IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 8, 2019, accepted December 19, 2019, date of publication December 23, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961626

A General Perceptual Infrared and Visible Image Fusion Framework Based on Linear Filter and Side Window Filtering Technology

HUIBIN YAN AND ZHONGMIN LI
School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
Corresponding authors: Huibin Yan (huibiny2019@gmail.com) and Zhongmin Li (zhongmli2018@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61263040, in part by the Scientific Research Foundation of Jiangxi Provincial Education Department, China, under Grant GJJ170602, in part by the China Scholarship Council under Grant 201608360166, and in part by the Postgraduate Innovation Special Foundation of Nanchang Hangkong University under Grant YC2018019.

ABSTRACT Recently, edge-preserving filters have achieved great success in infrared (IR) and visible (VI) image fusion field. However, most edge-preserving filters are complex. In this paper, with the side window filtering technology by which most filters can improve their edge-preserving capabilities, we propose a general perceptual IR and VI image fusion framework with simple linear filter. Firstly, the source images are decomposed into edge feature components, hybrid components and base components by using linear filter and its side window version. Then, these components are combined by max-absolute fusion rule and improved max-absolute fusion rule. Finally, the fused image is reconstructed by adding all the fused components. In our experiments, two popular linear filters, i.e., box filter and Gaussian filter, are used to verify the effectiveness of the proposed framework. Experimental results show that the proposed fusion framework can obtain better perceptual fusion results than compared methods.

INDEX TERMS Image fusion, side window filtering, linear filter, infrared images, visible images.

I. INTRODUCTION

Visible (VI) images can reflect clear details and background scenery, which can lead to better situation awareness. Infrared (IR) images can present obvious thermal object information, which are conducive to object detection and recognition. Infrared (IR) and visible (VI) image fusion is an important part of multi-sensor informa-tion fusion, which benefits many applications [1]. Taking advantage of all the information of the source image is the main core of image fusion. However, as for IR and VI image fusion, integrating all the information of IR and VI images will make the fusion results visually unpleasing because they are two different phenomena of the same scene.

In the past decades, many IR and VI image fusion methods have been proposed, and we divide them into three categories, i.e., multi-scale transform (MST)-based methods, deep learning (DL)-based methods, and other methods. DL-based methods [2]–[6] have emerged in recent years. Obviously, these methods would achieve better results in the near future. As for IR and VI image fusion, DL-based methods have some issues to be addressed. As there are no ground truth images, defining the fusion results is not an easy task. In addition, training images are limited. Other methods include gradient transfer-based methods [7], [8], representation learning-based methods [9], [10], and etc. However, the fusion results of these methods are visually unpleasing.

It is well known that MST-based methods [11]–[16] are the most researched fusion methods. Their main process includes three steps, i.e., decomposition, fusion, and reconstruction. MST-based methods can achieve general results in image fusion domain because their multi-scale processing mechanism is consistent with human visual perception. Generally speaking, the average fusion rule is used to fuse low-pass components of the source images. However, it will lead to low-contrast and poor visual performance of the fused image. In order to retain the object information of the IR image and the background information of the VI image simultaneously, Fu et al. proposed a fusion method based on non-subsampled contourlet transform (NSCT) and robust principal component

The associate editor coordinating the review of this manuscript and approving it for publication was Yuhao Liu.

analysis (RPCA) [11]. They firstly decomposed the source images into low-pass components and high-pass components by NSCT. Then RPCA was used to extract the saliency information of the source images, and they used it to guide the fusion of low-pass and high-pass components. Finally, they reconstructed the fused image by inverse NSCT. However, their method cannot do well in the low-light circumstances.

Recently, edge-preserving filters such as bilateral filter (BF) [17], weighted least squares filter (WLSF) [18], and rolling guidance filter (RGF) [19] are widely used in MST-based methods [12]–[14] as these filters have good capabilities of edge-preserving. Ma et al. proposed an IR and VI image fusion method to address several common defects of conventional methods [13]. Firstly, the source images were decomposed by RGF and Gaussian filter. Then, a saliency-based fusion rule was used to fuse the base layers, and an optimization-based fusion rule was applied to fuse the de-tail layers. Finally, the fused image was a linear combination of the fused base layer and detail layers. To achieve perceptually better fusion results than conventional MST-based methods, Zhou et al. proposed a hybrid MST-based method [14]. They firstly used BF and Gaussian filter to decompose the source images into edge feature components and texture detail components. Then three different fusion rules were adopted to fuse these components. Finally, the fused image was reconstructed by adding all the fused components. Indeed, their method can achieve perceptually better fusion results, and thus leads to better situation awareness. However, most edge-preserving filters are complex, especially BF. In this paper, we demonstrate that just using linear filter can obtain perceptual fusion results for IR and VI images with the side window filtering (SWF) technology.

In CVPR 2019, Yin and Gong et al. proposed a side window filtering (SWF) framework [20]. Most linear filters and nonlinear filters can significantly improve their edge-preserving capabilities by using this framework. This good property motivates us that using simple linear filter, for example, box filter [21], can accumulate edge features of the source images with the SWF technology, and we can use these edge features to guide the fusion. Firstly, we use linear filter and its side window version to decompose the source images into edge feature components, hybrid components and base components. Then, max-absolute fusion rule and improved max-absolute fusion rule are applied to fuse these decomposed components. Finally, the fused components are used to reconstruct the fused image. The effectiveness of the proposed fusion framework is verified by using two linear filters which are box filter and Gaussian filter. The main contributions of our work are outlined below.

(1) SWF technology is introduced into IR and VI image fusion for the first time, by using which most filters can improve their edge-preserving capabilities which are conducive to image fusion. For example, as Gaussian filter has no edge-preserving capability, the fused images obtained by fusion methods using decomposition methods based on Gaussian filter generally suffer from halo artifacts around the strong edges, especially in IR and VI image fusion. This problem can be well addressed with the SWF technology.
(2) We propose a general perceptual IR and VI image fusion framework based on linear filter and SWF technology to achieve visually satisfactory fusion. To the best of our knowledge, such a perceptual fusion framework has not been studied yet.
(3) Edge-preserving filters have been widely used in IR and VI image fusion, but they are complex. This paper demonstrates that just using simple linear filter can obtain state-of-the-art performance for the fusion of IR and VI images with the SWF technology.

The rest of this paper is outlined as follows. We introduce multi-scale decomposition based on linear filter and SWF technology in Section II. We elaborate on the proposed fusion framework in Section III. Experiments are conducted in Section IV. Some conclusions are given in Section V.

II. MULTI-SCALE DECOMPOSITION BASED ON LINEAR FILTER AND SIDE WINDOW FILTERING TECHNOLOGY

In CVPR 2019, Yin and Gong et al. proposed a side window filtering (SWF) framework [20]. Under SWF framework, most filters can significantly improve their edge-preserving capabilities. Box filtering and Gaussian filtering can be expressed as (1) and (2), respectively.

$$I_{BOX} = BOX(I, r_{box}), \tag{1}$$
$$I_{GAU} = GAU(I, r_{gau}, \sigma_{gau}), \tag{2}$$

where I represents the input image, I_{BOX} and I_{GAU} represent the filtered output images of box filter and Gaussian filter respectively, r_{box} represents the radius of the box filter, r_{gau} and σ_{gau} represent the radius and standard deviation of the Gaussian filter respectively, and BOX and GAU represent the box filtering function and Gaussian filtering function, respectively. The SWF versions of box filtering and Gaussian filtering can be expressed as (3) and (4), respectively.

$$I_{S-BOX} = S - BOX(I, r_{s-box}, itenum), \tag{3}$$
$$I_{S-GAU} = S - GAU(I, r_{s-gau}, \sigma_{s-gau}, itenum), \tag{4}$$

where I_{S-BOX} and I_{S-GAU} represent the filtered output images of SWF versions of box filter and Gaussian filter respectively, r_{s-box} represents the radius of SWF version of box filter, r_{s-gau} and σ_{s-gau} represent the radius and standard deviation of SWF version of Gaussian filter respectively, $S-BOX$ and $S-GAU$ represent the corresponding filtering functions respectively, and $itenum$ represents the number of iterations.

