

Perceptual Fusion of Infrared and Visible Image through Variational Multiscale with Guide Filtering

Xin Feng* and Kaiqun Hu*

Abstract

To solve the problem of poor noise suppression capability and frequent loss of edge contour and detailed information in current fusion methods, an infrared and visible light image fusion method based on variational multiscale decomposition is proposed. Firstly, the fused images are separately processed through variational multiscale decomposition to obtain texture components and structural components. The method of guided filter is used to carry out the fusion of the texture components of the fused image. In the structural component fusion, a method is proposed to measure the fused weights with phase consistency, sharpness, and brightness comprehensive information. Finally, the texture components of the two images are fused. The structure components are added to obtain the final fused image. The experimental results show that the proposed method displays very good noise robustness, and it also helps realize better fusion quality.

Keywords

Image Fusion, Guided Filter, Phase Consistency, Variational Multiscale Decomposition

1. Introduction

Infrared sensors are important to the reflection of real scenes. The resulting images are featured by low resolution and low signal-to-noise ratio. The details of the scene under certain conditions can be reflected clearly using visible light sensor, and the imaging is susceptible to the natural conditions such as illumination and weather [1].

Currently, infrared and visible light image fusion is based on various improved multi-scale analysis methods. Meng et al. [2] proposed an infrared and visible image fusion based on object region detection and the non-subsampled contourlet transform (NSCT), and by using the model that had proposed, objects and be clarified and details and visual artifacts in the fused image can be preserved. Zhu et al. [3] proposed an infrared and visible image fusion method based on an improved multi-scale top-hat transform model; this model highlights the target of infrared images and preserves the details of visible images, and its performance is better than certain conventional multi-scale transform image methods in this respect. Liu et al. [4] proposed a multi-scale geometric analysis combined with image fusion method in sparse representation (SR), and they obtained better results than that of solely SR or multi-scale geometric analysis. Zhang et al. [5] proposed a multi-exposure image fusion algorithm based on time-

* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received December 27, 2018; first revision July 26, 2019; accepted September 26, 2019.

Corresponding Author: Xin Feng (149495263@qq.com)

* College of Mechanical Engineering, Chongqing Technology and Business University, Chongqing, China; and Key Laboratory of Manufacturing Equipment Mechanism Design and Control of Chongqing, Chongqing, China ({149495263, 24507586}@qq.com)

domain median filtering. Although the traditional multi-scale method has certain noise robustness, the performance will decrease linearly when the noise is serious. In general, the structural components of infrared and visible light images mainly depict the basic contour structure of the image, and are most susceptible to noise interference. If the fusion rule is capable of filtering and maintaining the edge of the image, a clear structural fusion result can be obtained; and the texture components can reduce the effect of noise on the fusion results on infrared and visible light images by discarding the texture components on the smallest scale.

Based on this, this paper proposes an infrared and visible light image fusion method based on variational multi-scale decomposition. After the infrared and visible light images are decomposed to obtain structural components and texture components, the structural components are subject to fusion processing by means of guided filtering; After the texture components on the minimum two-layer scales are discarded, the phase-consistency, sharpness, and brightness synthesis information are adopted to integrate the weights of the fusion weights for the fusion of texture components; finally, the structural and textural component information are added together to obtain the fusion results. It can effectively alleviate the influence of noise in the fusion process by adopting the methods, while maintaining a clear edge contour structure, and excellent filtering effect is realized. The fusion method block diagram is shown in Fig. 1.

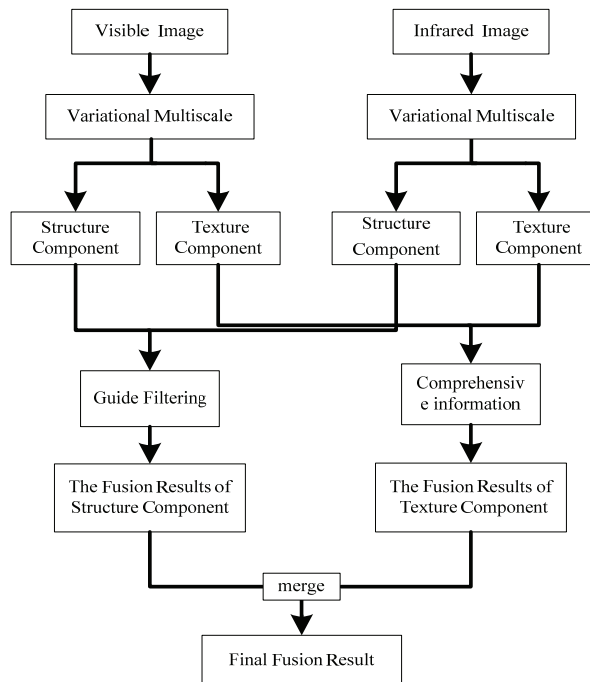


Fig. 1 .The block diagram of this paper.

