

Journal of Electronic Imaging

SPIDigitalLibrary.org/jei

Adaptive multistrategy image fusion method

Xiao-qing Luo
Zhan-cheng Zhang
Xiao-jun Wu



Adaptive multistrategy image fusion method

Xiao-qing Luo,^a Zhan-cheng Zhang,^b and Xiao-jun Wu^{a,*}

^aJiangnan University, School of Internet of Things, Wuxi, Jiangsu 214122, China

^bSuzhou University of Science and Technology, College of Electronic and Information Engineering, Suzhou, Jiangsu 215009, China

Abstract. Design of fusion rule is an important step in fusion process. Traditional single fusion rules are inflexible when they are being used to fuse feature-rich images. To address this problem, an adaptive multistrategy image fusion method is proposed. Its flexibility lies in the combination of a choose-max strategy and a weighted average strategy. Moreover, the region-based characteristics and the shift-invariant shearlet transform (SIST)-based activity measures are proposed to guide the selection of strategies. The key points of our method are: (1) Window-based features are extracted from the source images. (2) Use of the fuzzy c-means clustering algorithm to construct a region map in the feature difference space. (3) The dissimilarity between corresponding regions is employed to quantify the characteristic of regions and the local average variance of the SIST coefficients are considered as activity measures to evaluate the saliency of the related coefficient. (4) The adaptive multistrategy selection scheme is achieved by a sigmoid function. Experimental results show that the proposed method is superior to the conventional image fusion methods both in subjective and objective evaluations. ©2014 SPIE and IS&T [DOI: 10.1117/1.JEI.23.5.053011]

Keywords: image fusion; region; shift-invariant shearlet transform; characteristic.

Paper 14049 received Jan. 28, 2014; revised manuscript received Aug. 17, 2014; accepted for publication Aug. 27, 2014; published online Sep. 22, 2014.

1 Introduction

With the advance of imaging sensors and microelectronics, multimodality image fusion has emerged as a new and promising research area. With a proper fusion rule, multimodality images are combined into a single composite (i.e., fused image) for human and machine perception or further image processing tasks such as segmentation, feature extraction, and target recognition.^{1,2} Therefore, an effective fusion rule will improve the quality of a fused image.

At present, the most commonly used fusion rules can be divided into two types: the choose-max strategy and the weighted average strategy. For example, Zheng et al.³ combined the multiple sets of low/support value components using the choose-max strategy and the weighted average strategy. In Ref. 4, the detail subbands are combined by the choose-max strategy using a standard deviation measure and the approximation subbands are combined with the weighted average strategy with entropy measure. Li et al.⁵ applied homogeneity similarity for a multifocus image fusion, in which the absolute-maximum-choosing strategy and the average strategy are employed to fuse the detail and approximation subbands, respectively. Moreover, the choose-max strategies with a fire map of a pulse-coupled neural network are often used to fuse the subband coefficients.^{6,7} In addition to these pixel-based methods, several researchers argued that the fusion process based on regions is more robust and more easily to express the local structural characteristics of objects.^{8,9} Therefore, many region-based fusion methods have also been proposed in recent years. Correspondingly, the above-mentioned fusion rules have been widely extended into a region-based fusion method. For instance, Piella proposed a choose-max strategy using a local correlation measure to perform fusion of each

region.⁸ In Ref. 10, the priority of a region measured by energy, variance, or entropy of the regional wavelet coefficients is used to weight a region in the fusion process. In Ref. 11, Li and Yang fused the corresponding regions by the choose-max strategy with spatial frequencies (For simplicity, we call the method RSSF.).

It is evident that all the above-mentioned image fusion methods used only a single fusion rule to fuse all high/low-frequency subband coefficients or pixels. However, there are different dynamic ranges and correlations in the source image, and a single fusion rule may decrease the contrast of the fused image. A recent trend is multistrategy fusion rules. Examples include the works in Refs. 12 and 13, which employed a similarity index and threshold to distinguish the type of region between source images. Both the weighted average strategy and the choose-max strategy were employed. The former focuses on redundant information and the latter focuses on complementary information. Furthermore, to avoid tuning the threshold, Luo et al.¹⁴ used structural similarity (SSIM) to identify the type of region which further guides the selection of the fusion strategy.

Although the above multistrategy fusion methods have enhanced the quality of image fusion to some extent, there are still some drawbacks in the fusion process. First, for the redundant regions, when the redundant degree of the regions is large, the weighted average strategy is degenerated into the average strategy. On the contrary, when the redundant degree of regions is small, it reduces to the choose-max strategy. Therefore, the redundant degree of the regions should be considered in the weighted average strategy. Second, the source images endow with not only the local structural characteristics, but also the activity measure of individual coefficients. If the above two aspects are considered in the selection of a multistrategy fusion rule, the performance of the fused image will be more effective.

*Address all correspondence to: Xiao-jun Wu, E-mail: xiaojun_wu_jnu@163.com

In this paper, an adaptive multistrategy image fusion method is proposed. The whole architecture is constructed with a data plane and a feature plane. In the data plane, source images are decomposed into low-frequency coefficients and high-frequency coefficients with a shift-invariant shearlet transform (SIST). Since the low-frequency coefficients denote only the approximate information, the choose-max strategy with the local average energy is simply adopted to fuse them. Additionally, the high-frequency coefficients represent a lot of detailed information. Thus, a multistrategy fusion rule combined with the choose-max and weighted average strategies is executed to fuse these coefficients. In the feature plane, to incorporate the two strategies, source images are partitioned into several windows and the corresponding features are extracted, then these windows are clustered into regions. Moreover, the dissimilarity of regions and SIST-based activity measures are fed into a sigmoid function to achieve a flexible multistrategy fusion rule. The detailed diagram is depicted in Fig. 3.

The method consists of four main stages to: (1) obtain the region map; (2) quantify the characteristics of corresponding regions and distinguish the type of corresponding regions; (3) calculate the activity measure of the SIST coefficients; and (4) connect the fusion strategy selection with the characteristics of corresponding regions and the SIST-based activity measures by a sigmoid function. This paper is the extended version of our recent conference paper in International Conference on Pattern Recognition (ICPR).¹⁵

The remainder of this paper is organized as follows. Section 2 briefly reviews the principle of SIST. In Sec. 3, the framework of the proposed method is described. Section 4 explains how to obtain the region map. Section 5 discusses how to quantify the characteristics of corresponding regions and distinguish the type of corresponding regions. The proposed fusion rule is described in Sec. 6. Finally, a discussion of experimental results and conclusions is drawn in Secs. 7 and 8, respectively.

2 Principle of SIST

Wavelets are very efficient only when dealing with point-wise singularities. In higher dimensions, other types of singularities are usually present or even dominant, and wavelets are unable to handle them very efficiently. In order to overcome this limitation of traditional wavelets, their directional sensitivity has to be increased. SIST is one of the state-of-the-art multiscale decomposition means which has a rich mathematical structure similar to wavelet and the true two-dimensional (2-D) sparse representation for images with edges and a shift-invariance.^{16,17}

The SIST can be completed in two steps: multiscale partition and directional localization. In the first step, the shift-invariance, which means less sensitivity to the image shift, can be achieved by the nonsubsampling pyramid filter scheme in which the Gibbs phenomenon is suppressed to a great extent as a result of replacing down-samplers with convolutions. In the second step, the frequency plane is decomposed into a low-frequency subband and several

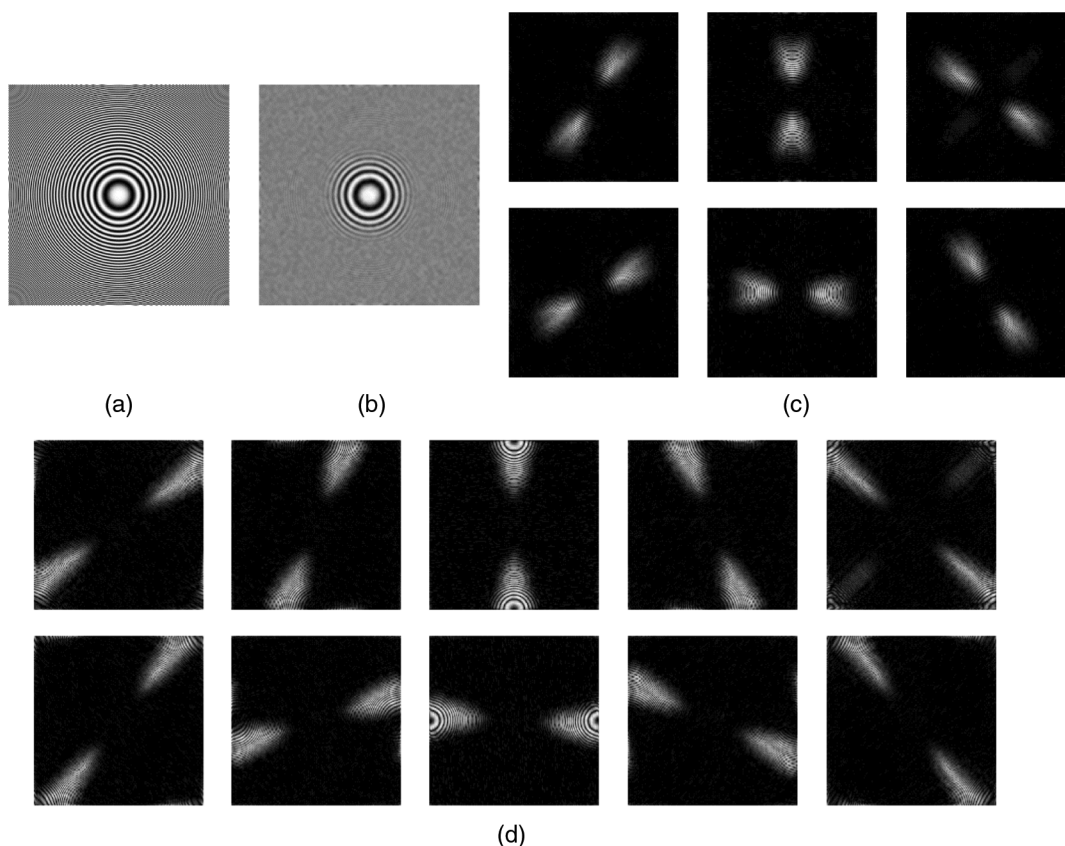


Fig. 1 Illustration of decomposing image zoneplate into two levels by SIST. (a) Zoneplate (256×256). (b) The approximate SIST coefficients at level 2. (c) Images of the detailed coefficients at level 1. (d) Images of the detailed coefficients at level 2.

trapezoidal high-frequency subbands by the shift-invariant shearing filters. To intuitively illustrate the principle of SIST and show its superiority in contrast to stationary wavelet transform (SWT), their two-level SIST and SWT decompositions of the zoneplate image are shown in Figs. 1 and 2, respectively. Here, the basic function of SWT is set as Symlets 4 (sym4) and the SIST parameter determining the number of directions is defined as [2, 3], so the direction numbers for each scale from coarse to fine are 6 and 10.

3 Framework of Proposed Method

Since this paper focuses on the research of image fusion, we assume that the source images to be fused have been geometrically registered. The system diagram of the proposed

method is depicted in Fig. 3. The general procedure of the proposed method can be divided into a feature plane and a data plane.

In the feature plane, the procedure is summarized as follows:

1. The source images A and B are divided into $M \times M$ windows w_i^A and w_i^B ($1 \leq i \leq \text{num}$), where num is the number of windows in an image. The features of w_i^A and w_i^B (f_i^A and f_i^B) are extracted and the difference between them is symbolized as \hat{f}_i . The detailed definitions of f_i^A and f_i^B are given in Sec. 4.1.
2. The region map can be obtained by segmenting the feature vectors f_i using fuzzy c-means (FCM).

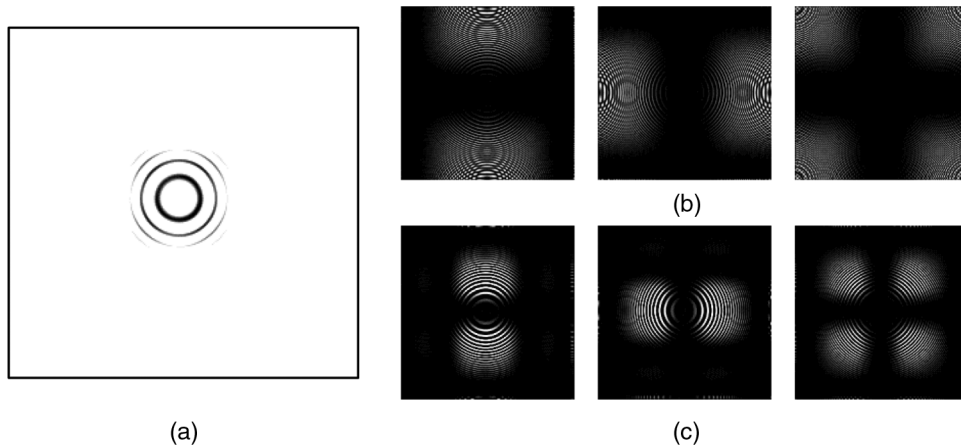


Fig. 2 Illustration of decomposing image zoneplate into two levels by SWT. (a) The approximate SWT coefficients at level 2. (b) Images of the detailed coefficients at level 1. (c) Images of the detailed coefficients at level 2.