Figs. 1 and 2 show the $3-level$ decomposition using box filter and its SWF version on a"Camp" image pair $\{IR, VI\}$, respectively.I^i_{BOX} and I^i_{S-BOX} $(i = 1, 2, 3, I \in \{IR, VI\})$ represent the $i-level$ filtered images obtained by box filter and its SWF version with $itenum_i$ $(i = 1, 2, 3)$, respectively.

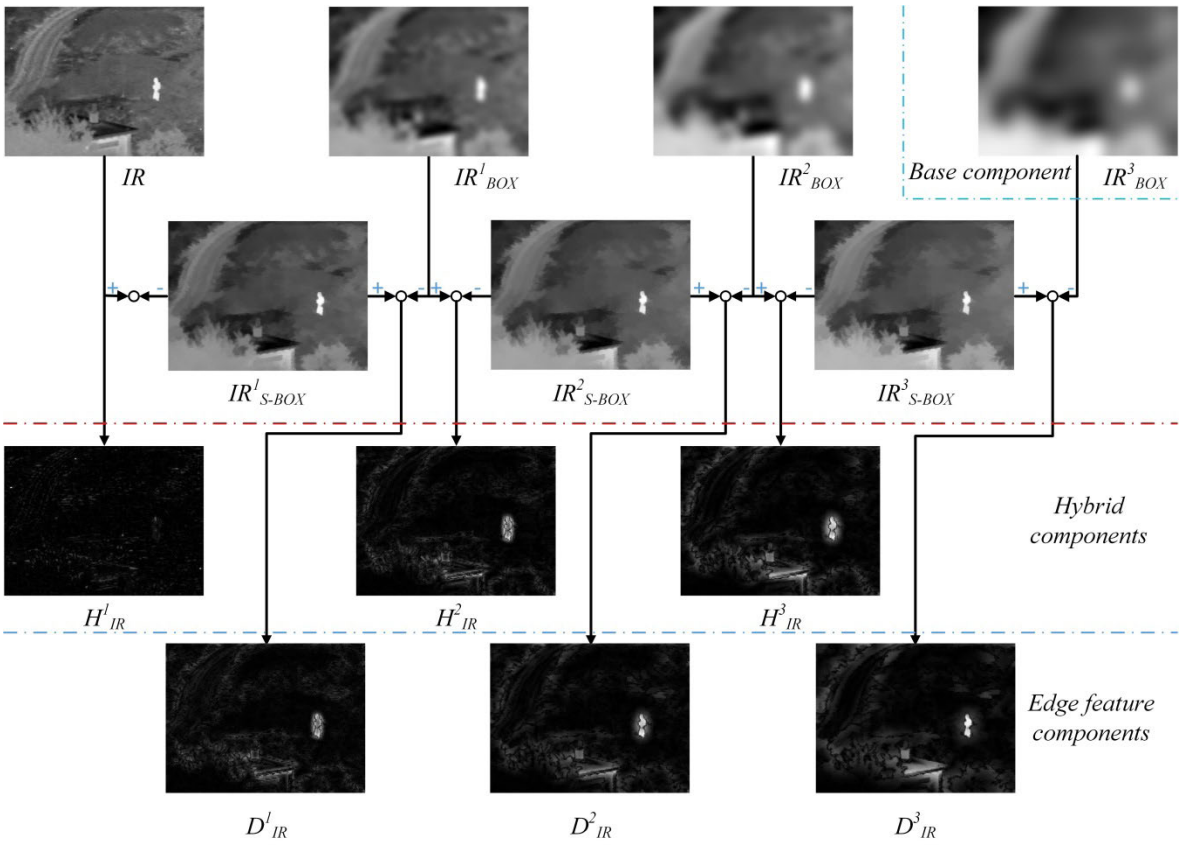


FIGURE 1. 3 – *level* decomposition of the "Camp" IR image using box filter and its SWF version, the grayscale of all the images is normalized to [0, 1] ($r_{box} = 7$, $r_{s-box} = 7$, $itenum_1 = 2$, $itenum_2 = 7$, and $itenum_3 = 12$).

It can be seen from Figs. 1 and 2 that the edge features and other details in the filtered images obtained by box filter are successively smoothed, but the edge features in the filtered images obtained by its SWF version are almost the same. Based on the above observation, we can accumulate the edge features by (5).

$$D_I^i = I_{S-BOX}^i - I_{BOX}^i \, (i = 1, 2, 3), \tag{5}$$

and the results are shown in Figs. 1 and 2. In addition, in Figs. 1 and 2, we also show the hybrid components which are obtained by (6). We named them hybrid components because they contain not only edge features but also other details. And IR_{BOX}^3 and VI_{BOX}^3 are the base components of the IR and VI image, respectively.

$$H_I^i = I_{BOX}^{i-1} - I_{S-BOX}^i \left(i = 1, 2, 3, I = I_{BOX}^0\right). \tag{6}$$

III. THE PROPOSED FUSION FRAMEWORK

Our proposed fusion framework mainly consists of three steps: decomposition, fusion, and reconstruction. Fig. 3 shows its schematic diagram. In the next, we use box filter and its SWF version to illustrate it. In addition to the parameters to be studied, the other parameter settings in Section III are consistent with Section IV.

A. DECOMPOSITION

As demonstrated in Figs. 1 and 2, we can decompose the source images $\{IR, VI\}$ into three parts, i.e., base components $\left\{B_{IR} = IR_{BOX}^L, B_{VI} = VI_{BOX}^L\right\}$, hybrid components $\left\{\sum_{i=1}^L H_{IR}^i, \sum_{i=1}^L H_{VI}^i\right\}$ and edge feature components $\left\{\sum_{i=1}^L D_{IR}^i, \sum_{i=1}^L D_{VI}^i\right\}$, L represents the decomposition level.

B. FUSION

As our goal is to achieve perceptual fusion results, it is appropriate to inject important edge feature components from the IR image into the VI image, which is demonstrated in [14]. Thus, we use edge feature components to guide the fusion. Firstly, the initial weights $V^i \, (i = 1, 2 \ldots L)$ are obtained by (7).

$$V^i = \begin{cases} \left|D_{IR}^i\right| - \left|D_{VI}^i\right|, & \left(\left|D_{IR}^i\right| - \left|D_{VI}^i\right|\right) > 0 \\ 0, & otherwise, \end{cases} \tag{7}$$

and we normalize it to [0, 1]. Secondly, a nonlinear function $S_\lambda(x)$ is used to adjust the values of $V^i \, (i = 1, 2 \ldots L)$ and the final weights $W^i \, (i = 1, 2 \ldots L)$ are obtained by (8).

