Fast filtering image fusion

Kun Zhan
Yuange Xie
Haibo Wang
Yufang Min

Fast filtering image fusion

Kun Zhan,[a,*] **Yuange Xie,**[a] **Haibo Wang,**[a] **and Yufang Min**[b]
[a]Lanzhou University, School of Information Science and Engineering, Lanzhou, Gansu, China
[b]Chinese Academy of Sciences, Northwest Institute of Eco-Environment and Resources, Lanzhou, Gansu, China

Abstract. Image fusion aims at exploiting complementary information in multimodal images to create a single composite image with extended information content. An image fusion framework is proposed for different types of multimodal images with fast filtering in the spatial domain. First, image gradient magnitude is used to detect contrast and image sharpness. Second, a fast morphological closing operation is performed on image gradient magnitude to bridge gaps and fill holes. Third, the weight map is obtained from the multimodal image gradient magnitude and is filtered by a fast structure-preserving filter. Finally, the fused image is composed by using a weighed-sum rule. Experimental results on several groups of images show that the proposed fast fusion method has a better performance than the state-of-the-art methods, running up to four times faster than the fastest baseline algorithm. © 2017 SPIE and IS&T [DOI: 10.1117/1.JEI.26.6.063004]

1 Introduction

Because different sensors have different imaging principles and optical lenses are confined by focal lengths and capture ranges, multimodal images of the same scene contain complementary features, such as edges, shapes, textures, etc. Image fusion aims to retain and integrate useful information from multimodal images into a composite image for interpretation, and a fused image with comprehensive details of the scene can be perceived more appropriately by humans and machines.

Filtering is the most fundamental process of image fusion, and it is very important to exploit structures and details within images for fusion. The basic image features contain shape, edge, and texture. Shape and edge are determined by its structures while texture is its details. Multiscale transforms are usually applied to image fusion for basic image feature extraction.[1–4] To exploit these features, Laplacian pyramid decompositions are first applied to image fusion.[5–7] Later, wavelet transforms and other multiscale transforms are widely used to fuse images from different sensing modalities.[8–10] The transform coefficients are determined by the predefined decomposition levels, which results in a few scales of details being represented with high-frequency coefficients in the transform domain.[11] These details may not have a direct relationship with original semantic features. What is more, fusion results of these transforms usually suffer from ringing effects and "halo" artifacts around the major structures because of using high-pass filtering.[12,13] In the spatial domain, the conventional spatial domain low-pass filters, e.g., Gaussian filter, can smooth texture but also structures. To prevent smoothing across structures while still smoothing texture, many structure-preserving filters have been developed recently, such as anisotropic diffusion,[14] bilateral filter,[15] weighted least squares filter,[16] L_0-smoothing filter,[17] guided filter,[18] etc. These structure-preserving filters can be used to realize multiscale decomposition as well as a Laplacian pyramid decomposition.[19–22] These methods combine structure-preserving filters into multiscale transforms, but they have a relatively high computational complexity and the original intensities of source images are not preserved in fusion results.[1–4]

Using anisotropic diffusion, a multiscale transform method is proposed for multimodal medical image fusion, and a data-specific multiscale geometrical analysis kernel is formulated for the decomposition.[20] However, anisotropic diffusion-based methods have a high computational complexity and tend to over-sharpen structures.[19,20] Farbman et al.[16] use the weighted least squares filter to construct a multiscale image decomposition for fusion multiexposure images, and they have demonstrated the weighted least squares filter is well-suited for progressive coarsening of images for multiscale detail extraction. The weighted least squares-based method requires the solution of a sparse linear system, which limits the performance of the technique.[23] Jointly using the bilateral filter and the nonsubsampled directional filter bank, Hu and Li[24] construct a multiscale representation for multisensor image fusion, and they mainly leverage the edge-preserving characteristic of the bilateral filter and the image directional feature extracted by the filter bank. Bilateral filter-based methods usually involve artifacts around the edge, e.g., a halo artifact. Zhao et al.[25] utilize L_0-smoothing filter to construct a multiscale decomposition method for image fusion, and the saliency extraction-based visual weight map is decomposed by L_0-smoothing filter. The L_0-smoothing filter-based method tends to lose small-scale details since it is a global optimization algorithm, which can preserve only the salient detail information.[20] The guided filter is applied to refine the weight map obtained by a two-scale decomposition, and images are fused by a weight-averaging rule.[22] The challenge of the guided filter-based method is the structure inconsistency between the guidance image and the target image.[26] These methods

*Address all correspondence to: Kun Zhan, E-mail: kzhan@lzu.edu.cn

combine the structure-preserving filter into multiscale transforms, and the major advantage of these kinds of methods is the ability to accurately separate fine-scale texture details, middle-scale edges, and large-scale spatial structures of an image. Because of the multiscale decomposition framework, they have a relatively high computational complexity, and the original intensities of source images are not preserved in fusion results.[1–4]

In this paper, we use fast filtering for image fusion in the spatial domain. In the spatial domain, most scales within images are processed by these structure-preserving filters simultaneously, i.e., details information is smoothed while structures are preserved. Due to the property, even the semantic information can be extracted from the low-level image.[27] Spatial domain fusion methods are different from fusing in the transform domains because multiscale transforms capture only limited image scales determined by the decomposition levels. In the spatial domain, we can directly process pixels rather than processing the transform domain coefficients. We mainly develop a fast structure-preserving filter and a fast morphological filtering to exploit the intrinsic structures in multimodal source images. Intrinsic structures have large scales and details have small scales. Because of the development of structure-preserving filters, the scales can be well processed by them. The proposed method is compared to the state-of-the-art methods, such as methods based on nonsubsampled contourlet transform (NSCT),[28] guided filter fusion (GFF),[22] multiscale transform using sparse representation (MSSR),[29] gradient transfer fusion (GTF),[30] and cross bilateral filter (CBF).[31] The experimental results indicate that the performance of the proposed fast filtering scheme is better than the state-of-the-art methods in terms of four performance measures and it is the fastest algorithm.

2 Fast Filtering

2.1 Structure-Preserving Filtering

Suppose that an input image is denoted by $I_p \in \mathbb{R}^{m \times n}$, where p is the pixel index and the image I_p has a mean of μ_k in a sliding window Ω_k centered at the pixel k, then, a zero-mean signal J_p is defined in Ω_k by

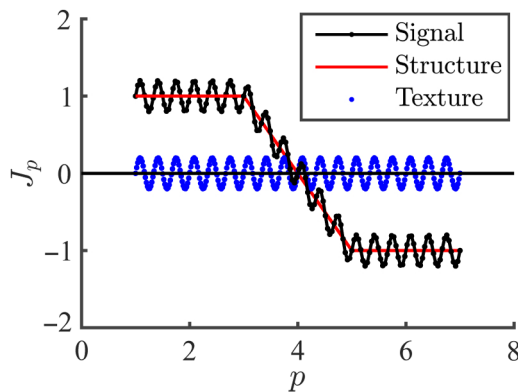$$J_p = I_p - \mu_k, \quad p \in \Omega_k. \tag{1}$$

Fig. 1 A 1-D example of a signal. The signal is decomposed to structure and texture.

As shown in Fig. 1, the texture T_p is usually a zero-mean signal. Then, the main structure of the image S_p is given by

$$S_p = J_p - T_p. \tag{2}$$

The structure S_p can be approximated from the input image J_p using a linear estimator[32]

$$\hat{S}_p = h_p \otimes J_p, \tag{3}$$

where $\otimes$ denotes the convolution and h_p denotes the transfer function of the linear system while J_p is its input signal and $\hat{S}_p$ is its output signal.

In a linear shift-invariant system, the signal estimation problem can be solved using the orthogonality principle,[32] i.e., the error $e_p = S_p - \hat{S}_p$ is perpendicular to J_p

$$E(e_p J_q) = 0. \tag{4}$$

Then

$$\begin{aligned} E(S_p J_q) &= E(\hat{S}_p J_q) \\ &= E(h_p \otimes J_p J_q) \\ &= \int_{-\infty}^{\infty} h_\tau E(J_{p-\tau} J_q) d\tau. \end{aligned} \tag{5}$$

The above equation can be rewritten by

$$R_p^{sj} = h_p \otimes R_p^j, \tag{6}$$

where R_p^{sj} is the cross correlation of S_p and J_p, and R_p^j is the autocorrelation of J_p.

Suppose that S_p is uncorrelated with T_p,[32] then, we have

$$R_p^{st} = 0. \tag{7}$$

Considering Eqs. (2) and (7), we have

$$R_p^{sj} = R_p^s = R_p^j - R_p^t. \tag{8}$$

It is straightforward to check that h_p can be obtained using Eqs. (6) and (8)

$$h_p \otimes R_p^j = R_p^j - R_p^t. \tag{9}$$

According to the correlation theorem,[33] we have

$$h_p = \frac{\sigma_k^2 - \sigma_\eta^2}{\sigma_k^2} \delta_p, \tag{10}$$

where σ_k^2 is the variance of J_p and σ_η^2 is the variance of T_p. Substituting Eq. (10) into Eq. (3), we have

$$\hat{S}_p = \frac{\sigma_k^2 - \sigma_\eta^2}{\sigma_k^2} J_p \quad = \frac{\sigma_k^2 - \sigma_\eta^2}{\sigma_k^2} (I_p - \mu_k). \tag{11}$$

Then, we use the estimated signal $\hat{S}_p$ to recover the signal $\hat{I}_p$ by

$$\hat{I}_p = \mu_k + \hat{S}_p = \mu_k + \frac{\sigma_k^2 - \sigma_\eta^2}{\sigma_k^2}(I_p - \mu_k). \tag{12}$$

Here, σ_k^2 is known while σ_η^2 is unknown.

Let v denote $\frac{\sigma_k^2 - \sigma_\eta^2}{\sigma_k^2}$

$$\hat{I}_p = \mu_k + \hat{S}_p = \mu_k + v(I_p - \mu_k). \tag{13}$$

We minimize the mean square error between I_p and $\hat{I}_p$ while maintaining the linear model [Eq. (13)]. Specifically, we seek a solution that minimizes the objective function in the window Ω_k

$$\min_v \sum_{p \in \Omega_k} \{[\mu_k + v(I_p - \mu_k) - I_p]^2 + \lambda v^2\}. \tag{14}$$

Here, λ is a regularization parameter penalizing large v.

Equation (14) is the linear ridge regression model, and its solution is given by

$$v = \frac{\sigma_k^2}{\sigma_k^2 + \lambda}. \tag{15}$$

Substituting Eq. (15) into Eq. (13), we have

$$\hat{I}_p = \mu_k + \frac{\sigma_k^2}{\sigma_k^2 + \lambda}(I_p - \mu_k), \quad p \in \Omega_k. \tag{16}$$

It is straightforward to check that pixels with variance larger than λ are preserved by Eq. (16), whereas patches with variance smaller than λ are smoothed. Equation (16) is a structure-preserving filter, which ensures that $\hat{I}_p$ has a structure only if I_p has a structure because of $\nabla \hat{I}_p = v \nabla I_p$.

If the intensity of a structure with very large variance σ_k^2 always changes sharply within Ω_k, then the structure can be preserved, i.e., if $\sigma_k^2 \gg \lambda$, then we have $\frac{\sigma_k^2}{\sigma_k^2 + \lambda} \approx 1$ and $\hat{I}_p \approx I_p$. The intensity is preserved while the pixel belongs to a main structure.
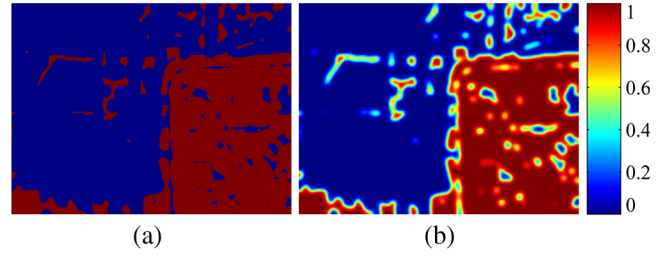
Fig. 3 Weight map: (a) before filtering, w_p; and (b) after filtering, $\hat{w}_p$.

