

## Research Article

# Infrared and Visible Image Fusion with Hybrid Image Filtering

Yongxin Zhang <sup>1</sup>, Deguang Li,<sup>1</sup> and WenPeng Zhu<sup>2</sup>

<sup>1</sup>*School of Information Technology, Luoyang Normal University, Luoyang 471934, China*

<sup>2</sup>*School of Information Science, Northwest University, Xi'an 710127, China*

Correspondence should be addressed to Yongxin Zhang; [tabo126@126.com](mailto:tabo126@126.com)

Received 13 March 2020; Revised 7 July 2020; Accepted 8 July 2020; Published 29 July 2020

Academic Editor: Peng-Yeng Yin

Copyright © 2020 Yongxin Zhang 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 an important technique aiming to generate a composite image from multiple images of the same scene. Infrared and visible images can provide the same scene information from different aspects, which is useful for target recognition. But the existing fusion methods cannot well preserve the thermal radiation and appearance information simultaneously. Thus, we propose an infrared and visible image fusion method by hybrid image filtering. We represent the fusion problem with a divide and conquer strategy. A Gaussian filter is used to decompose the source images into base layers and detail layers. An improved co-occurrence filter fuses the detail layers for preserving the thermal radiation of the source images. A guided filter fuses the base layers for retaining the background appearance information of the source images. Superposition of the fused base layer and fused detail layer generates the final fusion image. Subjective visual and objective quantitative evaluations comparing with other fusion algorithms demonstrate the better performance of the proposed method.

## 1. Introduction

Image fusion is an important technique of image enhancement that extracts different salient feature information from numerous images into one full enhanced image for increasing the amount of information and utilization of the image. In recent years, image fusion technology has been applied in several aspects such as multifocus, medical, remote sensing, infrared, and visible images [1], especially in the merging of infrared and visible images. The image generated by the infrared image according to the principle of thermal imaging has high contrast and mainly provides the saliency target information of the fused image, and the visible image mainly includes the accurate background information. The saliency target in the infrared image is important for target recognition, while the background texture data in the visible image are the key to environmental analysis and detail judgment. Infrared and visible image fusion provides more comprehensive information, which has important practical significance in military and civilian fields [2].

At present, researchers have proposed large amount of approaches for infrared and visible image fusion [3]. Spatial

and transform domain-based approaches are two popular approaches in this area. The spatial domain-based approaches mainly generate weights according to the characteristics of the original spatial information of the pixels or regions in the source image. These methods are usually simple and fast, but the edge blur is easy to occur. The transform domain-based method mainly includes two parts: image decomposition and fusion rules. Multiscale decomposition tools decompose the source images into different scale spaces for obtaining layers with different feature information. Fusion rules depending on the information characteristics of each layer guide the fusion of different layers. Then, the fusion result is obtained by the inverse operation of multiscale decomposition [2]. The transform domain-based methods have become a research hotspot for its good adaptability to the human visual system. Traditional multiscale transform methods have achieved good fusion results, such as pyramid [4], discrete wavelet [5], contourlet, non-subsampled contourlet [6], and non-subsampled shearlet transforms [7]. However, a large number of transform coefficients resulting in the complexity of parameter optimization compromise the fusion performance.

In recent years, deep learning has been utilized in the modeling of complicated relationship between data and extraction of distinctive features [2, 3]. The methods based on deep learning such as convolutional neural networks [8–10], adversarial network [11–14], and dictionary learning-based sparse representation [15] have achieved better fusion performance. Ma et al. proposed a new end-to-end model, termed as the dual-discriminator conditional generative adversarial network, for fusing infrared and visible images of different resolutions [16]. Chen et al. proposed a target-enhanced multiscale transform decomposition model for infrared and visible image fusion to simultaneously enhance the thermal target in infrared images and preserve the texture details in visible images [17]. Xu et al. presented a new unsupervised and unified densely connected network for infrared and visible image fusion [18]. Zhang et al. proposed a fast unified image fusion network for infrared and visible images based on proportional maintenance of gradient and intensity [19]. They are good at feature extraction and data reproduction. But the difficulties of ideal image selection and training, learning parameter settings, and the domain knowledge may compromise the fusion quality [8]. Despite the flexibility, rigidity, and robustness of the conventional infrared and visible image fusion approaches, some improvements could be attained in this area, yet. The current study concentrates on these improvements.

As is known, conventional transform domain-based methods suffer from complication of parameter optimization and cost in coefficient processing. Thus, an edge-preserving filter with spatial consistency and edge retention are introduced to image fusion. Edge-preserving filtering is an effective tool for image fusion, which enhances the edge information of the image and reduces artifacts around the edges. Local filter, global optimization, and hybrid filter based techniques are the three main methods in this area [2].

Bilateral filter [20], cross bilateral filter [21], iterative guided filtering [22], guided filtering (GFF) [23], gradient domain-guided filtering (GDGF) [24], and co-occurrence filter [25] (CoF) methods are the popular local filter-based methods. The visual quality of the fused images could be significantly enhanced through the mentioned filtering methods. The mean filter can well remove noise by averaging value of the neighboring pixels, so it may cause reduction of visual quality. The bilateral filter method can well smooth images and maintain their details except for the results in gradient reversal artifacts. To overcome the mentioned deficiencies, the guided filtering approach presented in [23] could be effectively employed. However, the near edges of the image could not be accurately described via the local linear model in a guided filtering approach. This approach may cause several halos in the fused images that degrade the fusion efficiency. The gradient domain-guided filtering method significantly improves the visual quality except for much time computation of three visual features. The CoF method can well extract the detail information of the original images and enhance the visual quality of the fused image, but

the iterative co-occurrence filtering on base layers for sharpening boundaries between textures costs much time.

Weighted least squares filtering approach [26, 27],  $L_1$  fidelity with  $L_0$  gradient [28], anisotropic diffusion [29], and fourth-order partial differential equation [30] are some well-known global optimization-based methods. The image smoothing could be achieved via the mentioned methods by solving an optimization problem with various fidelity terms or regularization. Although superior results and considerable enhancement of the fusion quality are always attained through these approaches, some iteration rules performed in them are time consuming. Besides, several parameters such as regularized factors, scale level, and synthetic weights should be adjusted in these methods that could be considered as another deficiency. On the other hand, various layers fused through the  $L_1$  fidelity with the  $L_0$  gradient method [27] by combining several levels with various weighted values may cause blocking artifacts during the fusion procedure.

The algorithms based on hybrid filter smooth the source images with two or more filters, such as Gaussian filter, bilateral filter [31], and rolling guidance filters [22]. These methods can well suppress the halos and retain details except for the lack in robust of adaptability.

As mentioned above, GFF can obtain better results with high computational efficiency except for representation of the image well near some edges. The local linear model used in the guided image filter improves the computational efficiency and gives the guided image filter the superiority in representation of background information of the source image. Unlike the across texture, the edges could be smoothed through CoF, which enables the extraction of texture data of the original images by CoF. Inspired by the advantage of GFF and CoF, a fusion method is presented for infrared and visible image with the guided filter and co-occurrence filter. The advantage of the guided filter in background information extraction and that of the co-occurrence filter in edge structure information extraction are combined to improve the fusion performance of infrared and visible image. The contributions of this study can be concluded as the following four aspects: (1) a novel infrared and visible image fusion approach using the guided filter and co-occurrence filter is proposed. (2) The guided filtering in base layers and co-occurrence filtering in detail layers enhance the fusion efficiency of the source images. (3) The base and detail layers are fused with the saliency maps constructed by the guided filter and co-occurrence filter, respectively. (4) The range filter of the normalized co-occurrence matrix is removed for improving the filtering speed.

The remaining parts of the current work are given as follows. Section 2 provides a detailed description of the guided filter and fast co-occurrence filter. Section 3 describes the presented approach. Section 4 presents the experimental results and their related discussions. Finally, Section 5 is devoted to conclusions and future work aspects.

## 2. Guided Filter and Improved Co-Occurrence Filter

**2.1. Guided Filter.** The output image  $Y(i)$  of the guided filter is a local linear transformation of the guided image  $I(i)$ :

$$Y(i) = a_p I(i) + b_p, \quad \forall i \in \omega_p, \quad (1)$$

where  $i$  denotes the position of the image,  $\omega_p$  is a partial rectangular window of  $(2r+1) * (2r+1)$  at center pixel  $p$ ,  $r$  is an input parameter of the GF, and  $a_p$  and  $b_p$  are obtained by minimizing the output image  $Y(i)$  and the input image  $X(i)$ :

$$E = \sum_{i \in \omega_p} \left[ (a_p I(i) + b_p - X(i))^2 + \lambda a_p^2 \right], \quad (2)$$

where  $\lambda$  is a regularization parameter given by the user and the optimal value is obtained by linear ridge regression:

$$a_p = \frac{1/|\omega| \sum_{i \in \omega_p} I(i) X_i - \mu_p \overline{X_p}}{\delta_p + \lambda}, \quad (3)$$

$$b_p = \overline{X_p} - a_p \mu_p,$$

where  $\mu_p$  and  $\delta_p$  represent the mean and variance of the guided image  $I$  in  $\omega_p$ , respectively,  $|\omega|$  is the total number of pixels in  $\omega_p$ , and  $\overline{X_p}$  is the mean of the input image  $X$  in  $\omega_p$ . The filtered output image is

$$Y(i) = \overline{a_i} I(i) + \overline{b_i}, \quad (4)$$

where  $a_i$  and  $b_i$  are the average of the different matrix windows:

$$\begin{aligned} a_i &= \frac{1}{|\omega|} \sum_{p \in \omega_i} a_p, \\ b_i &= \frac{1}{|\omega|} \sum_{p \in \omega_i} b_p. \end{aligned} \quad (5)$$

Large windows and regularization parameters can be used to fuse smooth regions containing background information, and the fusion efficiency is fast [32].