2. Variational Multi-scale Decomposition

Variational multiscale decomposition is frequently adopted to recover real images from noise images. It is an important research content in the domain of image processing research. It can be seen as a

decomposition of the image f . $f=u+v$ denotes that the former mainly represents the image structure component, and the latter mainly represents the texture component of the image [6]. For the input infrared light image f^A with the initial scale $\lambda=0.0005$, the following energy functional is defined:

$$(f_u^A, f_v^A) = \operatorname{arginf} \left\{ \begin{aligned} &E_\lambda(f^A, \lambda; f_u^A, f_v^A) = \\ &\|f_u^A\| + \lambda \|f^A - f_u^A - f_v^A\| + \|f_v^A\|_p \\ &\int_\Omega f_u^A = \int_\Omega f^A \end{aligned} \right\} \quad (1)$$

where λ denotes the initial decomposition scale. $f_{u_\lambda}^A$ denotes the structural component of the infrared light image at the scale λ , and $f_{v_\lambda}^A$ denotes the texture component of the infrared light image at the scale λ . $\operatorname{arginf}(\cdot)$ is the independent variable corresponding to the maximum likelihood function, f_u^A is the structural component of the input infrared light image, f_v^A is the texture component of the input infrared light image, and the residual of the image can be expressed as:

$$f_r^A = f^A - f_u^A - f_v^A \quad (2)$$

According to the above rules, constantly decomposition is performed to establish a multi-scale decomposition form, and obtain the structural components, texture components, and residual components of the final infrared light image:

$$f^A = f_{u_1}^A + f_{u_2}^A + f_{u_3}^A + \dots + f_{u_m}^A + f_{v_1}^A + f_{v_2}^A + f_{v_3}^A + \dots + f_{v_m}^A + f_{r_m}^A \quad (3)$$

Among them, m denotes the number of decomposition layers, $f_{u_m}^A$ and $f_{v_m}^A$ are the structural components and texture components of the m -layer decomposition of the visible light image, and $f_{r_m}^A$ is the residual difference component of the input infrared image after m -layer decomposition. Similarly, by taking the above steps for the visible light image, the corresponding structural component $f_{u_m}^B$, texture component $f_{v_m}^B$, and residual difference component $f_{r_m}^B$ of the visible light image can be obtained:

$$f^B = f_{u_0}^B + f_{u_1}^B + f_{u_2}^B + \dots + f_{u_m}^B + f_{v_0}^B + f_{v_1}^B + f_{v_2}^B + \dots + f_{v_m}^B + f_{r_m}^B \quad (4)$$

By discarding the texture components at the lowest two scales, the effect of noise on the fused image is reduced. The overall structural components f_u^A and f_u^B of infrared and visible light images are

$$f_u^A = \sum_{i=0}^m f_{u_i}^A \text{ and } f_u^B = \sum_{i=0}^m f_{u_i}^B, \text{ respectively, and the overall texture components } f_v^A \text{ and } f_v^B \text{ are}$$

$$f_v^A = \sum_{i=2}^m f_{v_i}^A \text{ and } f_v^B = \sum_{i=2}^m f_{v_i}^B, \text{ respectively.}$$

3. Structural Component Fusion

Guided filtering refers to a filtering algorithm based on the local linear model. It was proposed by He et al. [7] in 2013. It is a local multi-point filter and an effective boundary-preserving filter. Its main

advantage including its function of maintaining the edge of the image while filtering, and the algorithm complexity is linear and efficient. Therefore, the fusion rule is an ideal choice for the setting of the structural components.

