

## Article

# MRI Image Fusion Based on Sparse Representation with Patch-based Multiple Salient Features Measurement

Qiu Hu <sup>1</sup>, Weiming Cai <sup>1,\*</sup>, Shuwen Xu <sup>2</sup> and Shaohai Hu <sup>3</sup>

<sup>1</sup> School of Information and Engineering, Ningbo Technology University, Ningbo, China; hu\_qiu@nbt.edu.cn, caiwm@nit.zju.edu.cn

<sup>2</sup> Third Research Institute of China Electronics Technology Group Corporation, Beijing, China; hq13261670359@126.com

<sup>3</sup> Institute of Information Science, Beijing Jiaotong University, Beijing, China; shhu@bjtu.edu.cn

\* Correspondence: caiwm@nit.zju.edu.cn; Tel.: 13735564866

**Abstract:** Multimodal medical image fusion is a fundamental but challenging problem in the fields of brain science research and brain disease diagnosis, and it is challenging for sparse representation (SR)-based fusion to characterize activity level with single measurement and no loss of effective information. In this paper, the Kronecker-criterion-based SR framework is applied for medical image fusion with a patch-based activity level integrating salient features of multiple domains. Inspired by the formation process of vision system, the spatial saliency is characterized by textural contrast (TC), which is composed of luminance and orientation contrasts to promote more highlighted texture information to participate in the fusion process. As substitution of the conventional  $l_1$ -norm-based sparse saliency, a metric of sum of sparse salient features (SSSF) is used for promoting more significant coefficients to participate in the composition of activity level measure. The designed activity level measure is verified to be more conducive to maintain the integrity and sharpness of detailed information. Various experiments on multiple groups of clinical medical images verify the effectiveness of the proposed fusion method on both visual quality and objective assessment. Furthermore, the research work of this paper is helpful for further detection and segmentation of medical images.

**Keywords:** Multimodality medical image; Image fusion; Sparse representation (SR); Kronecker criterion; Activity level measure

## 1. Introduction

Over the past decades, medical image fusion has undergone major achievements in clinical diagnosis research [1]. The main purpose of medical image fusion is to combine the complementary information from various sensors to construct a new image to assist the diagnosis for medical experts. Despite simplicity of the idea, many challenges related to the theoretical background and the nature of medical images that need to be resolved. For instance, computed tomography (CT) imaging is informative for the dense tissues but lacks of soft tissue information. In contrast, magnetic resonance imaging (MRI) is more suitable for soft tissue but short of dense tissue information. More crucially, it tends to be ineffective to characterize the symptoms of different diseases with single imaging.

To overcome these challenges, varieties of image fusion methods have been proposed. Image content can be either visual (i.e. color, shape, texture) or textual (i.e. identify dataset appearing within image), some new advances in the fusion field consider above of the two aspects simultaneously [2, 3], and to further improve fusion performance, some new features such as different image moments [4-6] can also be used in image fusion.

The mainstream directions of image fusion mainly focus on the visual content including spatial domain [7, 8] and transform domain [9, 10]. The former usually addresses the fusion issue via image blocks or pixel-wise gradient information for the multi-focus fusion [11]-[13] and the multi-exposure

fusion [14]-[16] tasks. The latter merges the transform coefficients relevant to source images with different reconstruction algorithms to obtain the fused image, which is recognized to be effective for multimodal image fusion [17, 18]. The multi-scale transforms (MST)-based medical image fusion is a mainstream research direction. Dual-tree complex wavelet transform (DTCWT) [19], non-subsampled shearlet transform (NSST) [20] and non-subsampled contourlet transform (NSCT) [21] are the conventional MST methods for image fusion. In recent years, some novel MST-based methods have been proposed. Xia combines sparse representation with pulse coupled neural network (PCNN) in NSCT domain for medical image fusion [22]. Yin proposes a parameter-adaptive pulse coupled neural network in NSST domain (NSST-PAPCNN)-based medical image fusion strategy [23]. Dinh proposes a Kirsch compass operator with marine predator algorithm based method for medical image fusion [24].

Different from MST, the principle of SR is much in accord with the human visual system (HVS), and compared to the MST-based methods, two main distinctions are involved in the SR-based methods. For the former distinction, the fixed basis limits the MST-based methods to express significant features, while the SR-based methods are flexible to procure more intrinsic features with dictionary learning. For the latter distinction, the MST-based methods are sensitive to noise and mis-registration with large decomposition level settings, while the SR-based methods with overlapping patch-wise mode are robust to mis-registration, and this guarantees the accuracy of spatial location of tissues. Therefore, a widely range of research on SR-based medical image fusion is attracted in recent years [25]-[27].

However, there still remain drawbacks. Firstly, it may be insufficient to handle fine details with the over-complete dictionary, and the high redundant dictionary will lead to visual artifacts in fused result [28]. Secondly, the dictionary atoms are updated in column vector form result in the loss of correlation and structural information. In addition to the drawbacks of SR itself, it may reduce the fusion weight accuracy inevitably with unreasonable fusion strategy of coefficients. One issue focuses on the activity level measure which helps to recognize the distinct features in fusion process, and another issue concentrates on the integration of coefficients into counterparts of fused image. For the former issue, the  $l_1$ -norm mode is a conventional solution to describe the detailed information contained in sparse vectors [29], while the solution is insufficient to express the sparse saliency well since detail information characterizes activity level with same weight cannot be highlighted. Furthermore, as SR is an approximate technique, it tends to fail to reflect the salient features of sparse coefficient maps accurately with single measurement of activity level, thereby further leading to the loss of detailed information. For the latter issue, it may reduce the contrast of fused image with the weighted averaging rule, and the maximum absolute rule enables the fused image to absorb the main visual information of source images with the cost of minor information loss.

