A novel visible-infrared image fusion method based on visual enhancement and multiscale decomposition

Lingxiao Li¹*, Yong Feng² and Zezhong Ma³

¹Faculty of Science, Chongqing University of Technology, Chongqing, 40000, China
²Faculty of Computer Science, Chongqing University, Chongqing, 40000, China
³Faculty of Environment and Resources, Chongqing Technology and Business University, Chongqing, 40000, China
*Corresponding author’s e-mail: lilingxiao@cqut.edu.cn

Abstract. Aiming at the problems of low contrast, low signal-to-noise ratio and scattered energy in the field of optical detection and imaging, a visual enhancement and multi-scale decomposition method for fusion of visible and infrared images is proposed based on the visual characteristics of different bands’ image. Firstly, the infrared image with less texture information and low contrast is preprocessed to weaken the background noise and improve the visual contrast. On this basis, the multi-scale image is decomposed by detail preserving filter, and the saliency map of each scale image is obtained by saliency extraction method. Then, the multi-source images of each scale are fused with saliency map, and binary filtering fusion rules are adopted for the regions with salient details while weighted fusion rules are adopted for other regions. Finally, the images of all scale are reconstructed to get better fusion results. The experimental results show that this method can significantly improve the visual contrast of the fused object, and the objective evaluation indexes are superior to other comparison methods.

1. Introduction

Among various image fusion strategies, the most widely used is visible-infrared image fusion, which has obvious advantages in many aspects [1]. Visible infrared image fusion algorithms can be divided into six categories according to their different schemes and principles, respectively corresponding to multi-scale transformation method [2], sparse representation method [3], neural network method [4], sub-band space method [5], hybrid model method [6] and saliency based visual fusion method [7]. The multi-scale transformation method is the most common method in image fusion. It decomposes the input visible and infrared image into sub-images of different layers, and uses a certain weighting method to reconstruct the corresponding sub-images of each layer. The common image decomposition and reconstruction methods include wavelet transform [8], image pyramid [9], curvelet transform [10] and other improved methods. Sparse representation method describes the image through a series of over-complete dictionaries with sparse features through linear combination, which is the key to their good performance. Neural network method establishes image fusion model by simulating human brain's perception, which has good adaptability, fault tolerance and anti-noise ability. The basic idea of the sub-band space method is similar to the neural network method, but it uses a complete basis instead of an over complete basis to analyse the image in the spatial domain. In this kind of methods, principal component analysis (PCA), non-negative matrix factorization (NMF) and independent component analysis (ICF) are most used. The saliency vision fusion method considers that compared with the
background and flat areas in the image, human visual attention is usually more easily attracted by the
target or those image areas with rich details. Therefore, we can get the fusion image with better visual
effect through searching these regions of interest in the image. Hybrid model method combines the
advantages of the above methods to further improve the image fusion effect, which also uses some other
theoretical tools, such as total variation model, fuzzy theory and information entropy.

In this paper, a visible infrared image fusion method based on target visual enhancement is proposed
to achieve better visual fusion effect and higher target contrast. Rest of the paper is arranged as follows:
Firstly, Section 2 covers the specific description of proposed image fusion method; Secondly, Section 3
contains both experiment and analysis of the algorithm’s performance. Finally, the conclusion of the
entire work is given in Section 4.

2. Proposed method

2.1. Pre-processing of infrared images
In order to improve the contrast of infrared image and highlight the image details, we first use the image
defogging method [11] based on atmospheric scattering model to enhance the infrared image. The
preprocessed infrared image is shown in Fig.1 below. It can be seen that the enhanced image has higher
energy intensity, larger overall contrast and richer details, which is beneficial to subsequent fusion
processing.

2.2. Image saliency extraction
Image saliency extraction technology based on the human visual attention mechanism has now become
an important tool for various image analysis and processing. Among these saliency extraction methods,
the frequency tuning method [10] has been widely used because it can better reflect the region of interest
of the human eyes. In order to find the region of interest in each layer of the multi-scale decomposition,
we use the DOG operator to band-pass filter the image. By setting the standard deviation of two Gaussian
functions reasonably, the high-frequency noise and texture in the image could be filtered and the saliency
map \( S \) will be calculated as \( S(x, y) = |I_\mu - I_{ahc}(x, y)| \), Where \( I_\mu \) is the mean value of input image, \( I_{ahc} \)
is the smoothed result of Gaussian filter. The saliency map of each band of image can be obtained.
Visible and infrared image of the same scene and its corresponding saliency map are shown in Fig. 2,
respectively. In the areas of human visual interest, the image contrast has been enhanced, which makes
the image information more abundant.