Algorithm 2 Fast filtering image fusion.

Input: Input source images $I^{(1)} \in \mathbb{R}^{m \times n}$ and $I^{(2)} \in \mathbb{R}^{m \times n}$, where the superscript denotes the image index, radius r, and parameter λ.

Output: Fused image F.

1: Calculate the element number in each box window by
$c = \text{boxfilter}[\text{ones}(m, n), r]$.

2: **for** $q \in \{1, 2\}$ **do**

3:     Detect contrast and image sharpness by the gradient magnitude
$M = |\frac{\partial I^{(q)}}{\partial x}| + |\frac{\partial I^{(q)}}{\partial y}|$.

4:     Normalize M to $d \in [0, 1]$ by $d = \frac{M - \min(M)}{\max(M) - \min(M)}$.

5:     Perform the dilation operation by $d = \frac{\text{boxfilter}(d, r)}{c}$.

6:     Perform the closing operation by $g^{(q)} = 1 - \frac{\text{boxfilter}(1 - d, r)}{c}$.

7: **end for**

8: Obtain the weight matrix by $w = \text{step}[g^{(1)}, g^{(2)}]$.

9: Smooth the weight matrix by $w = \frac{\text{boxfilter}(w, r)}{c}$.

10: Calculate the mean of the weight matrix by $\mu = \frac{\text{boxfilter}(w, r)}{c}$.

11: Calculate the variance of the weight matrix by
$\sigma^2 = \frac{\text{boxfilter}(w. * w, r)}{c} - \mu. * \mu$.

12: Perform the structure-preserving filtering by
$\hat{w} = \mu + \frac{\sigma^2}{\sigma^2 + \lambda}. * (w - \mu)$.

13: Obtain the fused image by $F = \hat{w} I^{(1)} + (1 - \hat{w}) I^{(2)}$.

Algorithm 1 Structure-preserving filtering.

Input: Filtering input image I, radius r of the window Ω_k, and regularization λ.

Output: Filtering output image $\hat{I}$.

1: $\mu = f_{\text{mean}}(I)$;

2: $\sigma^2 = f_{\text{mean}}(I. * I) - \mu^2$;

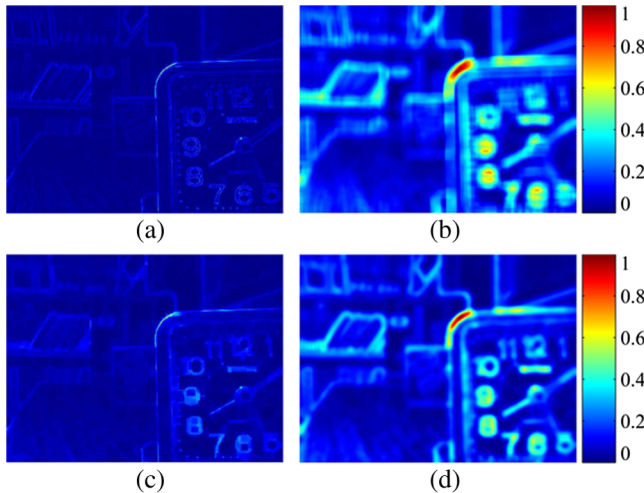3: $\hat{I} = \mu + \frac{\sigma^2}{\sigma^2 + \lambda}(I - \mu)$.

Fig. 2 (a) Gradient magnitude, d; (b) d is smoothed by averaging; (c) d is filtered by image closing operation; and (d) d is filtered by Eq. (23).

If the intensity is always not changed a lot in a textural region with much smaller variance σ_k^2 than structures, then these regions are smoothed by the linear mean filter, i.e., if $\sigma_k^2 \ll \lambda$, then we have $\hat{I}_p \approx \mu_k$ and $\frac{\sigma_k^2}{\sigma_k^2 + \lambda} \approx 0$. The intensity is smoothed by the mean filter, which is the most simple smoothing filter.[12]

2.2 Morphological Filtering

Salient structure detection may produce gaps at a continuous curve and holes in a homogeneous region. The morphological closing operation can well-bridge gaps and fill holes, and it is feasible to refine the structures of the images. The morphological closing operation can be expressed by

$$g_p = (d_p \oplus s) \ominus s, \tag{17}$$

where d_p is an input image, g_p is the output image, $\oplus$ denotes the morphological dilation operation, $\ominus$ denotes the morphological erosion operation, and s is a structuring element object.

According to the duality property of the morphological operations, dilation and erosion are duals of each other with respect to the complement operation, i.e., $A \ominus B = (A^c \oplus B)^c$, where A^c is the complement of A. Then, Eq. (17) can be rewritten by

$$g_p = [(d_p \oplus s)^c \oplus s]^c. \tag{18}$$

If the input image d_p is normalized in the range of [0, 1], then Eq. (18) can be expressed in the following equivalent form:

$$g_p = 1 - [1 - (d_p \oplus s)] \oplus s. \tag{19}$$

However, the dilation operation is a time-consuming procedure due to calculating the maximum value in the sliding structuring element s. It is straightforward to check that the convolution can obtain comparable results as similar as the dilation.[12] To accelerate the procedure, we use the convolution operation instead of the dilation operation. For the closing operation, we first perform the convolution on the normalized image to obtain the similar result of the dilation, and the similar operation of Eq. (19) is given by

$$g_p = 1 - [1 - (d_p \otimes s)] \otimes s. \tag{20}$$

2.3 Computation and Efficiency

The filtering process of Eq. (16) can be given by Algorithm 1. In Algorithm 1, $f_{mean}(\cdot)$ is a mean filter with a window of radius r. The box filter is equivalent to the convolution between an image and a square matrix of all ones,[18] and it can be computed efficiently in an $O(N)$ (where N is the pixel number) complexity using the integral image technique.[34] With the $O(N)$ time mean filter, the structure-preserving filter [Eq. (16)] is naturally $O(N)$ time.

Table 1 Quantitative evaluation of different fusion methods.

Images	Metrics	NSCT	GFF	MSSR	GTF	CBF	FFIF
Disk	$Q_p^{ab|f}$	0.6837	0.7054	0.6936	0.6279	0.6819	0.7124
	$Q_w^{xy|f}$	0.9243	0.9405	0.9304	0.9367	0.9231	0.9583
	Q_{MI}	0.8931	0.9732	0.9795	0.8878	0.9457	1.0573
	Q_{NCIE}	0.8244	0.8301	0.8313	0.8257	0.8275	0.8353
	Time (s)	79.4536	0.3692	69.6018	7.3069	50.1227	0.1024
Pepsi	$Q_p^{ab|f}$	0.7573	0.7866	0.7712	0.7530	0.7676	0.7901
	$Q_w^{xy|f}$	0.9533	0.9555	0.9470	0.9600	0.9421	0.9629
	Q_{MI}	0.9893	1.0393	1.0152	0.9567	1.0244	1.1332
	Q_{NCIE}	0.8295	0.8319	0.8323	0.8307	0.8309	0.8382
	Time (s)	67.9367	0.3147	64.1769	8.4516	42.5323	0.0879
Lab	$Q_p^{ab|f}$	0.7128	0.7143	0.7156	0.5822	0.7159	0.7197
	$Q_w^{xy|f}$	0.8670	0.9520	0.8871	0.8623	0.9329	0.9703
	Q_{MI}	1.0581	1.1661	1.1753	0.8342	1.1168	1.2339
	Q_{NCIE}	0.8321	0.8377	0.8382	0.8307	0.8349	0.8413
	Time (s)	86.5337	0.4317	57.6864	8.9642	50.0641	0.1030

Table 1 (*Continued*).

Images	Metrics	NSCT	GFF	MSSR	GTF	CBF	FFIF
Natocamp	$Q_p^{ab\|f}$	0.4212	0.4547	0.4404	0.3493	0.3670	0.5139
	$Q_w^{xy\|f}$	0.7634	0.8287	0.7504	0.7469	0.6464	0.9124
	Q_{MI}	0.2198	0.2424	0.3378	0.3380	0.2287	0.4611
	Q_{NCIE}	0.8033	0.8035	0.8044	0.8054	0.8035	0.8081
	Time (s)	19.6777	0.0459	9.7825	0.6911	12.5096	0.0157
Sandpath	$Q_p^{ab\|f}$	0.3862	0.5048	0.4507	0.3386	0.3377	0.6001
	$Q_w^{xy\|f}$	0.7926	0.9088	0.8359	0.8014	0.7146	0.9978
	Q_{MI}	0.1184	0.3199	0.3227	0.2193	0.2081	1.0095
	Q_{NCIE}	0.8024	0.8052	0.8058	0.8036	0.8034	0.8558
	Time (s)	67.4090	0.3238	60.4524	4.7001	44.1827	0.0874
Street	$Q_p^{ab\|f}$	0.5362	0.6530	0.6263	0.4473	0.5456	0.6707
	$Q_w^{xy\|f}$	0.6385	0.8590	0.8349	0.4260	0.6523	0.9767
	Q_{MI}	0.2015	0.2865	0.5563	0.2773	0.3062	0.8859
	Q_{NCIE}	0.8029	0.8039	0.8112	0.8034	0.8043	0.8327
	Time (s)	18.4305	0.0529	14.8249	0.6088	10.6250	0.0111
Medical A	$Q_p^{ab\|f}$	0.6315	0.6443	0.6668	0.6058	0.5535	0.6776
	$Q_w^{xy\|f}$	0.3995	0.8340	0.8231	0.7638	0.8673	0.9699
	Q_{MI}	0.5822	0.7303	0.8006	0.6965	0.7241	0.8536
	Q_{NCIE}	0.8079	0.8095	0.8103	0.8090	0.8099	0.8165
	Time (s)	17.1328	0.0418	17.8447	0.4422	10.5611	0.0091
Medical B	$Q_p^{ab\|f}$	0.5668	0.5999	0.6183	0.5135	0.6263	0.6450
	$Q_w^{xy\|f}$	0.4578	0.8559	0.7508	0.6594	0.9331	0.9688
	Q_{MI}	0.5618	0.6956	0.7420	0.6463	0.7766	0.9867
	Q_{NCIE}	0.8072	0.8084	0.8093	0.8072	0.8096	0.8114
	Time (s)	17.3855	0.0398	36.6682	0.5387	10.4869	0.0092
Medical C	$Q_p^{ab\|f}$	0.6296	0.6660	0.6459	0.5239	0.6210	0.6720
	$Q_w^{xy\|f}$	0.4023	0.8843	0.7635	0.7718	0.8841	0.9663
	Q_{MI}	0.5555	0.7426	0.7051	0.6518	0.7424	0.8915
	Q_{NCIE}	0.8070	0.8085	0.8082	0.8070	0.8087	0.8111
	Time (s)	18.5469	0.0472	11.3064	0.4648	10.5045	0.0110

The filtering process of Eq. (20) also can be realized using the box filter[18,34] when the structuring element s is set to a flat disk-shaped structuring element with a specified radius. Then, Eq. (20) is performed in an $O(N)$ time too in this paper.

