Infrared and visible image fusion based on two-scale decomposition and improved saliency detection

Ying An*, Zhao Hao and Zhaohui Huo

Chinese Flight Test Establishment, Xi’an, Shaanxi, 710089, China

*Corresponding author’s e-mail: 201721013@stumail.nwu.edu.cn

Abstract. Aiming at the problems of poor target saliency, loss of background information and time-consuming in image fusion, a fast image fusion algorithm combining two-scale decomposition and improved saliency detection is proposed. Mean filtering is used to decompose the source image into a base layer and a detail layer. The maximum symmetric surround (MSS) saliency detection algorithm is improved to obtain the dim suppressed MSS algorithm. dim suppressed MSS saliency detection and guided filtering is used to generate fusion rules for each layer. The inverse transformation of two-scale decomposition is used for the fusion sub-image of the base layer and the detail layer to obtain the final fusion result. Experimental results show that the algorithm consumes less time and has better fusion quality, which reflects the feasibility of the proposed algorithm.

1. Introduction

The visual image can capture the light reflection of the scene, usually with high resolution and rich detailed texture information, but it is highly dependent on light. The infrared image captures the thermal radiation information of the scene, which can clearly highlight the target and does not depend on the intensity of the light, but usually has low resolution and less detailed texture information. Infrared and visible image fusion can merge the complementary information of the same scene, which is an important branch of image fusion[1].

Many image fusion frameworks have been proposed. And multi-scale decomposition is effectively used in infrared and visible image fusion [2], including methods based on wavelet transform[3], contourlet transform[4] and various pyramid transforms, such as contrast pyramid[5], gradient pyramid[6]. These methods perform multi-scale decomposition on source images to extract image information. However, details of the image are smoothed, so that some information is lost and artifacts are generated in fused image. The non-subsampled contourlet transform can represent images in multiple scales and directions, and has shift invariance, which can produce better fusion effects. However, the multi-scale decomposition method of images has problems such as high computational complexity and poor real-time performance[7].

The core of image fusion based on multi-scale decomposition lies in the selection of transforms and fusion rules[2]. Recently, saliency analysis is adopted to obtain weight maps, which yields effective results in image fusion[8]. Saliency originated from visual uniqueness, unpredictability, rarity, or surprise, is often attributed to variations in image attributes such as color, gradient, edges, and boundaries[9]. One of the problems faced by image fusion is that the prominent target in the source image is not obvious in the fusion result. Therefore, using saliency detection for image fusion has obvious advantages. Achanta et al. proposed Frequency-tuned (FT) Salient Region Detection based on the center-peripheral operator of color features[10]. The maximum symmetric surround saliency
detection (MSS) proposed by Achanta et al. is an improved version of the FT algorithm. The area for calculating the saliency value of a pixel is improved from global to local, thus achieving better results[11]. However, the use of distance calculations makes the MSS algorithm tend to give larger saliency values to the brightest and darkest pixels. This is not in line with human intuitive visual perception, and introduces more artifacts in the fusion result.

Aiming at the above problems, a fast infrared and visible image fusion algorithm combining two-scale decomposition and improved MSS saliency detection is proposed. The mean filter is used to decompose the source image into a base layer and a detail layer, which reduces the time-consuming of the image fusion algorithm. Dim-suppressed MSS (DSMSS) is applied to the generation of fusion rules to improve the fusion quality.

2. Infrared and visible image fusion algorithm based on two-scale decomposition and improved saliency detection

The structure diagram of the infrared and visible image fusion algorithm based on two-scale decomposition and DSMSS saliency detection shown in Figure 1. In order to display clearly, the sub-picture of the detail layer of the structure picture is displayed after adding 150 to the original pixel. The algorithm flow mainly includes: (1) Two-scale decomposition; (2) DSMSS saliency detection; (3) Fusion rules.