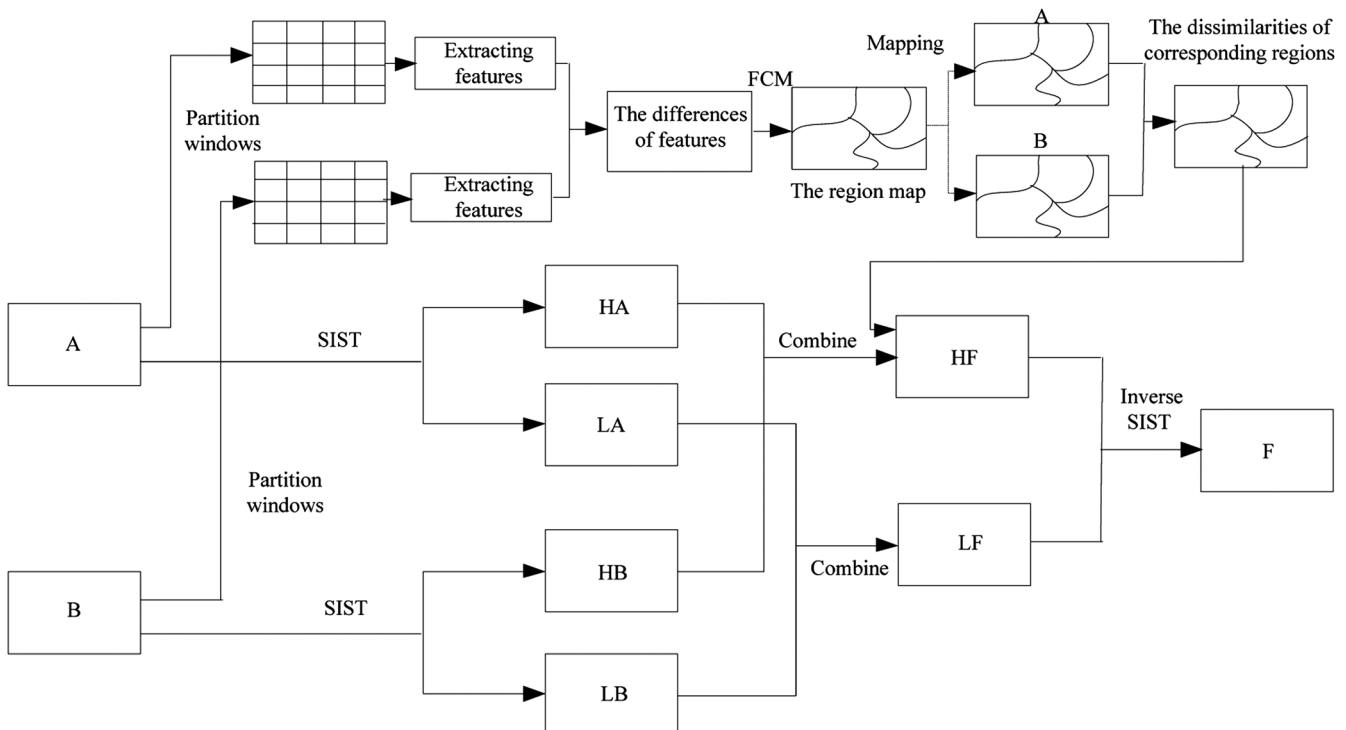


Fig. 3 Image fusion method using region segmentation and sigmoid function.

3. The region map is mapped into the source images A and B, and then the characteristics of corresponding regions (i.e., the dissimilarity) are calculated.

In the data plane, the procedure is summarized as follows:

1. The source images A and B are decomposed by SIST. Let LA and LB be the low-frequency coefficients of A and B, respectively. Then, let HA and HB be the high-frequency coefficients of A and B, respectively. LF is the fused low-frequency coefficients and HF is the fused high-frequency coefficients.
2. The high-frequency decomposition coefficients are fused by the adaptive multistrategy fusion rule with a sigmoid function which is parameterized by the dissimilarity calculated from the feature plane and the SIST-based activity measures. The low-frequency decomposition coefficients are simply fused by the choose-max strategy with the local average energy.
3. Finally, the fused image F is reconstructed by using inverse SIST.

In the following subsections, the proposed algorithm is explained in detail.

4 Generation of Region Map

In this section, we explain how to generate the region map. First, the source images are divided into windows and then the features of the windows are extracted. Second, the feature vectors are constructed in the feature difference space. Third, the region map is obtained by clustering the feature vectors using FCM.

4.1 Feature Extraction

In the process of feature-level image fusion, an important step is to extract features from the source images. Since the sharpness and edges of an image can be represented by the high-frequency coefficients, the features of windows are extracted not only from the pixels of a window but also the high-frequency coefficients of a window. The source images A and B are decomposed in the first level with six high-frequency subbands HA_j and HB_j ($1 \leq j \leq 6$). A, B, HA_j , and HB_j are divided into windows w_i^q and $H_{i,j}^q$ ($q = A, B; 1 \leq i \leq \text{num}$). In this paper, the variance¹⁸ (V_i^q), gradient⁶ (G_i^q), and the gray scale (WI_i^q) of w_i^q ¹⁹ are selected as the features of the spatial domain. The variance¹⁸ ($HV_{i,j}^q$) and energy¹⁸ ($HEN_{i,j}^q$) of $H_{i,j}^q$ are extracted as the features of the SIST domain. These features are defined as follows:

4.1.1 Variance

The variance reflects the relative degree of dispersion between the pixels in a window w_i^q . $I(x, y)$ is defined as the pixel intensity located at (x, y) and V_i^q denotes the variance of window w_i^q , therefore, we have

$$\text{Mean}_i^q = \frac{1}{M \times M} \sum_{x,y \in w_i^q} I(x, y), \quad (1)$$

$$V_i^q = \frac{1}{M \times M} \sum_{x,y \in w_i^q} [I(x, y) - \text{Mean}_i^q]^2. \quad (2)$$

The calculation of $HV_{i,j}^q$ is similar to V_i^q .

4.1.2 Gradient

The average gradient reflects the clarity and detailed information of the window.

$$G_i^q = \frac{1}{(M-1) \times (M-1)} \times \sum_{x=1}^{M-1} \sum_{y=1}^{M-1} \sqrt{\left\{ \left[\frac{\partial I(x, y)}{\partial x} \right]^2 + \left[\frac{\partial I(x, y)}{\partial y} \right]^2 \right\} / 2}. \quad (3)$$

4.1.3 Gray scale of window

The gray scale feature of the window is extracted by two dimensional principal component analysis (2-DPCA), which was proposed by Yang et al.¹⁹ In contrast with the PCA method, 2-DPCA can avoid reshaping the image window into an image vector so that the window structure is kept and the computational complexity is significantly reduced.

For the window set $w^q = [w_1^q, w_2^q, \dots, w_i^q, \dots, w_{\text{num}}^q]$, the mean window matrix is defined as

$$\text{WM}^q = \frac{1}{\text{num}} \sum_{i=1}^{\text{num}} w_i^q. \quad (4)$$

The covariance matrix of the window set is defined as

$$\text{WC}^q = \frac{1}{\text{num}} \sum_{i=1}^{\text{num}} (w_i^q - \text{WM}^q)(w_i^q - \text{WM}^q)^T. \quad (5)$$

The projection matrix P is formed by the eigenvectors corresponding to the first k largest eigenvalues of WC^q . The window set is projected into eigenspace by

$$\text{WI}_i^q = P^T w_i^q. \quad (6)$$

$\text{WI}_i^{q,k}$ is the gray scale feature of window w_i^q ($k = 1, 1 \leq i \leq \text{num}$).

4.1.4 Energy

$HEN_{i,j}^q$ and $H_{i,j}^q(x, y)$ are defined as the energy of SIST coefficients in the window $H_{i,j}^q$, and the SIST coefficient located at (x, y) , then we have

$$HEN_{i,j}^q = \sum_{x,y \in w_i^q} [H_{i,j}^q(x, y)]^2. \quad (7)$$

Next, we use the feature differences of the source images to construct feature vectors

$$\text{DV}_i = \text{abs}(V_i^A - V_i^B), \quad (8)$$

$$\text{DG}_i = \text{abs}(G_i^A - G_i^B), \quad (9)$$

$$\text{DI}_i = \text{abs}(\text{WI}_i^{A,1} - \text{WI}_i^{B,1}), \quad (10)$$

$$SV_{i,j} = \text{abs}(H_{V_{i,j}}^A - H_{V_{i,j}}^B), \quad (11)$$

$$SE_{i,j} = \text{abs}(HEN_{i,j}^A - HEN_{i,j}^B). \quad (12)$$

Each feature vector $\vec{f}_i (f_i \in R^{15}, 1 \leq i \leq \text{num})$ consists of 15 feature differences

$$\vec{f}_i = [SV_{i,1}, SE_{i,1}, \dots, SV_{i,6}, SE_{i,6}, DV_i, DG_i, DI_i]. \quad (13)$$

Let $\hat{f}_i (\hat{f}_i \in R^{15}, 1 \leq i \leq \text{num})$ be the final normalized feature vectors.

4.2 Regions Segmentation Using FCM

The feature vectors can reflect the characteristics between source images. To intuitively illustrate this fact, Fig. 4 shows histograms of SV_1 and DG from multifocus clock images [Figs. 4(a) and 4(b)]. As can be seen in Figs. 4(c) and 4(d), the distribution of the histograms is continuous, which means that SV_1 and DG have the ability to reflect the redundancy between source images. The histograms of other feature differences have the same properties. Considering the space limitation, these histograms are omitted.

Due to the local correlation of source images, the characteristics of a source are region related. Therefore, the following problem is to partition the feature vectors into clusters. Considering the fact that fuzzy clustering techniques have been effectively used in image processing, the FCM algorithm,²⁰ a well-known fuzzy clustering method, is adopted to cluster the feature vectors in this paper.

Assume that the number of clusters is c . The membership of the feature vector \hat{f}_i in the r 'th cluster is labeled as $\mu_{ri} (\mu_{ri} \in [0,1], 1 \leq r \leq c, 1 \leq i \leq \text{num})$. If $\mu_{ri} > \mu_{r'i}$ ($r, r' = 1, 2, \dots, c, r \neq r'$), then \hat{f}_i belongs to the r 'th cluster. As a result, the feature vectors \hat{f} are partitioned into c groups, i.e., the region map $R = \{R_1, R_2, \dots, R_r, \dots, R_c\}$. In this paper, let the number of clusters be 10 and let the size of window be 4×4 by the cross-validation technique. The discussions about the tuning of the parameters are given in Sec. 7.3.