**2.2. Improved Co-Occurrence Filter.** The co-occurrence filter places the value-domain Gaussian filter of the BF into the normalized co-occurrence matrix [33]. The pixel values of high co-occurrences are assigned with larger weights and are smoothed, and pixels of low co-occurrences have smaller weights and are not smoothed. The CoF is defined as follows:

$$Y_p = \frac{\sum_{q \in N(p)} w(p, q) * X_q}{\sum_{q \in N(p)} w(p, q)}, \quad (6)$$

where  $Y_p$  and  $X_p$  are the pixel values of the output and the input images, respectively,  $p$  and  $q$  denote indices of the pixels, and  $w(p, q)$  is the weight contributed by the pixel  $q$  to the output pixel  $p$ .

In a Gaussian filter,  $w(p, q)$  could be described as follows:

$$w(p, q) = e^{-d(p, q)^2 / 2\delta_s^2} \triangleq G_{\delta_s}(p, q), \quad (7)$$

where  $G_{\delta_s}(p, q)$  denotes a Gaussian filter,  $d(p, q)$  denotes the Euclidean distance of the pixels  $p$  and  $q$ , and  $\delta_s$  is the specific parameter.

In the BF,  $w(p, q)$  is defined as follows:

$$w(p, q) = G_{\delta_r}(p, q) \times e^{-|X_p - X_q|^2 / 2\delta_r^2}, \quad (8)$$

where  $\delta_r$  is a certain parameter.

CoF combines the normalized co-occurrence matrix with the range filter in BF to extend BF to handle the boundary. The formula is described as follows:

$$w(p, q) = G_{\delta_s}(p, q) * M(|X_p, X_q|), \quad (9)$$

$$M(|X_p, X_q|) = \frac{C(a, b)}{h(a)h(b)},$$

where  $M$  is a  $256 * 256$  matrix and  $C(a, b)$  denotes the co-occurrence of the counts  $a$  and  $b$ .  $h(a)$  and  $h(b)$  represent the corresponding frequencies of  $a$  and  $b$  in the image:

$$\begin{aligned} C(a, b) &= \sum_{p, q} e^{-d(p, q)^2 / 2\delta_s^2} [X_p = a] [X_q = b], \\ h(a) &= \sum_p [X_p = a], \end{aligned} \quad (10)$$

where  $\delta$  is the Gaussian filter parameter ( $\sqrt{15}$  by default) and  $[\ ]$  is a logical operation, and if the item inside is true, the result is 1, otherwise 0.

CoF gathers co-occurrence data from the image and filters out noise while maintaining sharp boundaries between different textures. However, a co-occurrence filter is utilized to calculate the co-occurrence value by means of local window combined with the Gaussian filter, which increases the filter rate. The time complexity of the co-occurrence matrix in the original co-occurrence filter is  $O(n \times r^2)$ , where  $n$  denotes the number of pixel points and  $r$  is the size of the local window.

To increase the filtering procedure speed, we remove the range filter of the normalized co-occurrence matrix and globally count the number of co-occurrences of the image. The  $C(a, b)$  of the improved co-occurrence filter (ICoF) is calculated as follows:

$$C(a, b) = \sum_{p, q} [X_p = a] [X_q = b]. \quad (11)$$

The number of intervals of the statistical pixel pair is determined experimentally to be 6. The simplified co-occurrence matrix time complexity is  $O(n)$ , which improves the filter speed.

The comparison of the filter results of CoF and ICoF is shown in Figure 1.

Due to the removal of the range filter, ICoF will produce spots in the texture, but ICoF still has the ability to maintain the edges. To further verify the edge retention capability of ICoF, the normalized image is input into two filters, and the results are presented in Figure 2.

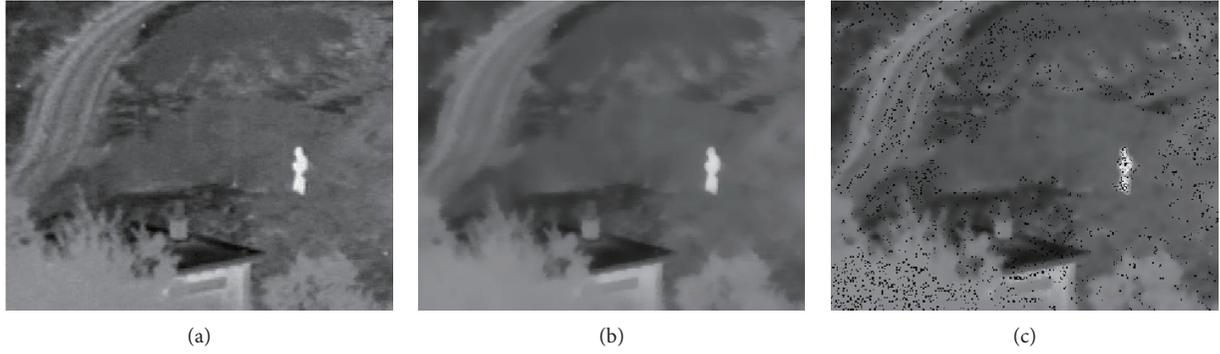


FIGURE 1: Comparison of filtering results using CoF and ICoF: (a) source image, (b) CoF, and (c) ICoF.



FIGURE 2: Comparison of normalized filtering results using CoF and ICoF: (a) source image, (b) CoF, and (c) ICoF.

As shown in Figure 2, CoF is smooth for the target person's head, leaves, grass edges, etc., while the ICoF edge retains more stringent. For the experimental comparison of the filter speeds of CoF and ICoF, the time consumption of the ten groups of experiments is averaged to attain the final time consumption. The experimental image size is  $270 * 360$ , and the results are given in Table 1. It is evident from the table that the filter speed of ICoF is about 50% smaller than that of CoF.

As mentioned above, the guided filter is good at background information extraction, and ICoF is good at filtering out noise while maintaining sharp boundaries between different textures. Thus, we make the most of their advantages to construct a novel fusion method for improving the fusion performance of the infrared and visible image.

### 3. Proposed Fusion Method

We propose a hybrid filtering-based fusion approach for infrared and visible images. The proposed fusion framework is described in Figure 3. We use the Gaussian filter to perform two-scale decomposition on the source image. The detail and the base layers are, respectively, fused by a weighted averaging method of the improved co-occurrence filter and guided filter, and finally the fusion result is obtained by superimposition.

**3.1. Two-Scale Decomposition.** The mean filter and Gaussian filter are usually used for two-scale decomposition. The Gaussian filter decomposition is sharper, and the edge

TABLE 1: CoF and ICoF time consumption comparison (unit: second).

Filter	CoF	ICoF
Time consuming	1.1308	0.5086

information obtained by the Gaussian filtering is more significant. Thus, the Gaussian filter is applied to our fusion framework for obtaining the base layer containing a large amount of background information. Assume that  $I_i$  represents the registered input image, where  $i = 1, 2$ . The base layer  $B_i$  of the two source images is calculated as follows:

$$B_i = I_i * G, \quad (12)$$

where  $G$  is the Gaussian filter with the standard deviation 5 while the filter window is adjusted as  $11 \times 11$ . The detail layer  $D_i$  including the edge structure can be obtained by subtraction of the base layer  $B_i$  containing the background information from the source image  $I_i$ :

$$D_i = I_i - B_i. \quad (13)$$

**3.2. Fusion of Base Layers.** It is easy to see that background information of the base layer is smooth. The Laplacian filter is applied to extract the saliency of the layer for obtaining a contrast map. Then, the Gaussian low-pass filter is applied to filter the high-frequency noise for a final saliency map  $BS_i$ . The multiple filtering is performed as follows:

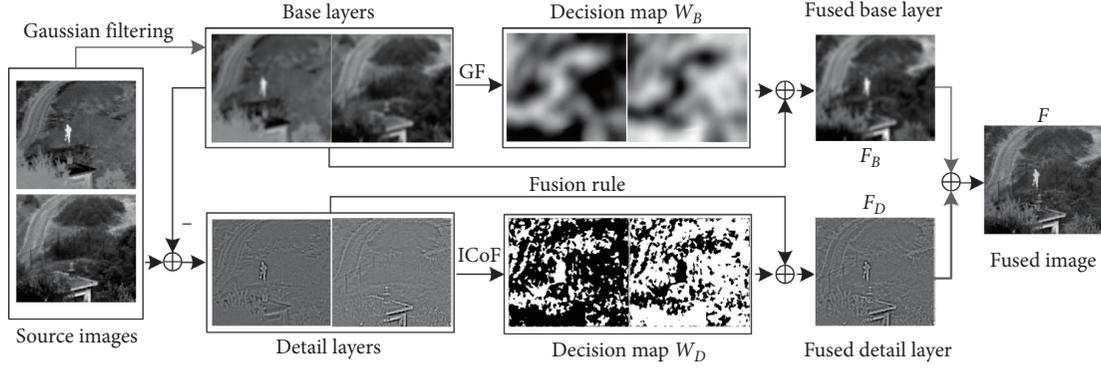


FIGURE 3: Block diagram of the image fusion algorithm based on GF-ICoF.

$$\begin{aligned} B_i H &= \text{Lap}(B_i), \\ BS_i &= G(B_i H), \end{aligned} \quad (14)$$

where  $\text{Lap}$  represents the Laplacian filter and the filtered operator is  $[0 \ 1 \ 0; 1 \ -4 \ 1; 0 \ 1 \ 0]$ .  $G$  denotes the Gaussian filter, and the filter window size is  $11 \times 11$ , and the standard deviation is 5. The initial binary weight maps  $BP_i^k$  of the base layer are obtained by comparing the saliency maps of the base layers. The pixels with high saliency have high weight:

$$BP_i^k = \begin{cases} 1, & \text{if } S_i^k = \max(BS_1^k, BS_2^k), \\ 0, & \text{otherwise,} \end{cases} \quad (15)$$

where  $k$  represents a pixel point, and a binary weight map is obtained by comparing pixel by pixel. The binary weight maps  $BP_i^k$  directly applied to the base layer fusion may produce artifacts and blurring because of spatial inconsistency. As is known, the guided filter can well preserve the edges between textures while smoothing the surrounding pixel values. Optimized weight maps with the guided filter can preserve texture information more while maintaining spatial consistency. Optimized weight maps are defined as follows:

$$W_i^B = \text{GIF}_{r,c}(BP_i, B_i), \quad (16)$$

where  $W_i^B$  represents the final base layer weight map, the input of the guided filter is  $BP_i$ , and the guided image is the original base layer  $B_i$ .

