IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 17, 2019, accepted January 29, 2019, date of publication February 4, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2897320

Poisson Reconstruction-Based Fusion of Infrared and Visible Images via Saliency Detection

JING LI[1], HONGTAO HUO[1], CHENHONG SUI[2], CHENCHEN JIANG[1], AND CHANG LI[3]

[1]Department of Information Technology and Cyber Security, People's Public Security University of China, Beijing 100038, China
[2]School of Opto-Electronic Information Science and Technology, Yantai University, Yantai 264000, China
[3]Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China

Corresponding author: Hongtao Huo (huohongtao@ppsuc.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC0803006, in part by the Ministry of Public Security Technology Research Program under Grant 2018JSYJA01, in part by the National Key Research and Development Program of China under Grant 2017YFC0822405, and in part by the National Natural Science Foundation of China under Grant 61601397.

ABSTRACT Saliency-based methods have been widely used in the fusion of infrared (IR) and visible (VIS) images, which can highlight the salient object region and preserve the detailed background information simultaneously. However, most existing methods ignore the salient information in the VIS image or they fail to highlight the boundaries of objects, which makes the final saliency map incomplete and the edges of the object blurred. To address the above-mentioned issues, we propose a novel IR and VIS images' fusion algorithm based on the Poisson reconstruction and saliency detection using the Dempster–Shafer (DS) theory. In detail, we mix the gradient using a mask map derived from the saliency map, which could avoid low contrast and halo effects in the results. Besides, both the intensity saliency of the IR image and the structural saliency of all source images are considered by DS to suppress some noise in the IR image. Thus, we could obtain smooth object contours and enhance the edge information of the salient region. Moreover, we also propose a novel probability mass function to calculate the probabilistic map in the process of applying DS to decrease the error from manually assigning the prior probability. Finally, the extensive qualitative and quantitative experiments have demonstrated the advantages and effectiveness of our method compared with other nine state-of-the-art IR and VIS image fusion methods.

INDEX TERMS Image fusion, object detection, context, image reconstruction.

I. INTRODUCTION

Image fusion is a popular technology that combines the complementary information from diverse sensors into one image in order to provide more support to a particular application, such as those in military affairs [1], [2], medical technology [3], [4] and artificial intelligence [5]. The single fused image contains more detailed information to enhance the application's performance and help us to comprehensively understand the scene [6]. The fusion of infrared and visible images has been a heavily researched topic for decades because it not only can increase visual perception but can also improve the accuracy of object detection and target recognition [7]. An infrared image could detect the target accurately since its sensitivity to the heat source and infrared sensors could work day and night. However, infrared images contain less detailed information due to their low spatial resolution. In contrast, visible images can capture some appearance information from the object and the background. Therefore, the fusion of infrared and visible images could provide clear contextual information regarding the target in the infrared image from the visible image in order to understand the scene better.

Various fusion methods of infrared and visible images have been proposed in recent years. Image fusion could be categorized into three levels: the pixel level, the feature level and the decision level [8]. The image fusion based on the pixel level, which combines the information from the source image directly, is the lowest level. However, pixel-level image fusion is still an important research field due to its efficiency and good human visual perception [9]. In addition, the fused image based on the pixel level has a close correlation with

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

the source images. In this paper, we fuse the infrared and visible images based on the pixel-level, and we assume that all source images involved in this paper have all been registered in advances.

Several research papers on image fusion deliver a complete summary of the previous works [10]–[12]. In the previous research, the earliest research on image fusion average the pixels of the source image [13], which is the simplest and most efficient method. However, this method only takes the average of the corresponding pixels from two raw images, which has poor performance for preserving the edges and leads to a low contrast fused image. Since then, the multi-scale transform (MST) scheme has been widely applied in a variety of fusion models due to its better expression of the visual attention system. The MST scheme is implemented by decomposing the source image into different scales, and then the multiscale information is fused with proper rules to reconstruct the final image via the inverse transform. Examples of these methods include the pyramid transform [14], [15], the wavelet transform [16], [17], and the non-subsampled contourlet transform [18], [19]. However, the MST-based methods ignore the different imaging characteristics of the visible and infrared images in order to use a single representation, and those methods lose some high frequency information in the process of the multiscale decomposition. As a result, the fused result cannot fully preserve all of the edges and suffers from ringing artifacts. After that step, the sparse representation (SR) is used for image fusion. Because SR-based methods usually aim to learn an overly complete dictionary, they could better represent the source image [20]–[22]. However, the SR-based methods do not consider the innovative features and salient features simultaneously, and as a result, the innovative features will be faint in the fused image when the source image has a low contrast. In general, how to preserve object edges better without reducing the image contrast is still a problem for the fusion of infrared and visible images. Thus, many scholars focus on saliency-based methods.

Saliency-based methods have been a heavily researched topic in the past several decades. According to the different processing mechanisms, saliency detection methods could be classified into three categories: biologically based methods, purely computational methods, and hybrid methods [23], [24]. The biologically based methods are motivated by simulating the human visual mechanism in order to compute the saliency map. The typical example is Itti's algorithm, which uses the center-surround operator to extract the color, intensity, and orientation feature maps, and then combines them into a saliency map with a unique scalar. Moreover, there are lots of similar methods based on biology [25]. However, this method imitates a simple human visual system, which leads to the saliency map having low resolution and poor detailed information. The purely computational methods calculate the saliency map without using biological vision principles and have the advantage of a fast processing speed [26], [27]. For instance, Bavirisetti and Dhuli [28] only use two-scale image decomposition to fuse the images through the saliency map. Liu *et al.* [21] use the joint sparse representation model to get the global and local saliency maps of the source images, and then integrate them using a weight fusion algorithm. Meng *et al.* [18] integrate the saliency map into the nonsubsampled contourlet transform for infrared and visible image fusion. In addition, for dynamic objects in image sequences, Han *et al.* [29] use Markov Random Fields to combine the high brightness values with motion features to generate the saliency maps, which are injected into the final fused image. For the static object, Sun *et al.* [30] use the Markov random field to derive the weight map of the source image, and then the weight map is used to fuse the image by solving the Poisson equation. The other category of methods is the one that combines ideas from the aforementioned two categories. Liu *et al.* [31] use either the Gaussian model or the kernel density estimation to express the region of interest (ROI), and both the color and spatial distributional features are integrated with the ROI to extract the saliency map. Zhang *et al.* [32] use the mean-shift segmentation algorithm and a quad mesh to over-segment the image, and then a color compactness measure is used to derive the saliency map directly. Besides, Fu *et al.* [33] propose saliency detection method based on normalized cut (Ncut) and adaptive multi-level region merging scheme, which can highlight the entire object in complex background. Fu *et al.* [34] firstly propose a scheme which completely learns a continuous conditional random field for saliency detection. They also provide an optimal way to integrate various unary saliency features with pairwise cues to discover the salient objects.

It should be noted that the saliency-based methods have achieved comparable performance for image fusion, and they can highlight the salient object region and preserve detailed information. However, most saliency-based methods extract the saliency information in one image but ignore the salient information among multi-images, which makes the saliency map incomplete and the edges of the object blurred [29]. In addition, some saliency-based methods over highlight the entire salient region but then fail to highlight the boundaries of objects. Besides, there are many pixel-based methods, which have achieved excellent results in image fusion. However, many traditional pixel-based methods such as multiscale transform methods [14], [17], [18] fuse the images both in the high frequency and low frequency domain. Those methods may reduce the contrast of the fused results.

