Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition

Xiaoye Zhang,[1] Yong Ma,[1] Fan Fan,[1] Ying Zhang,[2] and Jun Huang[1,2,*]

[1]Electronic Information School, Wuhan University, Wuhan 430072, China
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive NW, Atlanta, Georgia 30332, USA
*Corresponding author: junhwong@whu.edu.cn

Received 7 April 2017; revised 26 June 2017; accepted 29 June 2017; posted 30 June 2017 (Doc. ID 291835); published 21 July 2017

To retain the details of a visible image with a discernible target area, we propose a multi-scale decomposition image fusion method based on a local edge-preserving (LEP) filter and saliency detection. We first use a LEP filter to decompose the infrared and visible images. Then, a modified saliency detection method is utilized to detect the salient target areas of an infrared image, which determine the base layer's weights of fusion strategy. Finally, each layer is reconstructed to obtain a visually pleasing fused image. Comparison with 11 other state-of-the-art methods reveals the superiority of the proposed method in terms of quality and quantity results.

OCIS codes: (100.0100) Image processing; (100.2000) Digital image processing; (350.2660) Fusion.

https://doi.org/10.1364/JOSAA.34.001400

1. INTRODUCTION

Infrared imaging can resist illumination changes, smog, and disguise compared to visible imaging, making it suitable for searching hidden targets and identifying camouflages. However, it is difficult to identify the surroundings since the infrared image lacks details and reflects limited background information [1]. On the other hand, visible images record the information derived from the visible spectrum, which is different from the infrared image that reflects the thermal radiation intensity of objects. However, the visible images are greatly influenced by the lighting condition, which makes the target region in the visible image inconspicuous. To get an accurate, reliable, and comprehensive description of the scene information, the infrared image and the visible image could be combined to get a fused image [2–4], including both the infrared target and the details of the scenes, which is beneficial for subsequent tasks, such as detection and recognition [5,6]. The goal of image fusion is to capture complementary information from different types of images. For the infrared images, the major information is their intensity information, which is able to highlight the positions of objects (note that we have made an assumption that the objects have larger intensities than the background), while for the visible images, they contain rich texture information and hence help to further recognize the objects. Therefore, our goal of infrared/visible image fusion in this paper is to fuse the infrared intensity and visible texture.

Recently, a variety of image fusion algorithms have been proposed [7]. A popular class of algorithms is the multi-scale image fusion scheme. A multi-scale transform is a recognized tool that has been demonstrated to be effective in image fusion and other image processing applications. Generally, the multi-scale transform image fusion consists of three steps. First, it is used to obtain the multi-scale representations of the input images. Then, a fused multi-scale representation is obtained by fusion of the multi-scale representations of different images according to a specific fusion rule. Finally, the fused image is obtained by taking an inverse multi-scale transform on the fused representation. However, many multi-scale transform-based image fusion methods suffer from the following two defects. The first one is the ringing effect (halo and artifact) [8,9]. This is mainly because their basic filters do not preserve spatial consistency. The second one is the loss of contrast of the target [10]. Since most energy of an image is contained in the low-pass band, the "average" fusion rule tends to lose some energy in the source images.

To address these problems, we propose a novel multi-scale decomposition image fusion method based on a local edge-preserving (LEP) filter and saliency detection [9]. LEP is adept in preserving not only global salient edges but also the local salient edges. More precisely, we employ LEP to decompose both the infrared and visible images. Through preserving the local shape by the LEP filter, the fused image can avoid halo and artifact. On the other hand, a modified saliency detection method,

aiming at infrared images, is utilized to discriminate a target from the background. The saliency is used in the base layer's fusion strategy, which can well preserve the contrast of the target.

The main contributions of this paper include the following two aspects. First, we introduce an edge-preserving filter named LEP into multi-scale decomposition-based image fusion. Compared with existing methods, our fusion image achieves better results in retaining abundant details and preventing the halo and artifact. Second, according to the characteristics of an infrared image, we develop the saliency detection algorithm, and incorporate the saliency into the fusion strategy of the base layer, which prevents the fused image from losing the contrast.

The rest of this paper is organized as follows. Section 2 introduces related works. Section 3 describes our proposed method from three aspects: the LEP filter and multi-scale decomposition, saliency detection, and fusion strategy. Section 4 demonstrates the experimental comparison of our proposed method with several state-of-the-art methods. Section 5 gives the conclusion of this paper.

2. RELATED WORK

A. Multi-Scale Decomposition Filter

Compared with the single-scale-based decomposition methods, multi-scale analysis has the advantage of extracting and combining image features at different scales. Images can be decomposed into different scales by image filters. The transfer function of a linear filter, such as Gaussian [11], Laplacian pyramid (LP) [12], and the ratio of a low-pass pyramid [13], is fixed. These filters cannot adapt to the pixel value, which causes the artifacts and fuzzy details when fusing the image [8]. Another type of classic method, the wavelet method [14], suffers from some fundamental shortcomings, such as shift variance, aliasing, and lack of directionality. As a solution to these problems, the dual-tree complex wavelet transform (DTCWT) has been successfully applied for image fusion [15]. The main advantages of DTCWT are shift invariance and directional selectivity, which reduces the artifacts introduced by a wavelet. However, the wavelet-based approaches cannot well represent the curves and edges of images. To represent the spatial structures in the images more accurately, some novel multi-scale geometric analysis tools are introduced into image fusion. Nencini *et al.* successfully applied the curvelet transform (CVT) for the fusion of remote sensing images [16]. Li and Wang applied the nonsubsampled contourlet transform (NSCT) to fuse the biological images [17,18].

In recent years, edge-preserving filtering has been widely applied to construct multi-scale representations of images. For example, Farbman *et al.* constructed edge-preserving multi-scale image decompositions with a weighted least-squares (WLS) filter for the fusion of multi-exposure images [8]. Jiang and Wang applied the image fusion based on WLS to multi-focus, multi-sensor, and medical image fields [19]. Li *et al.* introduced a guided filtering-based method, which is efficient and effective in several image fusion applications [20]. Toet developed the guided filter into an iterative mode, which has a simple implementation and is efficient [21]. Leya and Biju introduced an efficient fusion method based on guided filtering and bilinear interpolation [22]. Pritika and Budhiraja focused on the multi-modal medical image fusion with modified fusion rules and a guided filter, and the fused images preserve good contrast and fine details [23]. Gan *et al.* developed a multi-scale decomposition infrared and visible image fusion method based on a phase congruency saliency map and guided filter (PC-GF) [24]. Zhou *et al.* proposed a multi-scale fusion method based on a hybrid multi-scale decomposition (HMSD) [25], furthermore they utilized the image enhancement method to achieve a visually pleasing result [26]. Although these methods can preserve the salient edge in the whole area when decomposing the image, they overlook the local edges and details [9]. For instance, the WLS considers the areas with globally relative large gradients as the salient edges. But a small gradient may also be a locally important edge and we believe these areas ought to be preserved. We utilize the LEP filter to decompose the source images, which can preserve the locally relative large gradient, so that the fused image will contain abundant details and prevent the artifacts.