It is assumed that, in a window $(2r+1) \times (2r+1)$ of size ω , the output O of the filter is always represented as the linear transformation of the pilot image I :

$$O_i = a_j I_i + b_j, \forall i \in \omega_j \quad (5)$$

In the fusion process, the structural components on different scales are respectively fused to obtain the fused structural components at different scales. f_u^A and f_u^B are the structural components of infrared and visible light images at a certain scale, and then f_u^A and f_u^B are compared based on each pixel to obtain weight maps w_1 and w_2 :

$$w_1 = \begin{cases} 0, & \text{otherwise} \\ 1, & \text{if } f_u^A > f_u^B \end{cases} \quad (6)$$

$$w_2 = \begin{cases} 0, & \text{otherwise} \\ 1, & \text{if } f_u^B > f_u^A \end{cases} \quad (7)$$

The structural component is obtained by through the calculation and the noise and the edges are not uniform in this respect. Then, guided filtering is performed on the weighted graphs w_1 and w_2 , and f_u^A and f_u^B are respectively acting as the infrared and visible guiding images to obtain new weighted graphs m_1 and m_2 :

$$m_1 = G_{r,\omega}(w_1, f_u^A) \quad (8)$$

$$m_2 = G_{r,\omega}(w_2, f_u^B) \quad (9)$$

And the filter core is expressed as:

$$G_{r,s} = \frac{1}{|\Omega|^2} \sum_{k:(i,j) \in \Omega_k} \left[1 + \frac{(G_i - \mu_k)(G_j - \mu_k)}{\sigma_k^2 + \zeta} \right] \quad (10)$$

where Ω is the radius of the pilot filter and is set to $r=3$, which ζ is the regularized adjustment parameter of the pilot filter and is set to $\zeta=10^{-6}$, σ_k^2 and μ_k represent the average and variance of the guide image within the radius respectively. With the new weight graphs m_1 and m_2 , the fused structure component f_u is obtained:

$$f_u = f_u^A \times m_1 + f_u^B \times m_2 \quad (11)$$

4. Texture Component Fusion

The texture component includes a large amount of texture information and detail information in the image. Considering the relationship between the intensity of the current subband coefficient and the

different subband coefficients in the same area, phase coherence information is adopted on the subbands of the infrared light and visible light texture components. Definition and image brightness information and other comprehensive information are adopted to calculate the texture component information of the image to be fused.

First of all, for the infrared and visible light component subbands, the phase consistency information is obtained as:

$$M_A^m(x, y) = \frac{\sum_k E_{A,\theta}^m(x, y)}{\varepsilon + \sum_n \sum_k S_{A,n,\theta}^m(x, y)} \quad (12)$$

$$M_B^m(x, y) = \frac{\sum_k E_{B,\theta}^m(x, y)}{\varepsilon + \sum_n \sum_k S_{B,n,\theta}^m(x, y)} \quad (13)$$

where m is the decomposition scale of the subband of the texture component, and $M_A^m(x, y)$, $M_B^m(x, y)$ are the phase coherence information of infrared and visible light images on the m -scale at the point (x, y) ; $E_{A,\theta}^m(x, y)$ and $E_{B,\theta}^m(x, y)$ denote the local energy information of the Fourier domain in the m -scale and θ -direction angles of the texture component sub-bands of the infrared and visible light images at the point (x, y) ; $S_{A,n,\theta}^m(x, y)$ and $S_{B,n,\theta}^m(x, y)$ denote the local amplitude information of the Fourier region of the infrared component and the visible light image at the point (x, y) in the m -scale and θ -direction angular subbands; ε is a very small positive value, which is adopted in case the denominator is zero. The sliding window is set as ϕ , and it is assumed that the center of the window is (x, y) , when the image is traversed to an arbitrary position. The definition information of the sub-band of the texture component of the visible light and infrared light image in the window is calculated as:

$$C_A^m(x_0, y_0) = [I_{h,A}^m(x, y) - I_{h,A}^m(x_0, y_0)]^2 \quad (x_0, y_0) \in \phi \quad (14)$$

$$C_B^m(x_0, y_0) = [I_{h,B}^m(x, y) - I_{h,B}^m(x_0, y_0)]^2 \quad (x_0, y_0) \in \phi \quad (15)$$

$I_{h,A}^m(x, y)$ and $I_{h,B}^m(x, y)$ are the texture component subbands of the m -scale at the point (x, y) of the infrared and visible light images respectively. $C_A^m(x_0, y_0)$ and $C_B^m(x_0, y_0)$ denote the amount of definition information of the texture component sub-bands at the m -scale dimension of the pixel point (x_0, y_0) respectively. At the same time, (x_0, y_0) is an arbitrary point within the sliding window ϕ . In this paper, according to the experience ϕ , the value is 11×11 .

