remote sensing

MDPI

Article

An Efficient Method for Infrared and Visual Images Fusion Based on Visual Attention Technique

Yaochen Liu, Lili Dong *, Yang Chen and Wenhai Xu

School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China; yaochen_liu@dlmu.edu.cn (Y.L.); chenyang123@dlmu.edu.cn (Y.C.); xuwenhai@dlmu.edu.cn (W.X.)
* Correspondence: donglili@dlmu.edu.cn

Received: 15 January 2020; Accepted: 25 February 2020; Published: 29 February 2020

Abstract: Infrared and visible image fusion technology provides many benefits for human vision and computer image processing tasks, including enriched useful information and enhanced surveillance capabilities. However, existing fusion algorithms have faced a great challenge to effectively integrate visual features from complex source images. In this paper, we design a novel infrared and visible image fusion algorithm based on visual attention technology, in which a special visual attention system and a feature fusion strategy based on the saliency maps are proposed. Special visual attention system first utilizes the co-occurrence matrix to calculate the image texture complication, which can select a particular modality to compute a saliency map. Moreover, we improved the iterative operator of the original visual attention model (VAM), a fair competition mechanism is designed to ensure that the visual feature in detail regions can be extracted accurately. For the feature fusion strategy, we use the obtained saliency map to combine the visual attention features, and appropriately enhance the tiny features to ensure that the weak targets can be observed. Different from the general fusion algorithm, the proposed algorithm not only preserve the interesting region but also contain rich tiny details, which can improve the visual ability of human and computer. Moreover, experimental results in complicated ambient conditions show that the proposed algorithm in this paper outperforms state-of-the-art algorithms in both qualitative and quantitative evaluations, and this study can extend to the field of other-type image fusion.

Keywords: image fusion; visual attention; saliency map

1. Introduction

Image fusion is an important branch of information fusion, which involves many research fields such as deep learning, image processing and computer vision [1–3]. Among them, the infrared and visible image fusion has great application value in the practical engineering. The visible image contains rich texture information and conforms to the human visual system. Infrared images distinguish targets from background based on differences in thermal radiation. By combining the complementary information of visible and infrared image, it is possible to generate fused images that are more conducive to human decision-making or computer vision tasks, which has been applied to many fields such as the military, target detection, surveillance and so on [4–9]. An excellent image fusion algorithm must contain the following conditions. First, the fused image can contain the useful information of the source image. Second, it gets a good robustness in complex environments such as noise. Third, it cannot generate artifacts that hinder human observation or application.

In recent years, scholars have proposed many infrared and visible image fusion algorithm through different schemes. They can be mainly divided into five categories including subspace-based methods [10–12], multi-scale transform-based methods [13–15], sparse representation-based

methods [16–18], deep learning-based methods [2,19,20] and other methods [21,22]. Next, the ideas of these methods are briefly introduced.

The subspace-based method first projects high-dimensional source image into low-dimensional space, and then fuses the information contained in the subspace, such as principal component analysis (PCA) [10], independent component analysis (ICA) [11], robust principal component analysis (RPCA) [12] and so on. Since processing subspace data consumes less time and memory than source images, this kind of method has the advantage of high computational efficiency. However, its stability is not good for application in complex environment. Multi-scale transform-based methods (MST) have been widely applied since it was introduced into the field of infrared and visible image fusion such as quaternion wavelet transform (QWT) [13], pyramid transform [14] and so on. Generally, this kind of method has three fusion steps [15]: first of all, the source image is decomposed into multiple scale, each of which contains different feature information. Then, the information of multiple scale is fused according to the designed fusion rule. Finally, the fused image is obtained by reconstruction. Spare representation-based methods aim to sparsely represent the source image by learning dictionary for image fusion. The fused image based on the spare representation method is very consistent with human visual perception. However, it lacks the reservation of details [16,17]. This kind of method has four fusion steps [18]: each source image is decomposed into several overlapping blocks. Then, a complete dictionary is learned from many high-quality natural images, and sparse coding is performed on each patch to obtain sparse representation coefficients. Third, sparse representation coefficients are fused according to a given fusion rule. Finally, the learned over complete dictionary is utilized to reconstruct the fused image. Deep learning-based methods are to imitate the behavioral perception mechanism of the human brain, which has strong adaptability and feature extraction ability [2]. However, this kind of method is computationally intensive and requires high hardware equipment [19,20]. In addition, there are other ideas and perspectives that inspire new image fusion method, such as entropy [21], total variation [22] and so on.

With the development of computer vision technology, saliency-based methods have been successfully implemented to infrared and visible image fusion because it effectively utilizes the complementary information of the source image. Saliency-based methods have four fusion steps including saliency region segmentation, designing a fusion rule of saliency region, designing a fusion rule of the background region, and reconstruction [23]. Meng et al. [24] used significant detection method to extract the interesting region, which can be mapped to the region of the fused image. Zhang et al. [25] utilized a super-pixel saliency model to obtain the interesting regions of the infrared image, which can retain the target information of the infrared image to the fused image. Liu et al. [26] integrated saliency detection into the fusion sparse representation framework and used global and local saliency maps to obtain the weight of reconstruction. Ma et al. [27] used the saliency maps to extract the targets from the base layers. Then, the least squares method was used to fuse the detail layers. However, existing saliency-based methods typically only extract significant targets in the infrared image during the fusion process. It may be inappropriate for the infrared and visible image fusion, as the interesting regions of the visible image and the weak targets cannot be captured. In addition, the weak activity level cannot be properly enhanced, which will lead to the fusion algorithm cannot be applied to complex environments. This is particularly true when noise appears in the background or when the source image has low contrast.

To overcome the above weaknesses, we propose a fusion algorithm based on visual attention technology in this paper. Specifically, visual attention technology be used to obtain visual features of the infrared and visible image at the same time, which ensures the fused image is appropriated for human or computer vision tasks. The feasibility and superiority of visual attention technique are also analyzed. Moreover, fusion rule is designed to combine the complementary features and enhance the weak features. It can enable the fused image to preserve more tiny details, which leads to accurate expression of the detection scene.

The main contributions of this work are the following three aspects. First, we propose a special visual attention system to extract visual features. The co-occurrence matrix is utilized to select a particular modality, and then a linear normalization method is used to fairly extract visual feature of each pixel. Secondly, we design a feature fusion strategy based on the saliency maps to combines the visual features. The saliency maps of the infrared and visible image obtained by the special visual attention system are used to integrate complementary information, and then the guided filter is utilized to decompose multi-scale information for enhancing weak features. Last but not least, experimental results in the public image fusion data set show that the proposed fusion algorithm has great robustness and can extend to the field of other-type image fusion.

The rest of this paper is organized as following. In Section 2, we briefly introduced visual attention technique for image fusion. In Section 3, the proposed fusion method in this paper is present in detail. A comparison and analysis of experimental results is presented in Section 4. Finally, the conclusions are presented in Section 5.

2. Visual Attention Technique for Image Fusion

In this section, we discussed the feasibility and advantages of visual attention technique in the field of infrared and visible image fusion. In addition, we also analyze the original VAM.

2.1. Feasibility

According to Section 1, it shows that the fusion framework is mainly composed of three parts including feature extraction, feature fusion and reconstruction. Among them, feature extraction is a key step, which determines the feature information contained in the fused image. When people observe images, the human visual system actively seeks interesting regions to reduce search tasks such as object detection and recognition, so the human brain's attention to the whole image is not balanced. Therefore, visual attention technique as a feature extraction method is theoretically feasible for image fusion because it can extract the visual features of the image by simulating the observation mechanism of the human eye.

We further explain feasible from the perspective of real application. The purpose of most existing fusion algorithms is to generate fusion images that help to perform human eye or computer vision tasks. We study the image fusion algorithm based on human visual characteristics, which can efficiently improve the visual sensory comfort of the fused image and help humans to monitor the complex environment. This is especially important in practical application such as military, surveillance. Therefore, the fusion image obtained by visual attention technique is also feasible in practical applications.

2.2. Superiority

The advantages of fusion images based on visual attention technique over existing fusion methods are two fold. First, it can effectively capture the interesting regions and remove a lot of redundant information in the source image, so that the fused image has a nice visual effect. Secondly, the human visual attention system can effectively extract the accurate information of targets from various interference information, which makes the fusion algorithm have strong stability. There have been various kinds of visual attention models to realize the simulation of visual attention systems, and it has been proved that the attention target can be accurately extracted even under the interference of noise. Therefore, compared with traditional methods, the algorithm based on visual attention technique has the potential to produce higher visual effects in the results, and also has great potential for better robustness in practical applications.

2.3. The Original VAM for Image Fusion

In order to extract visual attention features, Itti et al. [28] have established the visual attention model. The original VAM first generates intensity, orientation and colors saliency maps corresponding

to gray, texture and color features of the input image, and then fuses the saliency maps to obtain a gray image to represent the parts that are easily noticed. However, applying the original visual attention model directly to the field of image fusion may not be an effective method. Figure 1 shows a typical example. Figure 1a is an infrared image, and the interesting region is shown in the red box. Consider that the source images are gray images, only intensity and orientation modality are used. Its different modalities saliency maps are shown in Figure 1b–d. In saliency maps, the larger the pixels, the stronger the visual attention. It can be seen that Figure 1b can find the interesting region. Figure 1c cannot effectively extract the features of the target, but redundant information is extracted (as shown in the yellow box). Although Figure 1d retains the activity level of the interesting region, it is disturbed by the orientation modality. Figure 1e is the visible image whose different modalities saliency maps are shown in Figure 1f–h. We can see that a signal intensity or orientation saliency map can express the salient features from the visible image (as shown in the red box), but the saliency information is mutually suppressed in both modalities.

