Circuits, Systems, and Signal Processing
https://doi.org/10.1007/s00034-019-01131-z

Multi-scale Guided Image and Video Fusion: A Fast and Efficient Approach

Durga Prasad Bavirisetti[1] · Gang Xiao[1] · Junhao Zhao[1] · Ravindra Dhuli[2] · Gang Liu[3]

Received: 27 May 2018 / Revised: 23 April 2019 / Accepted: 25 April 2019

Abstract
In this paper, we propose a general purpose, simple and fast fusion algorithm based on guided image filter. The proposed method can well combine useful source image information into the fused image supported by multi-scale image decomposition, structure transferring property, visual saliency detection, and weight map construction. Multi-scale image decomposition is appropriate to represent and manipulate image features at various scales. Structure transferring property enabled by our algorithm can induce structures of one source image into the other. A new visual saliency detection based on guided image filter introduced in this paper is able to extract significant regions from visually different images of the same scene. The choice of weight maps helped to integrate the complementary information pixel by pixel at each scale. Experimental outcomes of the proposed method are compared and analyzed with traditional and recent guided image filter-based fusion algorithms in terms of visual quality, fusion metrics and run time. In addition, to enhance fusion results further we made an effort to find a suitable image and video enhancement algorithm. The fusion performance analysis clearly indicates that the proposed method is very promising along with less run time.

Keywords Edge preserving · Guided image filter · Image fusion · Multi-scale image decomposition · Structure transferring · Visual saliency

1 Introduction

Data fusion [1, 11, 30, 42], especially image fusion [1], has attracted the attention of researchers over the past few decades because of its wide variety of applications. In computer vision, image acquisition in various visual conditions [7, 9, 21] or spectral

✉ Gang Xiao
xiaogang@sjtu.edu.cn

Extended author information available on the last page of the article

Published online: 20 May 2019

ranges [1, 3, 35] of the same scene is quite common for accurate scene identification and representation. These individual captured images may provide redundant and complementary information. The information of these images should be combined into a single image by eliminating the redundant information for better scene understanding. For example, a visible camera [7] cannot focus more than one objects of a scene simultaneously due to the depth of field limitations. Specific objects in the scene need to be focused and captured one after the another. These captured images may provide unique information in a region whereas redundant information in another region. It is difficult to get the whole picture of a scene from these individual images. Hence, a single image which can provide all focused regions needs to be generated out of these.

Considering another scenario [1, 35] of capturing images during the night or under the poor light conditions, visible sensors can capture images well under better lighting conditions. But, in surveillance applications like military and navigation, surroundings need to be monitored day and night. Here, infrared or thermal cameras come for rescue by providing information about the scene in low visibility as a thermal map/infrared image based on temperature variations. However, either visible (VI) or infrared (IR) images alone are not sufficient to provide all the information about the scene. Therefore, the required details of the scene from both the images need to be integrated into a single image.

The process of combining or merging information of source images into a single image for better scene prediction and understanding is referred to as image fusion [1, 3, 7, 9, 21, 35]. Robotics, autonomous vehicles, visual inspection in industrial plants, visual surveillance, ambient intelligence, person re-identification, digital photography, medical imaging, remote sensing and biometric, etc., are various application areas of image fusion [7]. It can be performed at pixel, feature and decision levels. In this paper, we present a simple yet efficient pixel level fusion algorithm based on guided image filter or guided filter (GF).

The remainder of the paper is organized as follows. Section 2 reviews the background and state-of-the-art of image fusion algorithms based on GF. Section 3 briefs the preliminaries. Section 4 presents the proposed method. Experiments and analysis of results are discussed in Sect. 5. Section 6 concludes the paper.

2 Background on Image Fusion

Information of source images need to be associated into a single image due to four major aspects: fusion gain, fusion loss, fusion artifacts and run time. The amount of information transferred from source images into the fused image is referred to as fusion gain. Fusion loss is a measure of information loss during the fusion process. Visual information which is not relevant to that of source images introduced into the fused image is termed as fusion artifact. Run time is the amount of time required for executing the fusion algorithm.

An efficient image fusion algorithm has to transfer information from source images by maximizing the fusion gain and at the same time minimizing the fusion loss, fusion artifacts and run time. By considering these requirements, numerous fusion algorithms

were proposed. Since our interest is on image fusion based on GF, the background details of these algorithms are well presented here below.

Initially, Li et al. [24] proposed a fusion algorithm based on GF which we refer to as guided filter fusion (GFF) in this paper. In GFF method, source images are decomposed into base and detail layers based on an average filter. Then, saliency maps and corresponding initial weight maps are calculated using Laplacian and Gaussian operators. Next, initial weight maps are refined using GF to get final weight maps corresponding to each base and detail layers. Finally, base and detail layers are combined with these weights. This approach exhibited superior performance over the existing multi-scale decomposition-based fusion methods for many image fusion applications. However, it has some limitations which are addressed by various new fusion algorithms.

Gan et al. [12] mentioned two main problems with GFF. Firstly, saliency maps generated by the Laplacian operator in GFF do not preserve all the image features. In addition, GFF did not exploit the advantages of multi-scale decomposition. To address these, they proposed a fusion algorithm [12] based on weighted least squares filter, phase congruency, and GF. This method is similar to the GFF framework except for some differences. The average filter is replaced by weighted least squares filter in multi-scale decomposition mode while Laplacian and Gaussian operators are replaced with phase congruency. Qualitative and quantitative results are compared with GFF and other methods, but no run time comparison is performed. In addition, this method is tested on visible and infrared image datasets only.

Jameel et al. [17–19] proposed a series of image fusion algorithms based on GF. In [17], they developed an image fusion algorithm for medical images based on linear minimum square error estimator and modified weight map strategy to address various limitations of GFF. For instance, the Gaussian filter in GFF cannot remove Rician noise. Binary assignment in GFF may discard the effect of one value when both weights are equal. In [18], another algorithm for multi-focus images is proposed using the shift-invariant wavelet transform (SIWT) and GF for eliminating the noise effects. Here, instead of using an average filter in GFF, SIWT is considered for multi-scale decomposition, whereas remaining steps are similar to GFF. In [19], an alternative algorithm is developed for multi-focus images using the weight maps based on focus information.

Javed et al. [20] proposed a new method for MRI and CT image fusion by exploiting the fractal dimension and GF. The methodology adopted in these papers is similar to that of the GFF framework, by modifying some components for improved performance. In [33], Singh et al. proposed an image fusion algorithm based on GF and pyramid decomposition. This method follows an approach entirely different from the state-of-the-art fusion algorithms at that time. Here, only the multi-exposure image fusion problem is addressed. Pritika et al. [31] proposed an image fusion method for medical images using multi-scale guided image decomposition with layer dependent fusion rules. Zhou et al. [44] developed an image fusion method based on multi-scale guided image decomposition to fuse infrared and visible images. It is computationally expensive compared to that of the GFF approach.

Shuaiqi et al. [26] proposed two image fusion algorithms based on GF. The first one is the guided image fusion method performed in complex shearlet transform (CST) domain. This method may exhibit high computational complexity because it requires

a combination of various filters or algorithms. Added to this, no run time analysis is presented. Another fusion algorithm [36] is based on rolling guided image filter (RGF) and spiking cortical model (SCM) for medical images. This method also did not take run time calculation into the consideration.

Toet et al. [38] used RGF-based image decomposition and weights calculation using frequency tuned saliency maps for fusion. Another method along the same lines is presented in [37]. In both of these methods, experimental results are limited to infrared and visible image datasets; no comparison with GFF is presented and also no attention is paid to run time analysis.

Bavirisetti et al. [4] proposed a fusion method (GFS) based on GF and image statistics. This method is designed to apply on both single and multi-sensor images [4]. Even though results of GFS are reasonable, run time is very high since the fusion rule employed in this algorithm depends on the statistical properties of the neighborhood.

Except for GFF and GFS, other methods are application dependent. All the fusion methods based on GF depend on other tools and techniques for saliency map extraction or weight map construction process. Majority of them are computationally expensive and runtime comparison not given due priority. In addition, none of them are tested on video datasets.

By observing the drawbacks of existing image fusion methods based on GF, motivated by the advantages of multi-scale guided image decomposition and weight map construction based on self-saliency extraction process, we introduce a simple but computationally efficient fusion algorithm termed as multi-scale guided filtered-based fusion (MGF). The proposed MGF, inspired from [4], can well combine useful source image information into the fused image by exploiting the advantages of multi-scale image decomposition and structure transferring property of the GF, thereby developing a new visual saliency extraction and weight map construction process. Multi-scale image decomposition is appropriate to represent and manipulate image features at various scales. Structure transferring property of the GF enabling by our algorithm can transfer structures of one source image into the other. Visual saliency detection (VSD) process based on detail layer information with multi-scale GF developed in our algorithm can identify significant source image information. Weight map construction process based on visual saliency can integrate complementary information pixel by pixel. Hence, the proposed method is able to transfer pixel-wise complementary source image information into the fused image at each scale.

The contributions of this work are:

A. A new VSD algorithm based on GF is proposed for extracting the saliency information from visually different images.
B. A general-purpose image fusion algorithm based on GF is developed to address various applications of image fusion.
C. A single multi-scale GF is sufficient for extracting visual saliency as well as to perform image fusion. Hence, a drastic reduction in the complexity of the algorithm is achieved. However, existing GF-based fusion approaches use separate tools or techniques for feature extraction (edge and saliency information) and fusion.

D. Extensive simulations are carried out on 50 image datasets as well as on static and dynamic video datasets.
E. Suitable image and video enhancement algorithm to further enhance the visual content of the fusion results is also presented.

3 Preliminaries

This section covers an overview of GF and multi-scale guided image decomposition.

