Image Fusion with Guided Filtering

Shutao Li, *Member, IEEE,* Xudong Kang, *Student Member, IEEE,* and Jianwen Hu

***Abstract*— A fast and effective image fusion method is proposed for creating a highly informative fused image through merging multiple images. The proposed method is based on a two-scale decomposition of an image into a base layer containing large scale variations in intensity, and a detail layer capturing small scale details. A novel guided filtering-based weighted average technique is proposed to make full use of spatial consistency for fusion of the base and detail layers. Experimental results demonstrate that the proposed method can obtain state-of-the-art performance for fusion of multispectral, multifocus, multimodal, and multiexposure images.**

***Index Terms*— Guided filter, image fusion, spatial consistency, two-scale decomposition.**

I. Introduction

IMAGE fusion is an important technique for various image processing and computer vision applications such as feature extraction and target recognition. Through image fusion, different images of the same scene can be combined into a single fused image [1]. The fused image can provide more comprehensive information about the scene which is more useful for human and machine perception. For instance, the performance of feature extraction algorithms can be improved by fusing multi-spectral remote sensing images [2]. The fusion of multi-exposure images can be used for digital photography [3]. In these applications, a good image fusion method has the following properties. First, it can preserve most of the useful information of different images. Second, it does not produce artifacts. Third, it is robust to imperfect conditions such as mis-registration and noise.

A large number of image fusion methods [4]–[7] have been proposed in literature. Among these methods, multi-scale image fusion [5] and data-driven image fusion [6] are very successful methods. They focus on different data representations, e.g., multi-scale coefficients [8], [9], or data driven decomposition coefficients [6], [10] and different image fusion rules to guide the fusion of coefficients. The major advantage of these methods is that they can well preserve the details of different source images. However, these kinds of methods may produce brightness and color distortions since spatial consistency is not well considered in the fusion process. To make full use of spatial context, optimization based image fusion approaches, e.g., generalized random walks [3], and Markov random fields [11] based methods have been proposed. These methods focus on estimating spatially smooth and edge-aligned weights by solving an energy function and then fusing the source images by weighted average of pixel values. However, optimization based methods have a common limitation, i.e., inefficiency, since they require multiple iterations to find the global optimal solution. Moreover, another drawback is that global optimization based methods may over-smooth the resulting weights, which is not good for fusion.

To solve the problems mentioned above, a novel image fusion method with guided filtering is proposed in this paper. Experimental results show that the proposed method gives a performance comparable with state-of-the-art fusion approaches. Several advantages of the proposed image fusion approach are highlighted in the following.

1) Traditional multi-scale image fusion methods require more than two scales to obtain satisfactory fusion results. The key contribution of this paper is to present a fast two-scale fusion method which does not rely heavily on a specific image decomposition method. A simple average filter is qualified for the proposed fusion framework.
2) A novel weight construction method is proposed to combine pixel saliency and spatial context for image fusion. Instead of using optimization based methods, guided filtering is adopted as a local filtering method for image fusion.
3) An important observation of this paper is that the roles of two measures, i.e., pixel saliency and spatial consistency are quite different when fusing different layers. In this paper, the roles of pixel saliency and spatial consistency are controlled through adjusting the parameters of the guided filter.

The remainder of this paper is organized as follows. In Section II, the guided image filtering algorithm is briefly reviewed. Section III describes the proposed image fusion algorithm. The experimental results and discussions are presented in Section IV. Finally, Section V concludes the paper.

II. Guided Image Filtering

Recently, edge-preserving filters [12], [13] have been an active research topic in image processing. Edge-preserving smoothing filters such as guided filter [12], weighted least squares [13], and bilateral filter [14] can avoid ringing artifacts since they will not blur strong edges in the decomposition process. Among them, the guided filter is a recently proposed edge-preserving filter, and the computing time of which is independent of the filter size. Furthermore, the guided filter

Manuscript received January 3, 2012; revised January 13, 2013; accepted January 16, 2013. Date of publication January 30, 2013; date of current version May 22, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant 61172161, the Chinese Scholarship Award for Excellent Doctoral Student, and the Hunan Provincial Innovation Foundation for Postgraduate. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brendt Wohlberg.

The authors are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: (shutao_li@hnu.edu.cn; xudong_kang@163.com; hujianwen1@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2244222

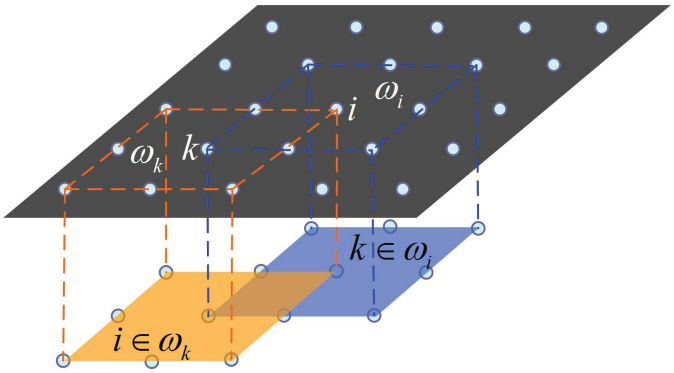


Fig. 1. Illustration of window choice.

is based on a local linear model, making it qualified for other applications such as image matting, up-sampling and colorization [12]. In this paper, the guided filter is first applied for image fusion.

In theory, the guided filter assumes that the filtering output O is a linear transformation of the guidance image I in a local window ω_k centered at pixel k.

$$O_i = a_k I_i + b_k \ \forall i \in \omega_k \tag{1}$$

where ω_k is a square window of size $(2r+1) \times (2r+1)$. The linear coefficients a_k and b_k are constant in ω_k and can be estimated by minimizing the squared difference between the output image O and the input image P.

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left((a_k I_i + b_k - P_i)^2 + \epsilon a_k^2 \right) \tag{2}$$

where ϵ is a regularization parameter given by the user. The coefficients a_k and b_k can be directly solved by linear regression [15] as follows:

$$a_k = \frac{\frac{1}{|\omega|}\sum_{i \in \omega_k} I_i P_i - \mu_k \overline{P}_k}{\delta_k + \epsilon} \tag{3}$$

$$b_k = \overline{P}_k - a_k \mu_k \tag{4}$$

where μ_k and δ_k are the mean and variance of I in ω_k respectively, $|\omega|$ is the number of pixels in ω_k, and $\overline{P}_k$ is the mean of P in ω_k. Next, the output image can be calculated according to (1). As shown in Fig. 1, all local windows centered at pixel k in the window ω_i will contain pixel i. So, the value of O_i in (1) will change when it is computed in different windows ω_k. To solve this problem, all the possible values of coefficients a_k and b_k are first averaged. Then, the filtering output is estimated as follows:

$$O_i = \overline{a}_i I_i + \overline{b}_i \tag{5}$$

where $\overline{a}_i = \frac{1}{|\omega|}\sum_{k \in \omega_i} a_k$, $\overline{b}_i = \frac{1}{|\omega|}\sum_{k \in \omega_i} b_k$. In this paper, $G_{r,\epsilon}(P, I)$ is used to represent the guided filtering operation, where r and ϵ are the parameters which decide the filter size and blur degree of the guided filter, respectively. Moreover, P and I refer to the input image and guidance image, respectively.

