Infrared and Visible Image Fusion using Multi-Scale Decomposition and Visual Saliency Map

Yunfan Chen, Han Xie, Donghoon Yeo, and Hyunchul Shin
Division of Electrical Engineering, Hanyang University, Ansan, Korea
shin@hanyang.ac.kr

Abstract— **In this paper, a novel infrared (IR) and visible (VIS) image fusion algorithm is proposed by using a joint-histogram weighted median filter (JH-WMF) and a mean filter (MF). The proposed method is based on a multi-scale decomposition (MSD) of an image into a base layer, a texture layer, and an edge layer, which is helpful to suppress the artifacts. Furthermore, a novel saliency detection algorithm is proposed for base layers fusion by combining JH-WMF and MF, which is more effective to reduce the loss of contrast. Experiments show that the proposed approach achieves better performance than other methods, in terms of subjective visual effect and objective assessment.**

Keywords —Image fusion; multi-scale decomposition; joint-histogram weighted median filter; mean filter

I. INTRODUCTION

Infrared and visible image fusion technology has wide applications in image processing and computer vision areas. Recently, many methods have been presented for IR and VIS image fusion. Among these, the multiscale image fusion [1-3] and optimization-based image fusion [4] are very successful methods. However, due to lack of consideration of spatial consistency during the process of fusion, these methods may produce distortions in brightness and color. Moreover, optimization-based methods may result in a blurred image. In this study, a novel image fusion algorithm is introduced that apply multi-scale decomposition (MSD) and joint-histogram weighted median filter (JH-WMF) [5] to overcome the aforementioned problems.

II. PROPOSED METHOD

The flowchart of the proposed approach is shown in Fig. 1. First, the input IR and VIS images are decomposed into base, texture, and edge layers. Second, the weight maps for the three layers fusion are caculated. Finally, multi-scale image reconstruction is implemented to generate the fused image.

A. Multi-scale image decomposition

As shown in Fig. 2, the input images are first decomposed into three scale representations by using JH-WMF and MF. JH-WMF is an edge-aware image filter that can well maintain edge features while smoothing small-scale details on each source image. Additionally, JH-WMF is very efficient since the computation complexity is $O(r)$ (r is the kernel size). The base layer of each input image is calculated by the following equation:

$$B_n = I_n * M \tag{1}$$

where I_n is the nth input image, M is the MF with the size of 31x31. Then, the detail layer can be obtained:

$$D_n = I_n - B_n \tag{2}$$

The edge layers are calculated by:

$$E_n = D_n * J \tag{3}$$

where J is the joint-histogram weighted median filter which is set with a default parameter. Then the texture layer can be easily obtained.

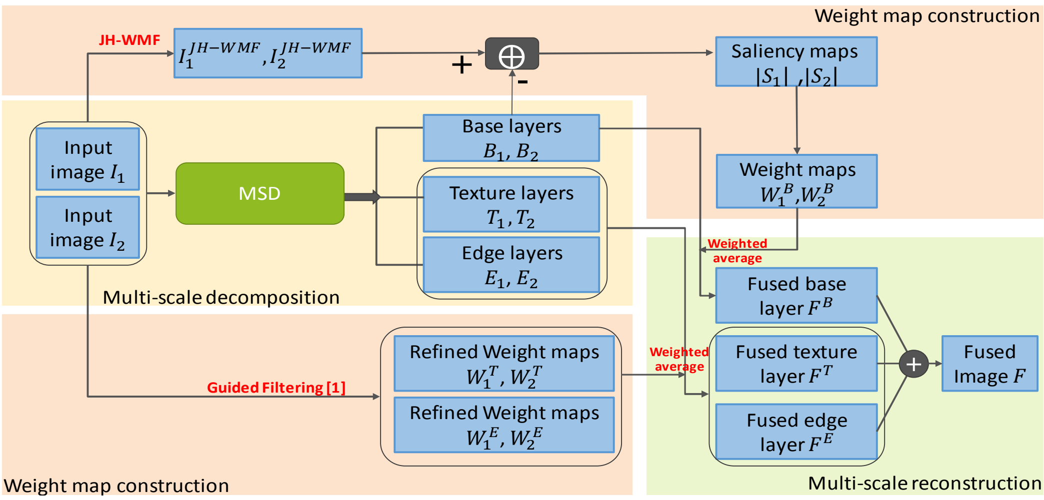

Fig. 1. The framework of our proposed method.

This material is based upon work supported by the Ministry of Trade, Industry & Energy (MOTIE, Korea) under Industrial Technology Innovation Program (10080619).

