IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 14, 2019, accepted November 6, 2019, date of current version December 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953786

Multi-Modal Sensor Medical Image Fusion Based on Multiple Salient Features With Guided Image Filter

WEISHENG LI[1], LINGHUI JIA[1], AND JIAO DU[2]
[1]Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China

Corresponding authors: Weisheng Li (liws@cqupt.edu.cn) and Jiao Du (dujiao19880429@126.com)

ABSTRACT In this paper, we propose an efficient image fusion algorithm using multiple salient features with guided image filter to prevent the problem of low contrast detail. First, we employ the guided image filter to decompose the input images into a series of smoothed and detailed images at different scales. Second, the salient features are extracted from the decomposed smoothed images and detailed images using two different algorithms: the spectral residual (SR) algorithm for extracting mainframe information and the graph-based visual saliency (GBVS) model for extracting gradient saliency information to construct the fusion rules. In addition, generalized intensity-hue-saturation (GIHS) is adopted to combine the decomposition coefficients. Finally, the fused image is reconstructed by the fused smoothed and fused detailed images. The experimental results demonstrate that the proposed algorithm can achieve better performance than other fusion methods in the domains of MRI-PET and MRI-SPECT fusion.

INDEX TERMS Spectral residual saliency, graph-based visual saliency, MRI-PET fusion, MRI-SPECT fusion.

I. INTRODUCTION

With the development of medical imaging technology, modern medical imaging provides multiple diagnostic images for clinical diagnosis, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) images. They focus on various aspects to provide information. CT and MRI images are anatomical images with a high spatial resolution that provide body contours and soft tissue information. PET and SPECT images are functional images with lower spatial resolution that contain color information that reflects the body's metabolic level. To overcome the limitations of single medical image information expression and to provide more comprehensive and complementary information for medical diagnosis and treatment, multimodal medical image fusion has been proposed. Multimodal medical image fusion obtains a single fused image in terms of human visual perception to increase the clinical applicability of medical images for the diagnosis and assessment of medical problems [1]. Brain diseases have a high incidence and present a high risk to people's lives. To provide additional auxiliary information for such diseases, this paper focuses on the fusion of anatomical images and functional images of brain diseases.

Research on multimodal medical image fusion methods has yielded promising results; however, various shortcomings are identified. The image fusion method [2] that is based on principal component analysis (PCA) transformation is simple and easy; however, the fusion result is prone to color distortion. The fusion result that is obtained via the image fusion method that is based on color space transformation [3] has low contrast. The image fusion methods that are based on the pyramid transform [4], [5] cannot capture the direction information. The image fusion methods that are based on the wavelet transform [6], [7] cannot represent the anisotropic features (such as lines and contours) accurately, and the fusion result has low contrast. The image fusion methods that are based on sparse representation [8]–[11] are time-consuming in the training and optimization of the dictionary. The image fusion methods that are based on neural networks [12], [13] rely on the process of adjusting parameters in the training of the network. In summary, the image fusion methods that are based

The associate editor coordinating the review of this manuscript and approving it for publication was Peng Liu.

on multiscale transformation (MST) have the advantage of extracting more salient features on multiple scales compared with other fusion methods. However, several shortcomings are identified, which can be roughly summarized into three categories: (1) higher time complexity; (2) the appearance of artifacts in the fused image; and (3) low contrast and little detail information. To overcome these problems and to obtain a superior fusion result, we propose an efficient image fusion algorithm that uses multiple salient features with a guided image filter.

Our method involves two key research hotspots: multiscale decomposition and reconstruction of images and construction of fusion rules. In the image decomposition step, considering that many available methods that are based on pyramid transformation and waveform transformation cannot maintain the shift consistency and spend a long time on the selection of parameters, our model uses GIF to decompose the source images into smoothed images and detailed images at various scales with the same resolution as the input images in the spatial domain. This approach can effectively maintain the shift consistency and enhance the edge information of the fused image with lower time complexity. In the fusion rule construction step, since the smoothed subband images that are obtained via the GIF decomposition contain large amounts of smooth information of the source images, we use the spatial residual significance detection algorithm, spectral residual approach (SR), to construct a fusion rule of smoothed images that can distinguish the important information from the redundant information. In contrast, the detailed subband images contain the salient features of the source images. To preserve the significant information (such as the brightness and color) of the source images, we use the graph-based visual saliency (GBVS) detection algorithm to construct the fusion rule of detailed images, which yields fusion results that more closely accord with the human visual system. In addition, to more completely preserve the details of the edges and textures of the images during the fusion process, we use GIHS in the coefficient recombination part of the fusion rule construction. In conclusion, our method uses GIF to decompose the source images into smoothed and detailed images of multiple scales and constructs fusion rules for smoothed and detailed subband images by utilizing SR and GBVS, respectively; finally, the fusion subband images are added to obtain the fused image.

Compared with the available approaches, the main contributions of the proposed method are as follows:

(1) By using the SR algorithm [14] to extract the salient features of the smoothed images with log spectrum, the saliency information of the source images is effectively derived by the combination of spectral residual and phase spectrum information with little distortion. And then, the Inverse Fourier Transform is used to construct a final saliency map in the spatial domain containing only information about the dominant orientations and scales that compose the image [15], [16]. Furthermore, the proposed method realizes superior performance in terms of a nonreference-image quality metric, namely, the natural image quality evaluator (*NIQE*), which is used to evaluate the distortion of the final fused images.

(2) Using methods of salient feature extraction and the GBVS model [17] to extract the salient features of the detailed images can preserve the global vital information, such as the brightness, gradient, and color information of the source images, thereby yielding fusion results that more closely accord with the human visual system [18]. In the next section (Section III), we demonstrate that in the salient features that are extracted from images, including natural and medical images, higher brightness information is perfectly preserved.

(3) Typically, the color information is neglected in the procedures of functional medical imaging, such as PET and SPECT, which is presented in pseudocolor. Color model transformations, namely, IHS and GIHS, are adopted for the fusion of single-channel images in gray and multichannel images in color to preserve the color information that is transferred from inputs to the final fused image [19]–[21]. In contrast to IHS, GIHS could overcome the color distortion problem that was encountered during the substitution operation.

The remainder of this paper is organized as follows: The related work is introduced in Section II. The theory of SR and the GBVS model are introduced in Section III. The algorithm for medical image fusion is proposed in Section IV. Section V presents the experimental evaluation. Finally, the conclusions of this work are presented in Section VI.

II. RELATED WORK

The available methods for multimodal image fusion can be divided into pixel-level, feature-level and decision-level fusion methods [22]. Pixel-level image fusion is the process of directly combining the source image's pixels, which can retain more of the source image's information, and most research is now aimed at pixel-level image fusion. In pixel-level image fusion, methods are roughly classified into the following categories: (1) image fusion methods that are based on MST [23], [24]. Methods of this type are in line with the layering process of the human visual system's cognitive outside world and can effectively extract more useful information on multiple scales. Image fusion methods that are based on MST consist of three steps: first, the source images are decomposed into multiple high-frequency and low-frequency coefficients after multiscale transformation; then, the corresponding fusion rules are used to fuse the low-frequency coefficients and the high-frequency coefficients; the final fused image is obtained by reconstructing the fused high-frequency coefficient and low-frequency coefficient. Commonly used MST decomposition tools include pyramid transformation, wavelet transformation, geometric analysis and edge-preserving filtering algorithms. The image fusion methods that are based on pyramid transformation [4], [5], [25] cannot capture the direction information of the image; the image fusion methods that are based on wavelet transform [6], [7], [26] can only capture limited direction information and cannot maintain

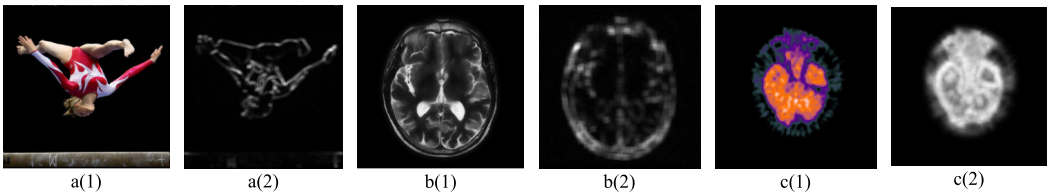

FIGURE 1. SR extraction of the significant features.

spatial translation consistency; the image fusion methods that are based on geometric analysis [27], [28] effectively solve the above problems, but still suffer from high time consumption. Edge-preserving filtering algorithms have been widely used in image fusion in recent years [29], [30], which have low time consumption and can maintain spatial translational consistency in the decomposition process. More importantly, the edge information of the images can be enhanced during the decomposition process; (2) image fusion methods that are based on morphological component analysis [31], [32]. Methods of this type consider the components of various spatial structures that the image contains. An image is decomposed into cartoon and texture components via sparse representation and fusion rules are constructed for cartoon coefficients and texture coefficients, which can express the source images more completely and accurately. However, these methods have a high time cost and it is difficult to optimize the training dictionary process; (3) image fusion methods that are based on variational technology [33], [34]. These methods construct an energy functional by using the relationship between the input images and the fusion result, and they obtain the fusion result by minimizing the functional. The main advantages of methods of this type are scalability, flexibility and the ability to incorporate multiple technologies (such as wavelets and random fields) into the variational model. However, the time complexity of solving the extreme value process is high; hence, the fusion process has poor real-time performance; and (4) image fusion methods that are based on deep learning [11], [35]. The fusion methods that are based on deep learning can learn more image features through the training network, which yields more abundant fusion results, and they have been widely used in multifocus image fusion research. Since no ground truth is available for the fusion of medical images, effectively labeling images and training on them is challenging in deep learning for multimodal medical image fusion.

