, where i is the gray value and p_i is the gray probability distribution.

For all the four metrics, the higher the objective indexes API, SD, AG, and EN are, the better the fusion effect will be.

The quantitative comparisons of the experimental results are presented in Table 3, and we show the maximum value in bold among different algorithms. It can be seen from the table that our method has higher fusion quality indicators than other methods in terms of fusion quality indexes API, SD, AG, and EN and has good correlation with the results of visual comparison.

In order to more intuitively show the advantages of the proposed method, we obtained the average value of each index of five groups of pictures (Camp, Road, Keptein, Meting, and Steamboat) and obtained the bar statistical chart of four indexes. As shown in Figure 9, it can be seen that the performance of the proposed method is the best among all indexes. We also record the run time of different fusion methods for each group of images. As shown in Table 4, our method is slightly slower, which is where we will further improve.

5. Conclusions

In this paper, an infrared and visible image fusion algorithm based on fast approximate bilateral filtering and local energy characteristics is proposed. The image decomposed by fast bilateral filter is smoother and contains less noise. According to the characteristics of each decomposition layer, we use different fusion regulation to fuse the decomposed image, which not only avoids the information loss of the traditional fusion rules, but also enriches the visual information of the fused image. On most occasions, the image looks more natural and contains less artificial information. Experimental results fully illustrate the superiority of the proposed algorithm. The comparison with other 10 fusion methods shows that our algorithm can better describe the most significant information in the image, improve the overall contrast of the image, and maintain the information of the source image to the greatest extent.

Data Availability

The TNO Image Fusion Dataset is available at https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Zongping Li was mainly responsible for data collating and draft writing. Wenxin Lei designed the method and validated the results. Xudong Li and Tingting Liao designed and debugged the program code. Jianming Zhang mainly reviewed and polished the article. All the authors have read and agreed on the current version of the article.

Acknowledgments

This work was supported by the Natural Science Foundation of China (61972056 and 61901061), the Basic Research Fund of Zhongye Changtian International Engineering Co., Ltd. (2020JCYJ07), the Research Fund of Hunan Provincial Education Department (19C0028 and 19C0031), the Enterprise-University Joint Postgraduate Scientific Research Innovation Fund of Hunan Province (QL20210205), and the Postgraduate Scientific Research Innovation Fund of CSUST (CX2021SS70).

References

- [1] A. C. Muller and S. Narayanan, "Cognitively-engineered multisensor image fusion for military applications," *Information Fusion*, vol. 10, no. 2, pp. 137–149, 2009.
- [2] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75–89, 2016.
- [3] S. Singh, D. Gupta, R. S. Anand, and V. Kumar, "Non-subsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network," *Biomedical Signal Processing and Control*, vol. 18, pp. 91–101, 2015.
- [4] C. Yu, B. Li, J. Xiao et al., "Astronomical data fusion: recent progress and future prospects - a survey," *Experimental Astronomy*, vol. 47, no. 3, pp. 359–380, 2019.
- [5] T. Wan, C. Zhu, and Z. Qin, "Multifocus image fusion based on robust principal component analysis," *Pattern Recognition Letters*, vol. 34, no. 9, pp. 1001–1008, 2013.
- [6] Li Guangxin, Xu Shuyan, and Z. Xin, *Fast Color-Transfer-Based Image Fusion Method for Merging Infrared and Visible Images*, pp. 77100S-77101S, SPIE, Orlando, Florida, 2010.