B. Saliency Detection

Saliency, as widely believed, is a bottom-up process that originates from visual distinctness, rarity, or surprise, and is often attributed to variations in image attributes such as color, gradient, edges, and boundaries [27]. The saliency detection algorithms can be broadly classified into local and global schemes.

Local contrast-based methods investigate the rarity of image regions with respect to local neighborhoods. Based on the highly influential biologically inspired early representation model introduced by Koch and Ullman [28], Itti proposed a saliency map model according to the primate early vision [29]. Ma and Zhang proposed an alternate local contrast analysis method for generating saliency maps, which was then extended using a fuzzy growth model [30]. Harel proposed a bottom-up visual saliency model to normalize the feature maps, highlight conspicuous parts, and permit combination with other importance maps [31]. The model is simple, biologically plausible, and easy to parallelize. More recently, Goferman simultaneously modeled local low-level clues, global considerations, visual organization rules, and high-level features to highlight salient objects along with their contexts [32]. Such methods using local contrast tend to produce higher saliency values near edges.

Global contrast-based methods evaluate saliency of an image region using its contrast with respect to the entire image. Achanta proposed a frequency tuned method that directly defines pixel saliency using a pixel's color difference from the average image color [33]. Cheng proposed a histogram-based contrast method (HC) to measure saliency [34]. The HC algorithm assigns saliency values based on color separation from all other image pixels to produce full-resolution saliency maps. They also proposed a region-based contrast (RC) algorithm, which is an improved version of HC [34]. RC first segments the input image into regions, and then assigns saliency values to them. The saliency value of a region is measured by the region's contrast and spatial distances to other regions in the image. Zhai and Shah (LC) define pixel-level saliency based on a pixel's contrast to all other pixels [35]. It is an efficient method which

considers the illumination's distinction between the salient region and background.

The local scheme may cause the globally dim areas to have higher saliency than those globally bright areas, which is improper in the infrared image fusion since the luminance reflects targets' thermal radiation. After analyzing the characteristics of infrared images, we develop the saliency detection method based on the LC algorithm and make it suitable for visible and infrared image fusion.

3. PROPOSED METHOD

We decompose the original images into a base layer and several detail layers. Aiming at the multi-sensor fusion of the infrared and visible images, we use different fusion strategies for different layers. This section describes the LEP filter and multi-scale decomposition, salience extraction, and fusion strategy.

A. LEP Filter and Multi-Scale Decomposition

Literally, edge-preserving smoothing means preserving the edges as well as smoothing the image. Let I denote the input image and B represent the smoothed image, which is derived from I and should be as close to I as possible. Edge-preserving smoothing requires that B should be as smooth as possible in the local area except across the edges. It obtains the image B by minimizing Eq. (1):

$$\mathbf{min}:\sum_{i\in w}(I_i - B_i)^2 + \frac{\alpha}{|\nabla I_i|^{\beta}}|\nabla B_i|^2, \tag{1}$$

where w is the window area. $\frac{\alpha}{|\nabla I|^{\beta}}$ is the coefficient balancing the two terms. β determines the coefficient's sensitivity to the gradient of I, and α is a free parameter. If the gradient of I is relatively larger at some location, the coefficient $\frac{\alpha}{|\nabla I|^{\beta}}$ will get smaller, then the first constraint will dominate, and B will be as close as possible to I, resulting in the salient edge preserved in B. On the other hand, if the gradient of I is relatively small, $\frac{\alpha}{|\nabla I|^{\beta}}$ will get larger, then the second constraint will dominate, and B will be as smooth as possible:

$$B_i = a_w I_i + b_w, \qquad i \in w, \tag{2}$$

where a_w and b_w are constant coefficients in the window w. Replacing B in Eq. (1) by Eq. (2) will get

$$\sum_{i\in w}(I_i - a_w I_i - b_w)^2 + \alpha|\nabla I_i|^{2-\beta}\cdot a_w^2. \tag{3}$$

Comparison between LEP and the guided filter reveals that LEP is adaptive to the gradient because of the coefficient $\alpha|\nabla I_i|^{2-\beta}\cdot a_w^2$, while the guided filter has a fixed parameter [9,20]. That means LEP has advantages in preserving edges. Now the optimization problem becomes a parameter estimating problem. The minimum of Eq. (3) can be found by setting the partial derivative of each parameter to zero. This linear least-squares solution is

$$\begin{cases} a_w = \frac{\sigma_w^2}{\sigma_w^2 + \frac{1}{N}\cdot\alpha\cdot\sum_{i\in w}|\nabla I_i|^{2-\beta}}, \\ b_w = \bar{I}_w - a_w\bar{I}_w, \end{cases} \tag{4}$$

where σ_w^2 is the variance of I in the window w and $\bar{I}_w$ is the mean of I in w. It can be easily deduced that a_w is always less than 1, so B is the smooth version of I. Each window contains N pixels, and the output of LEP is the mean of the N value of B_i:

$$B_i = \frac{1}{N}\sum_{k\in w}(a_k I_i + b_k) = \bar{a}_i I_i + \bar{b}_i, \qquad i \in \Omega, \tag{5}$$

where Ω is the area of the image and $\bar{a}_i$ is the average of the a_k in the neighborhood window, and the same with $\bar{b}_i$. Obviously, the larger window Ω becomes, the coarser image B will be.

WLS and LEP can get desirable results from both the smoothing image and preserving edges aspects; however, LEP provides a better result in preserving local edges [9].

LEP has two parameters, a and β, which can adjust the filter's sensitivity to gradient. When a and β are small, more gradients will be regarded as salient edges by LEP. Otherwise, when a and β are large, LEP will treat less gradient as salient edge, which means the filtered output will be oversmoothed.

A single LEP operation on the original image will give a base layer and a detail layer. However, some details cannot be fully extracted by a fixed scale, so, by iteratively applying LEP on the base layer, a multi-scale decomposition is utilized in our method. During the iteration, the local window is increasing, which results in progressive coarsening:

$$\begin{cases} B_{l-1} = \mathrm{LEP}_l(B_l), & \text{for } l = L, \ldots, 1, \quad \text{and} \quad B_L = I, \\ D_l = B_l - B_{l-1}, & \text{for } l = L, \ldots, 1, \end{cases} \tag{6}$$

where LEP_l denotes the filter function, and l is the scale levels. B_l and D_l are the base layer and the detail layer of l level. After the iterative process, the image is decomposed into one base layer and several detail layers:

$$I = B_0 + D_1 + D_2 + \cdots + D_L. \tag{7}$$

B. Saliency Extraction

Some saliency extraction methods, such as LC, HC and RC, take the contrast as the saliency feature. Zhai proposed the LC method, which defines the saliency value of pixel k of image I as $S(I_k) = \sum_{\forall I_i \in I}\|I_k - I_i\|$. Zhao uses this original method in his fusion strategy [36]. However, simply taking the contrast as the feature to extract saliency would retain the areas with the highest and lowest thermal radiation, which is inconsistent with the *a priori* knowledge that the targets are commonly hotter than the background. That may lead to highlighted areas at the non-target regions in fused images. So we modify the algorithm to retain the bright but not the dim regions. It should be noted that the infrared images in this paper are all in white-hot representation (as thermal images also can be used in a black-hot representation):

$$\begin{aligned} \mathrm{dis}(I_k, I_i) &= \begin{cases} I_k - I_i, & \text{if } I_k - I_i \geq 0, \\ 0, & \text{if } I_k - I_i < 0, \end{cases} \\ S(I_k) &= \sum_{\forall i}\mathrm{dis}(I_k, I_i), \end{aligned} \tag{8}$$

where i and k are the locations of pixels in the image. dis measures the luminance distance. Let $I_k = a_m$, and Eq. (8) is constructed to Eq. (9):