$$W_i = GAU(S_\lambda(V_i), r, \sigma), \tag{8}$$

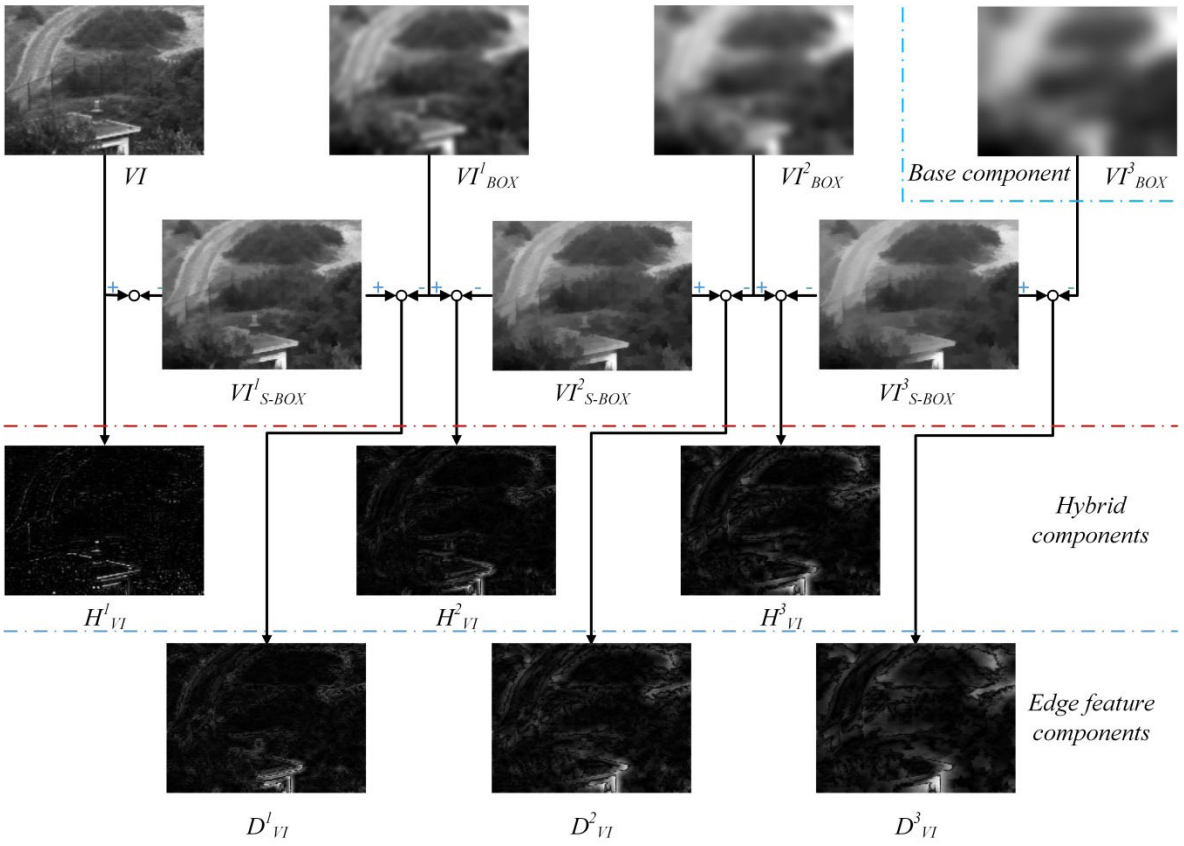
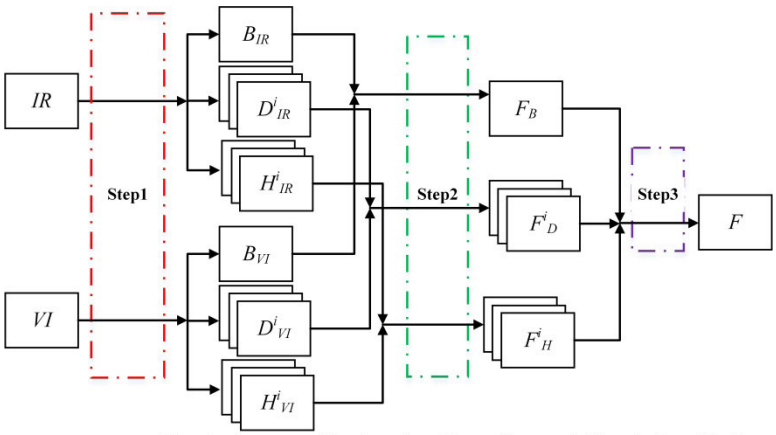


FIGURE 2. 3 – *level* decomposition of the "Camp" VI image using box filter and its SWF version, the grayscale of all the images is normalized to [0, 1] ($r_{box} = 7$, $r_{s-box} = 7$, $itenum_1 = 2$, $itenum_2 = 7$, and $itenum_3 = 12$).



Step1:multi-scale decomposition based on linear filter and side window filtering technology
Step2:fusion based on max-absolute fusion rule and improved max-absolute fusion rule
Step3:adding all the fused components
IR: the infrared image *VI*: the visible image *F*: the fused image
B_I: the base component of image *I*, *I* represents *IR or VI*
D^i_I: the *i*-level edge component of image *I*, *I* represents *IR or VI*
H^i_I: the *i*-level hybrid component of *I*, *I* represents *IR or VI*
F_B: the fused base component F^i_D: the fused *i*-level edge component
F^i_H: the fused *i*-level hybrid component

FIGURE 3. The schematic diagram of the proposed fusion framework.

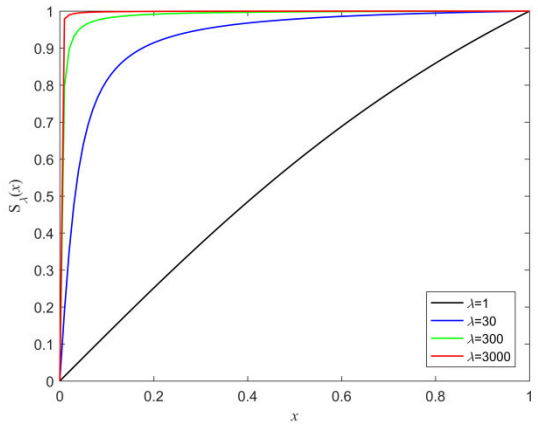


FIGURE 4. The plots of $S_\lambda(x)$ with different λ.

where r and σ represents the radius and standard deviation of the Gaussian filter, and we typically set $r = floor(3\sigma)$. Note that we use Gaussian filter to reduce noise and locally smooth $S_\lambda(V_i)$. $S_\lambda(x)$ is defined as (9).

$$S_\lambda(x) = \arctan(\lambda x)/\arctan(\lambda), \tag{9}$$

where $x \in [0, 1]$, and λ ($\lambda > 0$) is the function parameter.

The plots of $S_\lambda(x)$ with different λ are shown in Fig. 4. From Fig. 4, we can find that as λ increases, the function value $S_\lambda(x)$ of the same x increases. When $\lambda \to 0$, $S_\lambda(x) \to x$; when $\lambda \to \infty$, $S_\lambda(x) \to 1$, in this case, the fusion rule tends to be the max-absolute fusion rule. Thus,λ controls the amount of edge features injected into the fused image from the IR image. In order to better illustrate it, Fig. 5 and Fig. 6 show two fusion examples using box filter and its SWF version in the proposed fusion framework with different λ. Note that different λ will obtain different W_i for the same source images according to (7)-(9). From Fig. 5 and Fig. 6, we can find that as λ increases, the edge features labeled by red and green rectangles in Fig. 5(a) and Fig. 6(a) increases in the corresponding fused images. Finally, we can obtain the fused edge feature components according to (10).

$$F_D^i = W_i D_{IR}^i + (1 - W_i) D_{VI}^i \ (i = 1, 2 \ldots L). \tag{10}$$

The fused hybrid components are obtained by

$$\begin{cases} F_H^i = W_i H_{IR}^i + (1 - W_i) H_{VI}^i & (i = 2, 3, \ldots L) \\ F_H^i = \begin{cases} H_{IR}^i, & |H_{IR}^i| > |H_{VI}^i| \\ H_{VI}^i, & otherwise, \end{cases} & (i = 1) \end{cases} \tag{11}$$

In the next, we will explain why we use the max-absolute fusion rule instead of the weights W_1 to fuse the hybrid components at level $i = 1$. Fig. 7 shows the hybrid compo-nents and edge feature components of the IR and VI image at level $i = 1$. From Fig. 7, we can see that if we use the weights W_1 obtained by D_{IR}^1 and D_{VI}^1 to fuse H_{IR}^1 and H_{VI}^1, the small bright dots in the IR image will be lost to some extent. Fig. 8 shows the fusion results obtained by using the weights W_1 and the max-absolute fusion rule. As demonstrated in Fig. 8, the small bright dots are kept well in Fig. 8(d). While in Fig. 8(c), the two small bright dots in the top left corner (labeled by blue rectangles) are a little darker, and the small bright dot in the lower right corner (labeled by red rectangle) is lost.

The fused base component is obtained by (12) and (13).

$$W_B = GAU(S_\lambda(V_L), r_B, \sigma_B), \tag{12}$$

$$F_B = W_B B_{IR} + (1 - W_B) B_{VI}, \tag{13}$$

where $\sigma_B = 4\sigma_{gau}^L$,and $r_B = floor(3\sigma_B)$. We set $\sigma_B = 4\sigma_{gau}^L$ for the following consideration. As shown in Fig. 9, Fig. 9(a) and Fig. 9(b) tend to be smoother than Fig. 9(c) and Fig. 9(d), respectively, thus we set a larger σ value for base component fusion. Fig. 10(c) and Fig. 10(d) show the fusion results obtained by setting $\sigma_B = \sigma$ and $\sigma_B = 4\sigma_{gau}^L$, respectively. Obviously, $\sigma_B = 4\sigma_{gau}^L$ is more appropriate than $\sigma_B = \sigma$.

C. RECONSTRUCTION

The final fused image F can be reconstructed according to (14).

$$F = F_B + \sum_{i=1}^{L} F_D^i + \sum_{i=1}^{L} F_H^i \ (i = 1, 2 \ldots, L). \tag{14}$$

IV. EXPERIMENTS

This section contains four parts including instructions about the experiments, analysis of parameters, subjective results and objective results compared with several fusion methods.