With the main motivation to simultaneously better preserve the object edges and improve the image contrast of fused images, we proposed a novel method based on Poisson reconstruction and saliency detection using DS for infrared and visible image fusion.

In order to better preserve the object edges, the detection of saliency regions incorporates the structural saliency of both source images and the intensity saliency of infrared image by means of Dempster-Shafer theory. As is well-known, an infrared image is characterized by the intensities or the image's uniqueness, and the target has larger intensities

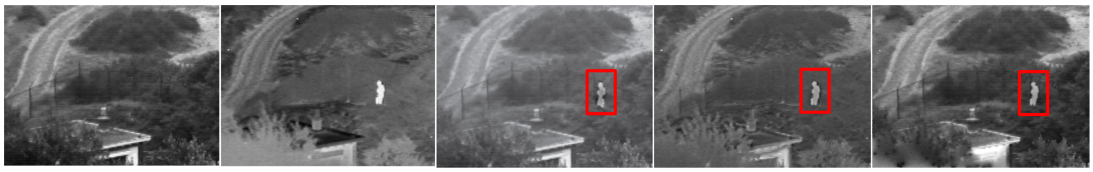


FIGURE 1. Schematic illustration of image fusion. From the left to right: the visible image, the infrared image, the fusion result of the Poisson image fusion [30], the fusion result of a popular sparse representation-based method [38], and the fusion result of our method.

than the background. Therefore, the intensity saliency of an infrared image is used to detect the saliency map in most methods, but the saliency map detected only from an infrared image always has incomplete contours or regions [29]. We use the structural saliency of an infrared image in an attempt to resolve this problem [35]. This is due to that structural saliency could indicate the salient edges and textures by using the image's gradient or compactness. In addition, the structural saliency of a visible image is used to suppress the noise in the infrared image.

Besides, in order to retain the salient foreground target information and the background information without reducing the contrast, the Poisson reconstruction is applied to this work. The major idea of this fusion method is to fuse the important structures of source images modeled as image gradient by solving a Poisson equation. Poisson reconstruction fuses the image from the high-frequency gradient domain in the original resolution without changing the low-frequency information [30] By this way, the fused images could have a proper contrast.

Our method has two major steps. In the first step, to better preserve the object edges, the structural saliency of both source images and the intensity saliency of infrared image are simultaneously incorporated to extract the saliency map using DS. To obtain an accurate probability for each feature in DS, a novel probability mass function is proposed. To the best of our knowledge, this report describes the first time that DS has been adopted to detect the salient regions. In the second step, the Poisson reconstruction is applied to improve the contrast of the fused result in the process of image fusion. Poisson reconstruction has been widely applied to image fusion [36], [37], which mixes the gradients using a weight map, which leads to the possibility that the gradient of a pixel may be compounded by two corresponding pixels, and the contrast of the result may be reduced. We mix the gradients of the fused image using a mask map derived from the saliency map rather than a weight map so that our method could use the raw gradient of the input image to avoid low contrast and halo effects in the results. To illustrate the major superiority of our method, an example is shown in Fig. 1. The left two images are the visible and infrared images, where the infrared image highlights the salient foreground target information and the visible image contains background information. The third image is the fusion result of a Poisson image fusion based on the Markov random field [30]. In the third image, the object has an incomplete edge, and we can observe some halo effects around the target object in the fusion results, which demonstrates the importance of preserving the object's edge. The fourth image is the fusion result of a popular method based on sparse representation [38]. In this image, though the saliency information and the background information are preserved simultaneously, the background has some blurring with a low contrast. The rightmost image is the fusion result of our method. We see that our result has smooth image contours and a better contrast.

The main contribution of our method includes the following three aspects. (i) To improve the contrast of the fused result and enhance the context of the target object, our method is biased in favor of the visible image from the Poisson reconstruction. We mix the gradient of the fused image using a mask map derived from the saliency map, rather than a weight map, such that our method can use the raw gradient of the input image to reconstruct the final result, which could avoid low contrast and halo effects in the result. (ii) We extract the saliency map using the intensity saliency of the infrared image (like several state-of-the-art methods) and the structural saliency of the infrared and visible images, which could suppress some noise in the infrared image, enhance the edge information of the salient region and improve the performance in terms of human visual perception. (iii) To obtain a convincing probability for each feature and decrease the error from manually assigning the prior probability, we propose a novel probability mass function to calculate the probabilistic map in the process of applying DS. It turns out that this function's results are more objective, and this method is better than the approaches in [39] and [40].

The rest of the paper is organized as follows. Section II presents the framework of our method and the formulation of the proposed method. We test our fusion method using a publicly available dataset and nine prior methods are compared to our approach in section III. The study's conclusions are presented in section IV.

II. METHOD

In this section, we first illustrate the proposed framework's architecture and data flows, and then introduce the definition of the probability mass function. After that step, we present our method of saliency region detection based on the DS. To this end, we introduce the Poisson image reconstruction.

A. FRAMEWORK OVERVIEW

Fig. 2. shows the architecture of the proposed approach. We divide our method into two major steps: saliency map detection and Poisson reconstruction. First, the gradients of

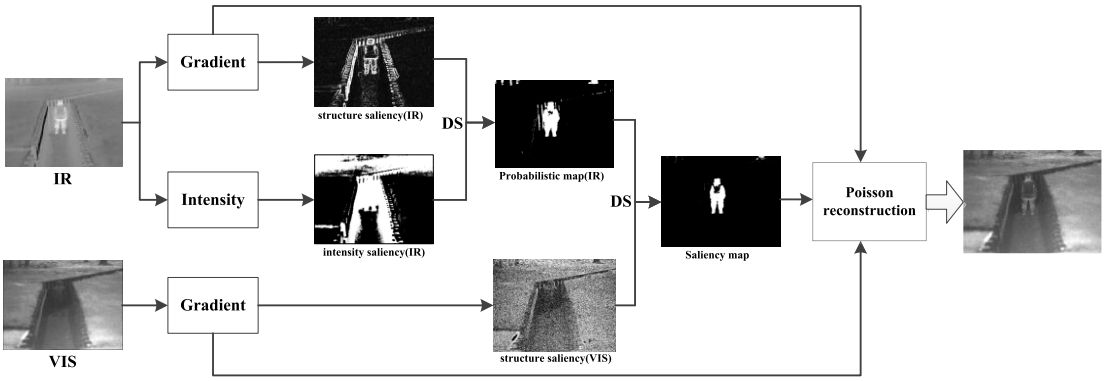


FIGURE 2. Framework of the proposed method.

the input images are calculated using the forward differences. Then, the structural saliency of the input image and the intensity saliency of the infrared image are calculated using a novel probability mass function, which will be discussed in the next section. The structural saliency of the input images and the intensity saliency of the infrared images are simultaneously taken as evidences to extract the saliency map using DS. After that, the saliency map is taken as the mask map to fuse the gradients of the source images. In the end, the final fused image is acquired from the fused gradient by solving the Poisson equation.