The multiple base layer weight maps are merged with the base layer to obtain the fused base layer  $F_B$ :

$$F_B = W_i^B * B_i. \quad (17)$$

**3.3. Fusion of Detail Layers.** The detail layer mainly includes the edge structure data of source images. Five classical saliency detection algorithms: spectral residual (SR) [34], frequency-tuned (FT) [35], maximum symmetric surround (MSS) [36], median mean difference (MMD) [37], and visual weight map (VWM) [22] are compared to search for a better measurement of the saliency of the edge structure in detail layers. The results are shown in Figure 4. It is evident that FT is better for the significant extraction of character and leaves except for the road and the grass. The MMD has improved

the saliency of the road, but the contrast of the grass is insufficient. SR has a lot of saliency positions, but not very continuous. The saliency of VWM is obvious, but there is still some information lost. The saliency of MSS is obvious in representing roads and grasses, and it is more comprehensive. Thus, we use MSS as the saliency extraction method of the detail layer containing a lot of edge structures.

MSS is defined as follows:

$$S(i, j) = \|X_\mu(i, j) - X_f(i, j)\|, \quad (18)$$

where  $S(i, j)$  is the saliency value of the output at position  $(i, j)$ .  $X_\mu(i, j)$  is the average of the CIELAB vectors of all pixels of the input image  $X(i, j)$ .  $X_f(i, j)$  is the corresponding CIELAB pixel vector of the input image after the Gaussian filter. The value range of the Gaussian filtered gray image in this paper is adjusted to  $[0, 100]$  for obtaining the brightness value as  $X_f(i, j)$ .  $\| \cdot \|$  indicates the  $L_2$  norm.  $X_\mu(i, j)$  is as follows:

$$X_\mu(i, j) = \frac{1}{H} \sum_{n=i-i_0}^{i+i_0} \sum_{m=j-j_0}^{j+j_0} X(n, m), \quad (19)$$

where  $i_0$  and  $j_0$  represent the offset and  $H$  is a subgraph, which is defined as follows:

$$\begin{cases} i_0 = \min(i, r - i), \\ j_0 = \min(j, r - j), \end{cases} \quad (20)$$

where  $r$  and  $c$  represent the width and height of the input image, and  $H = (2 \times i_0 + 1)(2 \times j_0 + 1)$ . The initial weight map  $DP_i^k$  is obtained by comparing the obtained saliency map. It is defined as follows:

$$DP_i^k = \begin{cases} 1, & \text{if } S_i^k = \max(DS_1^k, DS_2^k), \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

Binary weight maps directly applied for the fusion of detail layers may produce artifacts and blurring. Considering the advantage of CoF in texture feature extraction and the visual feature of the detail layer, we preserve the intertexture edge with the co-occurrence filter smoothing binary weight map. In order to improve the speed, we optimized CoF by removing the range filter of the normalized co-occurrence matrix. ICoF is defined as follows:

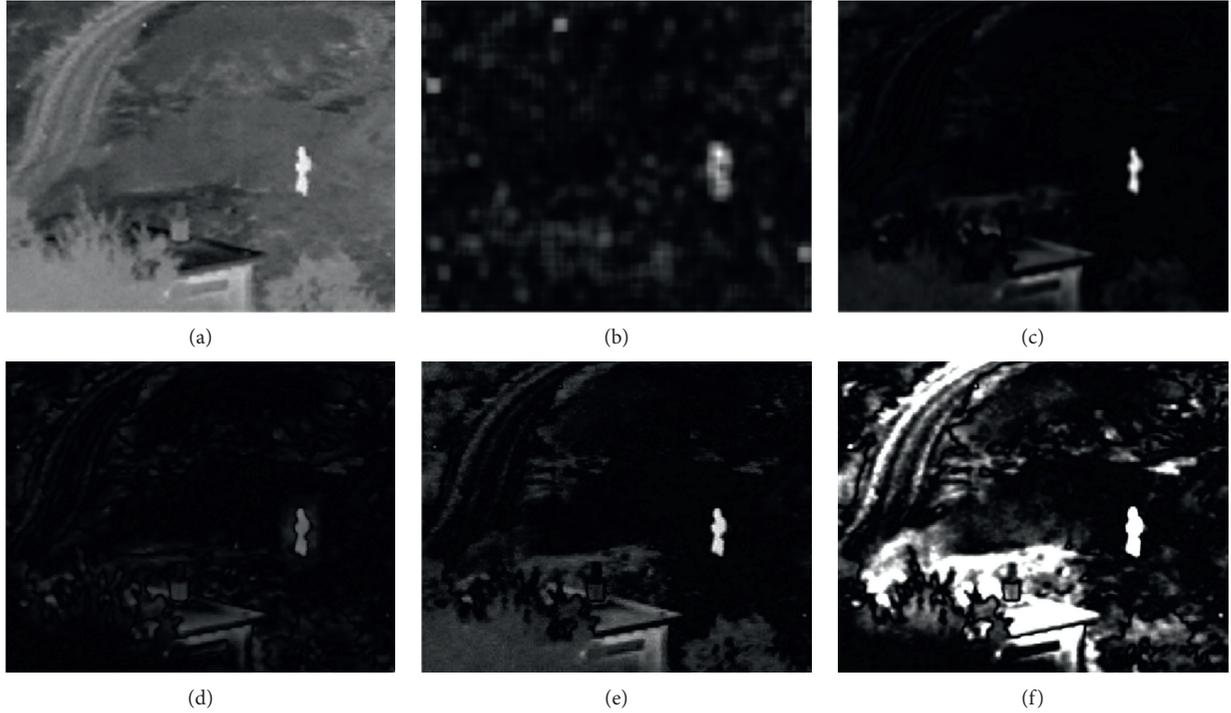


FIGURE 4: Significant comparison results: (a) source image, (b) FT, (c) SR, (d) MMD, (e) VWM, and (f) MSS.

$$W_i^D = \text{ICoF}_{\delta_s}(DP_i), \quad (22)$$

where  $\delta_s$  denotes the standard deviation of the spatial Gaussian filter in ICoF. Parameter settings are consistent with those of the original co-occurrence filter, and  $\delta_s^2 = 2 * \sqrt{15} + 1$ . The weight maps obtained guide the fusion of the detail layer for obtaining the fused detail layer  $F_D$ :

$$F_D = \sum_{i=1}^2 W_i^D * D_i. \quad (23)$$

**3.4. Two-Scale Image Reconstruction.** Since the detail layer and the base layer are obtained by subtraction, the reconstruction is obtained by merging the fusion results of the detail and the base layers:

$$F = F_B + F_D. \quad (24)$$

## 4. Experimental Results and Discussion

In order to evaluate the performance of the proposed method, the classical and recent proposed fusion algorithms are used for comparison. The experimental results are assessed with subjective evaluation and objective evaluation.

### 4.1. Experimental Settings

**4.1.1. Image Sets.** Figure 5 shows eight pairs of the test image used for the experiment of infrared and visible image fusion derived from three public datasets [37–39]. The three image datasets used to support the findings of this study are included

within the open data collection in <http://www.imagefusion.org/>, [https://figshare.com/articles/TNO\\_Image\\_Fusion\\_Dataset/1008](https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008), and <https://github.com/hanna-xu/RoadScene>. The upper images are infrared image, and the lower images are corresponding visible images.

**4.1.2. Compared Methods.** The fusion methods used for experiment in this paper are as follows: guided filter algorithm (GFF) [23], two-scale using saliency detection (TSD) [37], contrast pyramid algorithm (CP) [40], convolution sparse representation (CSR) algorithm [41], gradient-based transfer and total variation algorithm (GTTV) [42], weighted least squares optimization (WLS) algorithm [22], multiple visual feature measurement-based fusion (MVFMF) approach [24], co-occurrence filter-based fusion method (CoF) [25], cross bilateral filter (CBF) [21], target-enhanced multiscale transform decomposition algorithm (TEMTD) [17], and latent low-rank representation algorithm (LLRR) [43].

**4.1.3. Parameter Setting.** The parameters of the fast co-occurrence filter have been introduced in the previous section. For the setting of the two parameters  $r$  and  $\varepsilon$  of the guided filter, through multiple groups of experiments combined with subjective and objective evaluation analysis, the value of the two parameters and the size of the fused image are positively correlated. This paper introduces a window factor  $t$  which divides the sum of image rows and columns to obtain the final window size  $r$ :



FIGURE 5: Eight sets of images to be fused shown in the experiment: (a) NATO camp, (b) Kaptein man, (c) sandpath, (d) Kaptein rain, (e) soldiers with jeep, (f) bike with car, (g) crowds, and (h) backpacker.

$$r = \lfloor \frac{m+n}{t} \rfloor, \quad (25)$$

where  $m$  and  $n$  represent the corresponding rows and columns of the image, rounded up, and the value is integer. It is known that not all objective evaluation indicators can effectively reflect the fusion effect. In order to avoid different evaluation indicators for determining the value of the window factor, we use  $Q^{AB/F}$  as the objective feedback value of fusion results with different  $t$ . The final value of  $r$  is determined based on the trend of  $Q^{AB/F}$  of the fusion results going along with different  $t$ . The experimental results are shown in Figure 6. It can be seen from Figure 6 that  $Q^{AB/F}$  has reached a peak when  $t$  is set to 34. Thus, the window factor is determined to be 34.  $\varepsilon$  is obtained by dividing  $r$  by a constant  $g$ ,  $\varepsilon = (r/g)$ . According to the parameters of the guided filter fusion base layer in [19],  $g = 150$ .

**4.1.4. Evaluation Metrics.** To verify the performance of the presented algorithm comparing with existing approaches, we apply five evaluation metrics to the objective fusion assessment. They are quality assessment based on structural

similarity entropy (EN), standard deviation (SD), multiscale structural similarity (MS-SSIM) [44], normalized feature mutual information (NFMI) [45], edge retention fusion quality indicator  $Q^{AB/F}$  [46], and visual information fidelity degree VIFF [47]. These metrics describe the fusion results from different perspectives, and the larger of those values indicate that more source image information is retained, and better fusion performance is achieved [48].