A. INSTRUCTIONS ABOUT THE EXPERIMENTS

1) DATASET

Twenty one pairs of images widely used in IR and VI fusion field shown in Fig. 11 are adopted in our experiments, and most of them are downloaded from the TNO dataset [22]. Image pairs 11(i), (l), (m)-(n),and (q) are provided by Zhang, the first author of [23].

2) COMPARED METHODS AND PARAMETER SETTINGS

In order to verify the effectiveness of our proposed framework, five fusion methods, which can achieve state-of-the-art results, are selected to compare with our fusion framework. They are based on NSCT and RPCA (NSCT-RPCA) [11], gradient transfer fusion (GTF) [7], latent low-rank representation(LATLRR) [10], hybrid multi-scale decomposition (HMSD) [14], and ResNet and zero-phase component analysis (RESNET) [4]. The methods LATLRR and RESNET were proposed in just one year. The parameters of these five methods are set according to the corresponding original papers. In our fusion framework, we use two filters, which are box filter and Gaussian filter, to test in experiments. For convenience, the methods using the box filter and the Gaussian filter in our fusion framework are named HBOX and HGAU, respectively. The parameter settings of our method in Section IV are as follows: for box filter,

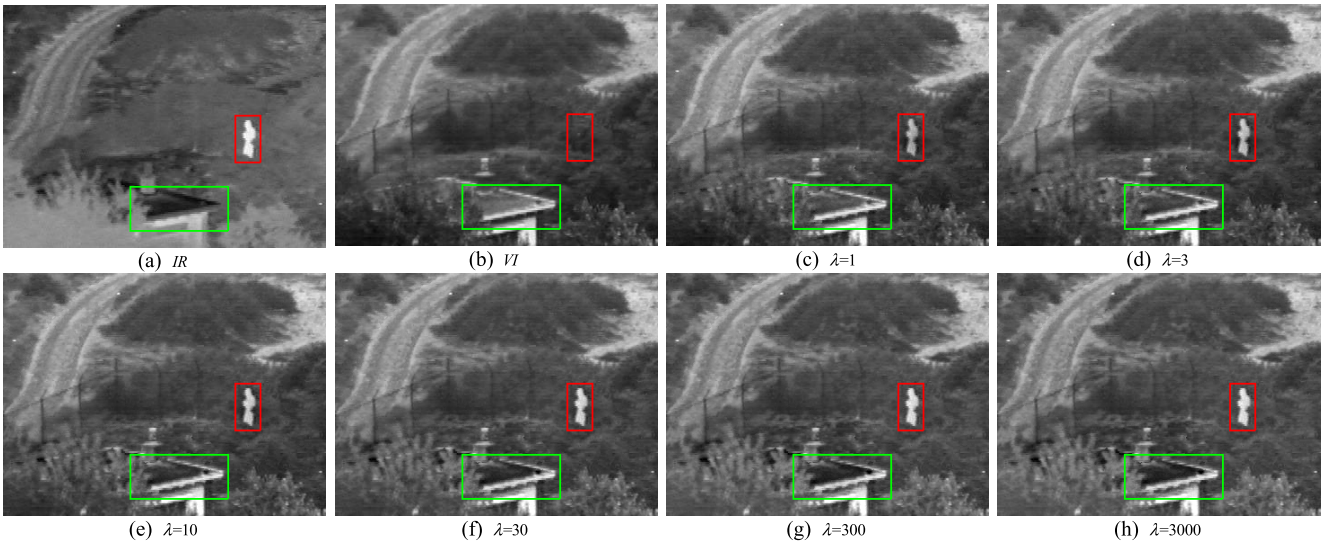


FIGURE 5. (a) and (b) are the IR and VI image respectively, (c) and (h) are the fusion results obtained by the proposed fusion framework using box filter and its SWF version with different λ, respectively.

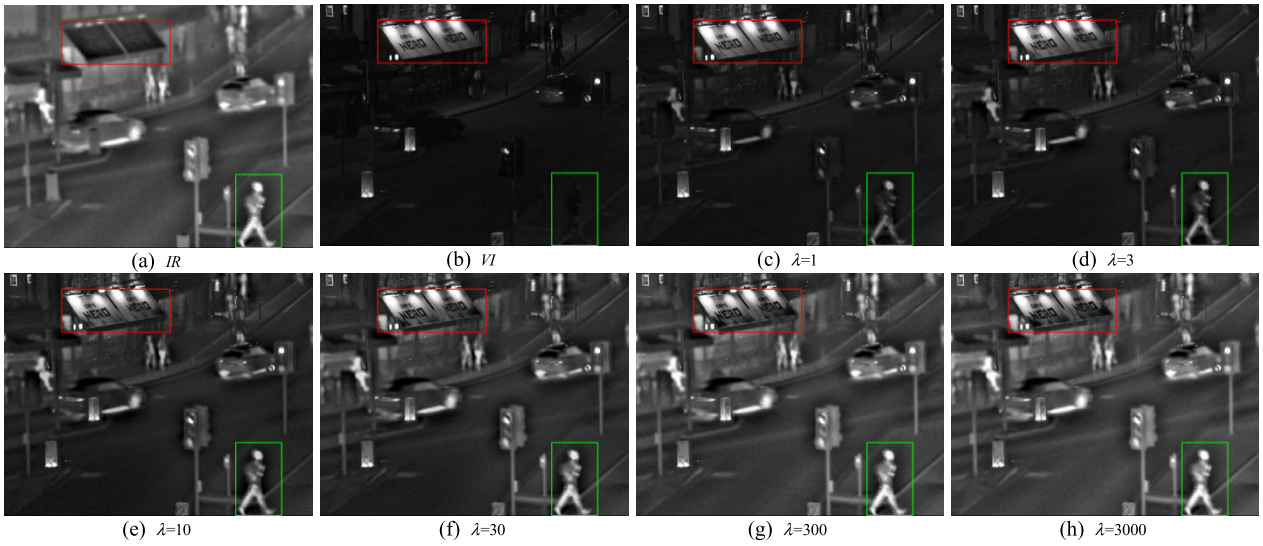


FIGURE 6. (a) and (b) are the IR and VI image respectively, (c) and (h) are the fusion results obtained by the proposed fusion framework using box filter and its SWF version with different λ, respectively.

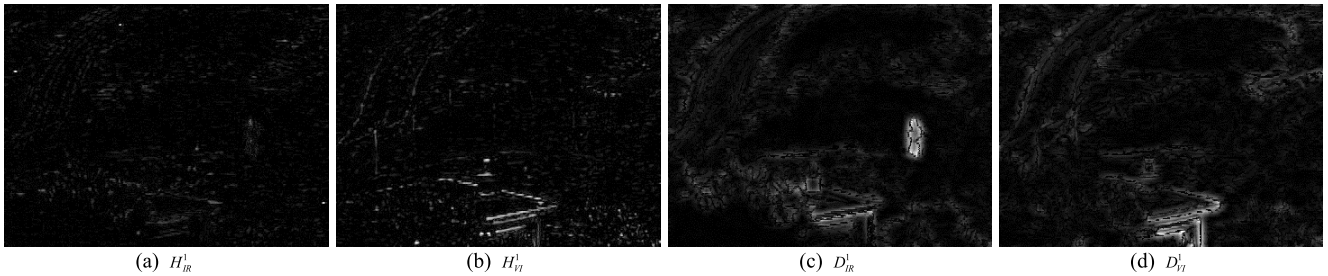


FIGURE 7. (a) and (b) are the hybrid components of the IR and VI image at level $i = 1$ respectively, (c) and (d) are the edge feature components of the IR and VI image at level $i = 1$, respectively.

$r_{box} = r_{s-box} = 7$; for Gaussian filter, $r_{gau} = r_{s-gau} = 7$, $\sigma_{gau}^{i+1} = \sigma_{s-gau}^{i+1} = 2\sigma_{gau}^{i} = 2\sigma_{s-gau}^{i}$ $(i = 1, 2, 3)$, and $\sigma_{gau}^{1} = \sigma_{\text{s-}gau}^{1} = 2$; and some identical parameters: $L = 4$, $itenum_i = 2 + s\,(i-1)$ $(i = 1, 2, 3, 4)$, $\lambda = 30$, $r = 3$, $\sigma = 1$, $\sigma_B = 4\sigma_{gau}^{L}$, and $r_B = floor\,(3\sigma_B)$, here s represents the step size of iteration number and we set it to 5. Note that

FIGURE 8. (a) and (b) are the IR and VI image respectively, (c) and (d) are the fusion results obtained by using the weights W_1 and the max-absolute fusion rule, respectively.