Furthermore, when the input is a color image, the filtering output can be obtained by conducting the guided filtering on the red, green, and blue channels of the input image, respectively. And when the guidance image I is a color image, the guided filter should be extended by the following steps. First, equation (1) is rewritten as follows:

(a) (b) (c) (d)

Fig. 2. Two examples of guided filtering. (a) and (c) are two input images of the guided filter. Image (b) is the filtered image ($r = 15$, $\epsilon = 0.3$), with image (a) serving as the input image and the guidance image simultaneously. Image (d) is the filtered image ($r = 10$, $\epsilon = 10^{-6}$), with images (a) and (c) serving as the guidance image and the input image, respectively.

$$O_i = \mathbf{a}_k^T \mathbf{I}_i + b_k \ \forall i \in \omega_k \tag{6}$$

where $\mathbf{a}_k$ is a 3×1 coefficient vector and $\mathbf{I}_i$ is a 3×1 color vector. Then, similar to (3)–(5), the output of guided filtering can be calculated as follows:

$$\mathbf{a}_k = (\Sigma_k + \epsilon U) \left(\frac{1}{|\omega|} \sum_{i \in \omega_k} \mathbf{I}_i p_i - \boldsymbol{\mu}_k \overline{p}_k \right) \tag{7}$$

$$b_k = \overline{p}_k - \mathbf{a}_k^T \boldsymbol{\mu}_k \tag{8}$$

$$O_i = \overline{\mathbf{a}}_i^T \mathbf{I}_i + \overline{b}_i \tag{9}$$

where Σ_k is the 3×3 covariance matrix of $\mathbf{I}$ in ω_k, and U is the 3×3 identity matrix.

For instance, Fig. 2(a) shows a color image of size 620×464. Guided filtering is conducted on each color channel of this image to obtain the color filtered image shown in Fig. 2(b) (for this example, Fig. 2(a) serves as the guidance image and the input image simultaneously). As shown in the close-up view in Fig. 2(b), the guided filter can blur the image details while preserving the strong edges of the image. Fig. 2(c) and (d) give another example of guided filtering when the input image and guidance image are different. In this example, Fig. 2(c) and (a) serve as the input image and the color guidance image, respectively. It can be seen that the input image shown in Fig. 2(c) is noisy and not aligned with object boundaries. As shown in Fig. 2(d), after guided filtering, noisy pixels are removed and the edges in the filtered image are aligned with object boundaries. It demonstrates that those pixels with similar colors in the guidance image tend to have similar values in the filtering process.

III. Image Fusion With Guided Filtering

Fig. 3 summarizes the main processes of the proposed guided filtering based fusion method (GFF). First, an average filter is utilized to get the two-scale representations. Then, the base and detail layers are fused through using a guided filtering based weighted average method.

A. Two-Scale Image Decomposition

As shown in Fig. 3, the source images are first decomposed into two-scale representations by average filtering. The base layer of each source image is obtained as follows:

$$B_n = I_n * Z \tag{10}$$

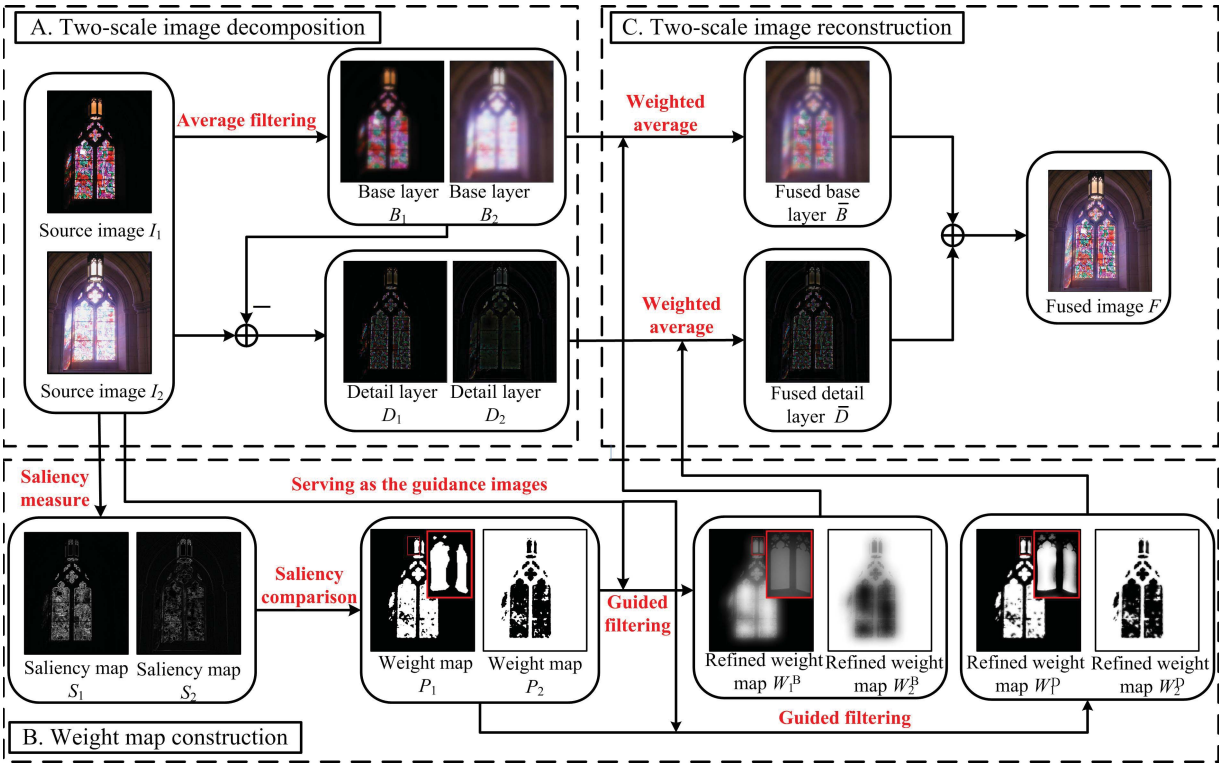


Fig. 3. Schematic diagram of the proposed image fusion method based on guided filtering.

where I_n is the nth source image, Z is the average filter, and the size of the average filter is conventionally set to 31×31 . Once the base layer is obtained, the detail layer can be easily obtained by subtracting the base layer from the source image.

$$D_n = I_n - B_n. \tag{11}$$

The two-scale decomposition step aims at separating each source image into a base layer containing the large-scale variations in intensity and a detail layer containing the small-scale details.

B. Weight Map Construction With Guided Filtering

As shown in Fig. 3, the weight map is constructed as follows. First, Laplacian filtering is applied to each source image to obtain the high-pass image H_n.

$$H_n = I_n * L \tag{12}$$

where L is a 3×3 Laplacian filter. Then, the local average of the absolute value of H_n is used to construct the saliency maps S_n.

$$S_n = |H_n| * g_{r_g,\sigma_g} \tag{13}$$

where g is a Gaussian low-pass filter of size $(2r_g + 1)(2r_g + 1)$, and the parameters r_g and σ_g are set to 5. The measured saliency maps provide good characterization of the saliency level of detail information. Next, the saliency maps are compared to determine the weight maps as follows:

$$P_n^k = \begin{cases} 1 \text{ if } S_n^k = \max\left(S_1^k, S_2^k, \ldots, S_N^k\right) \\ 0 \text{ otherwise} \end{cases} \tag{14}$$

where N is number of source images, S_n^k is the saliency value of the pixel k in the nth image. However, the weight maps obtained above are usually noisy and not aligned with object boundaries (see Fig. 3), which may produce artifacts to the fused image. Using spatial consistency is an effective way to solve this problem. Spatial consistency means that if two adjacent pixels have similar brightness or color, they will tend to have similar weights. A popular spatial consistency based fusion approach is formulating an energy function, where the pixel saliencies are encoded in the function and edge aligned weights are enforced by regularization terms, e.g., a smoothness term. This energy function can be then minimized globally to obtain the desired weight maps. However, the optimization based methods are often relatively inefficient.