**4.2. Qualitative Assessments.** Using qualitative result analysis, Figures 7–11 show the fusion results with different algorithms for the infrared and visible images of Figure 5.

The GFF algorithm can enhance the visual qualities of the fusion image by employing appropriate weights on the corresponding regions. Although the additional structure and detailed data could be transferred from the guidance image to the fused one via the GFF method, the images around certain edges could not be described appropriately through the local linear model in this method. Although no explicit constraint occurs from the edge-aware factor to treat edges, it may generate specific halos in the fused images. A

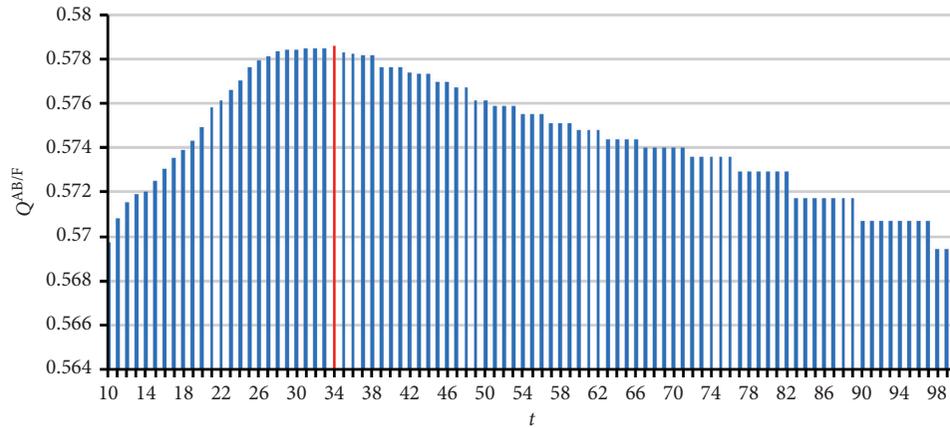


FIGURE 6:  $Q^{AB/F}$  values of different  $t$  fusion results.

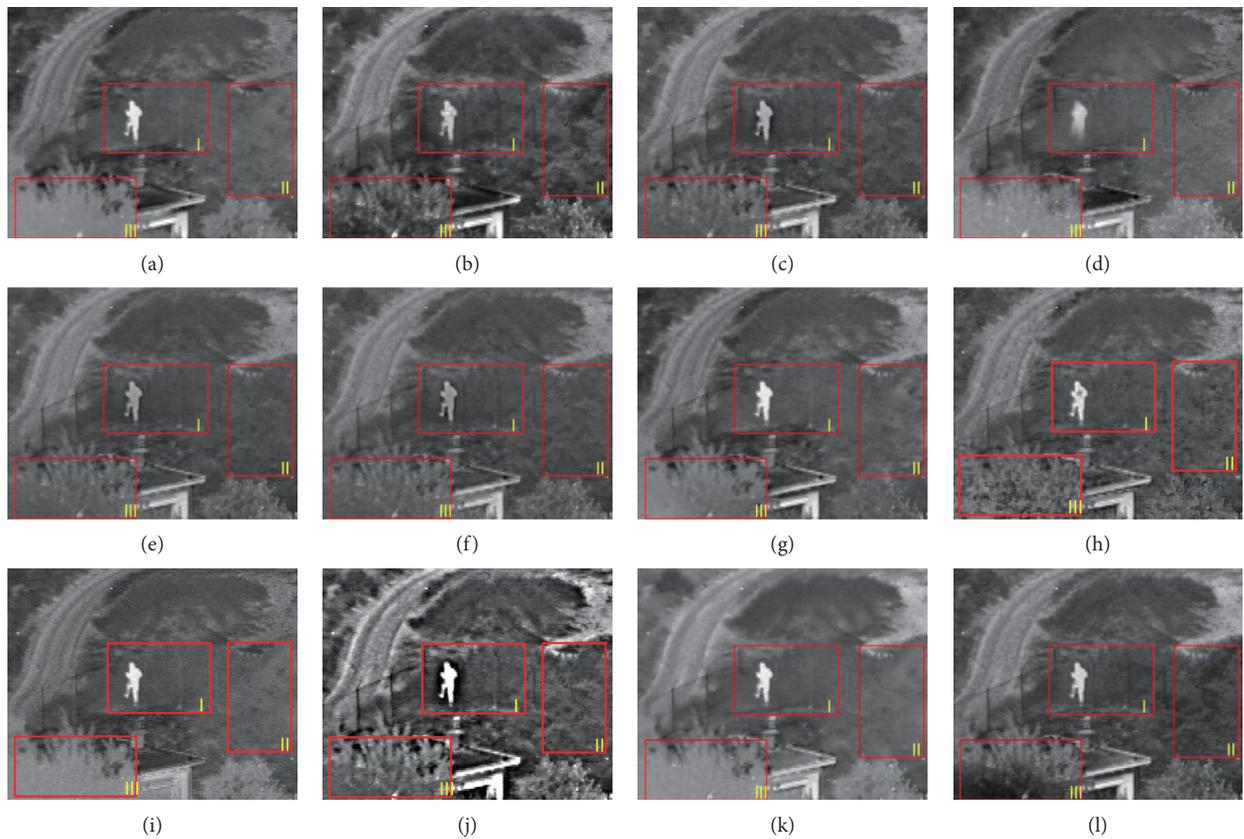


FIGURE 7: Fusion results of “NATO camp” with different fusion methods: (a) GFF, (b) CP, (c) TSSD, (d) GTTV, (e) CSR, (f) WLS, (g) MVFMMF, (h) CBF, (i) TEMTD, (j) LLRR, (k) CoF, and (l) GF-ICoF.

few blurring artifacts could be observed due to the GFF fusion such as the lost data of grass (boxes II and III) in Figure 7(a), the blurred branches (box I) in Figure 8(a), the reduced brightness of the trunk portion (box II) of the tree in Figure 9(a), the reduced brightness of the tent (box II) in Figure 10(a), the lost cloud information of the sky background (box I) in Figure 11(a), and the blurred street lamp (box I) and blurred building (box II) in Figure 12(a).

CP decomposed the source image into some decomposition layers with various resolution and spatial

frequencies. CP has achieved a good visual effect except for the lack in adaptability. The CP fusion results show some blurring, such as the low contrast in Figure 7(b), blurred branches (box I) in Figure 8(b), the black block appearing in the lower right corner (box III) in Figure 9(b), the low brightness of the jeep and soldiers in Figure 11(b), and the blurred street lamp (box I) and the blurred leg of the crowds (box II) in Figure 12(b). TSSD used a mean and median filter to extract visual saliency extraction. The weight map obtained from the saliency map guided the fusion of the base



FIGURE 8: Fusion results of “Kaptein man” with different fusion methods: (a) GFF, (b) CP, (c) TSSD, (d) GTTV, (e) CSR, (f) WLS, (g) MVFMF, (h) CBF, (i) TEMTD, (j) LLRR, (k) CoF, and (l) GF-ICoF.

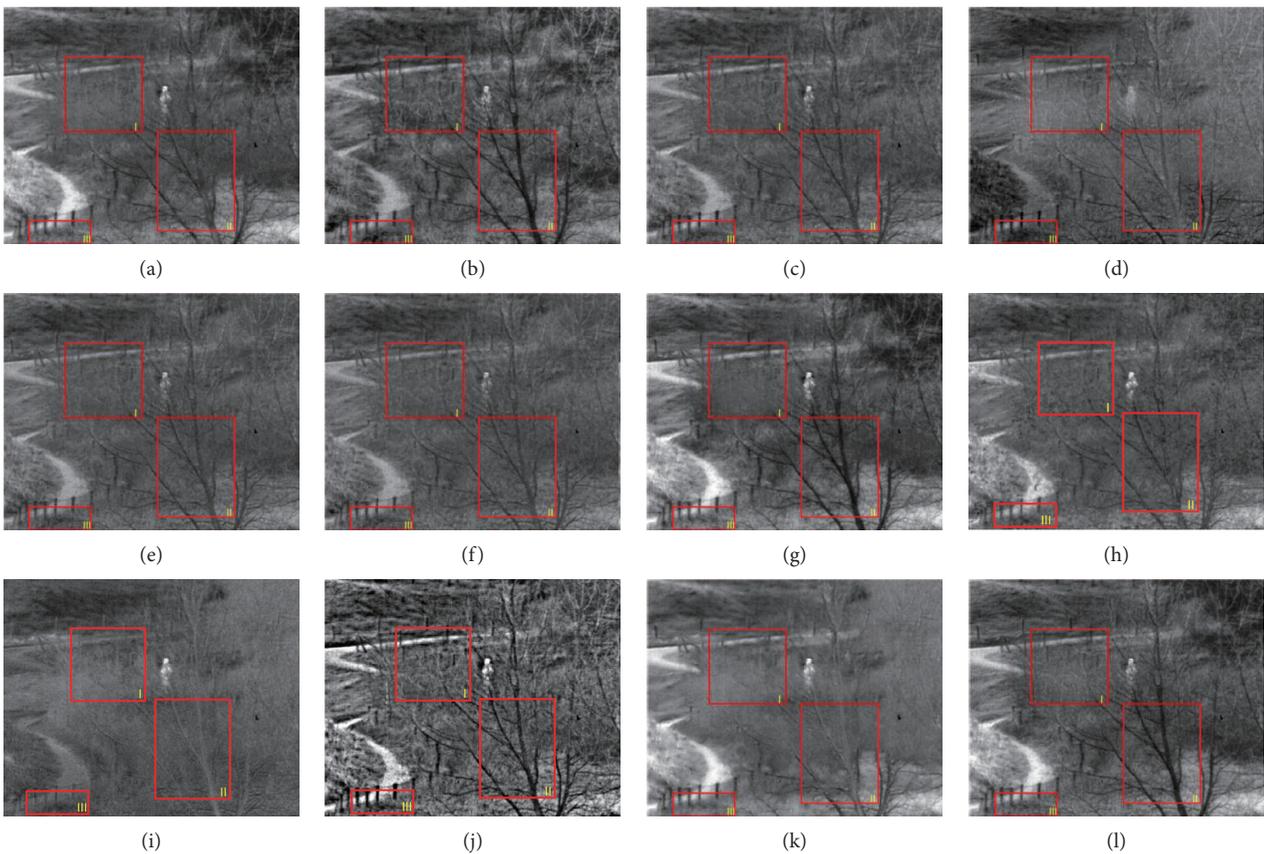


FIGURE 9: Fusion results of “sandpath” with different fusion methods: (a) GFF, (b) CP, (c) TSSD, (d) GTTV, (e) CSR, (f) WLS, (g) MVFMF, (h) CBF, (i) TEMTD, (j) LLRR, (k) CoF, and (l) GF-ICoF.