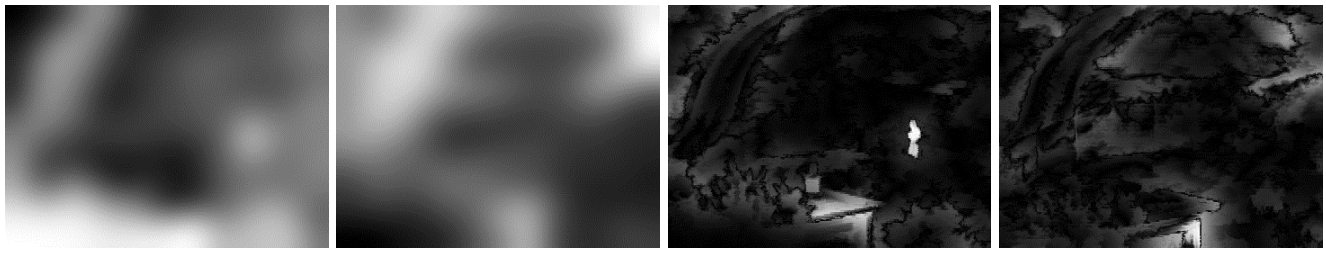


FIGURE 9. (a) and (b) are the L – level box filtered image of IR and VI image respectively, (c) and (d) are the L – level edge feature components of IR and VI image, respectively.

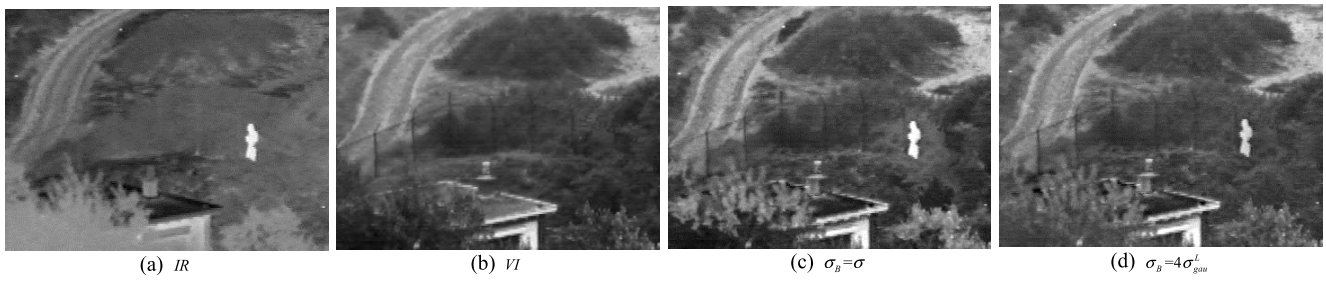


FIGURE 10. (a) and (b) are the IR and VI image respectively, (c) and (d) are the fusion results obtained by setting $\sigma_B = \sigma$ and $\sigma_B = 4\sigma^L_{gau}$, respectively.

the parameters $r_{box} = r_{s-box}$, $r_{gau} = r_{s-gau}$, L, and s are analyzed in Section IV-B, and other parameters are set to their empirical values [14].

3) EVALUATION METRICS

As there is no evidence that any image fusion evaluation metric is better than the others, four evaluation metrics are selected to comprehensively evaluate different methods. They are correlation-based metric (*SCD*) [24], structural similarity-based metric(*MS_SSIM*) [25], visual information fidelity-based metric (*VIFF*) [26], and human visual system (HVS) models-based metric (Q_{CB}) [27]. Metric *SCD* measures the sum of correlations of differences between the source images and the fused image. Metric *MS_SSIM* measures the amount of structural information transferred from the source images to the fused image. Metric *VIFF* measures the visual information fidelity between the source images and the fusion result. Metric Q_{CB} is closely matched to human perceptual evaluations. As our goal is to achieve percep-tually better fusion results, the four metrics are appropriate. For all the four metrics, a larger value means a better performance. The codes for them are publicly available and we keep the values of all the parameters in them unchanged. In addition, we also use a metric named *Sum* to comprehensively evaluate each method by cumulatively summing the score of each metric of each method. More details about *Sum* can be found in [28].

B. ANALYSIS OF PARAMETERS

In this subsection, we analyze the influence of three parameters on the fusion performance of our proposed framework, which are the decomposition level L, the filtering radius r(for box filter, $r = r_{box} = r_{s-box}$; for Gaussian filter, $r = r_{gau} = r_{s-gau}$), and the step size of iteration number s.

As demonstrated in [14],[29],[30], when the decomposition level L is too large, mis-registration will lead to artificial effects in the fusion results, especially in multi-focus image fusion. As we focus on IR and VI image fusion and most source images are all downloaded from the TNO dataset [22],

FIGURE 11. Source images used in our experiments.

which are all well registered, we do not consider this problem in this paper. In addition, as L increases, the computation load increases. On the other hand, when the decomposition level L is too small such as 1 or 2, enough spatial details cannot be extracted well [29], [30]. Fig. 12(a)–(e) show the fusion results obtained by HBOX with L from 1 to 5, respectively. It can be found that when L is equal to 1or 2, the mountain silhouette is not fused well; when L is larger than 3, the mountain silhouettes in Fig. 12(c)–(e) have no obvious difference. However, the larger the L, the higher the computational cost. Thus, a compromise on L should be made for the considering of extracting the spatial details and the computation cost. In most MST-based methods [14], [29], [30], L is recommended to set to 4.

For the filtering radius r, when the value is too large, the spatial details cannot be addressed well; when the value is too small, the thermal object regions which usually have certain size may not be fused well. Fig. 11(f)–(j) show the fusion results obtained by HBOX with r equal to 3, 5, 7, 9, 11, respectively. It can be found that the larger the r, the more blurred the windows labeled by green rectangles, which are enlarged in the lower left corner, and the more highlighted the thermal objects labeled by red rectangles, which are enlarged in the lower right corner. Thus, a compromise on r should be made for the considering of extracting the spatial details and highlighting the thermal objects. In [20], r is recommended to set to 7.

itenum is recommended to set to 10 in [20]. As our algorithm should accumulate the edge features, we increase *itenum* by step size s, and it is not necessary to set the step size of iteration number s to a larger value considering that the decomposition level L is usually set to a value larger