III. MATERIALS

A. SPECTRAL RESIDUAL APPROACH (SR)

Neurobiologists have studied the ability of humans to understand complex scenes in real time. They found that in the face of complex scenes, the human visual system can quickly and adaptively screen out the objects of interest and centrally process them. The target of human interest is the salient area that we want to study. In recent years, image salient features have been widely used in many fields, such as automatic detection [14], image retrieval [36], object recognition [37], image segmentation [38], and image fusion [29], [39]. SR is a purely mathematical method for detecting saliency, which divides the information that is carried by the image into a variable part and a redundant part. The redundant part represents the frequency information, which cannot be perceived by the human visual system. The variable part represents the significant and important information of the image. The process of calculating the image saliency feature map via SR is as follows.

First, Fourier transformation is performed on the source image I, and the image is converted from the spatial domain to the frequency domain. Then, the amplitude spectrum $A(f)$ and the phase spectrum $P(f)$ of the image are calculated. The expressions are as follows:

$$A(f) = \Re(\mathcal{F}(I)) \tag{1}$$

$$P(f) = \Im(\mathcal{F}(I)) \tag{2}$$

where $\mathcal{F}$ represents the Fourier transform, $\Re$ represents the real part of the complex matrix that is obtained via the Fourier transform, and $\Im$ represents the imaginary part of the complex matrix that is obtained via the Fourier transform.

Then, the residual spectrum $H(f)$ of the image is calculated. First, the logarithm of $A(f)$, which is obtained via formula (1), is calculated to obtain the log spectrum $L(f)$ of the image. Then, it is filtered using an average filter $h_n(f)$ with a kernel size of $n \times n$ and the value of $1/n^2$ to obtain an average spectrum. Finally, the average spectrum is subtracted from $L(f)$ to obtain $R(f)$. The expressions are as follows:

$$L(f) = \log(A(f)) \tag{3}$$

$$R(f) = L(f) - h_n(f) * L(f) \tag{4}$$

Finally, the inverse Fourier transform is applied to $R(f)$ and $P(f)$, and we use a Gaussian filter with a scale of $\delta = 8$ to convolve the results and to generate the final saliency map S. The expression is as follows:

$$S = g(x) * F^{-1}[\exp(R(f)) + P(f)]^2 \tag{5}$$

As shown in Fig. 1, we extracted the saliency map of the image via the SR method. (1) represents the original image, and (2) represents the salient image that is extracted via the SR method. According to the salient image, the SR algorithm can extract the contour information of the image's salient target.

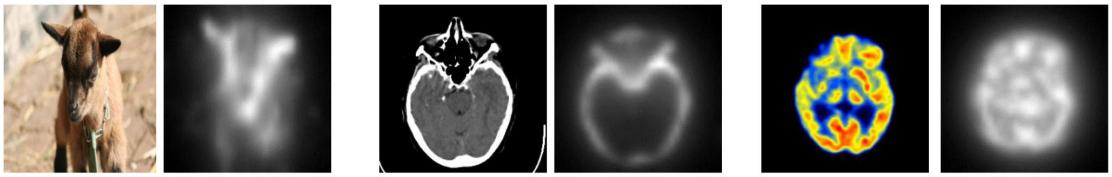

FIGURE 2. GBVS extraction of the significant features.

B. GRAPH-BASED VISUAL SALIENCY (GBVS)

GBVS is a saliency detection method that combines biological visual attention mechanisms and mathematical calculations, and it consists of two steps: forming activation maps on feature channels and normalizing them in a way that highlights their conspicuousness and enables combinations with other maps [17]. The characteristics and extraction methods that are adopted by the GBVS model are the same as those of the ITTI model [40]; however, the characteristic saliency calculation and the feature saliency maps are combined using the Markov chain algorithm.

First, the GBVS model uses Gaussian filtering to obtain the brightness, color and direction feature map M of the source image, which is consistent with the ITTI model. Then, we assume that all the pixels in the feature map M are connected to form a directed graph G_A, namely, each vertex in the graph is connected to n-1 other vertices. Moreover, the weight of the directed edge of any vertex p to q in G_A is defined as follows:

$$\omega_A(p, q) \triangleq d(p||q) \cdot F(p_x - q_x, p_y - q_y) \tag{6}$$

$$F(a, b) \triangleq \exp\{-\frac{a^2 + b^2}{2\sigma^2}\} \tag{7}$$

$$d(p||q) \triangleq |M(p) - M(q)| \tag{8}$$

where $d(p||q)$ represents the dissimilarity of points p and q in the feature map M; (p_x, q_y) and (q_x, q_y) are the spatial coordinates of nodes p and q, respectively; and σ is a free parameter.

To calculate the significance of the feature map M, we define a Markov chain on G_A and normalize the weights on the edges to the range of [0, 1]. Then, we define the points as states and the weights on the edges as transition probabilities. Therefore, the equilibrium state of the Markov chain can reflect the residence time of a random walker at each vertex [41]. The residence time can reflect the significance of a position; hence, the steady state of the Markov chain reflects the feature saliency state.

When using the GBVS model to obtain the feature saliency map, we normalize each feature map by constructing a Markov chain and superimpose all the normalized feature maps. We construct weighted directed graphs on the directed edges of feature saliency maps according to the following:

$$\omega_N(p, q) \triangleq A(p) \cdot F(p_x - q_x, p_y - q_y) \tag{9}$$

where $A(p)$ is the saliency of the vertex p. Then, we can obtain a normalized feature saliency map S by calculating the equilibrium state of the Markov chain.

As shown in Fig. 2, we use the GBVS model to obtain the salient images of several images.

Here, (1) represents the original image and (2) represents the salient image that is extracted via the GBVS method. According to the salient image, the GBVS algorithm can retain the more appealing visual information from the source images and can completely retain the region with higher brightness.

IV. ALGORITHM OF IMAGE FUSION

In Algorithm 1, the proposed method for MRI-PET fusion and MRI-SPECT fusion consists of three main parts: image multiscale decomposition, image fusion rule construction and image reconstruction. A and B are two input images, where A represents the MRI modal image and B represents the PET/SPECT modal image. We consider MRI-SPECT image fusion as an example and present a schematic diagram of the proposed fusion method framework in Fig. 3.

Algorithm 1. Algorithm of MRI-PET/SPECT image fusion

A. MULTISCALE IMAGE DECOMPOSITION

GIF [42] is an effective local linear translation-variant filter model, which has the same edge-preserving smoothing operator as the bilateral filter but performs better near the edges. As the computational complexity of the guided filter is independent of the filter kernel size, it has lower time [43] complexity with faster computing performance. In addition, it is assumed in GIF that there is a local linear model in the guided filter I and filter output O, which enables the filter output to be more structured and smoother than the input image P. GIF has many applications in image processing and computer vision, such as image detail enhancement [43] and image fusion [44].

In this paper, we built a multiscale image representation of the input images with GIF. Input image A or B is decomposed into a series of smoothed images C_i and detailed images $D_i(D_i = I - C_i)$ after L smoothing operations (smoothing radius $r = (2^1, 2^2, ...2^l)$) with GIF. Consistent with most methods that are based on the edge-preserving filter decomposition [11], [24], [26], this paper sets L to 3.

In image processing using guided filtering, the value of ε has a strong influence on the filtering results. If $\varepsilon = 0$, the filter has no effect. If $\varepsilon > 0$, GIF forms a weighted mean

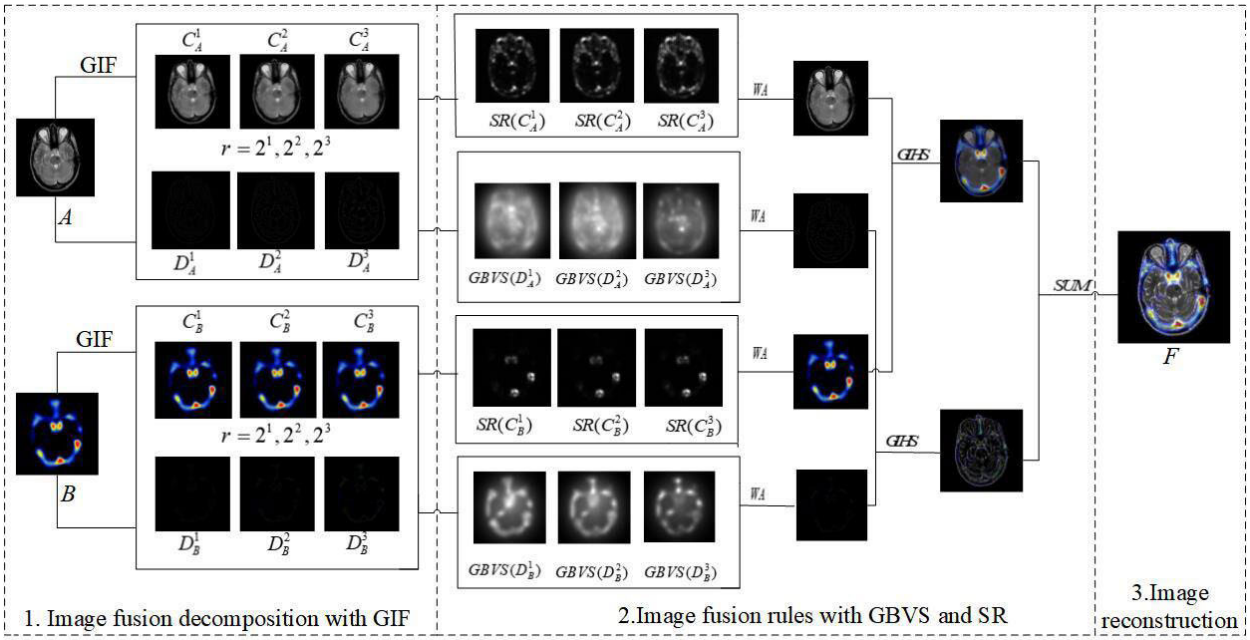

FIGURE 3. Schematic diagram of the proposed fusion framework.

