Information Sciences 430–431 (2018) 567–576

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Fusion of anatomical and functional images using parallel saliency features

Jiao Du[a], Weisheng Li[b,*], Bin Xiao[b]

[a] Chongqing Engineering Laboratory for Detection, Control and Integrated System, Chongqing Technology and Business University, Chongqing 400067, China

[b] Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

ARTICLE INFO

Article history:
Received 21 August 2016
Revised 29 November 2017
Accepted 4 December 2017
Available online 6 December 2017

Keywords:
Parallel saliency features
High-spatial-resolution edge
High-intensity color
Fusion of MRI-CBV and SPECT-Tc images
Fusion of MRI-T1 and PET-FDG images

ABSTRACT

An efficient method is proposed for fusion of anatomical and functional images by constructing the fused image through the combination of parallel saliency features in a multi-scale domain. First, the anatomical and functional images are decomposed into a series of smooth layers and detail layers at different scales by the average filter. Second, the parallel saliency features of both sharp edge and color detail are extracted to obtain the saliency maps. The edge saliency weighted map aims to preserve the high-spatial-resolution structural information using the Canny edge detection operator, while the color saliency weighted map extracts the high-intensity color detail using the context-aware operator. Finally, the fused image is reconstructed by the fused smooth layers and the fused detail layers using saliency maps. We demonstrate the application of the proposed method to a medical problem: Alzheimer's disease. Experimental results show that the proposed method for fusion of MRI-CBV and SPECT-Tc images and fusion of MRI-T1 and PET-FDG images successfully presents the pleasing fused medical images with high-spatial-resolution anatomical structural boundaries and high-intensity color detail.

1. Introduction

In the past several decades, a broad range of medical imaging techniques have become available in medical institutions. Tomographic techniques, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and Single Photon Emission Computed Tomography (SPECT), play an important role in determining the extent of human diseases, evaluating the effectiveness of treatments and guiding surgical procedures. Roughly, tomographic techniques are divided into two categories: anatomical imaging data and functional imaging data. Both CT and MRI images provide high-spatial-resolution anatomical structural information. In contrast to anatomical images, both PET and SPECT images display the high-intensity blood flow to tissues and organs [27]. The fusion of anatomical and functional images, such as fusion of CT and PET images, fusion of CT and PET images, fusion of MRI and PET images, and fusion of MRI and SPECT images [12,31,40], could provide additional advantages, including high-spatial-resolution of soft tissue, lower radiation dose and shorter examination time. For example, fusion of MRI and PET images is helpful in detecting hepatic metastasis [8] and fusion of MRI and SPECT images is useful in locating lesions in patients with tinnitus [9].

* Corresponding author.
E-mail address: liws@cqupt.edu.cn (W. Li).

https://doi.org/10.1016/j.ins.2017.12.008
0020-0255/

Medical image fusion is the process of merging multiple images from a single or multiple imaging modalities. The purpose of image fusion is to obtain a single final image that preserves specific features to increase the clinical applicability of medical images for diagnosis and assessment of medical problems [12]. Two directions can be explored in terms of the available methods for medical image fusion, to improve the quality of the fused images: advanced image decomposition schemes and effective image fusion rules. In this study, we focus on the latter. Image fusion rules refer to algorithms that seek to highlight the features of interest in images and suppress the insignificant features. The main aim of image fusion rules is the combination of multiple images into a single image. The effectiveness of the image fusion rules is closely related to the quality of the fused image. Prospective effective image fusion rules are generally classified into four categories: substitution, pulse coupled neural network (PCNN), sparse representation (SR), and local feature descriptor methods.

Substitution methods are easily to perform, such as principal component analysis (PCA) [26,32,45] and intensity-hue-saturation (IHS) [6,7,37]. A PCA-based fusion rule is related to a data-driven technique as well as higher-order statistics to reveal hidden saliency structures. PCA depends on the linear combination of vectors forming new irrelevant principle components. The advantage of PCA methods is their ability to preserve spatial information, e.g., adaptive PCA, spectral PCA, spatial PCA, and independent component analysis (ICA). The IHS fusion method, which is a popular technique in the remote sensing community, has been used in fusing panchromatic (Pan) and multispectral (MS) images. The intensity of an MS image is replaced by a Pan image with higher resolution. Based on the framework for fusion of high-spatial-resolution Pan and high-spectral-resolution MS images, IHS is introduced into the fusion of MRI and PET images. Both IHS and PCA can maintain the same spatial resolution as the anatomical image. However, the methods that are based on PCA introduce artificial effects, such as block effects, ringing effects and pseudo Gibbs effects, and the methods that are based on IHS cause the loss of an important saliency feature from one of the source images.

PCNN methods [22,35,41,47] are developed from the biological experimental observations of synchronous pulses in the visual cortex of the cat. Each neural cell is composed of an acceptance domain, a modulation domain and a pulse generator. The intensity of each pixel in the image corresponds to a neural cell in the PCNN framework. The main advantage of PCNN methods is that the information is processed in accordance with the human visual system. However, PCNN methods require a training stage for achieving the optimal image pixel values for fusing images [35]. For medical image fusion, methods that are based on PCNN apply a feedback neural network to fuse the input images that were obtained with different modalities. PCNN-based image fusion rules can be used in the wavelet domain, such as Discrete Wavelet Transform (DWT)-PCNN or the spatial domain, such as Internal Generative Mechanism (IGM)-PCNN and *m*-PCNN [22,41,47].

SR, which is an effective way to describe medical image signals by a linear combination of atoms using an over-complete dictionary [14,18,24], is widely used for fusion of medical images. The sparse coefficients are treated as the saliency features of the input medical images. The over-complete dictionary affects the performance of the SR-based fusion methods, whose aim is better representation of the input image signal. Unfortunately, it is argued that SR-based fusion methods are good at isolating common and innovative features, but are unsatisfactory for processing edge and texture information.

The local feature descriptor [5,20,43], such as smallest univalve segment assimilating nucleus (SUSAN), dense scale-invariant feature transform (DSIFT), and local extrema scheme (LES), are widely used in the field of medical image fusion. The SUSAN feature is proposed for fusing high-frequency bands that are obtained from the framelet transform. SUSAN is used as a tool for fusing multi-modal medical images, which is a feature descriptor of the image that is inspired by the human visual system. In the SUSAN feature [5], each image pixel corresponds to its local area of similar intensity. DSIFT [20] is adopted to measure the activity level of the original images in the feature space transform. The DSIFT local feature is constructed based on the histogram of gradient orientations in a cell. The experimental results show that the fusion method that uses DSIFT can not only measure the activity level but also match the mis-registered pixels between different original images. The mean of the local minima and maxima envelopes in LES [43] is used to measure the edge detail information of grayscale medical images. The main limitation of the local feature descriptor is that the detection is restricted to image features within local regions.