5 Quantization of Region Characteristics

The characteristic of corresponding regions can be quantified by the dissimilarity, which reflects the complementarity. In theory, the feature vector of fully redundant corresponding windows should be $\hat{f}_i = [0, 0, \dots, 0]$. Therefore, the dissimilarity of the corresponding region can be calculated as

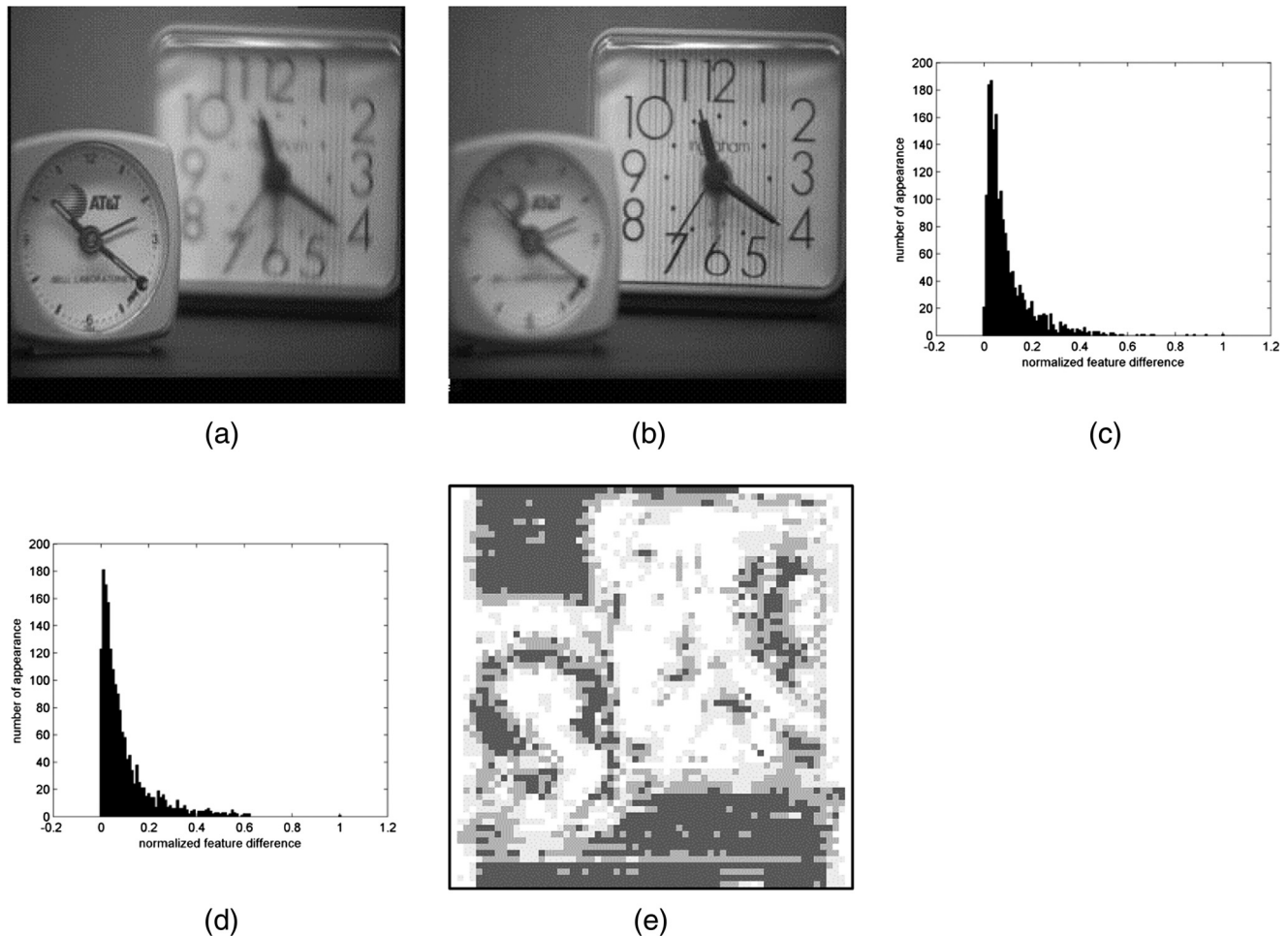


Fig. 4 The histograms of feature differences (SV_1 and DG) and the region map of multifocus clock images: (a) left focus clock image, (b) right focus clock image, (c) the histogram of SV_1 , (d) the histogram of DG , and (e) the region map using the proposed method.

$$d_r = \frac{\sum_{i \in R_r} \sqrt{\sum_{h=1}^{15} (\hat{f}_{ih})^2}}{V(R_r)}, \quad (14)$$

where $V(R_r)$ represents the number of feature vector of region r . The higher the values of d_r , the higher the complementarity is. To identify the redundancy or complementarity of a corresponding region, the complementary seed region is defined as the region with the largest d_r value (d_{\max}) and the redundant seed region is defined as the region with zero feature vectors (d_{\min}). Thus, the dissimilarity of the redundant seed region is 0 ($d_{\min} = 0$). Let $d = [d_{\min}, d_1, \dots, d_r, \dots, d_c]$. If d is normalized, then $d_{\max} = 1$ and $d_{\min} = 0$ ($d_r \in [0, 1]$). Denoting $\text{dist}_{\max r}$ as the distances of d_r to d_{\max} and $\text{dist}_{\min r}$ as the distances of d_r to d_{\min} , then

$$\text{dist}_{\max r} = d_{\max} - d_r, \quad (15)$$

$$\text{dist}_{\min r} = d_r - d_{\min}. \quad (16)$$

All distances can be denoted as

$$\text{dist}_{\max} = \{d_{\max 1}, d_{\max 2}, \dots, d_{\max c}\}, \quad (17)$$

$$\text{dist}_{\min} = \{d_{\min 1}, d_{\min 2}, \dots, d_{\min c}\}. \quad (18)$$

According to the comparison of $\text{dist}_{\max r}$ and $\text{dist}_{\min r}$, all regions are labeled as near- d_{\max} or near- d_{\min}

near- d_{\max} :

$$\begin{aligned} R_{\max} &= \{d_{\min r} \geq d_{\max r} | R_r \in R \equiv d_r \geq 0.5\} \\ &= \{R_{\max 1}, R_{\max 2}, \dots, R_{\max n1}\}, \end{aligned} \quad (19)$$

near- d_{\min} :

$$\begin{aligned} R_{\min} &= \{d_{\min r} < d_{\max r} | R_r \in R \equiv d_r < 0.5\} \\ &= \{R_{\min 1}, R_{\min 2}, \dots, R_{\min n2}\}. \end{aligned} \quad (20)$$

Then the regions are divided into $n1$ complementary parts (R_{\max}) and $n2$ redundant parts (R_{\min}), where $n1 + n2 = c$.

Figure 4(e) shows an example of the region map between Figs. 4(a) and 4(b), in which the regional attributes are represented using the different gray scale ($R_{\max} \rightarrow 255$, $R_{\min} \rightarrow [0, 255)$). The points with stronger brightness express more dissimilarity (i.e., complementarity).

6 Proposed Fusion Rule

Generally speaking, the purpose of image fusion is to preserve all useful information in the source images. In this section, we first briefly review the related knowledge then the details of our proposed multistrategy fusion rule are given.

6.1 Related Knowledge

6.1.1 Commonly used fusion strategy

The multistrategy fusion rule includes two commonly used fusion strategies, i.e., the choose-max fusion strategy and the weighted average fusion strategy, which can be described using Eqs. (21) and (22), respectively.

The choose-max fusion strategy can be written as

$$\text{CF}(x, y) = \begin{cases} \text{CA}(x, y) & aA(x, y) \geq aB(x, y) \\ \text{CB}(x, y) & aA(x, y) < aB(x, y) \end{cases}, \quad (21)$$

where $\text{CF}(x, y)$ is the fused coefficient located at (x, y) and $\text{CA}(x, y)$ and $\text{CB}(x, y)$ are the coefficients of source images located at (x, y) . $aA(x, y)$ and $aB(x, y)$ are the activity measures of $\text{CA}(x, y)$ and $\text{CB}(x, y)$. The salient feature of the coefficient (e.g., variance or gradient) is expressed by the so-called activity. The coefficient with a higher activity measure contains richer information than the others. It should be directly selected as the fused coefficient.

The weighted average fusion strategy can be written as

$$\text{CF}(x, y) = W_1 \text{CA}(x, y) + W_2 \text{CB}(x, y), \quad (22)$$

$$W_1 + W_2 = 1,$$

where the weighted factors W_1 and W_2 are calculated according to the specific task.

In general, when the source images are complementary, the choose-max strategy should be applied. Otherwise, the weighted average strategy should be employed.

6.1.2 Sigmoid function

A sigmoid function is a mathematical function having an ‘‘S’’ shape (sigmoid curve),¹⁸ which is shown in Fig. 5 and is defined as

$$\text{Sf}(k, e) = \frac{1}{1 + \exp[-k \ln(e)]}, \quad (23)$$

where k is the shrink factor and e is the variable; they jointly control the shape of the sigmoid curve.

We plot the sigmoid function Sf with different shrink factors k and e in Fig. 5. The shrink factor k controls the steepness of the sigmoid curve. There is a pair of horizontal asymptotes as $e \rightarrow \pm\infty$. For the same k , when e is very large or very small, Sf approaches $\text{Sf} = 1$ or $\text{Sf} = 0$. However, when e is closer to 1, Sf approaches $\text{Sf} = 0.5$. For the same e , when $k = +\infty$, Sf is equivalent to

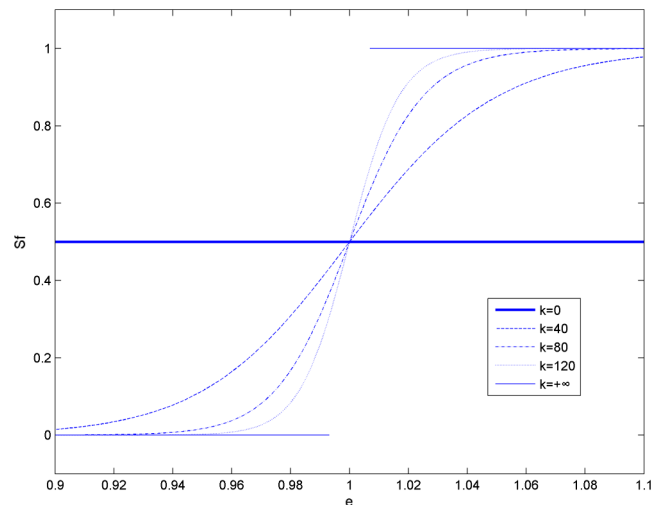


Fig. 5 Sigmoid curve with different shrink factors k and e .

$(1/2) + (1/2)\text{sgn}(e - 1)$, where $\text{sgn}(\cdot)$ is the sign function. When $k = 0$, Sf is equivalent to 0.5.

As can be seen from Fig. 5, the sigmoid function plays two roles: the selection role and the weighted average role, which are determined by k and e . This phenomenon is exactly in line with the choose-max strategy and the weighted average strategy. Moreover, k and e can be represented by the characteristic of the corresponding regions and the difference of activity measures of the corresponding coefficients. Therefore, using the sigmoid function to design a multistrategy fusion rule is appropriate.

6.2 Fusion of the Low-Frequency Subbands

The energy of the source image focuses on the low frequency part and the adjacent coefficients of the source image retain local correlation. Therefore, to get a high-contrast outcome, the low-frequency subbands are fused by the choose-max strategy with the local average energy in the proposed method

$$\text{LF}(x, y) = \begin{cases} \text{LA}(x, y) & \text{EA}(x, y) \geq \text{EB}(x, y) \\ \text{LB}(x, y) & \text{EA}(x, y) < \text{EB}(x, y) \end{cases}, \quad (24)$$

$$\text{EI}(x, y) = \frac{1}{L^2} \sum_{wi=x-\lfloor L/2 \rfloor}^{x+\lfloor L/2 \rfloor} \sum_{wj=y-\lfloor L/2 \rfloor}^{y+\lfloor L/2 \rfloor} [\text{LI}(wi, wj)]^2, \quad (25)$$

$$I = A \text{ or } B,$$

where $\text{LF}(x, y)$ denotes the fused low-frequency subband coefficient located at (x, y) , $\text{LA}(x, y)$ and $\text{LB}(x, y)$ are the low-frequency subband coefficients of the source images located at (x, y) . $\text{EA}(x, y)$ and $\text{EB}(x, y)$ are the local average energies of $\text{LA}(x, y)$ and $\text{LB}(x, y)$, which can be calculated by Eq. (25). L is the size of the local window. Our purpose is to maximally preserve the useful information of the source images.

6.3 Fusion of the High-Frequency Subbands

Since high-frequency subbands tend to contain a lot of image details (such as edges, area boundaries, and so on), the quality of the fusion rule for the high-frequency subband coefficients will obviously affect the fused result. To enhance the quality of the fused image, an adaptive multistrategy fusion rule with a sigmoid function is designed as follows:

$$\text{HF}_{r,j}^S(x, y) = w_1 H_{r,j}^{A,S}(x, y) + w_2 H_{r,j}^{B,S}(x, y),$$

$$w_1 = \frac{1}{1 + \exp\{-k_r \ln[e_{r,j}^S(x, y)]\}},$$

$$w_2 = 1 - w_1 = \frac{\exp\{-k_r \ln[e_{r,j}^S(x, y)]\}}{1 + \exp\{-k_r \ln[e_{r,j}^S(x, y)]\}},$$

$$e_{r,j}^S(x, y) = \frac{E_{r,j}^{A,S}(x, y)}{E_{r,j}^{B,S}(x, y)}, \quad (26)$$

where $\text{HF}_{r,j}^S(x, y)$ represents the fused high-frequency subband coefficient located at (x, y) of the S 'th level and j 'th high-frequency subband for the r 'th region. $\text{HA}_{r,j}^S(x, y)$ and $\text{HB}_{r,j}^S(x, y)$ have similar meanings. The weight w_1 is calculated by the sigmoid function, where k_r is the shrink factor

and $e_{r,j}^S(x, y)$ is the variable. To achieve the adaptive multistrategy fusion rule, k_r and $e_{r,j}^S(x, y)$ should be image dependent.

Considering the local correlation of pixels, the shrink factor based on the region is defined as the characteristics of the corresponding regions S_r . Let $E_{r,j}^{A,S}(x, y)$ and $E_{r,j}^{B,S}(x, y)$ represent the activity measures of the high-frequency coefficients located at (x, y) of the S 'th level and the j 'th high-frequency subband for the r 'th region of images A and B, respectively. Thus, their ratio $e_{r,j}^S(x, y)$ represents the difference of activity measures of the corresponding coefficients. k_r and $e_{r,j}^S(x, y)$ determine the role of the combining strategy. We discuss their details in the following subsections.