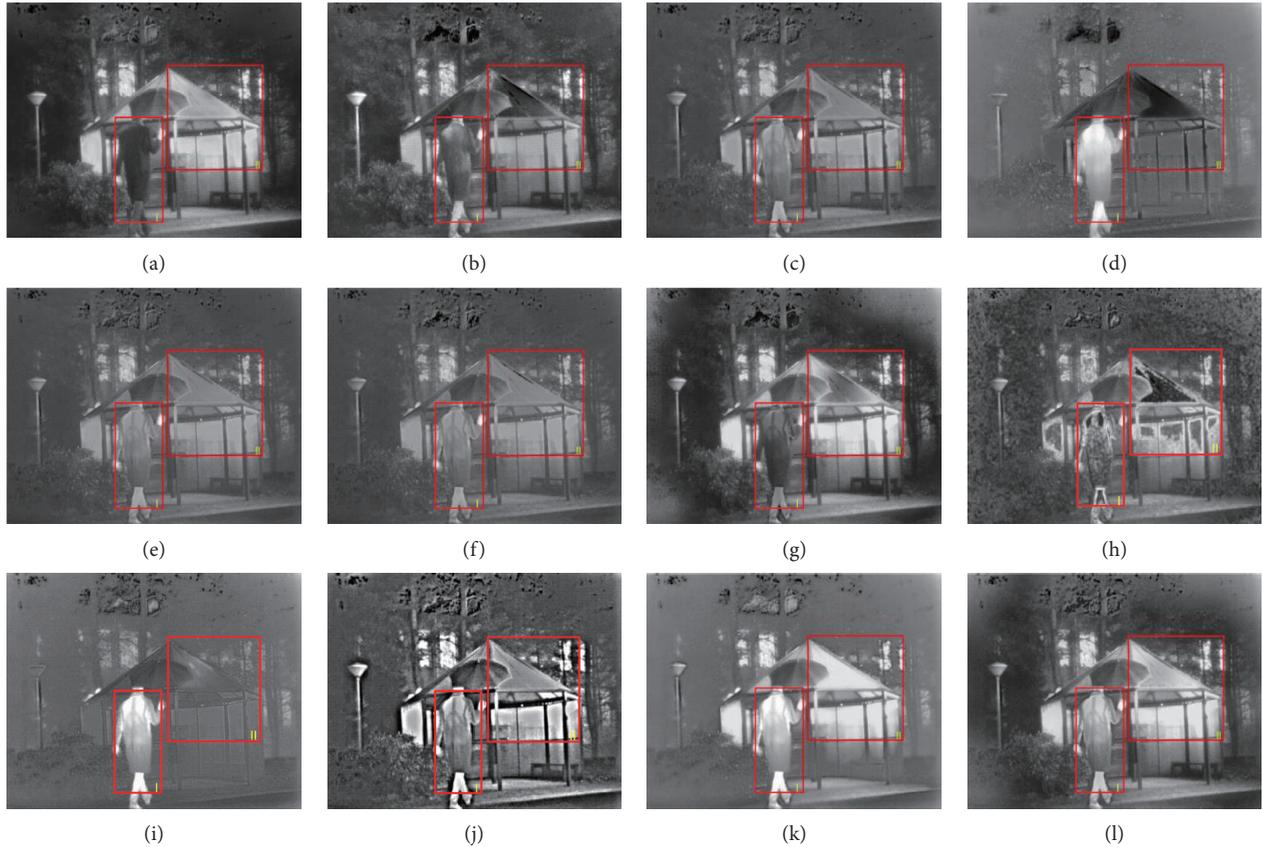


FIGURE 10: Fusion results of “Kaptein rain” with different fusion methods: (a) GFF, (b) CP, (c) TSSD, (d) GTTV, (e) CSR, (f) WLS, (g) MVFMM, (h) CBF, (i) TEMTD, (j) LLRR, (k) CoF, and (l) GF-ICoF.

and detail layers, which greatly improves the transfer ability of complementary data from the original images effectively. But the ability of the mean and median filter in saliency extraction compromised the fusion performance, such as the lost surrounding fence and grass information (boxes II and III) in Figure 7(c), the blurred grass around (box II) in Figure 8(c), the blurred trunk portion (box II) of the tree in Figure 9(c), the low brightness of the tent (box II) in Figure 10(c), the low brightness of the jeep and soldiers (box II) in Figure 11(c), and the low brightness of the crowds (box II) and street lamp (box I) in Figure 12(c).

The fusion process is converted to a  $l_1$ -TV minimization problem through the GTTV in which the fundamental intensity distribution in the infrared image and the gradient change in the visible image are preserved through the data fidelity and regularization terms, respectively. Moreover, the thermal radiation and the appearance data in the infrared and visible images could be preserved, respectively (except for the lack of adaptability). The blur appears such as the blurred branches (box I) in Figure 8(d), the blurred trunk (box II) of the tree in Figure 9(d), the lost information of the tent in Figure 10(d), the blurred jeep and soldiers (box II) in Figure 11(d), and the blurred crowds (box II) in Figure 12(d).

CSR used convolutional sparse representation to overcome detail preservation and strong sensitivity to misregistration in image fusion. It has obtained better fusion

performance for infrared and visible image. But the regularization parameters in the CSR model and the spatial size of dictionary filters lack in self-adaptability, which compromise the fusion efficiency such as the brightness of the target person (box I) in Figure 7(e), the dark background in Figure 8(e), the low brightness of the trunk portion (box II) of the tree in Figure 9(e), the low brightness of the jeep and soldiers (box II) in Figure 11(e), and the low brightness of the street lamp (box I) in Figure 12(e).

WLS employs the RGF and Gaussian filter to decompose input images to base and detail layers. WLS constructs weight maps by choosing various features of the IR and visible image. Effective visual information could be appropriately transformed into the fused image through the WLS while reducing the noise from the IR image. But the lack of parameter flexibility compromises the fusion quality, such as the low brightness of the target person (box I) in Figure 7(f), the dark background in Figure 8(f), the blurred trunk portion (box II) of the tree in Figure 9(f), the low-contrast vehicle and soldiers (box II) in Figure 11(f), and the low-contrast street lamp (box I) and crowds (box II) in Figure 12(f).

Three visual property measurements are utilized by the MVFMM to generate decision maps that are optimized through gradient domain-guided filtering. The visual description of the detailed data of the original images was provided through it. Although the fusion efficiency was

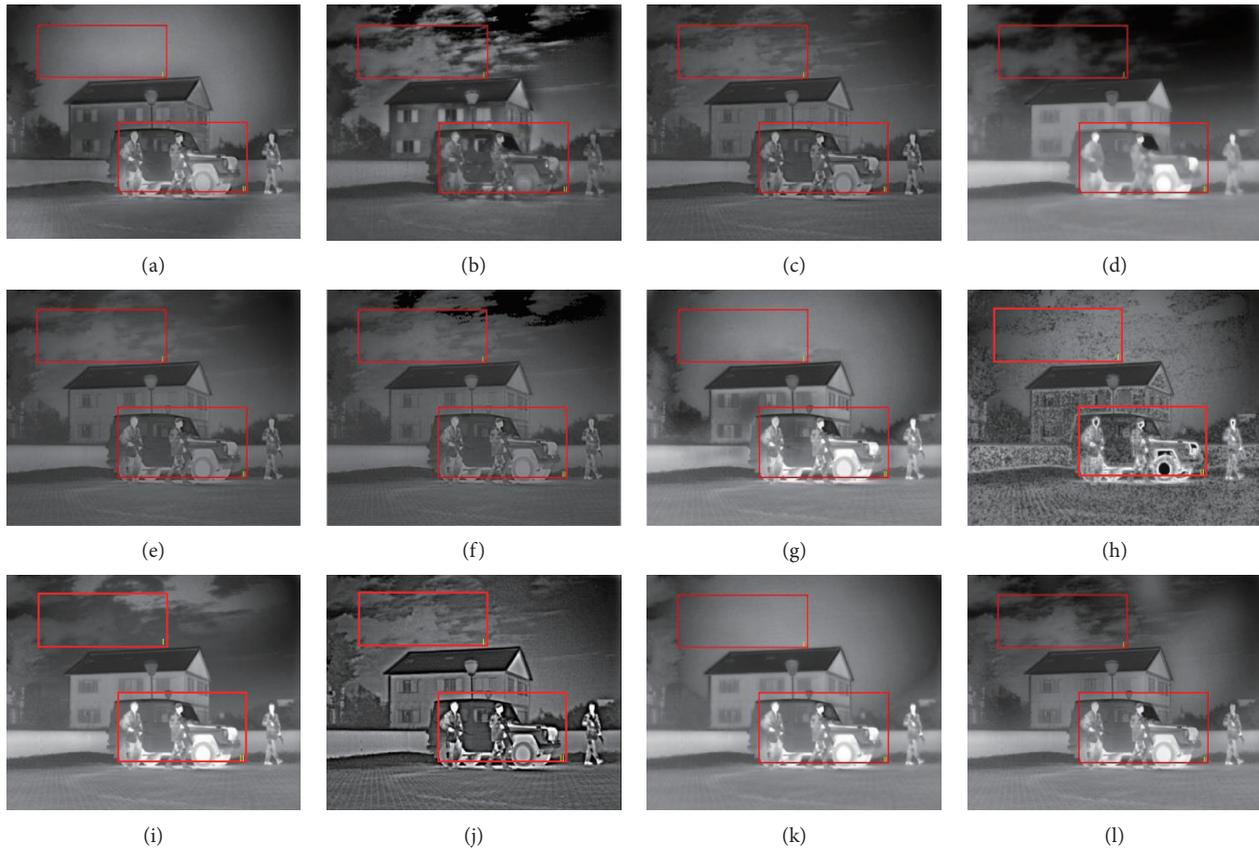


FIGURE 11: Fusion results of “soldiers with jeep” with different fusion methods: (a) GFF, (b) CP, (c) TSSD, (d) GTTV, (e) CSR, (f) WLS, (g) MVFMM, (h) CBF, (i) TEMTD, (j) LLRR, (k) CoF, and (l) GF-ICoF.

significantly enhanced via the MVFMM decision map construction model, the feasibility of the constant parameters for all original images with various values is not satisfactory in some instances. Slight blurs and artifacts could be observed in the fusion images such as the lost information of the grass (boxes II and III) in Figure 7(g), the blurred branches (box I) in Figure 8(g), the low brightness of the trunk portion (box II) of the tree in Figure 9(g), the information of the tent (box II) in Figure 10(g), the lost cloud information of the sky background (box I) in Figure 11(g), and the lost building information of the street lamp (box I) in Figure 12(g).