3 Image Fusion Scheme

The same region in different modality images has a different contrast and image sharpness. In the spatial domain, the gradient magnitude is a fast and simple way to detect the contrast and sharpness in intensity. The main reason that we use the gradient magnitude is to speed-up the process of image fusion. Because the gradient image contains rich texture and boundary information of image structure, we use the gradient magnitude to measure the saliency information. Gradient magnitude can accurately position edges and other details, and the use of forward difference can quickly realize a discrete gradient. Suppose that there are two input source images $I^{(1)}$ and $I^{(2)}$, where the superscript denotes the
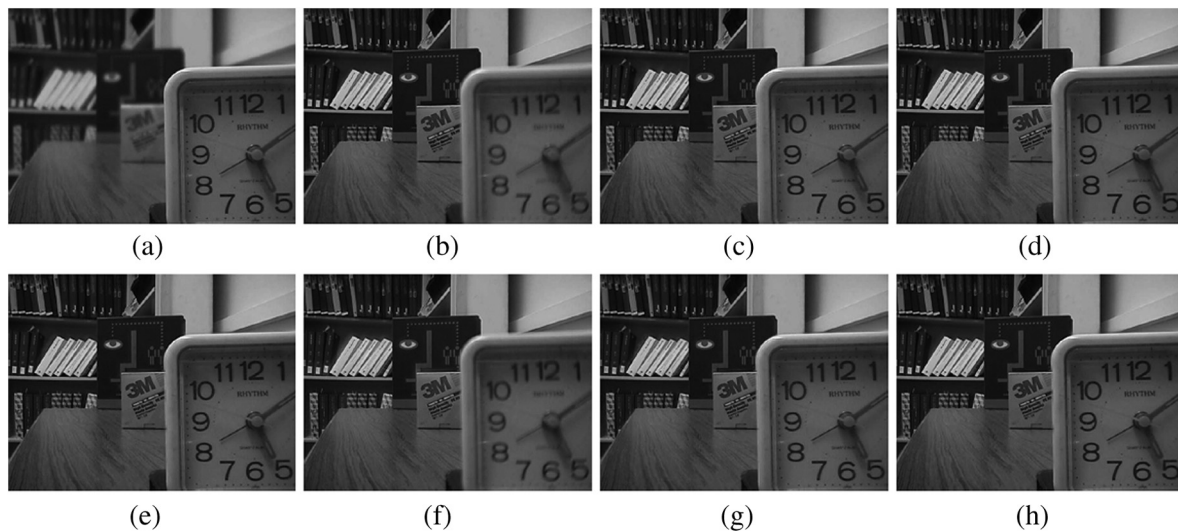

Fig. 4 Multifocus images disk and fused images obtained by different methods: (a) disk 1, (b) disk 2, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.
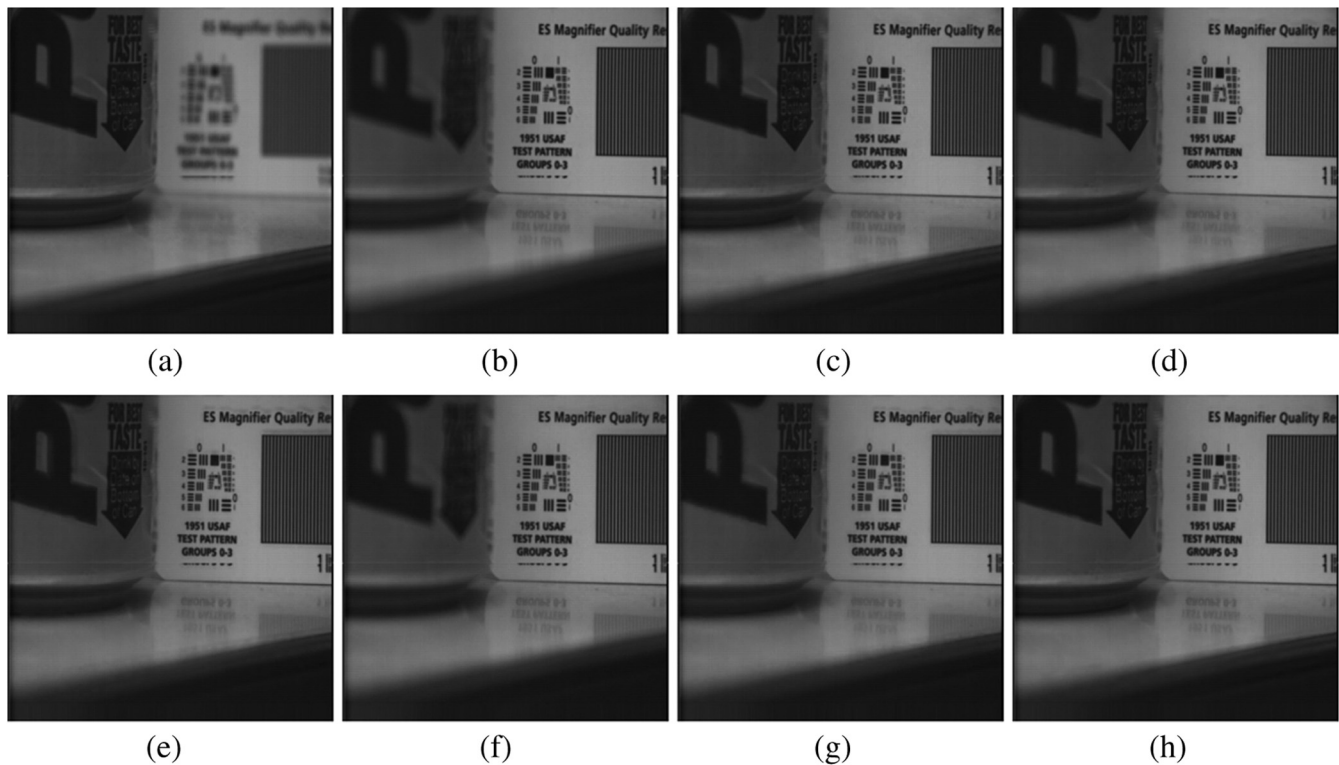

Fig. 5 Multifocus images Pepsi and fused images obtained by different methods: (a) Pepsi 1, (b) Pepsi 2, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.

image index. Approximating the gradient magnitude by absolute values is frequently used to detect contrast and image sharpness

$$M^{(q)} = \left| \frac{\partial I^{(q)}}{\partial x} \right| + \left| \frac{\partial I^{(q)}}{\partial y} \right|, \quad q \in \{1,2\}, \tag{21}$$

where (x, y) are the spatial coordinates, and the digital difference is $\frac{\partial I}{\partial x} = I(x + 1, y) - I(x, y)$.

Due to the postprocessing of the complement operation, a normalization needs to perform on the saliency map. We normalize $M^{(q)}$ to $d^{(q)}$ in the range of [0, 1] with linear function transformation[35]

$$d^{(q)} = \frac{M^{(q)} - \min[M^{(q)}]}{\max[M^{(q)}] - \min[M^{(q)}]}, \quad q \in \{1,2\}. \tag{22}$$

Because the gradient image contains rich texture and boundary information, we use the gradient magnitude for detecting contrast and image sharpness. Furthermore, a morphological closing operation is used to refine the gradient map since the salient structure detection may produce gaps at a continuous curve and holes in a homogeneous region. The morphological filtering can be given by

$$g^{(q)} = 1 - \{1 - [d^{(q)} \otimes s]\} \otimes s, \quad q \in \{1,2\}. \tag{23}$$
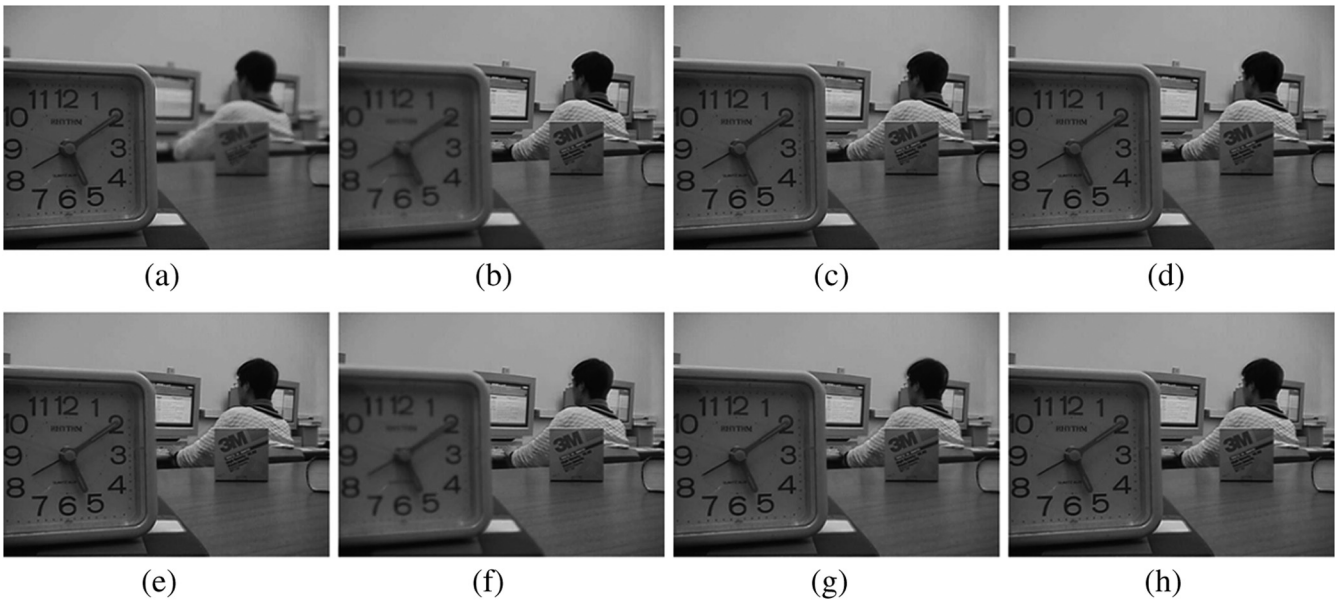

Fig. 6 Multifocus images lab and fused images obtained by different methods: (a) lab 1, (b) lab 2, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.
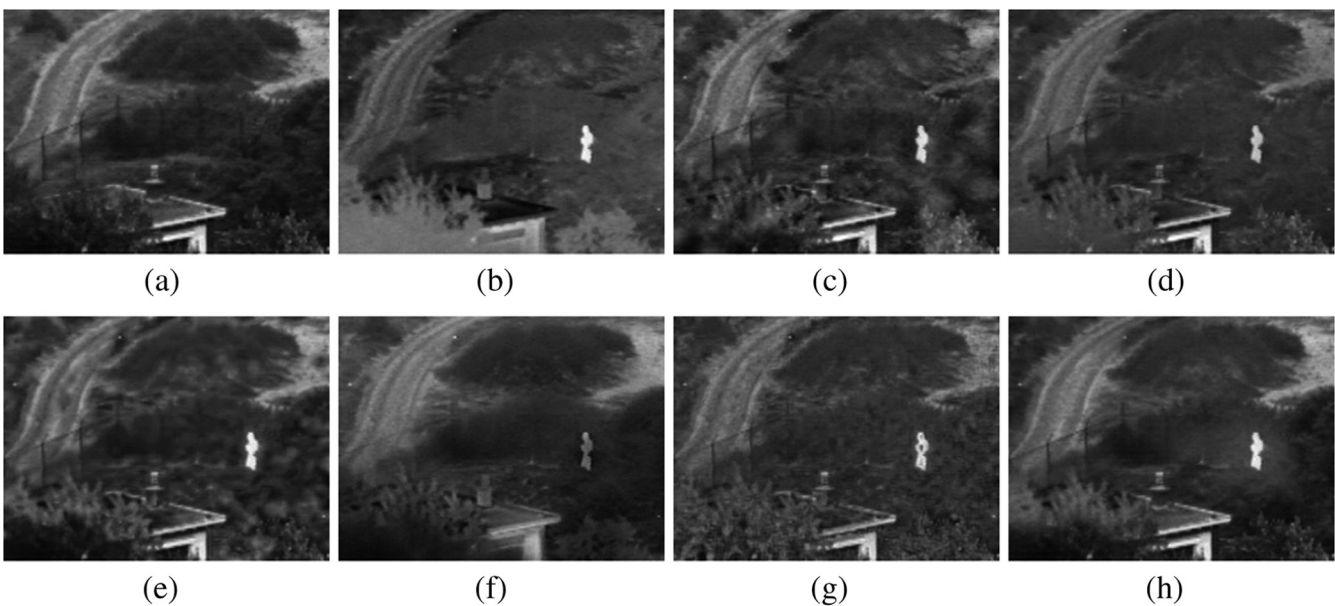

Fig. 7 Visible and infrared images Natocamp and fused images obtained by different methods: (a) visible, (b) infrared, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.