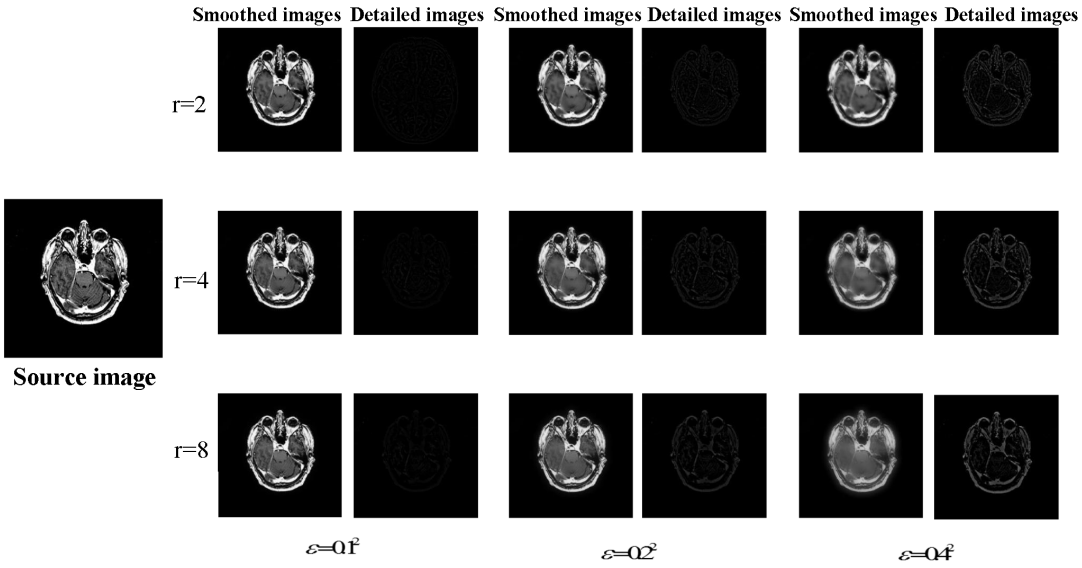

FIGURE 4. GIF filtering results with different parameters.

filter in the region with small changes in the pixel intensity, while in the region with large changes in the pixel intensity, the filtering effect on the image is very weak, which helps maintain the edges. However, if the value of ε is large, many details are filtered out.

As shown in Fig. 4, we experimented with the parameter values that are specified in the GIF reference [42] and when the parameter ε is set to 0.1^2, the smoothed subband images show a strong edge enhancement effect, but the detailed subband images contain less significant features; when ε is set to 0.4^2, the smoothness is too high and smooth information is divided among the detailed subband images. To preserve the detailed textural information while enhancing the edges, in our experiment, we set the radius r of the GIF decomposition schema as $r = 2, r = 4$ and $r = 8$, and set the regularization parameter as $\varepsilon = 0.2^2$ to obtain multiscale smoothed subband images and detailed sub-band images. To further evaluate our claim, this paper conducted 10 sets of MRI-SPECT fusion experiments and 10 sets of MRI-PET fusion experiments. The subjective fusion results are presented in Fig. 5. When $\varepsilon = 0.2^2$, the fusion result is the best, and the edges and the texture information of the fused image are clearer. According to the objective evaluation index values in Table 1, when $\varepsilon = 0.2^2$, the values of the mutual information (*MI*), the spatial frequency (*SF*), *NIQE*, the tone mapped image quality index (*TMQI*) and the edge

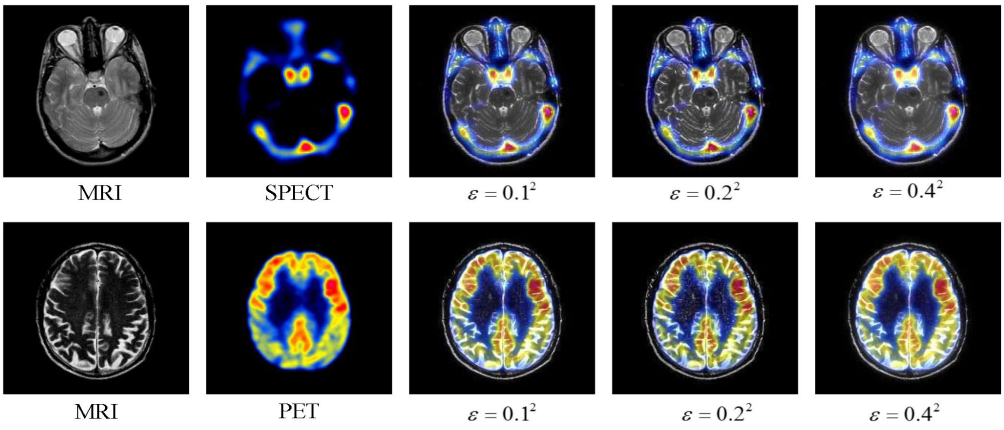

FIGURE 5. The subjective result with different parameters.

TABLE 1. The objective evaluation indicators with different parameters.

Image		MI	SF	NIQE	TMQI	EI
MRI-SPECT	ε=0.1^2	3.8667	0.0695	19.0662	0.7248	66.8518
	ε=0.2^2	**3.9556**	**0.0725**	**17.8306**	**0.7278**	**76.2768**
	ε=0.4^2	3.9340	0.0619	20.1099	0.7251	61.1809
MRI-PET	ε=0.1^2	4.3614	0.0796	19.6014	0.7754	77.8494
	ε=0.2^2	**4.3725**	**0.0820**	**17.6375**	**0.7825**	**83.4609**
	ε=0.4^2	4.3301	0.0681	21.0333	0.7792	67.4774

intensity (*EI*) (details on the objective indices are provided in Section 5) are all optimal; hence, the edge and texture information that is retained by the fusion result is richer. Therefore, we set $\varepsilon = 0.2^2$ in the subsequent experiments.

B. IMAGE FUSION RULES

Fusion rules play a crucial role in the image fusion process. An effective fusion rule is designed to highlight the important features in the original images while restricting the inessential features in the fusion process. In addition, commonly used fusion rules are typically composed of three parts: activity level measurement, coefficient grouping and coefficient combination [45].

(1) Activity level measurement: Since the smoothed images contain large amounts of smooth information of the source images, whose features are not readily identifiable, the SR method can be used to screen out the significant features via mathematical calculations. Hence, the fused image can contain more useful information. The activity levels are calculated as $\widehat{S}_A^i = SR(C_A^i(x,y))$, $\widehat{S}_B^i = SR(C_B^i(x,y))$. In addition, since the detailed images contain significant details of the source images, and features such as the brightness and color are important features of medical images, we use the GBVS algorithm to calculate the activity levels of the detailed images as $\check{S}_A^i = GBVS(D_A^i(x,y))$, $\check{S}_B^i = GBVS(D_B^i(x,y))$.

(2) Coefficient grouping: We group the coefficients by analyzing and comparing the activity levels of the smoothed and detailed images at various scales. The higher the level of coefficient activity, the more important the corresponding images are in the fused images. Furthermore, we use decision maps to record the results of the comparison of the coefficient activity levels. The sum of all decision maps is equal to 1 to ensure that there is no loss of original information. The method of constructing the decision maps is as follows:

$$\widehat{d}_A^i = \begin{cases} 1, & if \ \max\left(\widehat{S}_A^1, \widehat{S}_A^2, \cdots, \widehat{S}_A^n\right) = \widehat{S}_A^i \\ 0, & otherwise \end{cases} \tag{10}$$

$$\check{d}_A^i = \begin{cases} 1, & if \ \max\left(\check{S}_A^1, \check{S}_A^2, \cdots, \check{S}_A^n\right) = \check{S}_A^i \\ 0, & otherwise \end{cases} \tag{11}$$

and

$$\widehat{d}_B^i = \begin{cases} 1, & if \ \max\left(\widehat{S}_B^1, \widehat{S}_B^2, \cdots, \widehat{S}_B^n\right) = \widehat{S}_B^i \\ 0, & otherwise \end{cases} \tag{12}$$

$$\check{d}_B^i = \begin{cases} 1, & if \ \max\left(\check{S}_B^1, \check{S}_B^2, \cdots, \check{S}_B^n\right) = \check{S}_B^i \\ 0, & otherwise \end{cases} \tag{13}$$

(3) Coefficient combination: We use the weighted averaging method to combine the coefficients. With the weighted

Algorithm 1 Basic Steps of the Fusion Scheme

Input: image A, image B

Output: fused image F

Step 1: Multiscale image decomposition

1. Apply the L-level GIF algorithm to decompose A and B into smoothed images that contain smooth rough information and detailed images that contain edge information at levels $\left(C_A^i, D_A^i\right)$ and (C_B^i, D_B^i) with $i = 1, \cdots, L$.