Figure 1 Visual enhancement result of infrared image

Figure 2 Extraction of saliency map: (a) Infrared image; (b) Visible image; (c) Infrared saliency map;
(d) Visible saliency map
2.3. Multiscale decomposition of visible infrared image

After preprocessing the input image and extracting salient features, according to the image fusion framework of multi-scale decomposition, we use the Rolling Guidance Filter (RGF) to set different smoothing coefficients and get the decomposed image with k-level scale. Suppose the input visible and infrared band images are \( I_{\text{band1}} \) and \( I_{\text{band2}} \), the smooth image corresponding to the \( l \)-th scale could be described as:

\[
U_{l\text{band1}}^l = RGF(U_{l\text{band1}}^{l-1}, \sigma_{s}^{l-1}, \sigma_{r}, T), \quad l = 1, 2, \ldots, K
\]
\[
U_{l\text{band2}}^l = RGF(U_{l\text{band2}}^{l-1}, \sigma_{s}^{l-1}, \sigma_{r}, T), \quad l = 1, 2, \ldots, K
\]
\[
U_{0\text{band1}}^0 = I_{\text{band1}}, \quad U_{0\text{band2}}^0 = I_{\text{band2}}
\] (1)

Where \( \sigma_{r} \) is the guided filter weight coefficient, the value is set as 0.1; \( T \) is iterations and the value is set as 4. As for Gaussian smoothing parameter \( \sigma_{s} \), its value increases as the number of layers expands, and the corresponding expression is \( \sigma_{s}^{l+1} = 2 \sigma_{s}^{l} \). We can see the filtering result of the upper layer has more structure information than the lower layer. When the multi-scale decomposition reaches the last layer, the corresponding filtering result \( U^{K} \) is called the base layer and counted as \( B \). The corresponding detail layer image can be obtained by subtracting two adjacent layers of images, and the expression is described as:

\[
D_{l\text{band1}}^l = U_{l\text{band1}}^{l+1} - U_{l\text{band1}}^l, \quad l = 1, 2, \ldots, K.
\]
\[
D_{l\text{band2}}^l = U_{l\text{band2}}^{l+1} - U_{l\text{band2}}^l, \quad l = 1, 2, \ldots, K
\] (2)

Where \( D_{l\text{band1}}^l \) and \( D_{l\text{band2}}^l \) represent the detailed image corresponding to the current scale, the expression can be described as \( D_{l\text{band1}}^{l+1} = U_{l\text{band1}}^{K} \), \( D_{l\text{band2}}^{l+1} = U_{l\text{band2}}^{K} \). The reconstruction of original input image can be obtained by summing the accumulated image of each detail layer and the base image, finally the corresponding expression could be written as:

\[
I_{\text{band1}} = U_{\text{band1}}^{K} + \sum_{l=1}^{K} D_{l\text{band1}}^l = \sum_{l=1}^{K+1} D_{l\text{band1}}^l
\]
\[
I_{\text{band2}} = U_{\text{band2}}^{K} + \sum_{l=1}^{K} D_{l\text{band2}}^l = \sum_{l=1}^{K+1} D_{l\text{band2}}^l
\] (3)

2.4. Detail preserving fusion based on target detection

For each scale of the visible and infrared image to be fused, the traditional fusion methods only linear weighted the image. When the environment radiation is weak or the signal-to-noise of the target is low, this fusion strategy may lead to the phenomenon of scene distortion or even target submerged, which is not conducive to visual observation and subsequent detection. In order to increase the background information and not weaken the visual effect of target, we improve and design a detail preserving fusion strategy based on target detection. Using the Prewitt edge detection operator to calculate the gradients of saliency map \( S_{l\text{band1}}^l \) and \( S_{l\text{band2}}^l \), and obtain the gradient values \( G_{x} \) and \( G_{y} \) in the horizontal and vertical directions corresponding to each pixel. The maximum value of the gradient in the horizontal and vertical directions is taken as the gradient amplitude, which can be described as:

\[
G_{l\text{band1}}^l(i, j) = \max \{ G_{x_{l\text{band1}}}^l(i, j), G_{y_{l\text{band1}}}^l(i, j) \}
\]
\[
G_{l\text{band2}}^l(i, j) = \max \{ G_{x_{l\text{band2}}}^l(i, j), G_{y_{l\text{band2}}}^l(i, j) \}
\] (4)