As shown in Fig. 2, the gradient magnitude is filtered by different filters, and the filtered result of Eq. (23) preserves structure well. Morphological filtering is used to bridge gaps and fill holes since the salient structure detection may produce gaps and holes in a homogeneous region. The local average operation may smooth the details and structures simultaneously, which renders edges degraded and blurred. The main difference in the proposed algorithm from the averaging is the digits in Fig. 2, e.g., the eleven "11" in Fig. 2(b), and the proposed algorithm obtains a clearer structure of
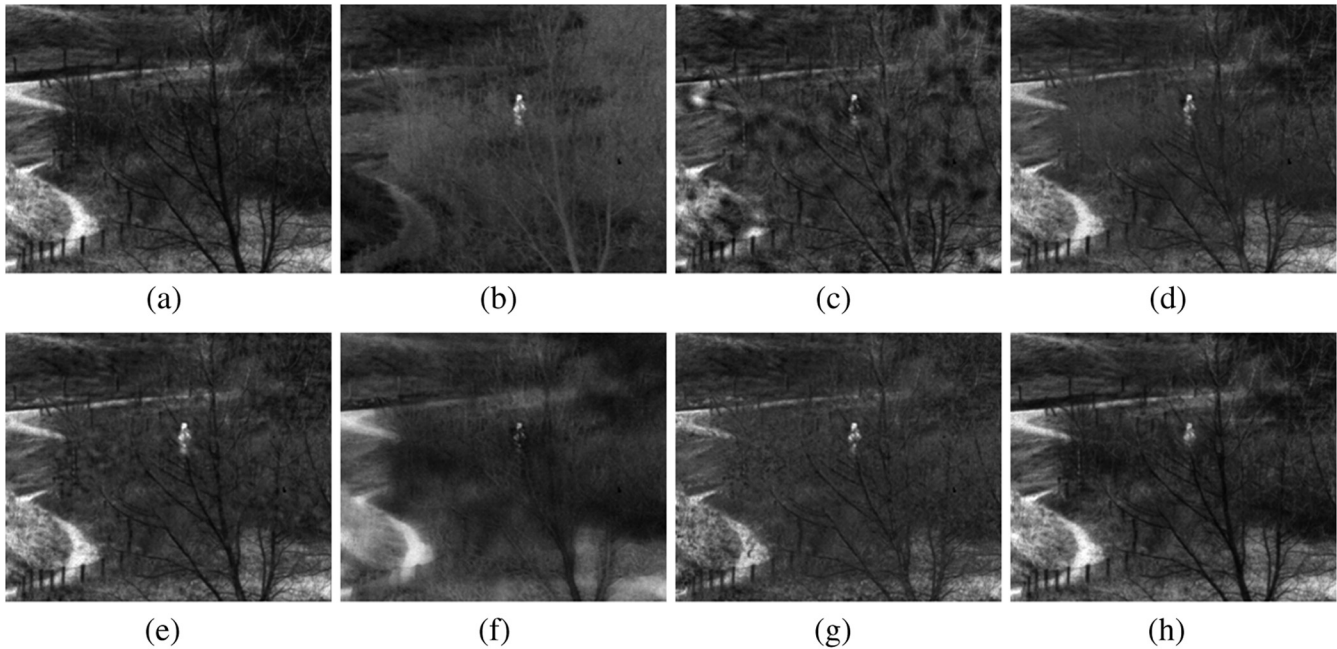

Fig. 8 Visible and infrared images sandpath and fused images obtained by different methods: (a) visible, (b) infrared, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.
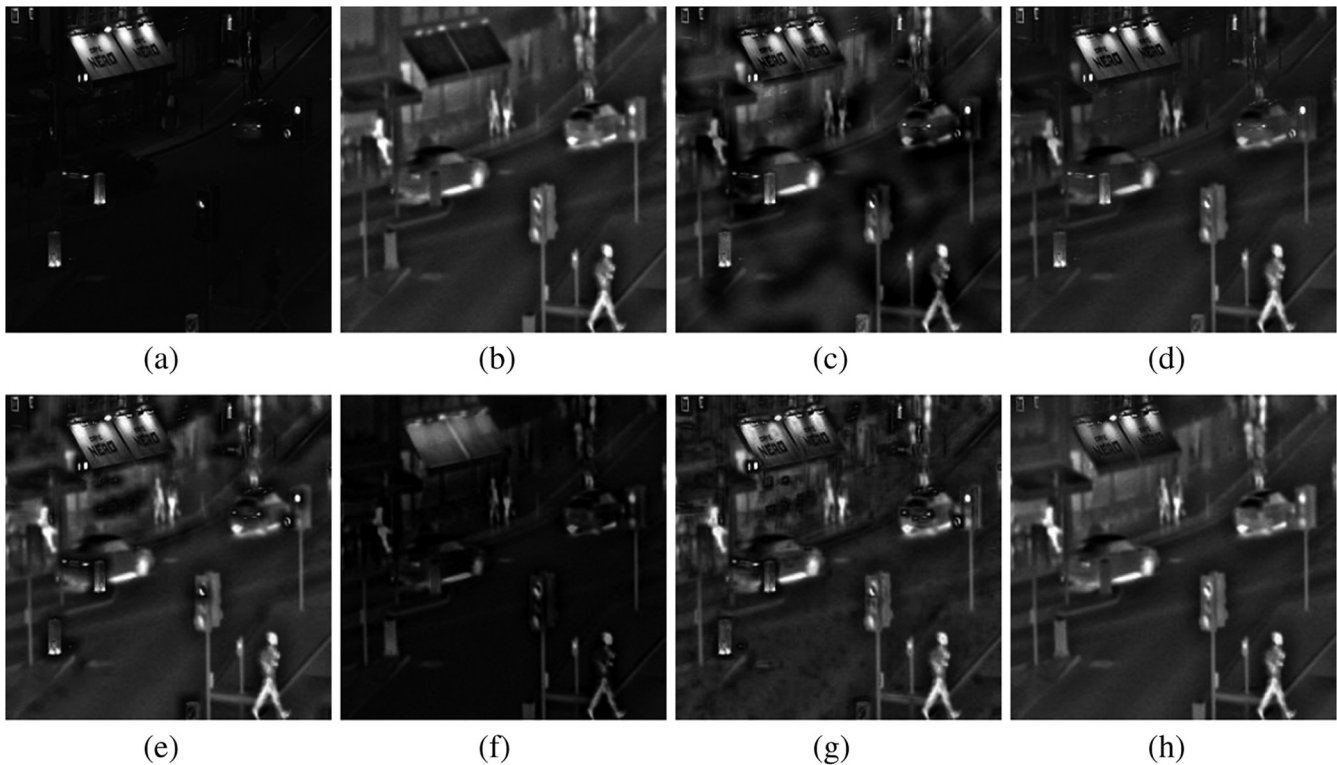

Fig. 9 Visible and infrared images street and fused images obtained by different methods: (a) visible, (b) infrared, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.
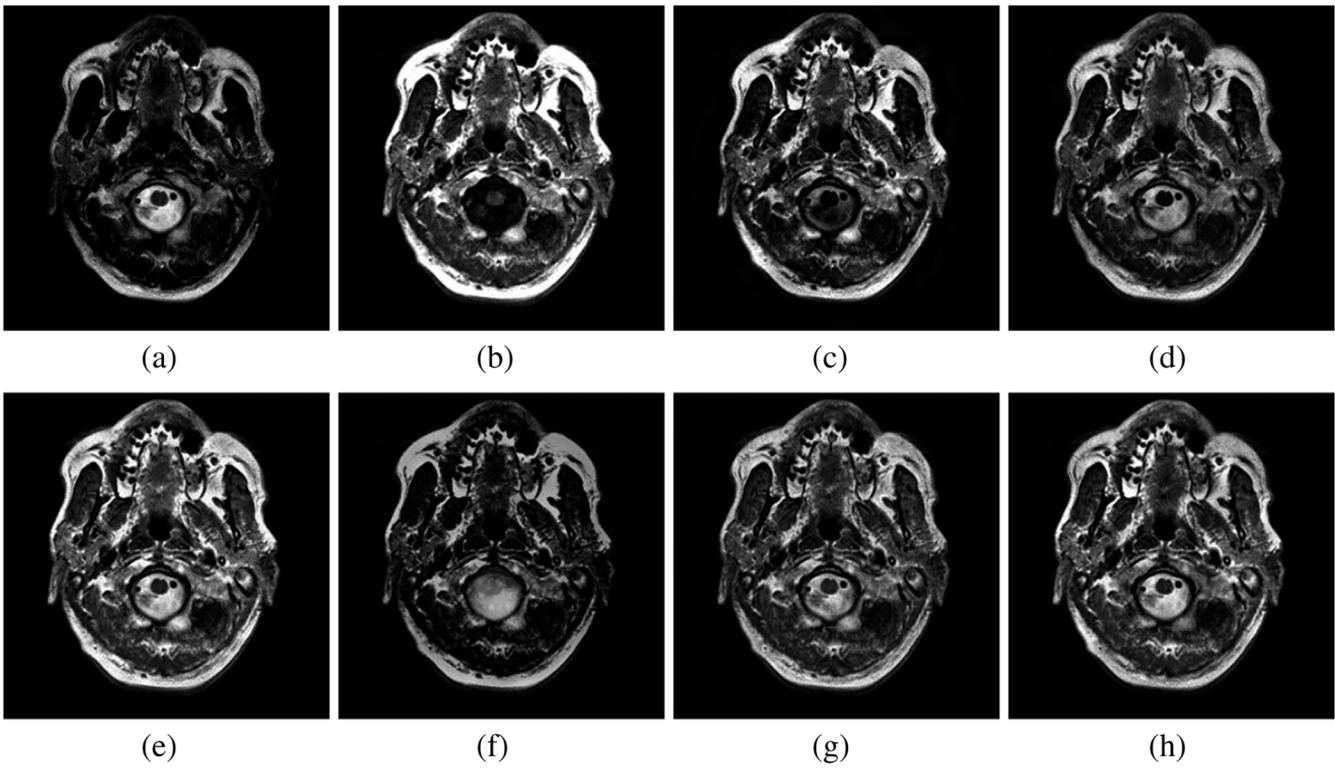
Fig. 10 Medical images A and fused images obtained by different methods: (a) MR-T_1, (b) MR-T_2, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.