3.1 An Overview of GF

GF performs smoothing by considering statistical properties of the neighborhood of a pixel into the consideration. It computes output like a linear time-invariant (LTI) filter. However, it uses another image for the guidance purpose. This additional image can be the input or its translated version. One can also choose a completely different image for this purpose. Like other edge-preserving filters, GF can also preserve edge information during the decomposition process which helps in avoiding ringing artifacts. This property makes GF useful in various applications like colorization, up-sampling and image matting. Besides edge-preserving property, GF also has another property called structure transferring property. If the guidance image is same as the input, then edge-preserving smoothing will be performed, whereas the structural behavior remains the same. On the other hand, when the guidance image differs from the input, the smoothing process is regulated by the guidance image structures. As depicted in Fig. 1a (left side), if both the input and guidance are Gaussian signals, then resultant signal also looks like Gaussian. Conversely, in Fig. 1a (right side), if the guidance signal (sinc function) differs from the input (Gaussian), then structures of the guidance signal controls or guides the smoothing process of the input signal. From this figure, we can clearly notice that side lobes of the sinc function are transferred to the Gaussian. The same process can also be observed in Fig. 1b using step and sinusoidal signals. From this illustration, it can be concluded that when we apply GF on images, edge structures can be transferred.

Multi-scale decomposition enables representation of edge information at various resolutions. Simultaneously operating GF can help in transferring edge structures at various scales. The primary requirement of a fusion algorithm is to transfer the information of one image into the other. From the above discussion, multi-scale GF operating in structure transferring mode can support this at various resolutions thereby improving the performance.

Algorithm 1: Guided image filter (GF) [24]

Input: input I, guide G, radius r, regularization ε

Output: output I_G

1: $\text{avg}_G = f_{\text{avg}}(G)$
$\text{avg}_I = f_{\text{avg}}(\text{I})$
$\text{corr}_G = f_{\text{avg}}(G \,.*\, G)$
$\text{corr}_{GI} = f_{\text{avg}}(G \,.*\, I)$

2: $\text{var}_G = \text{corr}_G - \text{avg}_G \,.*\, \text{avg}_G$
$\text{cov}_{GI} = \text{corr}_{GI} - \text{avg}_G .* \text{avg}_I$

3: $a = \text{cov}_{GI} \,./\, (\text{var}_G + \varepsilon)$
$b = \text{avg}_I - a \,.*\, \text{avg}_G$

4: $\text{avg}_a = f_{\text{avg}}(a)$
$\text{avg}_b = f_{\text{avg}}(b)$

5: $I_G = \text{avg}_a \,.*\, G + \text{mean}_b$

/*f_{avg} is an average filter*/

Let I and G represent input and guidance images, respectively. For a neighborhood size r and regulation parameter ε, the guided image filtering is given by $I_G = GF(I, G, r, \varepsilon)$. The parameters r and ε control the edge-preserving process. The algorithm to compute guided image filtering is depicted in Algorithm 1. For more details of GF, one can refer [15]. Now, we move onto the multi-scale guided image decomposition process.

3.2 Multi-scale Guided Image Decomposition and Reconstruction

Using GF, multi-scale decomposition can be performed. In the same way, we can also reconstruct the image without loss of information as shown in Fig. 2. Suppose b^{n-1} G^{n-1} are base layer and guidance images at $(n-1)$th level and r^n, ε^n are neighborhood size and regulation parameter at nth level, then base layer b^n can be calculated by performing guided image filtering on b^{n-1} by considering G^{n-1} as guidance image. Detail images d^n can be calculated by taking the difference of previous base layers b^{n-1} and current base layers b^n. These expressions are given by:

$$\begin{aligned} b^n &= GF\left(b^{n-1}, G^{n-1}, r^n, \varepsilon^n\right) \\ d^n &= b^{n-1} - b^n \end{aligned} \tag{1}$$

In (1), the input image is considered as a base layer at zeroth level ($b^0 = I$) and parameters r^n, ε^n at each level will be considered according to the application requirement.

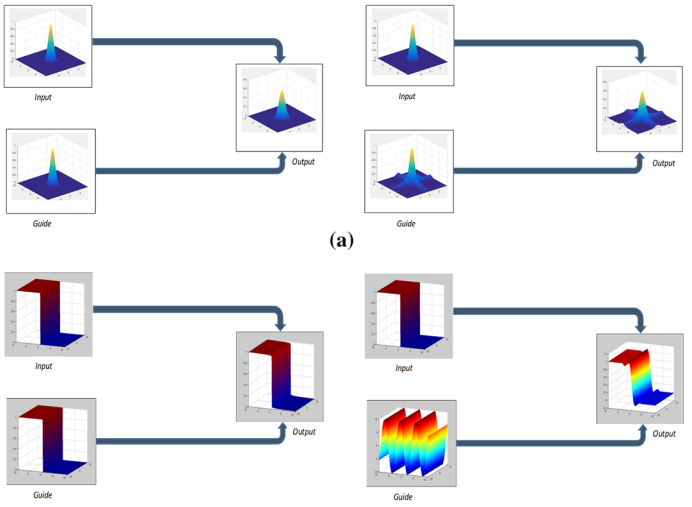

Fig. 1 Demonstration of structure transferring property. **a** Example 1: guided filtering with same Gaussian signal as input and guide (left), Guided filtering with input as Gaussian and guide as sinc function (right); **b** Example 2: guided filtering with same step signal as input and guide (left), guided filtering with input as step and guide as sinusoidal function (right)

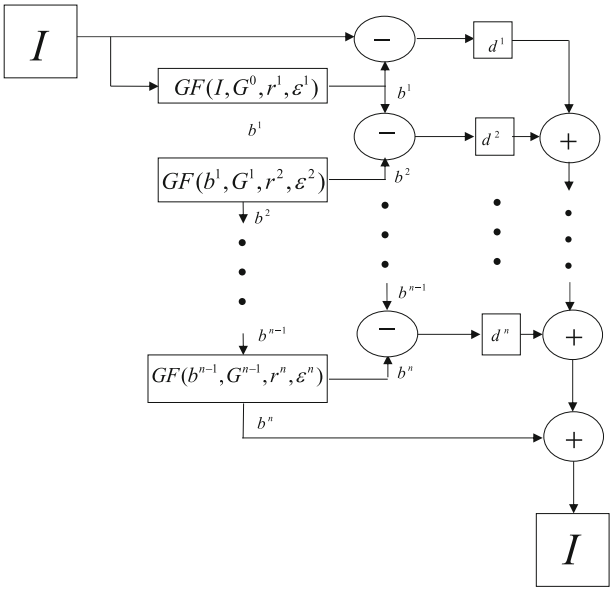

Fig. 2 Multi-scale guided image decomposition of image I

The input image I can be reconstructed from the decomposed base and detail layers as:

$$I = \sum_{k=1}^{n} d^k + b^n \tag{2}$$

This multi-scale decomposition process is also briefed in Algorithm 2.

Algorithm 2: Multi-scale guided image decomposition and reconstruction

Input: input image I, guidance image G^0,
radius $r^1,\ldots, r^n$, regularization $\varepsilon^1,\ldots, \varepsilon^n$

Output: reconstructed image I

1. $b^0 = I$
2. **for** $k = 1,2,3,\ldots,n$ **do**
   $b^k = \mathrm{GF}(b^{k-1}, G^{k-1}, r^k, \varepsilon^k)$
   $d^k = b^{k-1} - b^k$
   end for
3. $I = \sum_{k=1}^{n} d^k + b^n$

4 Proposed Method

The important steps involved in the proposed MGF method (Fig. 3) are summarized as:

A. Multi-scale decomposition of source images with the help of GF.
B. Generating saliency maps.
C. Computing weight maps corresponding to detail layers.
D. Combining the detail layers using weight maps.
E. Generating the final fused image.

A detailed discussion of each stage is presented in the following subsections. In addition, an example of a bottle dataset illustrating the outputs of each stage of the algorithm is presented in Fig. 4 for better understanding. Four levels are considered for image decomposition. Resultant images at various important stages of MGF are displayed in five steps as shown in Fig. 4.

They are source images, base layers, saliency maps, weight maps, and image reconstruction process. MGF method is also briefed in Algorithm 3.

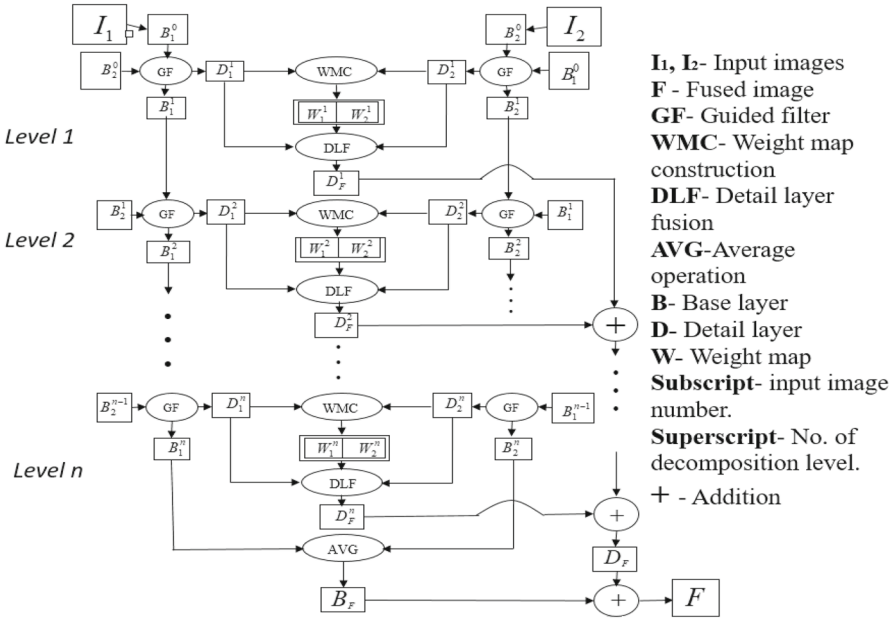

Fig. 3 Multi-scale guided image fusion (MGF)

4.1 Multi-scale Guided Image Decomposition

Let us consider pixel-wise co-registered source images $I_1(x, y)$ and $I_2(x, y)$ of the same size. We performed multi-scale decomposition of I_1 and I_2 using GF for obtaining base layers B_1^1, B_2^1 and detail layers D_1^1, D_2^1 as follows:

$$B_1^1 = GF\left(I_1, I_2, r^1, \varepsilon^1\right) \text{ and } B_2^1 = GF\left(I_2, I_1, r^1, \varepsilon^1\right)$$
$$D_1^1 = I_1 - B_1^1 \text{ and } D_2^1 = I_2 - B_2^1 \tag{3}$$

It can be observed that for filtering I_1, I_2 is used as a guidance image. Hence, structural information of I_2 is used to smooth I_1.