- The calculation of the probabilistic map (IR). The gradient and intensity of the target in the infrared image always have significant differences from the background. Based on this, the intensity saliency (IR) is generated by the intensity of the infrared image and the structural saliency (IR) is derived from the gradient of the infrared image. In the end, the intensity saliency (IR) and structural saliency (IR) are taken as two pieces of evidence to calculate the probabilistic map (IR) using the DS.
- The calculation of the structure saliency (VIS). We notice that the target's gradients in the visible image and infrared image usually change simultaneously. Therefore, the gradient of the visible image is also used to calculate the structural saliency (VIS), which can reduce the impact of the noise in the infrared image on the results.
- The calculation of the saliency map. To suppress the noise in the infrared image and obtain a pure saliency map, the structural saliency (VIS) is chosen as the third evidence to calculate the saliency map using the DS.
- Poisson image reconstruct. After calculating the saliency map, we mix the gradient of the infrared image with that of the visible image, and the final fused image is reconstructed by solving Poisson equations.

B. DEMPSTER-SHAFER THEORY

Dempster-Shafer theory is a popular theory to deal with the uncertainty problem, which could emphasize the event's objectivity and human's subjectivity. The distinctive character of DS is that the description of uncertainty is based on an ''interval estimation'', rather than a ''point estimation'', which could integrate the knowledge and data from different experts or multiple sensors [41]. In addition, DS also has great flexibility in distinguishing between unknown and uncertain aspects. We suppose Ω is the hypotheses space and the power set of Ω is denoted by 2^{Ω}, which contains all the classes and their possible unions. The imprecision and uncertainty in DS are expressed by the definitions of the plausibility (Pls) and belief (Bel), which are both obtained from a probability mass (m), and for each class A of 2^{Ω}, where $0 \leq \mathrm{m}(A) \leq 1$,

$$\begin{cases} m(A) = 0 \\ \sum_{A \in 2^{\Omega}} m(A) = 1. \end{cases} \tag{1}$$

For the imprecise knowledge, it could be represented by a nonzero probability mass to a compound hypothesis. The belief and plausibility functions can be defined for all $A \in 2^{\Omega}$ as follows:

$$Bel(A) = \sum_{B \subset A} m(B), \tag{2}$$

$$Pls(A) = \sum_{B \cap A = \emptyset} m(B). \tag{3}$$

Meanwhile, DS could combine the different evidentiary information from several sensors. For every sensor, the posterior probability (M) of all propositions (A_i) is calculated using the different periods' data that are acquired from the same sensor as follows:

$$M_S(A_i) = \frac{\sum_{\cap A_j = A_i} \prod_{1 \leq j \leq n} M_{sj}(A_i)}{\sum_{\cap A_j \neq \emptyset} \prod_{1 \leq j \leq n} M_{sj}(A_i)}, \tag{4}$$

where S is the sensor, and $1 \leq j \leq \mathrm{n}$. Then, all the sensors could be seen as a system to calculate the final posterior probability, as above.

C. DEFINITION OF PROBABILITY MASS FUNCTION

The definition of the probability mass is the critical part of any application of the DS since it provides a probability of a pixel being classified to salient regions. Probability mass function is a function that gives the probability that a discrete random

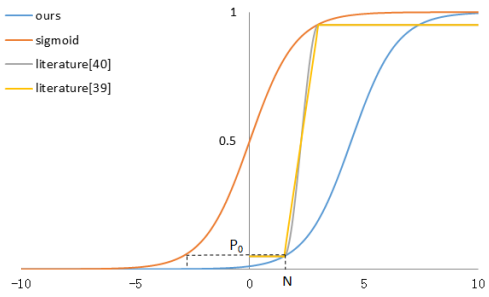


FIGURE 3. Three probability mass functions.

variable is exactly equal to some value [42]. A digitized intensity pattern is discrete and the intensity values are recorded as 'pixels' [43], and the values of pixel in digital image are taken in a discrete set of possible values [44]. So the image pixels we used could be seen as discrete random variables. In our research, the structures of some salient regional distributions in the spatial domain are different from that of the background region. Therefore, we need an appropriate probability mass function to distinguish their differences. Besides, the value of the probability mass should be distributed from 0 to 1 with a smooth distribution. There are many approaches to define the probability mass function in the state-of-the-art studies. In [40], the probability mass function is defined by a cubic parabola with a horizontal, which is shown in Fig. 3. In [39], they also proposed a probability mass function, which is shown in Fig. 3. In this paper, the proposed probability mass function actually is a function based on sigmoid function, which is shown in Fig. 3. Sigmoid function has been widely applied in machine learning, because it solves the binary classification problem very well. The value of sigmoid function at x represents the probability of x being classified to positive class ($y = 1$) [45]. However, in our experiment we find that if we use the sigmoid function directly, even if the pixel has a smaller gradient value (or intensity value), there is still a high probability of being classified to saliency, and this may lead to inaccurate saliency region. To overcome the above mentioned problem, we improve the sigmoid function by defining a threshold N. We assume that these pixels with gradient value (or intensity value) equals to N will have a smaller probability (P_0) of being classified to saliency map. The proposed probability mass function also guarantees that the larger the gradient value (or intensity value) of pixels, the greater probability of the pixels being classified to salient regions, which is defined as follows:

$$P(X_{i,j}) = \left[1 + e^{N - X_{i,j} - \ln(\frac{P_O}{1-P_O})}\right]^{-1}, \tag{5}$$

where $P(X_{i,j})$ is the probability of a pixel $X_{i,j}$ being classified to salient regions, and N is a threshold, and pixel $X_{i,j}$ denotes the gradient value or the intensity value, which is the random variable in the function to calculate probability. When the pixel value in the image is smaller than the threshold N, we assume that the assignment of a pixel to the saliency region will be very unlikely, which is modeled by a small probability P_0. Through a lot of experiments, P_0 is defined as 0.05, which means there is a small possibility that it is a salient region. In addition, the probability mass we proposed is real-valued, monotonic and differentiable. It has a non-negative first derivative that is bell shaped and maps variables between 0 and 1, which is shown in Fig. 3.

Furthermore, when we compare our probability mass with the literature [40] and literature [39], we assume that N equals literature [40]'s falling threshold of 1.5 (the rising threshold is 3.5). The results are illustrated in Fig. 3. The probability value of the method in the literature [39], [40] is abrupt near the threshold, and the probability change rate is so high that the overall distribution is less smooth. In addition, when the pixel $X_{i,j}$ is smaller than the falling threshold or higher than the rising threshold, the probability will stay the same with a small probability or a high probability, which means that the homogeneity of the probability will lead to an inappropriate local probability distribution. In addition, the threshold N needs to be defined only once in our method, which could guarantee that the probabilistic map is more objective. The fused results based on the three probability masses are compared in section III.

D. SALIENCY REGION DETECTION

Saliency region detection could be interpreted as a pixel labeling problem, where each pixel in the image is labeled as either salient or non-salient. In addition, the pixel labeling could be seen as a target recognition problem, which can be treated using DS. This is because the theory allows one to combine the evidences from different sources and arrive at a degree of belief that takes into account all the available evidence [41], [46]. Furthermore, the prior probability needed in DS is more intuitive and easier to obtain than the probability theory. The Dempster synthesis rule can continuously integrate new different knowledge and data to improve the recognition and extraction precision [46].