As shown in Fig. 1. First, a JH-WMF is applied on each input image to remove noise or artifacts:

$$I_n^{JH-WMF} = I_n * J \tag{4}$$

where J is a JH-WMF. Then, the saliency map of each input image is calculated as follows:

$$S_n = |I_n^{JH-WMF} - B_n| \tag{5}$$

Next, the weight maps are determined by comparing each saliency map:

$$\mathrm{P} = \begin{cases} |S_1|, & if\, |S_1| - |S_2| > 0 \\ 0, & else \end{cases} \tag{6}$$

$$W^B = g_{\sigma_c} * S_\lambda(\bar{P}) \tag{7}$$

where the Gaussian g_{σ_c} is applied for convolution to remove noise and locally smooth the weighting coefficients. Here, σ_c is set to 2. The S_λ is defined as

$$S_\lambda(\alpha) = \arctan(\lambda\alpha)\, / \arctan(\lambda) \tag{8}$$

The weighting maps W_1^T, W_2^T for texture layers fusion, and W_1^E, W_2^E for edge layers fusion are calculated by [6].

B. Multi-scale image reconstruction

First, the base, texture, and edge layers fused by:

$$F^B = W^B B_1 + (1 - W^B) B_2 \tag{9}$$

$$F^T = W_1^T T_1 + W_2^T T_2 \tag{10}$$

$$F^E = W_1^E E_1 + W_2^E\, E_2 \tag{11}$$

Then, the fused image F is obtained:

$$F = F^B + F^T + F^E \tag{12}$$

III. Experimental Results

Our experiments are implemented on three IR-VIS image pairs (Camp, kayak, and road) from multi-modal image database [7] and processed on a computer with 4.0 GHz CPU and 32 GB RAM. The proposed method is compared with two recent published IR-VIS fusion works GTF [4] and MSD-SR [1]. The metrics Q_G and Q_{CB} are used to assess the fusion performance. The metric Q_G is based on the gradient, which assesses how much the edge information of the source image successfully transferred to the fused image. Human perception inspired metric Q_{CB} is based on human visual system (HVS) models. For all the metrics, larger values express better performance. A good survey and comparative study of these quality metrics can be found in Z. Liu *et al.*'s work [8].

As shown in Table 1, we can see that our approach is superior or comparable to the other methods. In Table 2, we report a comparison between our method and the other methods in terms of their computation efficiency. Average computational time values are calculated over three pairs of the input images. Table 2 reveals that our approach has a better balance between the runtime and the fusion performance.

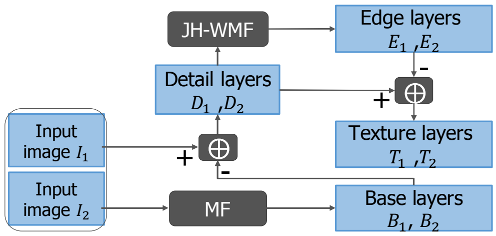

Fig. 2. Multi-scale decomposition (MSD)

TABLE I. THE QUANTITATIVE PERFORMANCE ASSESSMENTS OF DIFFERENT FUSION METHODS ON THE THREE PAIRS OF SOURCE IMAGES

Images	Index	Proposed	GTF	MST-SR
Camp	Q_G	**0.496**	0.397	0.370
	Q_{CB}	**0.541**	0.429	0.528
Kayak	Q_G	**0.604**	0.248	0.501
	Q_{CB}	**0.585**	0.330	0.517
Road	Q_G	**0.650**	0.344	0.479
	Q_{CB}	**0.519**	0.364	**0.519**

TABLE II. RUNTIME COMPARISION (UNIT: SECOND)

Images	Proposed	GTF	MST-SR
Camp	0.87	3.07	0.20
Kayak	1.08	4.76	0.24
Road	0.70	2.86	0.15
Average	**0.88**	3.56	0.20

IV. Conclusion

A novel infrared and visible image fusion method based on multi-scale image decomposition and visual saliency detection is introduced. Experimental results show that our approach achieves comparable or better performance when compared to those of the existing methods.

References

[1] Liu, Yu, Shuping Liu, and Zengfu Wang. "A general framework for image fusion based on multi-scale transform and sparse representation." Information Fusion 24 (2015): 147-164.

[2] Bavirisetti, Durga Prasad, and Ravindra Dhuli. "Two-scale image fusion of visible and infrared images using saliency detection." Infrared Physics & Technology 76 (2016): 52-64.

[3] Gan, Wei, et al. "Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter." Infrared Physics & Technology 72 (2015): 37-51.

[4] Ma, Jiayi, et al. "Infrared and visible image fusion via gradient transfer and total variation minimization." Information Fusion 31 (2016): 100-109.

[5] Zhang, Qi, Li Xu, and Jiaya Jia. "100+ times faster weighted median filter (WMF)." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2830-2837. 2014.

[6] Li, Shutao, Xudong Kang, and Jianwen Hu. "Image fusion with guided filtering." IEEE Transactions on Image Processing 22.7 (2013): 2864-2875.

[7] http://www.escience.cn/people/liuyu1/index.html

[8] Liu, Zheng, et al. "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study." IEEE transactions on pattern analysis and machine intelligence 34.1 (2012): 94-109.