The CBF extracts the detail information by using joint bilateral filtering and avoids the gradient reversal artifacts of the bilateral filter. It uses weighted average of pixel values to the fuse source image and improves the visual quality of the fused image. However, the weight construction may compromise the contrast of the fused image. The little blurs and the corresponding artifacts appear in the fused images such as the blurred target person (box I) in Figure 7(h), the blurred edge of the tree (box I) and the blurred figure of the man in Figure 8(h), the blurred trunk portion (box II) of the tree in Figure 9(h), the vehicle and soldiers (box II) in Figure 11(h), and the street lamp (box I) and crowds (box II) in Figure 12(h).

TEMTD is a novel infrared and visible image fusion method based on target-enhanced MST decomposition. It

deals with problem regarding the preservation of thermal radiation features in the traditional MST-based methods. It can simultaneously maintain thermal radiation characteristics in the infrared image and texture details in the visible image by using a specific fusion rule design. It uses the decomposed infrared low-frequency information to determine the fusion weight of low-frequency bands and highlight the target. The common “max-absolute” fusion rule is performed for fusion for high-frequency bands. The common “max-absolute” fusion rule may compromise the fusion performance, such as the blurred bush (boxes II and III) in Figure 7(i), the blurred branches (box II) in Figure 9(i), the blurred bush in the left bottom of Figure 10(i), the blurred sky in Figure 11(i), and the street lamp in Figure 12(i).

LLRR is a novel fusion framework for fusing infrared and visible images. A projection matrix  $L$  learned by latent low-rank representation is applied to extract detail parts and base parts of the input images at several representation levels. It extracts multilevel salient features by using latent low-rank representation. The final fused image is reconstructed by adaptive fusion strategies designed specifically for dealing with the detail parts and the base parts, respectively. The LLRR framework can be used to provide an efficient decomposition approach for extracting multilevel features for an arbitrary number of input images. The adaptive fusion rules improve the ability in texture and structural information extraction. The bush in Figure 7(j), the branch in

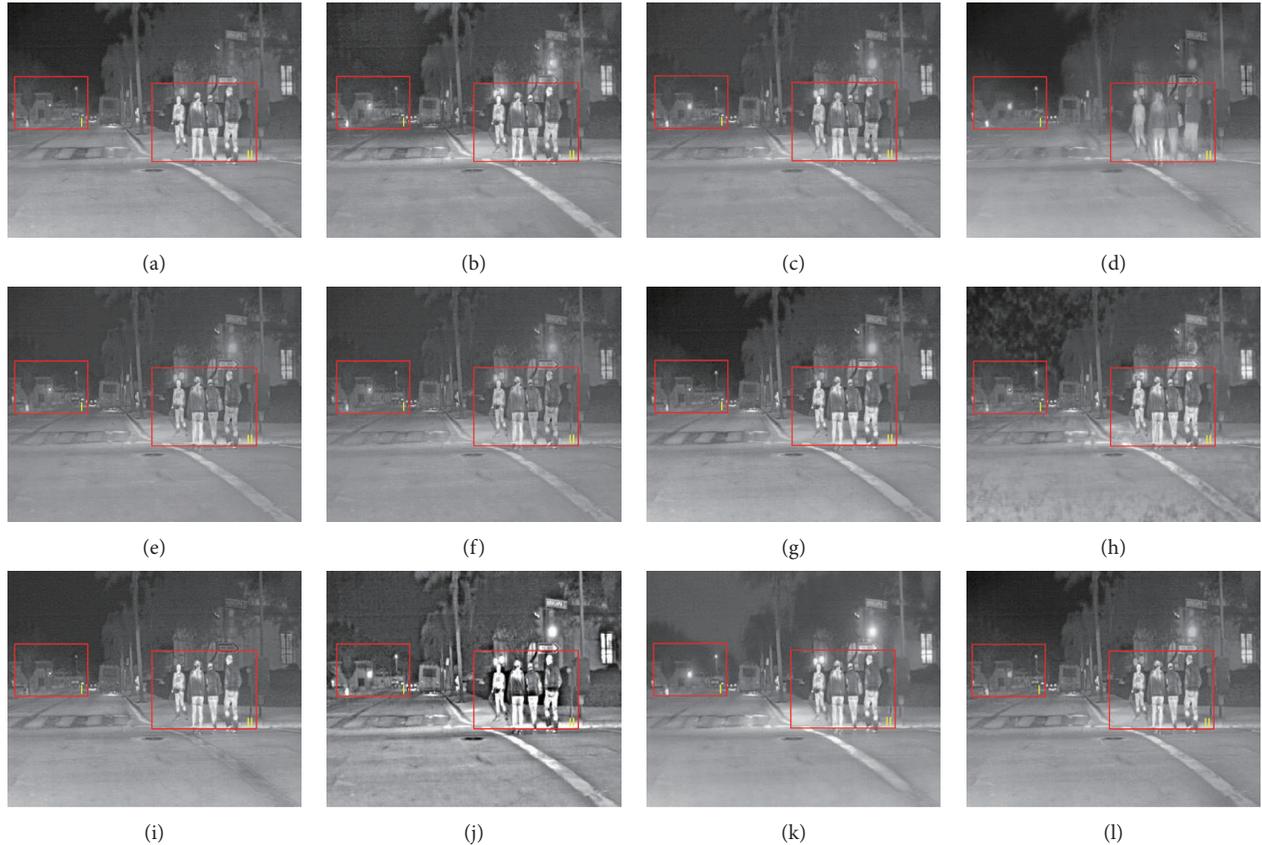


FIGURE 12: Fusion results of “soldiers with jeep” with different fusion methods: (a) GFF, (b) CP, (c) TSSD, (d) GTTV, (e) CSR, (f) WLS, (g) MVFMF, (h) CBF, (i) TEMTD, (j) LLRR, (k) CoF, and (l) GF-ICoF.

Figure 9(j), the bush in Figure 10(j), and the soldiers and jeep in Figure 11(j) all demonstrate high contrast. But there are information losses appear in the fused image, such as the black edge around the target man in Figures 7(j), 8(j), and 12(j).

The co-occurrence filter features of boundary detection and edge preservation could be utilized by the CoF for weight optimization, which improves the fusion quality of the base layers and detail layers. The same fusion rule of base and detail layers compromises the fusion performance, such as the lost information of the grass (boxes II and III) in Figure 7(k), the blurred grass (box II) in Figure 8(k), the blurred details of the branches (box I) in Figure 9(k), the blurred grass at the lower right corner (box II) in Figure 10(k), the lost cloud information of the sky background in Figure 11(k), and the lost information of the street lamp in Figure 12(k). We perform different image filtering for saliency extraction based on the visual feature of base layers and detail layers. By and large, compared with other algorithms, the proposed algorithm not only retains the saliency desired data of the infrared image but also preserves the background data in the visible image.

**4.3. Quantitative Assessments.** In the current subsection, seven quality measurement indices (as discussed in the previous subsection) are applied to the test images. The

obtained results are presented in Tables 2–6. Tables 2–6 demonstrate that the metric values of the presented approach are higher than the corresponding ones obtained with existing approaches, which demonstrates that further data could be preserved from the original images, and superior fusion quality could be obtained.

As is known,  $Q^{(AB/F)}$  shows the edge data of the fused image retained from the source image. NFMI measures the feature information of the fused image retention from the source one based on mutual information. MS-SSIM is based on structural similarity to measure the retention degree of the structural data of the fused image from the original one. EN indicates the information value contained in the fused image as a whole. SD reflects the contrast of the image with the difference degree between each pixel and the average value of the pixels. VIFF measures the degree of fusion of the effective visual information of each region.

Histogram of the average metric ( $Q^{AB/F}$ , NFMI, VIFF, and MS-SSIM) amounts of the fusion approaches is presented in Figure 13. It is easy to see that the average  $Q^{AB/F}$  values of GFF, MVFMF, and GF-ICoF are obviously higher than that of other methods, which demonstrates that GFF, MVFMF, and GF-ICoF can well preserve the edge data from original images to fused images. The average NFMI value of GFF, CSR, LLRR, MVFMF, CoF, and GF-ICoF are obviously higher than that of other fusion methods. It demonstrates that more significant feature data could be transformed from

TABLE 2: Fusion evaluation indicators for different algorithms for “NATO camp” and “Kaptein man.”