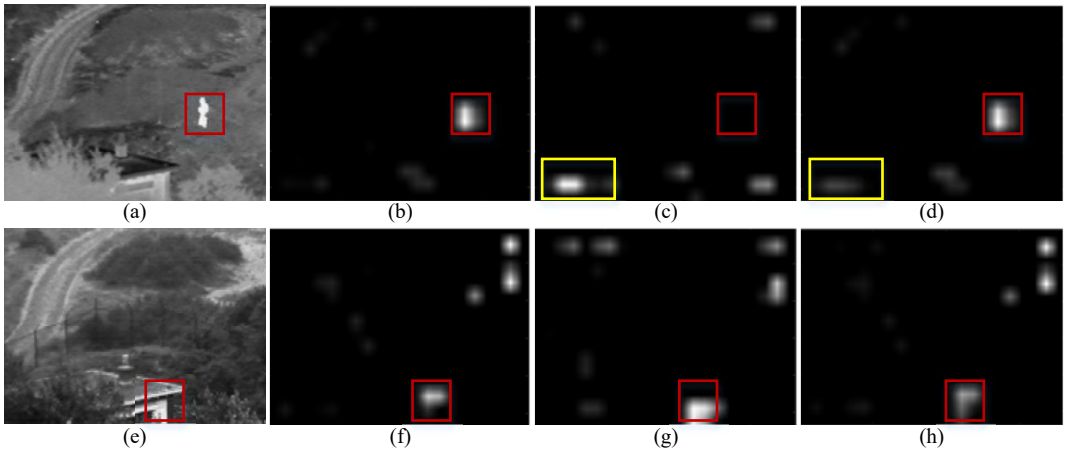

Figure 1. Different modalities saliency maps of the original VAM. (**a**) is an infrared image, its intensity, orientation and the combination of two modality saliency map are (**b**–**d**), respectively. (**e**) is a visible image, its intensity, orientation and the combination of two modality saliency map are (**f**–**h**), respectively.

Based on the above analysis, selecting both modalities to collect feature information is not an effective strategy because it causes the significant features of intensity and orientation to suppress each other and may introduce a lot of redundant features. However, single selection of intensity or texture features is also not a good strategy, which likely lead to loss of useful information. In addition, Figure 1 also shows that the original VAM suppressed the saliency of the weak activity position. The reason is that it adopts an iterative nonlinear normalization operator to simulate the feature competition scheme, which suppresses the weak activity location by the strong global peak.

3. Image Fusion Algorithm Based on Visual Attention Technique

Figure 2 shows the framework of the proposed infrared and visible image fusion algorithm. First, we propose a special visual attention system that is used to extract salient features. Then, feature fusion strategy based on visual saliency map is designed, which can combines the interesting region and enhances the texture details. The brief introduction of the proposed algorithm is given in Section 3.

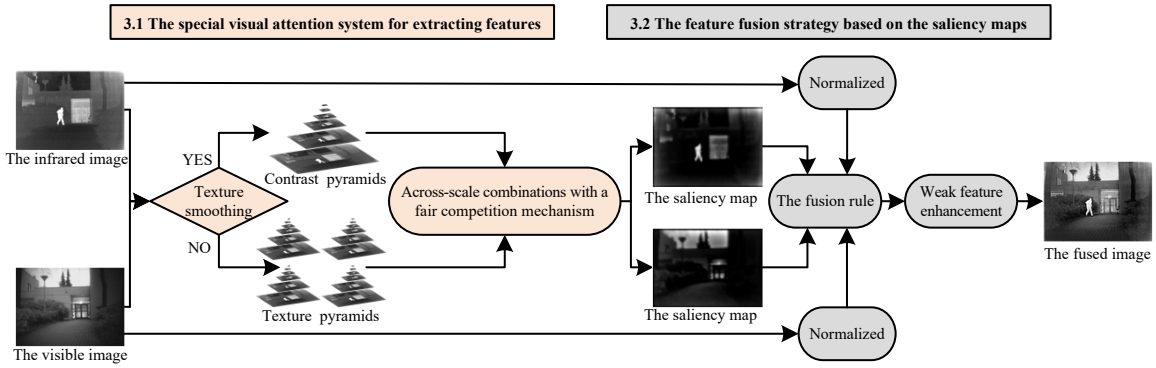

Figure 2. The proposed fusion framework.

3.1. The Special Visual Attention System for Extracting Features

According to Section 2, we can utilize the original VAM to generate the fused image. However, there are two disadvantages to this model. On the one hand, it doesn't automatically select the optimal modality, which may give unnecessary interference. On the other hand, due to the weak activity region suppression mechanism, it likely causes the background of the fused image to be smooth. Therefore, we propose a special visual attention system to extract the visual features of the source image for image fusion.

3.1.1. Modality Selection Based on Texture Complication Evaluation

Features collected by the intensity and orientation modality are different, so we must find an optimal modality. To solve this problem, we experimented on the TNO image fusion data set that is a public data set in the field of infrared and visible image fusion and contains many different military relevant scenarios. The observed results are as follows:

1. Collecting saliency information from intensity modality is an effective method when image texture smoothing. Since it is very sensitive to the image contrast, the intensity modality can use local contrast to measure the image activity level in the absence of direction information.
2. When the texture details are rich, only the orientation modality can be used to achieve the best effect. In texture-rich image, gradient information in different directions is strong. Therefore, when synthesizing the four directions features maps into a single saliency map, the saliency information is much stronger than the signal intensity modality.

We utilize co-occurrence matrix to quantize the texture complication. Different from other texture evaluation metrics, it takes advantage of the rotation invariance of texture feature and thus has strong resistance to noise [29]. The co-occurrence matrix $g(x,y)$ is normalized as follows:

$$g(x,y) = \frac{p(x,y)}{\sum_{x=0}^{N_g-1}\sum_{y=0}^{N_g-1} p(x,y)} \tag{1}$$

where $p(x,y)$ is the number of occurrences of pixel. N_g is the quantized gray level. For reducing computational complexity, we usually quantize the image to N_g=16.

Through the co-occurrence matrix, the local pattern and alignment rules of image can be analyzed, and then the second statistic-contrast is obtained. The equation is as follows:

$$con = \sum_{x=0}^{N_g-1}\sum_{y=0}^{N_g-1} (x-y)^2 \cdot g(x-y) \tag{2}$$

where *con* is the contrast. The large *con* means rich texture features. However, when the size of source images is different, the calculated texture complication may have large deviation. In order to overcome this problem, this paper will interpolate the image size to the same size (96×96) when performing texture complication evaluation. After experimenting on TNO data sets, the threshold *con* was found to be 0.314 which only work for military relevant scenarios. When *con* exceed the threshold, only the orientation modality is used to obtain SM. Otherwise, only intensity modality is used.

3.1.2. Across-Scale Combinations with a Fair Competition Mechanism

After modal selection, we will rely on the contrast or texture features of the image to generate saliency maps. In order to accurately evaluate the activity level of each pixel, this paper attempts to adopt a fair competition mechanism. The saliency map acquisition methods for the two modes are as follows:

(1) Intensity modality

First, the image is gaussian sampled to generate a gaussian pyramid I_σ, where the pyramid scale σ is in the range of $[0, 1, \ldots, 8]$.Then, the center-surround operator is utilized to generate feature maps. The equation is as follows:

$$I(c,s) = |I(c) \ominus I(s)| \tag{3}$$

where $\ominus$ indicates that the size of the two images is adjusted to be the same and then the matrix subtraction operation is performed, $s = c + \sigma$, $c \in \{2,3,4\}$, $\sigma \in \{3,4\}$. Therefore, we can get six intensity feature maps $I(c,s)$.

Then a linear normalization operator is used to simulate a fair competition mechanism which can reasonably measure the activity level of the targets and the background. The equation is as follows:

$$SM_c = Nor(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} I(c,s)) \tag{4}$$

where $Nor(\ )$ is linear normalization.$\oplus$ indicates that the size of the two images is adjusted to be the same and then the matrix addition operation is performed. SM_c is the intensity saliency map. In this way, a fair competition mechanism is formed so that the weak active in the background can also be evaluated.

(2) Orientation modality

The orientations pyramid $O(\theta)_\sigma$ is obtained by filtering I_σ in four angle with gabor filter:

$$O(\theta)_\sigma = I_\sigma * Gabor(\theta) \tag{5}$$

where $Gabor(\ )$ is the gabor filter and $\theta \in \{0°, 45°, 90°, 135°\}$.

Then, the center-surround operator also is utilized to generate feature maps. The equation is as follows:

$$O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)| \tag{6}$$

Therefore, we can get 24 orientation feature maps $O(c,s,\theta)$.

The feature maps in the four directions are also calculated using the fair competition mechanism to obtain four directions maps, and then summed and normalized to generate the final saliency map SM_o.The equation is as follows:

$$SM_o = Nor(\sum_{\theta \in \{0°,45°,90°,135°\}} Nor(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} O(c,s,\theta))) \tag{7}$$

Figure 3 shows the saliency maps from the original VAM and the special visual attention system. Figure 3a,f are the infrared and visible image, respectively. Figure 3b,g are the infrared and visible

intensity saliency maps by the original VAM. Figure 3d,i are infrared and visible the orientation saliency maps by the original VAM, respectively. We can see that the original VAM can effectively extract the visually significant areas in the image, but cannot accurately measure the activity level of the details in the background. Figure 3c,h are the infrared and visible intensity saliency maps by the special visual attention system, respectively. We can see that intensity modality not only effectively extract the strong interesting regions, but also accurately measure the activity level in the background. Figure 3e,j are the infrared and visible orientation saliency maps by the special visual attention system, respectively. We can see that orientation modality also can overcome the phenomenon of weak activity area suppression.