Step 2: Image fusion rules

1. Apply SR to detect the salient features of smoothed subband images C_A^i, C_B^i to obtain $\widehat{S}_A^i, \widehat{S}_B^i$, and compare the significant values of the pixels that correspond to the same spatial position to obtain decision maps $\widehat{d}_A^i, \widehat{d}_B^i$ (Eq. (10) and Eq. (12))
2. Apply GBVS to detect the salient features of detailed subband images D_A^i, D_B^i to obtain $\check{S}_A^i, \check{S}_B^i$, and compare the significant values of the pixels that correspond to the same spatial position to obtain decision maps $\check{d}_A^i, \check{d}_B^i$ (Eq. (11) and Eq. (13)
3. Use the decision maps as weight maps and use the weighted average method to fuse the coefficients of the smoothed layer and the detailed layer in source image A or B at various scales using Eqs. (14) and (15)
4. Use GIHS to fuse the coefficients of the smoothed layers and the detailed layers in A and B: $F^C = GIHS\left(F_A^C, F_B^C\right); F^D = GIHS\left(F_A^D, F_B^D\right)$ Eq. (16)

Step 3: Image reconstruction: Add the fused subband images to obtain the final fused image $F = F^C + F^D$ (Eq. (19)).

average (WA), the fused coefficient is a linear representation of the product of multiple input images and corresponding weights. In this paper, we use the decision maps as weight maps for the combination of the coefficients in smoothed and detailed layers of the source image at various scales to obtain the fused smoothed and fused detailed layers of source images A and B. F_A^C and F_A^D represent the fused smooth layer and the fused detail layer, respectively, of image A. Similarly, F_B^C represents the fused smoothed layer and F_B^D represents the fused detailed layer of image B. Therefore, the coefficient combination calculation method is summarized as follows:

$$F_A^C = \sum_{i=1}^{n} (\hat{d}_A^i * C_A^i); \quad F_A^D = \sum_{i=1}^{n} (\check{d}_A^i * D_A^i) \tag{14}$$

$$F_B^C = \sum_{i=1}^{n} (\hat{d}_B^i * C_B^i); \quad F_B^D = \sum_{i=1}^{n} (\check{d}_B^i * D_B^i) \tag{15}$$

F_A^C and F_A^D are subband images with higher resolution that contain more luminance information, and F_B^C and F_B^D are pseudocolor images with lower resolution that contain more color information. To preserve the texture and edge information of the gray images and the color information of the pseudo color images, we use GIHS [46] to recombine F_A^C and F_B^C to obtain F^C and to recombine F_A^D and F_B^D to obtain F^D. GIHS is a fusion method that is based on color space transformation, the main strategy of which is to replace the intensity components of low-resolution color images with high-resolution gray images, which can preserve the important features of gray images and color images more completely. The combination is expressed as follows:

$$F^C = GIHS\left(F_A^C, F_B^C\right); \quad F^D = GIHS\left(F_A^D, F_B^D\right); \tag{16}$$

The process of applying the GIHS recombination coefficient is as follows:

(1) Use a spatial transformation tool to transform the RGB color space $\left(R', G', B'\right)$ of the pseudo color images into the IHS color space (I, H, S), which is calculated as follows:

$$\begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -\sqrt{2}/6 & -\sqrt{2}/6 & 2\sqrt{2}/6 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \tag{17}$$

$$H = \tan^{-1}\left(\frac{v_2}{v_1}\right), \quad S = \sqrt{v_1^2 + v_2^2} \tag{18}$$

(2) Replace the component I that is obtained via the color space transformation of the pseudo color image using the high-resolution gray image I_{new};

(3) The replaced IHS color space image is inversely transformed into the RGB color space to obtain a fused image F.

C. IMAGE RECONSTRUCTION

Image reconstruction is used to restore the coefficients of the source images into the result image. Common reconstruction methods include the inverse transform, sparse representation and summation. We use the simplest summation method to reconstruct the image in our experiment, and the expression is as follows:

$$F = F^C + F^D \tag{19}$$

V. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETUP

To evaluate the performance of the proposed algorithm, 40 pairs of source images are collected from Whole Brain Web Site of the Harvard Medical School [47], which consisted of 10 pairs of MRI-PET images with mild Alzheimer's disease, 10 pairs of MRI-PET images with glioma disease, 10 pairs of MRI-SPECT images with cavernous hemangioma disease and 10 pairs of MRI-SPECT images with glioma disease. The size of each input image is 256×256, and each set of images is coregistered. The experimental environment is MATLAB 2010 on Windows 7, the processor is an Intel Core i5-4590 CPU, and the RAM size is 8 GB. We use eleven fusion methods for the comparative experiments: GIHS [43], PCA [2], nonsubsampled contourlet transform (NSCT) [48], image fusion with guided filtering (GFF) [41], multilevel local extrema (LES) [29], dense SIFT (DSIFT) [49], parallel saliency features (PSF) [39], LES+PSF [26], Zhu fusion

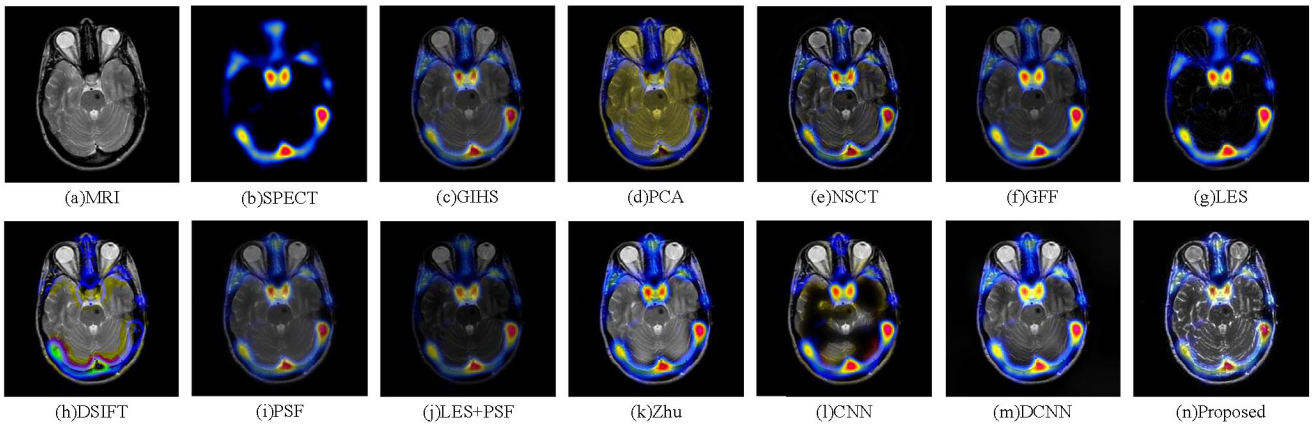

FIGURE 6. Subjective evaluation of MRI-SPECT fusion in cavernous hemangioma disease.

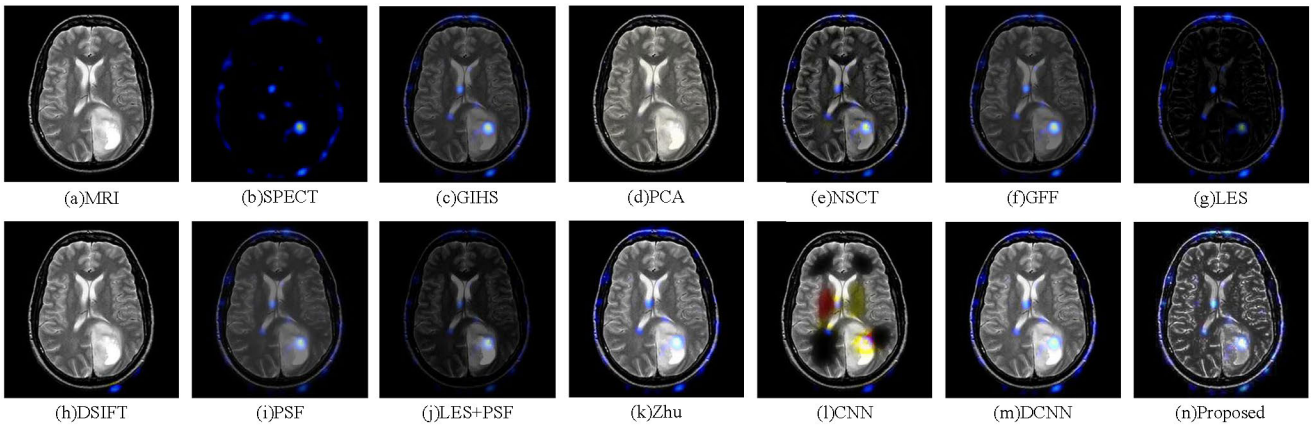

FIGURE 7. Subjective evaluation of glioma disease on MRI-SPECT fusion.

method [50], convolutional neural networks (CNN) [51], and deep CNN (DCNN) [12]. The parameter settings of the compared methods are specified as follows. The NSCT method adopts four decomposition levels with 4, 8, 8, and 16 directions from small to larger scale. The default parameters of guided filtering are used as image fusion rules in GFF: 45 and 0.3 for base layers and 7 and 10^{-6} for detail layers. LES adopts 5×5 sliding widow local image extrema. The decision map for high-pass subbands that is obtained via phase congruency (PC), local sharpness change (LSCM) and local energy (LE) with the weights 1, 2, and 2 is introduced in the Zhu method. Moreover, the parameter settings that are specified by the authors are adopted for the GIHS, PCA, DSIFT, PSF, LES+PSF, CNN, and DCNN methods.