2.1. Two-scale decomposition

In order to improve the fusion speed, mean filtering is used to decompose the source image into a base and a detail layer. The two-scale decomposition of infrared and visible images are as follows:

\[
\phi^B_n (x, y) = \phi^B_n (x, y) * \mu(x, y)
\]  

(1)
\[ \phi^B_{ir}(x, y) = \phi_{ir}(x, y) \ast \mu(x, y) \]  

(2)

Where \( \phi^B_{ir} \) and \( \phi^B_{vis} \) represent the base layer corresponding to the source infrared image \( \phi_{ir} \) and visible image \( \phi_{vis} \), \((x, y)\) represent the pixel, the window size of the mean filter is given by \( \mu \) and \( \ast \) represents the convolution operation.

The detail layer of the infrared and visible image is obtained by subtracting the source image and the corresponding base layer, the formulas are as follows:

\[ \phi^D_{ir}(x, y) = \phi_{ir}(x, y) - \phi^B_{ir}(x, y) \]  

(3)

\[ \phi^D_{vis}(x, y) = \phi_{vis}(x, y) - \phi^B_{vis}(x, y) \]  

(4)

Where \( \phi^D_{ir} \) and \( \phi^D_{vis} \) represent the base layer corresponding to the source infrared image \( \phi_{ir} \) and visible image \( \phi_{vis} \).

2.2. DSMSS saliency detection

MSS saliency detection uses symmetrical surround to change the bandwidth of the center surround filter near the image boundary. The algorithm retains the advantages of accuracy, high speed and simplicity. However, because the MSS algorithm calculates the Euclidean distance between the average of pixels in a specific area in the LAB space, both the brightest and darkest areas in the source image will appear in the salient image, which is not consistent with humans. Its application also introduces artifacts in image fusion. In this paper, the MSS saliency detection algorithm is improved, and the dark part of the image is suppressed in the display of the saliency map. Set the saliency value of the negative part of the mean filter minus the Gaussian filter to 0, so that the bright part of the original image will be displayed in the saliency map, and the dark part will not be displayed in the saliency map. Because the pixel value in the grayscale image only represents the brightness, this paper does not convert the image to LAB space. The improved MSS algorithm in this paper is called dim suppressed MSS (DSMSS). The DSMSS significance value of image \( I \) at \((x, y)\) is as follows:

\[ S(x, y) = \begin{cases} \| I_{\mu}(x, y) - I_f(x, y) \|, & \text{if } I_{\mu}(x, y) \geq I_f(x, y) \\ 0, & \text{otherwise} \end{cases} \]  

(5)

Where \( \| - \| \) represent Euclidean distance, \( I_f \) is the Gaussian filter value of the image \( I \) and \( I_{\mu}(x, y) \) is the mean value of the pixel \((x, y)\) in the maximum symmetrical surrounding sub-image which calculation formula is as follows:

\[ I_{\mu}(x, y) = \frac{1}{S_A} \sum_{(i,j)\in A} I(x, y) \]  

(6)

Where \( A \) is a region, and the pixels contained in it are as in formula (7); \( S_A \) is the area of \( A \), and the calculation formula is as in formula (8)

\[ A = \{(i, j) \mid x - x_0 \leq i \leq x + x_0, y - y_0 \leq j \leq y + y_0\} \]  

(7)

\[ S_A = (2x_0 + 1)(2y_0 + 1) \]  

(8)

Where the values of \( x_0 \) and \( y_0 \) are as follows:

\[ x_0 = \min(x, w - x) \]  

(9)

\[ y_0 = \min(y, h - y) \]  

(10)

Where \( w \) and \( h \) are the width and height of the image.
2.3. Fusion rules

The DSMSS saliency detection map $S_{ir}$ and $S_{vis}$ of the source infrared and visible images are obtained by formula (5) - (10). The initial weight map $M$ is generated by comparing the DSMSS saliency detection maps of the two source images:

$$M(x, y) = \begin{cases} 1, & S_{ir}(x, y) \geq S_{vis}(x, y) \\ 0, & S_{ir}(x, y) < S_{vis}(x, y) \end{cases}$$

(11)

Since the initial weight map is a binary map, directly serving as the fusion weight map will cause spatial inconsistencies in the fusion results. Therefore, applying guided filtering to the initial weight map makes the weight transition of the weight map smoother. Use the corresponding source infrared image as the guide map, and use different filtering parameters guide the initial weight map to obtain the fusion weight map of the base layer and the detail layer:

$$M_B = G(I_{ir}, M, r_B, \varepsilon_B)$$

$$M_D = G(I_{ir}, M, r_D, \varepsilon_D)$$

(12) (13)

Where $M_B$ and $M_D$ are the weight maps of the base layer and the detail layer respectively, $G(\bullet)$ represents the guided filtering algorithm, $I_{ir}$ is the source infrared image, $r_B$ and $\varepsilon_B$ are the guided filter parameters for generating the weight map of the base layer, $r_D$ and $\varepsilon_D$ are the guided filter parameters for generating the weight map of the detail layer.