In this paper, an interesting alternative to optimization based methods is proposed. Guided image filtering is performed on each weight map P_n with the corresponding source image I_n serving as the guidance image.

$$W_n^B = G_{r_1,\epsilon_1}(P_n, I_n) \tag{15}$$

$$W_n^D = G_{r_2,\epsilon_2}(P_n, I_n) \tag{16}$$

where r_1, ϵ_1, r_2, and ϵ_2 are the parameters of the guided filter, W_n^B and W_n^D are the resulting weight maps of the base and detail layers. Finally, the values of the N weight maps are normalized such that they sum to one at each pixel k.

The motivation of the proposed weight construction method is as follows. According to (1), (3) and (4), it can be seen that if the local variance at a position i is very small which means that the pixel is in a flat area of the guidance image, then a_k will become close to 0 and the filtering output O will equal to $\overline{P}_k$, i.e., the average of adjacent input pixels. In contrast, if the local variance of pixel i is very large which means that the pixel i is in an edge area, then a_k will become far from

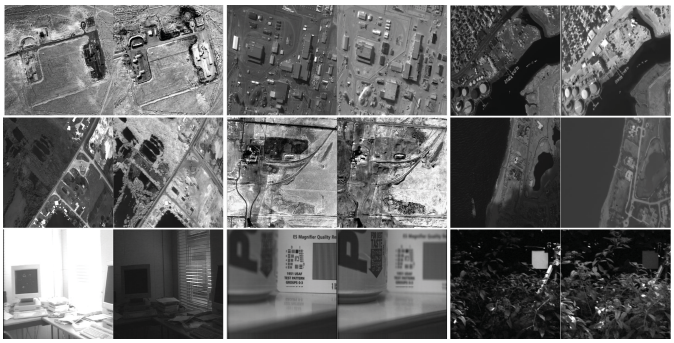


Fig. 4. Illustrations of nine pairs of testing images of the Petrović database.

zero. As demonstrated in [12], $\nabla O \approx \overline{a}\nabla I$ will become true, which means that only the weights in one side of the edge will be averaged. In both situations, those pixels with similar color or brightness tend to have similar weights. This is exactly the principle of spatial consistency.

Furthermore, as shown in Fig. 3, the base layers look spatially smooth and thus the corresponding weights also should be spatially smooth. Otherwise, artificial edges may be produced. In contrast, sharp and edge-aligned weights are preferred for fusing the detail layers since details may be lost when the weights are over-smoothed. Therefore, a large filter size and a large blur degree are preferred for fusing the base layers, while a small filter size and a samll blur degree are preferred for the detail layers.

C. Two-Scale Image Reconstruction

Two-scale image reconstruction consists of the following two steps. First, the base and detail layers of different source images are fused together by weighted averaging

$$\overline{B} = \sum_{n=1}^{N} W_n^B B_n \tag{17}$$

$$\overline{D} = \sum_{n=1}^{N} W_n^D D_n. \tag{18}$$

Then, the fused image F is obtained by combining the fused base layer $\overline{B}$ and the fused detail layer $\overline{D}$

$$F = \overline{B} + \overline{D}. \tag{19}$$

IV. Experiments and Discussion

A. Experimental Setup

Experiments are performed on three image databases, i.e., the Petrović database [16] which contains 50 pairs of images including aerial images, outdoor images (natural, industrial) and indoor images (with different focus points and exposure settings), the multi-focus image database which contains 10 pairs of multi-focus images, and the multi-exposure and multi-modal image database which contains 2 pairs of color multi-exposure images and 8 pairs of multi-modal images. The testing images have been used in many related papers [3]–[10], [17]–[21]. Fig. 4 shows 9 pairs of images of the Petrović database. Fig. 5 shows the multi-focus database.

Fig. 5. Multifocus image database composed by ten pairs of multifocus images.

Fig. 6. Multiexposure and multimodal image database composed by two pairs of multiexposure images and eight pairs of multimodal images.

Further, Fig. 6 shows the multi-exposure and multi-modal database.

The proposed guided filtering based fusion method (GFF) is compared with seven image fusion algorithms based on Laplacian pyramid (LAP) [8], stationary wavelet transform (SWT) [9], curvelet transform (CVT) [19], nonsubsampled contourlet transform (NSCT) [20], generalized random walks (GRW) [3], wavelet-based statistical sharpness measure (WSSM) [21] and high order singular value decomposition (HOSVD) [10], respectively. The parameter settings of these methods are as follows. Four decomposition levels, the "averaging" scheme for the low-pass sub-band, the absolute maximum choosing scheme for the band-pass sub-band and the 3×3 window based consistency check are adopted for the LAP, CVT, SWT, and NSCT method. Four decomposition levels with 4, 8, 8, 16 directions from coarser scale to finer scale are adopted for the NSCT method. Furthermore, the default parameters given by the respective authors are adopted for the GRW, WSSM and HOSVD based methods.

B. Objective Image Fusion Quality Metrics

In order to assess the fusion performance of different methods objectively, five fusion quality metrics, i.e., information theory based metric (Q_{MI} [22]), structure based metrics (Q_Y [23] and Q_C [24]) and feature based metrics (Q_G [25] and Q_P [26]) are adopted. A good survey and comparative study of these quality metrics can be found in Z. Liu *et al.*'s work [27]. The default parameters given in the related publications are adopted for these quality indexes.

1) Normalized mutual information Q_{MI} [22] is an information theory based metric. One problem with traditional mutual information metric [28] is that it is unstable and may bias the measure towards the source image with the highest entropy. Hossny *et al.* modified it to the normalized mutual information [22]. In this paper, Hossny *et al.*'s definition is adopted.

$$Q_{MI} = 2\left[\frac{MI(A,F)}{H(A)+H(F)} + \frac{MI(B,F)}{H(B)+H(F)}\right] \tag{20}$$

where $H(A)$, $H(B)$ and $H(F)$ are the marginal entropy of A, B and F, and $MI(A,F)$is the mutual information between the source image A and the fused image F.

$$MI(A,F) = H(A) + H(F) - H(A,F) \tag{21}$$

where $H(A,F)$ is the joint entropy between A and F, $H(A)$ and $H(F)$ are the marginal entropy of A and F, respectively, and $MI(B,F)$ is similar to $MI(A,F)$. The quality metric Q_{MI} measures how well the original information from source images is preserved in the fused image.