B. OBJECTIVE IMAGE QUALITY METRICS

The objective evaluation method uses mathematical formulas to calculate and obtain numerical values for the fused images [52], which measures the degree of distortion between the fused and input images. We selected eleven indicators for evaluating the preservation of the edges and textures, the level of detail contrast, and the loss of important information.

The spatial frequency (*SF*) [2] is used to measure the overall activity level of the space in the image. The larger the *SF* of the fused image, the richer the information it contains, and the better the fusion method has performed. It is defined as follows:

$$SF = \sqrt{RF^2 + CF^2} \tag{20}$$

where *RF* and *CF* represent the spatial row frequency and the spatial column frequency, respectively.

NIQE [53] evaluates the quality of the fused images when the type of fused image distortion is unknown. The smaller the *NIQE*, the higher the fusion performance. It is defined as follows:

$$NIQE = \sqrt{(\mu_1 - \mu_2)^T \cdot \left(\frac{\sigma_1 + \sigma_2}{2}\right)^{-1} \cdot (\mu_1 - \mu_2)} \tag{21}$$

where μ_1 and μ_2 represent the means of the natural image multivariate Gaussian model and the fused image multivariate Gaussian model, respectively, and σ_1 and σ_2 represent the covariances of the two models.

The tone-mapped image quality Index (*TMQI*) [54] is an index for measuring the degree of loss of the fused image contrast information and luminance information. The larger the value is, the better the fusion result.

$$TMQI\ (I_R, I_F) = aT^{\alpha} + (1 - a)M^{\beta} \tag{22}$$

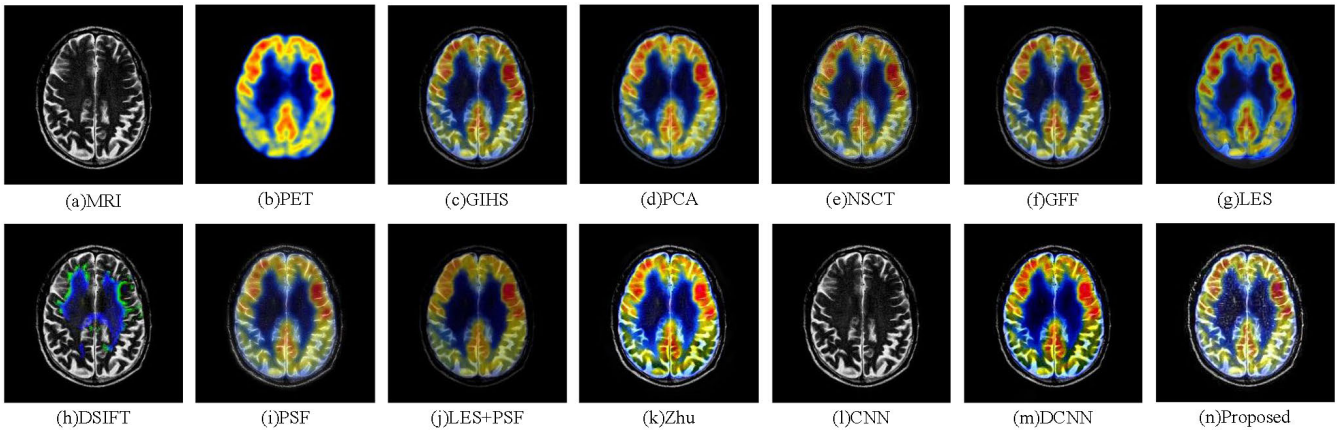

FIGURE 8. Subjective evaluation of mild Alzheimer's disease on MRI-PET fusion.

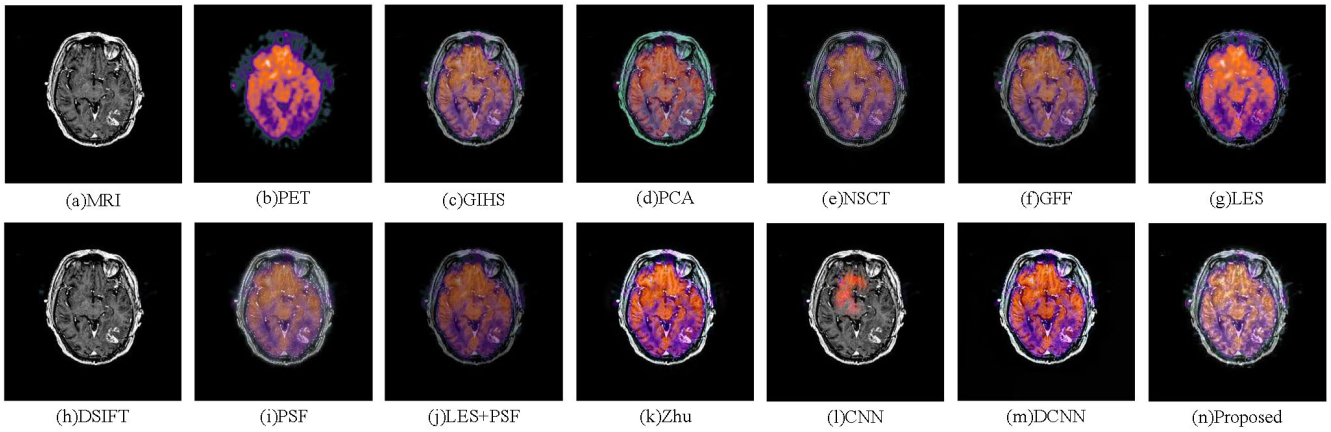

FIGURE 9. Subjective evaluation of glioma disease on MRI-PET fusion.

where T and M represent the structural fidelity and the statistical properties, respectively, of the image and the values of the constants are $a = 0.8012$, $\alpha = 0.3046$, and $\beta = 0.7088$.

The edge intensity (EI) [55] is used to detect the edges of the image and a larger value of EI corresponds to richer edge information in the image. It can be calculated as follows:

$$EI = \frac{1}{M \times N} \times \sum_{j=1}^{N} \sum_{i=1}^{M} \sqrt{G_x(i,j)^2 + G_y(i,j)^2} \quad (23)$$

where G_x and G_y are the results of the Sobel edge detection operator in the$x$- and y-directions, respectively, and M and N denote the numbers of rows and columns.

Mutual information (MI) [56] is used to calculate the information correlation between the fused image and the input source images. The larger the value ofMI, the larger the amount of source image information that is contained in the fused image. The calculation expression is as follows:

$$MI\left(I_R, I_F\right) = \frac{H(I_R) + H(I_F) - H(I_R, I_F)}{H(I_R) + H(I_F)} \quad (24)$$

where $H\,(I)$ and $H\,(I_R, I_F)$ represent a pixel probability distribution histogram of a single image and a joint pixel probability distribution histogram of two images, respectively.

The edge-dependent fusion quality index (Q_e) [57] measures the edge information:

$$Q_e\left(I_R, I_F\right) = w.Q_{wb}\left(I_1, I_F\right) + (1-w) \times Q_{wb}\left(I_2, I_F\right) \quad (25)$$

where Q_{wb} denotes average value between the reference image and the fused image. The larger the value of Q_e, the better the fused image is.

Entropy (EN) [58] measures the information in the fused image:

$$EN(I_F) = -\sum_{x=0}^{255} p_x\left(I_F\right) \ln p_x\left(I_F\right) \quad (26)$$

where p_i is the probability distribution of the pixels with value i over the total number of pixels.