Fig. 11 Medical images B and fused images obtained by different methods: (a) MR-T_1, (b) MR-T_2, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.
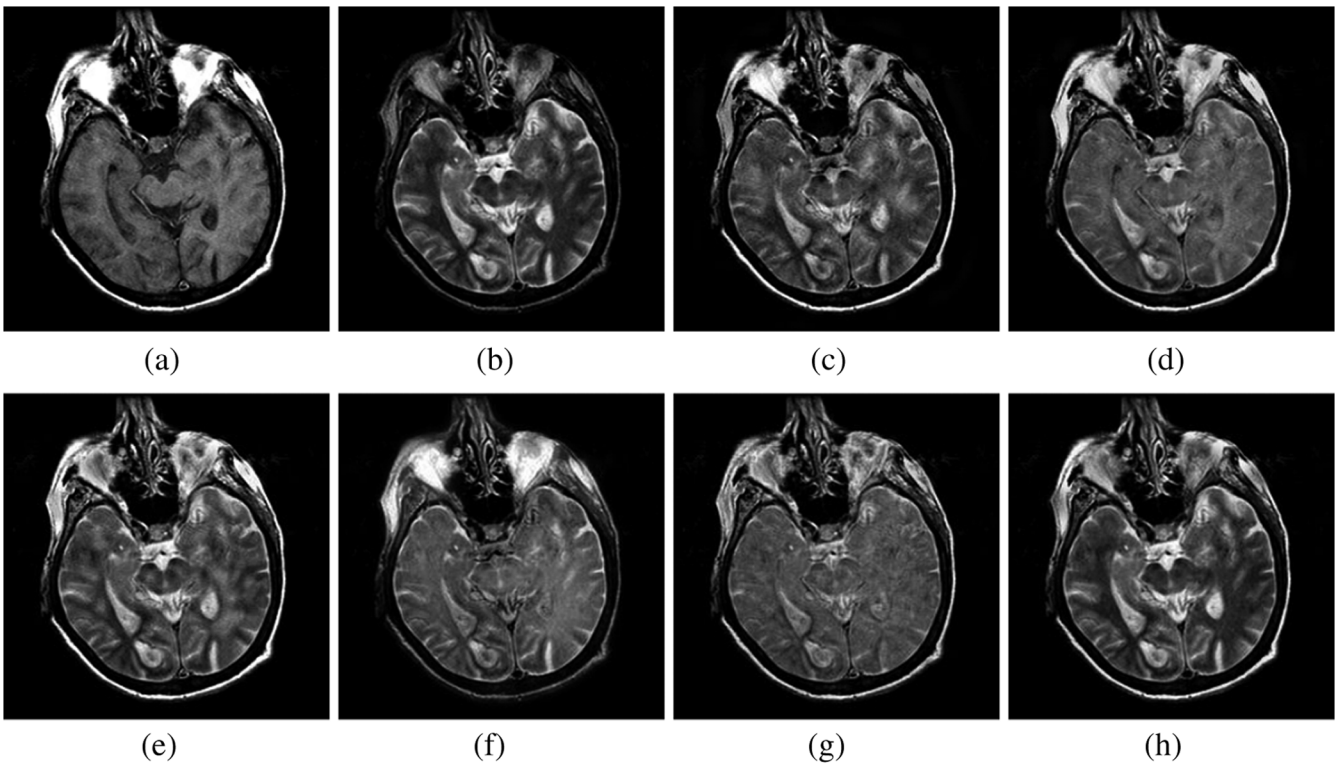
Fig. 12 Medical images C and fused images obtained by different methods: (a) MR-T_1, (b) MR-T_2, (c) NSCT, (d) GFF, (e) MSSR, (f) GTF, (g) CBF, and (h) FFIF.
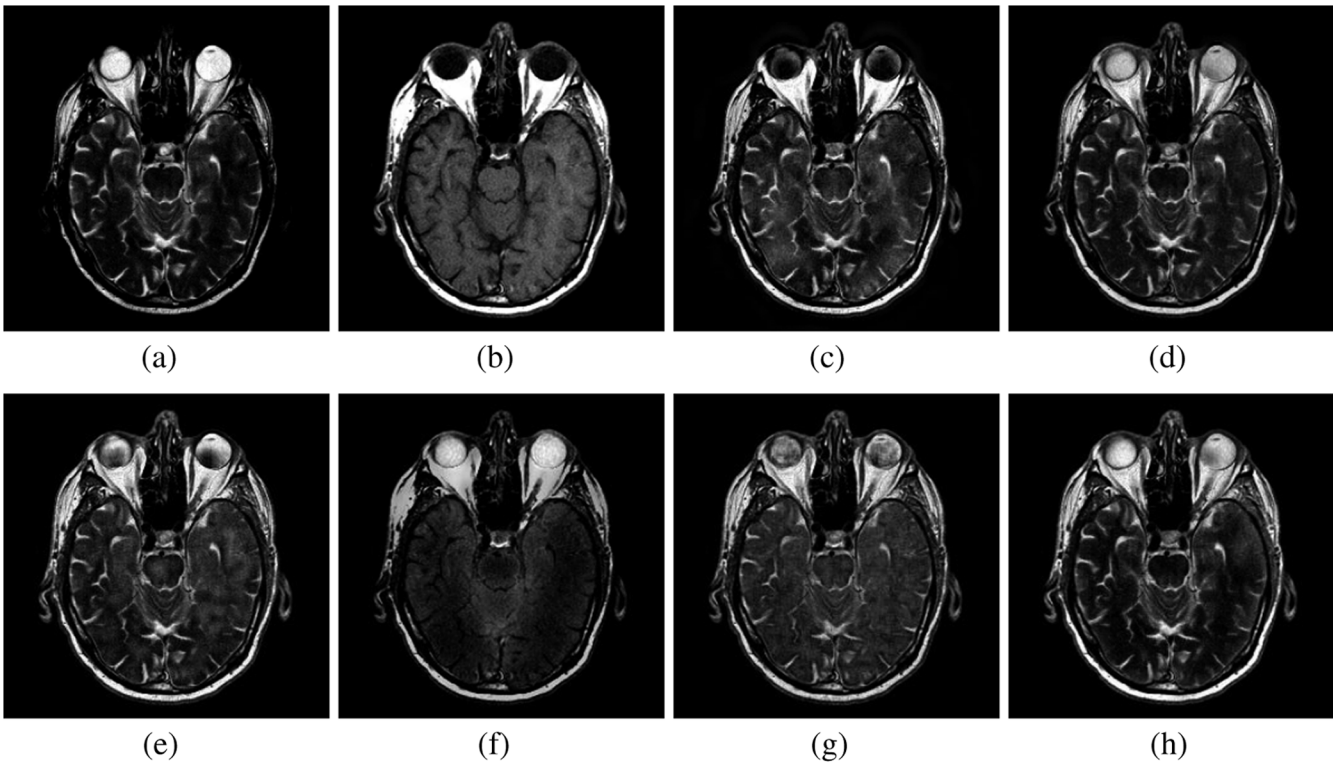
Table 2 Quantitative evaluation of different fusion methods for image sequences.

Images	Metrics	NSCT	GFF	MSSR	GTF	CBF	FFIF
Toy	$Q_p^{ab\|f}$	0.7056	0.7209	0.7235	0.6549	0.7211	0.7237
	$Q_w^{xy\|f}$	0.9173	0.9559	0.9354	0.9372	0.9206	0.9702
	Q_{MI}	1.0602	1.1580	1.1018	1.2139	1.0655	1.2150
	Q_{NCIE}	0.8310	0.8346	0.8329	0.8369	0.8309	0.8391
	Time (s)	149.0144	0.6710	1.6658	18.8374	75.9842	0.1884
Board	$Q_p^{ab\|f}$	0.4490	0.6028	0.6105	0.4598	0.5006	0.6723
	$Q_w^{xy\|f}$	0.7678	0.9479	0.9084	0.7741	0.7640	0.9832
	Q_{MI}	0.1912	0.6952	0.7184	0.3944	0.4740	0.9470
	Q_{NCIE}	0.8032	0.8184	0.8212	0.8056	0.8081	0.8357
	Time (s)	146.7246	0.7718	52.2784	10.5798	85.5720	0.2100
Medical	$Q_p^{ab\|f}$	0.5643	0.6135	0.6082	0.5548	0.5710	0.6268
	$Q_w^{xy\|f}$	0.4336	0.8729	0.6716	0.6309	0.8598	0.9582
	Q_{MI}	0.6370	0.7896	0.7624	0.7550	0.7876	1.1066
	Q_{NCIE}	0.8077	0.8094	0.8086	0.8088	0.8087	0.8158
	Time (s)	37.7614	0.0996	22.1736	0.1388	21.4638	0.0252

these digits with less blurred contours than the averaging, which indicates that the proposed method can well-protect the structure information.

The gradient magnitude is high if the pixel plays an important role in representing the scene and is low if pixel represents unimportant information. By comparing the saliency map, the weight map is determined by

$$w = \text{step}[g^{(1)}, g^{(2)}], \tag{24}$$

where $\text{step}[g^{(1)}, g^{(2)}]$ returns one for an element of w if the corresponding element of $g^{(1)}$ is a value larger than $g^{(2)}$, otherwise it returns zero.

Then, the structure-preserving linear filter is performed on w to obtain a desired weight map
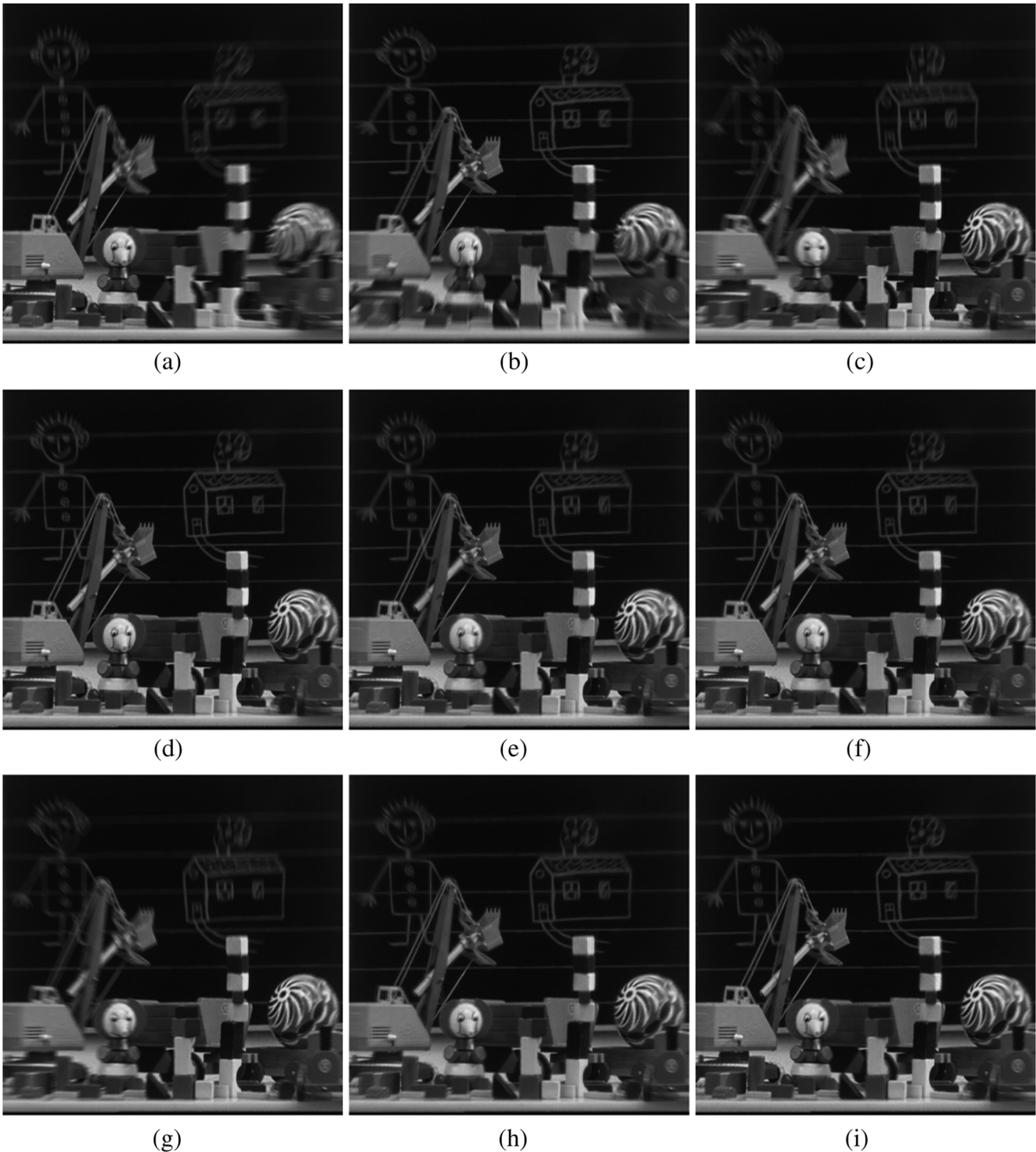

Fig. 13 Multifocus image sequence Toy-1, Toy-2, Toy-3, and fused images obtained by different methods: (a) Toy-1, (b) Toy-2, (c) Toy-3, (d) NSCT, (e) GFF, (f) MSSR, (g) GTF, (h) CBF, and (i) FFIF.

$$\hat{w}_p = \mu_k + \frac{\sigma_k^2}{\sigma_k^2 + \lambda}(w_p - \mu_k), \quad p \in \Omega_k. \tag{25}$$

Since the input of the linear filter is w_p, so the corresponding mean μ_k and variance σ_k are calculated in the sliding window of the image w_p.