Step 1 Source images	I_1 & I_2	
Step 2 Base layers	B_1^k $\forall k = 1, 2, ...4.$	
	B_2^k	
Step 3 Saliency maps	S_1^k	
	S_2^k	
Step 4 Weight maps	W_1^k	
	W_2^k	
Step 5 Image reconstruction	B_F, D_F & F	

Fig. 4 Top to bottom visual display of results of various steps involved in the proposed method with an example of a battle field dataset

Algorithm 3: Multi-scale Guided Image Fusion

Input: input image I_1 & I_2, radius $r^1, \ldots, r^n$, regularization $\varepsilon^1, \ldots, \varepsilon^n$

Output: fused image F

1. $B_1^0 = I_1, B_2^0 = I_2$
2. **for** $k = 1,2,\ldots,n$ **do**

$$B_1^k = GF\left(B_1^{k-1}, B_2^{k-1}, r^k, \varepsilon^k\right)$$
$$B_2^k = GF\left(B_2^{k-1}, B_1^{k-1}, r^k, \varepsilon^k\right)$$
$$D_1^k = B_1^{k-1} - B_1^k$$
$$D_2^k = B_2^{k-1} - B_2^k$$
$$S_1^k = \left|D_1^k\right|$$
$$S_2^k = \left|D_2^k\right|$$
$$W_1^k = \frac{S_1^k}{\sum_{n=1}^{2} S_n^k}$$
$$W_2^k = \frac{S_2^k}{\sum_{n=1}^{2} S_n^k}$$
$$D_F^k = W_1^k D_1^k + W_2^k D_2^k$$

end for

3. $D_F = \sum_{k=1}^{n} D_F^k$
4. $B_F = \frac{1}{2}\left(B_1^n + B_2^n\right)$
5. $F = B_F + D_F$

Similar operation can be observed on I_2 by taking I_1 as guidance image. The consecutive base layers are generated as:

$$\begin{aligned} B_1^k &= GF\left(B_1^{k-1}, B_2^{k-1}, r^k, \varepsilon^k\right) \\ B_2^k &= GF\left(B_2^{k-1}, B_1^{k-1}, r^k, \varepsilon^k\right), \; k = 1, \ldots, n \end{aligned} \tag{4}$$

where B_1^k, B_2^k are base layers of two source images at k level which depend on their previous level base layers B_1^{k-1}, B_2^{k-1}, respectively. It can be noted that B_1^0, B_2^0 represent two source images I_1 and I_2. The resultant base layers are displayed in step 2 of Fig. 4. Here, first and second rows demonstrate the base layers B_1^k and B_2^k, respectively.

Algorithm 4: VSD based on Multi-scale GF

Input: images I_1 & I_2, radius $r^1,\ldots,r^n$, regularization $\varepsilon^1,\ldots,\varepsilon^n$

Output: visual saliency S

1. $B_1^0 = I_1, B_2^0 = I_2$
2. **for** $k = 1,2,\ldots,n$ **do**

$$B_1^k = GF\left(B_1^{k-1}, B_2^{k-1}, r^k, \varepsilon^k\right)$$
$$B_2^k = GF\left(B_2^{k-1}, B_1^{k-1}, r^k, \varepsilon^k\right)$$
$$D_1^k = B_1^{k-1} - B_1^k$$
$$D_2^k = B_2^{k-1} - B_2^k$$
$$S_1^k = \left|D_1^k\right|$$
$$S_2^k = \left|D_2^k\right|$$

end for

3. $S_1 = \frac{1}{n}\sum_{k=1}^{n} S_1^k$

$S_2 = \frac{1}{n}\sum_{k=1}^{n} S_2^k$

4. $S = \max\left(S_1, S_2\right)$

Detail layers D_1^k, D_2^k represent the difference between previous and current level base layers:

$$D_1^k = B_1^{k-1} - B_1^k,$$
$$D_2^k = B_2^{k-1} - B_2^k. \tag{5}$$

These detail layers provide visually significant source image information which further useful for extracting visual saliency and weight map construction.

4.2 Visual Saliency Detection (VSD)

The process of extracting visually significant regions of a scene is referred to as VSD. This process attempts to mimic the human visual attention on particular objects of a scene. Itti et al. [16] developed the first VSD based on pyramid representation. Multi-scale representation of intensity, color, orientation channels is derived from the Gaussian pyramid. Saliency map is generated by combining these channels pixel by pixel after the pyramid reconstruction. This field is well explored because of the wide variety of applications using various tools and techniques. Our focus is only on multi-scale decomposition-based VSD techniques [16, 25, 27, 29]. A multi-scale decomposition-based VSD is proposed by Ma et al. using wavelets [27]. In this method: firstly, decomposed an image into wavelet coefficients for four levels. Secondly, determined the magnitude of wavelet coefficients to find salient regions. Later, calculated

saliency maps by averaging these coefficients in horizontal, vertical and diagonal directions. Then, the maximum of wavelet coefficients at various color spaces is computed. Finally, visual saliency is computed as the linear combination of these maximum saliencies at different levels.

Inspired by [16, 27], we propose a new multi-scale VSD process for the purpose of image fusion. This algorithm finds visual saliency of the entire scene from two visually different images. The steps are illustrated in Algorithm 4 and briefed as follows:

1. Decompose source images into base layers (B_1^k, B_2^k) and detail layers (D_1^k, D_2^k) using (4) and (5).
2. Compute visual saliencies S_1^k and S_2^k by taking the magnitude of detail layers D_1^k and D_2^k as: $S_1^k = \left|D_1^k\right|$ and $S_2^k = \left|D_2^k\right|$.
3. Find visual saliencies S_1 and S_2 corresponding to source images by averaging S_1^k and S_2^k i.e., $S_1 = \frac{1}{k}\sum_{k=1}^{n} S_1^k,\ S_2 = \frac{1}{k}\sum_{k=1}^{n} S_2^k$
4. Determine the final visual saliency of the entire scene as the maximum of S_1 and S_2 evaluated pixel-wise:$S = \max(S_1, S_2)$.

As displayed in step 3 of Fig. 4, the proposed VSD is identifying the visual saliency of each source image at various scales. In addition, it can represent the visual saliency of the entire scene from two different source images. Visual saliencies of bottle, leopard, aircraft, clock and mug fusion datasets are displayed in Fig. 5. For example, consider the bottle dataset as in the first row of Fig. 5. From source image 1 (Fig. 5a) and its corresponding visual saliency (Fig. 5b), it can be observed that all the visually important information (such as foreground bottle) is identified properly. Similarly, from source image 2 as well, the proposed VSD is identifying the visually prominent background information (such as bottle and other objects). Final visual saliency (S) is obtained by combining these individual saliencies (foreground and background information). It is displayed in Fig. 5f. We can make similar observations for the leopard, aircraft, clock and mug datasets. It can be noted that the proposed VSD can extract visual saliency of the entire scene from two different images by combining their individual saliencies. MGF fused image (all-in-one focused image) is shown in Fig. 5e for reference to verify the effectiveness of the algorithm. From the final visual saliency and fused image, one can conclude that our VSD is extracting the significant information of the visual scene from two different images.

Note that, saliency maps S_1, S_2, S are not used in the proposed MGF method. Because our aim in the fusion process is to combine detail images at each scale using the weight maps developed from the saliency information at that particular scale. So, as shown in Fig. 4 and Algorithm 3, we utilized local saliency maps S_1^k, S_2^k to combine detail image information at every kth level. These saliencies are extracting the significant information at each scale. However, as demonstrated in Fig. 5 and Algorithm 4, one can extract the saliency map of the whole scene from visually different images using the proposed VSD algorithm. Further, this scale specific saliency information is useful in building weight maps for the fusion purpose.

Fig. 5 Visual saliencies of various fusion datasets (bottle, leopard, aircraft, clock and mug) using the proposed VSD method: **a** Source image 1; **b** Visual saliency of (**a**); **c** Source image 2; **d** Visual saliency of (**c**); **e** Fused image generated by the proposed MGF method; **f** Final visual saliency S

4.3 Weight Map Calculation

Existing GF-based fusion methods make use of separate image feature extraction techniques such as saliency map and edge detection methods for the purpose of weight map generation. Unlike these, we are proposing a new weight map construction process by developing saliency maps directly from the detail layers. Owing to this approach, there is a drastic reduction in the computational complexity of the algorithm.