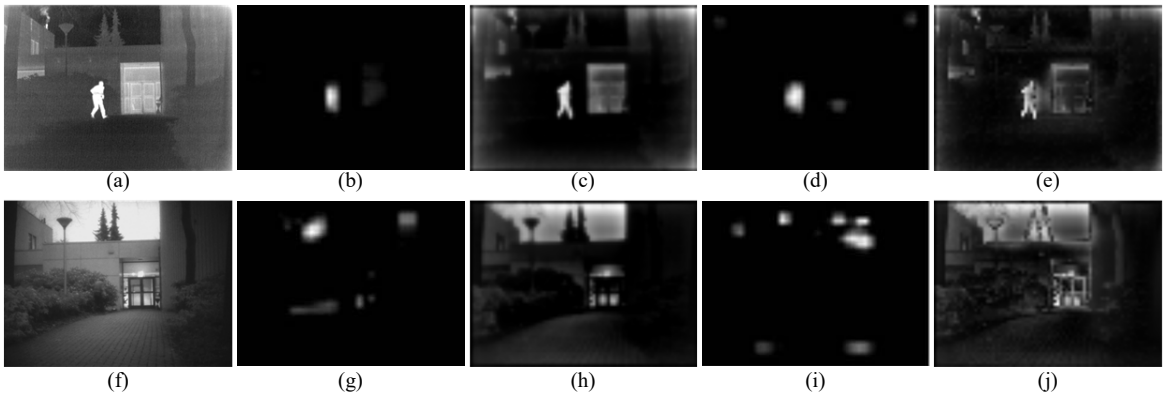

Figure 3. The saliency maps from the original VAM and the special visual attention system. (**a**,**f**) are the infrared and visible image, respectively. (**b**,**g**) are the infrared and visible intensity saliency maps by the original VAM. (**c**,**h**) are the infrared and visible intensity saliency maps by the special visual attention system. (**d**,**i**) are the infrared and visible orientation saliency maps by the original VAM. (**e**,**j**) are the infrared and visible orientation saliency maps by the special visual attention system.

3.2. Feature Fusion Strategy Based on the Saliency Maps

A survey of existing saliency-based methods by Meher et al. [23] shows that most saliency-based methods are to separately extract targets from the infrared image and then superimpose them into the visible image (accurate extraction of the target contour is often difficult), which not only loses a lot of complementary information in visible image, but also the noise in the visible image will greatly affect the robustness of the fusion algorithm. Different from existing methods, we use the proposed special visual attention system to extract the features of the infrared and visible image, respectively. Then, normalizing the source image to make sure that the input variables are used equally. Finally, combine the visual features contained in the saliency maps to get the fused saliency maps $The\_fused\_SM(x,y)$. The equation is as follows:

$$The\_fused\_SM(x,y) = \sum_{n=1}^{2} Nor[f_n(x,y)] \times \frac{SM_n(x,y)}{SM_1(x,y) + SM_2(x,y)} \tag{8}$$

where $f_n(x,y)$ is nth the source image, and its corresponding saliency map is $SM_n(x,y)$, $n \in \{1,2\}$. The range of linear normalization is $[0,1]$. It can be seen that the fused saliency map contains complementary information of the source image. However, the comparison method is used to fuse image features, which may result in smooth textures. It is not conducive to the observation of the weak targets. To solve this problem, detail features need to be appropriately enhanced.

Since the pixel value of the texture is low, we perform logarithmic transformation to emphasize the low gray value region, the transformation method is as follows:

$$S(x,y) = \log[1 + The\_fused\_SM(x,y)] \tag{9}$$

where $S(x,y)$ is transformed image.

Then, the guided filter is utilized to extract multi-scale detail information. Guided filter is an edge-preserving filter proposed by He et al. [30], which is widely used in image processing [31]. The equation is as follows:

$$R_i(x,y) = guided\_filter(\omega_i, \varepsilon_i) * S(x,y) \tag{10}$$

where $guided\_filter(\ )$ is guided filter, ω_i and ε_i are the filter window and coefficient, respectively. $R_i(x,y)$ is the ith output layer using $S(x,y)$ as both input and guidance image. As the scale of detail continues to increase, the time-consuming will grow linearly, so it is appropriate for i to be 3. The parameters ω_i and ε_i have been discussed in many literatures [30,32]. Therefore, due to the length of the article, it will not be explained in detail here.

We can combine the output layers to obtain the enhanced fusion saliency map $Enhanced\_SM(x,y)$, the equation is as follows:

$$Enhanced\_SM(x,y) = \sum_{i=1}^{N} \omega_i [S(x,y) - \log(R_i(x,y)+1)] + The\_fused\_SM(x,y) \tag{11}$$

where η_i is the weight coefficient, and its sum is 1.Then multiply the enhanced feature saliency map by 255 to get the fused image $Fused(x,y)$. In order to guarantee that all pixel values are between $[0,255]$, we also design an overflow judgment. The equation is as follows:

$$Fused(x,y) = \begin{cases} 255 & Fused(x,y) \geq 255 \\ 0 & Fused(x,y) \leq 0 \\ Fused(x,y) & 0 < Fused(x,y) < 255 \end{cases} \tag{12}$$

4. Experimental Results and Analyses

To test the effectiveness of the fusion algorithm in this paper, we utilize the most commonly used infrared and visible image fusion sets as experimental data. In addition, we compared with classic and state-of-the-art algorithms from qualitative and quantitative, respectively. The computational complexity of our proposed algorithm and comparative algorithms is also discussed. Finally, we have extended the proposed fusion algorithms to the field of medical, multi-focus and multi-exposure image fusion.

4.1. Experimental Settings

(1) Image sets

In experimenting, we selected seven pairs of visible and infrared images as the experimental sample, which was collected from the site: https://figshare.com/articles/TNO$_$Image$_$Fusion$_$Dataset/1008029. Figure 4 shows the seven pairs of images including "Soldier-in-trench", "Soldier-behind-smoke", "Kaptein-1123", "Airplane", "Road", "Bench", and "Kaptein-1654". Among them, "Soldier-in-trench" contains significant infrared targets and texture-smooth visible images. In "Soldier-behind-smoke", the visible image contains smoke, and the infrared image has the interesting region. "Kaptein-1123" not only has infrared targets but also contains rich background information in the visible image. The contrast of "Airplane" is very low. "Road" is a set of images taken at night. Both visible and infrared image in "Kaptein-1654" contain significant information. "Bench" contains significant infrared targets, but the background of the visible image has a lot of noise information. The size of images is 768×576, 768×576, 620×450, 595×328, 256×256, 620×450 and 280×280, respectively. Each image pair is pre-registered, which can fully verify the effect of the proposed algorithm from different scenes.

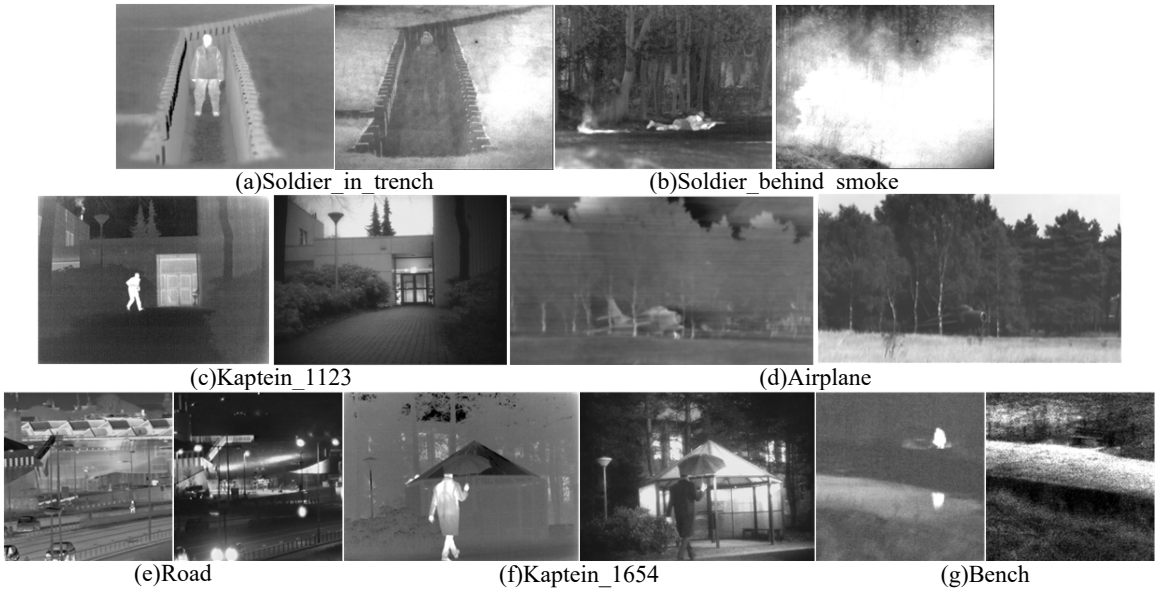

Figure 4. The infrared and visible image sets used in experiments. (**a**) The "Soldier-in-trench" image set. (**b**) The "Soldier-behind-smoke" image set. (**c**) The "Kaptein-1123" image set. (**d**) The "Airplane" image set. (**e**) The "Road" image set. (**f**) The "Kaptein-1654" image set. (**g**) The "Bench" image set.