As shown in Fig. 3, the filtered weight map $\hat{w}_p$ renders structures that look more natural than w_p. If the weight w_p is used to fuse images, subjectively, the fusion results may suffer from the blocking effect. As shown in Fig. 3, the filtered weight map $\hat{w}_p$ renders structures that look more natural than the result of w_p, then the fusion results of the proposed method have a natural structure. If the structure is very
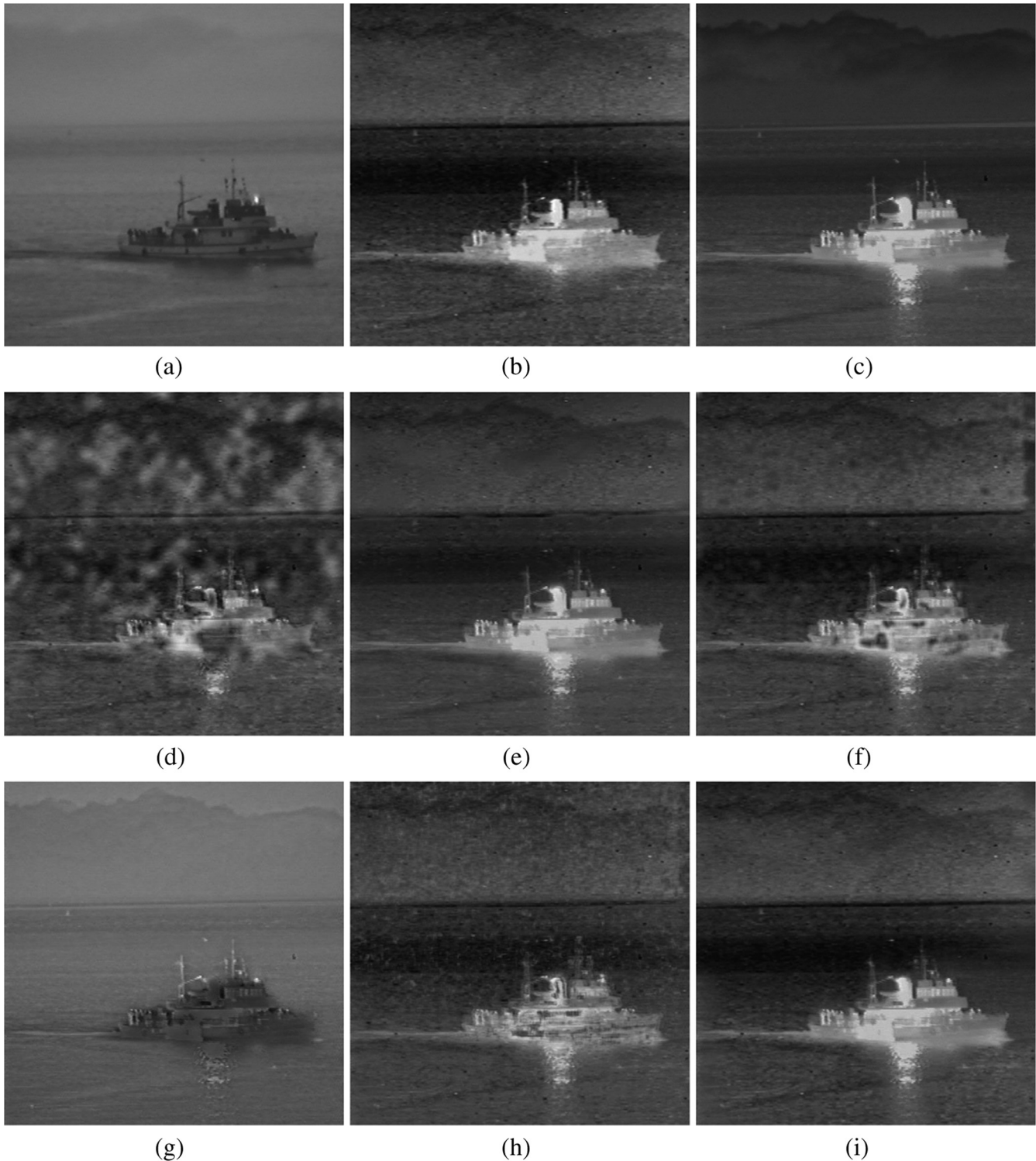

Fig. 14 Multispectral image sequence and fused images obtained by different methods: (a) visible, (b) near-infrared, (c) far-infrared, (d) NSCT, (e) GFF, (f) MSSR, (g) GTF, (h) CBF, and (i) FFIF.

complex, then using w_p hardly obtains a natural boundary. Finally, the fused image is obtained by the weight averaging on two input images

$$F = \hat{w}I^{(1)} + (1 - \hat{w})I^{(2)}. \tag{26}$$

The overall algorithm is given by Algorithm 2.

4 Experiments

4.1 Experimental Setup

Experiments are performed on three pairs of multifocus images, three pairs of the visible and infrared images, and three pairs of medical images. In addition, we use three different groups with more than two input images. All the
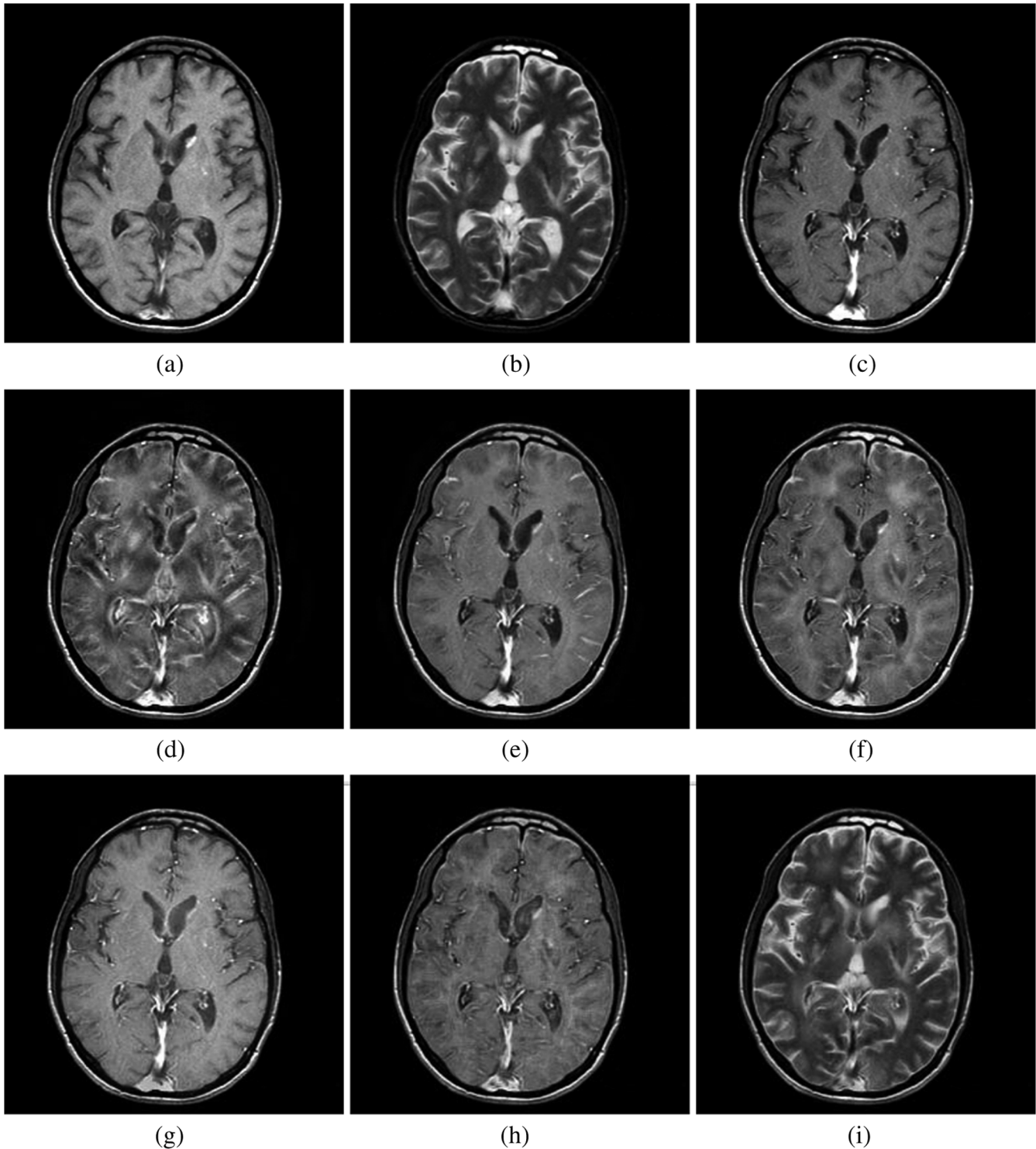

Fig. 15 Multimodal image sequence MR-T_1, MR-T_2, MR-Gad, and fused images obtained by different methods: (a) MR-T_1, (b) MR-T_2, (c) MR-Gad, (d) NSCT, (e) GFF, (f) MSSR, (g) GTF, (h) CBF, and (i) FFIF.

images have been aligned perfectly before fusion, and image registration needs to perform if images are not aligned well.[36,37] The proposed fast filtering image fusion (FFIF) method is compared with five state-of-the-art methods: NSCT,[28] GFF,[22] MSSR,[29] GTF,[30] and CBF.[31] For these methods, we adopt the default parameters given in their papers, respectively.

To evaluate the performance of the proposed fusion method, four objective image fusion performance metrics are adopted to evaluate the performances of different fusion methods, i.e., feature-based metric $Q_p^{ab|f}$,[38] structure-based metric $Q_w^{xy|f}$,[39] the normalized mutual information Q_{MI},[40] and nonlinear correlation information entropy Q_{NCIE}.[41,42] A comprehensive review of the metrics can be seen from Liu et al.'s literature.[42]

4.2 Experimental Results

First, experiments are conducted on the different two-modal images as shown in Figs. 4–12. To evaluate the performance of these results, the objective performances of different methods are presented in Table 1.