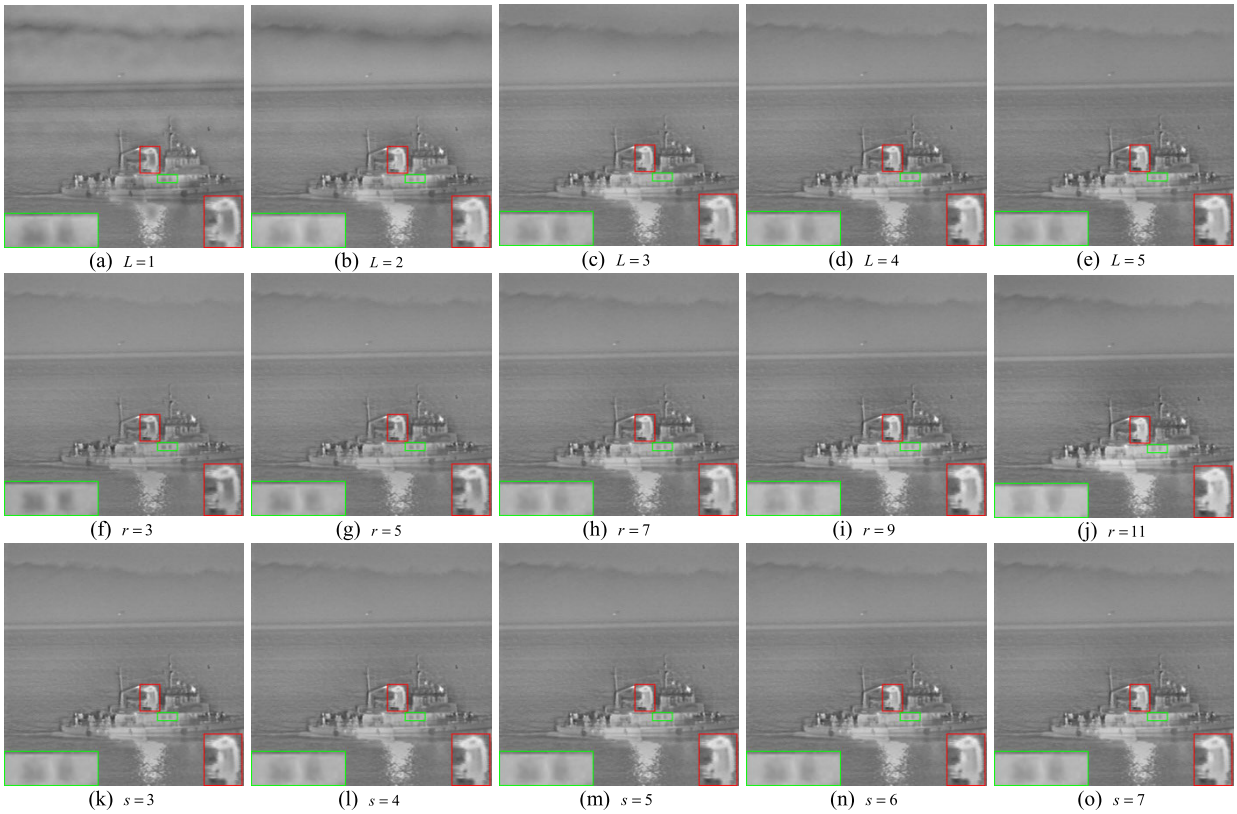


FIGURE 12. (a) and (e) are the fusion results obtained by H-BOX with different *L* respectively, (f) and (j) are the fusion results obtained by H-BOX with different *r* respectively, (k) and (o) are the fusion results obtained by H-BOX with different *s* respectively.

than 3 in most MST-based methods. Moreover, the larger the s, the higher the computation cost. Fig. 12(k)–(o) show the fusion results obtained by HBOX with s from 3 to7, respectively. It can be found that the results obtained using different s have no obvious difference. Considering that the larger the s, the higher the computation cost, we set s to 5.

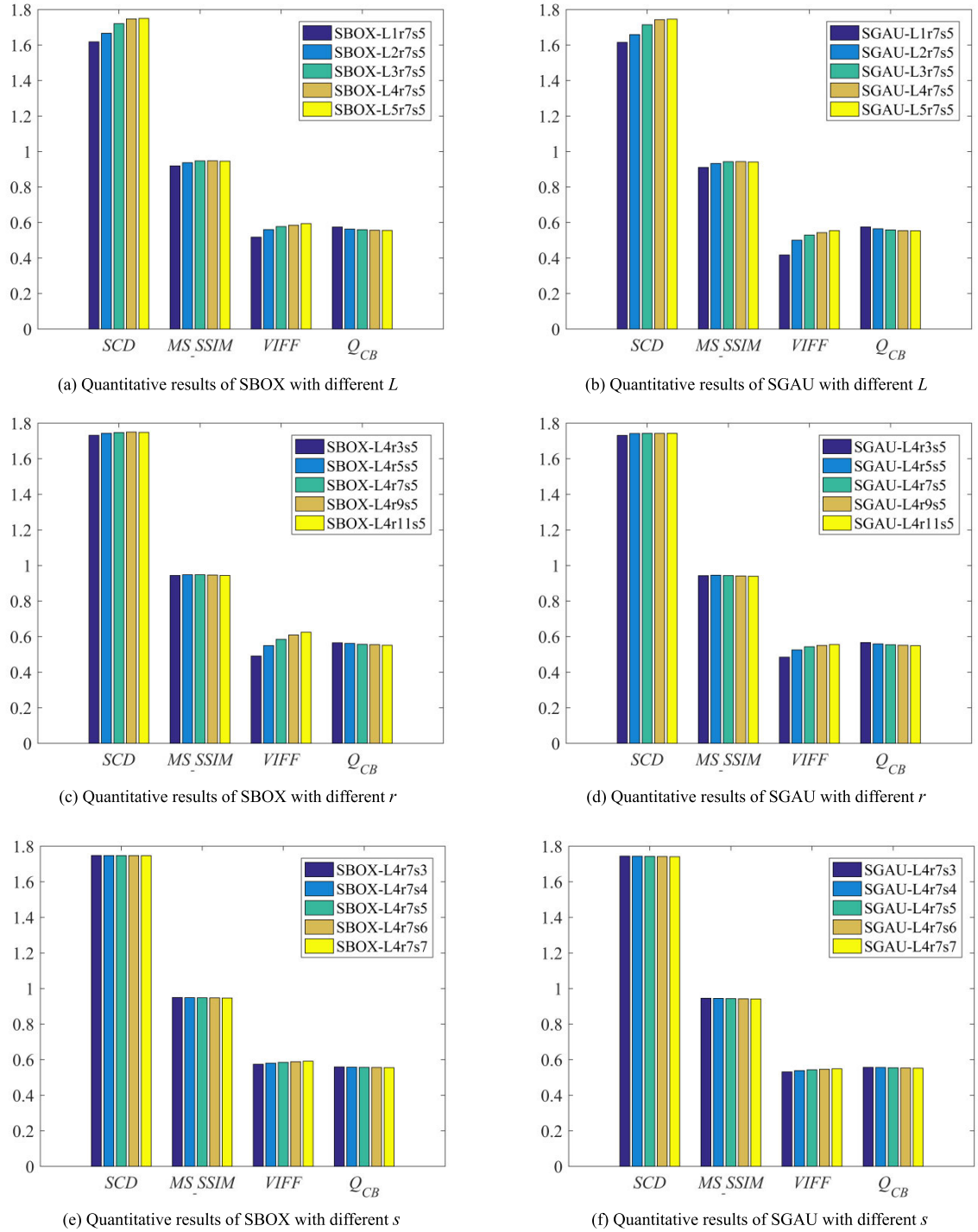
In the next, we experimentally studied the influence of these three parameters on the fusion performance of our proposed algorithm through quantitative comparisons using the above five evaluation metrics. For convenience, we name the method M (M represent HBOX or HGAU) with $L = 4$, $r = 7$, and $s = 5$ as ML4r7s5. For example, HBOXL4r7s5 stands for HBOX with parameters setting: $L = 4$, $r = 7$, and $s = 5$. When analyzing the effect of one parameter, we set the other two parameters to constant. For example, when analyzing the effect of L, we set the other two parameters to $r = 7$ and $s = 5$. Fig. 13 shows the quantitative results of SBOX and SGAU with different parameter combinations, and all the metric values in Fig. 13 are the average values on the twenty one pairs of images shown in Fig. 11. From Fig. 13(a) and (b), we can find that $L = 4$ and $L = 5$ are two more appropriate choices for both SBOX and SGAU. As shown in Fig. 13(c) and (d), $r = 7$ and $r = 9$ are two moderate choices for both SBOX and SGAU. From Fig. 13(e) and (f), we can find the fusion performance of both SBOX and SGAU is not sensitive to s. Combining the above analysis, in Section IV we set the three parameters as follows: $L = 4$, $r = 7$, and $s = 5$.

C. SUBJECTIVE RESULTS

We take two widely used image pairs, i.e., ''Octec'' and ''Road'', as examples to demonstrate the effectiveness of the proposed fusion framework, and the results are shown in Fig. 14 and Fig. 15, respectively. Some regions of the images are labeled with red or green rectangles, and some of them are enlarged for better comparisons.

Fig. 14(a) and (b) present two source images which were mainly captured under normal-light circumstance. Fig. 14(c)–(i) show the fusion results obtained by NSCT-RPCA, GTF, LATLRR, HMSD, RESNET, HBOX, and HGAU, respectively. Compared with GTF, LATLRR, and RESNET, our methods perform well in merging the ground. Compared with NSCT-RPCA and HMSD, our methods perform well in merging the windows labeled by red rectangle. In addition, the person in the fusion result obtained by NSCT-RPCA is a little darker than it in our fusion results.

