Multi-focus image fusion with half weighted gradient and self-similarity*

DU Chao-ben (杜超本)[1]**, **LIU Ying** (刘颖)[2]，**and GAO She-sheng** (高社生)[1]

1. School of Automation, Northwestern Polytechnical University, Xi'an 710129, China

2. Center for Image and Information Processing, Xi'An University of Posts and Telecommunications, Xi'an 710061, China

(Received 20 February 2018; Revised 7 April 2018)

In order to get a satisfactory image fusion effect, getting a focus map is very necessary and usually difficult to finish. In this paper, we address this problem with a half weighted gradient approach, aiming to obtain a direct mapping between focus map and source images. Based on the advantages of multi-scale weighted gradient, while abandoning the shortcomings of weighted gradient, a new multi-focus image fusion method called half weighted gradient and self-similarity (HWGSS) is proposed. Experimental results validate that the proposed algorithm can obtain state-of-the-art fusion performance in terms of both qualitative and quantitative evaluations.

Document code: A **Article ID:** 1673-1905(2018)04-0311-5

DOI https://doi.org/10.1007/s11801-018-8026-9

In recent years, various image fusion algorithms have been proposed. The image fusion algorithm based on the transform domain usually converts the source image to another feature domain, where the source image can be effectively fused. The most popular transform domain fusion algorithms are based on multi-scale transform (MST). Some representative examples include the Laplacian pyramid (LP)[1], the morphological pyramid (MP)[2], the discrete wavelet transform (DWT)[3], the dual-tree complex wavelet transform (DTCWT)[4] and the non-subsampled contourlet transform (NSCT)[5]. These methods share a common three-step framework, namely, decomposition, fusion and reconstruction.

Block-based fusion strategy decomposes the source images into blocks and each pair of blocks is fused with a designed activity level measurement like sum- modified-Laplacian (SML)[6]. Obviously, the size of the block has a great impact on the quality of the fusion results. From the earliest block-based approach[7], now many improved algorithms have emerged, such as adaptive block-based methods[8] using a differential evolution algorithm to obtain a fixed optimal block size. For the recently introduced method based on the quad-tree[9], the images can be adaptively divided into blocks of different sizes, according to the image content. Another type of spatial domain methods[10] is based on image segmentation by sharing the similar idea of block-based methods, but the fusion quality of these methods relies heavily on the segmentation accuracy.

To overcome the above-mentioned drawbacks, several state-of-the-art pixel-based image fusion algorithms have been proposed, such as guided filtering[11] and dense SIFT[12]. Guided filtering and dense SIFT firstly generate the fusion map by detecting the focused pixels from each source image, then, based on the modified decision map, the final fused image is obtained by selecting the pixels in the focus areas. A multi-focus image fusion method based on multi-scale weighted gradient is competitive with the most current state-of-the-art approaches introduced in Ref.[13]. Although these new algorithms can improve the visual quality of fused images, they may lose some of the original image information due to inaccurate fusion decision maps.

In this paper, we address this problem with a half weighted gradient and self-similarity (HWGSS) approach, aiming to obtain a direct mapping between focus map and source images. Experimental results validate that the proposed algorithm can obtain state-of-the-art fusion performance in terms of both qualitative and quantitative evaluations.

The most critical issue in gradient-based fusion is to capture the geometry of the fused image by combining the gradient from a different image. Let $I(x, y)$: $\Omega \rightarrow [1, M]^N$ be a multi-valued image with N gray-level components $I(x, y)$, $n=1,...N$, where $\Omega \hat{\mathrm{I}} R^2$ denotes the image domain. The coordinate (x, y) represents the gray value of a pixel at the source image. In order to describe the gradient field information of multi-valued images I, the difference in Euclidean space $\Omega \hat{\mathrm{I}} R^N$ is expressed as follows:

* This work has been supported by the National Natural Science Foundation of China (No.61174193).