6.3.1 Discussion of the k_r

With the same $e_{r,j}^S(x, y)$, the role of the sigmoid function is determined by k_r , which controls the steepness of the sigmoid curve. The selection of a strategy depends on the characteristic of the corresponding regions, therefore, we establish its connection with k_r by $0 \leq k_r \leq +\infty$, $0 \leq d_r \leq 1$,

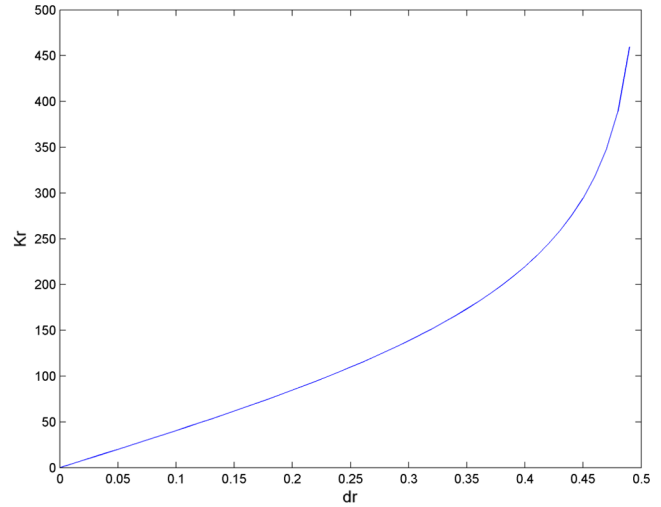


Fig. 6 The relationship between d_r ($0 \leq d_r < 0.5$) and k_r .

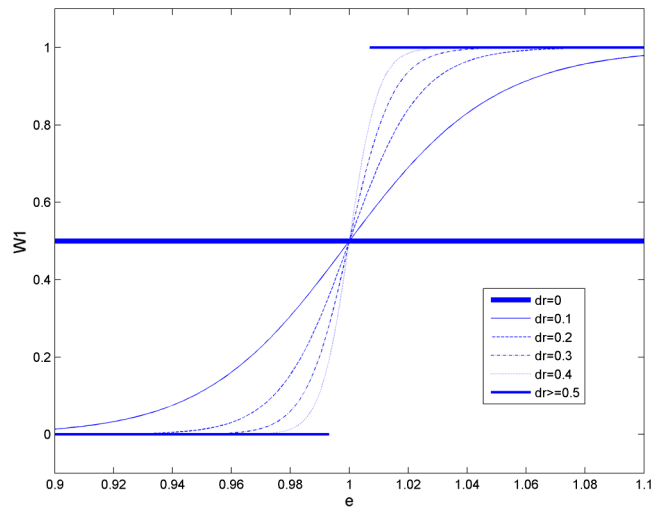


Fig. 7 The relationship between d_r ($0 \leq d_r < 0.5$) and W_1 .

$$k_r = +\infty, \quad \text{if } 0.5 \leq d_r \leq 1, \quad (27)$$

$$k_r = a * \ln\left(\frac{1 + 2d_r}{1 - 2d_r}\right), \quad \text{if } 0 \leq d_r < 0.5, \quad (28)$$

where a is a positive parameter and is fixed at 100 in the proposed method. When $0.5 \leq d_r \leq 1$, it means that the corresponding regions are complementary and the choose-max strategy should be adopted for these regions. This status is equivalent to the sigmoid function with $k_r = +\infty$. When $0 \leq d_r < 0.5$, it means that the corresponding regions are redundant and the weighted average strategy should be employed. This status is equivalent to the sigmoid function with the k_r as Eq. (28).

We then plot the relationship between d_r ($0 \leq d_r < 0.5$) and k_r in Fig. 6. As shown in Fig. 6, the value of k_r increases with the increase of d_r . Moreover, we plot the sigmoid function w_1 with different a d_r in Fig. 7. As can be seen from Fig. 7, for the redundant corresponding regions, with the increase of d_r , the sigmoid function becomes steeper. This change means that the fusion rule is gradually transformed from the weighted average strategy to the choose-max strategy. In this way, not only the type of region (complementarity or redundancy) but also the redundant degree is considered.

6.3.2 Calculation of e

With the same k_r , the variable $e_{r,j}^S(x, y)$ also affects the selection of the fusion strategy. According to the previous

discussion, the difference of activity measures of the corresponding high-frequency coefficients plays an important role in the strategy selection. Then, the $e_{r,j}^S(x, y)$ can be calculated as follows:

$$\text{Mean}_{r,j}^{I,S}(x, y) = \sum_{wi=x-[L/2]}^{x+[L/2]} \sum_{wj=y-[L/2]}^{y+[L/2]} H_{r,j}^{I,S}(wi, wj), \quad (29)$$

$$\begin{aligned} V_{r,j}^{I,S}(x, y) &= \frac{1}{L^2} \\ &= \sum_{wi=x-[L/2]}^{x+[L/2]} \sum_{wj=y-[L/2]}^{y+[L/2]} [H_{r,j}^{I,S}(wi, wj) - \text{Mean}_{r,j}^{I,S}(wi, wj)]^2, \end{aligned} \quad (30)$$

$$E_{r,j}^{I,S}(x, y) = \frac{1}{L^2} \sum_{wi=x-[\frac{L}{2}]}^{x+[\frac{L}{2}]} \sum_{wj=y-[\frac{L}{2}]}^{y+[\frac{L}{2}]} V_{r,j}^{I,S}(wi, wj), \quad (31)$$

$$e_{r,j}^S(x, y) = \frac{E_{r,j}^{A,S}(x, y)}{E_{r,j}^{B,S}(x, y)}, \quad I = A \text{ or } B, \quad (32)$$

where L is the size of the local window. $\text{Mean}_{r,j}^{I,S}(x, y)$ and $V_{r,j}^{I,S}(x, y)$ are the mean and variance of the local window centered at $H_{r,j}^{I,S}(x, y)$. $E_{r,j}^{I,S}(x, y)$ is the averaged variance of

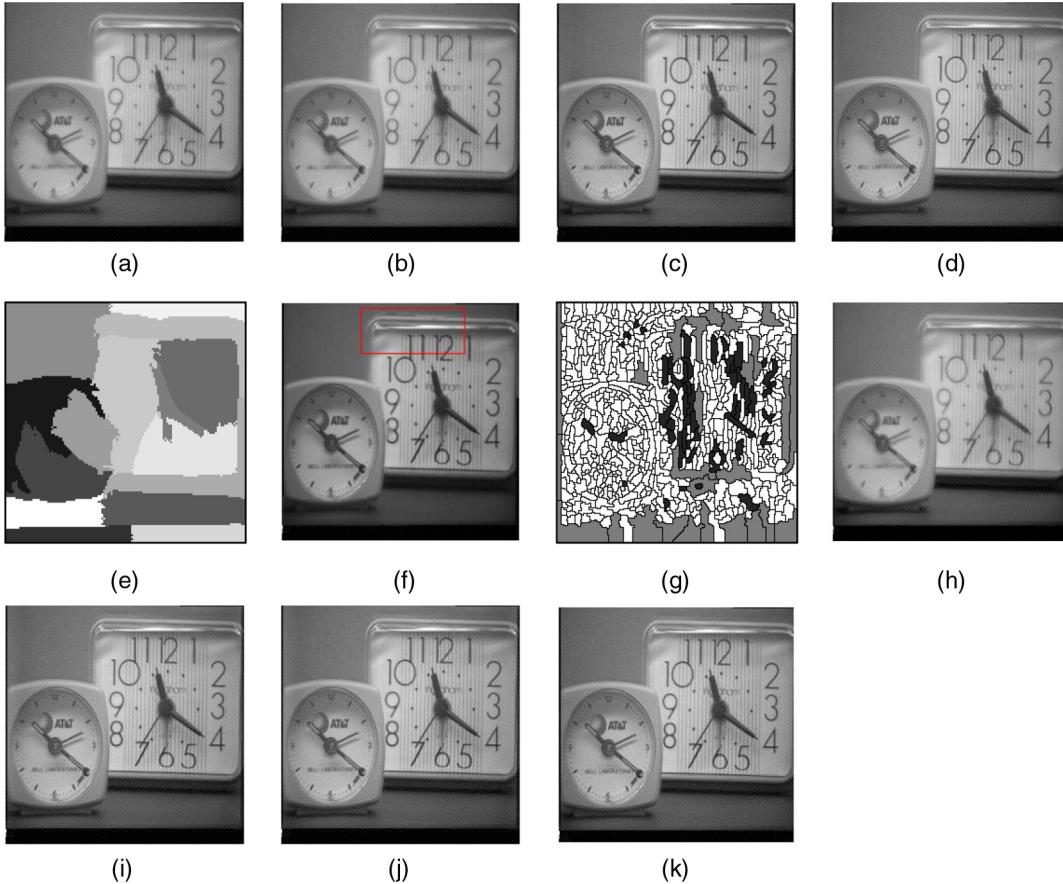


Fig. 8 The segmentation and fused results of multifocus images. (a) AVE, (b) PCA, (c) SIST, (d) GFFC, (e) the region map of RSSF, (f) RSSF, (g) the region map of RF_SSIM, (h) RF_SSIM, (i) the proposed method with $k_r = +\infty$, (j) the proposed method with $k_r = 80$, and (k) the proposed method.

Table 1 Objective comparison of fused results with different methods for multifocus images.

| Methods | G | EI | MI | EN | RCE | Qabf | IQM |
|---------------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|
| AVE | 8.8557 | 48.4706 | 7.1943 | 6.6776 | 0.9259 | 0.6185 | 0.9671 |
| PCA | 8.8535 | 48.4498 | 7.4022 | 6.7528 | 0.7806 | 0.6171 | 0.9671 |
| SIST | 11.1797 | 65.7593 | 6.0321 | 7.2171 | 1.2434 | 0.6646 | 0.9689 |
| GFFC | 10.6486 | 60.9055 | 6.2144 | 7.2683 | 0.8763 | 0.6752 | 0.9622 |
| RSSF | 11.0507 | 65.1473 | 7.2311 | 5.9230 | 0.0116 | 0.6743 | 0.9751 |
| RF_SSIM | 4.7832 | 44.5113 | 7.1210 | 7.2644 | 0.0507 | 0.5904 | 0.9671 |
| $k_r = +\infty$ | 11.2225 | 65.7157 | 6.2069 | 7.2696 | 0.8364 | 0.674 | 0.9768 |
| $k_r = 80$ | 11.2185 | 65.7035 | 6.2107 | 7.2665 | 0.8325 | 0.6755 | 0.9769 |
| The proposed method | 11.3646 | 66.7428 | 6.3394 | 7.2767 | 0.8270 | 0.6843 | 0.9770 |

Note: The bold values represent the best results.

the local window, which takes into account the neighbor dependency and can be considered as the activity measure of $H_{r,j}^{I,S}(x,y)$. Their ratio $e_{r,j}^S(x,y)$ represents the difference of the corresponding coefficients. When the difference is large, it means that the corresponding coefficients are

complementary and the choose-max strategy should be adopted. This status is equivalent to the sigmoid function with a very large or very little $e_{r,j}^S(x,y)$. When the difference is little, it means that the corresponding coefficients are redundant and the weighted average strategy should be