(2) Compared algorithms

The proposed algorithm base on visual attention technology (PROPOSE) is compared with seven image fusion algorithms based on gradient transfer (GTF) [22], convolutional neural network (DENSE) [33], guided filter (GFF) [32], latent low-rank representation (LATLRR) [34], visual saliency map and weighted least square optimization (VSM-WSM) [27], feature extraction and visual information preservation (FEVIP) [35] and discrete wavelet transform (DWT) [15], respectively. Among these compared algorithms, DWT is a classic multi-scale transform-based method that first divides the source images into two scales, then fuses the information contained in the two scales, and finally reconstruct the fused image. GFF uses the guided filter to obtain the fusion weight value, which is also a classic image fusion algorithm. In addition, we also compare with five state-of-art image fusion algorithms. VSM-WSM is a saliency-based method that utilizes the gaussian filter to divide the image into base and detail parts, and then fuses them by the least square method and the weighting method, respectively. FEVIP is also a saliency-based method that first reconstructs the infrared background by quadtree and Bedizer interpolation, then subtract the infrared image to obtain the target, and finally superimposes the visible image to obtain the fused image. DENSE, a deep learning-based method, uses convolutional neural network to extract various features and combine them to obtain fused results. LATLRR is a sparse representation-based method that utilize latent low-rank representation to decompose the source image into two layers and design different rules to obtain the fused image. GTF uses gradient transfer and total variation minimization to design decomposed method and fusion rules. These five methods were proposed in the last three years. These compared algorithm codes are derived from public data, and the parameters are the default.

The above seven image fusion algorithms can obtain desired fusion results, and the types of these algorithms are different. By comparing with these algorithms, the superiority of the proposed algorithm can be effectively shown.

(3) Computation platform

The proposed algorithm and the compared algorithms are all implemented on a PC-Windows 10 platform with Inter (R) Core (TM) i7-8700K @ 3.70 GHz processor, 16GB RAM, and CeForce GTX 1080 Ti. Besides, DENSE is performed on graphics processing unit (GPU), while other algorithms are programmed in Matlab.

4.2. Qualitative Evaluation

The qualitative evaluation for infrared and visible image fusion can be achieved by the visual effect of the fused image. The experimental results of the DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE are shown in Figures 5a–h–11a–h.

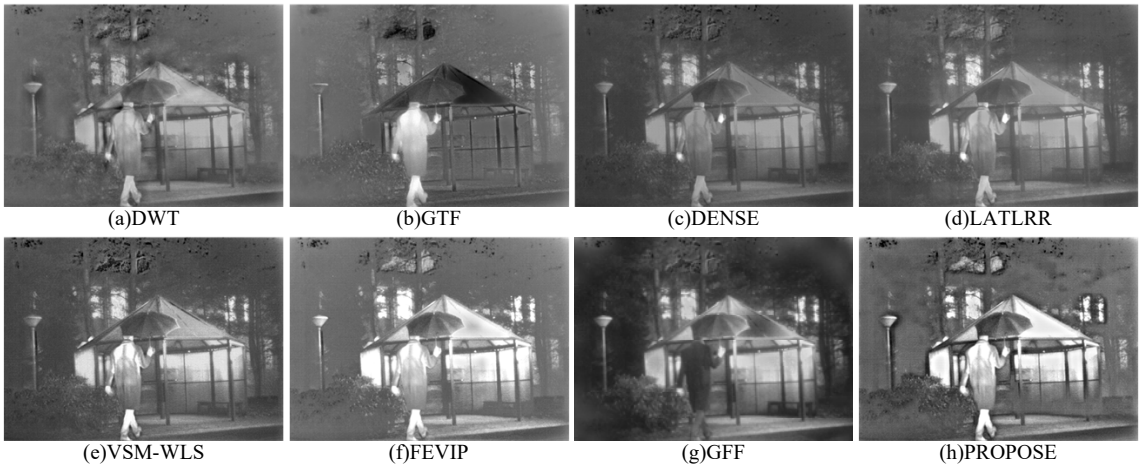

Figure 5. The fusion results of the "Kaptein-1654" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

Figure 5 shows the fusion results of the "Kaptein-1654" image set. Each fusion algorithm can accomplish the purpose of image fusion. However, different fusion algorithms may produce different fused images. The DWT result loses the significant information and tiny details because this algorithm cannot fully extract various features from the source image (see Figure 5a). The GTF result can preserve the interesting infrared region, but a lot of information contained in the visible image is lost. The DENSE and LATLRR results are better visually than Figure 5a, but these algorithms are also unable to retain the significant information. The VSM-WLS result can preserve the target, but the contrast is low. The background of the FEVIP result is overall bright, which leads to poor visual effects. The GFF results lose a lot of infrared information. However, the result of the proposed algorithm not only can better highlight the interesting region of the source image but also suit the human perception. In summary, the fusion result of the "Kaptein-1654" image set proves that the proposed algorithm can effectively combine the complementary information of the source image.

In addition to verifying the effect of retaining complementary information, it is necessary to test the ability of the proposed algorithm to preserve tiny details. Figure 6 shows the fusion results of the "Kaptein-1123" image set. It can be seen that the DWT fusion result is not only low in contrast but also blurry on the floor. The GTF fusion result can retain the significant information of the infrared image, but it loses a lot of visible background information. The DENSE and LATLRR fusion results are unclear in the texture areas. The FEVIP fusion result has artifacts in the sky. The GFF fusion result appears a lot of noise. The VSM-WLS fusion result is the best of the comparison results, but it cannot retain the salient and detailed regions in the visible image. However, the fusion result of the proposed algorithm not only preserves the interesting region but also has rich tiny details. In summary, the fusion results of the "Kaptein-1123" image set prove that the proposed algorithm can effectively retain the details of the source image, and the artificial information does not appear in the background.

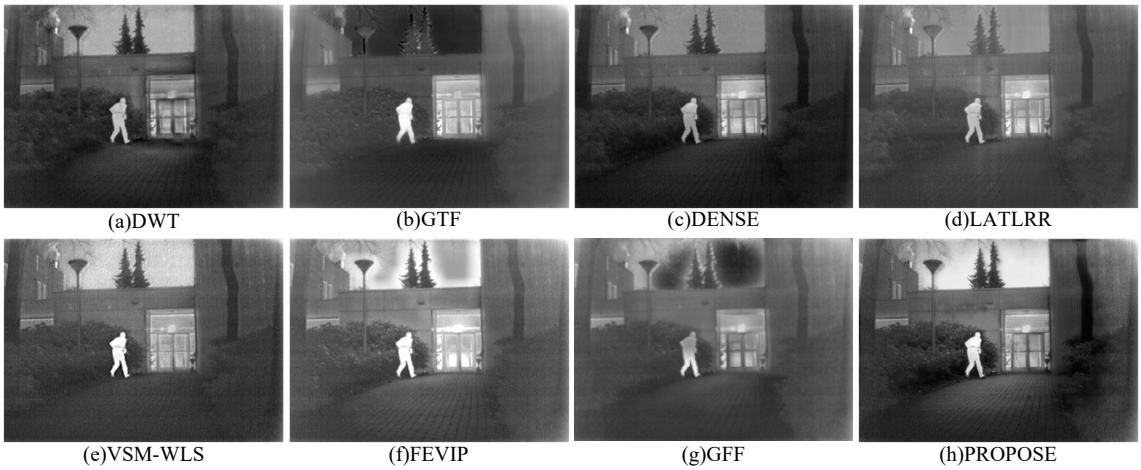

Figure 6. The fusion results of the "Kaptein-1123" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

We also experiment with the source images that contain noise to verify the robustness of the proposed algorithm. Figure 7 shows the fusion results of the "Bench" image set. The DWT, DENSE, LATLRR and GFF fusion results lose significant information and have low contrast. The GTF fusion result is disturbed by the noise of the source image. The VSM-WLS and FEVIP fusion results appear some noise in the background, which results in poor visibility. However, due to the better noise immunity of PROPOSE, the fusion result of the proposed algorithm is very clear and overcome noise interference. To further illustrate the robustness of the proposed algorithm, we also chose to experiment in a smoke-interfering environment. In Figure 8, the fusion results of GFF, GTF and the proposed algorithm can clearly observe the interesting region. On the contrary, other fusion results cannot see the target. However, the GTF fusion result can retain target because a lot of visible information is lost. Compared with the GFF fusion result, the proposed algorithm can observe a more complete significant information. In summary, the proposed algorithm can be applied to environments that contain noise.