Based on the above discussion, we adopt a promising signal decomposition model, known as the Kronecker-criterion-based SR [30], to the medical image fusion problem. The main contributions of this work are illustrated as follows:

- a) The Kronecker-criterion-based SR with a designed activity level measure integrating salient features of multiple domains will effectively reduce the loss of structural detailed information in fusion process.
- b) Inspired by the formation process of vision system, the spatial saliency by textural contrast consisted of luminance and orientation contrasts can promote more highlighted texture information to participate in fusion process.
- c) Compared with the  $l_1$ -norm based activity level measure in sparse vectors, the transform saliency by sum of sparse salient features can highlight more coefficients to composite activity level measure through sum of differences of the adjacent areas.

The rest of this paper is organized as follows. Section 2 gives a brief description of conventional sparse representation theory and the Kronecker-criterion-based SR, i.e. separable dictionary learning algorithm. The detailed fusion scheme is described in Section 3. Experimental results and discussion are given in Section 4. Finally, Section 5 concludes the paper.

## 2. Related work

### 2.1. SR-based image fusion

SR reflects the sparsity of natural signals with minimal sparse coefficients, and this is consistent with the principle of HVS [31]. Given  $y \in R^m$  in vector mode of signal sample  $Y \in R^{\sqrt{m} \times \sqrt{m}}$  and an over-complete dictionary  $D \in R^{m \times n}$  ( $m < n$ ), the objective function of dictionary learning consisted of fidelity term and penalty term is defined as

$$\arg \min_{D, \alpha} \frac{1}{2} \|y - D\alpha\|_2^2 + \lambda R(\alpha) \quad (1)$$

where  $\alpha$  means the sparse vector,  $\|\bullet\|_2^2$  means the L<sub>2</sub>-norm and  $\lambda$  means the regularization parameter of penalty term  $R(\alpha)$ . SR can be roughly divided into two categories as the greedy scheme (e.g. matching pursuit (MP) [32] and orthogonal matching pursuit (OMP) [33]) with  $R(\alpha) = \|\alpha\|_0$  and the convex optimization scheme (e.g. alternating direction method of multipliers (ADMM) [34]) with  $R(\alpha) = \|\alpha\|_1$ . The extremely high complexity inhibits the practicality of the convex optimization scheme, while the greedy scheme has superiority in this regard.

In the process of conventional SR-based image fusion, it may be insufficient to handle fine details with over-complete dictionary, since atoms (i.e. vectors) of the pre-trained over-complete dictionary are updated one-by-one with the methods of optimal directions (MOD) or k-singular value decomposition (K-SVD), and this can be understood as extracting image texture information from only one dimensional direction, while this breaks the potential correlations within image, and thus causes the obtained pre-trained dictionary unstructured. Meanwhile, the high redundant dictionary is sensitive to random noise and may cause visual artifacts. Therefore, there is a deviation between source and fused images to some extent.

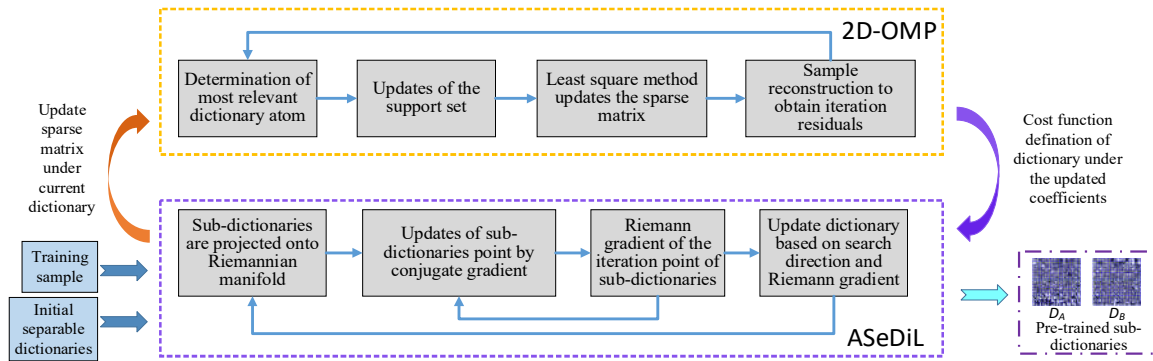
### 2.2. Separable dictionary learning algorithm

To overcome the aforementioned deficiencies of SR for image fusion, the Kronecker-criterion-based separable structure has received significant attention [30]. On premise of ensuring the quality of image reconstruction, the penalty in L<sub>0</sub>-norm with  $R(\alpha) = \|\alpha\|_0$ , and the corresponding objective function of separable dictionary learning is defined as

$$\arg \min_{D_A, S, D_B} \|S\|_0 \text{ such that } D_A S D_B^T = Y, S \in R^{n \times n} \quad (2)$$

where  $S$  means a sparse matrix. As cross product of the over-complete dictionary  $D$ , the sub-dictionaries  $D_A \in R^{\sqrt{m} \times n}$  and  $D_B \in R^{\sqrt{m} \times n}$  are obtained by the Kronecher product criterion, and for simplicity, we set the both sub-dictionaries with same size.