Weight maps are designed by normalizing the saliency maps as follows:

$$W_1^k = \frac{S_1^k}{\sum_{i=1}^{2} S_i^k}, \quad \forall k = 1, 2, \ldots, n.$$
$$W_2^k = \frac{S_2^k}{\sum_{i=1}^{2} S_i^k}, \quad \forall k = 1, 2, \ldots, n. \tag{6}$$

where k denotes the scale. It can be noted that we are calculating weights for each scale.

This complementary weight map construction process based on saliency information can automatically integrate detail layer information at each and every pixel as shown in step 4 of Fig. 4.

4.4 Detail Layer Fusion

The detail layers are integrated at each scale k with the help of weight maps W_1^k and W_2^k using a linear combination:

$$D_F^k = W_1^k D_1^k + W_2^k D_2^k. \tag{7}$$

Final detail layer D_F shown in step 5 of Fig. 4 is obtained by combining fused detail layers obtained at each scale:

$$D_F = \sum_{k=1}^{n} D_F^k. \tag{8}$$

This fused detail image D_F provides most of the visual information of the fused image.

4.5 Base layer Fusion

The fused base layer B_F is generated by taking the average of base layers at the final scale n as:

$$B_F = \frac{1}{2}\left(B_1^n + B_2^n\right). \tag{9}$$

As shown in step 5 of Fig. 4, B_F provides the contrast of the fused image.

4.6 Fused Image Reconstruction

Fused image is obtained by combining the base layer B_F and the detail layer D_F.

$$F = B_F + D_F. \tag{10}$$

As shown in Fig. 4, the fused image contains both contrast (from B_F) and visual information (from D_F) of the scene in a single image. Along with grayscale images, the proposed MGF algorithm is also applied on color images by implementing individually on red, green and blue color channels. Finally, all these channels are concatenated to get the fused color image. More details on color image fusion are available in [5].

5 Experiments and Results

To validate the proposed algorithm, we performed extensive simulations on images and videos. First, we discussed the experimental setup used for this. Then, based on the outcomes, we presented an analysis of visual quality, fusion metrics and run time. Finally, we made an effort to enhance fused images and videos further for better visual quality. We did not include this contrast enhancement concept in the proposed MGF to provide a fair comparison with other fusion methods.

5.1 Experiments

Experiments are carried out in MATLAB (2016b) environment using an Intel Core i7-6700HQ CPU with 2.60 GHz clock speed.

5.1.1 Image and Video Resources

Experiments are conducted on 50 images and 3 video datasets as shown in Fig. 6. Image datasets are divided into five classes based on application. They are multi-exposure (8 datasets), medical imaging (8 datasets), remote sensing (10 datasets), multi-focus (11 datasets) and visual surveillance (13 VI-IR datasets).

These datasets are collected from various web resources and are made available at personal website.[1] Two video datasets considered for these experiments are taken from the well-known TNO Image fusion dataset.[2] Another dataset is adopted from Sarnoff Corporation [41].

5.1.2 Fusion Metrics

Fusion results have to be validated to demonstrate the effectiveness of any fusion method. It can be done in terms of visual quality and fusion metrics. In any image processing scenario, the best way to justify the quality of the resultant image is by means of visual inspection. Quantitative analysis will be considered to further support the validation process. However, for quantitative analysis, there is no gold standard for choosing the right fusion metrics in the field of image fusion. In this paper, we consider both traditional and recent fusion metrics: average gradient (AG) [34], fusion symmetry (FS) [34], correlation coefficient (CC) [34], spatial frequency (SF) [34], universal quality index (Q_0) [43] and visual information fidelity (VIF) [14] for this purpose. AG measures sharpness and clarity of the fused image. FS quantifies the symmetry of the fused image with respect to the source images. CC finds the similarity of the fused image with that of the source images. SF is used to measure the overall activity level of regions in the fused image. Q_0 quantifies the distortion present in an image as a product of three distortions called correlation, luminance, and contrast. VIF measures the trade-off between information present in input images and to the source image information that can be extracted from the fused image. For better performance,

[1] https://sites.google.com/view/durgaprasadbavirisetti/datasets.

[2] https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.

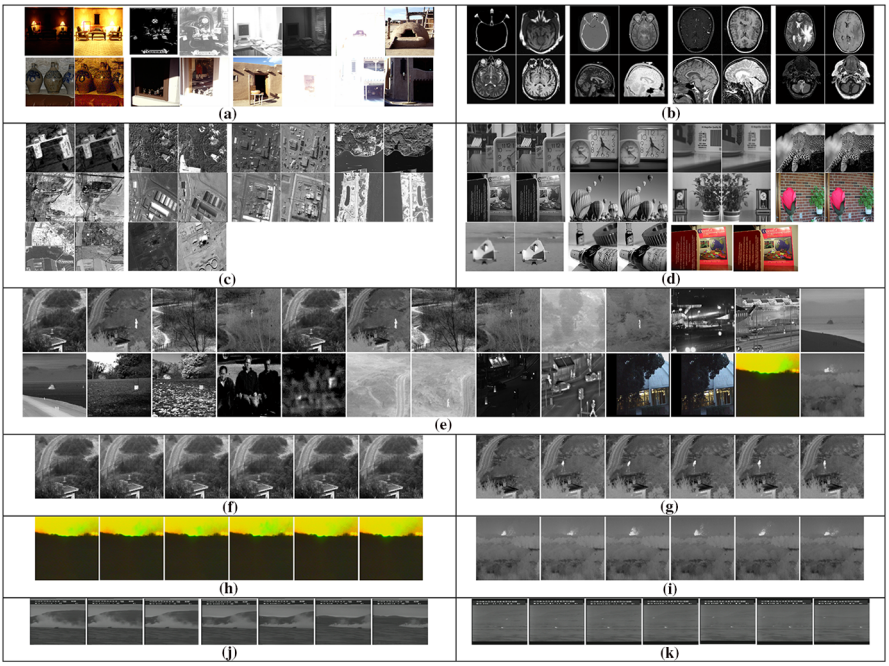

Fig. 6 Image and video datasets used for experiments. **a** Multi-exposure (8 datasets); **b** Medical (8 datasets); **c** Remote sensing (10 datasets); **d** Multi-focus (11 datasets); **e** VI-IR (13 datasets); **f**, **g** VI and IR image frames of Video dataset 1; **h**, **i** VI and IR image frames of Video dataset 2; **j**, **k** VI and IR image frames of Video dataset 3

all of these metrics should have the highest values. For more technical details, one can refer to [14, 34, 43].

5.1.3 Parameter Selection

In the proposed method, GF is utilized to separate base and detail layers. The performance of MGF depends on free parameters: filter size r, smoothing parameter ε and number decomposition levels k. Hence, these parameters have to be tuned appropriately. By following the analysis in [2], these parameters are tuned to $r = 9$, $\varepsilon = 10^3$ and $k = 4$, respectively.

5.2 Analysis of Results

This section is mainly divided into two parts, i.e., image analysis and video analysis. Image analysis is performed in terms of visual quality, quantitative analysis and computational time. However, qualitative analysis is considered for video datasets. It is sufficient to judge the performance of a fusion algorithm, after having quantitative results conducted over 50 image datasets of different classes to avoid the repetition.

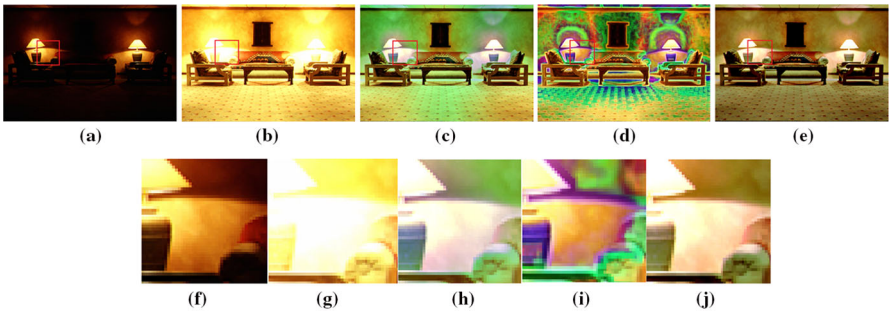

Fig. 7 Qualitative analysis of fused images of various fusion algorithms for multi-exposure chair dataset. **a** Underexposed image; **b** Over-exposed image; **c** GFF; **d** GFS; **e** MGF; **f**–**j** Zoom portion of (**a**–**e**) as highlighted in red rectangular box (Color figure online)

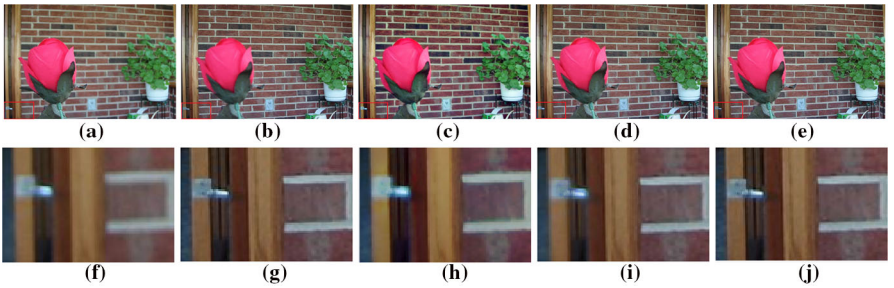

Fig. 8 Qualitative analysis of fused images of various fusion algorithms for multi-focus flower dataset. **a** Foreground focused image; **b** Background focused image; **c** GFF; **d** GFS, **e** MGF; **f**–**j** Zoom portion of (**a**–**e**) as highlighted in red rectangular box (Color figure online)

5.2.1 Image Analysis

Initially, the discussion about qualitative results is presented. Following that, quantitative analysis and computational time comparisons will be discussed.

i. Qualitative analysis

Experiments are performed on 50 image datasets. However, due to space constraint, one dataset from each class is considered for the qualitative analysis. To explain the visual quality of multi-exposure, multi-focus, medical imaging, remote sensing, and visual surveillance applications, datasets, chair, flower, brain, remote sensing, house shown in Figs. 7, 8, 9, 10 and 11 are considered.