Although many effective image fusion rules have been proposed in fusion methods, there is still large room for improvement. In the existing fusion methods, a single feature is used to extract the same information from medical imaging data that were obtained with different modalities. In this paper, we propose a new image fusion rule that uses parallel saliency features and considers extracting specific features from different medical images [44]. Parallel saliency features can be divided into two types: edge and color saliency features. Edge saliency feature (ESF) [4] obtained by the Canny edge detection descriptor is useful for preserving high-spatial-resolution soft tissue structural information of the human brain from anatomical images. Color saliency feature (CSF) [10] is used to obtain high-intensity color information from functional images by the context-aware color saliency descriptor. The motivation of the proposed image fusion rules is to extract not only the sharp edge information but also color detail information. Experimental results show the excellent high-intensity and high-spatial-resolution fused image using the proposed method in comparison with state-of-the-art multi-modal medical image fusion approaches.

The rest of this paper is organized as follows. In Section 2, the related work on saliency features is introduced. In Section 3, one method is presented for fusion of anatomical and functional images. Then, the experimental evaluation is presented in Section 4. Finally, Section 5 presents the study's conclusions.

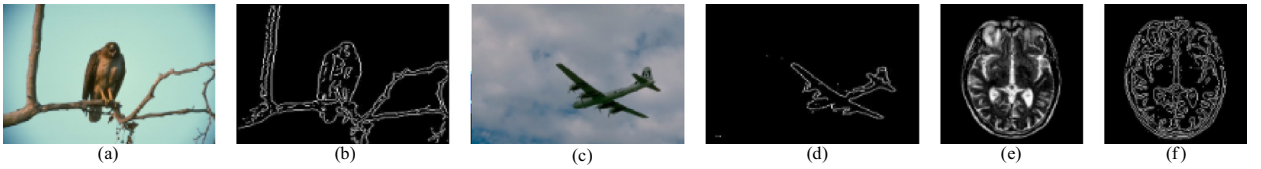

Fig. 1. Three examples of edge saliency maps. (a), (c), and (e) are three input images. (b), (d), and (f) are the edge saliency maps of input images (a), (c), and (e), respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Parallel saliency features

A saliency features is defined as a visual characteristic that distinguishes the salient object from its surroundings [1,3,21]. In this paper, parallel saliency features, including ESF [4] and CSF [10], are exploited to determine the salient of each coefficient at different scales in a medical imaging data. The brighter coefficient should be assigned higher saliency values. The high-spatial-resolution structural information of the input anatomical image is characterized by continuous sharp edge using ESF. Additionally, CSF is used to highlight the high-intensity color detail of the input functional image.

2.1. Edge saliency feature

Edge information describes important components of an image. The Canny edge map is detected as the local maxima in a two-scale multiplication [4]. The scale multiplication functions, namely, $P_I^x(x,y)$ and $P_I^y(x,y)$, are used as the responses of the edge detection filter in the x- and y-directions at a large scale s_2 and a small scale s_1. The multiplication function combines the advantages of the large and small scales. Therefore, the edge is much sharper while noise is better suppressed after application of function $M_I(x, y)$.

$$M_I(x,y) = \sqrt{P_I^x(x,y) + P_I^y(x,y)} \tag{1}$$

Then, the local maxima in function $M_I(x, y)$ should be used to eliminate redundant edges using a double-thresholding algorithm.

$$\mathrm{ESF}(I(x,y)) = M_I(x,y) - \sqrt{t_P^x + t_P^y} \tag{2}$$

Furthermore, when the input is a color image, the saliency map is obtained by conducting Canny edge detection on the intensity values, which are calculated by averaging the red, green, and blue channels of the input image. For instance, Fig. 1 shows three examples of edge saliency maps. Fig. 1(a) denotes the input color image. Canny edge detection is applied to the intensity image for obtaining the binary edge saliency map, as shown in Fig. (b). In Fig. 1(b), the contours of the bird and branch are clearly labeled with the pixel value of 1. Fig. 1(c) and (d) give another example of edge saliency map when the input image is a color image. In Fig. 1(d), the saliency object region in the edge saliency map is separated from the background. In addition, Fig. 1(e) presents an example of an anatomical image in grayscale. The result image highlights the high-spatial-resolution structure of the brain (Fig. 1(f)).

2.2. Color saliency feature

The context-aware saliency enhancement method is driven by the color stimulus [10]. Areas that have distinctive colors should be assigned higher values in the saliency map. Conversely, homogeneous areas should have lower saliency values. Pixel i is treated as salient only if the Euclidean distance $d_{color}(p_i, p_j)$ is large. Specifically, $d_{color}(p_i, p_j)$ denotes the Euclidean distance between patch p_i and its K ($K = 64$) most similar patches p_j ($j = 1, \cdots, K$) in *CIE* L^*a^*b color space. Multiple scales are utilized to conduct context-aware saliency enhancement to further decrease the salient of background pixels. The context-aware saliency is defined as

$$\mathrm{CSF}(I(x,y)) = \hat{S}_i G_i \tag{3}$$

where $\hat{S}_i$ is the salient of pixel i and G_i is a 2D Gaussian function that is positioned at the center of each image.

For instance, context-aware saliency maps are computed from three input color images in Fig. 2. The saliency maps emphasize the patterns in which the color is in high contrast. The background is endowed with lower intensity values, while the high-contrast color region is endowed with higher intensity values. In Fig. 2(b), the brown mushroom is captured due to its salient color. CSF marks as salient the entire mushroom as well as some of the stone, thus expressing the environment of the mushroom. In Fig. 2(e), the input functional image is mainly distributed over red and blue colors. Fig. 2(f) shows that higher intensity values in the saliency map correspond to the red and blue color regions from the input image. That is, the saliency map generated by CSF completely preserves the high-intensity regions of disease.