2) Yang *et al.*'s metric Q_Y uses structural similarity ($SSIM$) [29] for fusion assessment. It is defined as follows:

$$Q_Y = \begin{cases} \lambda_w SSIM(A_w, F_w) + (1-\lambda_w) SSIM(B_w, F_w), \\ \qquad \text{if } SSIM(A_w, B_w|w) \geq 0.75 \\ \max\left\{SSIM(A_w, F_w), SSIM(B_w, F_w)\right\}, \\ \qquad \text{if } SSIM(A_w, B_w|w) < 0.75 \end{cases} \tag{22}$$

where w is a window of size 7×7, A, B are the input images and F is the fused image, $SSIM$ is the structural similarity [29] and the local weight λ_w is calculated as follows:

$$\lambda_w = \frac{s(A_w)}{s(A_w) + s(B_w)} \tag{23}$$

where $s(A_w)$ and $s(B_w)$ are the variance of source images A and B within the window w, respectively. Q_Y measures how well the structural information of source images is preserved.

3) Cvejic *et al.*'s metric Q_C [24] is calculated as follows:

$$\begin{aligned} Q_C = {} & \mu(A_w, B_w, F_w) UIQI(A_w, F_w) \\ & + \left(1 - \mu\left(A_w, B_w, F_w\right)\right) UIQI(B_w, F_w) \end{aligned} \tag{24}$$

where $\mu(A_w, B_w, F_w)$ is calculated as follows:

$$\mu(A_w, B_w, F_w) = \begin{cases} 0, & \text{if } \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}} < 0 \\ \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}}, & \text{if } 0 \leq \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}} < 1 \\ 1, & \text{if } \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}} > 1 \end{cases} \tag{25}$$

σ_{AF} and σ_{BF} are the covariance between A, B and F, $UIQI$ refers to the universal image quality index [30]. The Q_C quality metric estimates how well the important information in the source images is preserved in the fused image, while minimizing the amount of distortion that could interfere with interpretation.

4) The gradient based index Q_G evaluates the success of edge information transferred from the source images to the fused image [25]. It is calculated as follows:

$$Q_G = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(Q^{AF}(i,j)\tau^A(i,j) + Q^{BF}(i,j)\tau^B(i,j)\right)}{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\tau^A(i,j) + \tau^B(i,j)\right)} \tag{26}$$

where $Q^{AF} = Q_g^{AF} Q_o^{AF}$. $Q_g^{AF}(i,j)$ and $Q_o^{AF}(i,j)$ are the edge strength and orientation preservation values at location (i,j), respectively, N and M are the width and height of the images, $Q^{BF}(i,j)$ is similar to $Q^{AF}(i,j)$. $\tau^A(i,j)$ and $\tau^B(i,j)$ reflect the importance of $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$, respectively.

5) The last quality metric is the phase congruency based index Q_P [26]. The phase congruency and the principal moments (maximum and minimum) which contain the information for corners and edges, are used to define the Q_P metric.

$$Q_P = (P_p)^\alpha (P_M)^\beta (P_m)^\gamma \tag{27}$$

where p, M and m denote phase congruency, maximum and minimum moments, respectively. α, β, and γ are the exponential parameters which are all set to 1 in this paper. The Q_P index computes how well the salient features of source images are preserved [26]. The larger the values of the five quality metrics described above are, the better the fusion results will be.

Fig. 7. Separate image database composed by two pairs of multispectral images, two pairs of multifocus images, four pairs of multimodal images, and two pairs of multiexposure images.

C. Analysis of Free Parameters

In this subsection, the influences of different parameters to objective fusion performances are analyzed with a separate image database shown in Fig. 7. Most images of which are public available[1]. The fusion performance is evaluated by the average values of five fusion quality metrics, i.e., Q_{MI}, Q_Y, Q_C, Q_G and Q_P. When analyzing the influence of r_1, other parameters are set to $\epsilon_1 = 0.3$, $r_2 = 7$, and $\epsilon_2 = 10^{-6}$. Then, when analyzing the influence of ϵ_1, other parameters are set to $r_1 = 45$, $r_2 = 7$, and $\epsilon_2 = 10^{-6}$. Next, r_2 and ϵ_2 are analyzed in the same way. As shown in Fig. 8, when fusing the base layers, it is preferred to have a big filter size r_1 and blur degree ϵ_1. When fusing the detail layers, the fusion performance will be worse when the filter size r_2 is too large or too small. In this paper, the default parameters are set as $r_1 = 45$, $\epsilon_1 = 0.3$, $r_2 = 7$, and $\epsilon_2 = 10^{-6}$. This fixed parameter setting can obtain good results for all images used in this paper, because the GFF method does not depend much on the exact parameter choice.

D. Experimental Results and Discussion

1) Comparison With Other Image Fusion Methods: Fig. 9(a1)–(a10) show two multi-spectral images and the fused images obtained by different methods. Furthermore, a close-up view is presented in the right-bottom of each sub-picture. As shown in Fig. 9(a10), the fused image obtained by the proposed guided filtering based fusion method (GFF) can well preserve the complementary information of different source

[1] http://imagefusion.org

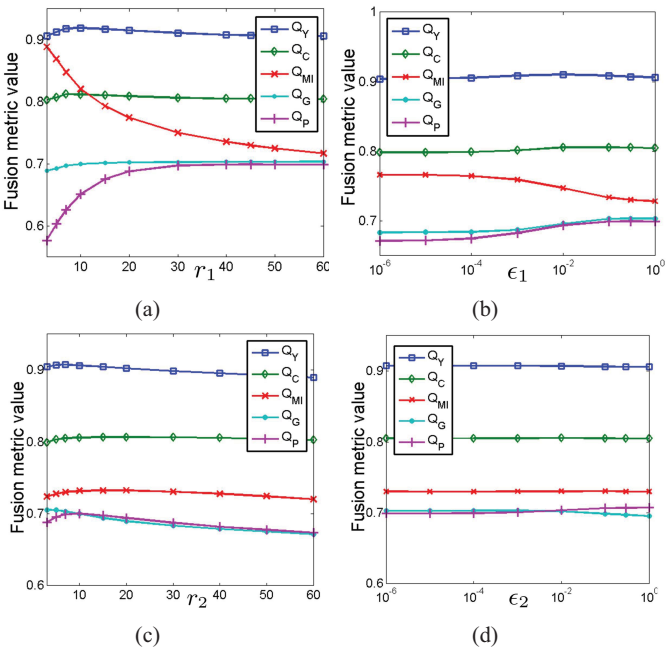


Fig. 8. Performance of the GFF method with different parameters, i.e., (a) r_1, (b) ϵ_1, (c) r_2, and (d) ϵ_2.

images such as the meniscus shaped road and the lines in the roof area (see Fig. 9(a1) and (a2)). However, the meniscus shaped road is not clear in the fusion results produced by other methods (see Fig. 9(a3)–(a9)). Fig. 9(b) gives an example of infrared and visual image fusion. From this figure, it can be seen that the GFF method can well preserve the brightness of the pedestrians, while other methods may decrease the brightness and contrast of the pedestrians (see Fig. 9(b3)–(b6)), produce serious artifacts (see Fig. 9(b8)) or lose important features (see Fig. 9(b7) and (b9)). Fig. 9(c1) and (c2) show two medical images captured using computed tomography (CT) and magnetic resonance imaging (MRI) respectively. The first one displays the bone structure and the other reveals the soft tissue of a human's head. From Fig. 9(c3)–(c6), it can be seen that the results produced by the CVT, SWT, LAP, and NSCT method may decrease the brightness of soft-tissue structures, and thus make some details invisible. As shown in Fig. 9(c7), the GRW based method does not work for this example, because fusion weights are over-smoothed in the global optimization process. The WSSM based method introduces serious artifacts to the fused image (see Fig. 9(c8)). The HOSVD based method loses some important bone structures (see Fig. 9(c9)). However, the GFF method can preserve these features and details without producing visible artifacts and brightness distortions.