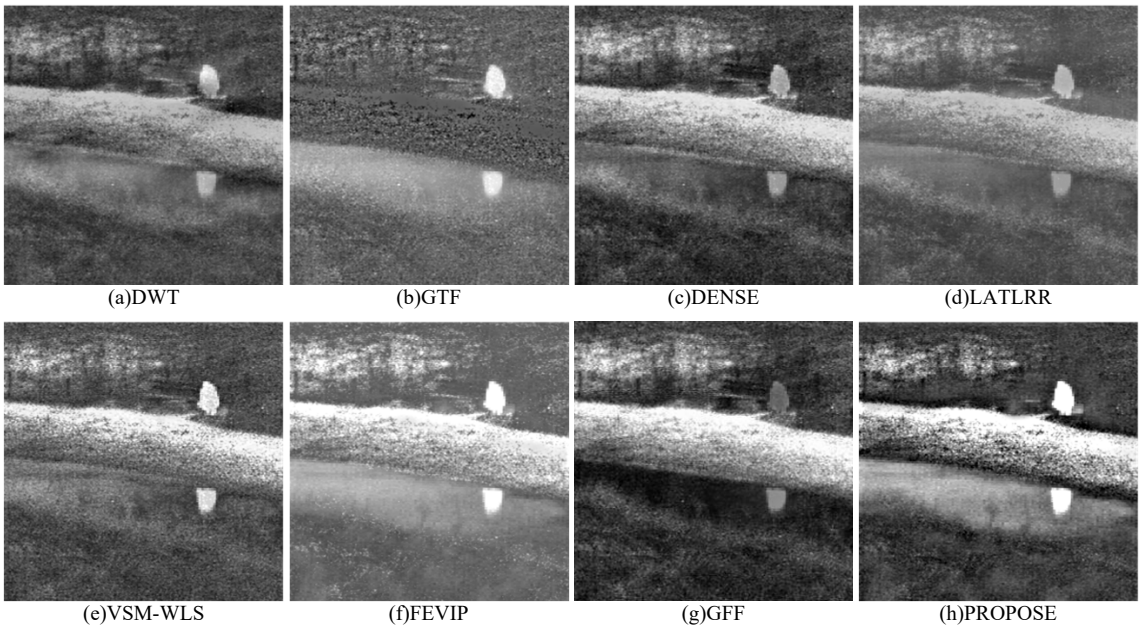

Figure 7. The fusion results of the "Bench" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

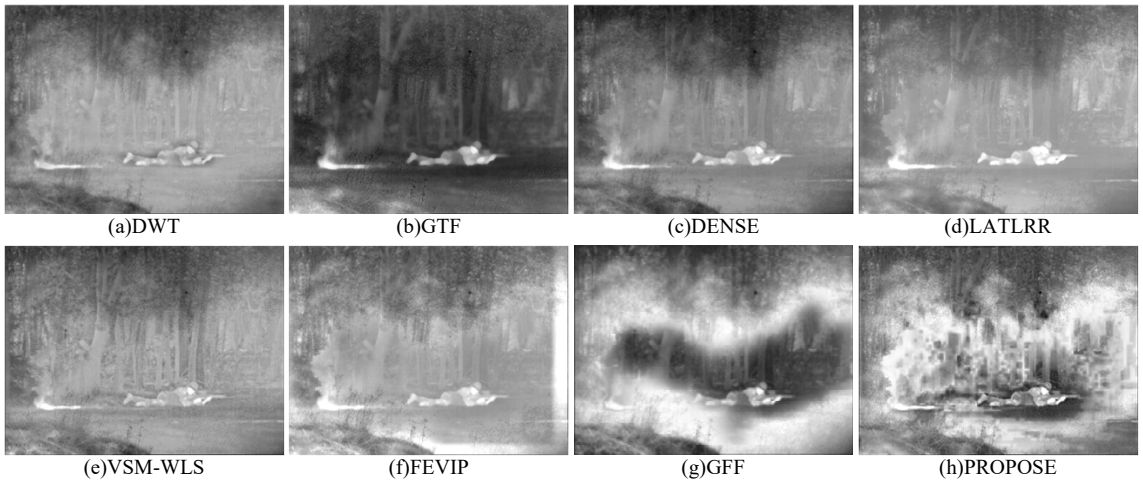

Figure 8. The fusion results of the "Soldier-behind-smoke" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

In addition, we also experimented with the source images taken at night. Figure 9 shows the fusion results of the "Road" image set. The fusion results of DWT, DENSE, LATLRR and GFF cannot highlight the interesting regions such as light and vehicles. The GTF fusion result is very blurred, which is not suitable for human eye observation. The fusion results of VSM-WLS and FEVIP have better contrast than other comparison fusion results. However, because the tiny features are properly enhanced, our fusion results are the clearest among all fusion results. In summary, the proposed algorithm is suitable for observation at night.

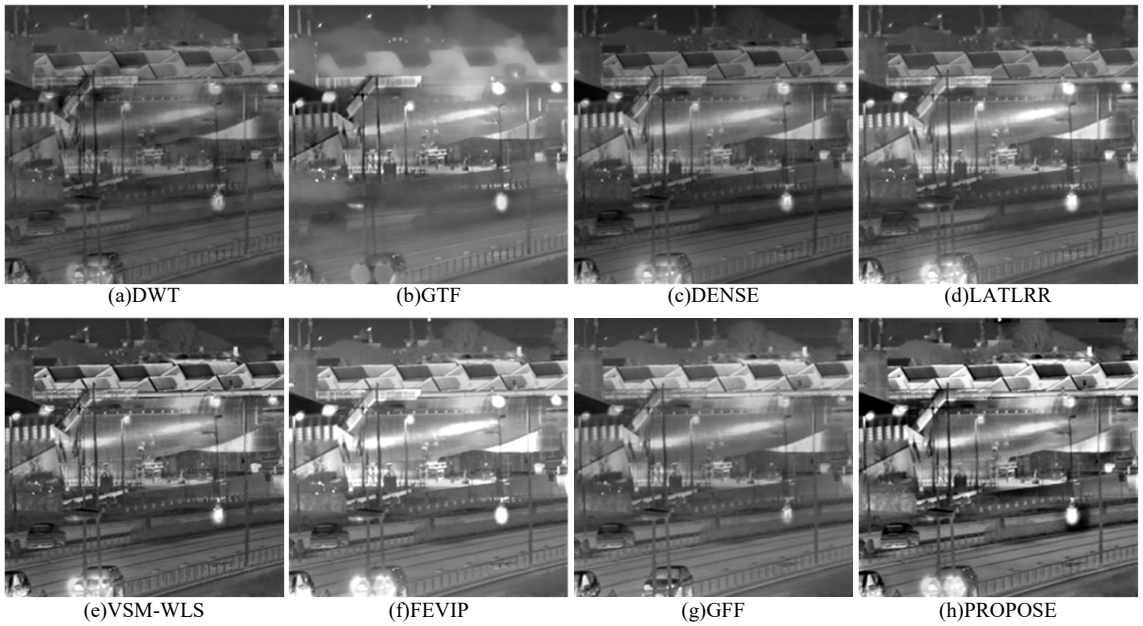

Figure 9. The fusion results of the "Road" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

Finally, we experiment with low contrast source images to test the effectiveness of the proposed algorithm. Figure 10 shows the fusion results of the "Airplane" image set. It can be seen that the fusion results of the DWT, LATTRR and VSM-WLS have low contrast. The GTF and GFF fusion results lose a lot of complementary information. The FEVIP fusion result has artifacts in the sky. However, because the proposed algorithm protects weak activity regions, the problem of low contrast is solved. Figure 11 shows the fusion results of the "Soldier-in-trench" image set. We can see that the fusion

result of the proposed algorithm is clearer than the comparison method. In summary, when the source images contrast is low, the fusion algorithm in this paper can still have a good effect.

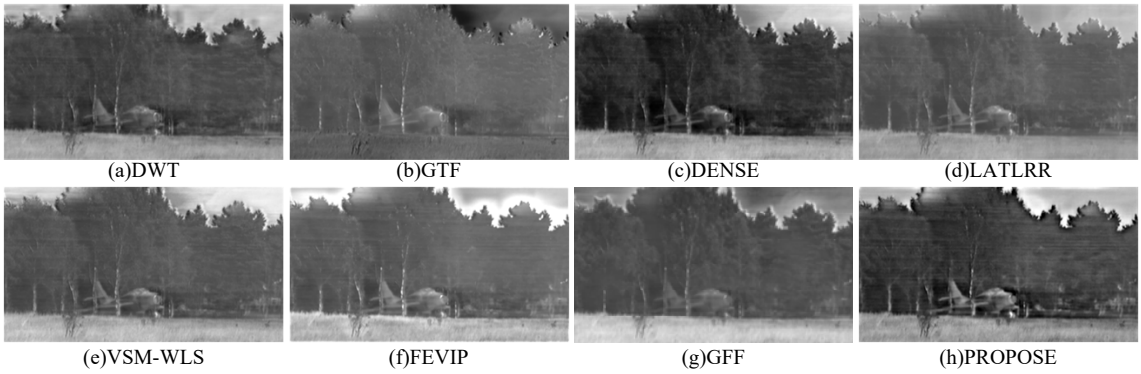

Figure 10. The fusion results of the "Airplane" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

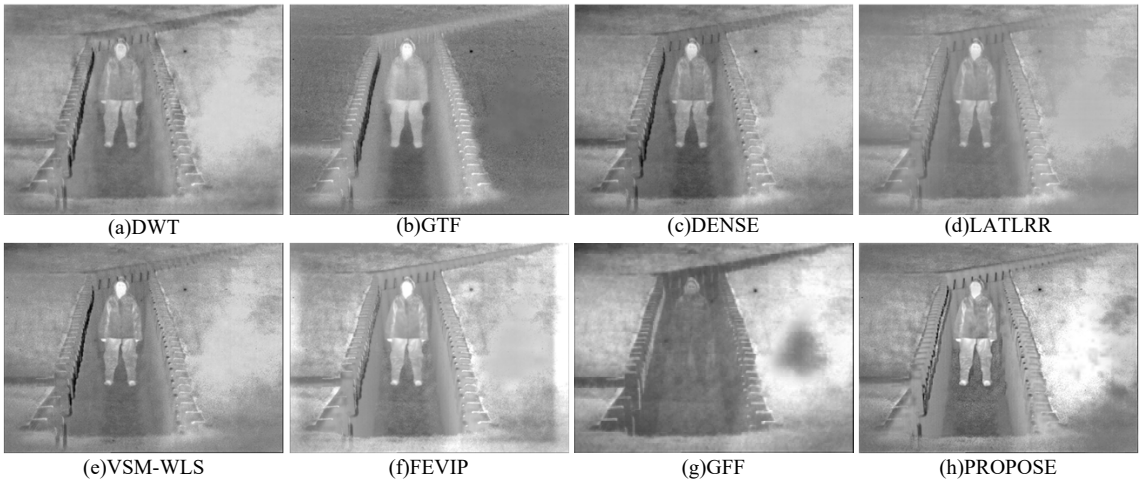

Figure 11. The fusion results of the "Soldier-in-trench" image set. (**a**–**h**) are the result of DWT, GTF, DENSE, LATLRR, VSM-WLS, FEVIP, GFF and PROPOSE, respectively.