** E-mail: dcbxjdaxue@163.com

$$\mathrm{d}I = \frac{\partial I}{\partial x}\mathrm{d}x + \frac{\partial I}{\partial y}\mathrm{d}y . \tag{1}$$

The squared norm of dI can be represented as:

$$\|\mathrm{d}I\|^2 = \sum_n \left(\frac{\partial I_n}{\partial x}\right)^2 \mathrm{d}x\mathrm{d}x + 2\frac{\partial I_n}{\partial x}\frac{\partial I_n}{\partial y}\mathrm{d}x\mathrm{d}y + \left(\frac{\partial I_n}{\partial y}\right)^2 \mathrm{d}y\mathrm{d}y = (\mathrm{d}x \;\; \mathrm{d}y)\boldsymbol{G}\begin{pmatrix}\mathrm{d}x\\ \mathrm{d}y\end{pmatrix}, \tag{2}$$

where $\boldsymbol{G} = \begin{pmatrix} \sum_n \left(\frac{\partial I_n}{\partial x}\right)^2 & \sum_n \frac{\partial I_n}{\partial x}\frac{\partial I_n}{\partial y} \\ \sum_n \frac{\partial I_n}{\partial x}\frac{\partial I_n}{\partial y} & \sum_n \left(\frac{\partial I_n}{\partial y}\right)^2 \end{pmatrix}$.

The characteristic value λ of the structural tensor $\boldsymbol{G}$ represents the rate of change of the multi-valued image at a certain point. Simultaneously, $\boldsymbol{G}$ is usually called the structure tensor. It is worth noting that $\boldsymbol{G}$ is a nonnegative matrix whose eigenvalues are real and nonnegative, and they reflect the change in the image at a given point. The maximum rate of change is given by the largest eigenvalue λ^+, and the smallest eigenvalue λ^- represents the minimum rate of change. The eigenvectors θ^+ and θ^- give the corresponding direction of the change.

For a gray-level image $I(x, y)$, we can easily calculate that $\lambda^+ = \|\nabla I\|^2$, $\lambda^- = 0$, and the eigenvector θ lies in the same direction with ∇I. Thus, the gradient of each component of a multi-valued image can be merged by using the largest eigenvalue λ^+ and the corresponding eigenvector θ^+ that are derived from the structure tensor $\boldsymbol{G}$. More specifically, the magnitude of the merged gradient $\nabla \hat{I}$ is calculated by $\|\nabla \hat{I}\| = \sqrt{\lambda^+}$ and the direction of $\nabla \hat{I}$ is specified by θ^+. However, it should be noticed that the sign of the eigenvector is not uniquely specified in the matrix $\boldsymbol{G}$. Indeed, the squared norm of dI in Eq.(2) achieves the same value in opposite directions. An effective solution to this problem is to find a reference gradient and make the direction of the merged gradient consistent with the direction of the reference gradient. As a result, the sign of the eigenvector θ^+ can be specified by sign($\theta^+ \cdot \nabla I_r$), where ∇I_r denotes the reference gradient. A simple choice of ∇I_r would be the averaged gradient of all inputs[13].

We assume that I_F is the fused image. So far, all traditional gradient-based fusion methods reconstruct I_F by making its gradient as close as possible to $\nabla \hat{I}$ [13]. This will result in the following energy functions being minimized.

$$E(I_F) = \iint_w \|\nabla I_F - \nabla \hat{I}\|^2 \mathrm{d}x\mathrm{d}y \tag{3}$$

subject to $0 \le I_F(x, y) \le M$. Its first variation is derived as

$$\mathrm{d}E(I_F) = 2\left(\mathrm{div}(\nabla \hat{I}) - \Delta I_F\right).$$

Thus, the fused image I_F can be simply reconstructed by applying the gradient descent method with $\delta E(I_F)$[12] or by solving the Poisson equation $\Delta I_F = \mathrm{div}(\nabla \hat{I})$, where Δ is a Laplace operator.

It is generally known that local image structure is closely related to the local gradient covariance. Considering a single-valued image $I(x, y)$, the gradient covariance matrix over an local window W_i is defined as follows:

$$\boldsymbol{C} = \begin{pmatrix} \sum_{X \in W_i} I_x^2(X) & \sum_{X \in W_i} I_x(X) I_y(X) \\ \sum_{X \in W_i} I_x(X) I_y(X) & \sum_{X \in W_i} I_y^2(X) \end{pmatrix}, \tag{4}$$

where $I_x(X)$ and $I_y(X)$ denote the gradients along x and y directions at a given position $X=(x, y)$. In order to obtain the representation of local image structure, we decompose $\boldsymbol{C}$ through eigenvalue decomposition as follows:

$$\boldsymbol{C} = \boldsymbol{V}\begin{pmatrix} s_1^2 & \\ & s_2^2 \end{pmatrix}\boldsymbol{V}^{\mathrm{T}} . \tag{5}$$