Fig. 10(a1) and (a2) show two color multi-focus images captured by a hand-hold camera. In this example, the source images are not perfectly registered due to camera movement. As shown in the close-up views of Fig. 10(a3)–(a9), the SWT, CVT, LAP, NSCT, GRW, and HOSVD based image fusion methods produce artificial edges in the fused images. Furthermore, some areas in the fused image obtained by the HOSVD based method are blurred (see the magnified leaf area of Fig. 10(a9)). However, as shown in Fig. 10(a8) and (a10), the WSSM based method and the GFF method can well preserve the focused areas of different source images without introducing any artifacts. This demonstrates that the GFF method and the WSSM based method are more robust to image mis-registration. Fig. 10(b1) and (b2) show two color multi-focus images which contain moving objects. From Fig. 10(b3)–(b9), it can be seen that other image fusion methods introduce different levels of artifacts in the girl area. Among these methods, the WSSM based method is more robust to image mis-registration, since the wavelet based saliency measure method performs well in detecting the real focused object. However, only considering pixel saliency, the resulting weights may be noisy and not aligned with real object boundaries and thus some artifacts are still visible in their fused image (see Fig. 10(b8)). The GFF method performs much better for Fig. 10(b), since it can make full use of the strong correlations between adjacent pixels to refine these noisy and inaccurate weights estimated by the saliency measure. This makes great contribution in removing the influence of image mis-registration. At last, Fig. 10(c1) and (c2) show two multi-exposure images captured with different exposure settings. As shown in the close-up views of Fig. 10(c3)–(c9), it can be seen that other fusion methods cause brightness and color distortions in the window and wall area. By contrast, the fused image obtained by the GFF method can well preserve the color and detail information in these areas (see Fig. 10(c10)). Based on the comparison above, it can be concluded that traditional multi-scale image fusion methods usually perform well in preserving image details while they may cause brightness or color distortions. The GRW based method can obtain good fusion results for multi-exposure images but may fail in other image fusion applications. The WSSM based method can well preserve the complementary information of different source images. However, it may produce serious edge artifacts when fusing multi-modal or multi-exposure images. The HOSVD based method can usually obtain fused images without artifacts and color distortions, but it may lose some important complementary information. By contrast, the GFF method can well preserve the complementary information of source images without producing artifacts and distortions.

Next, an artificial multi-focus image fusion example is presented in Fig. 11. For this example, the source images are created by respectively blurring the foreground and background objects of the cameraman image. As shown in Fig. 11(c), the SWT based method produces ringing artifacts around strong boundaries. The WSSM based method does not work well for this example since it cannot obtain edge-aligned weights (see Fig. 11(h)). For clearer comparison, the difference image in the box area is obtained by first subtracting each fused image from the reference all-focused image and then calculating the absolute values. The difference image is visualized by using a color map (blue means small differences, red means large differences) and presented in the right bottom of each image. As shown in Fig. 11(c)–(j), the result produced by the GFF method is much closer to the ideal reference image, whereas the results produced by other methods have different levels of distortions around edge areas.

At last, the objective performances of different methods are shown in Table I. It can be seen that, for the Petrović

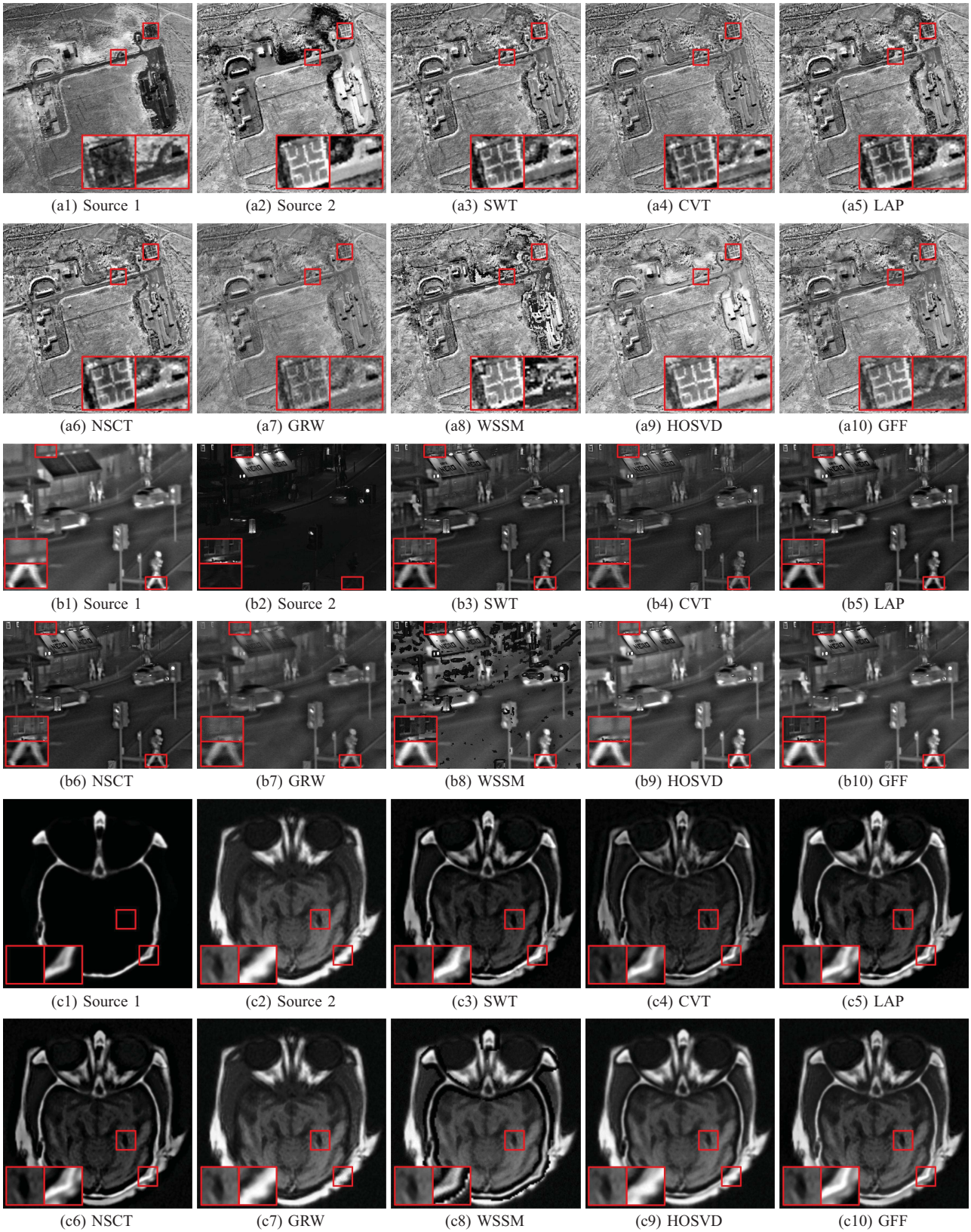


Fig. 9. Gray source images and the fused images obtained by different methods.