In conclusion, the qualitative evaluation results show that the proposed algorithm is suitable for application in various complex environments.

4.3. Quantitative Evaluation

The qualitative evaluation has the disadvantage of human intervention and time-consuming, therefore we also utilize the quantitative method to evaluate the fused images. Quantitative evaluation mainly relies on mathematical calculations to describe image features, which are a very reliable evaluation method. However, the fused images may have some noise, it causes the results of evaluation to be incorrect. To avoid this problem, we will employ multiple evaluation metrics to comprehensively evaluate the fused images. This subsection first introduces the concept of each metric and then the evaluation results are analyzed.

4.3.1. Quantitative Metrics

In recent years, a series of methods for quantitatively evaluating fused images have been proposed. Liu et al. [36] have surveyed the existing quantitative metrics for image fusion and pointed out that these metrics can be divided into three categories: information metrics, image texture metrics and human perception metrics. In this paper, we selected representative metrics from each category

including entropy (EN) [37], mutual information (MI) [38], spatial frequency (SF) [39] and visual information fidelity (VIF) [40]. Each metric is defined as follows:

(1) information metrics: EN and MI

EN means the richness of the information contained in the fused image. A larger EN value reflects the better performance in information content. This metrics can be calculated as:

$$EN = -\sum_{l=0}^{L-1} p_i \log p_i \tag{13}$$

where L is gray levels, p_i is the normalized histogram of the corresponding gray level in the fused image.

MI shows the amount of information that the source images convey to the fused image, which can evaluate the ability of the fusion algorithm to combine the complementary information. A larger MI value means that a lot of complementary information is transferred from the source image to the fused results. This metrics can be calculated as:

$$MI = MI_{A,F} + MI_{V,F} \tag{14}$$

$$MI_{X,F} = \sum_{x,f} p_{X,F}(x,f) \log \frac{p_{X,F}(x,f)}{p_X(x)p_F(f)} \tag{15}$$

where $MI_{A,F}$ and $MI_{V,F}$ are the amount of information that is transferred from infrared and visible to the fused image, respectively. $p_{X,F}(x,f)$ is the joint histogram of the source image X and the fused image F. $p_X(x)$ and $p_F(f)$ are the marginal histograms of X and F, respectively.

(2) image texture metrics: SF

SF measure the clarity of image texture. A larger SF value means the rich tiny details in the fused image. This metrics can be calculated as:

$$SF = \sqrt{RF^2 + CF^2} \tag{16}$$

$$RF = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - F(i,j-1))^2} \tag{17}$$

$$CF = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - F(i-1,j))^2} \tag{18}$$

(3) human perception metrics: VIF

The VIF is used to evaluate the visual effect of the fused image. The larger VIF value, the more consistent with human visual perception. This metric relies on natural scene statistical models, image signal distortion channels, and human visual distortion models.

4.3.2. Quantitative Evaluation Results

The quantitative evaluation results of all image fusion algorithms are shown in Table 1. The bold value in Table 1 represents the maximum value in the corresponding column, and the larger value indicates better performance.

Table 1. The results of quantitative evaluation for eight algorithms.

Group	Fusion Algorithm	Evaluation Index			
		EN	**MI**	**SF**	**VIF**
Kaptein-1654	DWT	6.4807	12.9614	8.3551	0.6842
	GFF	7.0315	14.0630	9.7603	0.7862
	VSM-WSM	6.7426	13.4853	12.158	0.7933
	FEVIP	6.6648	13.3297	11.803	0.8394
	DENSE	6.4133	12.8267	6.9027	0.6976
	GTF	6.5244	13.0488	9.1600	0.6558
	LATLRR	6.5546	13.1093	7.6490	0.6651
	PROPOSED	**7.0438**	**14.0877**	**14.314**	**0.9089**
Bench	DWT	7.0807	14.1614	18.0912	0.6409
	GFF	**7.4934**	**14.9868**	23.2069	0.8241
	VSM-WSM	7.1646	14.3293	26.3591	0.6546
	FEVIP	6.9297	13.8594	21.7905	0.6886
	DENSE	7.3496	14.6993	21.6091	0.6627
	GTF	6.7781	13.5562	21.8149	0.7237
	LATLRR	6.8550	13.7101	15.8557	0.5950
	PROPOSED	7.3676	14.7352	**27.5356**	**0.8357**
Kaptein-1123	DWT	6.9721	13.9442	8.2929	0.7879
	GFF	6.8563	13.7127	7.0714	0.7057
	VSM-WSM	6.9714	13.9429	10.580	0.8895
	FEVIP	7.1691	14.3383	9.2726	0.9552
	DENSE	6.9073	13.8147	7.1428	0.8230
	GTF	6.9581	13.9162	6.4738	0.7037
	LATLRR	6.7016	13.4032	6.4051	0.7232
	PROPOSED	**7.4212**	**14.8424**	**10.701**	**0.9851**
Soldier-in-trench	DWT	6.8856	13.7711	10.5937	0.8135
	GFF	7.1845	**14.8668**	13.0178	0.9179
	VSM-WSM	6.9738	13.9477	13.8672	**0.9587**
	FEVIP	6.9431	13.8864	11.7221	0.8934
	DENSE	6.9996	13.9993	10.1163	0.8834
	GTF	6.6015	13.2031	12.5306	0.8535
	LATLRR	6.5548	13.1097	8.11250	0.7282
	PROPOSED	**7.2061**	14.4122	**14.0425**	0.9365
Airplane	DWT	6.6942	13.3885	5.5777	0.7933
	GFF	6.4477	12.8954	5.2579	0.7515
	VSM-WSM	6.6104	13.2210	5.9202	0.8567
	FEVIP	6.7302	13.4606	7.1918	0.8798
	DENSE	7.0350	14.0700	6.1090	0.9516
	GTF	5.8563	11.7127	4.3989	0.6881
	LATLRR	6.4571	12.9143	4.2604	0.7423
	PROPOSED	**7.1444**	**14.2889**	**8.1186**	**1.0698**
Soldier_behind_smoke	DWT	6.9039	13.8079	8.5219	0.7425
	GFF	7.5263	15.1527	11.884	**0.9369**
	VSM-WSM	6.9735	13.9470	11.831	0.9064
	FEVIP	7.0271	14.0543	11.626	0.9149
	DENSE	7.0523	14.1046	7.8967	0.8117
	GTF	6.6015	13.2030	10.924	0.8302
	LATLRR	6.9239	13.8479	7.7548	0.7209
	PROPOSED	**7.6489**	**15.2979**	**15.063**	0.7842
Road	DWT	6.6485	13.2971	12.8952	0.5427
	GFF	7.1527	14.3056	17.7289	0.7165
	VSM-WSM	7.2656	14.5313	22.8475	0.6218
	FEVIP	7.3325	14.6650	21.5730	0.7106
	DENSE	7.0858	14.1717	13.9473	0.5809
	GTF	7.0878	14.1756	14.6903	0.5906
	LATLRR	7.1803	14.3606	16.4928	0.5532
	PROPOSED	**7.5860**	**15.1721**	**26.6966**	**0.7334**

First of all, the ability of all fusion algorithms to combine the complementary information and the information richness of fused images are evaluated by information metrics (MI and EN), respectively. The EN evaluation results show that the fused images of GTF contain the least amount of information because of the improper evaluation of the source images gradient. Besides, the fused images of DWT also have low values, the reason for this problem is that DWT cannot extract various image features. Due to better feature extraction capabilities, the other comparative fusion algorithms (GFF, VSM-WSM, FEVIP, DENSE and LATLRR) have rich information in the fused image. However, since the special visual attention system can measure the activity level of tiny details, the fused images of the proposed algorithm contain more information than the fusion result of the comparison fusion algorithm. For the

ability to combine source image information, the MI evaluation results show that GFF, VSM-WSM and FEVIP have higher evaluation values, which indicates that these three algorithms can better combine the source image information. However, FEVIP utilizes quadtree to reconstructs the infrared background, which may result in the loss of infrared information. GFF has poor robustness so that the features of different scenes cannot be accurately extracted by guided filtering. This may cause the redundant information to be transmitted to the fused image. Therefore, compared with FEVIP and GFF, the proposed algorithm achieves high performance, except that GFF has the highest value on the "Bench" and "Soldier-in-trench" because they contain much useless visible image information.