The local image structure is then related to the eigenvalues of this matrix. If both s_1 and s_2 (square roots of the eigenvalues) are small, there is no significant change in the local region, so this local region is flat. Otherwise, the local region contains certain structural information. In particular, if one of s_1 and s_2 values is large and the other is small, the significant change only occurs in one direction, so a ridge or edge shaped structure is indicated; if both s_1 and s_2 are large, the structure of this local region is sharp in two orthogonal directions, and thus it indicates a corner.

This approach is able to give a steady description of image structure which is robust to both blur and random noise. In our saliency measurement of local image structure, all the types of structures including ridge, edge and corner are concerned. Therefore, we define the image structure saliency Q as:

$$Q = \sqrt{(s_1 + s_2)^2 + \alpha (s_1 - s_2)^2} , \tag{6}$$

where $\alpha > -1$. In the flat region the value of Q is very small. The more salient the local structure is, the larger the Q value becomes. As α is increased to 1, the shape of the contour becomes a quarter of a circle. Since s_1^2 and s_2^2 are the two eigenvalues of matrix $\boldsymbol{C}$, we can obtain

$$s_1^2 + s_2^2 = \sum_{X \in W_i} \left[I_x^2(X) + I_y^2(X)\right] . \tag{7}$$

So, the sum of the eigenvalues reflects the sum of squares of the gradient magnitudes $\sum_{X \in W_i} \|\nabla I(X)\|^2$ in W_i. Assume s_1^2 and s_2^2 are the two eigenvalues derived from the corner structure in W_1, and s_3^2 is the non-zero eigenvalue for the edge structure in W_2, the

squares of Q values for W_1 and W_2 are calculated by

$$\begin{cases} Q_1^2 = (s_1 + s_2)^2 + \alpha (s_1 - s_2)^2 \\ Q_2^2 = s_3^2 + \alpha s_3^2 \end{cases} . \tag{8}$$

By comparing the magnitudes of gradients for all pixels in W_1 and W_2, we can approximately obtain $\sum_{X \in W_1} \|\nabla I(X)\|^2 = \sum_{X \in W_2} \|\nabla I(X)\|^2$, and thus $s_1^2 + s_2^2 = s_3^2$. Then we can obtain

$$Q_1^2 - Q_2^2 = 2(1-\alpha) s_1 s_2 . \tag{9}$$

Thus, the corner structure achieves higher Q values when $\alpha<1$, while the edge structure achieves relatively higher Q values when $\alpha>1$. In order to take more account of the corner structures, we choose $\alpha=0.5$ in this paper.

For each input image I_n at a given point (x, y), we assign a weight w_n that represents the saliency of local image structure by calculating the value of saliency Q_n in a neighborhood of point (x, y). Then, the weighted structure tensor is computed as:

$$\boldsymbol{G}_w = \begin{bmatrix} \sum_n \left(w_n \frac{\partial I_n}{\partial x} \right)^2 & \sum_n w_n^2 \frac{\partial I_n}{\partial x} \frac{\partial I_n}{\partial y} \\ \sum_n w_n^2 \frac{\partial I_n}{\partial x} \frac{\partial I_n}{\partial y} & \sum_n \left(w_n \frac{\partial I_n}{\partial y} \right)^2 \end{bmatrix} , \tag{10}$$

where w_n is normalized by $w_n = \dfrac{Q_n}{(\sum_{i=1}^{N} Q_i^2)^{1/2}}$.

Based on the structure saliency measure and the corresponding weighted structure tensor, we can obtain a weighted gradient-based fusion method. Similar to the fusion method with the structure tensor, in the proposed fusion method, the gradients of all input images are merged first with the weighted structure tensor $\boldsymbol{G}_w$, and then the merged gradient is used to reconstruct the fused image through variation approach.

As the multi-focus image includes both clear and unclear regions, the clear area fusion decision map in the fusion process is 1, while the unclear area fusion decision map is 0. Up to now, the state-of-the-art multi-focus image fusion is based on decision map.

The adaptive region is generated using the shared similarity of source images. The input images are first divided into multiple overlapped square patches, and similar patches are searched[14].