The steps of separable dictionary learning algorithm include sparse coding and dictionary update, and the dictionary optimization problems will be carried out by the extensional 2-dimension OMP (2D-OMP) greedy algorithm and the ASeDiL (Analytic Separable Dictionary Learning) algorithm to get the sparse coefficients and the pre-trained sub-dictionaries  $\{D_A, D_B\}$ , respectively through [35], and the dictionary pre-training model is shown in Figure 1.



**Figure 1.** Dictionary pre-trained model on Riemannian manifold.

The process of sparse coding consists of a four-step iterative loop, including the determination of most relevant dictionary atom, updates of the support set, updates of the sparse matrix  $S$ , and refactoring residual updates. To get the sparsest representation under current dictionary, the objective function of sparse coding is expressed as

$$\arg \min_S \|S\|_0 \text{ such that } \|D_A S D_B^T - Y\|_2^2 < \varepsilon \quad (3)$$

where  $\varepsilon$  means the tolerance of reconstruction error, and when  $\|D_A S_j D_B^T - Y_j\|_F^2 > \varepsilon$ , the condition of iterations will be terminated.

Combining the constraints with the  $L_2$ -norm of dictionary atoms equaling to 1 and no correlation of atoms in dictionary, log function is employed to fit the full rank and column irrelevance of sub-dictionaries. Then, the objective function of dictionary update is written as

$$\arg \min_{D_A, D_B} \|D_A S D_B^T - Y\|_2^2 + \omega[p(D_A) + p(D_B)] + \psi[h(D_A) + h(D_B)] \quad (4)$$

where  $\omega$  and  $\psi$  mean the fitting parameters, and  $h(D_A)$ ,  $p(D_A)$ ,  $h(D_B)$ ,  $p(D_B)$  are defined as

$$h(D_A) = -\frac{1}{m \log(m)} \log \det\left(\frac{1}{n} D_A D_A^T\right), p(D_A) = -\log(1 - ((D_A)^T (D_A))^2) \quad (5)$$

$$h(D_B) = -\frac{1}{m \log(m)} \log \det\left(\frac{1}{n} D_B D_B^T\right), p(D_B) = -\log(1 - ((D_B)^T (D_B))^2) \quad (6)$$

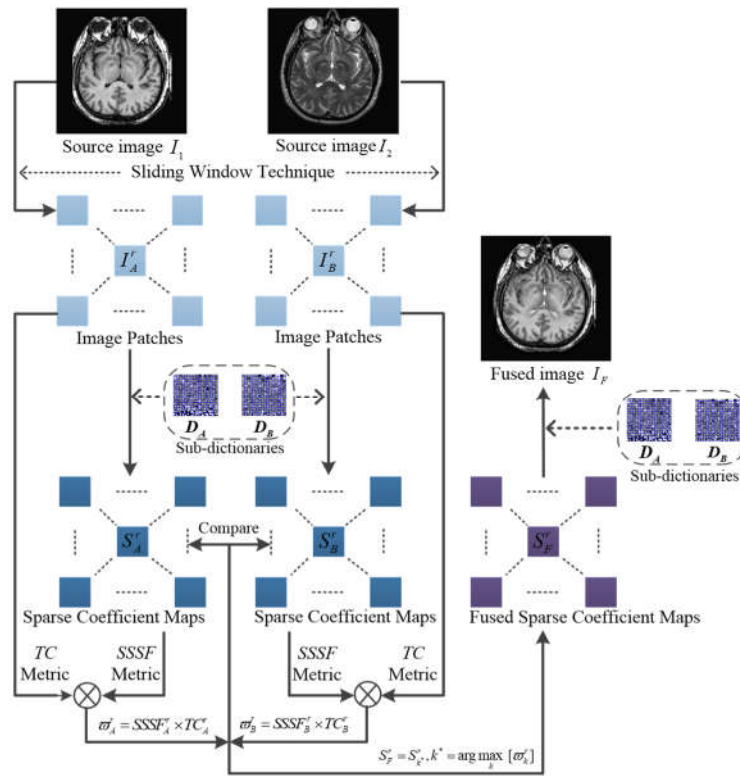
By means of geodesics on Riemannian manifold, the dictionary update adopts conjugate gradient method to correct the most rapid descent direction of the iteration point of dictionary update, which ensures the rapid convergence of cost function and improves the efficiency of dictionary update.

With the above separable structure, the obtained sparse matrix composed by correlation coefficients can characterize more textural and structural information, and this can not only increase dimensions of texture extraction without adding dictionary redundancy, but also ameliorate the accuracy of texture extraction with effective noise suppression performance. Through the above separable dictionary learning algorithm, the pre-trained sub-dictionaries can be obtained. Then, the pre-trained sub-dictionaries will participate in the subsequent transform saliency measure characterization process to extract features from source images.