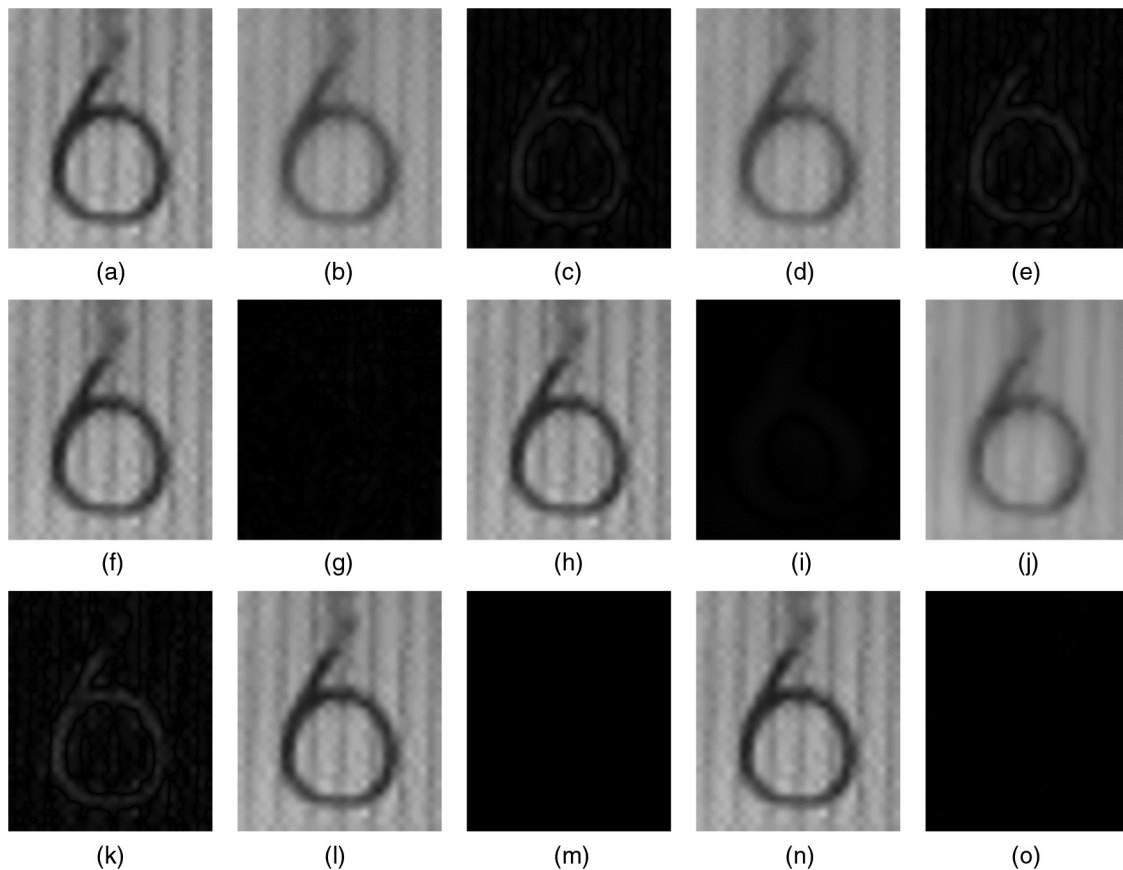


Fig. 9 Parts of Fig. 4(b) and the fused results of Figs. 8(a), 8(b), 8(c), 8(d), 8(f), 8(h), and 8(k) and the subtraction between parts of Fig. 4(b) and parts of the fused results of Figs. 8(a), 8(b), 8(c), 8(d), 8(f), 8(h), and 8(k). (a) imageA_detail, (b) AVE_detail, (c) AVE_sub, (d) PCA_detail, (e) PCA_sub, (f) SIST_detail, (g) SIST_sub, (h) GFFC_detail, (i) GFFC_sub, (j) RF_SSIM_detail, (k) RF_SSIM_sub, (l) RSSF_detail, (m) RSSF_sub, (n) the proposed method_detail, and (o) the proposed method_sub.

employed. This status is equivalent to the sigmoid function with the $e_{r,j}^S(x, y)$ near 1. This phenomenon can be seen from Fig. 7.

7 Experimental Results and Analysis

To verify the proposed method, it is compared with conventional and state-of-the-art fusion methods from the standpoint of visual perception and objective evaluation. The six compared fusion approaches are average-based (AVE-based) approach, PCA-based approach,¹ SIST-based approach,^{2,16} RSSF,¹¹ image fusion incorporating Gabor filters and fuzzy c-means clustering (For simplicity, we call the method GFFC),¹² and RF_SSIM.¹⁴ The choice of these algorithms is motivated by the following reasons: the typical fusion, region-based single strategy fusion, region-based multistrategy fusion, recent fusion methods, and easy reproducibility. Moreover, to demonstrate the superiority of the adaptive

k_r , two experiments with the fixed k_r ($k_r = +\infty, 80$) are executed. Furthermore, to test the superiority of SIST, the proposed method is compared with that using SWT and nonsampling contourlet transform (NSCT).

The quantitative comparison of different fusion algorithms should consider the following aspects: the edge intensity, the amount of information of the fused image, and the relationship between the source images and the fused images. Thus, the following image quality metrics (IQM)²¹⁻²⁵ are used in this paper:

1. The gradient (G) metric, which can reflect the clarity of the fusion image. The larger the G value, the clearer the fused result is.
2. Edge intensity (EI), which reflects the edge intensity of the fused image. The larger the EI value, the more details the fused result has.

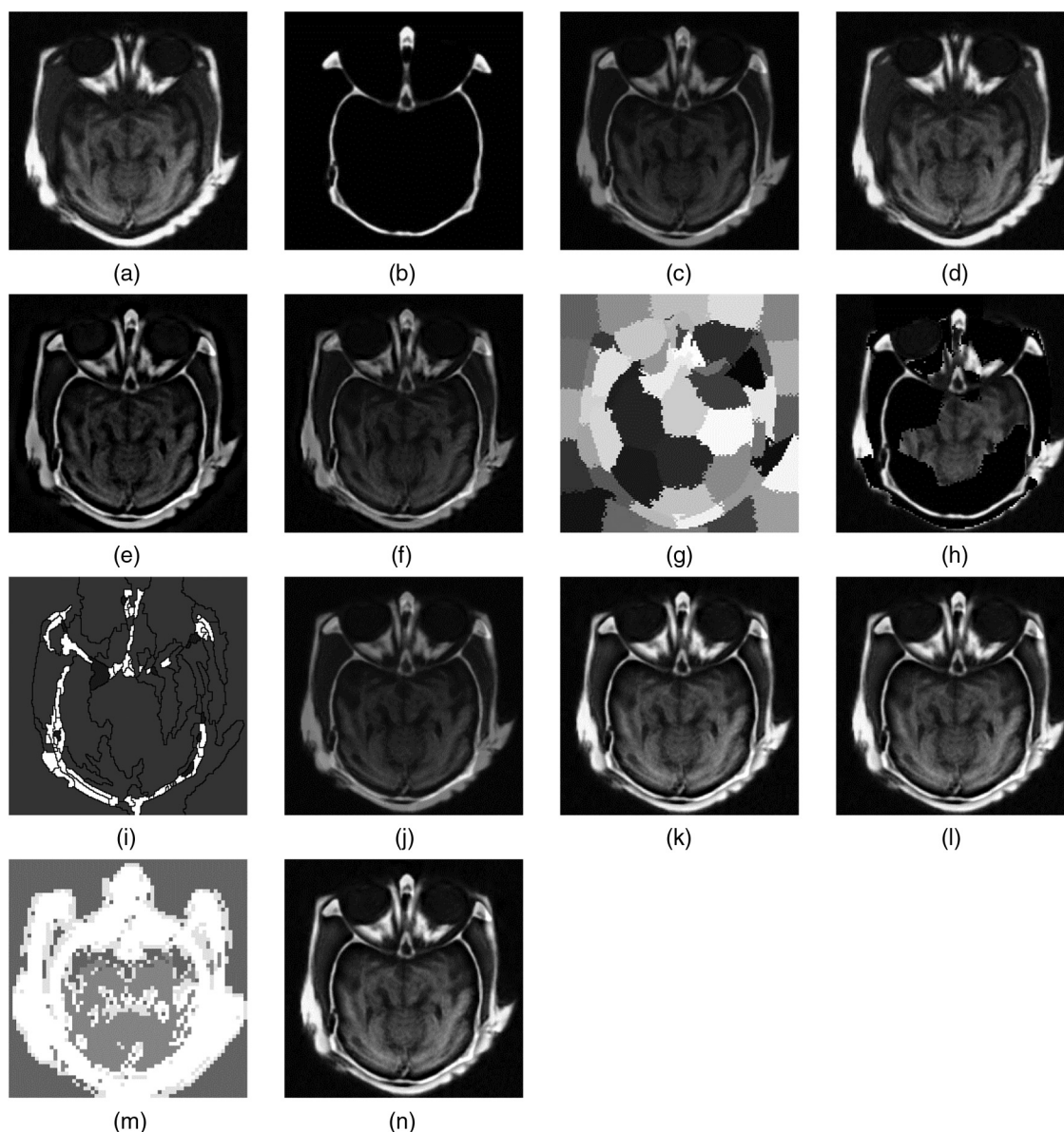


Fig. 10 The segmentation and fused results of medical images. (a) MRI image, (b) CT image, (c) AVE, (d) PCA, (e) SIST, (f) GFFC, (g) the region map of RSSF, (h) RSSF, (i) the region map of RF_SSIM, (j) RF_SSIM, (k) the proposed method with $k_r = +\infty$, (l) the proposed method with $k_r = 80$, (m) the region map of the proposed method, and (n) the proposed method.

3. Mutual information (MI), which reflects the total amount of information that the fused image inherited from the source images.
4. Entropy (EN) reflects the amount of information of the fused image. The larger the EN is, the more information the image carries if there is no noise in the image.
5. Overall cross entropy (RCE) is used to calculate the difference between the source images and the fused images. The smaller the RCE is, the better the fusion result that is obtained.
6. The Qabf metric considers the amount of edge information transferred from the source images to the fused images. The larger the Qabf value, the better the fused result is. It should be as close to 1 as possible.
7. An IQM-based metric, which is designed by modeling any image distortion as a combination of three factors: loss of correlation, radiometric distortion, and contrast distortion. The larger the IQM value, the better the fused result is. It should also be as close to 1 as possible.

For the SIST-based method and the proposed method, the pyramid filter of SIST is set as “maxflat (maximally flat filter)” and the source images are decomposed into four levels with the number of directions 6, 6, 10, and 10. Furthermore, the number of clusters $c = 10$, the size of the window $M \times M = 4 \times 4$ and the sliding window size $L = 5$ are applied to the proposed method. The effects of parameters (c, M, L) on the proposed method are discussed in Sec. 7.3.

7.1 Experiment 1: Performance Evaluation of the Proposed Fusion Method

7.1.1 Fusion results of multifocus images

Because the optical imaging cameras cannot capture objects at various distances all in focus, the multifocus images are fused to get a fused image with the focused parts of all the multifocus images. Figures 4(a) and 4(b) are a pair of multifocus images which have a common scene structure. Figure 8 and Table 1 show the fused images and evaluation results, and Fig. 9 shows the detailed blocks (the “6” in the right

clock) of the source images and the fusion results. Moreover, for a clearer comparison, the subtractions between the parts of Fig. 4(b) and the parts of the fusion results are also illustrated in Fig. 9. As shown in Fig. 9, Figs. 9(c) and 9(e) have a lot of residual information, which means that the traditional fusion methods (the AVE-based method and the PCA-based method) reduce contrast. This results from the fact that the AVE-based method just takes the pixel-by-pixel gray level average of the source images and the PCA-based method uses the weighted average rule with the eigenvalues of the covariance of the source images. Figures 8(c) and 8(f) are the fusion results of the SIST-based approach and RSSF. These methods employ a single fusion strategy. Although there is almost no residual information in Fig. 9(m), some incorrect fusion blocks are found in the rectangle region of Fig. 8(f). The reason for this lies in the fact that some incorrect segmentations worsen the region-based choose-max fusion rule. For example, as shown in Fig. 8(e), some pixels located at the upper boundary of the right clock are segmented into different regions. Figure 8(c) is the fused result of the SIST-based fusion approach, which adopts a coefficient-based choose-max fusion rule. Figure 9(g) shows that some information of Fig. 4(b) is lost in the fusion process. The fusion results of multistrategy fusion methods (GFFC and RF_SSIM) are shown in Figs. 8(d) and 8(h). As is evident, the contrast of the fused images is reduced, especially in Fig. 8(h). This is further illustrated by Figs. 9(i) and 9(k). This results because their multistrategy fusion rules ignore the activity difference among the coefficients in a region. As can be seen from Fig. 8(k), almost all the useful information of the source images has been transferred into the fused image by our proposed method. Moreover, we can see from Fig. 9(o) that it is almost black; this means that the residual image between Figs. 9(a) and 9(n) is very small. This observation further illustrates the information integration ability of the proposed method.

Considering the similar fused results of Figs. 8(i)–8(k), objective evaluations are performed and the results are listed in Table 1. It can be concluded that most of the objective

Table 2 Objective comparison of fused results with different methods for medical images.

| Methods | G | EI | MI | EN | RCE | Qabf | IQM |
|---------------------|---------------|----------------|---------------|---------------|---------------|---------------|---------------|
| AVE | 3.9869 | 40.5312 | 5.2315 | 5.9196 | 5.8537 | 0.4281 | 0.5799 |
| PCA | 5.6102 | 57.2228 | 6.3220 | 6.5953 | 6.1205 | 0.6584 | 0.7228 |
| SIST | 7.5993 | 76.8942 | 2.1638 | 6.0246 | 2.7043 | 0.7365 | 0.5920 |
| GFFC | 5.0099 | 49.6637 | 3.0257 | 6.0881 | 3.4530 | 0.5218 | 0.5849 |
| RSSF | 7.3646 | 71.2930 | 4.8933 | 5.5766 | 0.8956 | 0.6156 | 0.4309 |
| RF_SSIM | 3.9554 | 40.0386 | 4.5929 | 5.8168 | 1.4252 | 0.4284 | 0.3186 |
| $k_r = +\infty$ | 7.5349 | 76.5274 | 2.8973 | 6.6846 | 3.5521 | 0.7352 | 0.6660 |
| $k_r = 80$ | 7.5332 | 76.5101 | 2.8987 | 6.6847 | 3.5476 | 0.7355 | 0.6661 |
| The proposed method | 7.7319 | 78.5406 | 3.3536 | 6.8597 | 4.4816 | 0.7741 | 0.6798 |

Note: The bold values represent the best results.

evaluation results are in reasonable agreement with the visual effect. For example, the low contrasts of Figs. 8(a), 8(b), and 8(h) are quantified by the lagging indicators (G , EI , and Q_{abf}). The fused image of AVE, PCA, RSSF, and RF_SSIM obtained better results in terms of MI due to being performed in the spatial domain. RSSF directly copies original regions from the source images to the result image, so the pixel distributions of Fig. 8(f) are changed very little. The best RCE of the RSSF confirms this observation. By comparison, we can conclude that our proposed method outperforms the other methods using a regional activity measure based on a multistrategy fusion rule (GFFC and RF_SSIM).