Quality-aware clustering (QAC) [59] is used to predict features via machine learning. The smaller the QAC value is, the less distortion in the fused image. QAC consists of two stages: training and testing. In the stage of training, distortion models, such as Gaussian noise, blur, JPEG, and JPEG 2000 compression, are trained using machine learning methods. In the stage of testing, QAC is calculated as the sum of Z_l:

$$QAC(I_F) = \frac{1}{L} \sum\nolimits_{l=1}^{L} z_l \quad (27)$$

TABLE 2. Objective evaluation indexes of MRI-SPECT fusion in cavernous haemangioma.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AG*	$Q_{ab/f}$	*SSEQ*
GIHS	0.03300	21.25560	0.71270	34.63680	3.49080	0.71500	4.18898	**0.45832**	0.01228	0.40622	37.78393
PCA	0.04960	21.02940	0.68400	48.63110	3.48340	0.77194	4.39102	0.50787	0.01803	0.65042	**35.58704**
NSCT	0.05360	20.62660	0.69060	52.96470	2.83010	0.78046	4.57213	0.48420	0.01914	0.60984	41.00423
GFF	0.03130	21.64550	0.70990	33.75140	3.40040	0.69025	4.22519	0.49029	0.01143	0.33146	38.28709
LES	0.02950	23.46060	0.66870	29.97410	2.84110	0.38977	3.27826	0.49816	0.00981	0.21626	37.89465
DSIFT	0.05630	20.65730	0.67460	57.52520	3.47990	0.78656	4.37164	0.49110	0.02067	0.64924	37.06776
PSF	0.03480	24.75870	0.71390	32.73220	3.17630	0.64563	4.49308	0.49281	0.01229	0.29852	37.32381
LES+PSF	0.01940	22.41790	0.70880	20.20990	3.18100	0.42014	3.68832	0.49847	0.00708	0.11551	38.89017
Zhu	0.05510	20.49010	0.70020	58.80570	**4.66830**	0.82045	4.56524	0.52247	0.02052	**0.65429**	38.15438
CNN	0.06955	21.66154	0.69197	55.19325	3.56740	0.81038	4.32929	0.49727	0.01940	0.60619	41.83655
DCNN	0.07171	20.73718	0.70148	58.75840	4.04775	**0.82151**	**5.05797**	0.52288	0.02073	0.65041	37.23740
Proposed	**0.07900**	**17.83060**	**0.72130**	**78.39860**	3.07410	0.71520	5.01018	0.49119	**0.02972**	0.46664	36.72866

TABLE 3. Objective evaluation indexes of glioma disease on MRI-SPECT fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AG*	$Q_{ab/f}$	*SSEQ*
GIHS	0.04040	22.09090	0.67680	41.19440	3.81910	0.76575	4.64835	0.53420	0.01536	0.45038	37.99927
PCA	0.06800	21.69900	0.66700	65.84020	4.36720	0.88854	4.91814	0.52119	0.02525	**0.76912**	37.57273
NSCT	0.06540	21.31610	0.65840	63.73650	2.89880	0.83489	4.97192	0.55912	0.02395	0.68700	44.70582
GFF	0.03700	22.23320	0.67810	38.27080	3.74110	0.73085	4.61133	0.53514	0.01400	0.35506	38.17851
LES	0.03240	24.55720	0.64320	29.80190	1.97720	0.41861	3.21804	0.54624	0.01090	0.20160	**34.41474**
DSIFT	0.07080	21.55620	0.65850	70.09060	4.24580	0.88489	4.95451	0.52035	0.02645	0.75995	39.16359
PSF	0.04330	25.38450	0.67150	39.67330	3.38250	0.69611	5.01477	0.56625	0.01544	0.32599	41.21503
LES+PSF	0.02330	23.15600	0.65900	24.12180	3.40810	0.43296	4.09049	0.53599	0.00892	0.11435	36.62284
Zhu	0.07220	20.27490	0.68350	75.60160	**6.66380**	**0.89180**	5.04877	0.52179	0.02641	0.75163	39.32204
CNN	0.08885	21.68458	0.64557	70.70249	3.44717	0.88566	4.94487	**0.49325**	0.02631	0.70626	43.84889
DCNN	0.08824	21.40623	0.67124	71.17055	4.32927	0.89058	5.18319	0.52612	0.02649	0.75381	39.94318
Proposed	**0.09070**	**18.20430**	**0.68770**	**93.54610**	3.06030	0.78738	**5.53247**	0.53035	**0.03599**	0.53098	36.52058

$Q_{\mathrm{ab/f}}$ [60] measures the success of edge information transfer from the input images to the fused image. It is computed as follows:

$$Q_{ab/f} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left(Q_{I_1I_F}(x,y)w_{I_1}(x,y) + Q_{I_2I_F}(X,Y)w_{I_2}(x,y)\right)}{\sum_{x=1}^{M}\sum_{y=1}^{N}\left(w_{I_1}(x,y) + w_{I_2}(x,y)\right)} \tag{28}$$

where $Q_{I_1I_F}(x,y), Q_{I_2I_F}(x,y)$ are the preserved values of the edge and orientation information at location (x,y) and $w_{I_1}(x,y), w_{I_1}(x,y)$ reflect the importances of $Q_{I_1I_F}(x,y), Q_{I_2I_F}(x,y)$.

The average gradient (*AG*) [61] represents the details of the image contrast and texture features using horizontal gradient $\Delta_x I_F\,(i,j)$ and vertical gradient $\Delta_y I_F\,(i,j)$.

$$AG(I_F) = \frac{1}{(M-1)(N-1)} \times \sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\sqrt{\frac{(\Delta_x I_F\,(i,j))^2 + \left(\Delta_y I_F\,(i,j)\right)^2}{2}} \tag{29}$$

The spatial-spectral entropy-based quality (*SSEQ*) [61] is used to evaluate the quality of a distorted image across unknown distortion groups. *SSEQ* utilizes local spatial entropies and local spectral entropies to predict the image quality score. The distorted image is predicted using a 12-dimensional feature vector f. *SSEQ* is calculated by applying the function libsvm on the feature vector:

$$SSEQ = libsvm(f) \tag{30}$$

TABLE 4. Objective evaluation indicators of mild Alzheimer's disease on MRI-PET fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AG*	$Q_{ab/f}$	*SSEQ*
GIHS	0.04810	20.55130	0.74230	50.55950	4.29570	0.59352	4.31929	0.47873	0.01700	0.33990	45.41659
PCA	0.03390	20.42440	0.73450	44.81250	4.43300	0.51356	4.35097	0.46335	0.01353	0.20685	**40.00588**
NSCT	0.05660	21.18830	0.74340	49.19790	4.05340	0.52962	4.52910	0.54279	0.01844	0.24290	53.75627
GFF	0.04830	20.54570	0.74700	51.66820	4.27520	0.59984	4.39023	0.47754	0.01731	0.35131	48.60829
LES	0.02960	23.07940	0.72190	39.45570	4.06790	0.34853	4.43281	**0.45148**	0.01159	0.16754	53.30128
DSIFT	0.08520	19.22390	0.73950	82.85060	4.50660	**0.67302**	4.31912	0.47956	0.03245	0.65958	44.32562
PSF	0.05660	23.16190	0.75190	48.91850	4.14650	0.55066	4.68356	0.50837	0.01859	0.31348	51.11735
LES+PSF	0.03200	21.80480	0.73150	33.52380	4.22610	0.45700	3.98611	0.48316	0.01131	0.13858	44.73114
Zhu	0.07540	20.29440	0.75090	76.40630	4.16050	0.65586	4.79418	0.50210	0.02837	0.52519	41.03141
CNN	**0.11569**	19.46272	0.70502	77.80620	4.85475	0.67146	4.24883	0.48863	0.03124	**0.69552**	44.60043
DCNN	0.10163	20.13042	0.74729	74.65140	**5.15196**	0.63979	4.22233	0.48064	0.02577	0.54958	45.47658
Proposed	0.09060	**18.29300**	**0.75700**	**85.93820**	4.04980	0.59685	**4.97464**	0.49559	**0.03444**	0.46181	40.26992

TABLE 5. Objective evaluation indicators of glioma disease on MRI-PET fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AG*	$Q_{ab/f}$	*SSEQ*
GIHS	0.05770	21.97680	0.76110	37.55410	3.46570	0.66133	3.10135	0.42709	0.01526	0.46082	48.99267
PCA	0.06330	21.95270	0.75840	38.68920	3.46440	0.68222	3.18720	0.42268	0.01574	0.46158	48.38805
NSCT	0.06220	21.87220	0.75820	36.07380	3.30180	0.58687	3.29564	0.42563	0.01605	0.26142	50.38346
GFF	0.05450	21.90960	0.75860	36.41200	3.23480	0.59648	3.47397	0.43245	0.01420	0.31935	48.73516
LES	0.05560	23.83930	0.75030	36.28470	3.48580	0.38069	3.17386	**0.41399**	0.00742	0.11934	49.57926
DSIFT	0.08590	22.88240	0.75560	59.69510	3.66890	**0.75077**	2.97609	0.41582	**0.02534**	0.68422	50.22065
PSF	0.06320	25.22740	0.75750	34.66230	3.34360	0.58432	3.38394	0.43229	0.01478	0.27831	48.52984
LES+PSF	0.03610	23.71790	0.75900	24.48800	3.38260	0.52058	2.97260	0.42492	0.00971	0.12869	49.55252
Zhu	0.08550	22.82710	0.75470	58.21990	3.57230	0.59541	3.36644	0.42907	0.02351	0.63654	49.04351
CNN	**0.12618**	22.16384	0.74700	62.32321	4.08760	0.74998	2.99769	0.41550	0.02481	**0.69382**	50.20328
DCNN	0.12279	21.74402	0.74492	**63.02021**	**4.20535**	0.74479	**3.87881**	0.44498	0.02415	0.66094	49.25978
Proposed	0.08690	**18.95130**	**0.76750**	62.25210	3.41060	0.61221	3.66707	0.43375	0.02387	0.44682	**46.93417**

C. EXPERIMENTAL RESULTS ON THE DATABASE

1) EXPERIMENTAL RESULTS ON MRI-SPECT FUSION

Fig. 6 and Fig. 7 present the subjective results of MRI-SPECT image fusion for patients with cavernous hemangioma and glioma, respectively. According to Fig. 6 and Fig. 7, the results of our method show strong detail contrast enhancement effects, and the details of the texturesand edges are clearer and easier for human eyes to observ, shown in Fig. 6(n) and 7(n). Although the results of the PCA method and the DSIFT method satisfactorily preserve the textural details, they exhibit severe color distortion. Although the results of the LES method retain the colors of the SPECT, a large amount of MRI information is lost. The results of the LES+PSF method have lost brightness information. The results of the PSF methodhave lower detail contrast. Compared with other methods, the GIHS method, the NSCT method and the GFF method yield superior fusion results; however, the sharpness is not sufficient. The results of the Zhu and DCNN methods, which are shown in Fig. 7(k) and (m), exhibit superior colo retention, and the textures and edges are clearer. However, due to the increase in brightness, the results of our method have lower colo contrast in Fig. 7(n), compared to the results of. Zhu's method in Fig. 7(k). And, the superiority of the proposed method is with higher contrast textural information, compared to the other fusion methods. The fused result that is obtained via the CNN method, which is one of the deep learning methods, introduces color distortion, as shown in Fig. 7(l).