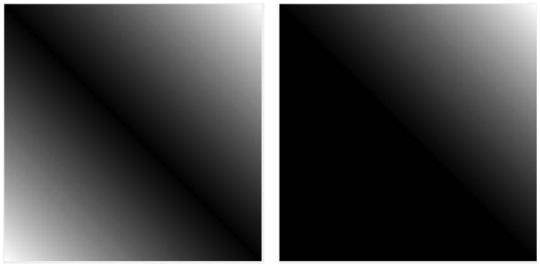
$$S(a_m) = \sum_{\forall i} \mathrm{dis}(k, i) = \sum_{n=0}^{255} f_n \cdot \mathrm{dis}(a_m, a_n), \tag{9}$$

where f_n is the frequency of pixel value a_n in the image. Because of $a_n \in [0, 255]$, the luminance distance metric $\mathrm{dis}(a_m, a_n)$ is also bounded in the range of $[0, 255]$. Since this is a fixed range, a distance map can be constructed, as shown in Fig. 1.

To well exhibit our saliency algorithm, we take the 2_men_in_front_of_house as the example shown in Fig. 2. The first image of the top row is the original infrared image, and the other two images represent saliency detection of LC and our method, respectively. Both methods can reserve the hottest objects, i.e., the two people in both scenes. However, the LC method also preserves the dark areas, such as the sky and the window. Our method can effectively extract the saliency, such as the people and outline of the building, and suppress the non-target areas, such as the sky in the infrared image. As a result, the target would get a good contrast in the fused image and the non-target areas keep clean.

C. Fusion Strategy

The base layer contains the most of an image's energy, so we use the saliency map in the base layer in order to make the target prominent. Since the targets in an infrared image are commonly brighter than the background and can be easily extracted, only the infrared image saliency is extracted. Via this process, the fusion image could well preserve the contrast of the target areas, resulting in a desirable visual effect. The base layer fusion strategy is shown as follows:



(a) LC method (b) Proposed method

Fig. 1. Illumination distance sketch map of the LC method and proposed method.

Fig. 2. Saliency detection result of LC and our method. From left to right, top to bottom, they are an infrared image, saliency map of the LC method and our method, visible image, fusion result of LC and our method.

$$B_f = B_{ir} \cdot S_{ir} + B_{vi} \cdot (1 - S_{ir}), \tag{10}$$

where S_{ir} is the normalized salient value. B_{ir} and B_{vi} are the base layer of an infrared and visible image, respectively. Our goal of infrared/visible image fusion in this paper is to fuse the infrared intensity and visible texture. The base layer fusion strategy is to keep the target area's energy of the infrared image. The latter item of the equation can ensure dominance of the infrared image in the target area. As a result, the target areas can preserve relatively high luminance in the fused image.

In the detail layers, although the commonly used methods, such as taking the maximum or the average values, are easily implemented, they can cause the halo effect and neglect some detail contexts [7]. The detail contexts are the places that have the high local contrast. Since the Laplacian energy reflects the local contrast [37], we use it as the detail fusion strategy. In that way, the fusion image can well preserve the detail information:

$$E_{\mathrm{lap}} = \sum_{\forall x \in w} \sum_{\forall y \in w} (E_{xy})^2, \tag{11}$$

$$\begin{aligned} E_{xy} = & -d(x-1, y-1) - 4d(x-1, y) - d(x-1, y+1) \\ & - 4d(x, y-1) + 20d(x, y) \\ & - 4d(x, y+1) - d(x+1, y-1) \\ & - 4d(x+1, y) - d(x+1, y+1), \end{aligned} \tag{12}$$

where x and y are the row and column pixel locations. E_{xy} is the Laplacian energy, w is the local window, and E_{lap} is the sum of the squared Laplacian energy of all pixels in the window w. We normalize the level l Laplacian energy of infrared and visible images, and get the weight M_{ir}^l and M_{vi}^l. Then M_{ir} and M_{vi} are utilized at each detail level as follows:

$$D_f^l = M_{ir}^l \cdot D_{ir}^l + M_{vi}^l. \tag{13}$$

Finally, the fusion image is synthesized with different weights at different scales:

$$I_f = B_f + \mu_1 D_f^1 + \mu_2 D_f^2 + \cdots + \mu_L D_f^L, \tag{14}$$

where I_f is the final result, and $\mu_l (l = 1, 2 \ldots L)$ is the weight of level l. By adjusting μ_l, the fusion image can obtain different visual effects. When μ_l is relatively small, the image will be smooth; otherwise, the image will be sharpened. The flow chart is shown in Fig. 3.

4. EXPERIMENTS

In this section, we test the performance of the proposed method on publicly available datasets, and compare it with 11 state-of-the-art fusion methods, namely, CVT [16], DTCWT [38], guided-filtering-based fusion (GFF) [20], LP [12], Laplacian pyramid with sparse representation (LPSR) [10], multi-resolution singular value decomposition (MSVD) [39], wavelet [40], WLS [8], NSCT [17], HMSD [25], and image fusion based on PC-GF [24]. All the comparative algorithms are implemented using the publicly available codes, where the parameters are set according to the original paper, and we try our best to tune the key parameters to obtain the best detail performance. In addition, we assume all the image pairs are pre-aligned to a pixel-to-pixel level; otherwise, some registration method could be used

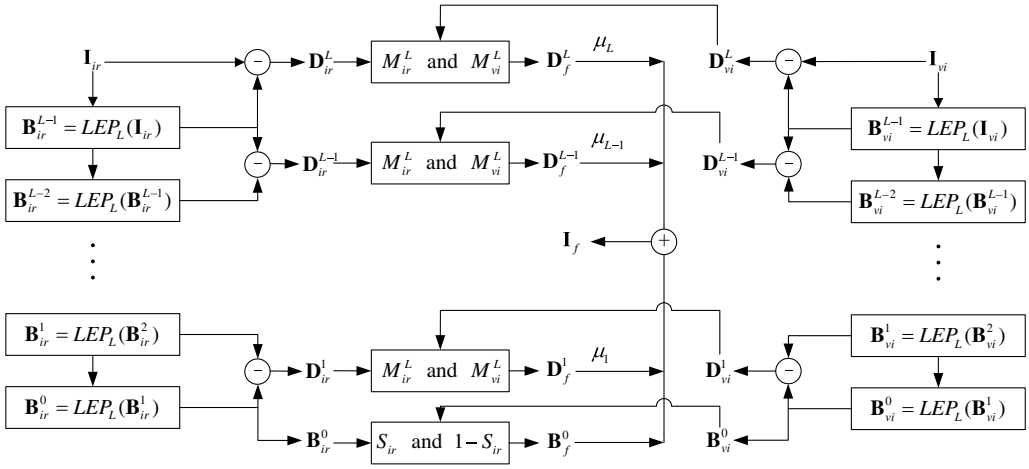


Fig. 3. Flow chart of the proposed method.

to achieve this goal [41–44]. The experiments are performed on a laptop with a 3.3 GHz Intel Core CPU, 8 GB memory, and using the MATLAB code.