We consider that the salient region not only has a greater intensity and gradient than the neighboring pixels in the infrared image, but it also has a greater gradient in the visible image. Therefore, the structural saliency of both source images and the intensity saliency of infrared image are simultaneously taken as evidences to calculate the saliency map using DS. Meanwhile, the effects of noise within the infrared image on the fusion result could be suppressed by using the gradient of the visible image. The image gradient is a directional change in the intensity or color of an image, which can be used to illustrate the differences between image pixels. In this paper, the gradient of the source image is computed by using the forward difference, so the pixels in gradient image are still the discrete random variables. Then, the structural saliency probabilistic map (M_G) is acquired through the probability mass function as follows:

$$M_G(G_{i,j}) = \left[1 + e^{N_G - G_{i,j} - \ln(\frac{P_O}{1-P_O})}\right]^{-1}, \tag{6}$$

where $G_{i,j}$ is the gradient value at pixel(i,j) and N_G represents the threshold of gradient which has a smaller probability (P_0) of being classified to salient regions. The salient region of the infrared image that is extracted relies on the human visual system, which is not very sensitive to the absolute luminance reaching the retina, but rather it responds to local intensity ratio changes [47]. Based on this, we proposed a novel method to calculate the saliency map. The salient map F for a source image I could be obtained as follows:

$$F_{i,j} = I_{i,j} - I_m, \tag{7}$$

where $F_{i,j}$ denotes the intensity value at pixel(i,j), $I_{i,j}$ is the original pixel value of the source image and I_m is the arithmetic mean pixel value of the image, so the pixels in salient map F are still the discrete random variables. Then, the intensity saliency probabilistic map (M_F) is obtained as follows:

$$M_F(F_{i,j}) = \left[1 + e^{N_F - F_{i,j} - \ln(\frac{P_o}{1-P_o})}\right]^{-1}, \tag{8}$$

where N_F is the threshold of intensity which has a smaller probability (P_0) of being classified to salient regions. Therefore, the structural saliency probabilistic map of the infrared image (M_{G_IR}) and the intensity saliency probabilistic map of the infrared image (M_{F_IR}) are used to calculate the probabilistic map of the infrared image (M_{IR}) by DS, as shown in Eq. (9). By combining the structural saliency and intensity saliency of the infrared image, the edge of the saliency map could be preserved well.

$$M_{IR} = \frac{M_{G\_IR} + M_{F\_IR}}{1 - M_{G\_IR} \times (1 - M_{F\_IR}) - M_{F\_IR} \times (1 - M_{G\_IR})}. \tag{9}$$

Then, we combine the M_{IR} with the structural saliency probabilistic map of the visible image (M_{G_VIS}) to calculate the final probabilistic map M by DS as Eq. (10), which could suppress the noise to derive a pure saliency map.

$$M = \frac{M_{G\_VIS} + M_{IR}}{1 - M_{G\_VIS} \times (1 - M_{IR}) - M_{IR} \times (1 - M_{G\_VIS})}. \tag{10}$$

E. IMAGE FUSION IN GRADIENT DOMAIN

In this paper, we combine the saliency detection with Poisson reconstruction to fuse infrared and visible images. Poisson image editing has been used in image processing for decades [30], [36], [37]. This is because it fuses the images in the high-frequency gradient domain without involving the low-frequency information, which could improve the contrast of the fused image. Compared with other works, we use the mask map rather than the weight map to mix the gradient. In this way, we could use the raw gradient of the source image to reconstruct the final result, which could avoid low contrast and halo effects in the results. First, the gradients of the input images are calculated using the forward difference. Then, a threshold λ is initialized to derive the binary saliency map (S) from M by using the following rule:

$$S(x,y) = \begin{cases} 0 & if\ M(x,y) < \lambda \\ 1 & otherwise. \end{cases} \tag{11}$$

The image fusion is conducted using the Poisson image reconstruction, which mixes the gradient of the source image with that of the destination image. Therefore, the mixed gradient is obtained by combining the gradients of the infrared image and the visible image, which is expressed as follows:

$$\nabla G(x,y) = S(x,y) \times \nabla G_{IR}(x,y) + (1 - S(x,y)) \times \nabla G_{VIS}(x,y), \tag{12}$$

where $\nabla G(x,y)$, $\nabla G_{IR}(x,y)$ and $\nabla G_{VIS}(x,y)$ are the gradients of the fused image, the infrared image, and the visible image, respectively, at pixel (x, y). Then, the mixed gradient is applied to construct the final fused image by Poisson image editing [48]. This method leads to a seamless result that preserves the gradient information of both images.

III. EXPERIMENTAL RESULTS

In this section, we test the performance of proposed method on a publicly available dataset, and compare it with nine previous methods. Fusion strategies are generally divided into several categories according to their adopted theories, i.e., multi-scale transform-based methods, sparse representation-based methods, deep learning based methods, subspace-based methods and other infrared and visible image fusion methods [10], [49], [50]. Based on this, we choose nine previous methods from all above categories to make comparisons with our method. The convolutional neural networks (CNN) [51] is representative deep learning based methods. The laplacian pyramid (LP) [14] and nonsubsampled contourlet transform (NSCT) [52] are representative multi-scale transform-based methods. The adaptive sparse representation (ASR) [38], laplacian pyramid with sparse representation (LPSR) [53] and sparse representation (SR) [54] are representative sparse representation-based methods. The fourth order partial differential equations (FPDE) [55] is representative subspace-based methods. The gradient transfer and total variation minimization (GTF) [9] and infrared feature extraction and visual information preservation (IFEVIP) [56] are representative other infrared and visible image fusion methods. These methods are conducted based on the publicly available codes with the parameters being initialized according to the original paper, and we try our best to tune some details. All the experiments are carried on a Macpro (Intel(R) Xeon(R) with E5-1650 v2 3.5 GHz CPU and 16GB memory) with the MATLAB software.

A. DATASETS AND SETTING

The images we used to test our method are from the TNO Image Fusion Dataset. This dataset contains multispectral (intensified visual, near-infrared, and long wave infrared or thermal) nighttime imagery of different military relevant scenarios that are registered with different

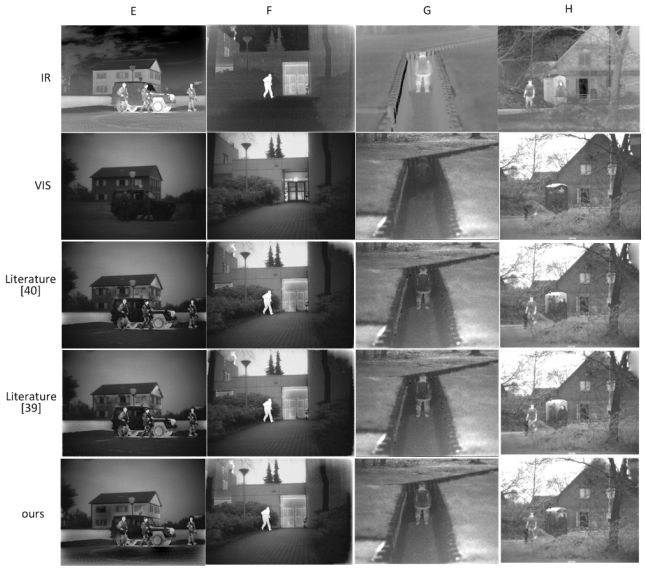


FIGURE 4. Qualitative comparisons of fused images with two probability mass functions.