- [7] G. Mingda et al., *Multiresolution Contrast Modulation Method Applied to Lidar 4-D Image Fusion*. In *Pattern Recognition Letters*, IEEE, pp. 224–227, Changchun, China, 2012.
- [8] S. Li, J. T. Kwok, and Y. Wang, “Multifocus image fusion using artificial neural networks,” *Pattern Recognition Letters*, vol. 23, no. 8, pp. 985–997, 2002.
- [9] P. J. Burt and E. H. Adelson, “The laplacian pyramid as a compact image code,” in *Readings in Computer Vision*, pp. 671–679, Morgan Kaufmann, 1987.
- [10] J. Du, W. Li, B. Xiao, and Q. Nawaz, “Union Laplacian pyramid with multiple features for medical image fusion,” *Neurocomputing*, vol. 194, pp. 326–339, 2016.
- [11] Z. Zhiqiang, L. Sun, and W. Bo, “Multi-scale weighted gradient-based fusion for multi-focus images,” *Information Fusion*, vol. 20, pp. 60–72, 2014.
- [12] B. D. Yu, J. L. Ming, and L. Jie, “Image fusion algorithm based on contrast pyramid and its performance evaluation,” *Applied Mechanics and Materials*, vol. 525, pp. 711–714, 2014.
- [13] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [14] E. Candes and D. DL, *Curvelets - A Surprisingly Effective Nonadaptive Representation for Objects with Edges*, Stanford Univ Ca Dept of Statistics, CA, United States, 2000.
- [15] A. L. Da Cunha, J. Zhou, and M. N. Do, “The nonsubsampling contourlet transform: theory, design, and applications,” *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, 2006.
- [16] J. Zhang, Y. Liu, H. Liu, and J. Wang, “Learning local-global multiple correlation filters for robust visual tracking with k filter redetection,” *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [17] J. Zhang, X. Jin, J. Sun, J. Wang, and K. Li, “Dual model learning combined with multiple feature selection for accurate visual tracking,” *IEEE Access*, vol. 7, pp. 43956–43969, 2019.
- [18] J. Zhang, J. Sun, J. Wang, and X.-G. Yue, “Visual object tracking based on residual network and cascaded correlation filters,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427–8440, 2021.
- [19] J. Zhang, Y. Liu, H. Liu, J. Wang, and Y. Zhang, “Distractor-aware visual tracking using hierarchical correlation filters adaptive selection,” *Applied Intelligence*, vol. 3, 2021.
- [20] J. Zhang et al., “Arun Kumar Sangaiah. Lightweight deep network for traffic sign classification,” *Annals of Telecommunications*, vol. 75, no. 7-8, pp. 369–379, 2020.
- [21] Y. Chen, L. Liu, J. Tao et al., “The improved image inpainting algorithm via encoder and similarity constraint,” *The Visual Computer*, vol. 37, no. 7, pp. 1691–1705, 2021.
- [22] J. Ma, Y. Ma, and C. Li, “Infrared and visible image fusion methods and applications: a survey,” *Information Fusion*, vol. 45, pp. 153–178, 2019.
- [23] Z. Siwang, He Yan, L. Yonghe, Li Chengqing, and Z. Jianming, “Multi-channel deep networks for block-based image compressive sensing,” *IEEE Transactions on Multimedia*, vol. 23, pp. 2627–2640, 2021.
- [24] Y. Liu, X. Chen, H. Peng, and Z. Wang, “Multi-focus image fusion with a deep convolutional neural network,” *Information Fusion*, vol. 36, pp. 191–207, 2017.
- [25] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, “FusionGAN: a generative adversarial network for infrared and visible image fusion,” *Information Fusion*, vol. 48, pp. 11–26, 2019.
- [26] Y. Liu et al., “Infrared and visible image fusion with convolutional neural networks,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 16, no. 3, Article ID 1850018, 2018.
- [27] H. Li and X. Wu, “DenseFuse: a fusion approach to infrared and visible images,” *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2018.
- [28] P. Sylvain and D. Frédo, “A fast approximation of the bilateral filter using a signal processing approach,” *International Journal of Computer Vision*, vol. 81, no. 1, pp. 24–52, 2009.
- [29] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 839–846, IEEE, Bombay, India, January 1998.
- [30] R. C. G. Woods and R. E. Woods, *Digital Image Processing*, pp. 177–193, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd ed. edition, 2007.
- [31] V. Petrovic and C. Xydeas, “Objective image fusion performance characterisation,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1866–1871, IEEE, Beijing, China, October 2005.
- [32] B. Kumar, “Image fusion based on pixel significance using cross bilateral filter,” *Signal, Image & Video Processing*, vol. 9, no. 5, pp. 1193–1204, 2015.
- [33] J. W. Roberts, J. Van Aardt, and F. Ahmed, “Assessment of image fusion procedures using entropy, image quality, and multispectral classification,” *Journal of Applied Remote Sensing*, vol. 2, no. 1, pp. 023522–0235228, 2008.
- [34] Z. Qiang and G. Bao-long, “Multifocus image fusion using the nonsubsampling contourlet transform,” *Signal Processing*, vol. 89, no. 7, pp. 1334–1346, 2009.
- [35] J. L. John et al., “Pixel- and region-based image fusion with complex wavelets,” *Information Fusion*, vol. 8, no. 2, pp. 119–130, 2005.
- [36] V. P. S. Naidu, “Image fusion technique using multi-resolution singular value decomposition,” *Defence Science Journal*, vol. 61, no. 5, p. 479, 2011.
- [37] D. P. Bavirisetti and R. Dhuli, “Two-scale image fusion of visible and infrared images using saliency detection,” *Infrared Physics & Technology*, vol. 76, pp. 52–64, 2016.
- [38] J. Ma, Z. Zhou, B. Wang, and H. Zong, “Infrared and visible image fusion based on visual saliency map and weighted least square optimization,” *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017.
- [39] D. P. Bavirisetti, G. Xiao, and G. Liu, “Multi-sensor image fusion based on fourth order partial differential Equations,” in *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–9, IEEE, Xi’an, China, July 2017.
- [40] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, “Multi-scale guided image and video fusion: a fast and efficient approach,” *Circuits, Systems, and Signal Processing*, vol. 38, no. 12, pp. 5576–5605, 2019.
- [41] J. Ma, C. Chen, C. Li, and J. Huang, “Infrared and visible image fusion via gradient transfer and total variation minimization,” *Information Fusion*, vol. 31, pp. 100–109, 2016.