A. Datasets and Settings

In our experiment, we first focus on qualitative and quantitative comparisons on the fusion performance of different methods on aligned infrared and visible image pairs. The surveillance images come from TNO Human Factors, which contains multi-spectral nighttime imagery of different military relevant scenarios, registered with different multi-band camera systems. We choose seven typical pairs with the names 2_men_in_front_of_house, bench, lake, sandpath, tank, Duine, and Nato_camp for qualitative illustration, while two image sequences named Duine and Nato_camp are further used for quantitative comparison. The two sequences contain 23 and 32 image pairs, respectively. In addition, the source images are enhanced in terms of their representation using the procedure of contrast stretching.

Several parameters need to be determined for our proposed algorithm: the LEP factor α, β, window w, and synthetic weight parameter μ_l. The effects of different LEP factors' values are demonstrated in Fig. 4 and the test image, derived from the WLS method [8], is used for demonstrating the results under different parameters by those authors. We finally set the parameters as $\alpha = 0.1$ and $\beta = 1$, which can smooth the output image and preserve the local salient edges at the same time. Windows w are set as $w_l = \{3, 7, 15\}(l = 1, 2, 3)$. The original images are decomposed into three levels, according to Eq. (14), and the synthetic weight parameters are set as $\mu_l = \{1.6, 1.1, 2.1\}(l = 1, 2, 3)$. The synthetic weight parameters work well within the range of [1, 2.5]. If the value is too large, some noise may be brought in the fused images.

B. Fusion Quality Assessment

The results of image fusion are typically assessed either subjectively or objectively. Recent studies have proposed various fusion quality assessment metrics. The basis for most of these metrics is the measurement of the transfer of a feature, such as edges and amount of information, from the source images into the new fused composite image [45]. In this paper, we evaluate the performances of different fusion methods using three metrics, i.e., entropy (EN) [10], spatial frequency (SF) [46], standard deviation (SD) [47], mean structural similarity index (MSSIM) [48], and normalized mutual information (NMI) [49]. The definitions of these four metrics are as follows:

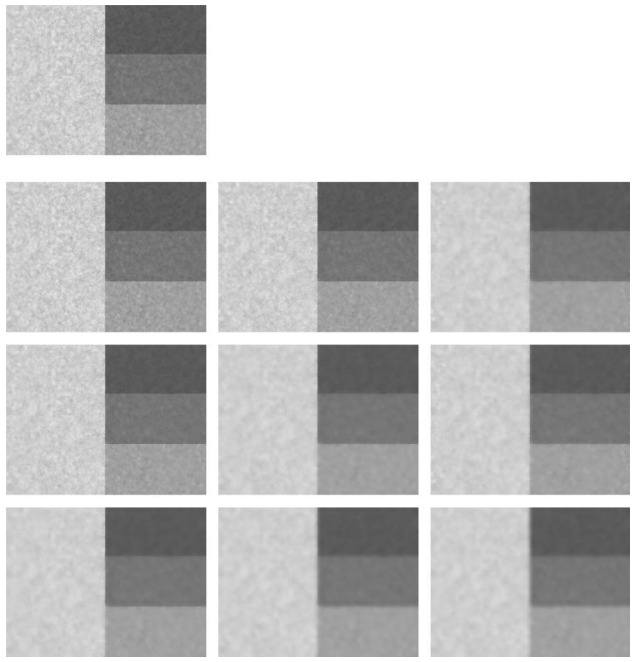


Fig. 4. Influence of different α and β values. The topmost image is the source image. The other nine images are the result of different α and β values. From left to right, the α is 0.01, 0.1, and 1, respectively. From top to bottom, the β is 0.01, 1, and 1.9, respectively.

EN: The EN that measures the amount of information contained in the fused image is defined as follows:

$$\mathrm{EN}(\mathbf{x}) = -\sum_{l=0}^{L-1} P_{\mathbf{x}}(l)\log_2 P_{\mathbf{x}}(l), \tag{15}$$

where L is the number of gray level, which is set to 256 in our experiments, and $P_{\mathbf{x}}(l)$ is the normalized histogram of the fused image $\mathbf{x}$.

SF: Spatial frequencies convey the information about the appearance of a stimulus. High spatial frequencies represent abrupt spatial changes in the image (such as edges), and generally correspond to configural information and fine details. Low spatial frequencies, on the other hand, represent global information about the shape (such as general orientation and proportions).

SD: Standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data values. In image processing, the high SD means the high contrast of the image.

MSSIM: The structural similarity index can be used to quantify the structural similarity between a source image A and a fused image F:

$$\mathrm{SSIM}_{x,y} = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{16}$$

where x and y represent local windows of size $M \times N$ in, respectively, A and F. μ_x and μ_y are the mean values, σ_x and σ_y are the standard deviation, and σ_{xy} is the standard covariance correlation. The SSIM is typically computed over a sliding window to compare local patterns of pixel intensities that have been normalized for luminance and contrast. The MSSIM index quantifies the overall similarity between a source image A and a fused image F:

$$\mathrm{MSSIM}_F^A = \frac{1}{N_\omega} \sum_{i=1}^{N_\omega} \mathrm{SSIM}_{x_i,y_i}, \tag{17}$$

where N_ω represents the number of local windows of the image. An overall image fusion quality index can then be defined as the mean MSSIM values between each of the source images and the fused result:

$$\mathrm{MSSIM}_F^{A,B} = \frac{\mathrm{MSSIM}_F^A + \mathrm{MSSIM}_F^B}{2}. \tag{18}$$

In this paper, MSSIM_F^V and MSSIM_F^I are the MSSIM of the fused result with visible and infrared images, respectively, and $\mathrm{MSSIM}_F^{V,I}$ is their mean.