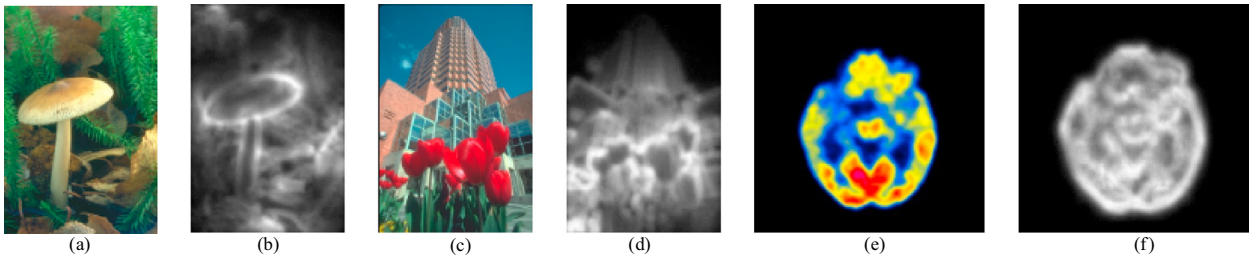

Fig. 2. Three examples of color saliency maps. (a), (c), and (e) are three input images. (b), (d), and (f) are the color saliency maps of input images (a), (c), and (b), respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Algorithm 1 Basic steps of the anatomical and functional image fusion scheme.

input: Anatomical image A, functional image B
output: Fused image F
Step 1: multi-scale image decomposition
$(\breve{c}_A^1, \bar{c}_A^1), (\breve{c}_A^2, \bar{c}_A^2), \cdots, (\breve{c}_A^n, \bar{c}_A^n) \leftarrow A, (\breve{c}_B^1, \bar{c}_B^1), (\breve{c}_B^2, \bar{c}_B^2), \cdots, (\breve{c}_B^n, \bar{c}_B^n) \leftarrow B$
Step 2: image fusion rules
(1) fuse smooth layers and detail layers using inter-scale dependencies
$\breve{c}_F^A \leftarrow \breve{c}_A^i \times \frac{\mathrm{EN}(\breve{c}_A^i)}{\sum_{i=1}^n \mathrm{EN}(\breve{c}_A^i)}$, $\bar{c}_F^A \leftarrow \bar{c}_A^i \times (\mathrm{ESF}(\bar{c}_A^i) == \max([\mathrm{ESF}(\bar{c}_A^1), \mathrm{ESF}(\bar{c}_A^2), \cdots, \mathrm{ESF}(\bar{c}_A^n)])$
$\breve{c}_F^B \leftarrow \breve{c}_B^i \times (\mathrm{CSF}(\breve{c}_B^i) == \max([\mathrm{CSF}(\breve{c}_A^1), \mathrm{CSF}(\breve{c}_A^2), \cdots, \mathrm{CSF}(\breve{c}_A^n)]), \bar{c}_F^B \leftarrow \bar{c}_B^i \times \frac{\mathrm{EN}(\bar{c}_B^i)}{\sum_{i=1}^n \mathrm{EN}(\bar{c}_B^i)}$
(2) calculate the fused image using intra-scale dependencies
$\breve{c}_F \leftarrow (\breve{c}_F^A + \breve{c}_F^B)./2, \bar{c}_F \leftarrow \max(\bar{c}_F^A, \bar{c}_F^B)$
Step 3: multi-scale image reconstruction
$F \leftarrow \breve{c}_F + \bar{c}_F$

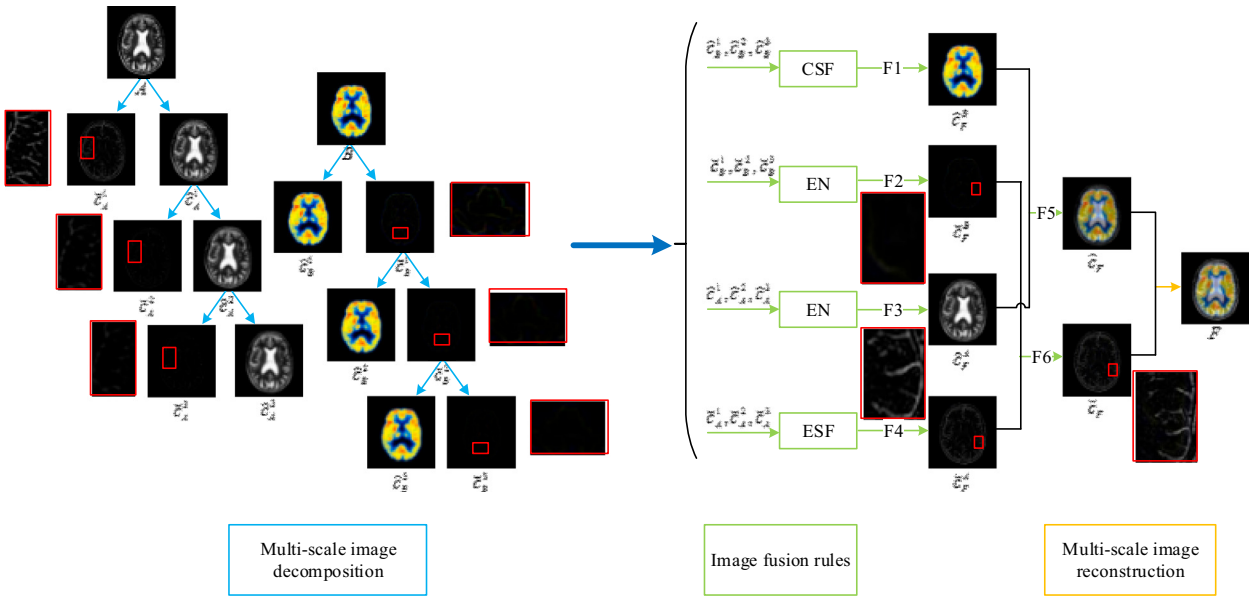

Fig. 3. Schematic diagram of the proposed methods based on saliency mapping. (CSF denotes the color saliency feature, ESF denotes the edge saliency feature, and EN denotes the entropy.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Proposed fusion method

The main process of Algorithm 1, which is proposed for fusion of anatomical and functional images is summarized in Fig. 3. The inputs of Algorithm 1 are an anatomical image A and a functional image B. First, multi-scale representations of anatomical and functional images are obtained by the average filter with a window size of 5×5 [15]. The input anatomical image A is decomposed into a series of smooth layers $\breve{c}_A^i$ and detail layers $\bar{c}_A^i$. And the input functional image B is decom-

posed into a series of smooth layers $\widehat{c}_B^i$ and detail layers $\breve{c}_B^i$. Then, the smooth and detail layers are fused using parallel saliency features. Using inter-scale dependencies, the fused smooth layer $\widehat{c}_F^A$ and the fused detail layer $\breve{c}_F^A$ of the anatomical image is obtained using entropy (EN) and ESF, respectively. The fused smooth layer $\widehat{c}_F^B$ and the fused detail layer of the functional image $\breve{c}_F^B$ is calculated by CSF and EN, respectively. And using intra-scale dependencies, the fused smooth layer $\widehat{c}_F$ and the fused detail layer $\breve{c}_F$ are obtained by the average and maximum functions (Note that the max function is defined in MATLAB). Finally, the output fused image F is obtained by the fused smooth layer $\widehat{c}_F$ and the fused detail layer $\breve{c}_F$.