In multi-exposure imaging, over and underexposed regions of a scene cannot be captured at the same time. As shown in Fig. 7, images captured under different lighting conditions (Fig. 7a, b) are considered. All properly exposed regions of these images have to be integrated into a single image by the fusion process. Fused images resulting from various algorithms: GFF, GFS and proposed MGF method, are displayed in Fig. 7c–e. Zoom portions of Fig. 7a–e (highlighted in red color rectangular box) are shown in Fig. 7f–j for better visual understanding. The GFF fused image (Fig. 7c) looks green (biased toward green channel) which is different from the source images.

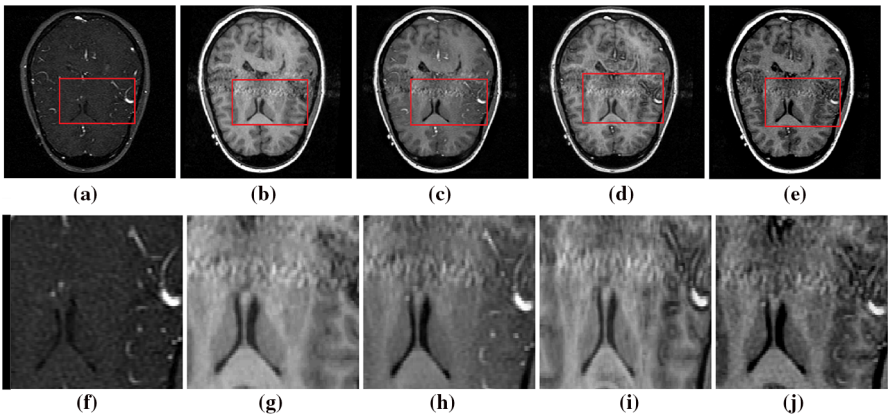

Fig. 9 Qualitative analysis of fused images of various fusion algorithms for medical brain dataset. **a** CT image; **b** MRI image; **c** GFF; **d** GFS; **e** MGF; **f–j** Zoom portion of (**a–e**) as highlighted in red rectangular box (Color figure online)

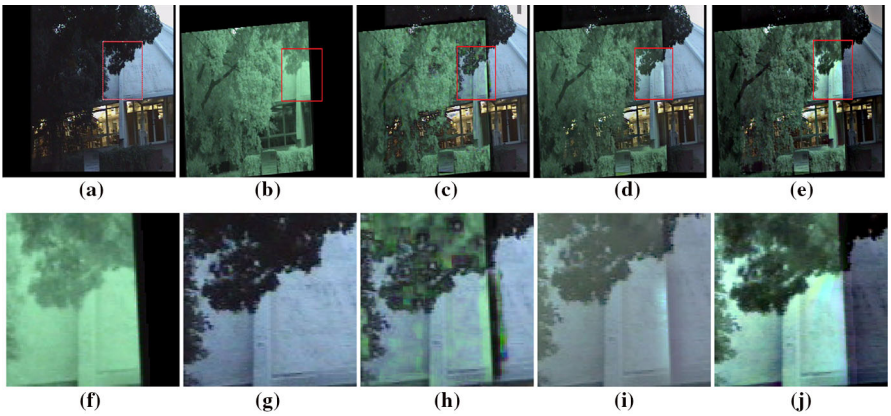

Fig. 10 Qualitative analysis of fused images of various fusion algorithms for VI-IR house dataset. **a** IR image; **b** Visible image; **c** GFF; **d** GFS; **e** MGF; **f–j** Zoom portion of (**a–e**) as highlighted in red rectangular box (Color figure online)

As shown in zoom portion (Fig. 7h), the fused image is also not clear. We can observe that wall and floor information is also looking in green color. Fused image (Fig. 7d, i) of GFS is visually distorted. Severe visual artifacts can be seen throughout the image. However, as shown in Fig. 7e, j, MGF is able to generate all properly exposed regions in a single image without introducing visual artifacts and color distortions.

Now, we look at multi-focus fusion challenges. In digital photography, only one object of a scene can be focused at a time. If we focus on one region, we may lose information about the other regions. However, we need all the regions of the scene to be focused. Therefore, individually focused images need to be combined into a single image. Foreground, background focused images and corresponding zoomed portions of the flower dataset are shown in Fig. 8a, b, f, g, respectively. Fused images

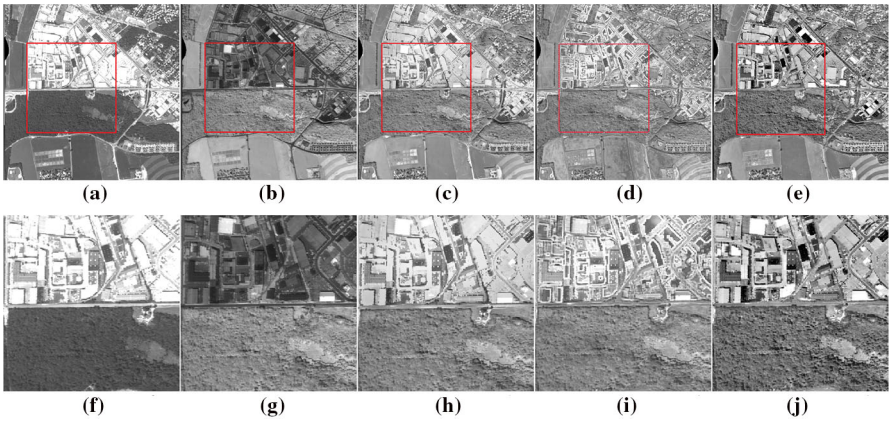

Fig. 11 Qualitative analysis of fused images of various fusion algorithms for remote sensing dataset. **a** Source image 1; **b** Source image 2; **c** GFF; **d** GFS; **e** MGF; **f**–**j** Zoom portion of (**a**–**e**) as highlighted in red rectangular box (Color figure online)

and respective zoomed portions of GFF, GFS and MGF algorithms are displayed in Fig. 8c–e, h–j, respectively. It can be seen from these visual results that MGF yields visually more focused regions with better contrast and edge information compared to GFF and GFS. Now, we move to medical imaging [6] where complementary information of a particular organ is captured using different modalities. As shown in Fig. 9a, b, computed tomography (CT) is used to capture hard tissue information such as the bone structure of a brain, whereas magnetic resonance imaging (MRI) is used to capture soft tissue information such as flesh. However, a radiologist needs both MRI and CT information together in an image for better diagnosis and treatment. Using image fusion technology, information of different images can be integrated. Figure 9c–e, h–j is fused images of GFF, GFS, and MGF respectively. As shown in figure, GFF is unable to combine both soft and hard tissue information. As can be seen, GFS is introducing some artifacts into the fused image. In contrast, MGF method is able to integrate both MRI and CT information in a single image with more details and fewer artifacts. Further, we can say that the proposed method is providing more information about both the soft tissue and hard tissues in the fused image compared to that of the remaining methods.

In remote sensing, multi-spectral imaging is used to capture different aerial information (Fig. 10a, b, f, g). Fusion process should combine all complementary information of multi-spectral images into a single image. Figure 10c–e, h–j is fused images and corresponding zoomed images of GFF, GFS, and MGF, respectively. It can be observed that the GFF fused image is not visually good and the result of GFS algorithm is visually distorted, whereas the proposed method's result contains all the complementary details of both multi-spectral images in the fused image.

In visual surveillance, target scene needs to be monitored all the time. Visual cameras can capture images only under proper lighting conditions. In low visibility or during the night, IR or near IR are used to capture images based on temperature variations present in the scene. However, neither VI nor IR can provide the entire scene

information. To detect and locate a particular activity or substance, complementary information in the images has to be combined. Figure 11a, b, f, g illustrates VI and IR images along with the zoomed portions of a house dataset. Figure 11c–e, h–j shows the fused images and corresponding zoomed portions of GFF, GFS and MGF methods, respectively. It can be noted that GFS fusion the result is severely distorted. Visual artifacts can be clearly observed around the edges. The visual contrast of the GFF fused image is very low. Many details are looking blur in the fused image. From the fused image and zoomed portions of the proposed method, one can observe that texture, edge and source image color information of MGF are well preserved and clearer when compared to other methods.

ii. Quantitative analysis

The main aim of the quantitative inspection is to check whether all the necessary information of source images is transferred into the fused image with fewer fusion artifacts. For a better assessment, we consider well-known and most frequently used as well as recent fusion metrics. These metrics can inspect and quantify the effectiveness of a fusion algorithm in possible ways.

Quantitative analysis of various datasets is tabulated in Tables 1, 2, 3, 4 and 5. For this, average fusion metric values are calculated over all datasets of that particular class. For example, the quantitative results of various fusion algorithms for multi-exposure imaging are calculated over 8 image datasets (Fig. 6a) as displayed in Table 1. Similarly, quantitative results of remaining multi-focus, medical, remote sensing, VI-IR datasets are presented in Tables 2, 3, 4 and 5, respectively.