NMI: Mutual information (MI) measures the amount of information that two images have in common [50]. The mutual information MI_F^A between a source image A and a fused image F is defined as

$$\mathrm{MI}_F^A = \sum_{i,j} P_{A,F}(i,j) \log \frac{P_{A,F(i,j)}}{P_A(i)P_F(j)}, \tag{19}$$

where $P_A(i)$ and $P_F(i)$ are the probability density functions in the individual images, and $P_{A,F}(i,j)$ is the joint probability density function. The traditional mutual information metric is unstable and may bias the measure toward the source image with the highest entropy. This problem can be resolved by computing the NMI as follows:

$$\mathrm{NMI}_F^{A,B} = \frac{\mathrm{MI}_F^A}{H_A + H_F} + \frac{\mathrm{MI}_F^B}{H_B + H_F}, \tag{20}$$

where H_A, H_B, and H_F are the marginal entropy of A, B, and F, and MI_F^A and MI_F^B represent the mutual information between, respectively, the source image A and the fused image F and between the source image B and the fused image F. A higher value of NMI indicates that more information from the source images is transferred to the fused image.

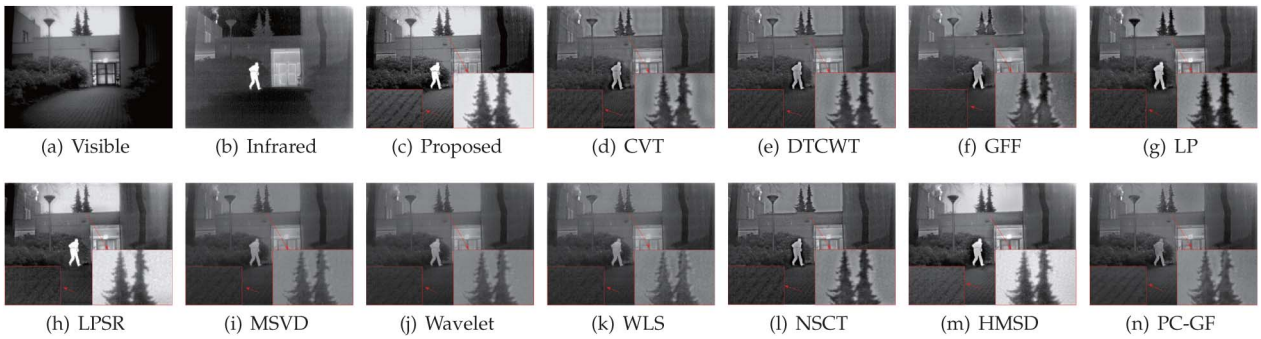
The fusion results are shown in Fig. 5. The experiments are conducted on seven typical pairs named 2_men_in_front_of_house, bench, lake, sandpath, tank, Duine, and Nato_camp for qualitative illustration, and the results are compared with that of CVT, DTCWT, GFF, LP, LPSR, MSVD, wavelet, WLS, NSCT, HMSD, and PC-GF. For each group of results, the first three rows present the original visible image, the original infrared image, and the fusion result of the proposed method, respectively. The rest of the rows correspond to the fusion results of the other methods. From the results, we can observe that our fused images seem like the visible images with abundant details and prominent target areas.

From the results, we can see that for the purpose of image fusion, all the methods work well. It is hard to judge which method is the best; however, we can generally evaluate the results from the following three aspects: (1) whether the image has the high contrast in the target area, (2) whether the image contains abundant details, and (3) the image quality in the non-target areas. As can be seen, targets in our results, such as humans (columns 1, 2, 4, 6, 7), are well highlighted. Our fused images contain abundant details, such as tree branches (columns 1, 3, 4, 5, 7). The non-target areas, such as the sky (column 1), are as clean as the original images.

To better illustrate the quality criterion, we present the partial enlarged detail images of the Kaptein_1123 scene in Fig. 6. The floor and trees are magnified to demonstrate the details. Obviously, the target is the running man. Comparing with the other methods, our method, LPSR and HMSD have the highest illumination in the running man's area, so that the target can be easily noticed. On the other hand, the bottom left corner of the magnified picture exhibits the texture on the floor, which is the detail in the fused image. Only our method can clearly discern the texture. Finally, whether the fused images have artifacts can be discriminated from the bottom right corner of the magnified picture. As we can see, our method, LPSR and HMSD show the good property in preventing the halo and artifacts.

Then we give the quantity assessment of these six scenes in Table 1. For each assessment criterion the greater value means better performance. In most scenes, the proposed method could achieve the highest value in EN, SF, SD, and MSSIM_F^V and relatively high value in NMI. That means the fused images have high contrast and abundant details and they obtain more information from the source images than most other methods. The MSSIM is used for measuring the similarity between two images. As we can see in Fig. 5, the fused results look like the visible images with abundant details and high luminance in the targets areas, so the proposed method works well in the MSSIM_F^V. Furthermore, we give the indices figures of quantitative comparisons of these methods on the Duine and Nato_camp sequences. Some examples of image pairs from the two sequences and the corresponding fusion results are

Fig. 5. Some fusion results of each algorithm. The scenes are 2_men_in_front_of_house, bench, lake, sandpath, tank, Duine, and Nato_camp. The first three rows are the visible image, infrared image, and our fused image; the other rows are the results of compared methods of CVT, DTCWT, GFF, LP, LPSR, MSVD, RP, wavelet, WLS, NSCT, HMSD, and PC-GF.



(a) Visible (b) Infrared (c) Proposed (d) CVT (e) DTCWT (f) GFF (g) LP

(h) LPSR (i) MSVD (j) Wavelet (k) WLS (l) NSCT (m) HMSD (n) PC-GF

Fig. 6. Fused image of the Kaptein_1123 scene. From left to right, top to bottom, they are the visible image, infrared image, the result of the proposed method, CVT [16], DTCWT [38], GFF [20], LP [12], LPSR [10], MSVD [39], wavelet [40], WLS [8], NSCT [17], HMSD [25], and PC-GF [24], respectively. The floor and two trees behind the building are magnified.

Table 1. Quantitative Assessments (the EN, SF, SD, $MSSIM_F^V$, $MSSIM_F^I$, $MSSIM_F^{V,I}$, and NMI in Each Cell) of Different Fusion Methods in Different Scenes[a]