Fig. 15(a) and (b) present two source images which were captured under low-light circumstance. Fig. 15(c)–(i) show the fusion results obtained by NSCT-RPCA, GTF, LATLRR, HMSD, RESNET, HBOX, and HGAU, respectively. The background of the fusion result obtained by NSCT-RPCA is too dark. The fusion result obtained by GTF is



(a) Quantitative results of SBOX with different L

(b) Quantitative results of SGAU with different L

(c) Quantitative results of SBOX with different r

(d) Quantitative results of SGAU with different r

(e) Quantitative results of SBOX with different s

(f) Quantitative results of SGAU with different s

FIGURE 13. Quantitative results of SBOX and SGAU with different parameter combinations.

over-smoothed. The fusion result obtained by RESNET has a low-contrast. Our methods perform better in merging the windows labeled by red rectangle than LATLRR. Compared with HMSD, our methods perform better in merging the part labeled by green rectangle.

All in all, the proposed fusion framework can achieve perceptual fusion results with more useful information. Moreover, the fusion results obtained by HBOX and HGAU on all source image pairs are given in the supplementary material file.

D. OBJECTIVE RESULTS

The quantitative results on "Octec", and "Road" image pairs are shown in Table 1 and Table 2 respectively, and the average quantitative results on all the image pairs are presented in Table 3. In Tables 1–3, the values in bold indicate the

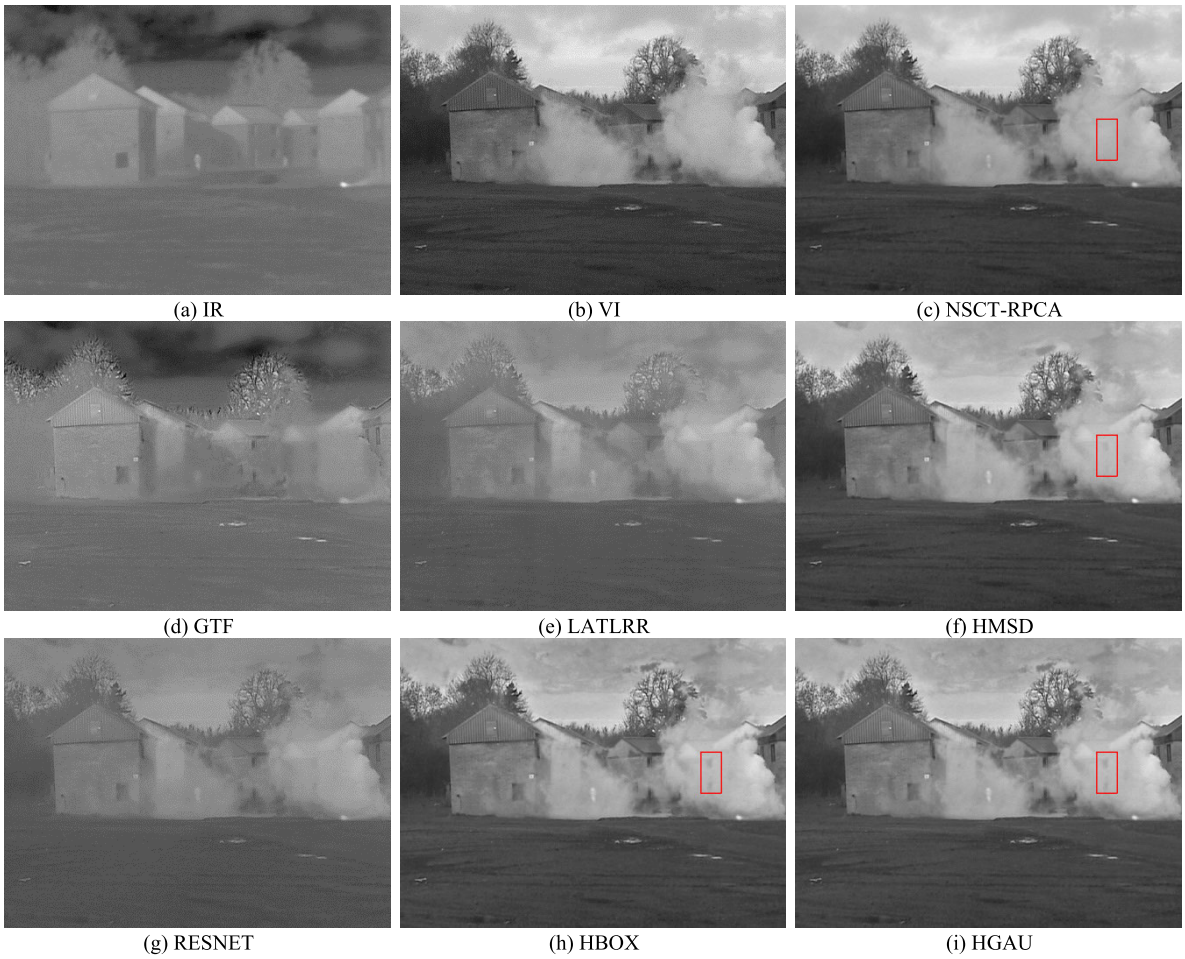


FIGURE 14. Fusion results on the "Octec" source image pair.

TABLE 1. Quantitative results on the "Octec" source image pair.

	NSCT-RPCA	GTF	LATLRR	HMSD	RESNET	HBOX	HGAU
SCD	1.2849	0.8302	1.6301	**1.8343**	1.5814	1.8061	1.8142
MS_SSIM	0.9103	0.8380	0.8759	0.9445	0.8916	**0.9563**	0.9554
VIFF	0.4574	0.1564	0.2467	0.5390	0.2610	**0.6283**	0.6049
Q_{CB}	**0.6181**	0.3673	0.4868	0.6052	0.5517	0.5849	0.5889
Sum	17	4	10	**23**	12	**23**	**23**

TABLE 2. Quantitative results on the "Road" source image pair.

	NSCT-RPCA	GTF	LATLRR	HMSD	RESNET	HBOX	HGAU
SCD	1.7216	1.1085	**1.7387**	1.7275	1.4762	1.7080	1.7057
MS_SSIM	0.8968	0.8953	0.9041	0.9332	0.8889	**0.9518**	0.9484
VIFF	0.5648	0.3992	0.4847	0.6332	0.3634	**0.6923**	0.6581
Q_{CB}	0.4209	0.3643	0.4756	0.4800	0.4883	**0.5022**	0.4965
Sum	14	6	17	20	9	**25**	21