Through the comprehensive analysis of the phase consistency and the factors such as definition and image brightness, the rules for subband fusion of the texture components are obtained as follows. The activity of the subband of the texture component is defined as:

$$H_A^m(x, y) = \sum_{(x,y) \in \phi} [M_A^m(x_0, y_0)]^{\alpha_1} \times [C_A^m(x_0, y_0)]^{\beta_1} \times |I_A^m(x, y)|^{\gamma_1} \quad (16)$$

$$H_B^m(x, y) = \sum_{(x,y) \in \phi} [M_B^m(x_0, y_0)]^{\alpha_1} \times [C_B^m(x_0, y_0)]^{\beta_1} \times |I_B^m(x, y)|^{\gamma_1} \quad (17)$$

$H_A^m(x, y)$ and $H_B^m(x, y)$ denote the active level of the texture component of the infrared and visible light

images on the m-scale, and $|I_A^m(x, y)|^{\gamma_1}$ and $|I_B^m(x, y)|^{\gamma_1}$ denote the brightness of the subbands of the texture component. a_1 , β_1 and γ_1 denote the weight information of phase consistency information, sharpness, and image brightness information respectively. Based on the experience values, the weight coefficient in experiment were 0.001, 1, and 1, respectively.

Then the weight of the fused texture components of infrared and visible light images is calculated:

$$Q_A^m(x, y) = \begin{cases} 0, & \text{otherwise} \\ 1, & Q_{S_A^{m,p}}(x, y) \geq \frac{x \times y}{2} \end{cases} \quad (18)$$

$$Q_B^m(x, y) = \begin{cases} 0, & \text{otherwise} \\ 1, & Q_{S_B^{m,p}}(x, y) \geq \frac{x \times y}{2} \end{cases} \quad (19)$$

$Q_{S_A^{m,p}}(x, y)$ and $Q_{S_B^{m,p}}(x, y)$ are the number of active pixels in the sliding window of the infrared and visible light images at the point (x, y) m-scale, where $x \times y$ is the size of the sliding window, and the expression of $Q_{S_A^{m,p}}(x, y)$ and $Q_{S_B^{m,p}}(x, y)$ are:

$$Q_{S_A^{m,p}}(x, y) = \begin{cases} (x_0, y_0) \in \phi | H_A^{m,p}(x_0, y_0) \geq \\ \max\{H_B^{m,p}(x_0, y_0), H_A^{m,p}(x_0, y_0)\} \end{cases} \quad (20)$$

$$Q_{S_B^{m,p}}(x, y) = \begin{cases} (x_0, y_0) \in \phi | H_B^{m,p}(x_0, y_0) \geq \\ \max\{H_B^{m,p}(x_0, y_0), H_A^{m,p}(x_0, y_0)\} \end{cases} \quad (21)$$

Finally, texture component fusion is performed:

$$I_{H,F}^m(x, y) = I_{H,A}^m(x, y) \times Q_A(x, y) + I_{H,B}^m(x, y) \times Q_B(x, y) \quad (22)$$

5. Fusion Results

In this experiment, the simulation experiment platform uses Intel Xeon CPU E3-1231 v3 @ 3.40 G 3.40 GHZ and its memory is 16 G, with PC Win7 and programming language MATLAB2010a. Due to the limited length of the paper, only a set of infrared visible images were adopted to verify the experimental results. At the same time, Gaussian white noise was artificially added to all images. The standard deviation was 10, 20. It can naturally introduce the denoising process into the fusion process, and has somehow denoising capability by adopting the transform domain method. Therefore, in this paper, it is mainly compared with the transform domain fusion method. Three representative methods for transform domain fusion are compared with the proposed method. That is, NSCT fusion method, SR fusion method, Shearlet transform (ST) fusion method, respectively. Among them, discrete wavelet transform (DWT), NSCT, ST decomposition layer is set to 4 layers, and wavelet base db4 is used in DWT method. In NSCT method, the direction filter is set to "vk", decomposition filter is set to "pyrex", and the direction of 4 layers decomposition are 4, 8, 8, and 16, respectively. The SR method image block size is set to 8×8 , the reconstruction error is set to 0.1, k-singular value decomposition (K-SVD) is used for training dictionary, and the dictionary size is 256. In terms of fusion rules, NSCT method, RT method, and ST

method integrate the fusion rules in literature [8,9] and [10], respectively.