According to the weight map of the corresponding layer, the fusion sub-image of the base layer and the detail layer is obtained by follows:

$$B_f = M_B \phi_{ir}^B + (1 - M_B) \phi_{vis}^B$$

$$D_f = M_D \phi_{ir}^D + (1 - M_D) \phi_{vis}^D$$

(14) (15)

Where $B_f$ and $D_f$ are the fusion sub-images of the base layer and the detail layer, respectively.

The fusion result image $I_f$ is obtained by the inverse transformation of two-scale decomposition, and the formula is as follows:

$$I_f = B_f + D_f$$

(16)

3. Experimental results and analysis

The TNO dataset is a public registered infrared and visible image fusion dataset. To demonstrate the effectiveness of the proposed method, our method is compared with five classical algorithms: adaptive sparse representation (ASR)[12], cross bilateral filter (CBF)[13], convolutional sparse representation (CSR)[14], weighted least square (WLS)[15] and latent low-rank representation (LatLRR)[16]. The experiment runs on the Intel(R) Core (TM) i5-6500 CPU @ 3.20GHz processor and 8G memory desktop. The operating system is 64-bit win10. The programming environment is MATLAB R2016a.

In the two-scale decomposition, the filter window size of the mean filter is $5 \times 5$. In the generation of the fusion rule, the guided filter parameters of the base layer are $r_B=5$, $\varepsilon_B=1$ and the guided filter parameters of the detail layer are $r_D=1$, $\varepsilon_D=0.1$. The subjective qualitative and objective quantitative comparisons of the experimental results of the method in this paper and the other five methods on the five sets of images are carried out. The source image is shown in Figure 2.
3.1. Subjective Evaluation

Figures 3 and 4 are the experimental results of the proposed method and the other five methods on the two sets of images.

It can be seen from Figure 3-4 that the background information of the fusion result of the ASR, CSR and LatLRR algorithms is partially lost, and the target is not sufficiently prominent. The CBF algorithm has spatial inconsistencies. The salient target of WLS performs well in the prominence, but the texture is not clear enough in the background information. The experimental results of the proposed algorithm have richer background texture information and more prominent target information.
3.2. Objective Evaluation

The proposed and comparison method are evaluated by four objective evaluation metrics, including entropy (EN)\cite{17}, standard deviation (SD), mutual information (MI)\cite{18}, and running time. The values of MI and SSIM are the average of the corresponding values of the fused image and the two source images. Table 1 shows the objective evaluation results of two groups of images. The best value is Bolded.

| Source image | Evaluation criteria | Algorithms | Proposed | ASR | CBF | CSR | WLS | LatLRR |
|--------------|---------------------|------------|----------|-----|-----|-----|-----|--------|
| Battle field | EN                  |            | 7.32     | 6.44| 6.55| 6.43| 6.60| 6.62   |
|              | SD                  |            | 44.94    | 25.86| 27.42| 25.75| 28.25| 30.94  |
|              | MI                  |            | 2.20     | 0.77| 0.76| 0.79| 0.79| 0.87   |
|              | Running Time(s)     |            | 0.13     | 134.68| 6.38| 44.15| 0.47| 21.31  |
| Kaptein_1123 | EN                  |            | 7.31     | 6.81| 6.78| 6.80| 6.97| 6.90   |
|              | SD                  |            | 68.41    | 34.50| 33.20| 34.39| 49.31| 37.73  |
|              | MI                  |            | 2.41     | 1.06| 0.95| 1.07| 1.02| 1.02   |
|              | Running Time(s)     |            | 0.32     | 394.59| 18.82| 133.46| 2.8 | 84.91  |

Table 1 shows that the proposed algorithm performs best in information entropy, standard deviation and mutual information. In terms of running time, compared with other algorithms, the proposed algorithm has the shortest fusion time and can better meet the requirements of real-time fusion environment. Consistent with subjective visual analysis, it shows that the fusion image of the proposed algorithm has higher image quality than other comparison algorithms.