In addition, the fixed k_r ($k_r = +\infty, 80$) and the adaptive k_r are compared. As can be seen from the last three rows of

Table 1, since an adaptive k_r makes the fusion strategy flexible, our proposed method with an adaptive k_r achieves higher performances than those with a fixed k_r .

Overall, although the result of the proposed method is slightly inferior to that of PCA in terms of MI and that of RSSF in terms of RCE, the result is superior to that of other methods in other terms. This means that the result image fused by our method contains more details and a greater amount of information.

7.1.2 Fusion results of medical images

It is easy for physicians to understand the lesion by reading images of different modalities with a multimodal medical

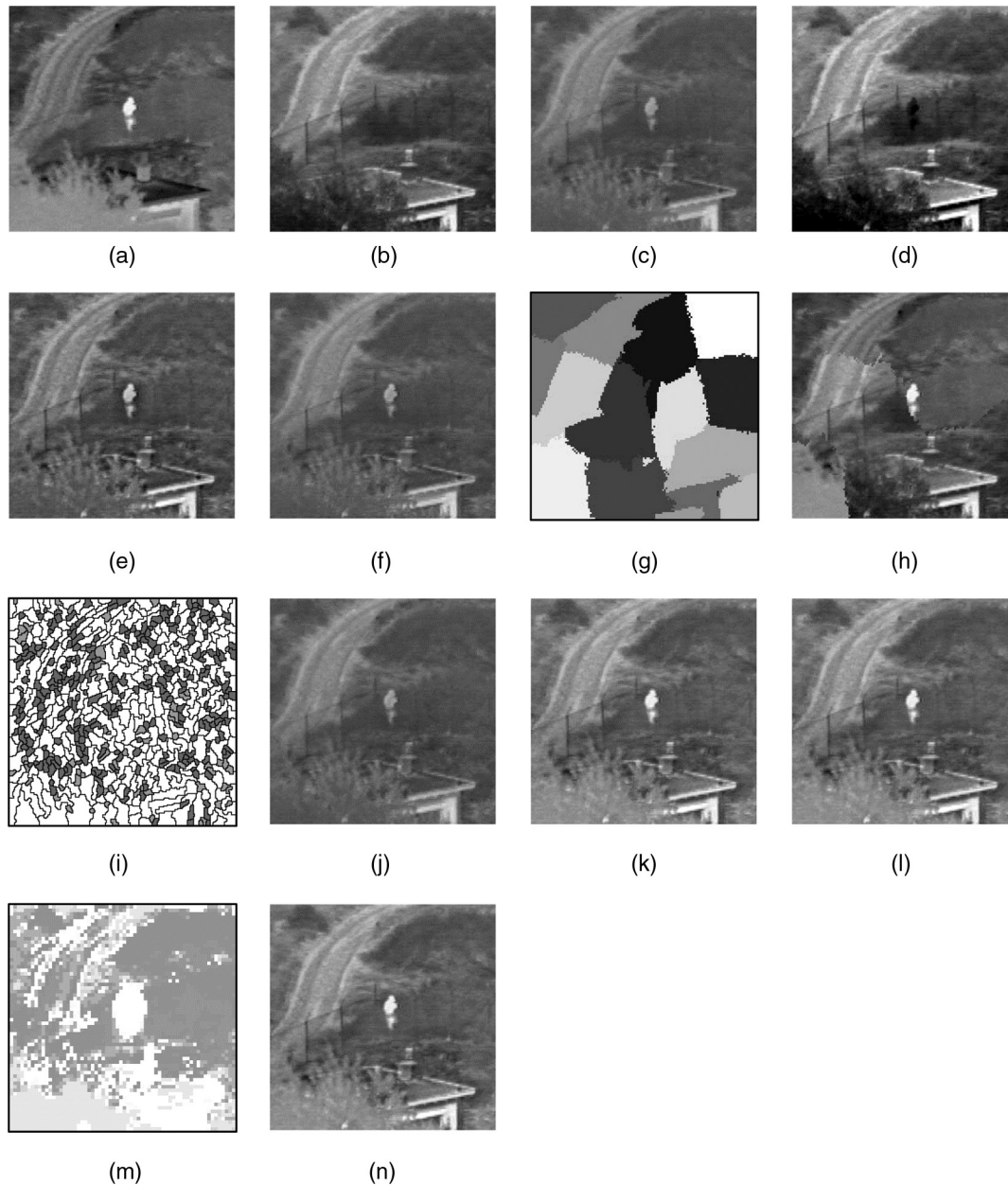


Fig. 11 The segmentation and fused results of infrared and visual images. (a) Infrared image, (b) visual image, (c) AVE, (d) PCA, (e) SIST, (f) GFFC, (g) the region map of RSSF, (h) RSSF, (i) the region map of RF_SSIM, (j) RF_SSIM, (k) the proposed method with $k_r = +\infty$, (l) the proposed method with $k_r = 80$, (m) the region map of the proposed method, and (n) the proposed method.

image fusion. For example, fused magnetic resonance imaging/computed tomography (MRI/CT) imaging can concurrently visualize anatomical and physiological characteristics of the human body for diagnosis and treatment planning.²⁶ In this section, the fused results of MRI/CT are shown in Fig. 10. It is not hard to see that the MRI [Fig. 10(a)] and CT [Fig. 10(b)] are obviously complementary, but the MRI contains more information.

By subjective evaluation, the result of the AVE-based method averaging the source images [Fig. 10(c)] leads to low contrast. A large amount of information exists in MRI. The weight with the eigenvalues of the covariance of the source images is biased toward MRI, thus the fused image of the PCA-based method [Fig. 10(d)] loses the information of the CT. Owing to the single choose-max strategy and the region-based activity measure, as shown in Fig. 10(h), RSSF is inclined to choose inappropriate regions. Due to the region based activity measure, the fusion result of RF_SSIM (a multistrategy fusion method) lacks the contrast. However, because our method adopts an adaptive multistrategy fusion rule which is based on local structural characteristics and a coefficient-based activity measure, our result achieves better visual perception.

The objective evaluations are listed in Table 2. It can be observed that the PCA method wins in terms of MI and IQM. This is because the PCA method is biased toward MRI and the amount of information of the MRI is far greater than that of the CT. The low contrast fused images of AVE, GFFC, and RF_SSIM yield worse performances in terms of G , EI, and Qabf; this fact is in line with the visual perception. Our proposed method has the best value in terms of the other four indices (i.e., G , EI, EN, and Qabf) among all seven indices.

7.1.3 Fusion results of infrared-visual images

Usually, visible images can provide spatial details of the background of the objects but cannot reveal the objects. However, infrared images can capture the objects but fail to reveal some background areas. The fused image displays

both the objects and spatial details of the background, which is a potential solution for improving target detection.

In Fig. 11, we show an example of the visual-infrared images. The luminance of the object (the person) in Figs. 11(c)–11(f) and 11(j) decreases compared with the infrared image Fig. 11(a). In Fig. 11(d), the object is changed into black from white. This change is disadvantageous to subsequent object processing. Although Fig. 11(i) preserves the luminance of the object and obtains the best values in terms of MI and Qabf, there are some apparent image stitches which seriously affect the visual perception. Obviously, our fused image has a better visual perception than the others. Moreover, our method also outperforms the others in terms of G , EI, and EN, which are given in Table 3. For the MI, our method gets the largest value except for those of PCA and RSSF. Furthermore, the objective evaluation confirms that the proposed method with an adaptive k_r is superior to the method with a fixed k_r in the fusion of infrared-visual images.

It should also be noticed that the improved performance of the proposed method is at the cost of an increasing computation complexity. As shown in the last column of Table 3, the consumed time of the proposed method is more than those of the other fusion methods. The increased time is mainly the result of the computation of the region map and the activity measure. This may limit its application in some real-time cases.

7.1.4 Fusion results of other images

To further evaluate the proposed method's robustness, a series of image fusion experiments are performed on five pairs of source images. Considering the limitation of space, Fig. 12 only shows the source images, the region maps, and the fused results of the proposed method. Moreover, in Table 4, only the fused results obtained by GFFC, RSSF, RF_SSIM, and the proposed method with $k_r = 80$ and $k_r = +\infty$ are given. Since the evaluation indices MI, RCE, and Qabf depend on the consistency of gray distribution or the pixel values between the fused images and source images, the simple copying source region method RSSF wins in

Table 3 Objective comparison of fused results with different methods for infrared and visual images.

| Methods | G | EI | MI | EN | RCE | Qabf | IQM | T (s) |
|---------------------|---------------|----------------|---------------|---------------|---------------|---------------|---------------|---------|
| AVE | 3.3481 | 29.9876 | 1.8725 | 6.3040 | 0.6208 | 0.3401 | 0.6184 | 1.2 |
| PCA | 5.5246 | 50.2872 | 4.3561 | 6.5213 | 0.4472 | 0.4014 | 0.4947 | 1.3 |
| SIST | 5.6794 | 51.1528 | 1.7532 | 6.6327 | 0.3762 | 0.4638 | 0.6310 | 5.8 |
| GFFC | 3.8845 | 34.1798 | 1.8256 | 6.3391 | 0.6608 | 0.3903 | 0.6224 | 53.5 |
| RSSF | 5.3727 | 48.9421 | 6.3428 | 6.9785 | 0.1560 | 0.5844 | 0.4636 | 10.2 |
| RF_SSIM | 3.6728 | 31.9834 | 1.8052 | 6.3160 | 0.5977 | 0.3437 | 0.5128 | 20.5 |
| $k_r = +\infty$ | 5.6840 | 51.3021 | 2.0485 | 7.0086 | 0.1486 | 0.4565 | 0.5634 | 63.5 |
| $k_r = 80$ | 5.6806 | 51.2826 | 2.0537 | 7.0088 | 0.1489 | 0.4566 | 0.5634 | 60.3 |
| The proposed method | 5.7006 | 51.4643 | 2.1552 | 7.0190 | 0.1488 | 0.4700 | 0.5709 | 90.1 |

Note: The bold values represent the best results.

term of these indices. However, the fused result of RSSF often contains apparent image stitches, which seriously affect the visual perception. From the other four metrics, the proposed method provides the best objective data, which illustrates that more details from the source images are retained. It has been verified that the proposed adaptive multistrategy fusion rule is more beneficial for fusion results than these methods, including the single fusion strategy (RSSF), multistrategy fusion rules based on regional activity measure (GFFC, RF_SSIM), and our method with a fixed k_r ($k_r = +\infty, 80$).