Method	NATO camp						Kaptein man					
	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD
GFF	0.533	0.903	0.173	6.663	0.704	27.433	0.545	0.942	0.280	6.858	0.879	32.898
CP	0.481	0.886	0.403	6.824	0.870	32.776	0.455	0.935	0.333	6.632	0.919	34.143
TSSD	0.434	0.889	0.398	6.469	0.927	26.380	0.448	0.932	0.371	6.648	0.933	33.347
GTTV	0.405	0.885	0.218	6.694	0.787	27.380	0.408	0.906	0.316	6.683	0.836	38.173
CSR	0.463	0.898	0.331	6.335	0.890	24.462	0.314	0.937	0.184	6.957	0.798	32.259
WLS	0.417	0.891	0.350	6.374	0.906	24.479	0.580	0.928	0.213	6.967	0.831	32.254
MVFMF	0.511	0.904	0.222	6.563	0.721	26.482	0.511	0.942	0.308	6.593	0.910	35.139
CBF	0.389	0.871	0.273	6.450	0.740	26.016	0.388	0.897	0.199	6.777	0.733	33.193
TDMTD	0.402	0.885	0.175	6.600	0.752	26.727	0.468	0.938	0.225	6.766	0.843	35.712
LLRR	0.312	0.886	0.414	6.779	0.849	31.343	0.328	0.934	0.350	7.154	0.858	45.318
CoF	0.390	0.890	0.330	6.733	0.831	30.130	0.406	0.938	0.287	7.139	0.892	50.986
GF-ICoF	0.547	0.916	0.424	6.790	0.926	32.020	0.562	0.942	0.357	7.172	0.923	51.693

TABLE 3: Fusion evaluation indicators for different algorithms for “sandpath” and “Kaptein rain.”

Method	Sandpath						Kaptein rain					
	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD
GFF	0.530	0.853	0.256	6.840	0.886	33.654	0.585	0.911	0.376	7.027	0.889	48.277
CP	0.470	0.830	0.318	6.735	0.883	27.118	0.513	0.892	0.441	6.807	0.863	37.091
TSSD	0.416	0.839	0.354	6.342	0.931	21.779	0.493	0.891	0.416	6.442	0.925	30.622
GTTV	0.495	0.841	0.144	6.510	0.825	22.312	0.400	0.881	0.160	6.524	0.745	28.765
CSR	0.471	0.855	0.257	6.224	0.895	20.124	0.525	0.900	0.340	6.354	0.896	28.560
WLS	0.385	0.845	0.296	6.244	0.899	20.188	0.470	0.893	0.347	6.400	0.897	28.664
MVFMF	0.590	0.880	0.260	7.046	0.898	38.281	0.583	0.906	0.419	7.155	0.880	47.532
CBF	0.356	0.831	0.281	6.520	0.808	27.560	0.432	0.864	0.328	6.878	0.651	36.413
TDMTD	0.315	0.846	0.096	6.148	0.675	17.498	0.363	0.888	0.131	6.139	0.695	26.014
LLRR	0.276	0.850	0.311	7.163	0.855	38.771	0.313	0.917	0.439	7.192	0.832	48.288
CoF	0.344	0.846	0.230	6.650	0.801	30.492	0.452	0.898	0.341	6.671	0.846	39.839
GF-ICoF	0.602	0.883	0.324	7.164	0.947	38.965	0.598	0.920	0.489	7.262	0.937	48.362

TABLE 4: Fusion evaluation indicators for different algorithms for “soldiers with jeep” and “bike with car.”

Method	Soldiers with jeep						Bike with car					
	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD
GFF	0.585	0.914	0.366	6.934	0.857	30.873	0.581	0.862	0.239	7.465	0.851	45.442
CP	0.294	0.873	0.411	6.724	0.795	30.393	0.504	0.827	0.331	7.332	0.810	41.419
TSSD	0.495	0.904	0.460	6.552	0.947	23.523	0.466	0.835	0.309	6.655	0.902	25.477
GTTV	0.324	0.906	0.395	7.536	0.875	56.795	0.321	0.850	0.133	7.600	0.647	48.860
CSR	0.490	0.913	0.347	6.471	0.919	22.147	0.513	0.856	0.250	6.548	0.861	23.755
WLS	0.458	0.897	0.388	6.494	0.896	26.253	0.470	0.840	0.259	6.552	0.870	23.618
MVFMF	0.643	0.915	0.468	7.187	0.853	39.617	0.580	0.861	0.227	7.565	0.842	49.134
CBF	0.352	0.834	0.384	7.018	0.595	34.169	0.463	0.808	0.231	7.510	0.758	48.269
TDMTD	0.558	0.917	0.566	7.227	0.841	43.175	0.450	0.841	0.155	7.108	0.688	39.293
LLRR	0.296	0.907	0.560	7.094	0.850	37.829	0.361	0.831	0.249	7.339	0.854	44.219
CoF	0.448	0.910	0.507	7.148	0.891	38.904	0.424	0.834	0.177	7.074	0.778	40.135
GF-ICoF	0.611	0.920	0.566	7.628	0.948	58.919	0.569	0.856	0.258	7.536	0.877	49.241

the source image to the fused one through the five methods transfer compared with existing fusion approaches. Moreover, the mean VIFF and MS-SSIM values of TSSD and GF-ICoF are obviously greater than that of existing fusion approaches. It shows that TSSD and GF-ICoF can transfer more significant structure data from original images compared with existing fusion techniques.

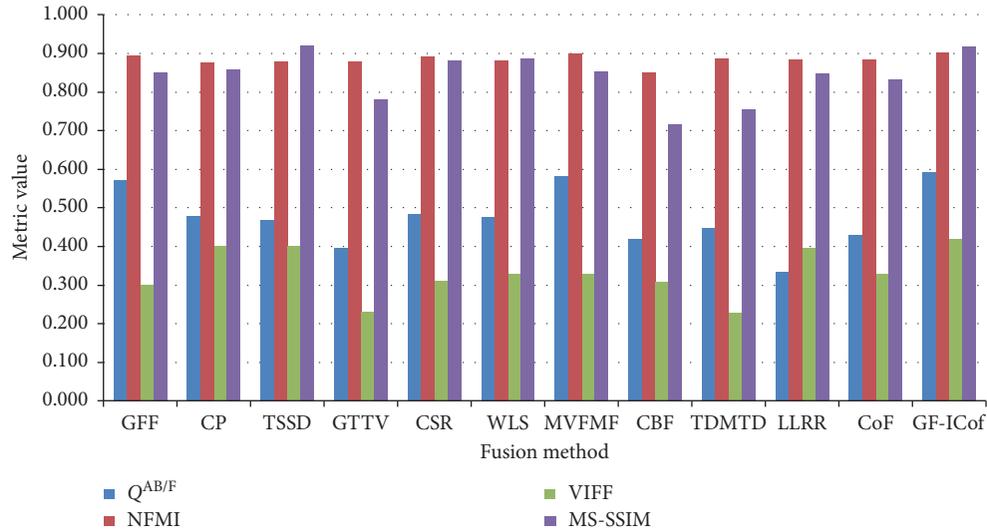
Histogram of the mean metric (EN, SD) amounts of the fusion approaches is depicted in Figure 14. It is easy to see that the average SD values of MVFMF, LLRR, CoF, and GF-ICoF are obviously greater than the corresponding values obtained by other fusion approaches. It demonstrates that the fusion results of MVFMF, LLRR, CoF, and GF-ICoF have higher contrast with better visual quality compared

TABLE 5: Fusion evaluation indicators for different algorithms for “crowds” and “backpacker.”

Method	Crowds						Backpacker					
	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD	$Q^{AB/F}$	NFMI	VIFF	EN	MS-SSIM	SD
GFF	0.627	0.915	0.426	7.357	0.886	47.692	0.593	0.869	0.291	7.245	0.844	37.960
CP	0.587	0.910	0.610	7.126	0.934	43.394	0.525	0.851	0.358	7.490	0.786	44.898
TSSD	0.513	0.901	0.530	6.884	0.942	39.869	0.484	0.852	0.367	6.938	0.856	31.128
GTTV	0.391	0.911	0.267	7.277	0.813	47.854	0.423	0.851	0.214	7.319	0.714	43.100
CSR	0.555	0.912	0.460	6.814	0.923	38.998	0.543	0.866	0.313	6.851	0.863	29.101
WLS	0.531	0.904	0.458	6.829	0.923	38.854	0.490	0.856	0.318	6.832	0.863	28.546
MVFMF	0.638	0.915	0.466	7.412	0.899	49.064	0.597	0.868	0.273	7.505	0.818	44.921
CBF	0.526	0.906	0.504	7.280	0.839	44.999	0.447	0.808	0.273	7.544	0.606	45.049
TDMTD	0.548	0.910	0.316	7.148	0.839	45.578	0.468	0.859	0.176	6.874	0.710	35.952
LLRR	0.384	0.899	0.507	7.292	0.850	49.598	0.395	0.845	0.326	7.422	0.839	45.265
CoF	0.503	0.909	0.544	6.989	0.903	41.952	0.461	0.851	0.218	6.978	0.731	44.146
GF-ICoF	0.651	0.916	0.596	7.450	0.923	49.601	0.586	0.865	0.332	7.459	0.864	45.591

TABLE 6: Time consumption of nine algorithms (unit: second).

Method	GFF	CP	TSSD	GTTV	CSR	WLS	MVFMF	CBF	TDMTD	LLRR	CoF	GF-I CoF
Time	0.638	0.106	0.151	6.651	211.002	3.190	0.634	4.023	2.963	159.165	33.875	3.158

FIGURE 13: Histogram of the average metric ( $Q^{AB/F}$ , NFMI, VIFF, and MS-SSIM) values of the fusion methods.

with existing fusion algorithms. The mean EN values of GFF, MVFMF, CoF, and GF-ICoF are greater than that of existing fusion techniques. The comparison results of the fused image quality evaluation indices demonstrate the superiority of the developed fusion approach. The proposed method has sufficient retention for the edge and structure data from the original image. The fused image obtained with the presented approach has the best visual quality in terms of contrast. Overall, the approach is preferred to existing algorithms in accordance with different evaluation metrics.