Select an appropriate threshold \( T \), and perform binary segmentation according to the calculated value of the gradient in the saliency map to obtain the corresponding binary saliency distribution map. The expression is written as:
\[\begin{align*}
P'_{\text{band}1}(i, j) &= \begin{cases} 1, G'_{\text{band}1}(i, j) > T'_{\text{band}1} \\ 0, G'_{\text{band}1}(i, j) \leq T'_{\text{band}1} \end{cases} \\
0, G'_{\text{band}2}(i, j) \leq T'_{\text{band}2} \\
1, G'_{\text{band}2}(i, j) > T'_{\text{band}2} \\
0, G'_{\text{band}2}(i, j) \leq T'_{\text{band}2} \end{align*}\]

Where \( P'_{\text{band}1} \) and \( P'_{\text{band}2} \) are the binary segmentation results of saliency images in different bands of the \( l \)-th layer, respectively. \( T'_{\text{band}1} \) and \( T'_{\text{band}2} \) are segmentation thresholds, and their size is equal to the mean value of the corresponding saliency map. According to \( P'_{\text{band}1} \) and \( P'_{\text{band}2} \), the traditional fusion strategy can be improved and the corresponding expression is:

\[M_j = \begin{cases} D'_{\text{band}1}, & \text{if } P'_{\text{band}1} = 1 \text{ and } P'_{\text{band}2} = 0 \\ D'_{\text{band}2}, & \text{if } P'_{\text{band}1} = 0 \text{ and } P'_{\text{band}2} = 1 \\ \max \{D'_{\text{band}1}, D'_{\text{band}2}\}, & \text{if } P'_{\text{band}1} = 1 \text{ and } P'_{\text{band}2} = 1 \\ \left\lbrack (D'_{\text{band}1}S'_{\text{band}1} + D'_{\text{band}2}(1 - S'_{\text{band}1})) + (D'_{\text{band}2}S'_{\text{band}2} + D'_{\text{band}1}(1 - S'_{\text{band}2})) \right\rbrack / 2, & \text{others} \end{cases}\]

The improved fusion method uses binary filtering to fuse the areas which detected as salient details, while the weighted fusion method is used for non-salient details that \( P'_{\text{band}1} = 1 \) and \( P'_{\text{band}2} = 0 \), the visible light band image contains a higher amount of information, which may be the target or the edge of significant details, therefore the gray value of the corresponding position in the image is retained. When \( P'_{\text{band}1} = 0 \) and \( P'_{\text{band}2} = 1 \), the gray value of the corresponding infrared band image is preserved. When \( P'_{\text{band}1} = 1 \) and \( P'_{\text{band}2} = 1 \), this indicates that the same location has high information content in both bands. In order to avoid mutual interference, the band image with the largest gray value in the current location is selected. Finally when \( P'_{\text{band}1} = 0 \) and \( P'_{\text{band}2} = 0 \), this indicates that the location has no high information in both bands, and may be a flat background area. At this time, the pixel is fused by weighted fusion to ensure the authenticity of the fused image.

According to the above fusion process of visible and infrared images, the fusion results of each scale \( M_j \) can be obtained, and the final result could be calculated by weighted reconstruction of each scale’s fusion results as follows:

\[M = \alpha_1 M_1 + \ldots + \alpha_K M_K + \alpha_{K+1} M_{K+1} = \sum_{l=1}^{K+1} \alpha_l M_l\]

Where \( M \) is the final fusion image, \( \alpha_l (l = 1, 2, \ldots, K+1) \) is the reconstruction weight coefficient corresponding to the fusion image of different scales. Setting a reasonable fusion weight coefficient is conducive to improving the visual effect of the final result. \( \alpha_l \) is generally set between \([0,1]\).

3. Experiment and analysis

3.1. Visual effect evaluation

In order to show the fusion effect of our algorithm, we select two groups of typical data from different fusion image databases [11] to test the fusion performance, the corresponding names are: (1) "UNcamp" database and (2) "Traffic" database. As a contrast, some other typical image fusion algorithms are used to compare the data fusion effects of the above different scenes, including laplacian pyramid algorithm (LP), wavelet transform based algorithm (Wavelet), non-subsampled contourlet transform based algorithm (NSCT), direction information excitation neural network method (OI-PCNN), hybrid multiscale decomposition (HMSD) and directional discrete cosine transform and principal component analysis (DDCTPCA). For the image fusion comparison of different scenes, the fusion results of the corresponding algorithms are shown in Fig. 3-4. Picture (a) and (b) are the visible image and infrared image corresponding to the same scene respectively; Picture (c) is the result of LP, (d) is the result of Wavelet, (e) is the result of NSCT, (f) is the result of OI-PCNN, (g) is the result of HMSD, (h) is the result of DDCTPCA, and finally (i) is the result of proposed method.
Figure 3 "Uncamp" fusion results: (a) visible image; (b) infrared image; (c) LP result; (d) Wavelet result; (e) NSCT result; (f) OI-PCNN result; (g) HMSD result; (h) DDCTPCA result; (i) Proposed method’s result
3.2. Objective index evaluation