### 3. Proposed fusion method

The framework of the proposed method is shown in Figure 2. Suppose that there are  $K$  pre-registered source images denoted as  $I_k, k \in \{1, 2, \dots, K\}$ , the  $r$ -th overlapping image patch of the  $k$ -th source image  $I_k^r$  is obtained through the sliding window technique, and the corresponding sparse coefficient map  $S_k^r$  is learned through sparse coding process in the separable dictionary learning algorithm with the pre-trained sub-dictionaries  $\{D_A, D_B\}$ . The proposed SR-based medical image

fusion with measurement integrating spatial saliency and transform saliency consists of the following two steps.



**Figure 2.** Overall framework of the proposed method.

### 3.1. The measurement of activity level for fusion

#### 3.1.1. Spatial saliency by textural contrast

In general, salient area is recognized by vision system from retina to visual cortex, and one of the early information in retina is luminance contrast, and orientation contrast in visual cortex is involved to understand the context at higher levels. We are inspired to attempt to express spatial saliency by allowing textural contrast defined by luminance contrast and orientation contrast.

First of all, the luminance contrast is defined by considering the distinctive of intensity attribute between each pixel and the corresponding image patch. To increase useful dynamic ranges to suppress high contrast effectively in the background, the  $n$ -th order statistic is applied as

$$LC_k^r(x, y) = \left| \hat{\mu}_k^r - \frac{1}{M} \sum_{(x, y) \in \Phi} I_k^r(x, y) \right|^n \quad (7)$$

where  $\hat{\mu}_k^r$  denotes the mean of luminance values over the  $r$ -th patch in the  $k$ -th source image  $I_k^r$ .  $\Phi$  and  $M$  represent a  $3 \times 3$  neighborhood with pixel  $(x, y)$  centered and its size, respectively.

Along with luminance contrast, the local image structure is captured by orientation contrast through weighted structure tensor. It is worth noting that we focus on weighted gradient information rather than gradient itself, and this highlights main features of source images. The weighted structure tensor can effectively summarize the dominant orientation and the energy along this direction based on the weighted gradient field as

$$G_k^r(x, y) = \begin{bmatrix} \sum_{k=1}^K \left( a_k^r(x, y) \frac{\partial I_k^r}{\partial x} \right)^2 & \sum_{k=1}^K a_k^r(x, y) \frac{\partial I_k^r}{\partial x} \frac{\partial I_k^r}{\partial y} \\ \sum_{k=1}^K a_k^r(x, y) \frac{\partial I_k^r}{\partial x} \frac{\partial I_k^r}{\partial y} & \sum_{k=1}^K \left( a_k^r(x, y) \frac{\partial I_k^r}{\partial y} \right)^2 \end{bmatrix} \quad (8)$$

where  $\partial I_k^r / \partial x$  and  $\partial I_k^r / \partial y$  denote the gradients along  $x$  and  $y$  directions, respectively at given pixel  $(x, y)$ . The weight function  $\omega_k(x, y)$  is calculated by

$$\omega_k(x, y) = \frac{LSM_k^r(x, y)}{\sqrt{\sum_{k=1}^K (LSM_k^r(x, y))^2}} \quad (9)$$

where  $LSM_k^r(x, y)$  means the local salient metric, which reflects the importance of pixel  $(x, y)$  by computing the sum of intensity around it, defined as

$$LSM_k^r(x, y) = \sum_{(x, y) \in \Phi} \left( \left| \frac{\partial I_k^r(x, y)}{\partial x} \right| + \left| \frac{\partial I_k^r(x, y)}{\partial y} \right| \right) \quad (10)$$

where  $|\cdot|$  denotes the absolute value operator. The local salient metric is sensitive to the edges and texture while insensitive in the flat part. To express local image structure, the weighted structure tensor as a semi-definite matrix can be decomposed by eigenvalue decomposition as

$$G_{I_k^r} = V \begin{pmatrix} \beta_1^2 & \\ & \beta_2^2 \end{pmatrix} V^T \quad (11)$$

The orientation contrast related to the eigenvalues of  $\beta_1$  and  $\beta_2$  of this matrix is calculated as

$$OC_k^r(x, y) = \sqrt{(\beta_1 + \beta_2)^2 + \eta(\beta_1 - \beta_2)^2} \quad (12)$$

where  $\eta > -1$ , this parameter can determine the relative emphasis of the orientation contrast to the corner-like structures effectively.

Since it is assumed that salient area contains luminance contrast and orientation contrast as mentioned, the texture contrast is then defined with the two parts as

$$TC_k^r(x, y) = LC_k^r(x, y) \times OC_k^r(x, y) \quad (13)$$

Here, each part is smoothed by Gaussian filtering as in [36] and  $TC_k^r$  is normalized to [0,255] for gray-scale representation.

### 3.1.2. Spatial saliency by textural contrast

Compared with the conventional transform-based activity level measure which uses the  $L_1$ -norm to describe the detail information contained in sparse vectors, SSSF metric is able to highlight more significant coefficients to participate in the composition of activity level measure through the sum of differences of adjacent areas, which is defined as

$$SSSF_k^r(x, y) = \sum_{p=-P}^P \sum_{q=-Q}^Q [LSSF_k^r(x+p, y+q)]^2 \quad (14)$$

where  $P$  and  $Q$  determine a sliding window with the size of sparse matrix equal to the  $r$ -th patch in the  $k$ -th source image  $I_k^r$ . The local sparse salient feature (LSSF) metric is defined by the sparse saliency diversity of adjacent pixels and is calculated as

$$LSSF_k^r(x, y) = \sqrt{\sum_{(m, n) \in \Phi} [S_k^r(x, y) - S_k^r(m, n)]^2} \quad (15)$$

where  $\Phi$  denotes a square window centered with certain sparse coefficient that corresponding to the pixel  $(x, y)$  in the source image patch  $I_k^r$ .