7.2 Experiment 2: Advantages of SIST Over SWT and NSCT

To verify the superiority of SIST, the proposed method is evaluated in different transform domains [SWT and non-subsampled contourlet transform (NSCT)] on the multifocus clock images [Fig. 4(a) and 4(b)] and MRI/CT images [Figs. 10(a) and 10(b)]. In our experiments, the images are all decomposed into four levels by SWT (with the basic function of sym4), NSCT, and SIST. The decomposition level of NSCT is set as [2, 2, 3, 3]. The pyramid filter “9-7” (Gaussian or Laplacian pyramid decomposition filter)

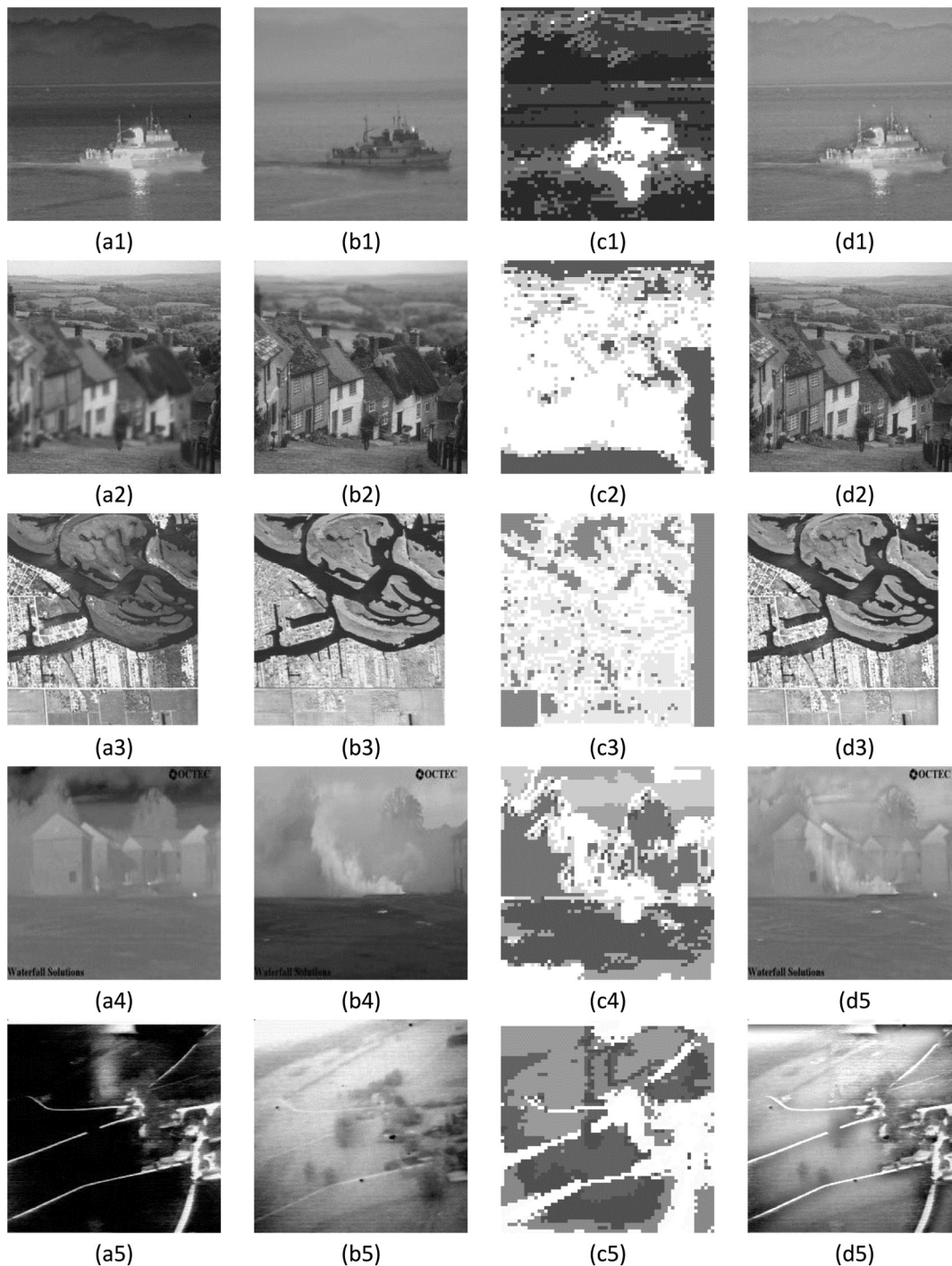


Fig. 12 The region map (c) and fused results (d) of source images (a) and (b) for the other five fusion models.

Table 4 Objective comparison of fused results with different methods for the other five kinds of images.

| Row | Methods | G | EI | MI | EN | RCE | Qabf | IQM |
|-----|---------------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|
| 1 | GFFC | 2.5030 | 19.1531 | 1.7370 | 4.8796 | 3.4849 | 0.5106 | 0.2429 |
| | RSSF | 2.9936 | 24.1457 | 7.6596 | 5.3016 | 2.3407 | 0.6837 | 0.3115 |
| | RF_SSIM | 2.9736 | 22.4230 | 1.6876 | 5.4488 | 2.2049 | 0.3707 | 0.3985 |
| | $k_r = +\infty$ | 2.9723 | 24.1265 | 2.5478 | 5.8245 | 2.2976 | 0.5145 | 0.4012 |
| | $k_r = 80$ | 2.9712 | 24.1076 | 2.5639 | 5.8312 | 2.2932 | 0.5141 | 0.4011 |
| | The proposed method | 2.9941 | 24.6837 | 2.6638 | 5.9292 | 2.2755 | 0.5291 | 0.4057 |
| 2 | GFFC | 9.0403 | 68.9348 | 7.6037 | 7.4419 | 0.0078 | 0.7739 | 0.9812 |
| | RSSF | 9.3988 | 72.8833 | 9.4811 | 7.4563 | 0.0075 | 0.7837 | 0.9894 |
| | RF_SSIM | 6.3186 | 55.1418 | 7.2995 | 7.4048 | 0.0079 | 0.6993 | 0.9855 |
| | $k_r = +\infty$ | 9.5975 | 74.1305 | 8.8239 | 7.4511 | 0.0086 | 0.7763 | 0.9894 |
| | $k_r = 80$ | 9.5947 | 74.1049 | 8.8302 | 7.4509 | 0.0086 | 0.7760 | 0.9894 |
| | The proposed method | 9.6856 | 74.7176 | 8.7227 | 7.4581 | 0.0094 | 0.7765 | 0.9896 |
| 3 | GFFC | 13.5482 | 101.711 | 4.7248 | 7.1364 | 0.4248 | 0.6163 | 0.9096 |
| | RSSF | 15.3697 | 120.133 | 7.9555 | 7.0803 | 0.3956 | 0.7091 | 0.9080 |
| | RF_SSIM | 11.8372 | 90.2916 | 4.5514 | 7.1076 | 0.4468 | 0.5599 | 0.9163 |
| | $k_r = +\infty$ | 16.1784 | 125.133 | 4.1323 | 7.2312 | 0.3826 | 0.6361 | 0.9261 |
| | $k_r = 80$ | 16.1772 | 125.123 | 4.1325 | 7.2322 | 0.3824 | 0.6358 | 0.9261 |
| | The proposed method | 16.1861 | 125.161 | 4.1326 | 7.2328 | 0.3823 | 0.6362 | 0.9361 |
| 4 | GFFC | 2.3742 | 20.2089 | 3.4018 | 6.2072 | 2.1039 | 0.5145 | 0.6015 |
| | RSSF | 2.7561 | 23.9129 | 7.4507 | 5.9900 | 0.4874 | 0.6290 | 0.6308 |
| | RF_SSIM | 2.6396 | 21.6345 | 3.1522 | 5.4524 | 1.9532 | 0.4838 | 0.4430 |
| | $k_r = +\infty$ | 2.8450 | 24.7367 | 3.5270 | 6.4237 | 1.7445 | 0.5405 | 0.6533 |
| | $k_r = 80$ | 2.8439 | 24.7304 | 3.5274 | 6.4217 | 1.7415 | 0.5406 | 0.6533 |
| | The proposed method | 2.8539 | 24.7397 | 3.5369 | 6.4395 | 1.7334 | 0.5409 | 0.6533 |
| 5 | GFFC | 9.6882 | 67.6715 | 2.7741 | 7.1743 | 3.2466 | 0.5154 | 0.4027 |
| | RSSF | 9.5176 | 74.6358 | 6.6581 | 5.9680 | 1.7514 | 0.6617 | 0.4068 |
| | RF_SSIM | 6.3234 | 47.2927 | 3.2792 | 7.1526 | 0.9412 | 0.4007 | 0.3801 |
| | $k_r = +\infty$ | 11.3284 | 83.4447 | 2.6649 | 7.5098 | 3.5562 | 0.5647 | 0.4012 |
| | $k_r = 80$ | 11.3264 | 83.4357 | 2.6656 | 7.5097 | 3.5637 | 0.5663 | 0.4036 |
| | The proposed method | 11.3462 | 84.4354 | 2.6659 | 7.5398 | 3.4451 | 0.5775 | 0.4108 |

Note: The bold values represent the best results.

and the directional filter “pkva” (directional filter banks decomposition filter) are selected in the NSCT domain. Figures 8(k) and 10(n) show the fusion results of the proposed method using SIST in Figs. 4(a), 4(b), 10(a), and 10(b). Figure 13 shows the fusion results of the proposed method using SWT and NSCT in Figs. 4(a), 4(b), 10(a), and 10(b). Tables 5 and 6 provide an objective comparison among different transformation domains. The results presented in these tables show that, in general, the proposed fusion method has the best performance in the SIST domain. This is mainly because the SIST transform does not restrict the number of directions for the shearing and its computation is more efficient than NSCT and SWT.

7.3 Experiment 3: the Effects of Parameters on the Proposed Method

7.3.1 Effect of the number of clusters on the proposed method

In this experiment, the effect of the number of clusters on the performance of the proposed method is investigated. Due to space limitation, only Figs. 11(a), 11(b), 12(a2), and 12(b2) are given for examples. The reason is that they represent

two typical kinds of source images: Figs. 11(a) and 11(b) represent the images with different clarity and different information, Figs. 12(a2) and 12(b2) represent the same information but with different clarity. Let $M \times M = 4 \times 4$, $L = 5$ and the number of clusters varies from 2 to 50, the corresponding EN, RCE, and Qabf performance metrics are plotted in Figs. 14(a), 14(b), and 14(c), respectively. As the number of clusters increases ($c \geq 10$), the EN and RCE values of the two fused images and the Qabf values of Figs. 12(a2) and 12(b2) only vary by a small amount. When the number of clusters is 10, the Qabf values of Figs. 11(a) and 11(b) achieve their largest values. However, the time consumption increases with the increase of the number of clusters. To balance the time consumption and the quality of image fusion, the number of clusters is set as 10 in our proposed method.

7.3.2 Effect of window size on the proposed method

Let the window size vary from 4 to 30 under the conditions of $c = 10$ and $L = 5$.

Correspondingly, the values of EN, RCE, and Qabf are plotted in Figs. 15(a), 15(b), and 15(c), respectively. As

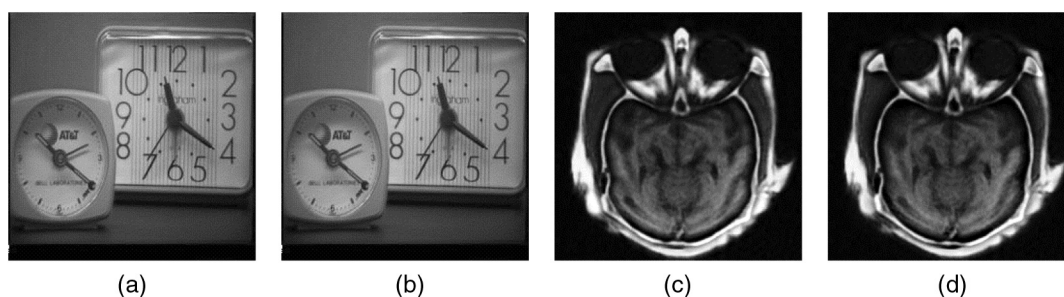


Fig. 13 The fusion results of clock images [e.g., Figs. 4(a) and 4(b)] and medical images [e.g., Figs. 10(a) and 10(b)] using the proposed method in different transformation domains. (a) The fusion results of clock images using the proposed method with SWT. (b) The fusion results of clock images using the proposed method with NSCT. (c) The fusion results of medical images using the proposed method with SWT. (d) The fusion results of medical images using the proposed method with NSCT.

Table 5 The advantages of using SIST versus SWT and NSCT [Figs. 4(a) and 4(b)].

| The proposed method | G | EI | MI | EN | RCE | Qabf | IQM |
|---------------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|
| Using SWT | 11.3185 | 66.5561 | 6.3285 | 7.2185 | 0.6949 | 0.6752 | 0.9766 |
| Using NSCT | 11.3268 | 66.6764 | 6.1955 | 7.3171 | 0.9228 | 0.6717 | 0.9768 |
| Using SIST | 11.3644 | 66.7428 | 6.3294 | 7.2767 | 0.8270 | 0.6842 | 0.9770 |

Note: The bold values represent the best results.

Table 6 The advantages of using SIST versus SWT and NSCT [Figs. 10(a) and 10(b)].

| The proposed method | G | EI | MI | EN | RCE | Qabf | IQM |
|---------------------|---------------|----------------|---------------|---------------|---------------|---------------|---------------|
| Using SWT | 7.6317 | 77.4339 | 4.5546 | 6.7712 | 5.2380 | 0.7412 | 0.6791 |
| Using NSCT | 7.5244 | 76.3924 | 3.4962 | 6.7778 | 4.3564 | 0.7660 | 0.6793 |
| Using SIST | 7.7308 | 78.5406 | 3.3534 | 6.8597 | 4.4816 | 0.7740 | 0.6797 |

Note: The bold values represent the best results.