3.1. Multi-scale image decomposition

An average filter smoothing operation of the input image C yields a smooth layer $\widehat{c}$, which contains the cartoon component of the original image, and a detail layer $\breve{c}$, which represents the high-frequency information, such as edge information [15]. Compared to a single-scale image representation, a multi-scale image representation can retain the saliency features of the input images at different scales [16]. A multi-scale representation of the input image is obtained by recursively extracting a series of detail layers from the filtered smooth image after the average filter is applied. After n recursive average filter smoothing operations, we obtain multi-scale image representations of input image, which include n smooth layers $\widehat{c}^1, \widehat{c}^2, \cdots, \widehat{c}^n$ and n detail layers $\breve{c}^1, \breve{c}^2, \cdots, \breve{c}^n$:

$$A = \sum_{i=1}^{n} (\widehat{c}_A^i + \breve{c}_A^i) \text{ and } B = \sum_{i=1}^{n} (\widehat{c}_B^i + \breve{c}_B^i) \tag{4}$$

3.2. Image fusion rules

Effective image fusion rules should not discard any saliency information from the input images. Image fusion rules normally consist of three components: (1) activity level measurement, in which the salient of each coefficient is highlighted in a decomposed image; (2) coefficient grouping, in which the coefficients are clustered in a specific way; and (3) coefficient combination, in which coefficients from input medical images with different modalities are combined into a single coefficient [29,33].

(1) Activity level measurement: Due to their specific features, anatomical and functional images are usually treated using different activity level measurements. Parallel saliency features are used as the activity level measurement method in image fusion rules. Anatomical images can provide high-spatial-resolution structural information of the brain. Therefore, ESF attempts to preserve the sharp edge among inter-scale pixels in the detail layer of the anatomical image, namely, $\breve{a}_A^i = \mathrm{ESF}(\breve{c}_A^i(x,y))$. An activity level measure is based on the EN for the smooth layer of the anatomical image, namely, $\widehat{a}_A^i = \mathrm{EN}(\widehat{c}_A^i(x,y))$. In contrast to an anatomical image, a functional image represents the high-intensity metabolism of the brain in pseudo-color. CSF is exploited to measure the saliency information of each pixel in the smooth layer of the functional image: $\widehat{a}_B^i = \mathrm{CSF}(\widehat{c}_B^i(x,y))$. For the detail layer of the functional image, EN is used to measure the activity of each pixel: $\breve{a}_B^i = \mathrm{EN}(\breve{c}_A^i(x,y))$.

(2) Coefficient grouping: The decision map is adopted to group the coefficients from the inputs. The value of the decision map determines the importance of the coefficients in the fused image. The sum of all decision maps equal 1, which ensures that no original information is lost. The decision maps are constructed as follows:

$$\widehat{d}_A^i = \widehat{a}_A^i / \sum_{i=1}^{n} \widehat{a}_A^i;\ \breve{d}_A^i = \begin{cases} 1, \text{if } \max(\left[\breve{a}_A^1, \breve{a}_A^2, \cdots, \breve{a}_A^n\right]) == \breve{a}_A^i \\ 0, \text{otherwise} \end{cases} \tag{5}$$

and

$$\widehat{d}_B^i = \begin{cases} 1, \text{if } \max(\left[\widehat{a}_B^1, \widehat{a}_B^2, \cdots, \widehat{a}_B^n\right]) == \widehat{a}_B^i; \\ 0, \text{otherwise} \end{cases} \breve{d}_B^i = \breve{a}_B^i / \sum_{i=1}^{n} \breve{a}_B^i \tag{6}$$

(3) Coefficient combination:

F1-F4: As shown in Fig. 3, F1-F4 represent the image fusion rules using inter-scale dependencies. Each pixel in the fused image is replaced by the higher activity level of the input images. To find better saliency information, an inter-scale strategy is adopted to address the coefficients of the same layer at different scales: $\widehat{c}^1, \widehat{c}^2, \cdots, \widehat{c}^n$ and $\breve{c}^1, \breve{c}^2, \cdots, \breve{c}^n$. Inter-scale dependencies between the coefficients may preserve more spatial information since the same spatial location is represented in the input images. The fused smooth layers of anatomical and functional images are calculated by $\widehat{c}_F^A = \sum_{i=1}^{n} (\widehat{c}_A^i \times \widehat{d}_A^i)$ and $\widehat{c}_F^B = \sum_{i=1}^{n} (\widehat{c}_B^i \times \widehat{d}_B^i)$. Similarly, the fused detail layers $\breve{c}_F^A$ and $\breve{c}_F^B$ are obtained by multiplying the multi-scale coefficients with their decision maps.

F5-F6: As shown in Fig. 3, we adopt intra-scale dependencies to combine image coefficients. Intra-scale dependencies between the coefficients may be used to discover the spatial clusters of the image coefficients. Average and maximum strategies are chosen to obtain the fused smooth layer $\widehat{c}_F$ and the fused detail layer $\breve{c}_F$ (F5-F6) [25,34].

3.3. Multi-scale image reconstruction

In this subsection, the fused image is obtained by a multi-scale image reconstruction scheme. The fusion methods in the frequency domain usually apply the inverse transform scheme that is used in multi-scale image decomposition. However, in the proposed method, we use simple arithmetic to construct the fused image in the spatial domain. In contrast to the fusion method in the frequency domain, the multi-scale image reconstruction scheme in the spatial domain performs better in terms of computational complexity. In the proposed method, the fused image F is obtained by the arithmetical addition of the fused smooth layer $\widehat{c}_F$ and the fused detail layer $\breve{c}_F$.

4. Empirical evaluation

4.1. Experimental settings

To evaluate the performance of the proposed method, it has been implemented in MATLAB R2010b. The experiments are performed on testing imaging data that were downloaded from Whole Brain Atlas database (available at [13]). The Whole Brain Atlas [13] is a benchmark database for evaluating the performance of multi-modal medical image fusion methods, which was established by Keith A. Johnson and J. Alex Becker at Harvard Medical School. Importantly, all the images of the database are co-aligned. The testing data that are used in the experiments include anatomical images (MRI-T1 and MRI-CBV) and functional images (PET-FDG and SPECT-Tc). MRI-T1 images are sagittal T1-weighted axial MRI images [11]. MRI-CBV images utilize changes in cerebral blood volume (CBV) in brain areas to detect lesion regions of MRI images [28]. PET-FDG images use ^{18}F-fluorodeoxyglucose (FDG) to detect lesions in PET images. PET-FDG images play a vital role in detecting functional information, rather than structural information [36]. SPECT-Tc images use specific technetium (Tc) to obtain functional information about the metabolic status of a tumor [23] in SPECT images. The proposed method in this paper (PRO) is compared with six methods, which are based on PCA [32], IHS [37], PCNN [35], SR [24], DSIFT [20], and LES [43], on fusion of MRI-CBV and SPECT-Tc images and fusion of MRI-T1 and PET-FDG images.