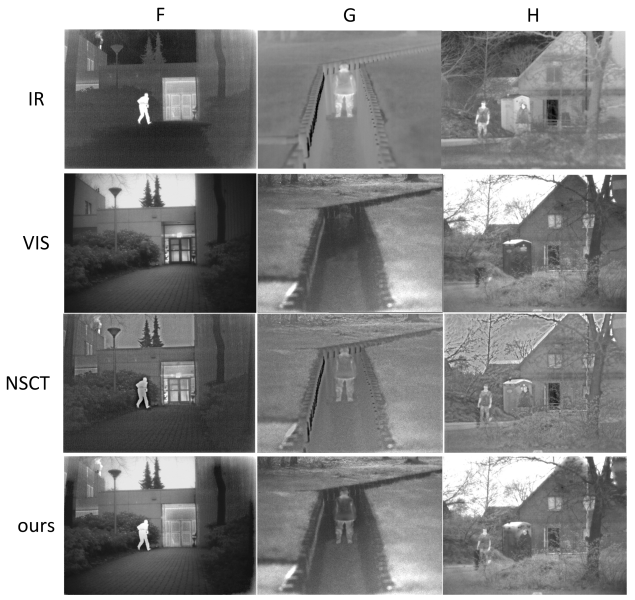


FIGURE 5. Qualitative comparisons of the fused images using our method and NSCT.

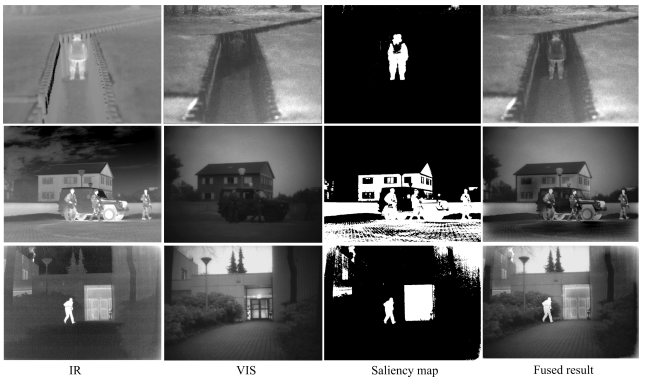


FIGURE 6. Intermediate results of saliency maps.

multiband camera systems (including Athena, DHV, FEL, and TRICLOBS), and is available at https://figshare.com/articles/TNO_Image_Fusion_Dataset. Eight typical image pairs are chosen to assess the performance of the proposed method, and the image pairs we used in this paper have been aligned in advances.

We first show some intermediate results of saliency maps in the Fig. 6. The image quality evaluation could evaluate the performances of the infrared and visible image fusion methods. In addition, the quality evaluation methods could be classified as subjective and objective evaluation methods [10], [57], [58]. The subjective evaluation methods play an important role in fusion quality evaluation because they focus on the image's distortion, details, and completeness. Nevertheless, it is necessary to introduce objective evaluation methods since they can quantitatively assess the fused image quality without being biased by observers or interpreters. Therefore, we use multiple metrics to make a comprehensive evaluation of the different fusion methods. In this paper, we choose four metrics. Entropy (EN) [59], [60] measures the amount of information contained in a fused image based on information theory. The standard deviation (SD) [61] is based on the statistical concept that reflects the distribution and contrast of the fused image. Mutual information (MI) [62], [63] measures the information that is transferred from the source image to the fused image, and $Q^{AB/F}$ [62] measures the amount of edge information that is transferred from source images to the fused image. Larger values of all these metrics represent the better performance of the fusion method.

Parameter Initialization: In this paper, there are three parameters that should be initialized, i.e., N_G, N_F, and λ in section II. N_G controls the probability of a pixel being classified to the salient region in the gradient field. According to the experiments, $N_G = 1$ is applicable, and when N_G is larger than 1, the saliency map will be incomplete. Furthermore, when $N_G = 0$, there are more no saliency region are contained in the fusion result than the situation of $N_G = 1$. Meanwhile, N_F should be fixed as 0 after our experiments because $N_F \geq 0$ means that it is possible for the pixel that has a larger intensity than the mean intensity of the image to be defined by the saliency map. In the end, to preserve more target information we adopt $\lambda = 0.5$ after numerous experiments during our study.

B. QUALITATIVE COMPARISONS

To get an intuitive impression of the proposed method's performance, nine excellent methods are selected for the qualitative comparison. These methods are mentioned at the beginning of section III. We first perform experiments on four typical image pairs in order to compare the result of our probability mass function with that of literature [39], [40], which is mentioned in section II, as shown in Fig. 4. We could find that both the results that use literature [39], [40]'s probability mass

TABLE 1. Quantitative comparisons of images in FIGURE 4.

		E	F	G	H
EN	ours	**5.9766**	**5.4412**	**5.8528**	5.2215
	literature[40]	5.8871	5.3935	5.8339	5.3385
	literature[39]	5.9107	5.4033	5.8326	**5.3410**
SD	ours	85.8604	**95.5277**	69.9803	**71.4809**
	literature[40]	84.1689	95.196	**70.314**	71.2374
	literature[39]	**88.5819**	94.8477	70.2989	70.7301
MI	ours	0.2279	**0.4217**	**0.394**	0.3028
	literature[40]	0.218	0.4216	0.3869	**0.3041**
	literature[39]	**0.2545**	0.4211	0.3905	0.3035
$Q^{AB/F}$	ours	0.5877	**0.511**	**0.8333**	0.5522
	literature[40]	0.5764	0.504	0.7935	**0.6081**
	literature[39]	**0.5975**	0.4978	0.7741	0.5869

functions can obtain similar fused result, and they cannot well preserve the details in the source image. Notably, the salient region in the infrared image is not extracted completely, such as the ground information in the first row, the man's detailed information in the third row, and the background information located in the top-right of the image in the fourth row. This illustrates that the probability mass function we proposed is appropriate to be used to calculate the probability for each pixel of the image. Furthermore, the advantage will be more obvious if the salient region has similar intensities with the background.

Then, we perform experiments on three typical image pairs to compare our method with a popular fusion method, known as the NSCT [52]. The results are illustrated in Fig. 5. The top two rows are the infrared images and the visible images. The third row is the fusion results of the NSCT. On the one hand, it can be seen from the results that the NSCT cannot well preserve the background information from the visible image. For example, the sky in the first and third columns is not preserved in the final fused images and the same occurs with the trench in the second column. On the other hand, the NSCT fails to highlight the target information, and the men in all three image pairs have a lower contrast than our results. The fourth row is the fusion results of ours, which well preserves both the salient foreground object information and the background information. In addition, the fused images of our method look similar to high-resolution infrared images with clear highlighted targets, which will be beneficial for understanding the image.