Secondly, the details of the fused image are evaluated by texture metric (SF). The evaluation results show that the DWT and GTF evaluation results have poor performances. Among them, DWT can't extract the significant features, which lead to texture smoothing. The GTF fusion results have low texture complexity because a lot of visible information is lost. But other comparative fusion algorithms have good gradient information. However, compared with GFF, VSM-WSM, FEVIP, DENFUSE and LATLRR, the evaluation results of the proposed algorithm have great advantages. This is because the fusion algorithm in this paper not only can better extract the salient features of the image, but also design a feature combination strategy to retain the extracted interesting region in the fused image. In conclusion, the proposed algorithm can obtain the fused image with clear texture.

Thirdly, the human visual perception of the fused image is evaluated by human perception metrics (VIF). The evaluation results show that the fused images of the saliency-based methods (VSM-WSM, FEVIP) have better visual in the fusion results of all comparison algorithms because LATLRR, DWT, GFF, GTF and DENSE cannot effectively extract interesting region. However, VSM-WSM and FEVIP use a simple weighted method to fuse the extracted features, which results in the fused image with low contrast. Compared with VSM-VSM and FE-VIP, the fused images of the proposed algorithm has better performance. This is because the feature fusion method designed in this paper can better combine the extracted features, and we also enhance the tiny features.

Moreover, in order to better appear the quantitative performance of the proposed algorithm, Figure 12 also shows the average assessment results of each algorithms. The average score on Figure 4 is shown here. We can see that the EN and MI values of PROPOSE are the largest because the special visual attention system in this paper can effectively extract the complementary information of the source images. SF also has the best performance because the proposed algorithm can properly enhance the details. Finally, since the visual features can be effectively extracted, the VIF of PROPOSE has the highest score.

4.4. Computational Costs

In addition to qualitative and quantitative evaluation, we need to measure the computational cost of the fusion algorithm, which determines the practical application value of these image fusion algorithms. Running time is used to estimate the computational cost of all fusion algorithms.

The evaluation results of each algorithm processing 7 image sets are listed in Table 2, where the bold value denotes the maximum value in the corresponding column, and the larger value indicates better performance. It can be seen that the longest time is the LATLRR, because this algorithm contains a large number of parameters in the LATLRR model. The time of GTF is wasted by traversing the pixels multiple times to obtain gradient information. The VSM-WSM repeats the filtering operation, resulting in an increase in time complexity. The DWT and GFF algorithms have low computational complexity and therefore take less time. Although DENSE performs multi-layer feature extraction, the convolution layer size is small, therefore this image fusion algorithm is faster. Compared with other algorithms, the computational efficiency of the proposed algorithm is second to FEVIP that utilizes an efficient quadtree decomposition strategy, but the fusion algorithm in this paper wastes a lot of computing resources on the guided filter and the co-occurrence matrix. However, efficient programming through C++ without using MATLAB can increase the speed of the fusion algorithm, and as the hardware

conditions continue to increase, real-time programs based on the proposed algorithm will not become a problem.

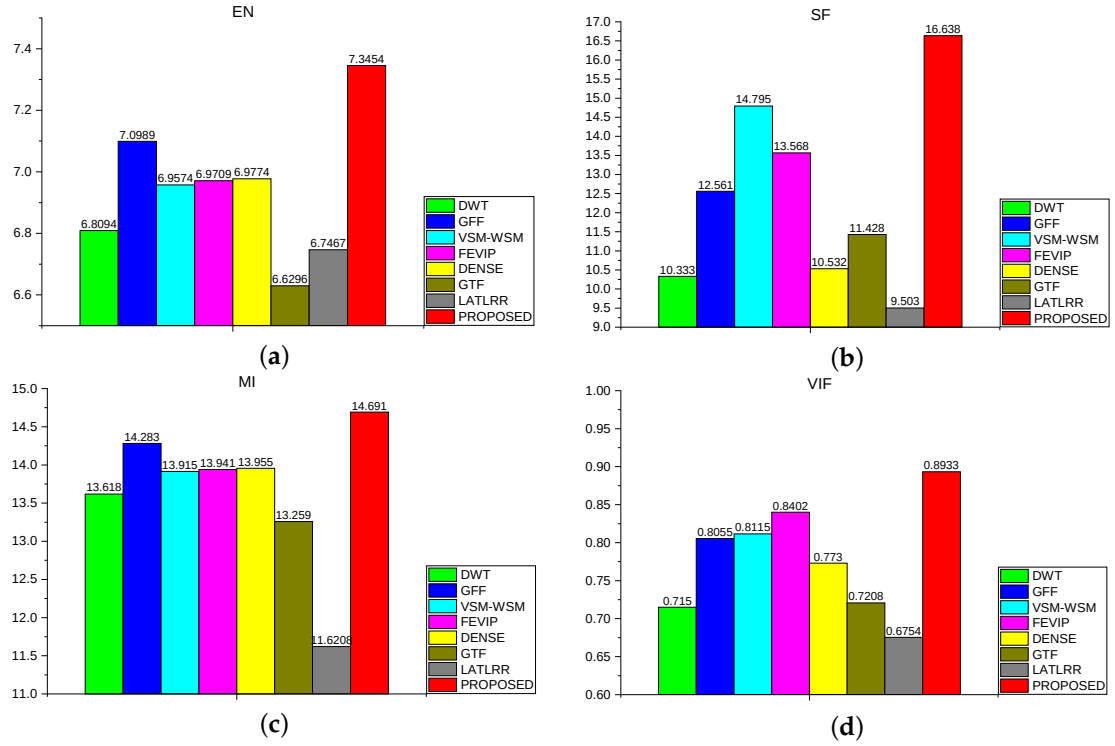

Figure 12. The average quantitative assessment of each fusion method. (**a**) The evaluation results of EN. (**b**) The evaluation results of SF. (**c**) The evaluation results of MI. (**d**) The evaluation results of VIF.

Table 2. The comparison of the computational costs.

Fusion Algorithm	Kaptein-1654	Bench	Kaptein-1123	Soldier-in-Trench	Airplane	Soldier_behind_SMOKE	Road
DWT	0.2417	0.0981	0.2085	0.3058	0.1647	0.2735	0.1009
GFF	0.2047	0.0628	0.2042	0.3086	0.1669	0.3046	0.0709
VSM-WSM	1.6626	0.2464	1.6521	2.8538	1.1455	2.8557	0.2132
FEVIP	**0.0827**	**0.0467**	**0.0904**	**0.0997**	**0.0761**	**0.1074**	0.0500
DENSE	0.4263	0.3906	0.6727	0.4877	0.6018	0.5047	0.3663
GTF	3.5350	0.2946	3.1396	6.2277	3.1988	3.8056	0.2673
LATLRR	58.971	13.081	59.234	108.09	42.076	110.11	10.999
PROPOSED	0.2054	0.0514	0.2161	0.3570	0.1534	0.3443	**0.0487**

4.5. Extension to Other-Type Image Fusion Field

To further exhibit robustness of the fusion algorithm, we extend the proposed algorithm to multi-modal image fusion, multi-medical image fusion and multi-exposure image fusion. Different types of image have their own characteristics, so the framework and factors considered are also different during the image fusion process. However, since the proposed framework has certain universality and the visual attention technology is not susceptible to complex environments, the fusion algorithm in this paper can be extended to other types of images.

The fusion results of the multi-medical images are shown in Figure 13a. It can be seen that the fusion result can effectively retain the complementary information of the source image.

The fusion results of the multi-modal images are shown in Figure 13b. It can be seen that the fusion result is clearer than the source image. This is because the results obtained by the special visual attention system is similar to the fusion image obtained by the weighted average method, and then the quality of the detail information is improved by the feature fusion strategy.

The fusion results of the multi-exposure image are shown in Figure 13c. It can be seen that the proposed algorithm can successfully solve the over-exposure problem by extracting the exposed regions and retain the details of the source image at the same time.

The discussion and analysis of the experimental results prove that the fusion framework in this paper is a reasonable way. Therefore, future research on fusion algorithms can be continued using this framework.