### 3.2. Fusion scheme

Combining the transform saliency and the spatial saliency of image patch, the proposed activity level measure is defined as

$$\varpi_k^r(x, y) = SSSF_k^r(x, y) \times TC_k^r(x, y) \quad (16)$$

where  $\varpi_k^r$  is the measurement result of the source image patch  $I_k^r$ . Then, the maximum weighted activity level measure is used to achieve the fused coefficient map as



$$S_F^r(x,y)=S_{k^*}^r(x,y),\; k^*=\arg\max_k[\varpi_k^r(x,y)] \tag{17}$$

Then, the fusion result is obtained by sparse reconstruction as

$$I_F^r=D_AS_F^rD_B^T \tag{18}$$

The final fused image  $I_F$  is constituted by the overall fused image patches.

4. Experiments

4.1. Experimental setting

4.1.1. Source images

In our experiments, three categories “Acute stroke”, “Hypertensive encephalopathy” and “Multiple embolic infarctions” of clinical multimodal image pairs from the Whole Brain Atlas Medical Image (WBAMI) Database are used as test images. The spatial resolution of all test images is set to  $256\times 256$ , and to make sure the registration has been realized, we take the feature-based registration algorithm for each pair, i.e. the method of complementary Harris feature point extraction based on mutual information [37], which has strong robustness and is able to adapt to various image characteristics and variations.

4.1.2. Objective evaluation metrics

Table 1. Objective assessment of different fusion methods.

Category	Metirc	Symbol	Description
Texture features preservation based metrics	Normalized mutual information	$Q_{NMI}$	It measures the mutual information of a fused image and source images.
Edge-dependent sharpness based metrics	Normalized weighted performance index	$Q^{AB/F}$	It measures the amount of edge and orientation information of the fused image using the Sobel edge detection operator.
	Overall image quality index	$Q_0$	It evaluates structural distortions of the fused image.
Comprehensive evaluation based metrics	Weighted fusion quality index	$Q_W$	It values the structure similarity by addressing coefficient correlation, illumination, and contrast.
	Structural similarity index	$Q_S$	It determines the structural similarity by taking comparisons of luminance, contrast, and structure.
	Universal image index	$Q_U$	It is designed by modeling image distortion as a combination of the loss of correlation, luminance distortion, and contrast distortion.

Since there are limitations of a single objective metric to reflect fusion result accurately, thus the six popular objective metrics, namely, Xydeas–Petrovic index [38], the structural similarity index  $Q_S$  [39], the universal image quality index  $Q_U$  [40], the overall image quality index  $Q_0$  [40], the weighted fusion quality index  $Q_W$  [41], and the mutual information index  $Q_{NMI}$  [42], are adopted to evaluate fusion performance in this letter. The higher scores of the above metrics, the better fusion result of the corresponding fusion method. The classification of these metrics are shown in table 1.

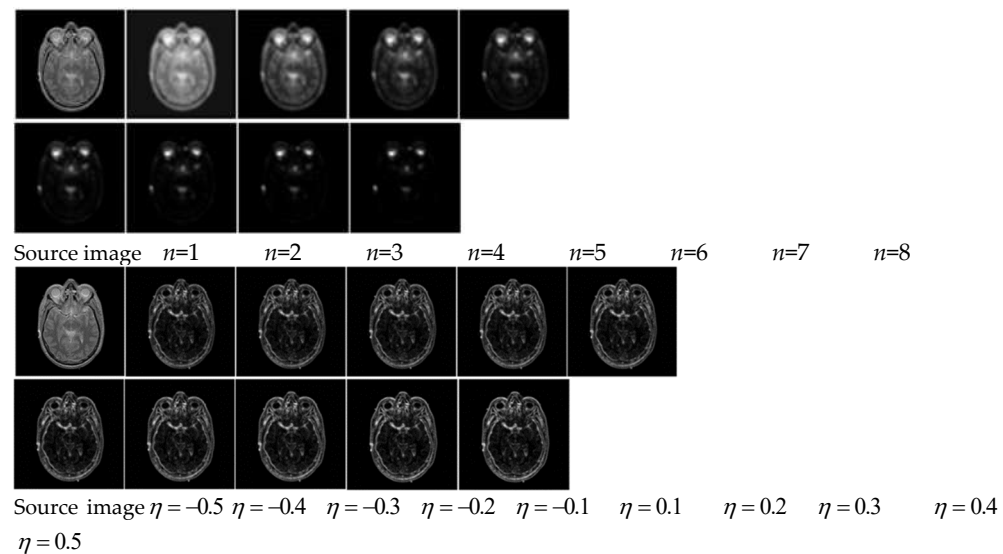
#### 4.1.3. Methods for comparison

Since it is inspired by the transform domain based method [29] to design the activity level measure and fusion scheme in our method, to make a fair and clear comparison, the conventional  $l_1$ -norm based scheme [29] and the sum of sparse salient features (SSSF) based scheme are used in the context of this letter to verify the advantages of the proposed activity level measure. Meanwhile, in each of the medical image fusion categories, some latest published representative medical image fusion methods, such as LRD [43], NSST-MSMG-PCNN [44] and CSMCA [45], are used for comparison with the proposed method. The competitors adopt the default parameters in the corresponding literatures.