The experimental results of the qualitative comparison between our method and the nine state-of-the-art methods are presented in Fig. 7. Eight image pairs are assessed in this step. By learning from the results, our method preserves the target information and the background information simultaneously. In addition, it also preserves more edge information of the salient targets in the fused image, such as images A, B, E, F, and H. All the objects in those images have sharp boundaries and high contrast. Meanwhile, our algorithm inherits plenty of textural information from the visible image, which leads to a great image with high contrast and clear details. In addition, the fusion performances of CNN varied for different image pairs, and the fused results on image B, C, D and E could obtain comparable fusion performances with our method. Whereas the other image pairs using CNN could not well preserve detail information compared with our fusion method. The quantitative evaluations of the fused images of CNN are shown in Table 2-5, and it can be clear seen that the metric values of CNN are lower than our methods in most cases.

C. QUANTITATIVE COMPARISONS

In this section, we first present the quantitative comparisons of our probability mass function and literature [39], [40]'s method, which are shown in Table 1. We can see that the metric values of the images that are fused using our probability mass function have a better performance than that of the other two methods.

Then, the quantitative comparisons of the nine methods that mentioned above are implemented using eight image pairs, which are shown in the Fig. 7. We choose four indexes to evaluate the fused images, including the EN, SD, MI and $Q^{AB/F}$. The quantitative results are presented in the Tables 2-5, where the best values are highlighted with bold.

The evaluation results of the Entropy (EN) for all the methods are shown in Table 2, and there are five results of our method that have the largest respective entropy values. A larger entropy value indicates a better performance for the fused image. This means that our results contain much more information than the other methods in most cases because the results of our method preserve the target information

FIGURE 7. Qualitative fusion results on the eight image pairs of A-H. From top to bottom: infrared images, visible images, the results of our method, IFEVIP, FPDE, ASR, GTF, LP, LPSR, NSCT, SR and CNN, respectively.

and the background information simultaneously and inherit plenty of textural information from the visible image. This is consistent with the qualitative comparisons. The evaluation results of the standard deviation (SD) of all algorithms are presented in Table 3. The SD reflects the distribution and contrast of the fused image. For all the eight test images, there are four results of our method that have the largest value among all the respective results, which are for images A, B, E and F. While FPDE method gets higher scores in C and D image pairs compared to our method, because the fused results of FPDE method have relatively higher contrast than our method. Our method is biased in favor of the visible image from the Poisson reconstruction to fuse the images, when visible images (like C and D image pairs) have low contrast, the fused results may have relatively low contrast. But our method gets relatively high scores compared to the

TABLE 2. Quantitative comparisons of the ENs of the images in FIGURE 7.

	A	B	C	D	E	F	G	H
ASR	4.7516	3.9686	3.7271	3.2855	5.4906	5.0005	5.2469	5.3373
FPDE	5.0843	4.1269	3.8955	3.883	5.5273	5.2368	5.3036	5.4237
GTF	5.315	4.2231	3.87	3.8675	6.2458	5.3468	5.7434	5.5918
IFEVIP	5.0411	4.1353	3.9335	3.6258	5.9894	5.3065	5.377	5.4576
LP	4.9964	4.1181	3.8009	3.4286	5.6674	5.122	5.3395	5.47
LPSR	5.3017	4.1935	3.7885	**3.9454**	**6.2596**	5.3492	5.6889	5.4953
NSCT	4.9137	4.067	3.7777	3.3711	5.5484	5.088	5.2853	5.4191
SR	5.0844	4.373	3.7793	3.4035	5.4999	5.3176	5.8029	5.6349
CNN	5.3148	4.2756	3.9053	3.6182	6.1025	5.2859	5.837	**5.7808**
ours	**5.3411**	**4.3787**	**3.9708**	3.1118	5.9766	**5.4412**	**5.8528**	5.2215

TABLE 3. Quantitative comparisons of the SDs of the images in FIGURE 7.

	A	B	C	D	E	F	G	H
ASR	83.5979	83.3943	40.5609	78.5069	82.0263	87.8845	62.5876	61.3029
FPDE	80.16	72.1346	**74.4711**	**93.0196**	70.3193	95.5119	58.4968	66.5681
GTF	86.9918	82.8858	39.4814	86.4619	82.8127	90.2903	64.7225	68.2825
IFEVIP	71.4453	72.4756	32.1448	64.9773	69.3274	79.3361	52.0889	51.047
LP	84.6076	84.3965	40.779	78.7593	82.6587	88.7894	63.3551	62.5539
LPSR	86.9069	82.4954	40.5102	82.1872	85.5222	89.8546	63.3361	69.8993
NSCT	84.0702	83.9365	40.7492	78.616	82.259	88.246	62.8337	61.7316
SR	83.2453	81.6187	40.5308	66.6647	84.6287	92.3024	**70.5521**	**72.1002**
CNN	87.3139	85.1728	40.0415	67.0809	79.4389	89.043	69.3142	70.0365
ours	**87.6601**	**87.1001**	43.0783	63.9191	**85.8604**	**95.5277**	69.9803	71.4809

TABLE 4. Quantitative comparisons of the MIs of the images in FIGURE 7.

	A	B	C	D	E	F	G	H
ASR	0.2488	0.5037	0.1234	0.1307	0.2393	0.3226	0.3105	0.2779
FPDE	0.2872	0.4954	0.1977	0.1954	0.2477	0.337	0.3228	0.2861
GTF	0.3554	0.5942	0.1456	0.1868	0.2542	0.35	0.3587	0.3309
IFEVIP	0.252	0.5617	0.1416	0.1672	0.2182	0.3854	0.1843	0.1972
LP	0.2676	0.5756	0.149	0.1341	0.2222	0.2967	0.3073	0.2669
LPSR	0.354	0.6132	0.159	**0.2617**	**0.2603**	0.3347	0.3474	**0.3312**
NSCT	0.2576	0.5479	0.1414	0.13	0.2224	0.3057	0.3106	0.2724
SR	0.3118	0.6147	0.1603	0.214	0.1552	0.3712	**0.3952**	0.3162
CNN	0.2195	0.5211	**0.2402**	0.1776	0.1869	0.3328	0.3652	0.2965
ours	**0.3556**	**0.6834**	0.2285	0.163	0.2279	**0.4217**	0.394	0.3028

other methods in the rest image pairs. It demonstrates that the results of our method have a higher contrast than other approaches. It is interesting that images G and H have the largest SD values, but the visual effects are poorer than our method. Therefore, it is necessary to evaluate the image quality by using multiple indexes. Note that our method and

TABLE 5. Quantitative comparisons of the $Q^{AB/F}$ of the images in FIGURE 7.

	A	B	C	D	E	F	G	H
ASR	0.5551	0.6314	0.6983	0.7811	0.6576	0.5809	0.7672	0.6449
FPDE	0.5822	0.4479	0.5471	0.64	0.6317	0.648	0.7256	0.5933
GTF	0.6132	0.6721	0.6883	0.7684	0.6967	0.6649	0.8016	0.6614
IFEVIP	0.5859	0.6467	0.5876	0.6946	**0.7146**	0.6235	0.6658	0.5763
LP	0.6155	**0.6762**	0.6913	0.7801	0.7035	**0.6791**	0.7976	0.6624
LPSR	0.6143	0.6739	0.6894	0.77	0.6977	0.6624	0.8054	0.6624
NSCT	0.5971	0.6691	0.6785	0.7852	0.6891	0.6641	0.7839	0.6405
SR	0.5279	0.5369	**0.7119**	0.7486	0.6022	0.5544	0.8218	**0.714**
CNN	0.5076	0.4919	0.5313	0.6116	0.591	0.5565	0.751	0.5689
ours	**0.6644**	0.6281	0.6904	**0.7889**	0.5877	0.511	**0.8333**	0.5522

TABLE 6. Running time (seconds) of different image fusion methods.