image database, the HOSVD based method gives the largest quality indexes for Q_Y, Q_P, and Q_{MI}. Moreover, the Q_{MI} value of the HOSVD based method is always the largest for all image databases. This means that the HOSVD method can well preserve the original information of different source images. However, the Q_{MI} index cannot measure whether the complementary information of different source images is well preserved. In other words, if the fused image looks very

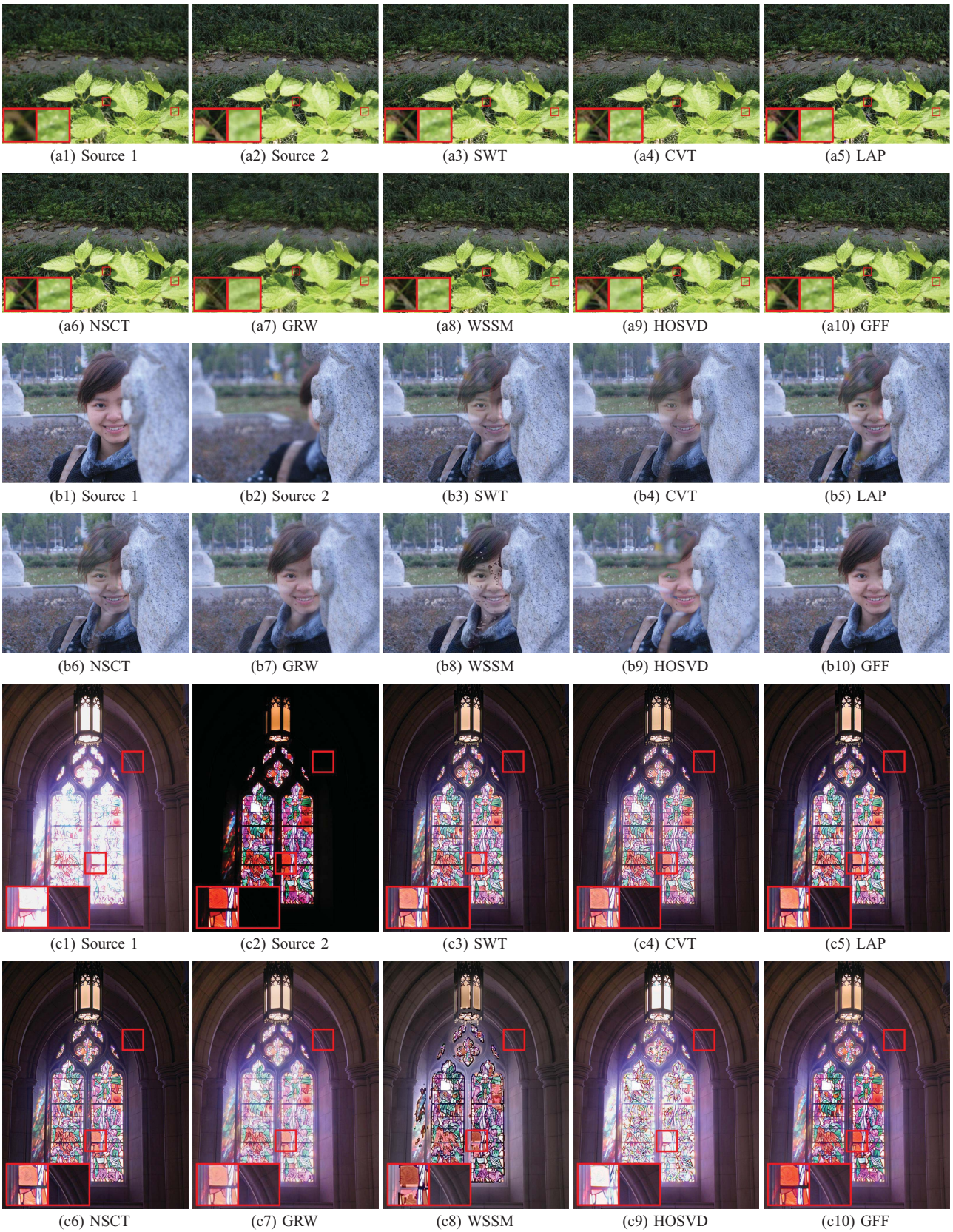


Fig. 10. Color source images and the fused images obtained by different methods.

close to one of the source images (fusion is turned off), the Q_{MI} value will also be very large and thus a very big Q_{MI} is not always a good thing. In our opinion, the five quality metrics should be considered together to evaluate the real fusion performance. It should be noticed that the HOSVD based method gives a very unstable and bad performance of

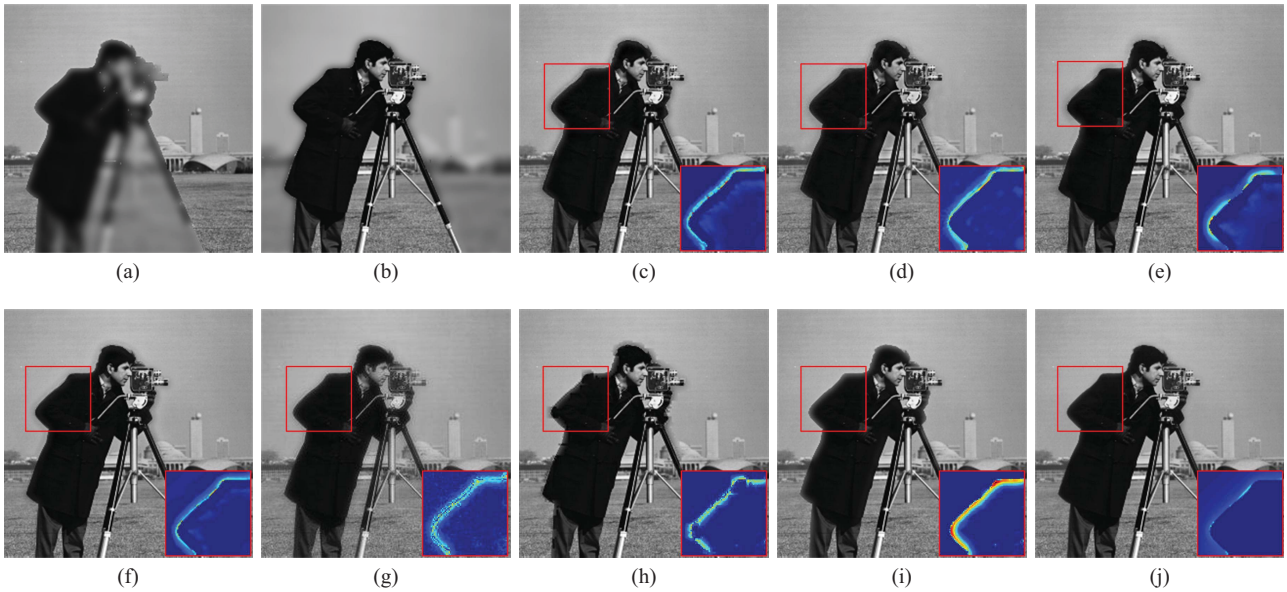


Fig. 11. Artificial multi-focus images and fused images obtained by different methods. The inset box in the lower right corner is the difference image between the fused image and the reference image (visualized by using a color map). (a) Source 1. (b) Source 2. (c) SWT. (d) CVT. (e) LAP. (f) NSCT. (g) GRW. (h) WSSM. (i) HOSVD. (j) GFF.