Fig. 2 is the result of the experiment in this paper when the noise standard deviation is 10. Fig. 2(a) and 2(b) are the original images of the third group of samples to be fused. Fig. 2(c) and 2(d) are schematic diagrams of the third group of samples to be fused after adding 10 standard deviation Gaussian white noise; Fig. 2(e)–2(i) are the four fusion methods for the purpose of comparison, and the fusion of the paper as a result of the method. It can be seen that due to the small amount of added noise, the fusion results of several methods are relatively ideal, as they are basically capable of maintaining the edges and details, and all of them have displayed certain capacity of noise suppression. The NSCT method, the Shearlet method, and the method mentioned in this paper are relatively smooth and have displayed relatively better visual effects.

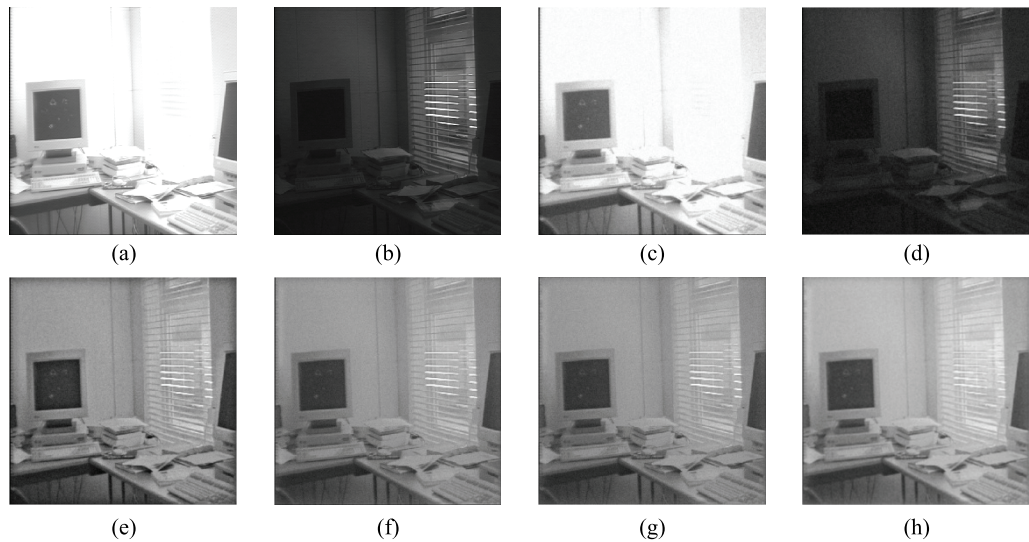


Fig. 2. Comparison of the third group of fusion results (noise standard deviation is 10): (a) original visible light image, (b) original infrared light image, (c) visible light and noise image, (d) infrared light and noise image, (e) NSCT method fusion result, (f) RT method fusion result, (g) Shearlet method fusion result, (h) fusion results of proposed method in this paper.

In order to highlight the noise suppression effect of this method, the standard deviation of Gaussian white noise added is increased to 20 and the others remain unchanged as shown in Fig. 3. Since the serious noises can be effectively suppressed by adopting most of existing methods for multi-scale fusion methods, the guide filtering is adopted in the method to set the fusion rules on the structural components, which effectively removes noise pollution at the edges, and discards the lowest two in the texture component. The texture components on the scale have shown the best noise suppression performance compared to the other types. It can be seen that the fusion results of the DWT method, the ST method and the NSCT method are very poor, and the noise particles have been heavily contaminated to the structural edge and texture information of the fusion result. The fusion of the method in this paper is obviously outperform other methods for transform domain fusion. It can effectively suppress the noise while retaining the edge and texture information and has the best visual effect. And after the partial amplification of the fusion results of the above four methods can be clearly seen (as shown in Fig. 4), and the image obtained through this method looks significantly smoother than that obtained using other methods, and edge and texture

information is much better. In summary, the visual effect of this method is better than the traditional transform domain method in the presence of noise pollution, especially when the noise is greater than 20 standard deviations, namely, the fusion method of this paper is more prominent when the original image is seriously polluted by noise.