Methods	IFEVIP	FPDE	ASR	GTF	LP	LPSR	NSCT	SR	CNN	ours
Run Time	0.4526	2.2091	475.9933	6.4557	0.4068	0.3904	5.5946	86.8019	84.9965	63.7893

LPSR could obtain comparable fusion performances and produce the best MI values in most cases. Because our method combines the salient regions in the infrared image with the background information in the visible image, the final fused images always have much mutual information. The largest MI indicates that there is a significant amount of information that is transferred from the source image to the fused image, and our method has the largest results on images A, B and F. For $Q^{AB/F}$, our method also has an excellent performance because we first mix the gradient of the salient region and the background region, and then the image is reconstructed by solving the Poisson equations. As a result, the edge in the source image is preserved very well. Furthermore, the SR and LP methods could produce comparable results to our method. In the end, we can conclude that our method tends to not only preserve the salient region but also provide abundant background information to the fused image.

In addition, it is difficult to analyze the computational complexity of both the compared methods and our fusion method theoretically. Nevertheless, we have evaluated the computational complexity of the compared methods and our fusion method by time consumption, and the running time of all methods on the eight image pairs (A-H) are given in Table 4, where each value denotes the mean of run time of a certain method on the eight image pairs. The methods of IFEVIP, FPDE, LP and LPSR run faster than others according to Table 6. The main time consumption of GTF lies in the optimization process of the gradient transfer fusion model. NSCT decomposes images into different scales, and it uses a nonlinear mapping function to obtain the directional subband coefficients, then it reconstructs the fused image by the inverse multiscale transform, which makes it take more time than these of IFEVIP, FPDE, LP and LPSR. Besides, the time consumption of ASR and SR mainly lie in training the dictionary for image fusion. The time consumption of CNN fusion methods mainly lies in training to encode a direct mapping from source images to the weight map, thus the running time of ASR, SR and CNN is more than these of IFEVIP, FPDE, LP, LPSR, GTF, NSCT and our method. Moreover, our method is based on Poisson reconstruction and saliency detection using the Dempster-Shafer theory (DS). The time consumption of our method mainly lies in the process of Poisson reconstruction, and it needs plenty of time to reconstruct the fused image by solving the Poisson equation. In summary, our method generally takes more time than these of IFEVIP, FPDE, LP, LPSR, GTF and NSCT, and it takes less time than ASR, SR and CNN. In addition, our method can obtain the best image fusion performance in most circumstance. Therefore, our method can achieve a desirable balance between image fusion performance and time consumption.

IV. CONCLUSION

In this paper, we design an efficient approach for infrared and visible image fusion based on saliency detection and Poisson reconstruction. In addition, we proposed a novel probability mass function before using the DS. By taking advantage of the accuracy of saliency map detection and the desirable performance of Poisson reconstruction, the results show that our method could effectively retain the salient foreground target information and the background information, and the fused image has smooth image contours with better contrast. Experimental results demonstrate that both the salient region and the contextual information are well-incorporated

into the fused image. The experimental results indicate that our method performs excellently. Nevertheless, our method is biased in favor of the visible image from the Poisson reconstruction to fuse the images, when visible image has low contrast, the fused image may have a relatively low contrast. So in the future work, the algorithm will incorporate more features as evidences to acquire a more accurate saliency map to improve the contrast of results, when visible images have a low contrast. Furthermore, a new dataset with higher resolution images from https://www.ino.ca/en/ video-analytics-datasets should be applied into the further research.

REFERENCES

[1] Y. Li, C. Tao, Y. Tan, K. Shang, and J. Tian, "Unsupervised multilayer feature learning for satellite image scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 157–161, Feb. 2016.

[2] S. Yin, L. Cao, Y. Ling, and G. Jin, "One color contrast enhanced infrared and visible image fusion method," *Infr. Phys. Technol.*, vol. 53, no. 2, pp. 146–150, 2010.

[3] G. Bhatnagar, Q. M. J. Wu, and Z. Liu, "A new contrast based multimodal medical image fusion framework," *Neurocomputing*, vol. 157, pp. 143–152, Jun. 2015.

[4] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 19, pp. 4–19, Sep. 2014.

[5] R. Singh, M. Vatsa, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition," *Pattern Recognit.*, vol. 41, no. 3, pp. 880–893, 2008.

[6] L. Tang, F. Meng, Q. Wu, N. L. Sowah, K. Tan, and H. Li, "Salient object detection and segmentation via ultra-contrast," *IEEE Access*, vol. 6, pp. 14870–14883, 2018.

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

[8] G. Piella, "A general framework for multiresolution image fusion: From pixels to regions," *Inf. Fusion*, vol. 4, no. 4, pp. 259–280, 2003.

[9] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, Sep. 2016.

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

[11] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016.

[12] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques—An introduction, review and comparison," *Isprs J. Photogramm. Remote Sens.*, vol. 62, no. 4, pp. 249–263, 2007.

[13] A. Toet, "Hierarchical image fusion," *Mach. Vis. Appl.*, vol. 3, no. 1, pp. 1–11, 1990.

[14] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *Readings Comput. Vis.*, vol. 31, no. 4, pp. 671–679, 1987.

[15] A. V. Vanmali and V. M. Gadre, "Visible and NIR image fusion using weight-map-guided Laplacian–Gaussian pyramid for improving scene visibility," *Sadhana*, vol. 42, no. 7, pp. 1063–1082, 2017.

[16] P. Chai, X. Luo, and Z. Zhang, "Image fusion using quaternion wavelet transform and multiple features," *IEEE Access*, vol. 5, pp. 6724–6734, 2017.

[17] L. Xu, J. Du, and Z. Zhang, "Infrared-visible video fusion based on motion-compensated wavelet transforms," *IET Image Process.*, vol. 9, no. 4, pp. 318–328, 2014.

[18] F. Meng, M. Song, B. Guo, R. Shi, and D. Shan, "Image fusion based on object region detection and non-subsampled contourlet transform," *Comput. Elect. Eng.*, vol. 62, pp. 375–383, Aug. 2017.

[19] H.-X. Liu, T.-H. Zhu, and J.-J. Zhao, "Infrared and visible image fusion based on region of interest detection and nonsubsampled contourlet transform," *J. Shanghai Jiaotong Univ.*, vol. 18, no. 5, pp. 526–534, 2013.

[20] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.

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

[22] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Inf. Fusion*, vol. 40, pp. 57–75, Mar. 2017.

[23] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.

[24] J. Ren, Z. Liu, X. Zhou, G. Sun, and C. Bai, "Saliency integration driven by similar images," *J. Vis. Commun. Image Represent.*, vol. 50, pp. 227–236, Jan. 2017.