Fig. 12. Performance comparison of different weight map construction methods. (a) AVG-ABS. (b) AVG-GD. (c) GB-ABS. (d) GB-GD.

Q_C (ranking as the second worst) for the Petrović image database. Moreover, the Q_P and Q_G of the HOSVD based method are also relatively bad (ranking as the fourth and fifth, respectively) for the multi-exposure and multi-modal database. This means that the HOSVD based method may perform not good in some aspects. As shown in Figs. 9 and 10, the HOSVD based method fails in preserving the important complementary information in most of the examples. By contrast, although the objective performance of the GFF method is not always the best, it has a very stable performance in terms of all five quality metrics (always ranking as top second). Thus, it is demonstrated that state-of-the-art fusion performance can be achieved by the GFF method.

2) Comparison With Classic Weight Construction Methods: Fig. 12 shows the fused images obtained by different weight construction approaches, i.e., taking the average of the base layers (AVG), selecting the absolute maximum of the detail layers (ABS), the proposed guided filtering based base layer fusion method (GB), and the proposed guided filtering based detail layer fusion method (GD). As shown in Fig. 12(a) and (c), the AVG-ABS and GB-ABS methods produce artificial points and edges in the window area. This means that the ABS scheme is not qualified for the proposed method since it cannot produce edge aligned weights which can effectively avoid edge artifacts. Further, as shown in Fig. 12(a) and (b), the wall area looks quite dark in the fusion results obtained by the AVG-GD and AVG-ABS methods. It demonstrates that the AVG scheme is not a good method for fusing the base layers since it may produce color and brightness distortions. However, the fused image obtained by the GB-GD method (see Fig. 12(d)) can well preserve the color and edge information of the source images shown in Fig. 10(c1) and (c2). Therefore, the proposed weight construction method is quite helpful in improving fusion performance. Specifically, the GB method ensures the well preservation of brightness and color information, while the GD method ensures that the details of source images can be well preserved without causing artifacts.

The proposed weight construction approach is also compared against classical fusion rules through using fusion quality indexes. Experiments are performed on all 70 testing images shown in Figs. 4–6. As shown in Table II, the proposed weight construction method gives the largest

TABLE I
QUANTITATIVE ASSESSMENT OF DIFFERENT IMAGE FUSION METHODS. THE NUMBERS IN PARENTHESES DENOTE THE NUMBER OF IMAGE PAIRS THAT THIS METHOD BEATS OTHER METHODS

Source Images	Index	SWT	CVT	LAP	NSCT	GRW	WSSM	HOSVD	GFF
Petrović database	Q_Y	0.862(0)	0.813(0)	0.868(0)	0.864(0)	0.696(0)	0.809(0)	**0.967(38)**	0.934(12)
	Q_C	0.745(0)	0.724(0)	0.744(0)	0.751(2)	0.645(0)	0.708(1)	0.691(7)	**0.804(40)**
	Q_G	0.632(1)	0.560(0)	0.644(0)	0.633(1)	0.446(0)	0.617(1)	0.648(24)	**0.657(23)**
	Q_P	0.525(2)	0.439(0)	0.516(1)	0.510(0)	0.355(0)	0.347(0)	**0.628(30)**	0.594(17)
	Q_{MI}	0.391(0)	0.380(0)	0.398(0)	0.390(0)	0.383(0)	0.710(5)	**0.910(42)**	0.570(3)
Multifocus database	Q_Y	0.915(0)	0.894(0)	0.922(0)	0.911(0)	0.761(0)	0.877(0)	0.955(4)	**0.964(6)**
	Q_C	0.818(0)	0.798(0)	0.816(0)	0.829(1)	0.724(0)	0.779(0)	**0.847(7)**	0.835(2)
	Q_G	0.681(0)	0.661(0)	0.698(1)	0.673(0)	0.519(0)	0.668(0)	0.685(1)	**0.714(8)**
	Q_P	0.734(0)	0.721(0)	0.772(1)	0.744(0)	0.559(0)	0.698(0)	0.740(0)	**0.801(9)**
	Q_{MI}	0.849(0)	0.814(0)	0.904(0)	0.840(0)	0.778(0)	0.865(1)	**1.063(7)**	0.953(2)
Multiexposure and multimodal database	Q_Y	0.717(0)	0.738(0)	0.792(0)	0.798(0)	0.717(0)	0.827(1)	**0.953(9)**	0.914(0)
	Q_C	0.648(0)	0.674(0)	0.695(0)	0.715(0)	0.674(0)	0.741(2)	0.764(3)	**0.801(5)**
	Q_G	0.605(0)	0.575(0)	0.693(1)	0.672(0)	0.474(0)	0.638(2)	0.620(1)	**0.704(6)**
	Q_P	0.540(0)	0.501(0)	0.602(0)	0.588(0)	0.439(0)	0.362(0)	0.551(3)	**0.661(7)**
	Q_{MI}	0.509(0)	0.538(0)	0.542(1)	0.542(0)	0.552(0)	0.755(0)	**1.015(9)**	0.597(0)

TABLE II
QUANTITATIVE ASSESSMENTS OF DIFFERENT WEIGHT MAP CONSTRUCTION METHODS

Indexes	AVG-ABS	GB-ABS	AVG-GD	GB-GD
Q_Y	0.824	0.853	0.923	**0.934**
Q_C	0.739	0.755	0.807	**0.808**
Q_G	0.651	0.652	**0.672**	**0.672**
Q_P	0.498	0.502	0.631	**0.632**
Q_{MI}	0.489	0.573	0.540	**0.630**

TABLE III
RUNNING TIME IN SECONDS (THE SECOND ROW) AND MEMORY CONSUMPTION IN MB (THE THIRD ROW) OF DIFFERENT ALGORITHMS ON A PAIR OF IMAGES OF SIZE 512×512. ALL METHODS ARE IMPLEMENTED IN MATLAB FOR COMPARISON

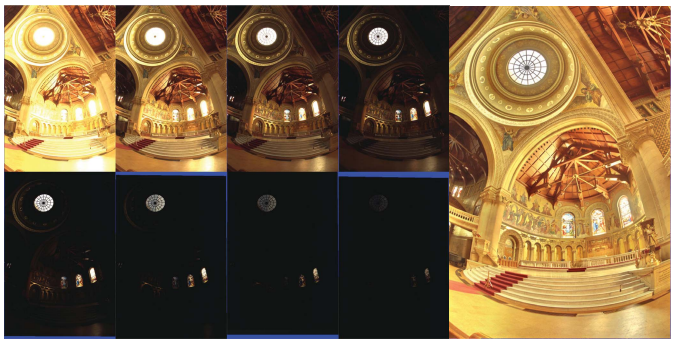
SWT	CVT	LAP	NSCT	GRW	WSSM	HOSVD	GFF
1.61	2.6	0.02	12.16	0.04	155.5	66.5	1.16
7.8	4.2	1.4	22	1.2	4	1.3	2.2

average values of all five fusion metrics. However, it should be noticed that the differences of some objective quality indexes, i.e., Q_Y, Q_C, Q_G, and Q_P are quite small when comparing the AVG-GD method with the GB-GD method. The reason is that the differences between Fig. 12(b) and (d) appear as brightness and color differences. Structure and feature based indexes are not sensitive to such differences and thus fail in detecting the improvement of fusion performance (see Fig. 12(b) and (d)). However, the Q_{MI} metric still works in this situation since it can measure the preservation of brightness information. It can be seen that, compared with the AVG-GD method, the GB-GD method gives much better fusion performance in terms of Q_{MI}. This example shows that there is no objective fusion metric that totally coincides with subjective perception. In other words, different fusion metrics reflect different aspects of a fusion algorithm.