Table 1 Quantitative analysis of various fusion algorithms for multi-focus datasets

Method	AG	FS	CC	SF	Q_0	VIF
GFF	9.237024	1.937163	0.974	18.25782	0.841501	0.922087
GFS	9.204458	1.942676	0.974373	16.29798	0.833828	0.91236
MGF	10.85881	1.98493	0.97828	18.71182	0.86322	0.96927

Table 2 Quantitative analysis of various fusion algorithms for Multi-exposure datasets

Method	AG	FS	CC	SF	Q_0	VIF
GFF	12.71254	1.73288	0.617496	25.64845	0.620965	0.80098
GFS	14.265	1.739333	0.423285	24.911	0.605413	0.73196
MGF	15.83335	1.82717	0.80826	27.06536	0.67147	0.883

Table 3 Quantitative analysis of various fusion algorithms for Medical datasets

Method	AG	FS	CC	SF	Q_0	VIF
GFF	11.51847	1.879605	0.85209	21.64628	0.49133	0.546469
GFS	11.31004	1.877259	0.830096	23.32347	0.466835	0.523823
MGF	12.53311	1.9277	0.859475	24.45315	0.513425	0.660813

Table 4 Quantitative analysis of various fusion algorithms for Remote sensing datasets

Method	AG	FS	CC	SF	Q_0	VIF
GFF	15.97062	1.82572	0.70915	23.42539	0.632445	0.50247
GFS	17.47344	1.88976	0.68393	24.47569	0.606423	0.51303
MGF	19.00531	1.94914	0.76935	27.98636	0.66806	0.84708

Table 5 Quantitative analysis of various fusion algorithms for VI-IR datasets

Method	AG	FS	CC	SF	Q_0	VIF
GFF	7.222273	1.75501	0.55964	10.36962	0.60033	0.37052
GFS	8.147	1.83714	0.52841	9.78957	0.47797	0.42032
MGF	8.73423	1.87724	0.64218	11.92994	0.643345	0.75721

As discussed in Sect. 5, an algorithm with the highest fusion metric values yields better performance. As shown in Table 1, for the multi-focus dataset, AG = 10.85881, FS = 1.98493, CC = 0.97828, SF = 18.71182, Q_0 = 0.86322, and VIF = 0.96927 metric values of the proposed method are higher than the GFF and GFS methods. In the same way, AG = 15.83335, FS = 1.82717, CC = 0.80826, SF = 27.06536, Q_0 = 0.67147, and VIF = 0.883 metric values of the MGF are higher than the remaining methods for the multi-exposure dataset. Similarly, quantitative analysis for remaining medical (Table 3), remote sensing (Table 4) and VI-IR (Table 5) datasets can also be done. From the quantitative analysis, it is obvious that the proposed MGF algorithm performance is superior.

iii. Run time

The run time analysis of various fusion algorithms is shown in Fig. 12. They are calculated by taking the average over datasets of an identical class. Run time of GFF, GFS and MGF is 0.8836, 5.795 and 0.4822 s, respectively, for multi-exposure class. We can see that the run time of the GFS is huge compared to GFF and MGF. This is due to that the fusion rule employed in GFS is based on statistical properties of the neighborhood. However, it can be observed that the proposed MGF algorithm is approximately two times faster than the GFF. Similarly, for other classes as well, MGF algorithm is yielding results quickly when compared to GFF and GFS.

5.2.2 Video Analysis

The visual analysis is done on two co-registered video datasets as shown in Figs. 6, 13 and 14. The first one is a static video in which both VI & IR cameras are fixed and the target is moving with respect to the time. The dimension of this video dataset is $360 \times 270 \times 32$. Here, 360, 270 and 32 represent the width of the frame, the height of the frame and number of frames, respectively. Due to the space constraint, only 6 random frames (1, 7, 14, 21, 27, 32) of the video are considered for the visual quality analysis. Figure 13a_1–a_6, b_1–b_6 shows VI and IR video frames, respectively.

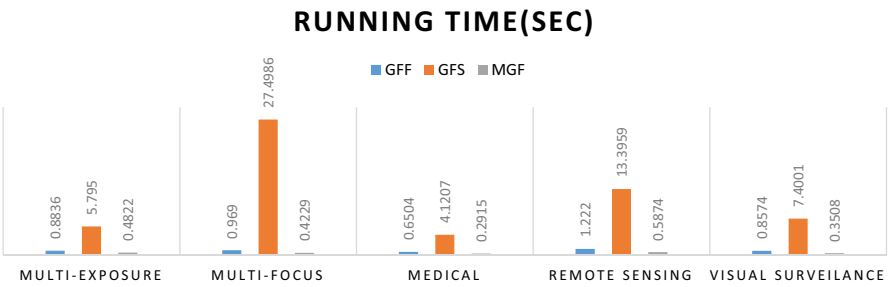

Fig. 12 Run time (s) of GFF, GFS and MGF fusion algorithms for multi-exposure, multi-focus, medical, remote sensing and visual surveillance datasets

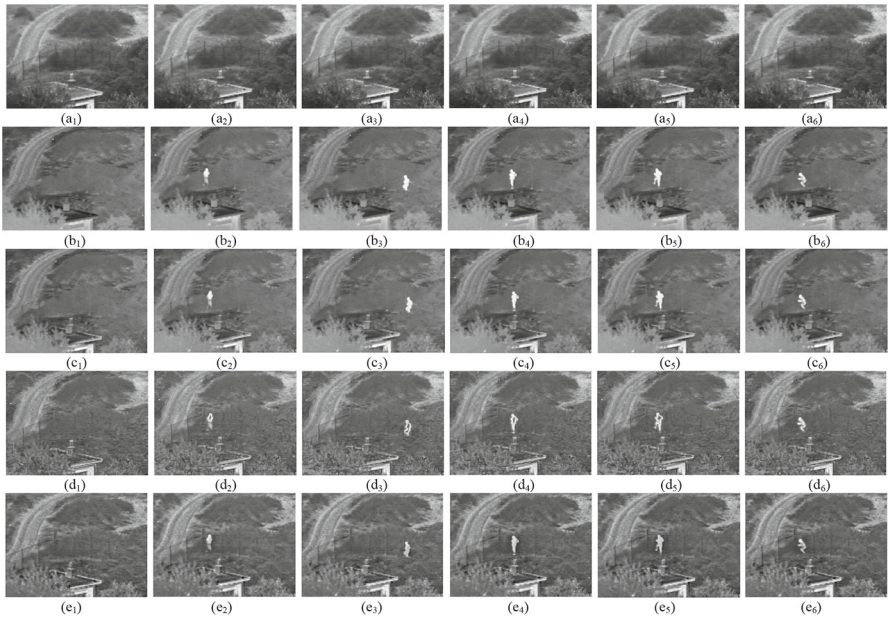

Fig. 13 Qualitative analysis of various fusion algorithms for static video dataset. Six random image frames (1, 7, 14, 21, 27, 32) of the video are considered for visual quality analysis. $\mathbf{a_1}$–$\mathbf{a_6}$ VI video frames; $\mathbf{b_1}$–$\mathbf{b_6}$ IR Video frames; $\mathbf{c_1}$–$\mathbf{c_6}$ Fused results of GFF; $\mathbf{d_1}$–$\mathbf{d_6}$ Fused results of GFS; $\mathbf{e_1}$–$\mathbf{e_6}$ Fused results of MGF

Similarly, Fig. 13c_1–c_6, d_1–d_6, e_1–e_6 illustrates the fusion results of GFF, GFS, and MGF methods.

VI video frames convey information such as fencing, building, trees, and path about the battlefield. IR or thermal video frames convey information about a person moving near the fencing. Almost in all frames, the scene is identical since cameras are static.

For better understanding of the scene and localization of the person in the battlefield, both visual and thermal image information has to be combined. Figure 13c_1–c_6 is fused frames of the GFF algorithm. We can observe that GFF is not able to integrate visible information well into the fused video. For example, information about trees, fencing and battlefield is missing in the resultant video. Figure 13d_1–d_6 is fused images

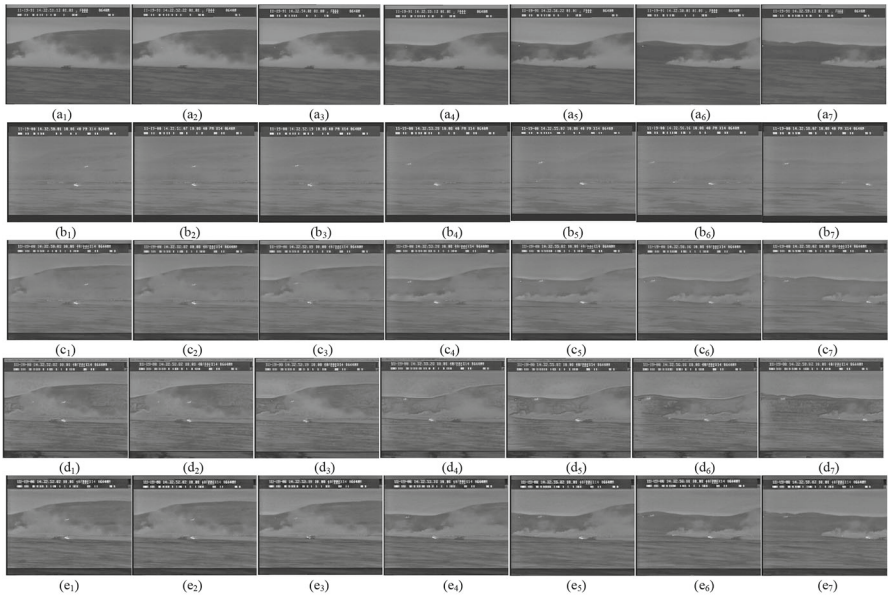

Fig. 14 Qualitative analysis of various algorithms for dynamic video dataset. Seven random image frames (1, 26, 52, 78, 104, 130 157) are considered for analysis. $\mathbf{a_1}$–$\mathbf{a_7}$ VI video frames; $\mathbf{b_1}$–$\mathbf{b_7}$ IR video frames; $\mathbf{c_1}$–$\mathbf{c_7}$ fused results of GFF; $\mathbf{d_1}$–$\mathbf{d_7}$ Fused results of GFS; $\mathbf{e_1}$–$\mathbf{e_6}$ Fused results of MGF

of the GFS. From these video frames, it is evident that resultant video is suffering from artifacts and loss of source information. Figure 13e_1–e_6 represnts fused image sequence of the MGF. From the sequence, it is clear that MGF method is able to integrate all the battlefield and person information with fewer fusion artifacts compared to GFF and GFS.