From the above experimental comparison, it can be found the proposed method in this paper has good fusion effect for visible and infrared images in different scenes. The fusion image conforms to the visual characteristics of human eyes, and the image contrast has been greatly improved while retaining the image target information. In order to make a more comprehensive and objective evaluation of the image fusion effect, several objective evaluation indexes are selected to quantitatively calculate and compare the above fusion results. The selected indexes include information entropy (IE), mutual information (MI), spatial frequency (SF) and edge retention $Q^{AB/F}$. In order to intuitively show the performance of each method, the average value of each index in different scenes is taken as the standard, and the corresponding bar chart of each algorithm is drawn as the quantitative evaluation standard in Fig. 5. It can be seen that compared with other methods, the proposed visible-infrared image fusion method in this paper has the best average value of each index, which proves this method has good fusion effect, strong adaptability, and can effectively and stably fuse multi-band images for different scenes.

Figure 4 "Traffic" fusion results: (a) visible image; (b) infrared image; (c) LP result; (d) Wavelet result; (e) NSCT result; (f) OI-PCNN result; (g) HMSC result; (h) DDCTPCA result; (i) Proposed method’s result

Figure 5 The mean value of each objective index: (a) Information Entropy (IE); (b) Mutual Information (MI); (c) Spatial frequency (SF); (d) Edge retention $Q^{AB/F}$
4. Conclusion

Aiming at the multi-source image fusion technology of the optical detection imaging system, this paper proposed a visible-infrared image fusion method based on target vision enhancement. In the fusion process, binary filtering fusion rules are used for the regions with salient details, while weighted fusion rules are used for the regions with non-salient details, in this way each scale image can be reconstructed to obtain the desired fusion results. Through the fusion experiments on two test scenes, the fusion effect is evaluated from both visual effect and objective index, and compared with other methods, the effectiveness of this method is finally verified.

Acknowledgment

This work was supported by Scientific Research Foundation of Chongqing University of Technology (NO. 2020ZDZ002), Guangxi Key Laboratory of Trusted Software (NO. kx202006) and Zhejiang Lab (NO. 2021KE0AB01).

References

[1] Duan C, Wang Z, Xing C, et al. (2021) Infrared and visible image fusion using multi-scale edge-preserving decomposition and multiple saliency features[J]. Optik-International Journal for Light and Electron Optics, 228(1):165-175.
[2] Yang Y, Liu J, Huang S, et al. (2021) Infrared and Visible Image Fusion via Texture Conditional Generative Adversarial Network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 28(99):164-169.
[3] Xing C, Wang M, Dong C, et al. (2020) Using Taylor Expansion and Convolutional Sparse Representation for Image Fusion[J]. Neurocomputing, 402(4): 437-455.
[4] Cheng B, Jin L, Li G. (2018) Infrared and visual image fusion using LNSST and an adaptive dual-channel PCNN with triple-linking strength[J]. Neurocomputing, 3(10):135-147.
[5] Nair R, Singh T. (2021) MAMIF: multimodal adaptive medical image fusion based on B-spline registration and non-subsampled shearlet transform[J]. Multimedia Tools and Applications, 12(4): 79-105.
[6] Vishwakarma A, Bhuyan M K. (2019) Image Fusion Using Adjustable Non-subsampled Shearlet Transform[J]. IEEE Transactions on Instrumentation and Measurement, 68(9):3367-3378.
[7] Yang Y, Zhang Y, Huang S, et al. (2020) Infrared and Visible Image Fusion Using Visual Saliency Sparse Representation and Detail Injection Model[J]. IEEE Transactions on Instrumentation and Measurement, 53(99):1-15.
[8] Ren L, Pan Z, Cao J, et al. (2021) Infrared and visible image fusion based on weighted variance guided filter and image contrast enhancement[J]. Infrared Physics & Technology, 114(2):103-112.
[9] Xin J, Qian J, Shang Y, et al. (2018) Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain[J]. Infrared Physics & Technology, 88(7):1-12.
[10] Zhang S, Li X, Zhang X, et al. (2021) Infrared and visible image fusion based on saliency detection and two-scale transform decomposition[J]. Infrared Physics & Technology, 114(5):73-82.
[11] Nawaz M, Raja G, Qasim M. (2021) A New Approach for Dehazing and Enhancement of Infrared Images[C] In: 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST). Islamabad, pp. 545-551.