Here, a shared self-similarity of source images is defined as follows to generate an adaptive region of pixels for fusion. Given a reference patch $P_r \in R^{m \times m}$ and a region $R(r) \in R^{n \times n}$ centered at pixel r, the similarity of any candidate patch $P_q \in R^{m \times m}$ to the P_r is defined as

$$n_q = \left\| P_q - P_r \right\|_F , \tag{11}$$

where $\|\cdot\|$ denotes the Frobenius norm. By sorting the n_q by the descending order for all the patches in this region, the most k similar patches to P_r are found and the collection of this patches is expressed as $L_R(r) = \{P_{q_1}, P_{q_2}, \mathbf{L}, P_{q_k}\}$. Similar patches shared by both L_{R^A} and L_{R^B} are

$$L_R^S(r) = L_{R^A}(r) \cap L_{R^B}(r) , \tag{12}$$

where $L_R^S(r)$ is one of the adaptive regions for fusion, and $R^A(r)$ and $R^B(r)$ denote the same region of source images. The locations and the number of similar patches of each adaptive region vary with the shared similarity of source images.

The schematic diagram of our algorithm is shown in Fig.1. In this study, we mainly consider the situation that there are only two pre-registered source images. To deal with more than two multi-focus images, one can fuse them one by one in series. It can be seen from Fig.1 that our method consists of four steps: focus detection, initial segmentation, consistency verification and fusion. In the first step, the two source images are fed to an HWGSS model to output a score map, which contains the focus information of source images. Particularly, each coefficient in the score map indicates the focus property of a pair of corresponding patches from two source images. Then, a focus map with the same size of source images is obtained from the score map by averaging the overlapping patches. In the second step, the focus map is segmented into a binary map with a threshold of 0.5. In the third step, we refine the binary segmented map with a popular consistency verification strategy, namely, small region removal, to generate the final decision map. In the last step, the fused image is obtained with the final decision map using the pixel-wise weighted-average strategy.

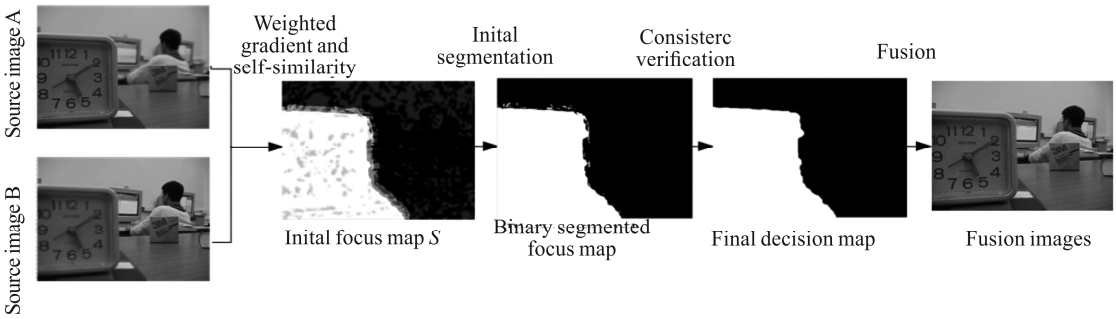

Fig.1 Schematic diagram of the proposed multi-focus image fusion algorithm

The outline of the proposed method is illustrated in Fig.2. First, adaptive regions are generated using image similarity information. Next, we combine with the half weighted gradient to measure the clarity in adaptive regions. Finally, a pixel is fused based on decision map from source images.

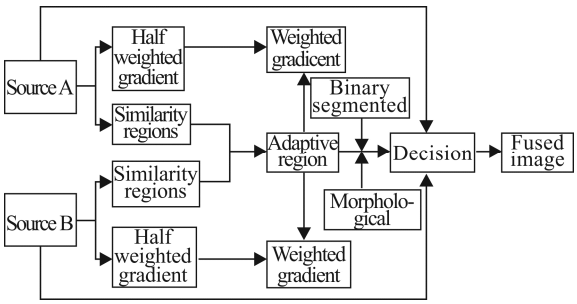

Fig.2 The framework of fusion method

Step1: calculate the gradient weight of the image to be fused according to Eqs. (1) —(10).

Step2: according to Eqs.(11) and (12), calculate the adaptive region of the image to be fused.