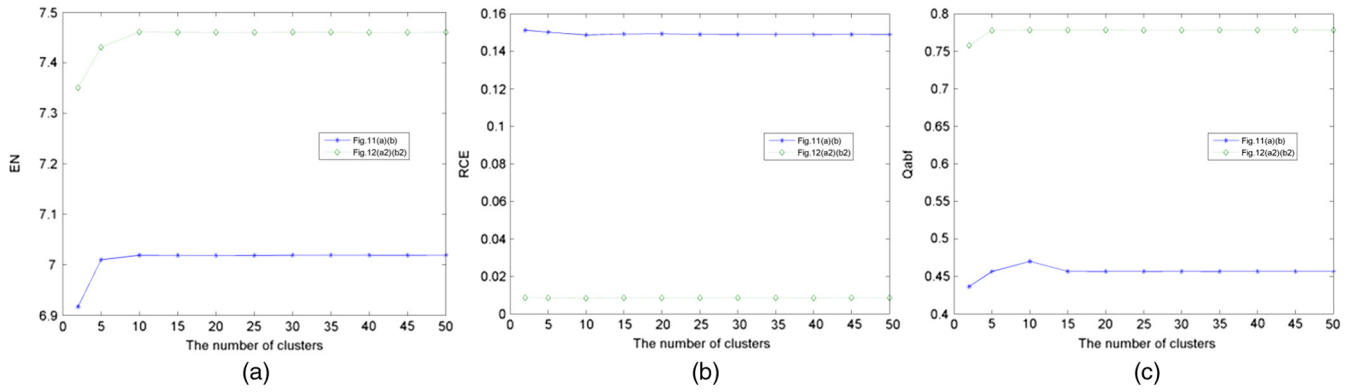


Fig. 14 The effect of the number of clusters on the proposed method. (a) The effect of the number of clusters on EN. (b) The effect of the number of clusters on RCE. (c) The effect of the number of clusters on Qabf.

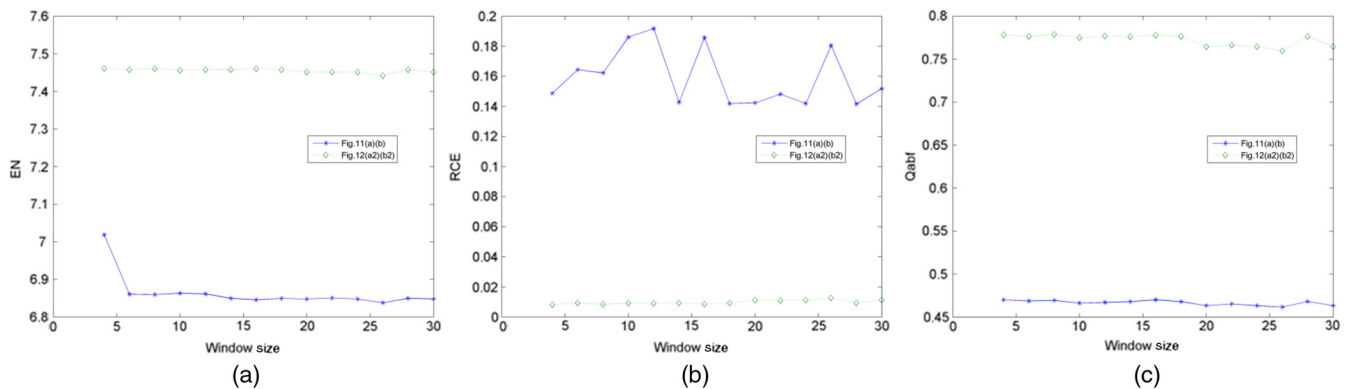


Fig. 15 The effect of the number of clusters on the proposed method. (a) The effect of window size on EN. (b) The effect of window size on RCE. (c) The effect of window size on Qabf.

can be seen, for Figs. 12(a2) and 12(b2), the values of EN, RCE, and Qabf are stable with the increase of window size. For Figs. 10(a) and 10(b), the EN value (from 6 to 30) and the Qabf value (from 4 to 30) vary slightly. However, the RCE value is volatile. By the statistical analysis of Figs. 11(a) and 11(b), we can find that the EN metric has the largest value and the RCE metric shows better fusion properties when the window size of Figs. 10(a) and 10(b) is 4. Thus, a window size of 4×4 is a reasonable selection.

7.3.3 Effect of sliding window size on the proposed method

This experiment explores the effect of sliding window size L on the performance of the proposed method. With the fixed window size 4×4 and $c = 10$, the sliding window size varies from 3 to 23 in Fig. 16. As can be seen from Fig. 16, the variation of the sliding window size has little effect on EN, RCE, and Qabf for Figs. 12(a2) and 12(b2). For Figs. 11(a) and 11(b), the metrics have a small fluctuation with the increase of the sliding window size and

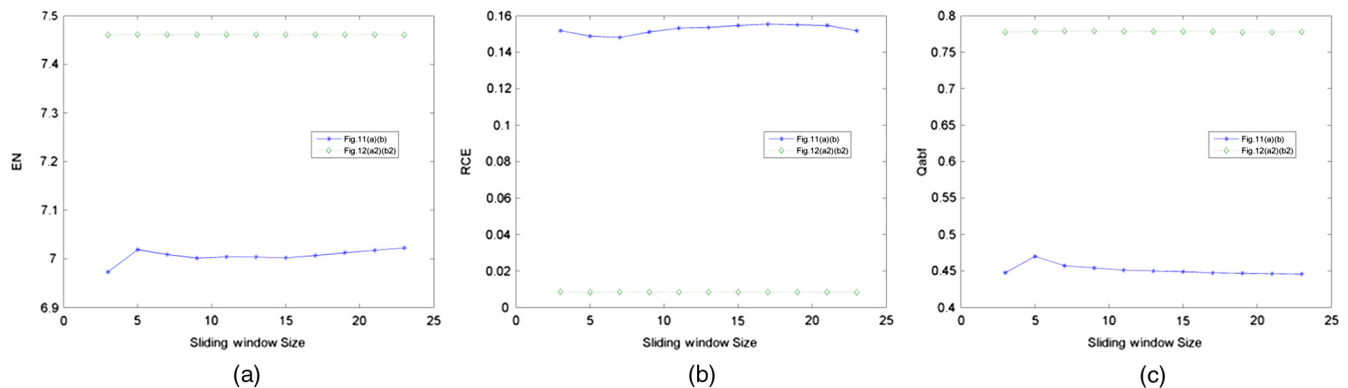


Fig. 16 The effect of sliding window size on the proposed method. (a) The effect of sliding window size on EN. (b) The effect of sliding window size on RCE. (c) The effect of sliding window size on Qabf.

the best performance is achieved with the condition of $L = 5$. Therefore, the sliding window size in our proposed method is assigned as 5.

8 Conclusion

In this paper, an adaptive multistrategy image fusion method has been proposed. A multiscale image decomposition tool and a multistrategy fusion rule are the two key components in the proposed method. The SIST is adopted as the multiscale analysis tool in the proposed image and the source images are decomposed into low-frequency subbands and high-frequency subbands. The choose-max fusion strategy is employed to fuse the low-frequency subbands which contain the approximate information of the source images. An adaptive multistrategy fusion rule with a sigmoid function has been proposed to distinguish the attributes of the high-frequency subbands and can fuse them automatically. The dissimilarity of corresponding regions and the activity measure of the high-frequency coefficient are employed to identify the attributes between the high-frequency coefficients. These are used as the variables of the sigmoid function. They determine the curve of the sigmoid function with different steepnesses, which correspond to the different fusion strategies. By using the sigmoid function, the adaptive selection of the fusion strategy is achieved. Several sets of experimental results demonstrate the validity, flexibility, and generality of the proposed method in terms of both visual quality and objective evaluation. It should be noted that although the proposed method achieves better results, the computational complexity is a bit high since several techniques have been incorporated into the proposed method. Thus, how to optimize time consumption will be one of our future works. The other future work is to investigate more effective functions for adaptive fusion strategy selection.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61103128 and 61373055, the China Postdoctoral Science Foundation under Grant No. 2013M541601, and Postdoctoral Research Funds of Jiangsu province under Grant No. 1301079C. The authors would like to thank the anonymous reviewers for their detailed reviews, valuable comments, and constructive suggestions.

References

- H. Yesou, Y. Besnus, and J. Rolet, "Extraction of spectral information from landsat TM data and merger with SPOT panchromatic imagery-A contribution to the study of geological structures," *ISPRS J. Photogramm. Remote Sens.* **48**(5), 23–26 (1993).
- H. Li, S. Manjunath, and S. Mitra, "Multi sensor image fusion using the wavelet transform," *Graph. Models Image Process.* **57**(3), 235–245 (1995).
- S. Zheng et al., "Multisource image fusion method using support value transform," *IEEE Trans. Image Process.* **16**(7), 1831–1840 (2007).
- J. Tian and L. Chen, "Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure," *Signal Process.* **92**(9), 2137–2146 (2012).
- H. Li et al., "Multifocus image fusion and denoising scheme based on homogeneity similarity," *Opt. Commun.* **285**(2), 91–100 (2012).
- C. Shi, Q. G. Miao, and P. F. Xu, "A novel algorithm of remote sensing image fusion based on shearlets and PCNN," *Neurocomputing* **117**, 47–53 (2013).
- G. S. El-taweel and A. K. Helmy, "Image fusion scheme based on modified dual pulse coupled neural network," *IET Image Process.* **7**(5), 407–414 (2013).
- G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Inf. Fusion* **4**(4), 259–280 (2003).
- Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- T. Chen, J. P. Zhang, and Y. Zhang, "Remote sensing image fusion based on ridgelet transform," in *Proc. of Int. Conf. on Geoscience and Remote Sensing Symposium*, Seoul, Korea, pp. 1150–1153 (2005).
- S. Li and B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Image Vision Comput.* **26**(7), 971–979 (2008).
- X. J. Wu, D. X. Su, and X. Q. Luo, "A new similarity function for region based image fusion incorporating Gabor filters and fuzzy c-means clustering," *Proc. SPIE* **6625**, 66250Z (2007).
- Q. Zhang et al., "Similarity-based multimodality image fusion with shiftable complex directional pyramid," *Pattern Recogn. Lett.* **32**(13), 1544–1553 (2011).
- X. Y. Luo, J. Zhang, and Q. H. Dai, "A region image fusion based on similarity characteristics," *Signal Process.* **92**(5), 1268–1280 (2012).
- X. Luo, Z. Zhang, and X. Wu, "Image fusion using region segmentation and sigmoid function," in *Proc. of the 22nd Int. Conf. on Pattern Recognition (ICPR)*, pp. 1049–1054, Stockholm, Sweden (2014).
- L. Wang, B. Li, and L. Tian, "Multi-modal medical image fusion using the inter-scale and intra-scale dependencies between image shift-invariant shearlet coefficients," *Inf. Fusion* **19**(9), 20–28 (2014).
- G. Easley, D. Labate, and W. Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. A* **25**(1), 25–46 (2008).
- J. L. Liang et al., "Image fusion using higher order singular value decomposition," *IEEE Trans. Image Process.* **21**(5), 2898–2909 (2012).
- J. Yang et al., "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(1), 131–137 (2004).
- R. P. Nikhil, P. Kuhu, and M. K. James, "A possibilistic fuzzy c-means clustering algorithm," *IEEE T. Fuzzy Syst.* **13**(4), 517–530 (2005).
- J. Saeedi and K. Faez, "Fisher classifier and fuzzy logic based multi-focus image fusion," in *Proc. of IEEE Int. Conf. on Intelligent Computing and Intelligent Systems*, Shanghai, China, pp. 420–425 (2009).
- G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.* **38**(7), 313–315 (2002).
- C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.* **36**(4), 308–309 (2000).
- Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002).
- X.-Q. Luo and X.-J. Wu, "A new metric of image fusion based on region similarity," *Opt. Eng.* **49**(4), 047006 (2010).
- A. Polo, F. Cattani, and A. Vavassori, "MR and CT image fusion for post implant analysis in permanent prostate seed implants," *Int. J. Radiat. Oncol. Biol. Phys.* **60**(5), 1572–1579 (2004).

Xiao-qing Luo received a PhD degree in pattern recognition and intelligent systems from Jiangnan University, Wuxi, China, in 2010. She currently teaches in the School of Internet of Things as an associate professor of Jiangnan University. Her current research interests are image fusion, pattern recognition, and other problems in image technologies. She has published more than 40 technical articles in these areas.

Zhan-cheng Zhang received a PhD degree from the School of Information Technology, Jiangnan University, in 2011. From 2012 to 2013, he was a postdoctoral with the Chinese Academy of Sciences. Since 2014, he has been an assistant professor with the College of Electronic and Information Engineering, Suzhou University of Science and Technology. He is the author of 30 articles and holds five patents. His research interests include pattern recognition and image fusion.

Xiao-jun Wu received a PhD degree in pattern recognition and intelligent systems from Nanjing University of Science and Technology, Nanjing, China, in 2002. He joined the School of Information Engineering (now renamed as School of IoT Engineering), Jiangnan University, in 2006, where he is a professor. His current research interests are pattern recognition, computer vision, fuzzy systems, neural networks, and intelligent systems. He has published more than 150 papers in his fields of research.