Table 2 and Table 3 show the objective results of MRI-SPECT image fusion for patients with cavernous hemangioma and glioma, respectively, which are based on eleven objective image quality metrics. The value of each index is the average of 10 sets of experimental results, and according to Table 2, our method has the best results in terms of *SF*, *NIQE*, *TMQI*, *EI,* and *AG*; the best results for each index are marked in bold. Thus, the method preserves

TABLE 6. Objective scoring of cavernous haemangioma disease on MRI-SPECT fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AVG*	$Q_{ab/f}$	*SSEQ*	SUM_1
GIHS	4	6	10	5	9	5	3	**12**	6	5	7	72
PCA	6	7	3	6	8	7	7	3	5	11	**12**	75
NSCT	7	10	4	7	1	8	10	11	7	8	2	75
GFF	3	5	9	4	6	4	4	10	3	4	4	56
LES	2	2	1	2	2	1	1	4	2	2	6	25
DSIFT	9	9	2	9	7	9	6	8	10	9	10	88
PSF	5	1	11	3	4	3	8	7	4	3	8	57
LES+PSF	1	3	8	1	5	2	2	5	1	1	3	32
Zhu	8	11	6	11	**12**	11	9	2	9	**12**	5	96
CNN	10	4	5	8	10	10	5	6	8	7	1	74
DCNN	11	8	7	10	11	**12**	**12**	1	11	10	9	102
Proposed	**12**	**12**	**12**	**12**	3	6	11	9	**12**	6	11	**106**

TABLE 7. Objective scoring of glioma disease on MRI-SPECT fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AVG*	$Q_{ab/f}$	*SSEQ*	SUM_2
GIHS	4	5	9	5	8	5	4	6	4	5	8	63
PCA	7	6	6	7	11	10	5	10	7	**12**	9	90
NSCT	6	10	3	6	2	7	8	2	6	7	1	58
GFF	3	4	10	3	7	4	3	5	3	4	7	53
LES	2	2	1	2	1	1	1	3	2	2	**12**	29
DSIFT	8	8	4	8	9	8	7	11	10	11	6	90
PSF	5	1	8	4	4	3	9	1	5	3	3	46
LES+PSF	1	3	5	1	5	2	2	4	1	1	10	35
Zhu	9	11	11	11	**12**	**12**	10	9	9	9	5	**108**
CNN	11	7	2	9	6	9	6	**12**	8	8	1	79
DCNN	10	9	7	10	10	11	11	8	11	10	4	101
Proposed	**12**	**12**	**12**	**12**	3	6	**12**	7	**12**	6	11	105

more detail information without the reference input image. The results of the Zhu method are the best in terms of *MI* and $Q_{ab/f}$; hence, this method has the highest correlation with the two source images. However, it is outperformed by our method in retaining edge and texture information. DCNN, which is a deep learning method, realizes the best performance in terms of the Q_e and *EN* metrics. According to Table 3, the proposed method realizes the best performance in terms of metrics *SF*, *NIQE*, *TMQI*, *EI*, *EN* and *AG*. Thus, the final fused image contains high brightness information and high contrast gradient information. The fusion results that were obtained via the Zhu method are the best in terms of metrics *MI* and Q_e, namely, the Zhu method preserves the original information from the input images.

2) EXPERIMENTAL RESULTS ON MRI-PET FUSION

Fig. 8 and Fig. 9 present the subjective results of MRI-PET image fusion for patients with mild Alzheimer's disease and glioma, respectively. In Fig. 8 and Fig. 9, the results of the proposed method have large advantages in the preservation of the detail information such as edges and textures, and the detail contrast enhancement effect is strong, which is convenient for human visual observations; however, the colo contrast is lower. The PCA method loses edge information. Although the LES method completely retains the color information of PET images, the texture and edge information of MRI images is largely lost. Although the DSIFT method preferably preserves details such as the textures and edges of MRI images, the color information of PET images is lost. The PSF method has lower color contrast and detail contrast due to the excessive brightness information. The LES+PSF method loses color and brightness information. Although the GIHS, NSCT and GFF methods can completely preserve the color and detail information of the source images, the detail contrast is not high, and the result is not sufficiently clear. The CNN method could not preserve color information from the

TABLE 8. Objective scoring of mild Alzheimer's disease on MRI-PET fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AVG*	$Q_{ab/f}$	*SSEQ*	SUM_3
GIHS	4	5	6	6	8	6	6	9	4	6	6	66
PCA	3	7	4	3	9	3	5	11	3	3	**12**	63
NSCT	6	4	7	5	2	4	9	1	6	4	1	49
GFF	5	6	8	7	7	8	7	10	5	7	4	74
LES	1	2	2	2	3	1	8	**12**	2	2	2	37
DSIFT	9	11	5	11	10	**12**	4	8	11	11	9	**101**
PSF	7	1	11	4	4	5	10	2	7	5	3	59
LES+PSF	2	3	3	1	6	2	1	6	1	1	7	33
Zhu	8	8	10	9	5	10	11	3	9	9	10	92
CNN	**12**	10	1	10	11	11	3	5	10	**12**	8	93
DCNN	11	9	9	8	**12**	9	2	7	8	10	5	90
Proposed	10	**12**	**12**	**12**	1	7	**12**	4	**12**	8	11	**101**

TABLE 9. Objective scoring of glioma disease on MRI-PET fusion.

	SF	*NIQE*	*TMQI*	*EI*	*MI*	Q_e	*EN*	*QAC*	*AVG*	$Q_{ab/f}$	*SSEQ*	SUM_4
GIHS	4	7	11	6	7	8	4	6	5	7	8	73
PCA	7	8	8	7	6	9	6	9	6	8	11	85
NSCT	5	10	7	3	2	4	7	7	7	3	1	56
GFF	2	9	9	5	1	6	10	3	3	5	9	62
LES	3	2	3	4	8	1	5	**12**	1	1	4	44
DSIFT	9	4	5	9	10	**12**	2	10	**12**	11	2	86
PSF	6	1	6	2	3	3	9	4	4	4	10	52
LES+PSF	1	3	10	1	4	2	1	8	2	2	5	39
Zhu	8	5	4	8	9	5	8	5	8	9	7	76
CNN	**12**	6	2	11	11	11	3	11	11	**12**	3	93
DCNN	11	11	1	**12**	**12**	10	**12**	1	10	10	6	**96**
Proposed	10	**12**	**12**	10	5	7	11	2	9	6	**12**	**96**

input PET image in pseudo color, as shown in Fig. 8(l). Meanwhile, the CNN method introduces color distortion, as shown in Fig. 9(l). The results of the Zhu method are similar to the source images, and the texture and edge information can be preserved more completely; however, the contrast is lower compared to our method.

Table 4 and Table 5 present the objective results of MRI-PET image fusion for patients with mild Alzheimer's disease and glioma, respectively. Each value in the table represents the average of 10 sets of fusion results, and the best results are shown in bold. According to Table 4, our method has the highest values of indicators *NIQE*, *EI*, *TMQI*, *EN*, and *AG*; hence, the fusion result of our method has the highest fidelity and contains more information and higher detail contrast. The DCNN method has the highest value of the *MI* index. Additionally, according to Table 5, the proposed method obtains lowest value of the *SSEQ* metric, namely, the proposed method introduces little distortion of the type that is defined in *SSEQ*. CNN and DCNN outperform the other fusion methods on the eleven selected metrics.

3) DISCUSSION

According to the subjective evaluation results and the objective evaluation results of MRI-SPECT fusion and MRI-PET fusion, our method yields images with the highest contrast, which makes the edges and textures clearer and the details richer; however, the color contrast is lower. Although the DSIFT method retains the structural information more effectively, substantial color information is lost. The colors of the LES method are consistent with those of the source image; however, substantial structural information is lost. The CNN and DCNN fusion methods introduce color distortion. The Zhu method preserves the colors more effectively and retains the structural information more completely; however, the contrast is lower and the texture and edge retention is lower compared to the proposed method. The remaining

TABLE 10. Objective scoring of multi-model sensor medical image fusion ($SUM=SUM_1+SUM_2+SUM_3+SUM_4$).

	GIHS	PCA	NSCT	GFF	LES	DSIFT	PSF	LES+PSF	Zhu	CNN	DCNN	Proposed
SUM	274	313	238	245	135	365	214	139	372	339	389	**408**

methods are substantially outperformed by our method in terms of both the objective and subjective results. In summary, our method yields satisfactory results in MEI-SPECT fusion and MRI-PET fusion.