Step3: at the edge of the gradient weight, the adaptive region is used to optimize it, and the initial decision map *S* is obtained.

Step4: the decision map *T* is obtained by segmenting initial decision map *S*.

Step5: use mathematical morphology to optimize *T* to get the final decision map *D*.

Step6: use Eq.(14) to reconstruct the fused image.

In order to get a more accurate decision map, initial decision map *S* needs to be further processed in this article. Correspondingly, using a fixed threshold β=0.5 to segment *S* into binary segmented map *T*. The map *T* is given as follows:

$$T(x,y)=\begin{cases}1, & S(x,y)>0.5\\ 0, & \text{otherwise}\end{cases}. \tag{13}$$

From Fig.1, it can be seen that the binary map *T* may contain some misclassified pixels, and these error categories can be easily removed through the small area clear strategy. The fused feature map sometimes contains some very small holes. When this happens, we should also use morphological processing.

Combined with the final fusion decision map *D*, the fused image *F* is obtained according to the pixel weighted average rule as follows:

$$F(x,y)=D(x,y)A(x,y)+(1-D(x,y))B(x,y). \tag{14}$$

To verify the validity of the proposed HWGSS-based fusion algorithm, eight pairs of multi-focus images (including colour images and greyscale images) are used in our experiments. The proposed fusion method is compared with four state-of-the-art multi-focus image fusion methods, which are the MWGF[13], SSDI[14], DCNN[15], and DSIFT[12].

Since the fused images are difficult to categorize between good and bad ones, to further prove the validity of the proposed model for multi-focus image fusion, we mainly compare the decision maps that are produced by a variety of methods. The comparison results of eight pairs of input source images are shown in Fig.3. The final decision maps displayed in the fifth column of Fig.3 obtained from the proposed method are very precise in the boundary, which results in higher visual quality fusion results shown in the last column of Fig.3. The accuracy of the decision map determines the quality of the fusion image, so it can be seen from the visual aspect of Fig.4 that the fusion effect of the proposed method is better than that of other methods.

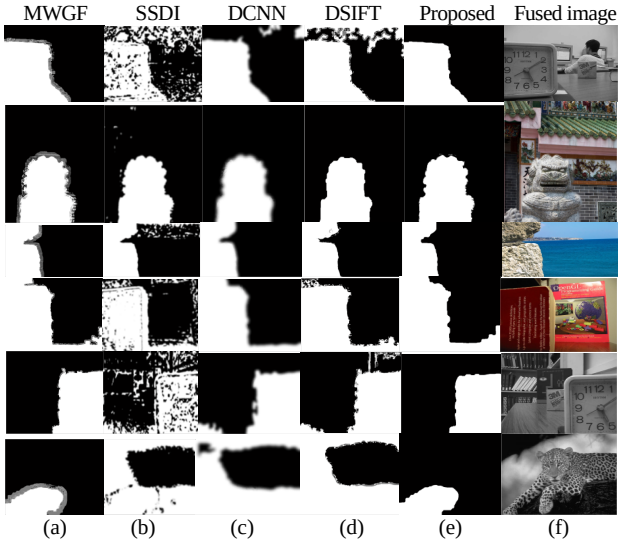

Fig.3 Five different ways to get the decision map matrix *D* and the fusion image: (a) Lab; (b) Temple; (c) Seascape; (d) Book; (e) Desk; (f) Leopard

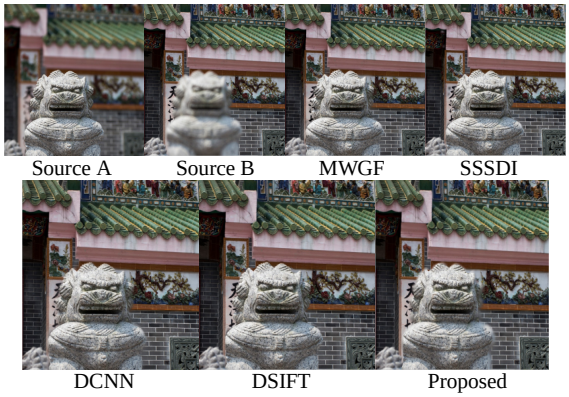

Fig.4 The fusion results of five different algorithms

To prove the validity and practicability of the proposed algorithm, the two indexes of mutual information *MI* and $Q^{AB/F}$ are used as the objective evaluation measure of information fusion performance[16]. The objective performances on the fused images using the five fusion methods are listed in Tab.1, from which we observe that

the HWGSS-based method outperforms all of the other fusion methods, which further proves the validity of HWGSS for image fusion. At the same time, the time complexity of this paper is the least. It is very suitable for engineering applications.