Figures 4(a) and 4(b) show two "disk" images that contain different focused regions. As shown in Figs. 4(c)–4(h), the results of GTF are not good since it cannot obtain clear digits on the clock. It is straightforward to check that MSSR- and CBF-based methods cause artifacts in the area of the book. FFIF obtains a better visual effect than other methods [see Fig. 4(h)], which indicates that the output image contains original focused regions of the input images. For the image brightness, GFF decreases image brightness, while FFIF well-preserves brightness. As shown in Figs. 5(a) and 5(b), we use the "Pepsi" images. Similarly, the words of "ES magnifier quality" are not fused well with these methods as shown in Figs. 5(c)–5(g). For the "lab" images [see Figs. 6(a) and 6(b)], the fusion results are shown in Figs. 6(c)–6(h). Fused results of NSCT, MSSR, and GTF methods have lower lightness and sharpness than the others. GFF and CBF make the fused images have visual artifacts in the student's hair. FFIF has a good contrast and spatial consistency. As shown in Table 1, the performance of FFIF always outperforms other methods in terms of the four evaluation metrics.
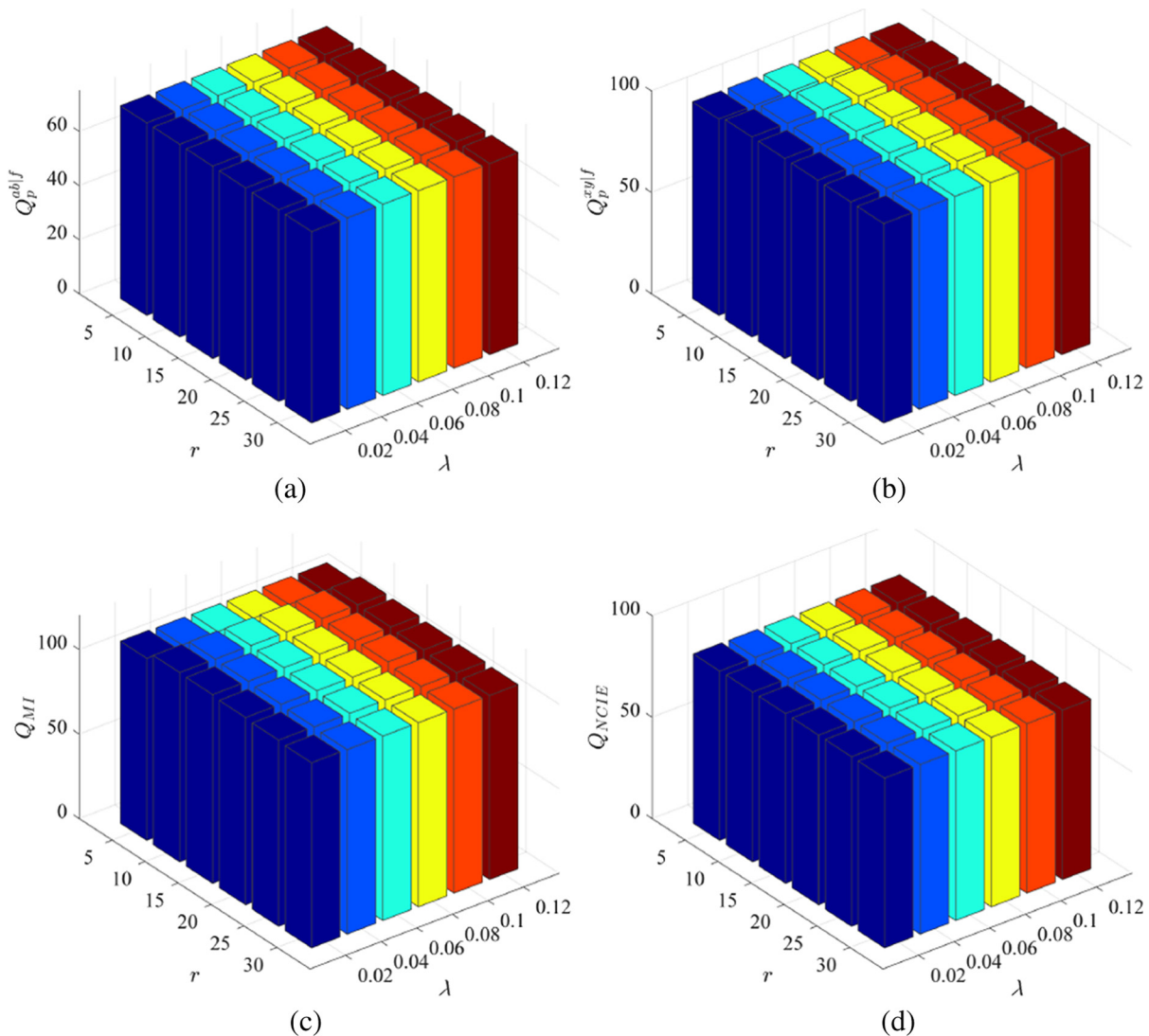

Fig. 16 Metrics with respect to different r and λ on multifocus image disk: (a) $Q_p^{ab|f}$, (b) $Q_p^{xy|f}$, (c) Q_{MI}, and (d) Q_{NCIE}.

For the visible and infrared image fusion, Figs. 7(a) and 7(b) show two source "Natocamp" images. The fused results of different methods are shown in Figs. 7(c)–7(h). GTF and NSCT can obtain dissatisfactory results in terms of the texture and target. Since Figs. 7(c) and 7(f) render some details invisible and lose some original features and structures. The fused results of MSSR and CBF have higher brightness than other methods, and they can well-preserve the texture and details of the source images. However, the details of the background are not easy to distinguish. GFF can obtain good fusion results for the visible and infrared images, but it can be seen that the method introduces unclear woods structure and decreases the contrast. By contrast, the FFIF-fused images have a distinct boundary and good visual perception. FFIF ensures that structures of source images are well-preserved. Figures 8(a) and 8(b) show a "sandpath" image. Similarly, as shown in Figs. 8(c)–8(h), FFIF performs well in preserving source image information. NSCT, GTF, and CBF-based methods [see Figs. 8(c), 8(f), and 8(g)] lose different levels of information in Fig. 8(a), As shown in Fig. 8(d), the GFF result is not very good in this case,

because the fused image produces the brightness and even causes uneven shade. In short, MSSR and FFIF usually perform well in preserving the image details. In fact, FFIF has the better performance. Moreover, Fig. 9 shows the third example of the visible and infrared image fusion, the street is not clear in the fusion results produced by other methods except FFIF [see Figs. 9(c)–9(g)], FFIF renders some details visible and well-preserves the brightness of the source images. As presented in Table 1, FFIF with the largest quality indices for $Q_p^{ab|f}$, $Q_w^{xy|f}$, Q_{MI}, and Q_{NCIE} among all of comparison methods implies that structures can be well-preserved by FFIF.

The medical source images and fused results with different methods are shown in Figs. 10–12. These images are from different magnetic resonance imaging (MRI) sensors, and we use three pairs of T_1-weighted and T_2-weighted MRI images. It can be seen from Figs. 10–12 that FFIF can obtain satisfactory results with more natural edges, better spatial consistency and the complementary information of different images, such as the structure of the bone and the area of the brain. Furthermore, the computational efficiency
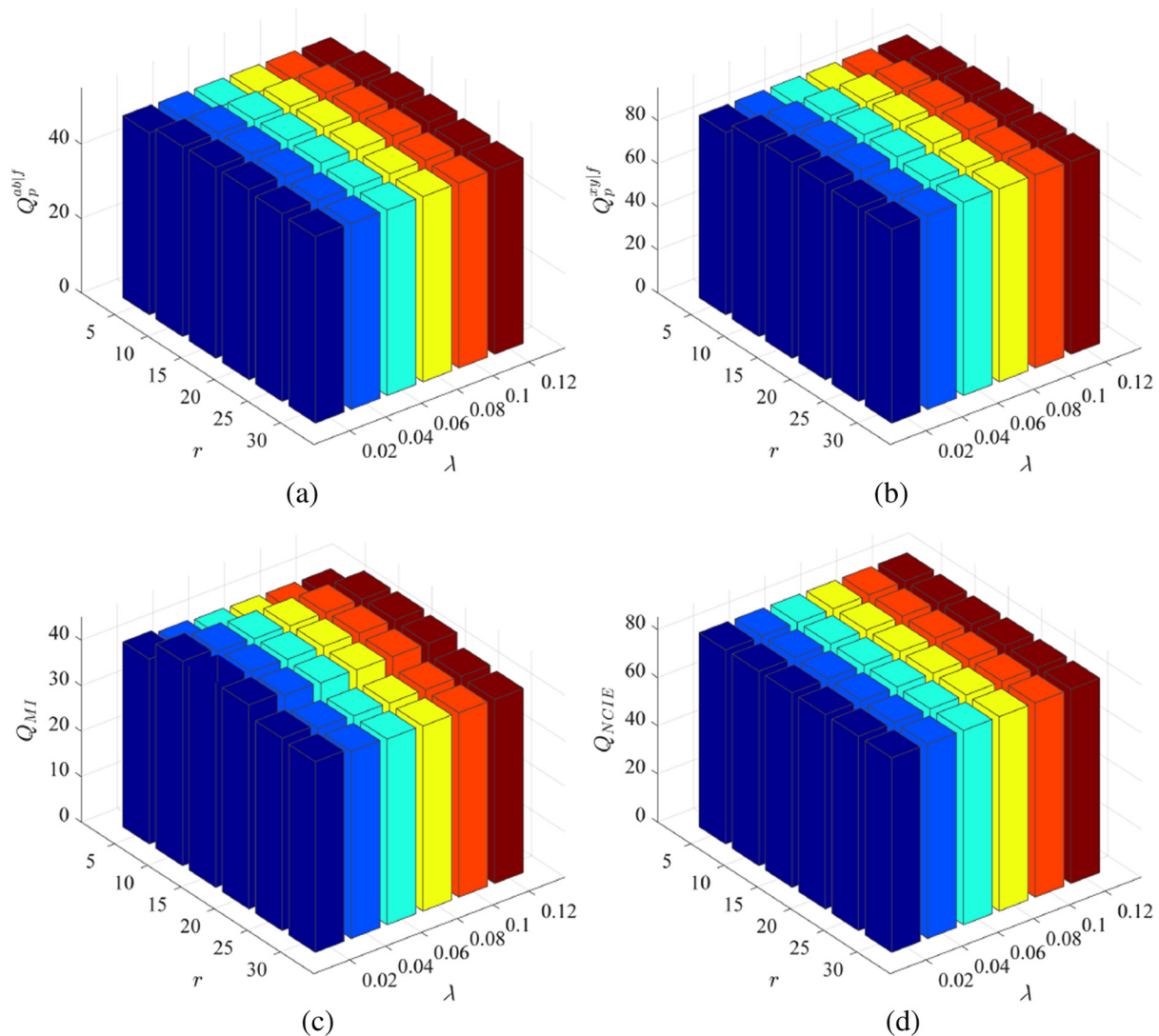

Fig. 17 Metrics with respect to different r and λ on multispectral image Natocamp: (a) $Q_p^{ab|f}$, (b) $Q_p^{xy|f}$, (c) Q_{MI}, and (d) Q_{NCIE}.

and quantitative evaluation of the different image fusion methods are shown in Table 1, and FFIF has better performance than others.

Second, we conduct experiments to demonstrate the effectiveness of the proposed method for image sequences, and three groups of three source images and their fused images are shown in Figs. 13–15, respectively. First, we fuse the top two images and second, we fuse the result with the third one. For the multifocus images, FFIF performs very well. For example, details in "Toy" are clearly presented in the fused image. Figures 14(a)–14(c) show three multispectral images, the three images are visible, near-infrared, and far-infrared images. The fusion results obtained by different fusion methods are shown in Figs. 14(d)–14(i). It can be seen from Fig. 14 that FFIF works well in keeping details of scenes, such as people on the board and brightness. Similarly, for the multimodal medical source images [MR-T_1 and MR-T_2, and MR-Gad (T_1-weighted, after Gd-DTPA)], FFIF has better results than others in terms of the bone structure and visual quality. Table 2 indicates that the performance of FFIF outperforms other methods in terms of $Q_p^{ab|f}$, $Q_w^{xy|f}$, Q_{MI}, and Q_{NCIE} and running time.