The second dataset is a dynamic video in which both target and cameras are moving. Here, the observing scene is varying from right to left with respect to time. The dimension of the second video dataset is $252 \times 240 \times 157$. Seven random frames (1, 26, 52, 78, 104, 140 157) are considered for visual quality analysis. Figure 14a_1–a_7 shows VI video frames. Figure 14b_1–b_7 displays image frames of IR video. Similarly, Fig. 14c_1–c_7, d_1–d_7, e_1–e_7 illustrates fusion results of GFF, GFS, and MGF, respectively.

As shown in Fig. 14a_1–a_7, VI video frames convey information such as mountain, fog, and water. IR or thermal video frames (Fig. 14b_1–b_7) mainly provide information about ship and aircraft which are not able to captured using VI camera. Both VI and IR image sequences are providing complementary information. For better detection and localization of vehicles in the scene, both VI and IR image information has to be combined. Figure 14c_1–c_7 is fused frames of VI and IR videos using GFF algorithm.

From results, it is clear that GFF is not able to combine all the VI and IR information into the fused video. For example, fog, vehicles and water information is not clear in the resultant video. Figure 14d_1–d_7 is fused image frames of the GFS method. Even though GFS is able to combine VI and IR information, unnecessary information or

artifacts are introduced into the resultant sequence. Figure 14e_1–e_7 is the fused image sequence of the MGF. We can observe that the proposed method is able to integrate both fog and vehicle information into the fused image sequence with fewer artifacts compared to GFF and GFS.

5.2.3 Image and Video Contrast Enhancement

In some scenarios, even if the fusion algorithm is transferring the required information of source images into the fused image, details may not be highlighted properly due to the poor contrast. To emphasize the contrast of the fused images further, we explored the state-of-the-art contrast enhancement algorithms. We considered promising and reliable contrast enhancement techniques based on histogram equalization (HE) for this purpose. They are HE [13], bi-histogram equalization (BHE) [22], recursively separated histogram equalization based on mean (RMSHE) [8], parametrically bi–bi-histogram equalization with variable enhancement degree (BBHEwVED) [28], recursively separated weighted histogram equalization (RSWHE) [23] and contrast-limited adaptive histogram equalization (CLAHE) [45]. For all the contrast enhancement methods mentioned above, default parameter settings are considered for the experimentation. These enhancement algorithms are implemented; the visual quality analysis of 10 image datasets and one video dataset is presented in Figs. 15 and 16. Source image 1 and source image 2 are displayed in Fig. 15a (first column) and b (second column). MGF fused images are shown in Fig. 15c (third column). Contrast enhancement results of MGF fused images using HE, BHE, RMSHE, BBHEwVED, RSWHE, and CLAHE algorithms are displayed in Fig. 15d–i, respectively. As shown in Fig. 15d, HE is increasing the contrast of fused images. But HE is also enhancing the noise. Furthermore, it is making dark areas more dark and bright areas even more bright. BHE is designed to enhance the image while reducing the noise and preserving the mean brightness value. However, this method is smoothing the output image without enhancing the darker regions as displayed in Fig. 15e. RMSHE method needs many recursive steps for processing and generating the output. Moreover, fused images (Fig. 15f) are not enhanced much. Contrast-enhanced images of BBHEwVED are shown in Fig. 15g. It is clear from the results that this method is introducing some artifacts. Added to this, bright regions look much brighter due to the tone saturation. RSWHE is giving visually better results (Fig. 15h), but dark regions need to be enhanced further. Conversely, CLAHE is providing visually qualitative results compared to remaining methods as shown in Fig. 15i. By examining various images and video datasets, we found that CLAHE is the best choice for image and video enhancement if we wish to enhance the fusion results of the MGF algorithm further. From the visual quality analysis (Figs. 15, 16), we infer that the MGF method with the combination of CLAHE provides superior visual quality for both images and videos.

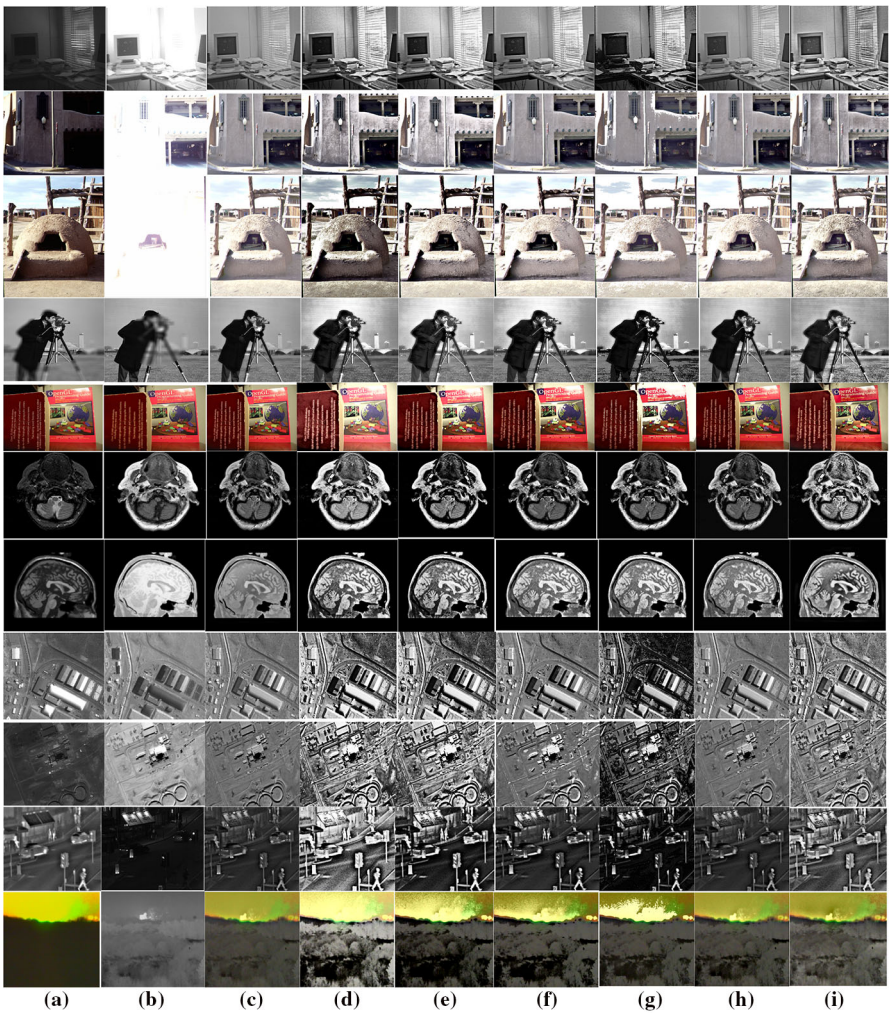

Fig. 15 Comparison of visual quality of various contrast enhancement algorithms: **a** Source image 1; **b** Source image 2; **c** MGF fused image; **d** Resultant image after applying HE on (**c**); **e** Resultant image after applying BHE on (**c**); **f** Resultant image after applying RMSHE on (c); **g** Resultant image after applying BBHEwVED on (**c**); **h** Resultant image after applying RSWHE on (**c**); **i** Resultant image after applying CLAHE on (**c**)

6 Conclusion

A fast, efficient, general-purpose image and video fusion method is proposed. A new VSD algorithm is also developed. This VSD algorithm can extract visual saliency of the whole scene from the individual source images. The main contributions of this work are as follows:

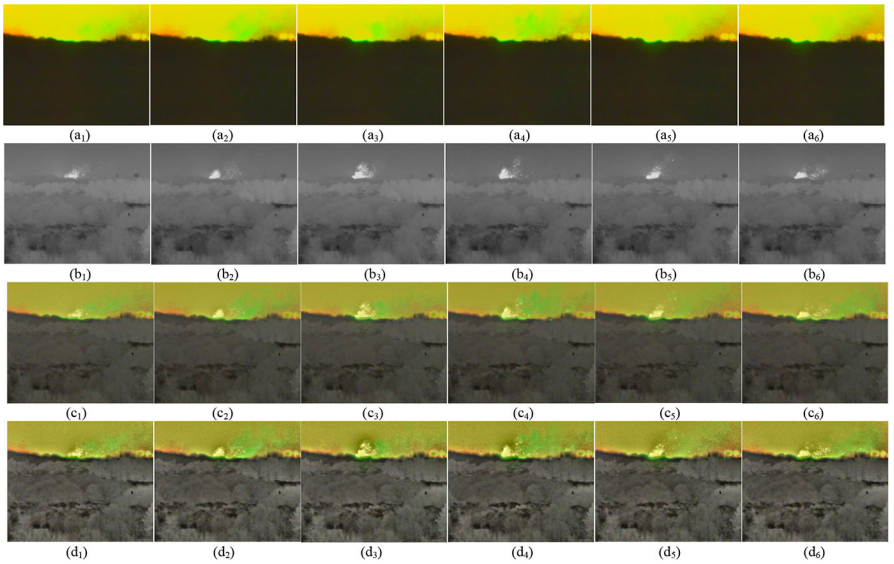

Fig. 16 Qualitative analysis of video enhancement results for fire dataset. Random image frames (1, 6, 12, 18, 24, 28) are considered for analysis. $\mathbf{a_1}$–$\mathbf{a_6}$ VI video frames; $\mathbf{b_1}$–$\mathbf{b_6}$ IR video frames; $\mathbf{c_1}$–$\mathbf{c_6}$ Fused video frames using the proposed MGF algorithm; $\mathbf{d_1}$–$\mathbf{d_6}$ Contrast-enhanced results after applying CLAHE algorithm on ($\mathbf{c_1}$–$\mathbf{c_6}$)

- MGF method is easy to implement since it requires only GF in multi-scale decomposition mode. However, existing fusion methods based on GF may need other tools such as filters or transform techniques.
- The proposed method utilizes detail layer information for the saliency extraction and weight map construction process. However, current GF-based methods depend on other techniques such as saliency detection or edge extraction techniques for weight map construction.
- Proposed MGF is verified on 50 image datasets and 3 video datasets. The results demonstrated clearly that this method is fast and efficient.