4.2. Objective image quality metrics

In addition to the subjective evaluation, objective image quality metrics are adopted for evaluating the performances of the fusion methods. The purpose of objective image quality metrics is to predict whether the qualities of the fused image are in accordance with human vision [2]. Conventionally, objective image quality metrics are divided into three categories: full-reference, reduced-reference and no-reference [17]. When the original reference image is available, four full-reference objective image quality metrics (structural similarity (*SSIM*), peak-signal-to-noise ratio (*PSNR*), visual saliency-induced index (*VSI*), and tone-mapped image quality index (*TMQI*) [38,39,46,48]) are used in this paper to measure the information that is transferred from the reference images without distortion I_R (the input anatomical and functional images$A$and$B$) to the distorted image I_F (the final fused image$F$) [49]. The higher the values of the full-reference objective image quality metrics, the better the performance of the fusion method. The four objective image quality metrics are briefly introduced as follows:

The *SSIM* metric [38] measures how well the structural information of input images is preserved. *SSIM* is computed by averaging $SSIM(I_A, I_F)$ and $SSIM(I_B, I_F)$. The task of *SSIM* is separated into three components: luminance l, contrast c and structure s.

$$SSIM(I_R, I_F) = [l(I_R, I_F)]^a \cdot [c(I_R, I_F)]^b \cdot [s(I_R, I_F)]^c \tag{7}$$

where the default values of a, b and c are 1/3.

The *PSNR* metric [39] is widely used in the image processing literature to evaluate the influence of noise in images. *PSNR* is computed by the average of $PSNR(I_A, I_F)$ and $PSNR(I_B, I_F)$.

$$PSNR(I_R, I_F) = 10 \cdot \lg \left[(X \times Y)^2 / \left(\sum_{x=1}^{X} \sum_{y=1}^{Y} [I_R(x,y) - I_F(x,y)]^2 / (X \times Y) \right)^{1/2} \right] \tag{8}$$

The *VSI* metric [48] is used to predict the ability of the visual saliency map to characterize the visual importance of a local region, between the reference image and the fused image. For a given image pixel position $i_{position} = (x, y)$, *VSI* is defined as the average of $VSI(I_A, I_F)$ and $VSI(I_B, I_F)$.

$$VSI(I_R, I_F) = \frac{\sum_{i_{position} \in \Omega} S_{I_R, I_F}(i_{position}) \cdot \max\left(VS_{I_R}(i_{position}), VS_{I_F}(i_{position})\right)}{\sum_{i_{position} \in \Omega} \max\left(VS_{I_R}(i_{position}), VS_{I_F}(i_{position})\right)} \tag{9}$$

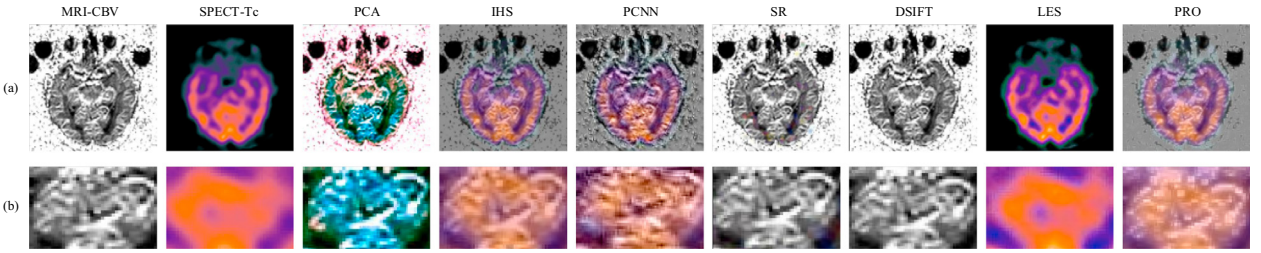

Fig. 4. The first example of fusion of MRI-CBV and SPECT-Tc images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

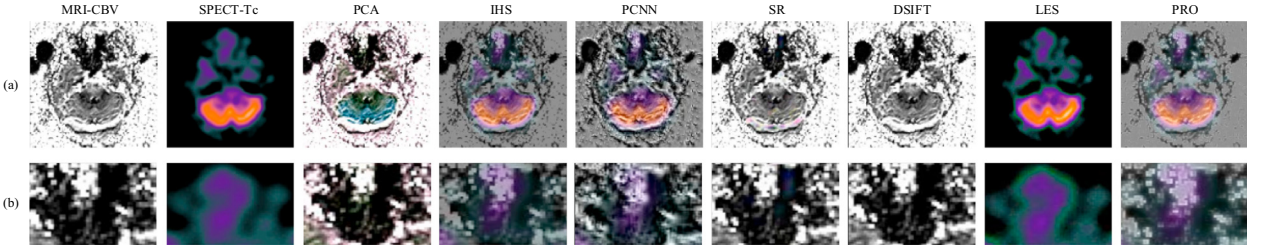

Fig. 5. The second example of fusion of MRI-CBV and SPECT-Tc images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $S_{I_R,I_F}(i_{position})$ refers to the local similarity of I_R and I_F, $VS_{I_R}(i_{position})$ and $VS_{I_F}(i_{position})$ refer to the visual saliency map of $I_R$and I_F, respectively.

The *TMQI* metric [46] aims to measure the brightness and contrast saliency features between the input image and the fused image. *TMQI* is defined as the average of $TMQI(I_A, I_F)$ and $TMQI(I_B, I_F)$.

$$TMQI(I_R, I_F) = aT^{\alpha} + (1-a)M^{\beta} \tag{10}$$

where S stands for the multi-scale structural fidelity between the reference image I_R and the fused (distorted) image I_F, and N is the statistical naturalness of the fused image I_F. The default values of the parameters in *TMQI* are $a = 0.8012$, $\alpha = 0.3046$ and $\beta = 0.7088$.