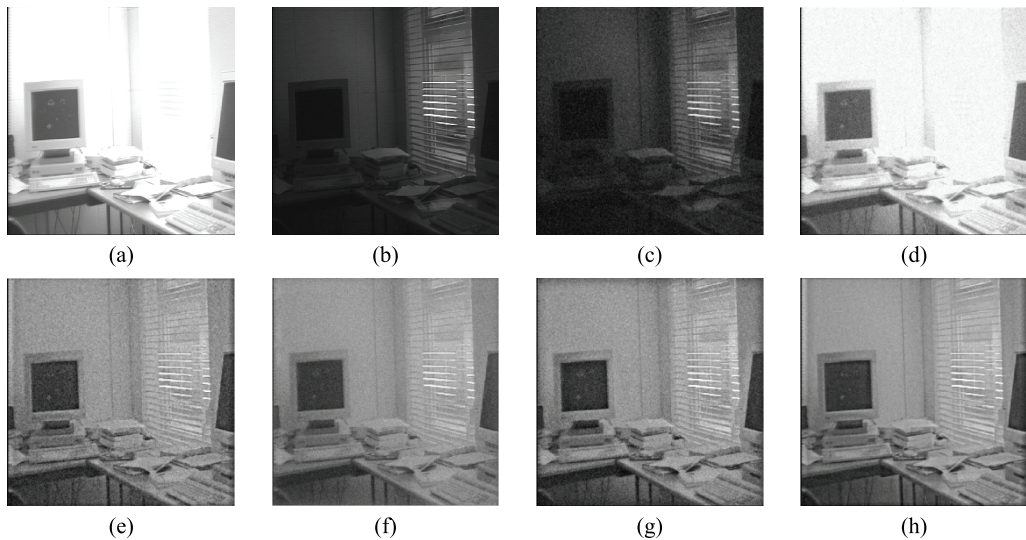


Fig. 3. Comparison of the third group of fusion results (noise standard deviation is 20): (a) original visible light image, (b) original infrared light image, (c) visible light and noise image, (d) infrared light and noise image, (e) NSCT method fusion result, (f) RT method fusion result, (g) Shearlet method fusion result, and (h) fusion results of proposed method in this paper.

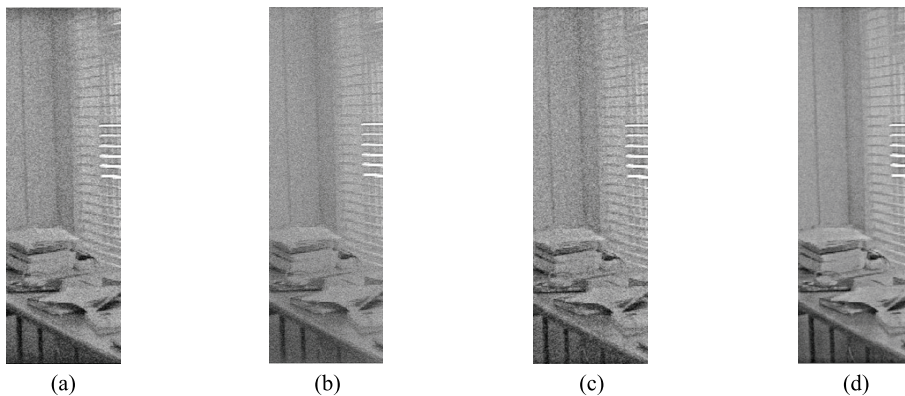


Fig. 4. The partially enlarged views of the fusion results in Fig. 3.

In terms of the objective evaluation indicators, five common fusion indicators are adopted to evaluate the objective quality of various fusion methods. In the aspect of the information theory, the mutual information index is adopted, and the gradient-based fusion index is adopted for the image feature. As for the structure, the fusion index based on structural similarity is adopted and the index in human visual sensitivity is also used. The objective evaluation results of the quality of the three sets of image fusion results are shown in Table 1.