Tab.1 The comparison of *MI*, $Q^{AB/F}$ values and time of five different algorithms

		MWGF	SSDI	DCNN	DSIFT	Proposed
Lab	*MI*	8.061 8	8.141 2	8.600 8	8.520 1	8.867 1
	$Q^{AB/F}$	0.714 7	0.752 8	0.757 3	0.758 5	0.765 6
	Time	5.827 8	19.701 4	287.199 2	23.649 1	3.411 0
Temple	*MI*	5.965 5	7.089 6	6.889 5	7.351 4	7.433 7
	$Q^{AB/F}$	0.750 1	0.763 4	0.759 0	0.764 3	0.753 1
	Time	13.642 5	15.173 9	248.668 0	8.440 2	3.326 8
Seascape	*MI*	7.140 4	7.482 4	7.628 5	7.948 7	8.037 6
	$Q^{AB/F}$	0.705 9	0.711 0	0.711 3	0.712 6	0.712 5
	Time	11.467 1	11.117 6	179.488 4	5.909 2	2.709 7
Book	*MI*	8.236 8	8.400 8	8.779 6	8.662 3	8.312 5
	$Q^{AB/F}$	0.724 0	0.726 0	0.727 7	0.713 4	0.738 3
	Time	22.501 8	25.075 4	403.118 1	8.631 2	2.792 9
Desk	*MI*	7.911 5	7.539 9	8.043 8	8.216 5	8.895 8
	$Q^{AB/F}$	0.729 8	0.729 3	0.734 2	0.736 4	0.793 6
	Time	6.095 4	17.944 9	320.292 1	12.217 4	2.783 9
Leopard	*MI*	9.947 4	10.888 7	10.879 2	10.922 6	10.945 9
	$Q^{AB/F}$	0.817 5	0.817 1	0.797 3	0.806 9	0.824 4
	Time	3.661 1	11.213 9	172.759 7	3.235 7	2.013 5

In this article, a new multi-focus image fusion method called HWGSS is proposed. We can experimentally verify that the focus/decision map obtained from the HWGSS is reliable, which can lead to high-quality fusion results.

References

[1] P. Burt and E. Adelson, IEEE Trans. Commun. **31**, 532 (1983).

[2] A. Toet, Pattern Recognit. Lett. **9**, 255 (1989).

[3] H. Li, B. Manjunath and S. Mitra, Graphical Models Image Process. **57**, 235 (1995).

[4] J. Lewis, R. Oí Callaghan, S. Nikolov, D. Bull and N. Canagarajah, Inf. Fusion **8**, 119 (2007).

[5] Q. Zhang and B. Guo, Signal Process. **89**, 1334 (2009).

[6] W. Huang and Z. Jing, Pattern Recognit. Lett. **28**, 493 (2007).

[7] S. Li, J. Kwok and Y. Wang, Inf. Fusion **2**, 169 (2001).

[8] V. Aslantas and R. Kurban, Expert Syst. Appl. **37**, 8861 (2010).

[9] X. Bai, Y. Zhang, F. Zhou and B. Xue, Inf. Fusion **22**, 105 (2015).

[10] C. Du and S. Gao, IEEE Access **5**, 15750 (2017).

[11] S. Li, X. Kang and J. Hu, IEEE Trans. Image Process. **22**, 2864 (2013).

[12] Y. Liu, S. Liu and Z. Wang, Inf. Fusion **23**, 139 (2015) .

[13] Z. Zhou, S. Li and B. Wang, Inf. Fusion **20**, 60 (2014).

[14] Guo D, Yan J.W and Qu X, Optics Communications **338**, 138 (2015).

[15] Liu Y, Chen X, Peng H and Wang Z, Inf. Fusion **36**, 191 (2017).

[16] Yong Yang, Song Tong, Shuying Huang and Pan Lin, IEEE Sensors Journal **15**, 2824 (2015).