4.3. Subjective evaluation of Alzheimer's disease

4.3.1. Fusion of MRI-CBV and SPECT-Tc images

We perform five sets of experiments for fusing MRI-CBV and SPECT-Tc images to evaluate the performance of the proposed method in this paper. Figs. 4 and 5 show two clinical examples of Alzheimer's disease using the input MRI-CBV and SPECT-Tc images. The proposed method is compared with six existing fusion methods: PCA, IHS, PCNN, SR, DSIFT, and LES. Fig. 4(a) and (b) list the input MRI-CBV, SPECT-Tc images and the fused results that are obtained using different methods. The fused image using the proposed method contains both the high-spatial-resolution brain skeleton structural information from the input MRI-CBV image and high-intensity information from the input SPECT-Tc image, simultaneously (Fig. 4). In contrast to the proposed method, the PCA method fails to extract the saliency features from the functional images and yields results with color distortion. The images that are obtained using the SR and DSIFT methods show lower signal intensity indicated by the distorted color. The PCNN method produces obvious edge artifacts. By carefully observing the magnified regions in Fig. 4(b), the intensity values in the fused image that is obtained using the IHS method are of lower contrast. Fig. 5 shows another example of Alzheimer's disease. Fig. 5(b) shows the magnified regions of the fused images in Fig. 5(a). The gray and white matters of the brain disappear in the fused image that is obtained using the LES method. Additionally, the PCA method reduces the color detail information and blurs the sharp edge indicated by black color. The result images that is obtained using the IHS, PCNN, and proposed method integrate the functional SPECT-Tc color information with the MRI-CBV spatial detail information.

4.3.2. Fusion of MRI-T1 and PET-FDG images

Five sets of MRI-T1 and PET-FDG images are chosen as the testing data. Figs. 6 and 7 show two clinical examples of Alzheimer's disease. Figs. 6(a) and 7(a) give the fused images using various methods for the fusion of MRI-T1 and PET-FDG images. In addition, Figs. 6(b) and 7(b) show the magnified regions of the fused results. The MRI-T1 image provides high-spatial-resolution structural information of visceral organs. The PET-FDG image emphasizes the high-intensity functional information of local activity within tissue classes. The SR and DSIFT methods improve the visual quality of the input MRI-T1 image by preserving high-spatial-resolution structural boundaries information, such as edge and texture. While, these

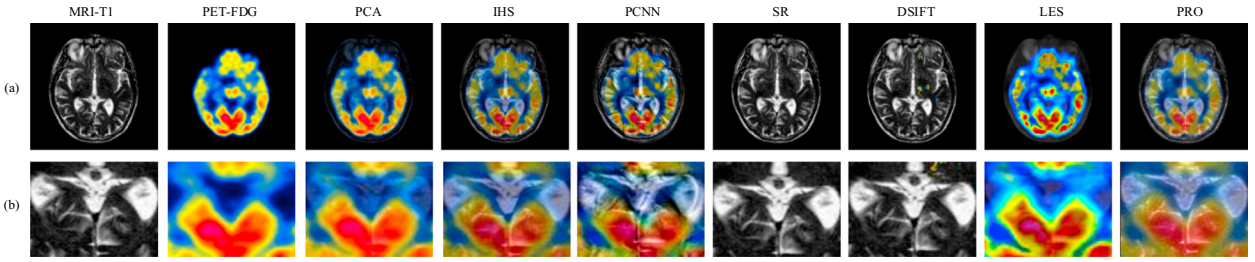

Fig. 6. The first example of fusion of MRI-T1 and PET-FDG images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

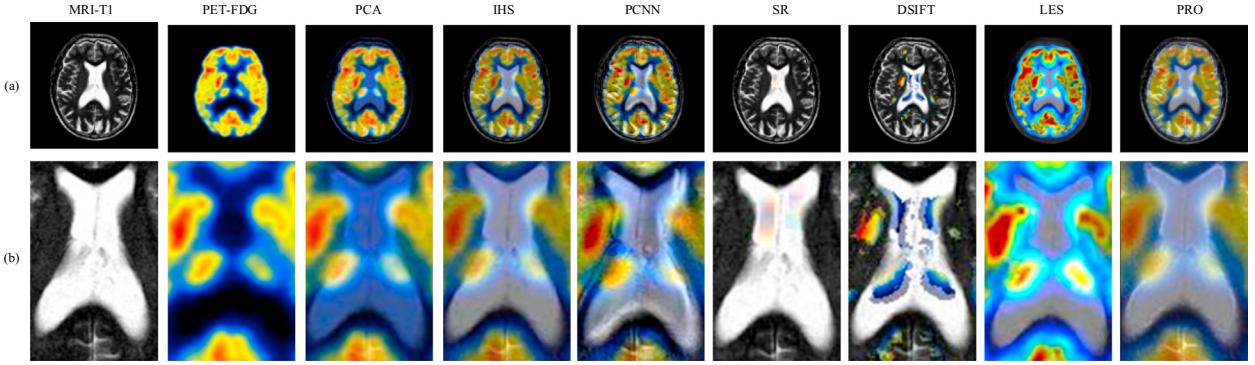

Fig. 7. The second example of fusion of MRI-T1 and PET-FDG images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Average metrics of MRI-CBV and SPECT-Tc image fusion.

	SSIM	*PSNR*	*VSI*	*TMQI*
PCA	0.5526	69.3592	0.9624	0.5243
IHS	0.5679	57.3423	0.9675	0.7249
PCNN	0.3625	56.5043	0.9669	0.5381
SR	**0.6067**	77.0902	0.9557	0.5314
DSIFT	0.6066	80.9744	0.9543	0.5214
LES	0.6033	**81.9707**	0.9576	0.5275
PRO	0.5390	57.9472	**0.9683**	**0.7565**

methods could not reveal the high-intensity activity information of the input PET-FDG image. The result image that is obtained using the LES method induces color distortion. It is noticeable that the IHS and proposed methods outperform the PCA, PCNN, SR, DSIFT, LES methods in intensity enhancement and preserving the PET-FDG unique color feature. However, in Fig. 6(b), the intensity is higher in the result image that is obtained using the proposed method, compared to that obtained using the IHS method. Fig. 7 shows another example of the fusion of MRI-T1 and PET-FDG images. The PCNN method induces pseudo edge in the fused image. The SR method shows high-contrast anatomical lesions while less color information. For the DSIFT method, the white matter of the brain is corrupted by the color of the input PET-FDG image. The proposed method performs better in preserving high-intensity skeleton.