Table 1. Fusion evaluation index

Fusion method	Q_{MI}	Q_G	Q_Y	Q_{CB}	Time (s)
Noise standard deviation is 10					
NSCT	0.791	0.679	0.839	0.769	54
RT	0.780	0.680	0.856	0.799	98
ST	0.699	0.690	0.817	0.776	43
This paper	0.743	0.685	0.829	0.756	34
Noise standard deviation is 20					
NSCT	0.597	0.499	0.638	0.568	78
RT	0.569	0.548	0.667	0.594	116
ST	0.540	0.498	0.624	0.574	59
This paper	0.695	0.640	0.791	0.656	61

From the above Table 1, we can see that when 10 standard deviation Gaussian white noise is added, this method has shown a few advantages in aspect of fusion index comparison, followed by RT fusion method and NSCT fusion method. However, when the noise is increased to 20 standard deviations, the advantages of the method become obvious in terms of gradient, structural similarity, visual sensitivity, and mutual information indicators. The merits of the method is obvious for the traditional transform domain when dealing with noise-contaminated original images. Meanwhile, as the complexity of the pilot filter algorithm is linear, regardless of the size of the set window, it has high efficiency. In terms of subjective and objective synthesis, this method can be applied to the fusion of noisy infrared and visible light images, and it has certain advantages over traditional transform domain fusion methods.

6. Conclusions

In this paper, an infrared and visible light image fusion method based on variational multiscale decomposition is proposed. Guided filtering was used to fuse the structural components. The fusion weight rules were constructed using the phase consistency, definition, and brightness synthesis information on the texture components. Compared with the traditional methods for infrared and visible image fusion, it not only effectively overcomes the noise interference in the fusion process, but also obtains better texture details while maintaining the edge structure, and displays certain subjective and objective quality by adopting the method. However, due to the complexity of the guided filtering algorithm, it takes a long time if the algorithm is used. In the case of high real-time demand, the algorithm is difficult to meet the requirements, so the next study needs to further improve the computational efficiency of the algorithm.

Acknowledgement

The paper is supported by National Natural Science Foundation of China (No. 31501229, 61861025), Chongqing Nature Science Foundation for Fundamental Science and Frontier Technologies (No. cstc2018jcyjAX0483), Key Laboratory of Chongqing Technology and Business University (No. KFJJ2019076), and project of science and technology research program of Chongqing Education Commission of China (No. KJQN201900821).

References

- [1] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: a survey," *Information Fusion*, vol. 45, pp. 153-178, 2019.
- [2] F. Meng, M. Song, B. Guo, R. Shi, and D. Shan, "Image fusion based on object region detection and non-subsampled contourlet transform," *Computers & Electrical Engineering*, vol. 62, pp. 375-383, 2017.
- [3] P. Zhu, X. Ma, and Z. Huang, "Fusion of infrared-visible images using improved multi-scale top-hat transform and suitable fusion rules," *Infrared Physics & Technology*, vol. 81, pp. 282-295, 2017.
- [4] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Information Fusion*, vol. 36, pp. 191-207, 2017.
- [5] B. Zhang, X. Lu, H. Pei, H. Liu, Y. Zhao, and W. Zhou, "Multi-focus image fusion algorithm based on focused region extraction," *Neurocomputing*, vol. 174, pp. 733-748, 2016.
- [6] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Information Fusion*, vol. 31, pp. 100-109, 2016.
- [7] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397-1409, 2013.
- [8] Q. Zhang and B. L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Processing*, vol. 89, no. 7, pp. 1334-1346, 2009.
- [9] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884-892, 2010.
- [10] Q. G. Miao, C. Shi, P. F. Xu, M. Yang, and Y. B. Shi, "A novel algorithm of image fusion using shearlets," *Optics Communications*, vol. 284, no. 6, pp. 1540-1547, 2011.



Xin Feng <https://orcid.org/0000-0001-8793-3775>

He received Ph.D. degrees in School of Computer Science and Engineering from Lanzhou University of Technology in 2012. Since March 2013, he has been teaching in the School of Chongqing Technology and Business University.



Kaiqun Hu <https://orcid.org/0000-0002-5590-2584>

She received her Ph.D. degree from China Agriculture University in 2011. Now she is a lecturer in Chongqing Technology and Business University. Her main research direction is the technology of computer control and image processing.