Scenes	Indexes	Proposed	CVT	DTCWT	GFF	LP	LPSR	MSVD	Wavelet	WLS	NSCT	HMSD	PC-GF
2 men in front of house	EN	6.9632	6.6638	6.6546	**7.2518**	6.7831	7.1301	6.4700	6.4261	6.5106	6.6949	7.0773	6.4121
	SF	**7.8831**	5.7364	5.7463	5.6928	5.8793	5.8303	5.3659	3.6895	5.1383	5.8504	6.4003	4.2342
	SD	**53.135**	21.917	27.434	44.620	30.155	52.146	23.850	23.211	24.479	27.968	45.914	21.986
	NMI	0.5860	0.6692	0.6747	0.6553	0.6794	0.6744	0.3461	0.4387	0.4801	0.4539	0.4329	0.4407
	$MSSIM_F^V$	0.8835	0.8724	0.8789	**0.9698**	0.8724	0.8976	0.8294	0.8270	0.8995	0.8606	0.8718	0.7499
	$MSSIM_F^I$	0.3726	0.5750	0.5767	0.4568	0.5891	0.5535	0.6951	0.7470	0.6048	0.6124	0.5570	**0.7877**
	$MSSIM_F^{V,I}$	0.6281	0.7237	0.7278	0.7133	0.7308	0.7256	0.7623	**0.7870**	0.7521	0.7365	0.7144	0.7688
Bench	EN	7.4699	6.9685	6.9390	**7.5215**	7.1144	7.4337	6.6913	6.5319	6.8030	6.9607	7.3898	6.7532
	SF	**12.344**	10.611	10.599	10.585	10.650	10.658	10.059	7.4912	9.7313	10.600	11.034	8.2173
	SD	30.385	28.659	32.242	35.196	60.080	33.297	**48.063**	27.158	27.180	44.111	29.330	37.280
	NMI	0.3934	0.4604	0.5077	0.5064	0.4655	0.4594	0.5132	0.4717	0.4850	**0.5905**	0.5117	0.5157
	$MSSIM_F^V$	0.7651	0.7062	0.8333	0.8445	0.8829	0.7169	0.8820	0.9119	0.9172	**0.9896**	0.9199	0.9106
	$MSSIM_F^I$	0.3761	**0.5205**	0.3061	0.2501	0.2297	0.5121	0.3651	0.5708	0.5701	0.4742	0.5672	0.5600
	$MSSIM_F^{V,I}$	0.4807	0.5387	0.5391	0.5641	0.5455	0.5459	0.5706	0.6134	0.5697	0.5473	0.5563	**0.6145**
Lake	EN	**7.4016**	6.6741	6.6578	7.3687	6.7550	7.1976	6.5683	6.5328	6.6130	6.6852	7.1569	6.5045
	SF	**8.1626**	5.6114	5.5899	5.7126	5.6665	5.7088	5.0750	3.1921	4.9603	5.6369	6.3033	3.5541
	SD	**48.063**	27.158	27.180	44.111	29.330	37.280	24.684	24.101	25.763	27.381	37.926	25.096
	NMI	0.5132	0.4717	0.4850	**0.5905**	0.5117	0.5157	0.3724	0.4481	0.5094	0.4802	0.4696	0.4355
	$MSSIM_F^V$	0.8820	0.9119	0.9172	**0.9896**	0.9199	0.9106	0.8584	0.8410	0.9306	0.9100	0.9171	0.7904
	$MSSIM_F^I$	0.3651	0.5708	0.5701	0.4742	0.5672	0.5600	0.7021	0.7731	0.6036	0.5889	0.5304	**0.7913**
	$MSSIM_F^{V,I}$	0.6236	0.7414	0.7437	0.7319	0.7435	0.7353	0.7803	**0.8071**	0.7671	0.7494	0.7238	0.7908
Sandpath	EN	**7.2040**	6.4548	6.4059	6.8402	6.5898	7.1107	6.1616	6.0840	6.6190	6.4548	7.1324	6.4192
	SF	**9.8059**	7.0677	6.9077	6.6894	7.7551	7.7702	6.8572	4.8114	5.6727	7.4500	8.6213	5.8087
	SD	**40.987**	23.082	22.405	33.654	25.627	39.371	19.201	18.458	19.403	23.016	39.152	24.752
	NMI	**0.5029**	0.4030	0.4313	0.4982	0.4479	0.4551	0.3493	0.4085	0.4185	0.4384	0.4028	0.4380
	$MSSIM_F^V$	**0.9012**	0.7282	0.8761	0.8719	0.8443	0.8716	0.7680	0.7717	0.7750	0.8531	0.8521	0.8397
	$MSSIM_F^I$	0.2367	0.4728	0.4523	0.4756	0.4649	0.4331	0.6445	**0.6941**	0.6636	0.5059	0.4518	0.5964
	$MSSIM_F^{V,I}$	0.5690	0.6678	0.6642	0.6737	0.6546	0.6524	0.7062	**0.7329**	0.7193	0.6795	0.6520	0.7180
Tank	EN	7.8740	7.4061	7.4032	7.7938	7.4472	**7.9162**	7.2433	7.1859	7.3162	7.5401	7.8945	7.7494
	SF	**12.181**	11.372	11.229	10.942	11.415	11.595	10.873	8.6654	10.037	11.348	11.706	10.394
	SD	**76.021**	51.331	50.908	61.049	54.508	69.035	39.379	38.575	40.701	51.008	69.487	58.459
	NMI	0.4894	0.3907	0.4265	**0.5144**	0.4597	0.4710	0.2272	0.4077	0.4249	0.4358	0.4007	0.4599
	$MSSIM_F^V$	**0.7539**	0.6165	0.6161	0.5948	0.6129	0.6938	0.4714	0.5159	0.4600	0.6171	0.7032	0.7071
	$MSSIM_F^I$	0.1124	0.1376	0.1446	0.2643	0.1572	0.1583	0.1915	0.2277	**0.3200**	0.1588	0.1776	0.1651
	$MSSIM_F^{V,I}$	0.4331	0.3770	0.3804	0.4396	0.3851	0.4260	0.3315	0.3718	0.3900	0.3879	**0.4404**	0.4361
Keptein_1123	EN	**7.2888**	6.7778	6.7052	6.8575	6.7761	7.2024	6.5531	6.5170	6.5845	6.7208	7.1628	6.7485
	SF	**6.5129**	5.1898	5.0874	4.6267	5.3146	5.3116	4.7023	3.3338	3.9948	5.1615	6.0900	3.8423
	SD	**58.915**	34.158	33.615	32.898	36.264	56.012	31.619	31.454	31.477	33.880	51.728	32.069
	NMI	0.4610	0.3948	0.4134	**0.4796**	0.4539	0.4496	0.2776	0.4131	0.4306	0.4423	0.4000	0.4292
	$MSSIM_F^V$	**0.8274**	0.6750	0.6834	0.6417	0.6626	0.6660	0.7424	0.7776	0.7318	0.6809	0.6858	0.7949
	$MSSIM_F^I$	0.5174	0.7848	0.7948	**0.8399**	0.8029	0.7793	0.7681	0.7982	0.8222	0.8083	0.7548	0.7537
	$MSSIM_F^{V,I}$	0.6724	0.7299	0.7391	0.7408	0.7328	0.7226	0.7552	**0.7879**	0.7770	0.7446	0.7203	0.7743

[a]The highest values are shown in bold.