4.4. Objective evaluation of Alzheimer's disease

Tables 1 and 2 present the objective evaluation of Alzheimer's disease in terms of four full-reference objective image quality metrics: *SSIM, PSNR, VSI*, and *TMQI*. Each value in the table is the average of five image pairs. Table 1 lists the objective comparison of MRI-CBV and SPECT-Tc images fusion. The proposed method obtains the largest values of the *VSI* metric, that is, the graph-based visual saliency (GBVS), gradient modulus (GM), and chrominance information maps of the fused image that is obtained using the proposed method are closest to those of the input MRI-CBV and SPECT-Tc images. For the *TMQI* metric, the proposed method performs best, as well. From Table 2, the proposed method performs best in terms of the *VSI* and *TMQI* metrics. The higher the *TMQI* metric is, the better the performance of the fusion method is. That means the fused images that is obtained using the proposed method could best preserve both the structural fidelity and the intensity statistics of input medical imaging data. Furthermore, the SR method obtains the largest value in terms of the *SSIM*

Table 2
Average metrics of MRI-T1 and PET-FDG image fusion.

	SSIM	PSNR	VSI	TMQI
PCA	0.8130	70.7146	0.9598	0.6918
IHS	**0.8217**	65.3056	0.9634	0.7095
PCNN	0.7106	63.8162	0.9642	0.6672
SR	0.7999	**81.2812**	0.9442	0.6303
DSIFT	0.7873	71.6168	0.9443	0.6814
LES	0.7695	66.4990	0.9539	0.6945
PRO	0.7981	64.8873	**0.9645**	**0.7171**

metric for the fusion of MRI-CBV and SPECT-Tc images. The IHS method performs best in preserving structural information from the input MRI-T1 and PET-FDG images. The LES and SR methods are robust to noise, as they obtain the best values of the *PSNR* metric.

4.5. Discussion

The proposed method utilizes a new image fusion rule, which uses parallel features, namely ESF and CSF, to measure the activity levels of the images. The results in Tables 1 and 2 and Figs. 4–7 can be summarized as follows:

(1) From Table 1, the proposed method obtains the best results in terms of two objective image quality metrics: *VSI* and *TMQI*. According to Figs. 4 and 5, the result images are with high-spatial-resolution and high-intensity that are obtained using the proposed method. That is, the subjective evaluation is in accordance with the objective evaluation. The *SSIM* metric measures whether the structural information of the input images is well preserved in the result image. The fused images that are obtained using the SR and DISFT methods seem the same as the input MRI-T1 image, as shown in Figs. 4 and 5. As a result, the SR and DSIFT methods obtain the largest values in terms of the *SSIM* metric. However, the result images poorly preserve the high-intensity functional information from the SPECT-Tc image.

(2) Figs. 6 and 7 show clinical examples of the fusion of anatomical images, which are obtained by the MRI-T1 scanner, and functional images, which are obtained by the PET-FDG scanner. The *PSNR* metric of the SR method is larger, compared with the other existing methods. The PCNN method introduces obvious color distortion in Figs. 6 and 7. Correspondingly, the PCNN method obtains lower values in terms of the *TMQI* and *VSI* metrics. The proposed method perfectly preserves the sharp edge of white and gray matters from the input anatomical images and the high intensity of color detail from the input functional images, in Figs. 6 and 7. Table 2 verifies the effectiveness of the proposed method, since it achieves the largest values of the *TMQI* and *VSI* metrics.

5. Conclusions

This paper has proposed a fusion method using a new image fusion rule, which is based on the parallel saliency features: ESF and CSF. The innovative aspects of the proposed method are the following: (1) It uses two different saliency features for preserving sharp edge information from anatomical images and preserving color detail information from functional images. (2) The experiments are performed on images that are obtained by several different modalities: MRI-T1, MRI-CBV, PET-FDG, and SPECT-Tc. The results suggest that the proposed fusion method lead to the high-spatial-resolution and high-intensity fused images by the fusion of anatomical and functional images based on subjective and objective evaluations of Alzheimer's disease in comparison with other image fusion methods.

However, the proposed method in this paper has some limitations. First, saliency features are only used to measure the activity levels of MRI, PET, and SPECT images. Saliency features exhibit great potential for fusion of anatomical and functional images. Utilizing saliency features in measuring the activity levels of much more medical imaging data will be a topic of future research [30,49]. Then, the results of the subjective evaluation of the fusion methods are not consistent with those of the objective evaluation. The superiority of the proposed method is indicated by its better performance in terms of the selected objective quality metrics. However, the statistical significance of the differences between different methods is unknown. In the future, significant testing tools will be adopted to evaluate the performances of fusion methods in terms of objective image quality metrics [19,42].

Acknowledgments

This work was supported in part by Natural Science Foundation of China (No.61272195, 61201383, 61472055, U1401252), National Science and Technology Major Project (2016YFC1000307-3), Program for New Century Excellent Talents in University of China (NCET-11-1085), Chongqing Outstanding Youth Fund (cstc2014jcyjjq40001) and Chongqing Research Program of Application Foundation and Advanced Technology (cstc2012jjA1699).

References

[1] B. Alexe, T. Deselaers, V. Ferrari, What is an object? in: 2010 IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 73–80.