#### 4.1.4. Algorithm parameters setting

For the proposed method, to obtain the pre-trained sub-dictionaries, we choose  $10^4$  patches of size  $8 \times 8$  from different uncorrupted images as training dataset, and the training patches are normalized with zero mean and unit  $L_2$ -norm, and the initial sub-dictionaries are obtained by the MATLAB function *randn* with normalized columns. Following the experimental setup of the previous work in [35], the spatial size of sliding window is set to  $8 \times 8$ , the patch-wise step size is set to 1 to keep shift invariant of SR, the two Kronecker-criterion-based separable dictionaries are set to the same size of  $8 \times 16$ , the tolerance of reconstruction error  $\varepsilon$  is set to 0.01.

In addition to the above general settings, variable  $n$  and variable  $\eta$  are the key parameters to affect the luminance contrast and the orientation contrast separately, and the parameters setting through quantitative experiments are shown in Figure 3. It can be seen that variable  $n$  will affect luminance contrast and the retention of effective information in subsequent fusion results. On the basis, we set  $n=3$  as a compromise. As increase of variable  $\eta$ , texture structure of the source image is clearer, and it is conducive for extracting orientation contrast information. On the basis, we set  $\eta = 0.5$ .



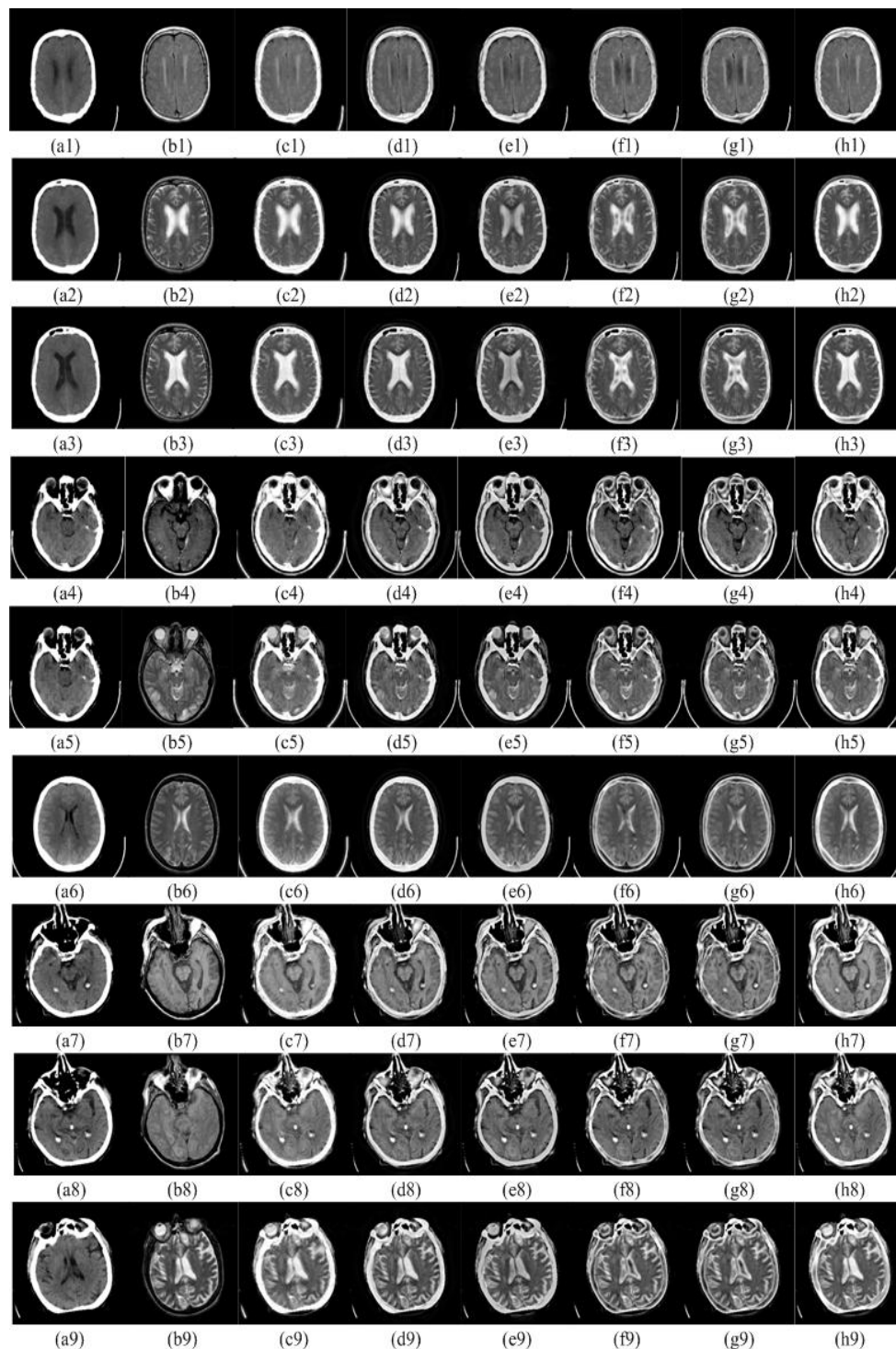
**Figure 3.** Parameters setting through quantitative experiments: The first line indicates the effect of  $n$  on luminance contrast, and the second line indicates the effect of  $\eta$  on orientation contrast.