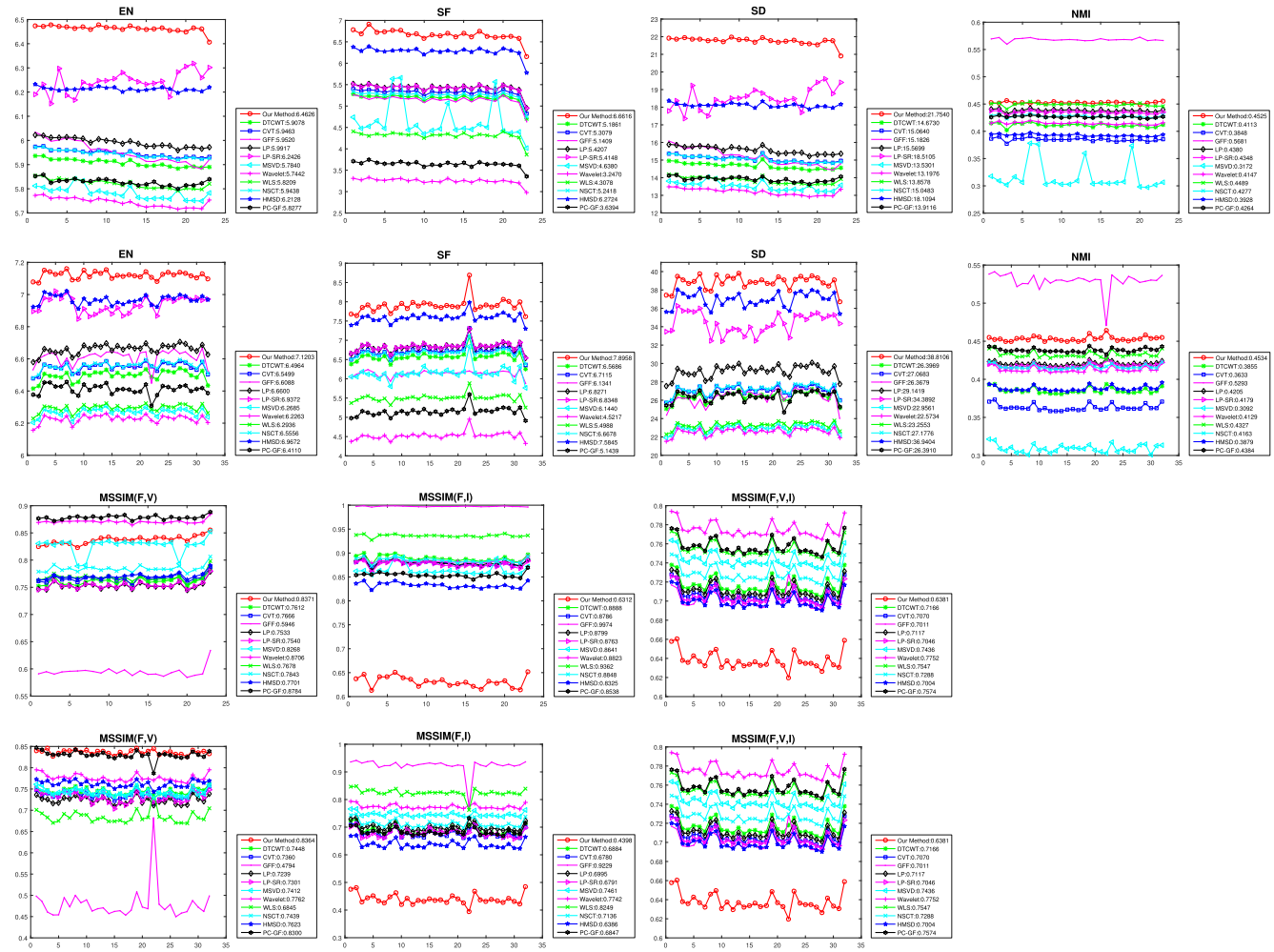


Fig. 7. Quantitative analysis with the sequence datasets. The top figures are the Duine sequence and the bottom figures are the Nato_camp sequence. The assessment criteria are entropy (EN), spatial frequency (SF), standard deviation (SD), mean structural similarity (MSSIM_F^V, MSSIM_F^I, $\text{MSSIM}_F^{V,I}$), and normalized mutual information (NMI). The proposed method is marked with red circles; the compared algorithms are CVT, DTCWT, GFF, LP, LPSR, MSVD, wavelet, WLS, NSCT, HMSD, and PC-GF, respectively. For each assessment criterion the greater value means better performance.

shown in the last two columns of Fig. 5. From the results, we can observe how situational awareness can be improved by fusing the infrared and visible images in these two scenes. In a visible image, distinguishing a person from his background is difficult; however, this person becomes more distinct in an IR image. Similarly, the easily distinguishable background in the visible image becomes nearly indiscoverable in the IR image. Our fused image provides an enhanced rendition of the complete scene provided by both original images.

The quantitative comparisons on the two sequences are given in Fig. 7. We see that the results are similar on the two sequences. Our method marked with red circles consistently has the highest ENs, SFs, and SDs on all image pairs, followed by the LPSR and HMSD methods. In the NMI metric result, the proposed method gets the second highest scores, and the GFF shows the best performance.

The run-time comparison of 12 algorithms on the two sequences are given in Table 2, where the images are all of size 270 × 360, and each value denotes the mean and variance of

Table 2. Run Time Comparison of 12 Algorithms on the Duine and Nato_Camp Sequences, Where Each Value Denotes the Mean and Standard Deviation of Run Time of a Certain Method on a Sequence (Unit: Second)

Method	Duine	Nato_camp
LP	0.0043 ± 0.0007	0.0043 ± 0.0006
LPSR	0.0111 ± 0.0026	0.0087 ± 0.0005
WLS	0.0863 ± 0.0048	0.0886 ± 0.0056
GFF	0.0871 ± 0.0067	0.0927 ± 0.0091
DTCWT	0.1170 ± 0.0093	0.1195 ± 0.0022
Proposed Method	0.1494 ± 0.0085	0.1575 ± 0.0056
Wavelet	0.1550 ± 0.0382	0.1592 ± 0.0018
MSVD	0.1674 ± 0.0031	0.1695 ± 0.0024
HMSD	0.5441 ± 0.0558	0.5492 ± 0.0328
CVT	0.6592 ± 0.0299	0.6692 ± 0.0060
NSCT	1.4385 ± 0.0092	1.4402 ± 0.0096
PC-GF	2.7571 ± 0.0908	2.7972 ± 0.2530

run time of the corresponding method on a sequence. The proposed method can achieve comparable efficiency compared with the other 11 methods.

5. DISCUSSION

A multi-scale decomposition and saliency-map-based infrared and visible image fusion method is proposed in this paper. The fused image has abundant details and prevents artifacts and halo. Through coefficients adjustment, the base layer of the multi-scale decomposition smoothes the oscillating details and preserves edges, and detail layers contain abundant details. The targets in the fused image preserve high luminance and make them easily to discern. Since the base layer image contains most of the energy, we take the modified saliency map into the base layer fusion strategy. The saliency modification is worked on the infrared image, since the luminance in the infrared image reflects the thermal radiation. It is also noted that the nontarget areas could keep clean as visible images, because these areas' ratio is relatively low in the saliency map, and the visible image dominates the areas as a result.