4. Conclusion

Aiming at the problem that infrared and visible image fusion algorithm is too time-consuming to be applied in real-time fusion environment, this paper proposes an infrared and visible image fusion algorithm combining two-scale decomposition and saliency detection. The proposed algorithm uses mean filtering to decompose the source image at two scales. Compared with the traditional multi-scale image, the decomposition time is greatly shortened. In addition, the MSS algorithm is improved and applied to the fusion rule generation of the base layer and the detail layer, which better preserves the salient target in the source infrared image and the rich background texture information in the source visible image. From the perspective of human vision, the fusion result of the proposed algorithm introduces fewer artifacts, has richer background information and more prominent targets. From the perspective of objective quantitative evaluation, compared with comparative experiments, the proposed algorithm has higher information entropy, standard deviation and mutual information, and the proposed algorithm consumes less time.

References

[1] Li S, Kang X, Fang L, Yin H. (2017) Pixel-level image fusion: A survey of the state of the art. Information Fusion, 33:100-112.
[2] Ma J, Ma Y, Li C. (2019) Infrared and visible image fusion methods and applications: A survey. Information Fusion, 45(02): 153–178.

[3] Pajares G, Jesús Manuel de la Cruz. (2004) A wavelet-based image fusion tutorial. Pattern Recognition, 37(9):1855-1872.

[4] Qu X B, Xie G F, Yan J W, Zhu ZQ, Chen BG. (2008) image fusion algorithm based on neighbors and cousins information in nonsampled contourlet transform domain. In: International Conference on Wavelet Analysis & Pattern Recognition.

[5] Li M, Dong Y. (2014) Image fusion algorithm based on contrast pyramid and application. In: International Conference on Mechatronic Sciences.

[6] Qu X J, Zhang F, Zhang Y. (2013) Feature-Level Fusion of Dual-Band Infrared Images Based on Gradient Pyramid Decomposition. Applied Mechanics & Materials, 347-350:2380-2384.

[7] Wu C, Chen L. (2020) Infrared and visible image fusion method of dual NSCT and PCNN. PLoS ONE, 15(9): e0239535.

[8] ZHANG Y, WEI W, YUAN Y. (2019) Multi-focus image fusion with alternating guided filtering. Signal, Image and Video Processing, 13(4): 727–735.

[9] Cheng M M , Mitra N J , Huang X , Torr P.H.S. Hu S. (2015) Global Contrast Based Salient Region Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3):569-582.

[10] Achanta R, Hemami S, Estrada F, Süsstrunk S. (2009) Frequency-tuned Salient Region Detection. IEEE Conference on Computer Vision & Pattern Recognition. In: IEEE.

[11] Achanta R, Sabine Süsstrunk. (2010) Saliency detection using maximum symmetric surround. In: International Conference on Image Processing. IEEE.

[12] Liu Y, Wang Z. (2014) Simultaneous image fusion and denoising with adaptive sparse representation. IET Image Processing, 9(5): 347–357.

[13] Kumar B K S. (2015) Image fusion based on pixel significance using cross bilateral filter. Signal, image and video processing, 9(5): 1193–1204.

[14] Liu Y, Chen X, Ward R K, Wang ZI. (2016) Image fusion with convolutional sparse representation[J]. IEEE signal processing letters, 23(12): 1882–1886.

[15] Ma J, Zhou Z, Wang B, Zong H. (2017) Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Physics & Technology, 82: 8–17.

[16] Li H, Wu XJ. (2018) Infrared and visible image fusion using Latent Low-Rank Representation. arXiv preprint arXiv:1804.08992.

[17] Aardt V J. (2008) Assessment of image fusion procedures using entropy, image quality, and multispectral classification. Journal of Applied Remote Sensing, 2(1): 1-28.

[18] Qu GH, Zhang DL, Yan PF. (2002) Information measure for performance of image fusion. Electronics letters, 38(7): 313-315.