#### 4.2. Comparison to other fusion methods

The subjective visual and objective metrics are used to evaluate the proposed method. The comparison experiment contains three categories of clinical multimodal medical images, including "Acute stroke" with 28 pairs of CT/MR-PD and CT/MR-T2, "Hypertensive encephalopathy" with 28 pairs of CT/MR-Gad and CT/MR-T2, and "Multiple embolic infarctions" with 60 pairs of CT/MR-PD, CT/MR-T1 and CT/MR-T2.



#### 4.2.1. Subjective Visual Evaluation

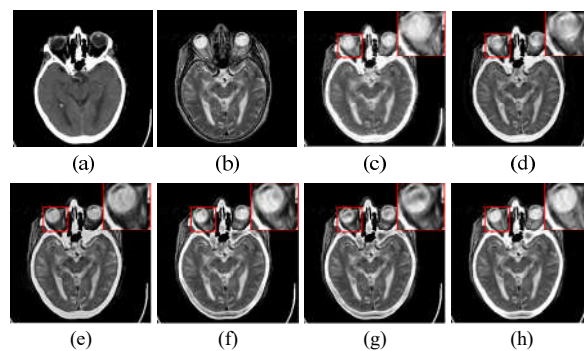


**Figure 4.** Source images and the corresponding fusion results with nine pairs of CT/MRI images: (a1), (b1) image group 1 (CT and MR-PD); (a2), (b2) image group 2 (CT and MR-T2); (a3), (b3) image group 3 (CT and MR-T2); (a4), (b4) image group 4 (CT and MR-Gad); (a5), (b5) image group 5 (CT and MR-T2); (a6), (b6) image group 6 (CT and MR-T2); (a7), (b7) image group 7 (CT and MR-T1); (a8), (b8) image group 8 (CT and MR-PD); (a9), (b9) image group 9 (CT and MR-T2); fused images (c1)-(c9) LRD-based method; fused images (d1)-(d9) NSST-MSMG-PCNN-based method; fused images (e1)-(e9) CSMCA-based method; fused images (f1)-(f9)  $l_1$ -norm-based method; fused images (g1)-(g9) SSSF-based method; and fused images (h1)-(h9) the proposed method.

In the experiments of multimodal medical image fusion, CT and MRI image fusion is the most common, since the information provided by CT and MRI images can create a good supplement, while the multimodal combination category can be expanded to other types with the fusion method of this

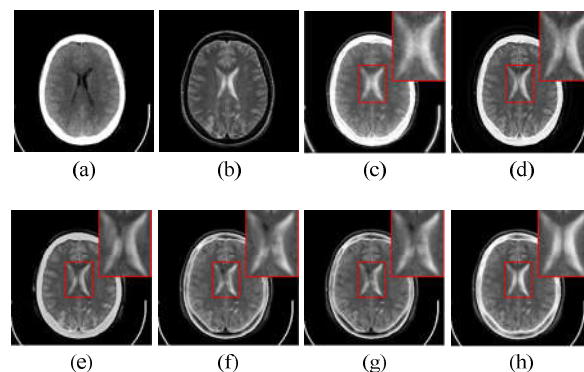
paper. Figure 4 shows the randomly selected nine groups of multimodal medical fused image in subjective visual experiments, and the first three groups belongs to “Acute stroke”, the second three groups belongs to “Hypertensive encephalopathy”, and the last three groups belongs to “Multiple embolic infarctions”. To better intuitively reflect the superiority of the proposed method, one group of typical fusion example is selected from each of the three WBAMI categories to conduct a detail analysis of the amplification of representative regions, as shown in Figure 5- Figure 7, respectively.

The CT/MR-T2 fusion results and the red box selections of the proposed method and competitors are shown in Figure 5. The fusion results of LRD and NSST-MSMG-PCNN are blurred since artificial interference is unsuppressed (see (c) and (d) in Figure 5), while CSMCA,  $l_1$ -norm, SSSF and Proposed as SR-based methods are robust to artificial interference, and the fused edges are more distinct (see (e), (f), (g) and (h) in Figure 5). However, the luminance loss of CSMCA causes the reduction of contrast (see (e) in Figure 5), and CSMCA,  $l_1$ -norm and SSSF are in the situation of partial details reduction (see (e), (f) and (g) in Figure 5). In contrast, more details from source images are extracted by the proposed method with artificial interference suppressed effectively (see (h) in Figure 5).



**Figure 5.** The CT/MR-T2 image pair from “Acute stroke” category and the corresponding fusion results with different methods: (a) and (b) are the CT image and MR-T2 image, respectively; (c) is the fusion result of LRD; (d) is the fusion result of NSST-MSMG-PCNN; (e) is the fusion result of CSMCA; (f) is the fusion result of  $l_1$ -norm; (g) is the fusion result of SSSF; (h) is the fusion result of the proposed method.