Compared to the input images, one output fused image that was obtained via the proposed method is enhanced with high contrast and high brightness. The enhanced fused image destroys the original pixel distribution. Therefore, the large difference between one input and one output fused image results in lower similarity in terms of the pixel probability, namely, the proposed method yields a lower *MI* value. Tables 6-9 present the scoring results [30] in accordance with the metrics values in Tables 2-5. Importantly, the last line in Tables 6-9 corresponds to the total values of the eleven metrics. According to Tables 6, 8, and 9, the proposed method obtains the highest total values. Besides, Table 7 shows that fusion method proposed by Zhu obtains the best prize in terms of glioma disease on MRI-SPECT fusion. Furthermore, according to Table 10, the proposed method realizes the highest performance in the objective evaluations, which constitute an easy approach for evaluating fusion methods.

VI. CONCLUSION

This paper proposes an effective multimodal medical image fusion method that is based on SR and GBVS saliency, which can enhance the detail contrast of fused images while retaining the important information of the source images. Using the SR algorithm to extract the salient features of smoothed images is beneficial for distinguishing the important information from the redundant information, which causes the fused image to retain more useful information. Employing the GBVS algorithm to extract the salient features of the detailed images has the advantage of selecting the detail information from the source images, which renders the color and brightness information of the fused image more abundant. In addition, by using the GIHS algorithm for the combination of decomposition coefficients, more texture and edge information of the high-resolution images can be preserved. The experimental results demonstrate that the method can be effectively applied to MRI-PET fusion and MRI-SPECT fusion since the fusion results have more complete structures and color characteristics while the edges and textures are clearer. However, to date, this method has only been applied to brain medical images, and whether it is suitable for medical images of other parts can continue to be explored in future research.

REFERENCES

[1] J. Du, W. Li, K. Lu, and B. Xiao, "An overview of multi-modal medical image fusion," *Neurocomputing*, vol. 215, pp. 3–20, Nov. 2016.
[2] C. He, Q. Liu, H. Li, and H. Wang, "Multimodal medical image fusion based on IHS and PCA," *Proc. Eng.*, vol. 7, pp. 280–285, Jan. 2010.
[3] S. Daneshvar and H. Ghassemian, "MRI and PET image fusion by combining IHS and retina-inspired models," *Inf. Fusion*, vol. 11, no. 2, pp. 114–123, 2010.
[4] M. J. Li, Y. B. Dong, and X. L. Wang, "Image fusion algorithm based on gradient pyramid and its performance evaluation," *Appl. Mech. Mater.*, vol. 525, pp. 715–718, Feb. 2014.
[5] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
[6] V. Bhateja, H. Patel, A. Krishn, A. Sahu, and A. Lay-Ekuakille, "Multimodal medical image sensor fusion framework using cascade of wavelet and contourlet transform domains," *IEEE Sensors J.*, vol. 15, no. 12, pp. 6783–6790, Dec. 2015.
[7] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, 2004.
[8] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*, vol. 13, no. 1, pp. 10–19, 2012.
[9] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.
[10] H. Cheng, Z. Liu, L. Yang, and X. Chen, "Sparse representation and learning in visual recognition: Theory and applications," *Signal Process.*, vol. 93, no. 6, pp. 1408–1425, 2013.
[11] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.
[12] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. 20th Int. Conf. Inf. Fusion*, Jul. 2017, pp. 1–7.
[13] S. Monica and S. Sahoo, "Pulse coupled neural networks and its applications," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3965–3974, 2014.
[14] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[15] N. Imamoglu, W. Lin, and Y. Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Trans. Multimedia*, vol. 15, no. 1, pp. 96–105, Jan. 2013.
[16] M. Song, C. Chen, S. Wang, and Y. Yang, "Low-level and high-level prior learning for visual saliency estimation," *Inf. Sci.*, vol. 281, pp. 573–585, Oct. 2014.
[17] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 545–552.
[18] L. Zhang, A. Li, Z. Zhang, and K. Yang, "Global and local saliency analysis for the extraction of residential areas in high-spatial-resolution remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3750–3763, Jul. 2016.
[19] C.-L. Chien and W.-H. Tsai, "Image fusion with no gamut problem by improved nonlinear IHS transforms for remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 651–663, Jan. 2014.
[20] S. Chen, R. Zhang, H. Su, J. Tian, and J. Xia, "SAR and Multispectral image fusion using generalized IHS transform based on à Trous wavelet and EMD decompositions," *IEEE Sensors J.*, vol. 10, no. 3, pp. 737–745, Mar. 2010.
[21] G. Bhatnagar, Q. M. J. Wu, and Z. Liu, "Human visual system inspired multi-modal medical image fusion framework," *Expert Syst. Appl.*, vol. 40, pp. 1708–1720, Apr. 2013.
[22] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jun. 2017.
[23] P. Gomathi and B. Kalaavathi, "Medical image fusion based on multiscale transforms," *J. Med. Imag. Health Inform.*, vol. 7, no. 2, pp. 478–484, 2017.
[24] Z.-G. Wang, W. Wang, and B. Su, "Multi-sensor image fusion algorithm based on multiresolution analysis," *Int. J. Online Eng.*, vol. 14, no. 6, pp. 44–57, 2018.

[25] M. Li and Y. Dong, "Image fusion algorithm based on contrast pyramid and application," in *Proc. Int. Conf. Mechatronic Sci. Electr. Eng. Comput. (MEC)*, Dec. 2013, pp. 1342–1345.
[26] J. Wu, X. Ding, Y. Chen, Z. Wang, and Y. Su, "Image fusion algorithm based on wavelet transform and fuzzy reasoning," *Proc. SPIE*, vol. 6787, Nov. 2007, Art. no. 67870O.
[27] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.
[28] P. Ganasala and V. Kumar, "Multimodality medical image fusion based on new features in NSST domain," *Biomed. Eng. Lett.*, vol. 4, no. 4, pp. 414–424, Dec. 2014.
[29] Z. Xu, "Medical image fusion using multi-level local extrema," *Inf. Fusion*, vol. 19, no. 11, pp. 38–48, 2013.
[30] J. Du, W. Li, X. Bin, and N. Qamar, "Medical image fusion by combining parallel features on multi-scale local extrema scheme," *Knowl.-Based Syst.*, vol. 113, pp. 4–12, Dec. 2016.
[31] Y. Jiang and M. Wang, "Image fusion with morphological component analysis," *Inf. Fusion*, vol. 18, pp. 107–118, Jul. 2014.
[32] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, Jul. 2018.
[33] W. Zhao and H. Lu, "Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2283–2294, Sep. 2017.
[34] W.-W. Wang, P.-L. Shui, and X.-C. Feng, "Variational models for fusion and denoising of multifocus images," *IEEE Signal Process. Lett.*, vol. 15, pp. 65–68, Jan. 2008.
[35] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.
[36] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2Photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, no. 5, p. 124, 2009.
[37] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun./Jul. 2004, p. II.
[38] W. Wang, J. Shen, and F. Porikli, "Saliency-aware geodesic video object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3395–3402.
[39] J. Du, W. Li, and B. Xiao, "Fusion of anatomical and functional images using parallel saliency features," *Inf. Sci.*, vol. 430, pp. 567–576, Mar. 2018.
[40] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[41] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Springer, 2012.
[42] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[43] N. Liu and D. Zhao, "Detail enhancement for high-dynamic-range infrared images based on guided image filter," *Infr. Phys. Technol.*, vol. 67, pp. 138–147, Nov. 2014.
[44] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[45] R. Shen, I. Cheng, and A. Basu, "Cross-scale coefficient selection for volumetric medical image fusion," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1069–1079, Apr. 2013.
[46] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.
[47] K. A. Johnson and J. A. Becker, *The Whole Brain Altas*. [Online]. Available: http://www.med.harvard.edu/aanlib/
[48] W. W. Kong, Y. J. Lei, Y. Lei, and S. Lu, "Image fusion technique based on non-subsampled contourlet transform and adaptive unit-fast-linking pulse-coupled neural network," *IET Image Process.*, vol. 5, no. 2, pp. 113–121, 2011.
[49] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Inf. Fusion*, vol. 23, pp. 139–155, May 2015.
[50] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019.
[51] Y. Liu, X. Chen, H. Peng, and Z. F. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36 pp. 191–207, Jul. 2017.
[52] Y. Yang, X. Wang, T. Guan, J. Shen, and L. Yu, "A multi-dimensional image quality prediction model for user-generated images in social networks," *Inf. Sci.*, vol. 281, pp. 601–610, Oct. 2014.
[53] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'Completely Blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
[54] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
[55] S. Tang, C. Shen, and G. Zhang, "Adaptive regularized scheme for remote sensing image fusion," *Frontiers Earth Sci.*, vol. 10, no. 2, pp. 236–244, 2016.
[56] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'Information measure for performance of image fusion'," *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.
[57] S. Zheng, W.-Z. Shi, J. Liu, G.-X. Zhu, and J.-W. Tian, "Multisource image fusion method using support value transform," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1831–1839, Jul. 2007.
[58] Q.-G. Miao, C. Shi, P.-F. Xu, M. Yang, and Y. B. Shi, "A novel algorithm of image fusion using shearlets," *Opt. Commun.*, vol. 284, no. 6, pp. 1540–1547, Mar. 2011.
[59] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 995–1002.
[60] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. Int. Conf. Image Process.*, Sep. 2003, pp. III–173.
[61] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.