[2] A.D. Angelo, Z. Li, A.B. Mauro, A full-reference quality metric for geometrically distorted images, IEEE Trans. Image Process. 19 (4) (2010) 867–881.
[3] H. Ao, N. Yu, Edge saliency map detection with texture suppression, in: IEEE Computer Society International Conference on Image & Graphics, 2011, pp. 309–313.
[4] P. Bao, D. Zhang, X. Wu, Canny edge detection enhancement by scale multiplication, IEEE Trans. Pattern Anal. Mach. Intell. 27 (9) (2005) 1485–1490.
[5] G. Bhatnagar, Q.M.J. Wu, Z. Liu, Human visual system inspired multi-modal medical image fusion framework, Exp. Syst. Appl. 40 (5) (2013) 1708–1720.
[6] M. Choi, A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter, IEEE Trans. Geosci. Remote Sens. 44 (6) (2006) 1672–1682.
[7] S. Daneshvar, H. Ghassemian, MRI and PET image fusion by combining IHS and retina-inspired models, Inf. Fusion 11 (2010) 114–123.
[8] O.F. Donati, T.F. Hany, C.S.R. CS, G.K. von Schulthess, B. Marincek, B. Seifert, D. Weishaupt, Value of retrospective fusion of PET and MR images in detection of hepatic metastases: comparison with 18 F-FDG PET/CT and Gd-EOB-DTPA-enhanced MRI, J. Nucl. Med. 51 (5) (2010) 692–699.
[9] M. Farhadi, S. Mahmoudian, F. Saddadi, et al., Functional brain abnormalities localized in 55 chronic tinnitus patients: fusion of SPECT coincidence imaging and MRI, J. Int. Soc. Cerebral Blood Flow Metab. 30 (4) (2010) 864–870.
[10] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1915–1926.
[11] M.C. Hernández, K.J. Ferguson, F.M. Chappell, et al., New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images, Eur. Radiol. 20 (7) (2010) 1684–1691.
[12] A.P. James, B.V. Dasarathy, Medical image fusion: a survey of the state of the art, Inf. Fusion 19 (2014) 4–19.
[13] K.A. Johnson, J.A. Becker, The Whole Brain Altas, 2011 [Online]. Available: http://www.med.harvard.edu/aanlib/.
[14] M. Kim, D.K. Han, H. Ko, Joint patch clustering-based dictionary learning for multimodal image fusion, Inf. Fusion 27 (2015) 198–217.
[15] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013) 2864–2875.
[16] H. Li, X. Li, Z. Yu, et al., Multifocus image fusion by combining with mixed-order structure tensors and multiscale neighborhood, Inf. Sci. (2016) 25–49.
[17] L. Li, Y. Yan, Z. Lu, et al., no-reference quality assessment of deblurred images based on natural scene statistics, IEEE Access 5 (99) (2017) 2163–2171.
[18] S.A. Li, H. Yin, Multimodal image fusion with joint sparsity model, Opt. Eng. 50 (6) (2011) 409–421.
[19] Z. Liu, E. Blasch, V. John V, Statistical comparison of image fusion algorithms: recommendations, Inf. Fusion 36 (2017) 251–260.
[20] Y. Liu, S. Liu, Z. Wang, Multi-focus image fusion with dense SIFT, Inf. Fusion 23 (2015) 139–155.
[21] T. Liu, J. Sun, N.N. Zheng, et al., Learning to detect a salient object, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[22] Z. Liu, H. Yin, Y. Chai, et al., A novel approach for multimodal medical image fusion, Exp. Syst. Appl. 41 (16) (2014) 7425–7435.
[23] D.J. Loeffelbein, E. Mielke, A.K. Buck, et al., Impact of nonhybrid ^{99m}Tc-MDP-SPECT/CT image fusion in diagnostic and treatment of oromaxillofacial malignancies, Mol. Imaging Biol. 12 (1) (2010) 71–77.
[24] J. Ma, C. Chen, C. Li, et al., Infrared and visible image fusion via gradient transfer and total variation minimization, Inf. Fusion 31 (2016) 100–109.
[25] Q.G. Miao, C. Shi, P.F. Xu, et al., A novel algorithm of image fusion using shearlets, Opt. Commun. 284 (6) (2011) 1540–1547.
[26] N. Mitianoudis, T. Stathaki, Pixel-based and region-based image fusion schemes using ICA bases, Inf. Fusion 8 (2) (2007) 131–142.
[27] S. Mitra, B.U. Shankar, Medical image analysis for cancer management in natural computing framework, Inf. Sci. 306 (2015) 111–131.
[28] T. Mueggler, F. Razoux, H. Russig, et al., Mapping of CBV changes in 5-HT_{1A} terminal fields by functional MRI in the mouse brain, Eur. Neuropsychopharmacol. 21 (4) (2011) 344–353.
[29] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Inf. Fusion 4 (4) (2003) 259–280.
[30] L. Qu, S. He, J. Zhang, et al., RGBD salient object detection via deep fusion, IEEE Trans. Image Process. 26 (5) (2017) 2274–2285.
[31] F. Sebastian, Reducing attenuation and motion artefacts in hybrid PET/MR imaging, (2015) 1–141.
[32] H.R. Shahdoosti, H. Ghassemian, Combining the spectral PCA and spatial PCA fusion methods by an optimal filter, Inf. Fusion 27 (2016) 150–160.
[33] R. Shen, I. Cheng, A. Basu, Cross-scale coefficient selection for volumetric medical image fusion, IEEE Trans. Biomed. Eng. 60 (4) (2013) 1069–1079.
[34] R. Singh R, A. Khare, Fusion of multimodal medical images using Daubechies complex wavelet transform – a multiresolution approach, Inf. Fusion 19 (3) (2014) 49–60.
[35] M.M. Subashini, S.K. Sahoo, Pulse coupled neural networks and its applications, Exp. Syst. Appl. 41 (8) (2014) 3965–3974.
[36] Y. Song, W. Cai, H. Huang, et al., Lesion detection and characterization with context driven approximation in thoracic FDG PET-CT images of NSCLC studies, IEEE Trans, Med. Image 33 (2) (2014) 408–421.
[37] Y.M. Tu, S.C. Su, H.C. Shyu, et al., A new look at IHS-like image fusion methods, Inf. Fusion 2 (3) (2001) 177–186.
[38] Z. Wang, A.C. Bovik, H.R. Sheikh, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[39] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2011) 1185–1198.
[40] L. Wang, B. Li, L.F. Tian, EGGDD: An explicit dependency model for multi-modal medical image fusion in shift-invariant shearlet transform domain, Inf. Fusion 19 (11) (2014) 29–37.
[41] Z. Wang, Y. Ma, Medical image fusion using m-PCNN, Inf. Fusion 9 (2) (2008) 176–185.
[42] Y. Xia, H. Leung, Performance analysis of statistical optimal data fusion algorithms, Inf. Sci. (2014) 808–824.
[43] Z. Xu, Medical image fusion using multi-level local extrema, Inf. Fusion 19 (11) (2014) 38–48.
[44] Y. Yang, Y. Que, S. Huang S, et al., Multiple visual features measurement with gradient domain guided filtering for multisensor image fusion, IEEE Trans. Instrum. Meas. 99 (2017) 1–13.
[45] S. Yang, M. Wang, L. Jiao, Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis, Inf. Fusion 13 (2012) 177–184.
[46] H. Yeganeh, Z. Wang, Objective quality assessment of tone-mapped images, IEEE Trans. Image Process. 22 (2) (2013) 657–667.
[47] X. Zhang, X. Li, Y. Feng, et al., Image fusion with internal generative mechanism, Exp. Syst. Appl. 42 (5) (2015) 2382–2391.
[48] L. Zhang, Y. Shen, H. Li, VSI: a visual saliency-induced index for perceptual image quality assessment, IEEE Trans. Image Process. 23 (10) (2014) 4270–4281.
[49] Y. Zheng, E.A. Essock, B.C. Hansen, et al., A new metric based on extended spatial frequency and its application to DWT based fusion algorithms, Inf. Fusion 8 (2) (2007) 177–192.