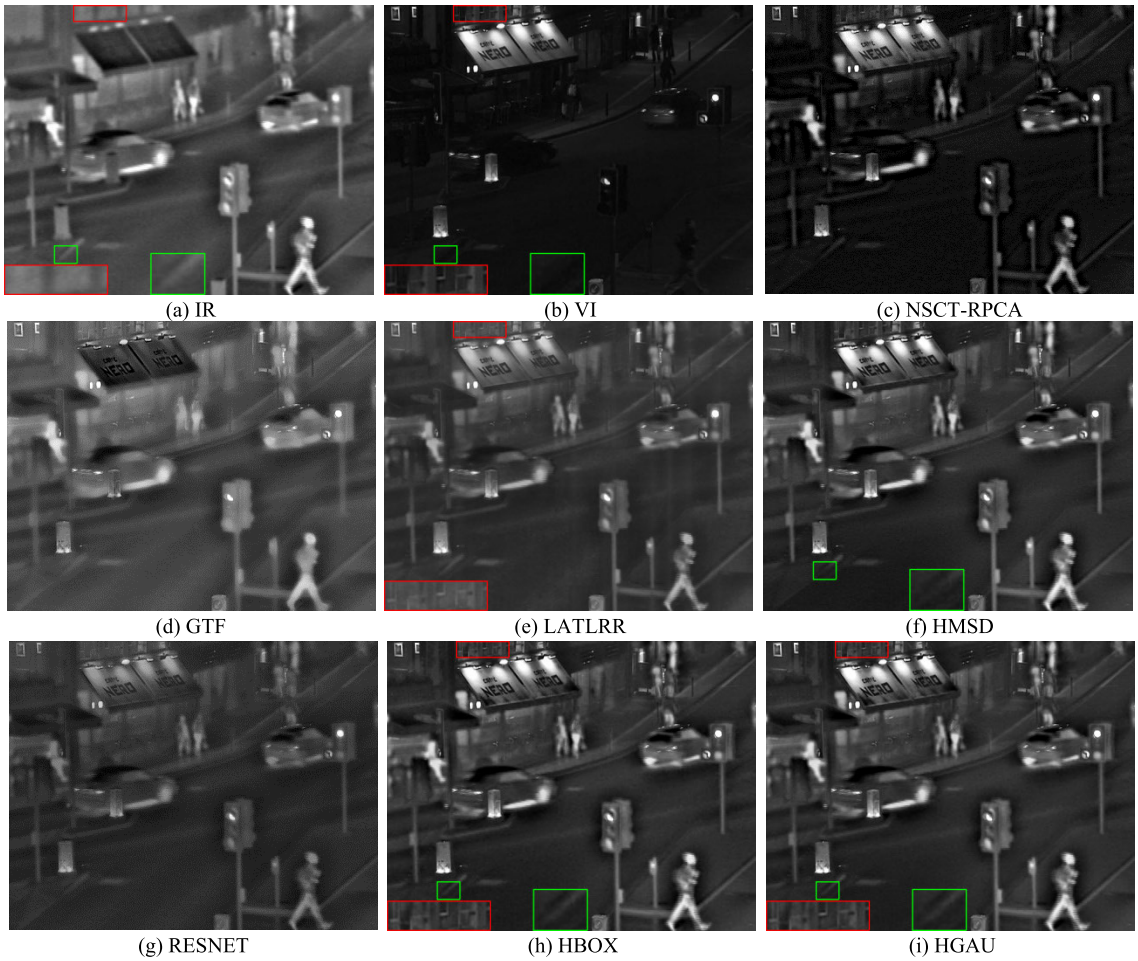


FIGURE 15. Fusion results on the "Road" source image pair.

TABLE 3. Average quantitative results on all the source image pairs.

	NSCT-RPCA	GTF	LATLRR	HMSD	RESNET	HBOX	HGAU
SCD	1.5994	0.9956	1.6857	1.7024	1.5708	**1.7476**	1.7429
MS_SSIM	0.9249	0.8093	0.8857	0.9339	0.8720	**0.9479**	0.9435
VIFF	0.4771	0.2138	0.3587	0.5045	0.2890	**0.5852**	0.5431
Q_{CB}	0.5511	0.4217	0.4970	0.5539	0.5014	**0.5570**	0.5545
Sum	15	4	12	20	9	**28**	24

best for each metric, and the values in red indicate the second for each metric. It can be seen from Table 1 that our methods outperform other methods in terms of MS_SSIM and $VIFF$. We can see from Table 2 that our methods outperform other methods in term of MS_SSIM, $VIFF$, and Q_{CB}. Although the performance of the four metrics is not exactly the same for different images, our method has certain advantages considering the average results on all the source image pairs, which can be seen from Table 3. In addition, the performance of metric *Sum* in Tables 1–3 further verifies the effectiveness of our methods. The quantitative results demonstrate that our fusion results can retain more useful information, and are more consistent with human visual perception. Moreover, the quantitative results on all the source image pairs are given in the supplementary material file.

V. CONCLUSION

In this paper, we propose a general perceptual fusion framework with linear filter and side window filtering (SWF) technology. In our framework, the source images are firstly decomposed by linear filter and its SWF version into edge feature components, hybrid components, and base components. Then, max-absolute fusion rule and improved max-absolute fusion rule are used to merge these compo-nents.

Finally, the fusion result is obtained by adding all the fused components. In our experiments, we use two linear filters, i.e., box filter and Gaussian filter, to validate the effectiveness of the proposed fusion framework. The qualitative results indicate that the proposed fusion framework can achieve perceptual fusion results with more useful information. The quantitative results demonstrate that our fusion results can retain more useful information, and are more consistent with human visual perception.

Some limitations of our fusion framework are as follows: (1) There are many parameters in our proposed framework. In Section IV-B, although the parameter settings of the proposed framework are given experimentally, they do not take into account the correlation between the parameters. In recent years, due to the rise of deep learning, it will bring new ideas to solve this problem. (2) The statistical difference between different methods is unknown. Therefore, in our future work, statistical significance tools (such as non-parametric Friedman test) can be used to analyze them [31], [32]. In addition, since multi-modal medical image fusion has certain similarities with IR and VI image fusion, it is also a future research work whether the proposed framework is suitable for multi-modal medical image fusion.

REFERENCES

[1] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[2] 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, Jan. 2018, Art. no. 1850018.

[3] 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.

[4] 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.

[5] 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.

[6] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, and J. Wu, "Infrared and visible image fusion via detail preserving adversarial learning," *Inf. Fusion*, vol. 54, pp. 85–98, Feb. 2020.

[7] 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.

[8] Y. Ma, J. Chen, C. Chen, F. Fan, and J. Ma, "Infrared and visible image fusion using total variation model," *Neurocomputing*, vol. 202, pp. 12–19, Aug. 2016.

[9] C. H. Liu, Y. Qi, and W. R. Ding, "Infrared and visible image fusion method based on saliency detection in sparse domain," *Infr. Phys. Technol.*, vol. 83, pp. 94–102, Jun. 2017.

[10] H. Li and X.-J. Wu, "Infrared and visible image fusion using Latent Low-Rank Representation," unpublished.

[11] Z. Fu, X. Wang, J. Xu, N. Zhou, and Y. Zhao, "Infrared and visible images fusion based on RPCA and NSCT," *Infr. Phys. Technol.*, vol. 77, pp. 114–123, Jul. 2016.

[12] W. Gan, X. Wu, W. Wu, X. Yang, C. Ren, X. He, and K. Liu, "Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter," *Infr. Phys. Technol.*, vol. 72, pp. 37–51, Sep. 2015.

[13] 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.

[14] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Inf. Fusion*, vol. 30, pp. 15–26, Jul. 2016.

[15] X. Luo, Z. Zhang, B. Zhang, and X. Wu, "Image fusion with contextual statistical similarity and nonsubsampled shearlet transform," *IEEE Sensors J.*, vol. 17, no. 6, pp. 1760–1771, Mar. 2017.

[16] B. Cheng, L. Jin, and G. Li, "General fusion method for infrared and visual images via latent low-rank representation and local non-subsampled shearlet transform," *Infr. Phys. Technol.*, vol. 92, pp. 68–77, Aug. 2018.

[17] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.

[18] Z. Farbman, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *Trans. Graph.*, vol. 27, no. 3, p. 67, Aug. 2008.

[19] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 815–830.

[20] H. Yin, Y. Gong, and G. Qiu, "Side window filtering," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 8758–8766.

[21] Y. Gong, B. Liu, X. Hou, and G. Qiu, "Sub-window box filter," in *Proc. IEEE Vis. Commun. Image Process.*, Dec. 2018, pp. 1–4.

[22] A. Toet. (2015). *TNO Image Fusion Dataset*. [Online]. Available: http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029

[23] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visual image fusion through infrared feature extraction and visual information preservation," *Infr. Phys. Technol.*, vol. 83, pp. 227–237, Jun. 2017.

[24] 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.

[25] 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.

[26] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion*, vol. 14, pp. 127–135, Apr. 2013.

[27] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

[28] J. Du, W. Li, X. Bin, and N. Qamar, "Medical image fusion by combining parallel features on multi-scale local extrema scheme," *Knowl.-Based Syst.*, vol. 113, pp. 4–12, Dec. 2016.

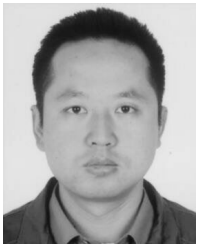
[29] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.

[30] 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.

[31] J. Du, W. Li, and B. Xiao, "Anatomical-functional image fusion by information of interest in local laplacian filtering domain," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5855–5866, Dec. 2017.

[32] J. Du, W. Li, and H. Tan, "Intrinsic image decomposition-based grey and pseudo-color medical image fusion," *IEEE Access*, vol. 7, pp. 56443–56456, 2019.

HUIBIN YAN was born in Jiangxi, China, in 1993. He received the B.S. degree from Wuyi University. He is currently pursuing the M.S. degree with the School of Information Engineering, Nanchang Hangkong University. His current interests include image fusion and saliency detection.



ZHONGMIN LI received the Ph.D. degree in communication and information system from Wuhan University, in 2008. He is an Assistant Professor with Nanchang Hangkong University. He is the author of more than 30 articles. His current research interests include target detection and tracking, image fusion, image retrieval, and wireless communication.