3) Computational Efficiency Analysis: The computing time and memory consumption of different image fusion methods are compared in Table III. Experiments are performed on a computer equipped with a 2.50 GHz CPU and 4 GB memory. The codes of the proposed GFF[2] method, LAP[3], GRW[4], and WSSM[5] based methods are available online. As shown in Table III, the GFF method is not as efficient as the LAP method and the GRW method due to the inefficient MATLAB implementation of the guided filter. Encouragingly, the C++ implementation of the GFF method takes only 0.04 seconds for fusing an image of size $512 \times 512 \times 2$. More importantly, through exploiting integral image technique [31], the GFF method actually has a linear time complexity $O(N)$ and thus can be implemented in real-time through GPU programing. Furthermore, the memory consumption (the maximum of data which needs to be kept in memory) is also an important factor affecting the computing time. If the input image is of size 512×512, a four-scale NSCT will transform it into a multi-scale and multi-orientation representation of size $512 \times 512 \times 37$. In contrast, for an image of the same size, the GFF method transforms it into a two-scale representation of size $512 \times 512 \times 2$ which is very small.

4) Color Image Sequence Fusion: Furthermore, Fig. 13 gives a multi-exposure image sequence fusion example. Fig. 13(a) shows eight images of the multi-exposure sequence. As shown in Fig. 13(b), the fused image obtained by the GFF method can well preserve the color and detail information of different exposures. This example demonstrates that the GFF method works well when there are more than two source images.

[2]http://xudongkang.weebly.com
[3]http://www.metapix.de/toolbox.htm
[4]http://www.ualberta.ca/~rshen/papers/tip11/
[5]http://www.mathworks.com/matlabcentral/fileexchange/36231



(a) (b)

Fig. 13. Image sequence fusion with more than two source images. (a) Eight images of the original multiexposure sequence (sixteen images in total). (b) Fused image of the sequence. Image courtesy of Paul Debevec.

V. Conclusion

We have presented a novel image fusion method based on guided filtering. The proposed method utilizes the average filter to get the two-scale representations, which is simple and effective. More importantly, the guided filter is used in a novel way to make full use of the strong correlations between neighborhood pixels for weight optimization. Experiments show that the proposed method can well preserve the original and complementary information of multiple input images. Encouragingly, the proposed method is very robust to image registration. Furthermore, the proposed method is computationally efficient, making it quite qualified for real applications. At last, how to improve the performance of the proposed method by adaptively choosing the parameters of the guided filter can be further researched.

Acknowledgment

The authors would like to thank the Editor Dr. Brendt Wohlberg and anonymous reviewers for their insightful comments and suggestions, which have greatly improved this paper.

References

[1] A. A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Inf. Fusion*, vol. 8, no. 2, pp. 114–118, Apr. 2007.

[2] D. Socolinsky and L. Wolff, "Multispectral image visualization through first-order fusion," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 923–931, Aug. 2002.

[3] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, Dec. 2011.

[4] S. Li, J. Kwok, I. Tsang, and Y. Wang, "Fusing images with different focuses using support vector machines," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1555–1561, Nov. 2004.

[5] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[6] D. Looney and D. Mandic, "Multiscale image fusion using complex extensions of EMD," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1626–1630, Apr. 2009.

[7] M. Kumar and S. Dass, "A total variation-based algorithm for pixel-level image fusion," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2137–2143, Sep. 2009.

[8] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[9] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proc. Int. Conf. Image Process.*, vol. 3, Washington, DC, USA, Oct. 1997, pp. 288–291.

[10] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2898–2909, May 2012.

[11] M. Xu, H. Chen, and P. Varshney, "An image fusion approach based on markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5116–5127, Dec. 2011.

[12] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, Heraklion, Greece, Sep. 2010, pp. 1–14.

[13] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67-1–67-10, Aug. 2008.

[14] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, Jul. 2002.

[15] N. Draper and H. Smith, *Applied Regression Analysis*. New York, USA: Wiley, 1981.

[16] V. Petrović, "Subjective tests for image fusion evaluation and objective metric validation," *Inf. Fusion*, vol. 8, no. 2, pp. 208–216, Apr. 2007.

[17] G. Piella, "Image fusion for enhanced visualization: A variational approach," *Int. J. Comput. Vision*, vol. 83, pp. 1–11, Jun. 2009.

[18] S. Li, X. Kang, J. Hu, and B. Yang, "Image matting for fusion of multi-focus images in dynamic scenes," *Inf. Fusion*, vol. 14, no. 2, pp. 147–162, 2013.

[19] L. Tessens, A. Ledda, A. Pizurica, and W. Philips, "Extending the depth of field in microscopy through curvelet-based frequency-adaptive image fusion," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 1, Apr. 2007, pp. 861–864.

[20] Q. Zhang and B. Guo, "Multifocus image fusion using the non-subsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, Jul. 2009.

[21] J. Tian and L. Chen, "Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure," *Signal Process.*, vol. 92, no. 9, pp. 2137–2146, Sep. 2012.

[22] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'information measure for performance of image fusion'," *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.

[23] C. Yang, J. Zhang, X. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, no. 2, pp. 156–160, Apr. 2008.

[24] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *Int. J. Signal Process.*, vol. 2, no. 3, pp. 178–182, Apr. 2005.

[25] C. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

[26] J. Zhao, R. Laganiere, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *Int. J. Innovative Comput. Inf. Control*, vol. 3, no. 6, pp. 1433–1447, Dec. 2007.

[27] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.

[28] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, Mar. 2002.

[29] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[31] F. C. Crow, "Summed-area tables for texture mapping," in *Proc. SIGGRAPH 84. 11th Annu. Conf. Comput. Graph. Interact. Tech.*, vol. 18, no. 3, pp. 207–212, Jan. 1984.

Shutao Li (M'07) received the B.Sc., M.Sc, and Ph.D degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively. In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, Egham, U.K., working with Prof. J.-S. Taylor. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 160 refereed papers.

His current research interests include information fusion, pattern recognition, and image processing.

Xudong Kang (S'12) received the B.Sc. degree from Northeast University, Shenyang, China, in 2007. He is pusuring the Ph.D degree in electrical engineering from Hunan University, Chansha, China.

His current research interests include image fusion, image super-resolution, pansharpening, and hyperspectral image classification.

Jianwen Hu received the B.Sc. degree from Inner Mongolia University of Science and Technology, Baotou, China, in 2008. He is currently pursuing the Ph.D degree in electrical engineering from the Hunan University, Changsha, China.

His research interests include image processing, information fusion, and sparse representation.