Many authors [10, 32, 39] recommended and suggested that object recognition [32], classification [10] and visual tracking [39, 40] performance will be improved if one uses fused results instead of individual VI and IR videos. We would like to subject our research in these directions.

Acknowledgments This work is sponsored by National Program on Key Basic Research Project (2014CB744903), National Natural Science Foundation of China (61673270), Shanghai Pujiang Program(16PJD028), Shanghai Industrial Strengthening Project (GYQJ-2017-5-08), Shanghai Science and Technology Committee Research Project (17DZ1204304) and Shanghai Engineering Research Center of Civil Aircraft Flight Testing. We would like to thank our postdoctoral researcher Dr. Xingchen Zhang for English proof reading of the manuscript.

References

1. D.P. Bavirisetti, R. Dhuli, Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform. IEEE Sens. J. **16**(1), 203–209 (2016)
2. D.P. Bavirisetti, R. Dhuli, Two-scale image fusion of visible and infrared images using saliency detection. Infrared Phys. Technol. **76**, 52–64 (2016)
3. D.P. Bavirisetti, R. Dhuli, Multi-filtering based edge preserving image fusion technique. Int. J. Serv. Technol. Manage **23**(4), 275–289 (2017)
4. D.P. Bavirisetti, V.K. Kollu, X. Gang, R. Dhuli, Fusion of MRI and CT images using guided image filter and image statistics. Int. J. Imaging Syst. Technol. **27**, 227–237 (2017)
5. D.P. Bavirisetti et al., A new image and video fusion method based on cross bilateral filter, in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE (2018)
6. G. Bhatnagar, Q.M.J. Wu, Z. Liu, Directive contrast based multimodal medical image fusion in NSCT domain. IEEE Trans. Multimed. **15**(5), 1014–1024 (2013)
7. R.S. Blum, Z. Liu (eds.), *Multi-sensor image fusion and its applications* (CRC Press, Boca Raton, 2005)
8. S. Chen, A. Ramli, Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. IEEE Trans. Consum. Electron. **49**(4), 1301–1309 (2003)
9. J.-C. Chiang et al., High-dynamic-range image generation and coding for multi-exposure multi-view images. Circuits Syst. Signal Process. **36**(7), 2786–2814 (2017)
10. R.R. Colditz et al., Influence of image fusion approaches on classification accuracy: a case study. Int. J. Remote Sens. **27**(15), 3311–3335 (2006)
11. N.D. Duong, S.D. Tio, A.S. Madhukumar, A cooperative spectrum sensing technique with dynamic frequency boundary detection and information-entropy-fusion for primary user detection. Circuits Syst. Signal Process. **30**(4), 823–845 (2011)
12. W. Gan et al., 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)
13. R.C. Gonzalez, R.E. Woods, in *The Book, Digital Image Processing [M]. Publishing house of electronics industry 141.7* (2002)
14. Y. Han et al., A new image fusion performance metric based on visual information fidelity. Inform. Fusion **14**(2), 127–135 (2013)
15. K. He, J. Sun, X. Tang, Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397–1409 (2013)
16. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
17. A. Jameel, A. Ghafoor, M.M. Riaz, Improved guided image fusion for magnetic resonance and computed tomography imaging. *Sci. World J.* (2014). https://doi.org/10.1155/2014/695752
18. A. Jameel, A. Ghafoor, M.M. Riaz, Wavelet and guided filter based multifocus fusion for noisy images. Optik Int. J. Light Electron Opt. **126**(23), 3920–3923 (2015)
19. A. Jameel, A. Ghafoor, M.M. Riaz, All in focus fusion using guided filter. Multidimension. Syst. Signal Process. **26**(3), 879–889 (2015)
20. U. Javed et al., Weighted fusion of MRI and PET images based on fractal dimension. Multidimension. Syst. Signal Process. **28**(2), 679–690 (2017)
21. F.T. Jhohura, T. Howlader, S.M. Rahman, Bayesian fusion of ensemble of multifocused noisy images. Circuits Syst. Signal Process. **34**(7), 2287–2308 (2015)
22. Y. Kim, Contrast enhancement using brightness preserving bi-histogram equalization. IEEE Trans. Consum. Electron. **43**(1), 1–8 (2002)
23. M. Kim, M. Chung, Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement. IEEE Trans. Consum. Electron. **54**(3), 1389–1397 (2008)
24. S. Li, X. Kang, H. Jianwen, Image fusion with guided filtering. IEEE Trans. Image Process. **22**(7), 2864–2875 (2013)
25. J. Li et al., Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans. Pattern Anal. Mach. Intell. **35**(4), 996–1010 (2013)
26. S. Liu et al., Image fusion based on complex-shearlet domain with guided filtering. Multidimension. Syst. Signal Process. **28**(1), 207–224 (2017)
27. X. Ma et al, Saliency analysis based on multi-scale wavelet decomposition, in 2013 16th *International IEEE Conference on Intelligent Transportation Systems-(ITSC)*. IEEE (2013)

28. K. Murahira, T. Kawakami, A. Taguchi, Modified histogram equalization for image contrast enhancement, in 4th *International Symposium on Communications, Control and Signal Processing* (ISCCSP). IEEE, pp. 1–5 (2010)
29. S. Pachori, S. Raman, *Multi-scale Saliency Detection using Dictionary Learning*. arXiv preprint arXiv: 1611.06307 (2016)
30. M. Peng et al., Fault diagnosis of analog circuits using systematic tests based on data fusion. Circuits Syst. Signal Process. **32**(2), 525–539 (2013)
31. Pritika, S. Budhiraja, Multimodal medical image fusion based on guided filtered multi-scale decomposition. Int. J. Biomed. Eng. Technol. **20**(4), 285–301 (2016)
32. S. Singh et al., Infrared and visible image fusion for face recognition, in *Proceedings of SPIE*, vol. 5404 (2004)
33. H. Singh, V. Kumar, S. Bhooshan, A novel approach for detail-enhanced exposure fusion using guided filter. *Sci. World J.* (2014). https://doi.org/10.1155/2014/659217
34. P. Shah, S.N. Merchant, U.B. Desai, Multifocus and multispectral image fusion based on pixel significance using multiresolution decomposition. Signal Image Video Process. **7**(1), 95–109 (2013). https://doi.org/10.1007/s11760-011-0219-7
35. P. Shah et al., Multimodal image/video fusion rule using generalized pixel significance based on statistical properties of the neighborhood. Signal Image Video Process. **8**(4), 723–738 (2014)
36. L. Shuaiqi, Z. Jie, S. Mingzhu, Medical image fusion based on rolling guidance filter and spiking cortical model. *Comput. Math. Methods Med.* (2015). https://doi.org/10.1155/2015/156043
37. A. Toet, Iterative guided image fusion. Peer J. Comput. Sci. **2**, e80 (2016)
38. A. Toet, M.A. Hogervorst, Multiscale image fusion through guided filtering. SPIE Security + Defence. *Int. Soc. Opt. Photonics* (2016)
39. N. Xu et al., Object tracking via deep multi-view compressive model for visible and infrared sequences, in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE (2018)
40. N. Xu et al., Relative object tracking algorithm based on convolutional neural network for visible and infrared video sequences, in *Proceedings of the 4th International Conference on Virtual Reality*. ACM (2018)
41. C. Zhang, A.A. Sufi, Color enhancement in image fusion, in *IEEE Workshop on Applications of Computer Vision, 2008. WACV 2008*. IEEE (2008)
42. T. Zhang et al., A novel method of signal fusion based on dimension expansion. Circuits Syst. Signal Process. **37**(10), 4295–4318 (2018). https://doi.org/10.1007/s00034-018-0760-5
43. W. Zhou, A.C. Bovik, A universal image quality index. IEEE Signal Process. Lett. **9**(3), 81–84 (2002)
44. Z. Zhou et al., Fusion of infrared and visible images for night-vision context enhancement. Appl. Opt. **55**(23), 6480–6490 (2016)
45. K. Zuiderveld, *Contrast limited adaptive histogram equalization. Graphic gems IV* (Academic Press Professional, San Diego, 1994), pp. 474–485

Affiliations

Durga Prasad Bavirisetti[1] · Gang Xiao[1] · Junhao Zhao[1] · Ravindra Dhuli[2] · Gang Liu[3]

Durga Prasad Bavirisetti
bdps1989@gmail.com

Junhao Zhao
jhzhao_16@sjtu.edu.cn

Ravindra Dhuli
ravindra.d@vitap.ac.in

Gang Liu
lukelg@gmail.com

1 School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China

2 School of Electronics Engineering, VIT University, Amaravati, A.P, India

3 School of Automation Engineering, Shanghai University of Electrical Power, Shanghai, China