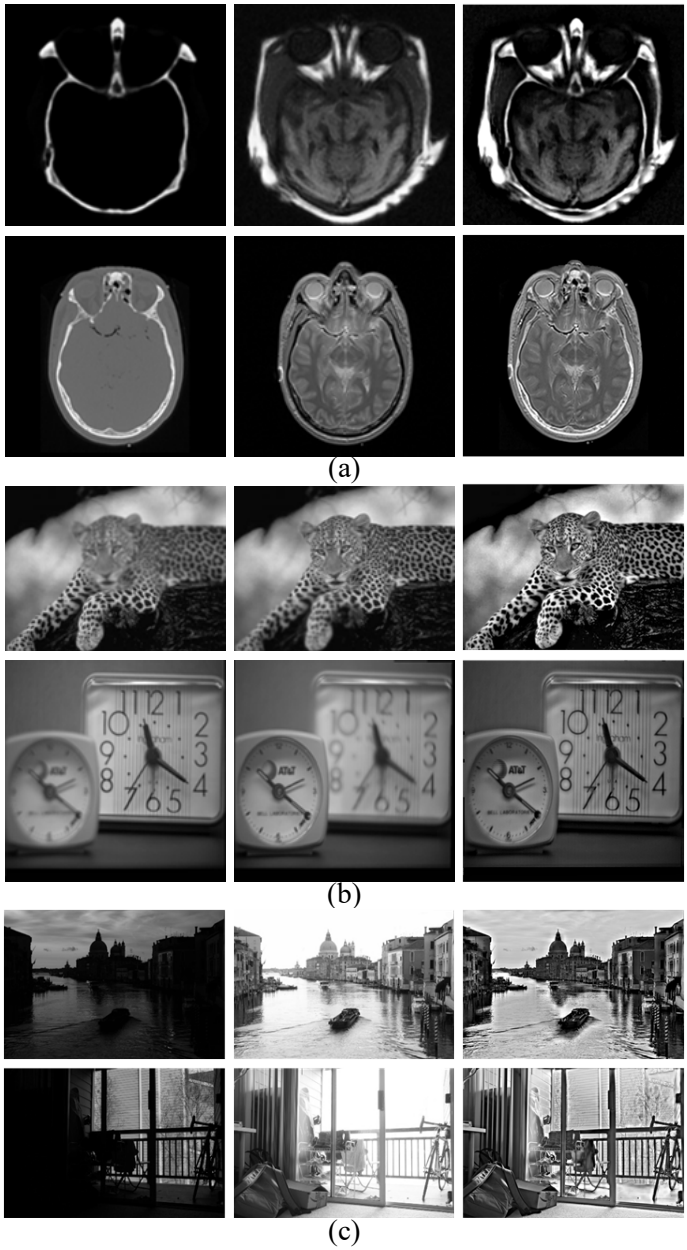

Figure 13. Fusion results of the other three modal images. (**a**) Medical image fusion. (**b**) Multi-focus image fusion. (**c**) Multi-exposure image fusion.

4.6. Algorithm Limitation Analysis

The proposed fusion algorithm still has some limitations that may weaken its performance under certain conditions.

1. Optimal modality selection threshold. As can be seen from Section 3.1, the optimal modality has played a key role in the proposed algorithm. However, for different data sets, the contrast

and texture features of the interesting region are different, so thresholds need to be adjusted for different data sets. In order to resolve this issue, we can experiment on many different data sets, and then the threshold empirical equation can be fitted according to the experimental results. In this way, we can automatically select thresholds for different data sets.

2. Manual parameter selection. As can be seen from Section 3.2, this paper proposes that the feature fusion strategy has the problem of manually designing parameters, which result in the algorithm not being able to run automatically. One possible solution is to match the gray-scale histogram. If there is an over-enhancement phenomenon, adjust the feedback.

5. Conclusions

In this paper, we propose a visible and infrared image fusion algorithm based on visual attention. The proposed algorithm first uses the co-occurrence matrix to select a particular modality. Then, a fair competition mechanism was utilized to obtain the saliency maps. Moreover, a feature fusion strategy is designed to fuse the visual features and appropriately enhance the tiny features, which ensure that the proposed algorithm can be applied to real environment. Experimental results show that the proposed fusion algorithm has advantages in both quantitative and qualitative evaluation, and can be extended to other types of images. Although the proposed method may have some disadvantages (as described in Section 4.6), we have proposed corresponding solutions. In conclusion, the image fusion algorithm in this paper is meaningful and worthwhile. In the future, we will utilize the color modality of visual attention technology to study other-type of color image fusion, such as multispectral image.

Author Contributions: Y.L. and L.D. conceived the proposed algorithm and wrote the paper; Y.C. designed and performed the experiments; W.X. revisited the paper and provided technical guidance. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (61701069) and the Fundamental Research Funds for the Central Universities of China (3132019340, 3132019200).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ardeshir Goshtasby, A.; Nikolov, S.G. Image fusion: Advances in the state of the art. *Inf. Fusion* **2007**, *8*, 114–118. [CrossRef]
2. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2017**, *42*, 158–173. [CrossRef]
3. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [CrossRef]
4. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [CrossRef]
5. Pohl, C.; Van Genderen, J. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854. [CrossRef]
6. Zhang, Z.; Xin, B.; Deng, N.; Xing, W.; Cheng, Y. An investigation of ramie fiber cross-section image analysis methodology based on edge-enhanced image fusion. *Measurement* **2019**, *145*, 436–443. [CrossRef]
7. Palsson, F.; Sveinsson, J.; Ulfarsson, M. Sentinel-2 Image Fusion Using a Deep Residual Network. *Remote Sens.* **2018**, *18*, 1290. [CrossRef]
8. Liu, J.; Pan, C.; Wang, G. A Novel Geometric Approach to Binary Classification Based on Scaled Convex Hulls. *IEEE Trans. Neural Netw.* **2009**, *20*, 1215–1220.
9. Pan, X.; Li, L.; Yang, H.; Liu, Z.; Yang, J.; Zhao, L.; Fan, Y. Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing* **2017**, *229*, 88–99. [CrossRef]
10. Li, H.; Liu, L.; Huang, W. An improved fusion algorithm for infrared and visible images based on multi-scale transform. *Infrared Phys. Technol.* **2016**, *74*, 28–37. [CrossRef]

11. Cvejic, N.; Bull, D.; Canagarajah, N. Region-based multimodal image fusion using ICA bases. *IEEE Sens. J.* **2007**, *7*, 743–751. [CrossRef]
12. Bouwmans, T.; Javed, S.; Zhang, H. On the Applications of Robust PCA in Image and Video Processing. *Proc. IEEE* **2018**, *106*, 1427–1457. [CrossRef]
13. Chai, P.; Luo, X.; Zhang, Z. Image fusion using quaternion wavelet transform and multiple features. *IEEE Access* **2017**, *5*, 6724–6734. [CrossRef]
14. Bulanon, D.M.; Burks, T.F.; Alchanatis, V. Image fusion of visible and thermal images for fruit detection. *Biosyst. Eng.* **2009**, *103*, 12–22. [CrossRef]
15. Zhang, Z.; Blum, R.S. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proc. IEEE* **1999**, *87*, 1315–1326. [CrossRef]
16. Yang, B.; Li, S. Multifocus image fusion and restoration with sparse representation. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 884–892. [CrossRef]
17. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [CrossRef]
18. Zhang, Q.; Liu, Y.; Blum, R.; Blum, R.; Han, J.; Tao, D. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Inf. Fusion* **2018**, *40*, 57–75. [CrossRef]
19. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [CrossRef]
20. Liu, Y.; Dong, L.; Ji, Y.; Xu, W. Infrared and Visible Image Fusion through Details Preservation. *Sensors* **2019**, *19*, 4556. [CrossRef]
21. Zhao, J.; Cui, G.; Gong, X. Fusion of visible and infrared images using global entropy and gradient constrained regularization. *Infrared Phys. Technol* **2017**, *81*, 201–209. [CrossRef]
22. Ma, J.; Chen, C.; Li, C. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [CrossRef]
23. Meher, B.; Agrawal, S.; Panda, R.; Abraham, A. A survey on region based image fusion methods. *Inf. Fusion* **2019**, *48*, 119–132. [CrossRef]
24. Meng, F.; Song, M.; Guo, B. Image fusion based on object region detection and non-subsampled contourlet transform. *Comput. Electr. Eng.* **2017**, *62*, 375–383. [CrossRef]
25. Zhang, B.; Lu, X.; Pei, H. A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled shearlet transform. *Infrared Phys. Technol.* **2015**, *73*, 286–297. [CrossRef]
26. Liu, C.; Qi, Y.; Ding, W. Infrared and visible image fusion method based on saliency detection in sparse domain. *Infrared Phys. Technol.* **2017**, *83*, 94–102. [CrossRef]
27. Ma, J.; Zhou, Z.; Wang, B. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [CrossRef]
28. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
29. Tahir, M.A.; Bouridane, A.; Kurugollu, F. Accelerating the computation of GLCM and Haralick texture features on reconfigurable hardware. In Proceedings of the IEEE International Conference on Image Processing, Singapore, 24–27 October 2004.
30. He, K.; Sun, J.; Tang, X. Guided image filter. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [CrossRef]
31. Zhang, W.; Dong, L.; Pan, X.; Zhou, J.; Qin, L.; Xu, W. Single Image Defogging Based on Multi-Channel Convolutional MSRCR. *IEEE Access* **2019**, *7*, 72492–72504. [CrossRef]
32. Li, S.; Kang, X.; Wang, Z. Image fusion with guided filtering. *IEEE Trans. Image Process* **2013**, *22*, 2864–2875. [PubMed]
33. Prabhakar, K.R.; Srikar, V.S.; Babu, R.V. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4714–4722.
34. Li, H.; Xiao, J. Infrared and visible image fusion using Latent Low-Rank Representation. *arXiv* **2019**, arXiv:1804.08992.
35. Zhang, Y.; Zhang, L.; Bai, X.; Zhang, L. Infrared and Visual Image Fusion through Infrared Feature Extraction and Visual Information Preservation. *Infrared Phys. Technol.* **2017**, *83*, 227–237. [CrossRef]

36. Liu, Z.; Blasch, E.; Xue, Z. Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 94–109. [CrossRef] [PubMed]
37. Bai, X.; Zhou, F.; Xue, B. Edge preserved image fusion based on multiscale toggle contrast operator. *Image Vis. Comput.* **2011**, *29*, 829–839. [CrossRef]
38. Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. *IEEE Trans. Commun.* **1995**, *43*, 2959–2965. [CrossRef]
39. Cheng, B.; Jin, L.; Li, G. Infrared and visual image fusion using LNSST and an adaptive dual-channel PCNN with triple-linking strength. *Neurocomputing* **2018**, *310*, 135–147. [CrossRef]
40. Roberts, J.; Van Aardt, J.; Ahmed, F. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *J. Appl. Remote Sens.* **2008**, *2*, 023522.