The time complexity of each algorithm is obtained on the running time of 10 tests. The selected image size is  $270 \times 360$ . The results are shown in Table 6. It is evident that the time consumption of CSR is highest. The parameter training and dictionary learning cost much time. The extraction of detail parts and base parts of the source images is implemented by

learning with latent low-rank representation, which results in higher time consumption of LLRR. The algorithm models of CP, TSSD, and GFF are simple and small time consumptive. The time consumption of WLS, CBF, TDMTD, and MVFMF is a little more than the previously mentioned methods. CSR spends a lot of time in sparse representation, which leads to excessive time overhead. The time cost of the propose method and the GTTV algorithm tends to be similar. Because the time consumption of the CoF is a little more, the CoF fusion algorithm uses an iterative way to fuse the base layer, which further increases the time spent for the fusion. Compared with the CoF fusion algorithm, the time efficiency of this work is increased by about 90%. In terms of the subjective and objective results, we can conclude that the algorithm is an effective infrared and visible image fusion algorithm.

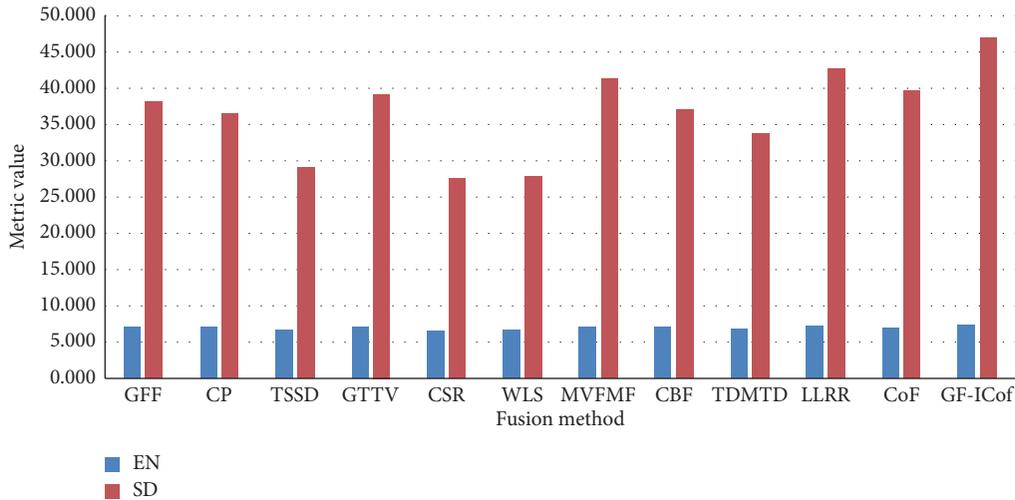


FIGURE 14: Histogram of the average metric (EN and SD) values of the fusion methods.

## 5. Conclusions

As conventional infrared and visible image fusion methods suffer from low contrast and background texture loss, a novel fusion approach is presented using guided and improved co-occurrence filters. The advantage of the guided filter in background information extraction and the advantage of the co-occurrence filter in edge structure information extraction are combined to improve the fusion performance of the infrared and visible image. The co-occurrence filter is improved by removing the range filter and globally synthesizing the co-occurrence information. The filtering time of the co-occurrence filter is reduced by half while preserving across texture edges. The qualitative assessments demonstrate that the fusion results of the proposed method can retain the thermal radiation and appearance data in the infrared and visible images, respectively. The quantitative comparisons on seven metrics with recent fusion approaches indicate that more significant edge and structure data could be transformed from the original image to the fused one through the presented approach. Future work will further improve the speed of the proposed method and apply it to image fusion applications such as medical and remote sensing.

## Data Availability

The two image datasets used to support the findings of this study are included within the open data collection in <http://www.imagefusion.org/> and [https://figshare.com/articles/TNO\\_Image\\_Fusion\\_Dataset/1008](https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to thank Kang Xudong, Ma Jiayi, Zhang Ping, Liu Yu, Ma Jinlei, Qu Xiaobo, Chen Jun, Li Hui,

Baviriseti D. P., and Jevnisek R. J. for providing their codes. This work was supported by the National Natural Science Foundation of China (no. 61802162), the Scientific and Technological Project of Henan Province (no. 192102210122), the Key Industrial Innovation Chain Project in Industrial Domain of Shaanxi Province (no. 2017ZDCXL-GY-03-01-01), and the Primary Research & Development Plan of Shaanxi Province (nos. 2019ZDLSF07-02 and 2019ZDLGY10-01).

## References

- [1] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: a survey of the state of the art," *Information Fusion*, vol. 33, pp. 100–112, 2017.
- [2] J. Ma, M. Yong, and L. Chang, "Infrared and visible image fusion methods and applications: a survey," *Information Fusion*, vol. 45, pp. 153–178, 2018.
- [3] X. Jin, Q. Jiang, S. Yao et al., "A survey of infrared and visual image fusion methods," *Infrared Physics & Technology*, vol. 85, pp. 478–501, 2017.
- [4] H. Jin and Y. Wang, "A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization," *Infrared Physics & Technology*, vol. 64, no. 3, pp. 134–142, 2014.
- [5] X. Han, L. L. Zhang, L. Y. Du et al., "Fusion of infrared and visible images based on discrete wavelet transform," in *Proceedings of the Photoelectronic Technology Committee Conferences*, International Society for Optics and Photonics, July 2015.
- [6] L. Yan and T. Z. Xiang, "Fusion of infrared and visible images based on edge feature and adaptive PCNN in NSCT domain," *Acta Electronica Sinica*, vol. 44, no. 4, pp. 761–766, 2016.
- [7] P. Jiang, Q. Zhang, J. Li et al., "Fusion algorithm for infrared and visible image based on NSST and adaptive PCNN," *Laser and Infrared*, vol. 44, no. 1, pp. 108–112, 2014.
- [8] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "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.
- [9] H. Li and X. J. Wu, "Densefuse: a fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2018.

- [10] C. Xie and X. Li, "Infrared and visible image fusion: a region-based deep learning method," *Intelligent Robotics and Applications*, pp. 604–615, Springer, Cham, Switzerland, 2019.
- [11] H. Li, X. J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, Sydney, Australia, pp. 2705–2710, 2018.
- [12] 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.
- [13] J. Ma, P. Liang, W. Yu et al., "Infrared and visible image fusion via detail preserving adversarial learning," *Information Fusion*, vol. 54, pp. 85–98, 2020.
- [14] H. Xu, P. Liang, W. Yu et al., "Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 3954–3960, Beijing, China, 2019.
- [15] G. He, J. Ji, D. Dong, J. Wang, and J. Fan, "Infrared and visible image fusion method by using hybrid representation learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 11, pp. 1796–1800, 2019.
- [16] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4980–4995, 2020.
- [17] J. Chen, X. Li, L. Luo, X. Mei, and J. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Information Sciences*, vol. 508, pp. 64–78, 2020.
- [18] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: a unified densely connected network for image fusion," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12484–12491, 2020.
- [19] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12797–12804, 2020.
- [20] J. Hu and S. Li, "The multiscale directional bilateral filter and its application to multisensor image fusion," *Information Fusion*, vol. 13, no. 3, pp. 196–206, 2012.
- [21] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1193–1204, 2015.
- [22] 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.
- [23] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 22, no. 7, pp. 2864–2875, 2013.
- [24] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multiple visual features measurement with gradient domain guided filtering for multisensor image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 691–703, 2017.
- [25] P. Zhang, Y. Yuan, C. Fei, T. Pu, and S. Wang, "Infrared and visible image fusion using co-occurrence filter," *Infrared Physics & Technology*, vol. 93, pp. 223–231, 2018.
- [26] J. Zhao, Q. Zhou, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition," *Infrared Physics & Technology*, vol. 56, pp. 93–99, 2013.
- [27] Y. Jiang and M. Wang, "Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter," *IET Image Processing*, vol. 8, no. 3, pp. 183–190, 2014.
- [28] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Optics Communications*, vol. 341, pp. 199–209, 2015.
- [29] D. P. Bavisetti and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 203–209, 2015.
- [30] D. P. Bavisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proceedings of the 20th International Conference on Information Fusion (Fusion)*, IEEE, Xi'an, China, pp. 1–9, 2017.
- [31] 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," *Information Fusion*, vol. 30, pp. 15–26, 2016.
- [32] K. He, J. Sun, and X. Tang, *Guided Image filtering*, European Conference on Computer Vision, Springer, Berlin, Germany, 2010.
- [33] R. J. Jevnisek and S. Avidan, "Co-occurrence filter," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3184–3192, Honolulu, HI, USA, 2017.
- [34] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Minneapolis, MN, USA, pp. 1–8, 2007.
- [35] R. Achanta, S. Hemami, F. Estrada et al., "Frequency-tuned salient region detection," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 1597–1604, Miami, FL, USA, 2009.
- [36] R. Achanta and S. Süsstrunk, "Saliency detection using maximum symmetric surround," in *Proceedings of the 2010 IEEE International Conference on Image Processing*, IEEE, Taipei, Taiwan, pp. 2653–2656, 2010.
- [37] D. P. Bavisetti 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] <http://www.imagefusion.org/>.
- [39] [https://figshare.com/articles/TNO\\_Image\\_Fusion\\_Dataset/1008](https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008).
- [40] M. Li and Y. Dong, "Image fusion algorithm based on contrast pyramid and application," in *Proceedings of the 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, IEEE, Shenyang, China, pp. 1342–1345, 2013.
- [41] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.
- [42] 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.
- [43] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: a novel decomposition method for infrared and visible image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4733–4746, 2020.
- [44] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1398–1402, IEEE, Pacific Grove, CA, USA, 2003.
- [45] M. Haghighat and M. A. Razian, "Fast-NFMI: non-reference image fusion metric," in *Proceedings of the IEEE 8th*

*International Conference on Application of Information and Communication Technologies (AICT)*, IEEE, Astana, Kazakhstan, pp. 1–3, 2014.

- [46] V. Petrovic and C. Xydeas, “Objective image fusion performance characterisation,” in *Proceedings of the Tenth IEEE International Conference on Computer Vision*, 2005.
- [47] Y. Han, Y. Cai, Y. Cao, and X. Xu, “A new image fusion performance metric based on visual information fidelity,” *Information Fusion*, vol. 14, no. 2, pp. 127–135, 2013.
- [48] R. Alais, P. Dokládál, A. Erginay et al., “Fast macula detection and application to retinal image quality assessment,” *Biomedical Signal Processing and Control*, vol. 55, Article ID 101567, 2020.