4.3 Computational Time Analysis

The computational time with the different fusion methods on several images is presented in Table 1. All the experiments are implemented on a PC with 3.6-GHz CPU and 12.0-GB memory. It can be seen from Table 1 that FFIF is very fast. FFIF costs only 0.009 s for the medical images, and NSCT, CBF, and MSSR methods required more computational time than other methods because the mean in a sliding window (structure-preserving filtering and morphological filtering) can be computed by a box filter.

4.4 Influence of r and λ

There are two parameters r and λ for FFIF. r denotes the radius of the sliding window and λ denotes the degree of the smoothing, respectively. However, they both have influence on the fusion quality and computational efficiency of FFIF. In detail, r determines filtering radius of the structuring element in morphological closing operation and λ is the smoothing degree of the linear filtering. In this section, the influence of these parameters on the fusion performance is analyzed in Figs. 16–18. Experiments are performed on
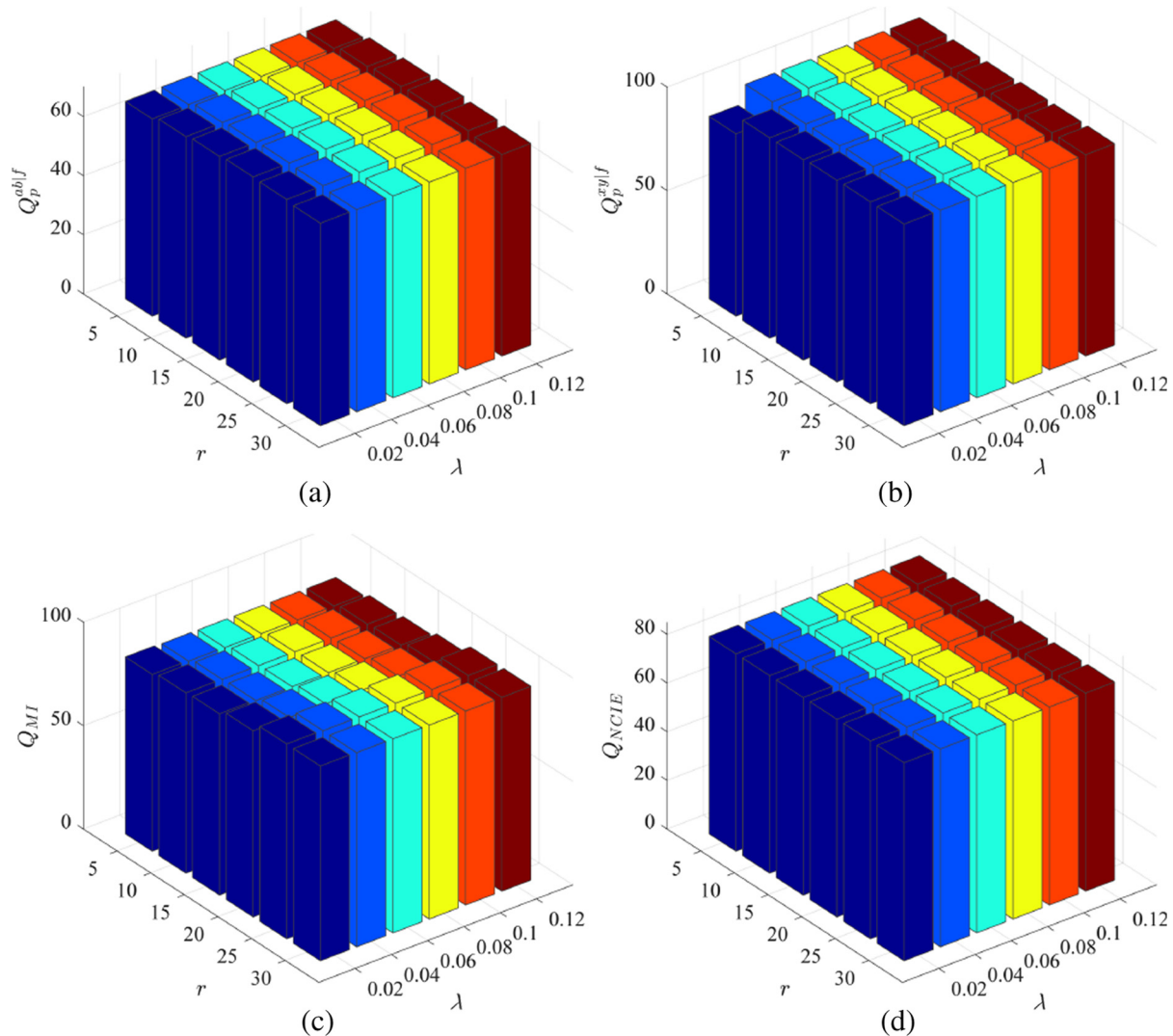

Fig. 18 Metrics with respect to different r and λ on medical image A: (a) $Q_p^{ab|f}$, (b) $Q_p^{xy|f}$, (c) Q_{MI}, and (d) Q_{NCIE}.

three different types of images, i.e., disk, Natocamp, and medical A, respectively. $Q_p^{ab|f}$ $Q_w^{xy|f}$, Q_{MI}, and Q_{NCIE} of FFIF are measured with different parameter settings. Specifically, Figs. 16–18 show the influence of r by varying it from 5 to 30 while λ is varying from 0.02 to 0.12. It can be seen from Figs. 16–18 that the amplitude of the column changes slightly indicating that the performance of FFIF is not sensitive to the setting of the r and λ. The default parameter setting of FFIF for the medical and multifocus image fusion is set as $r = 6$ and $\lambda = 0.06$ since it can give good fusion quality on the subjective and objective evaluation. In addition, the default parameter setting for the visible and infrared image fusion is $r = 16$ and $\lambda = 0.06$. The experimental results show that good fusion performance is obtained with these parameters.

5 Conclusion

In this paper, we propose a spatial image fusion method based on fast filtering. The proposed method uses the discrete gradient magnitude to detect contrast and image sharpness, and it is refined with a fast morphological filtering operation. Moreover, we utilize a structure-preserving filter to obtain a desired weight map in the spatial domain. Experimental results of different images show that the performance of the FFIF method outperforms other state-of-the-art image fusion approaches in terms of both visual performance and objective metrics. More importantly, the proposed method is not sensitive to the setting of the parameters and costs less time, which renders it easy to apply to real-time applications.

References

1. Z. Zhang and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proc. IEEE* **87**(8), 1315–1326 (1999).
2. G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Inf. Fusion* **4**(4), 259–280 (2003).
3. G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.* **37**(9), 1855–1872 (2004).
4. S. Li, B. Yang, and J. Hu, "Performance comparison of different multiresolution transforms for image fusion," *Inf. Fusion* **12**(2), 74–84 (2011).
5. P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition," *Proc. SPIE* **0575**, 173–181 (1985).
6. A. Toet, L. J. Van Ruyven, and J. M. Valeton, "Merging thermal and visual images by a contrast pyramid," *Opt. Eng.* **28**(7), 789–792 (1989).
7. P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proc., Fourth Int. Conf. on Computer Vision*, pp. 173–182, IEEE (1993).
8. R. Redondo et al., "Multifocus image fusion using the log-Gabor transform and a multisize windows technique," *Inf. Fusion* **10**(2), 163–171 (2009).
9. T. Li and Y. Wang, "Biological image fusion using a NSCT based variable-weight method," *Inf. Fusion* **12**(2), 85–92 (2011).
10. Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus images," *Inf. Fusion* **20**, 60–72 (2014).
11. M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Process.* **14**(12), 2091–2106 (2005).
12. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed., Pearson Education, London, United Kingdom (2007).
13. F. Zhang et al., "Segment graph based image filtering: fast structure-preserving smoothing," in *Proc. of the IEEE Int. Conf. on Computer Vision*, pp. 361–369 (2015).
14. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990).
15. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth Int. Conf. on Computer Vision*, Vol. **6**, pp. 839–846, IEEE (1998).
16. Z. Farbman et al., "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graphics* **27**(3), 67 (2008).
17. L. Xu et al., "Image smoothing via L_0 gradient minimization," *ACM Trans. Graphics* **30**(6), 174 (2011).
18. K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2013).
19. Y. Jiang and M. Wang, "P-M equation based multiscale decomposition and its application to image fusion," *Pattern Anal. Appl.* **17**(1), 167–178 (2014).
20. Q. Wang et al., "Robust multi-modal medical image fusion via anisotropic heat diffusion guided low-rank structural analysis," *Inf. Fusion* **26**, 103–121 (2015).
21. Y. Jiang and M. Wang, "Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter," *IET Image Proc.* **8**(3), 183–190 (2014).
22. S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.* **22**(7), 2864–2875 (2013).
23. E. S. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graphics* **30**(4), 69 (2011).
24. J. Hu and S. Li, "The multiscale directional bilateral filter and its application to multisensor image fusion," *Inf. Fusion* **13**(3), 196–206 (2012).
25. J. Zhao et al., "Detail enhanced multi-source fusion using visual weight map extraction based on multi scale edge preserving decomposition," *Opt. Commun.* **287**, 45–52 (2013).
26. W. Liu et al., "Robust guided image filtering," arXiv preprint arXiv:1703.09379 (2017).
27. L.-C. Chen et al., "Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 4545–4554, IEEE (2016).
28. G. Bhatnagar, Q. J. Wu, and Z. Liu, "Directive contrast based multimodal medical image fusion in NSCT domain," *IEEE Trans. Multimedia* **15**(5), 1014–1024 (2013).
29. Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion* **24**, 147–164 (2015).
30. J. Ma et al., "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion* **31**, 100–109 (2016).
31. B. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal Image Video Process.* **9**(5), 1193–1204 (2015).
32. J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, New Jersey (1990).
33. S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, New Jersey (1993).
34. F. C. Crow, "Summed-area tables for texture mapping," *ACM Trans. Graphics* **18**(3), 207–212 (1984).
35. T. Shibata, M. Tanaka, and M. Okutomi, "Unified image fusion based on application-adaptive importance measure," in *IEEE Int. Conf. on Image Processing (ICIP '15)*, pp. 1–5, IEEE (2015).
36. J. Ma et al., "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.* **48**(3), 772–784 (2015).
37. J. Ma et al., "Robust l_2e estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.* **63**(5), 1115–1129 (2015).
38. C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.* **36**(4), 308–309 (2000).
39. C. Yang et al., "A novel similarity based quality metric for image fusion," *Inf. Fusion* **9**(2), 156–160 (2008).
40. M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'information measure for performance of image fusion'," *Electron. Lett.* **44**(18), 1066–1067 (2008).
41. T. Stathaki, *Image Fusion: Algorithms and Applications*, Academic Press, Cambridge, Massachusetts (2008).
42. Z. Liu et al., "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(1), 94–109 (2012).

Kun Zhan received his BSc and PhD degrees from Lanzhou University, China, in 2005 and 2010, respectively. Currently, he works at Lanzhou University. His main research interests are image processing and computer vision.

Yuange Xie received her BSc degree in electronic information science and technology from Shanxi Normal University, China, in

2015. She is currently a master's student at Lanzhou University. Her main research interest is image fusion.

Haibo Wang received his BSc degree in electronic and communication engineering from Shandong Agricultural University, China, in 2015. He is currently a master's student at Lanzhou University. His main research interest is hyperspectral image classification.

Yufang Min received her BSc and MSc degrees from Lanzhou University, China, in 2005 and 2008, respectively. She is now a PhD candidate at the Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences. Her main research interest is remote sensing.