The quantitative analysis demonstrates the proposed method's good performance on the indices of EN, SF, SD, and NMI. The proposed method could preserve high luminance of targets and abundant details, and achieve a relatively high-contrast fused image. However, in the $\mathrm{MSSIM}_F^{A,B}$ and MSSIM_F^{I}, the proposed method, HMSD, and LPSR do not perform as well as the other methods. Because the structural similarity is a perception-based model that considers image degradation as perceived change in structural information and the fused images of these methods all look like visible images but unlike infrared images, so they perform well on the MSSIM_F^{V} but not so well on the MSSIM_F^{I}.

The proposed method still has a shortcoming in that it assumes the targets have relatively higher thermal radiation than background, so that the saliency map extraction method is based on the luminance of infrared images. However, in some cases, such as the fire scene, the environmental temperature is higher than people, and the proposed method may not be able to achieve a desirable result.

6. CONCLUSION

Aiming at the infrared and visible image fusion, we propose a multi-scale fusion method based on the LEP filter and saliency detection. We introduce the LEP filter into a multi-scale image fusion framework, and develop a LC-based saliency algorithm to adapt to the infrared image characteristics in which the target commonly has relatively high thermal radiation. Our fused result looks like a visible image with abundant details and high luminance on the target, which would be helpful for target recognition. The qualitative comparison reveals that our method can enhance the target, preserve details, and prevent artifacts. The quantitative analysis illustrates that the proposed method can be competitive with or even outperform some state-of-the-art methods.

The surveillance images come from TNO Human Factors; see Dataset 1, Ref. [51], which contains multi-spectral nighttime imagery of different military-relevant scenarios, registered with different multi-band camera systems.

Funding. National Natural Science Foundation of China (NSFC) (61503288, 61605146); China Postdoctoral Science Foundation (2016M592385, 2016T90725); Fundamental Research Funds for the Central Universities of China (2042016KF0017); Ph.D. Programs Foundation of Ministry of Education of the People's Republic of China (MOE) (20120142110088).

REFERENCES

1. J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," Inf. Fusion **31**, 100–109 (2016).
2. V. Tsagaris and V. Anastassopoulos, "Fusion of visible and infrared imagery for night color vision," Displays **26**, 191–196 (2005).
3. A. Toet, J. K. IJspeert, A. M. Waxman, and M. Aguilar, "Fusion of visible and thermal imagery improves situational awareness," Displays **18**, 85–95 (1997).
4. Y. Ma, J. Chen, C. Chen, F. Fan, and J. Ma, "Infrared and visible image fusion using total variation model," Neurocomputing **202**, 12–19 (2016).
5. J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," Pattern Recognit. **48**, 772–784 (2015).
6. Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," IEEE Trans. Image Process. **26**, 2545–2560 (2017).
7. S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: a survey of the state of the art," Inf. Fusion **33**, 100–112 (2017).
8. Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *ACM Transactions on Graphics (TOG)* (ACM, 2008), Vol. **27**, p. 67.
9. B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," IEEE Trans. Image Process. **22**, 70–79 (2013).
10. Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," Inf. Fusion **24**, 147–164 (2015).
11. D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-7**, 165–177 (1985).
12. P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **31**, 532–540 (1983).
13. A. Toet, "Image fusion by a ratio of low-pass pyramid," Pattern Recognit. Lett. **9**, 245–253 (1989).
14. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989).
15. N. Kingsbury, "The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement," in *9th European Signal Processing Conference (EUSIPCO)* (IEEE, 1998), pp. 1–4.
16. F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," Inf. Fusion **8**, 143–156 (2007).
17. T. Li and Y. Wang, "Biological image fusion using a NSCT based variable-weight method," Inf. Fusion **12**, 85–92 (2011).
18. Q. Xiao-Bo, Y. Jing-Wen, X. Hong-Zhi, and Z. Zi-Qian, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," Acta Automat. Sinica **34**, 1508–1514 (2008).
19. Y. Jiang and M. Wang, "Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter," IET Image Process. **8**, 183–190 (2014).
20. S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," IEEE Trans. Image Process. **22**, 2864–2875 (2013).
21. A. Toet, "Iterative guided image fusion," PeerJ. Comput. Sci. **2**, e80 (2016).
22. L. G. Leya and V. G. Biju, "A modified image fusion using guided filtering and bilinear interpolation," Int. J. Sci. Res. Eng. Technol. **4**, 412–419 (2015).

23. Pritika and S. Budhiraja, "Multimodal medical image fusion using modified fusion rules and guided filter," in *International Conference on Computing, Communication & Automation (ICCCA)* (IEEE, 2015), pp. 1067–1072.
24. 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," Infrared Phys. Technol. **72**, 37–51 (2015).
25. 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 **30**, 15–26 (2016).
26. Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," Appl. Opt. **55**, 6480–6490 (2016).
27. L. Itti and C. Koch, "Computational modelling of visual attention," Nat. Rev. Neurosci. **2**, 194–203 (2001).
28. C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," in *Matters of Intelligence* (Springer, 1987), pp. 115–141.
29. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998).
30. Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the Eleventh ACM International Conference on Multimedia* (ACM, 2003), pp. 374–381.
31. J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Neural Information Processing Systems (NIPS)* (2006), Vol. **1**, p. 5.
32. S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1915–1926 (2012).
33. R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)* (IEEE, 2009), pp. 1597–1604.
34. M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," IEEE Trans. Pattern Anal. Mach. Intell. **37**, 569–582 (2015).
35. Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM International Conference on Multimedia* (ACM, 2006), pp. 815–824.
36. 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 Phys. Technol. **56**, 93–99 (2013).
37. W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," Pattern Recognit. Lett. **28**, 493–500 (2007).
38. J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," Inf. fusion **8**, 119–130 (2007).
39. V. Naidu, "Image fusion technique using multi-resolution singular value decomposition," Defence Sci. J. **61**, 479–484 (2011).
40. L. J. Chipman, T. M. Orr, and L. N. Graham, "Wavelets and image fusion," in *Proceedings of the International Conference on Image Processing* (IEEE, 1995), Vol. **3**, pp. 248–251.
41. J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," IEEE Trans. Image Process. **25**, 53–64 (2016).
42. K. Yang, A. Pan, Y. Yang, S. Zhang, S. H. Ong, and H. Tang, "Remote sensing image registration using multiple image features," Remote Sens. **9**, 581 (2017).
43. J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," IEEE Trans. Image Process. **23**, 1706–1721 (2014).
44. J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," Pattern Recognit. **46**, 3519–3532 (2013).
45. N. Cvejic, T. Seppanen, and S. J. Godsill, "A nonreference image fusion metric based on the regional importance measure," IEEE J. Sel. Top. Signal Process. **3**, 212–221 (2009).
46. M. Bar, "Visual objects in context," Nat. Rev. Neurosci. **5**, 617–629 (2004).
47. D.-C. Chang and W.-R. Wu, "Image contrast enhancement based on a histogram transformation of local standard deviation," IEEE Trans. Med. Imaging **17**, 518–531 (1998).
48. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).
49. M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'information measure for performance of image fusion'," Electron. Lett. **44**, 1066–1067 (2008).
50. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, 2012).
51. http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.