[25] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[26] X. Zhou, Z. Liu, C. Gong, and W. Liu, "Improving video saliency detection via localized estimation and spatiotemporal refinement," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2993–3007, Nov. 2018.

[27] Z. Liu, J. Li, L. Ye, G. Sun, and L. Shen, "Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2527–2542, Dec. 2017.

[28] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infr. Phys. Technol.*, vol. 76, pp. 52–64, May 2016.

[29] J. Han, E. J. Pauwels, and P. De Zeeuw, "Fast saliency-aware multi-modality image fusion," *Neurocomputing*, vol. 111, pp. 70–80, Jul. 2013.

[30] J. Sun, H. Zhu, Z. Xu, and C. Han, "Poisson image fusion based on Markov random field fusion model," *Inf. Fusion*, vol. 14, no. 3, pp. 241–254, 2013.

[31] Z. Liu, Y. Xue, H. Yan, and Z. Zhang, "Efficient saliency detection based on Gaussian models," *IET Image Process.*, vol. 5, no. 2, pp. 122–131, Mar. 2011.

[32] X. Zhang, Z. Ren, D. Rajan, and Y. Hu, "Salient object detection through over-segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 1033–1038.

[33] K. Fu, C. Gong, I. Y.-H. Gu, and J. Yang, "Normalized cut-based saliency detection by adaptive multi-level region merging," *IEEE Trans Image Process*, vol. 24, no. 12, pp. 5671–5683, Dec. 2015.

[34] K. Fu, I. Y.-H. Gu, and J. Yang, "Saliency detection by fully learning a continuous conditional random field," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1531–1544, Jul. 2017.

[35] P. Zhang, Y. Yuan, C. Fei, T. Pu, and S. Wang, "Infrared and visible image fusion using co-occurrence filter," *Infr. Phys. Technol.*, vol. 93, pp. 223–231, Sep. 2018.

[36] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum, "Drag-and-drop pasting," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 631–637, Jul. 2006.

[37] J. Wen, Y. Li, and H. Gong, "Remote sensing image fusion on gradient field," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2006 pp. 643–646.

[38] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, 2015.

[39] Y. Cao, H. Wei, H. Zhao, and N. Li, "An effective approach for land-cover classification from airborne lidar fused with co-registered data," *Int. J. Remote Sens.*, vol. 33, no. 18, pp. 5927–5953, 2012.

[40] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection," *Inf. Fusion*, vol. 6, no. 4, pp. 283–300, 2005.

[41] A. P. Dempster, *Upper and Lower Probabilities Induced by a Multivalued Mapping*. Berlin, Germany: Springer, 2008.

[42] W. J. Stewart, *Probability, Markov Chains, Queues, and Simulation. The Mathematical Basis of Performance Modeling*. Princeton, NJ, USA: Princeton Univ. Press, 2009, pp. 40–60.

[43] M. A. Sutton, C. Mingqi, W. H. Peters, Y. J. Chao, and S. R. McNeill, "Application of an optimized digital correlation method to planar deformation analysis," *Image Vis. Comput.*, vol. 4, no. 3, pp. 143–150, 1986.

[44] R. Klette, *Concise Computer Vision*. New York, NY, USA: Springer, 2014, pp. 1–41.

[45] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *Proc. Int. Workshop Artif. Neural Netw.*, 2005, pp. 195–201.

[46] J. Inglis, "A mathematical theory of evidence," *Technometrics*, vol. 20, no. 1, p. 106, Jan. 1976.

[47] J. M. Dicarlo and B. A. Wandell, "Rendering high dynamic range images," *Proc. SPIE*, vol. 3956, pp. 392–401, May 2000.

[48] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *Proc. ACM SIGGRAPH*, San Diego, CA, USA, 2003, pp. 313–318.

[49] Y. Liu, X. Chen, H. Peng, and Z. F. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36 pp. 191–207, Jul. 2017.
[50] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.
[51] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 3, p. 1850018, 2018.
[52] A. L Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.
[53] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.
[54] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.
[55] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–9.
[56] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visual image fusion through infrared feature extraction and visual information preservation," *Infr. Phys. Technol.*, vol. 83, pp. 227–237, Jun. 2017.
[57] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48 pp. 11–26, Aug. 2019.
[58] D. Liu, P. An, R. Ma, C. Yang, L. Shen, and K. Li, "Three-dimensional holoscopic image coding scheme using high-efficiency video coding with kernel-based minimum mean-square-error estimation," *J. Electron. Imag.*, vol. 25, no. 4, p. 043015, 2016.
[59] J. W. Roberts, F. B. Ahmed, and J. A. Van Aardt, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, p. 023522, 2008.
[60] X. Zhou, L. Zhi, G. Sun, L. Ye, and X. Wang, "Improving saliency detection via multiple kernel boosting and adaptive fusion," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 517–521, Apr. 2016.
[61] Y.-J. Rao, "In-fibre Bragg grating sensors," *Meas. Sci. Technol.*, vol. 8, no. 4, p. 355, 1997.
[62] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, Mar. 2002.
[63] D. Liu, P. An, R. Ma, C. Yang, and L. Shen, "3D holoscopic image coding scheme using HEVC with Gaussian process regression," *Signal Process., Image Commun.*, vol. 47, pp. 438–451, Sep. 2016.

JING LI received the M.Sc. degree in geographic information system and remote sensing from Beijing Forestry University, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree with the Department of Information Technology and Cyber Security, People's Public Security University of China.

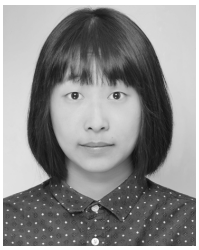
His research interests include remote sensing, image processing, and pattern recognition.

HONGTAO HUO received the M.S. and Ph.D. degrees in computer science and technology from Beijing Forestry University, Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the Department of Information Technology and Cyber Security, People's Public Security University of China, China. He has published several papers in various refereed journals. His research has been supported by the National Key Research and Development Program of China and the Ministry of Public Security Technology Research Program. His research interests include image processing, pattern recognition, remote sensing application technology, and image forensics.

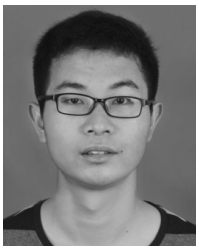
CHENHONG SUI received the Ph.D. degree in spatial information science and technology from the Huazhong University of Science and Technology, Wuhan, China, in 2015.

She is currently an Associate Professor with the School of Opto-Electronic Information Science and Technology, Yantai University, Yantai, China. Her research interests include remote sensing image processing, pattern recognition, and image quality assessment.



CHENCHEN JIANG received the bachelor's degree in geographic information system from the Jinling College, Nanjing University, Nanjing, China, in 2017. She is currently pursuing the master's degree with the Institute of Information Technology and Cyber Security, People's Public Security University of China.

Her research interests include data mining in crime, public security, remote sensing image classification, and target detection.



CHANG LI received the B.S. degree in information and computing science from the Wuhan Institute of Technology, Wuhan, China, in 2012, and the Ph.D. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, in 2018.

He is currently a Lecturer with the Department of Biomedical Engineering, Hefei University of Technology, Hefei, China. His current research interests include the areas of biomedical signal processing, hyperspectral image analysis, computer vision, pattern recognition, and machine learning.