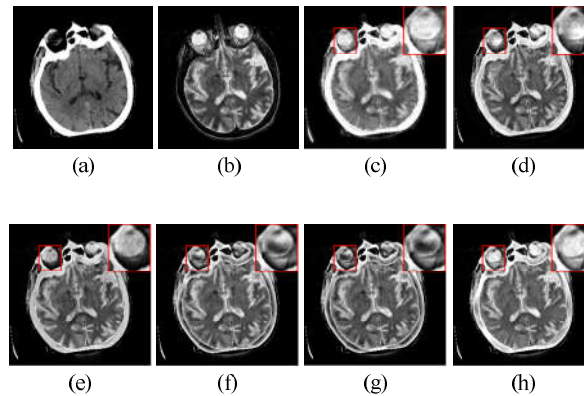
The CT/MR-T2 fusion results and the red box selections of the proposed method and competitors are shown in Figure 6. We can clear see that the results of LRD and NSST-MSMG-PCNN are disturbed by noise (see (c) and (d) in Figure 6). CSMCA,  $l_1$ -norm and SSSF lost a significant amount of structural information (see (e), (f) and (g) in Figure 6). In contrast, the proposed method performs better in structural integrity and robustness to artificial interference (see (h) in Figure 6).



**Figure 6.** The CT/MR-T2 image pair from “Hypertensive encephalopathy” category and the corresponding fusion results with different methods: (a) and (b) are the CT image and MR-T2 image, respectively; (c) is the fusion result of LRD; (d) is the fusion result of NSST-MSMG-PCNN; (e) is the fusion result of CSMCA; (f) is the fusion result of  $l_1$ -norm; (g) is the fusion result of SSSF; (h) is the fusion result of the proposed method

The CT/MR-T2 fusion results and the red box selections of the proposed method and competitors are shown in Figure 7. It is clear that artifacts appear when using LRD method (see (c) in Figure 7). NSST-MSMG-PCNN, CSMCA,  $l_1$ -norm and SSSF have lost luminance, and all of them are in the situation of partial details reduction (see (d), (e), (f) and (g) in Figure 7). In contrast, the proposed

method is obviously superior to competitors in luminance and detail information retention (see (h) in Figure 7).



**Figure 7.** The CT/MR-T2 image pair from “Multiple embolic infarctions” category and the corresponding fusion results with different methods: **(a)** and **(b)** are the CT image and MR-T2 image, respectively; **(c)** is the fusion result of LRD; **(d)** is the fusion result of NSST-MSMG-PCNN; **(e)** is the fusion result of CSMCA; **(f)** is the fusion result of  $l_1$ -norm; **(g)** is the fusion result of SSSF; **(h)** is the fusion result of the proposed method.

Through the subjective comparison experiments, it is hard to contain completed information for the SR-based image fusion with single measurement of activity level, such as the CSMCA,  $l_1$ -norm and SSSF, while the proposed method could not only retain luminance and detail information from source images, but also performs better in robustness to artificial interference to keep the fused edges more distinct. Therefore, the proposed method offers better subjective visual performance than the competitors.

#### 4.2.2. Objective Quality Evaluation

Objective quality evaluation is an important approach to evaluate fusion performance. Table 2 reports the objective assessment results of the proposed method and competitors. The average scores over all test examples from each of the three WBAMI categories are calculated, and the highest value of each row shown in bold indicates the best fusion performance. It can be seen that the proposed method performs best in all six metrics through the “Acute stroke” category with 28 pairs of multimodal medical images. In the “Hypertensive encephalopathy” category with 28 pairs of multimodal medical images, except  $Q_0$  ranking second, the other five metrics of the proposed method are the best. In the “Multiple embolic infarctions” category with 60 pairs of multimodal medical images, metric ranks second and the other five metrics of the proposed method are the best. As a whole, the average results of the six metrics of the proposed method are the best in the three clinical categories experiments. Therefore, based on the above subjective analysis and objective evaluation, the proposed method has considerable advantages over the latest published methods of LRD, NSST-MSMG-PCNN and CSMCA.

Furthermore, with no changing the fusion framework, the ablation experiments are carried out to verify the universal advantages of the proposed method over the  $l_1$ -norm- and SSSF-based schemes, which only consider the transform domain situation of activity level measure. Through the six commonly used fusion metrics,  $Q_{NMI}$  metric of the proposed method has the most obvious advantage over the two ablation competed experiments, which indicates that the proposed new activity level measure plays a significant role in the retention of texture information of source images. Furthermore, it is worth noting that the SSSF-based scheme has a slightly significant superiority than the  $l_1$ -norm-based scheme over all test examples, and this reveals the reasonability of SSSF as a substitute of  $l_1$ -norm to participate in the construction of activity level measure in transform domain.

**Table 2.** Objective assessment of different fusion methods.

WBAMI	Metrics	LRD	NSST-		$l_1$ -norm	SSSF	Proposed
			MSMG-PCNN	CSMCA			
<b>Acute stroke</b> (28 pairs of CT/MR -PD, CT/MR-T2)	$Q^{AB/F}$	0.4821	0.5187	0.5513	0.5863	0.5844	<b>0.5880</b>
	$Q_S$	0.7244	0.6972	0.7254	0.7359	0.7366	<b>0.7418</b>
	$Q_U$	0.6709	0.4628	0.5862	0.6803	0.6809	<b>0.6866</b>
	$Q_0$	0.3008	0.2984	0.3038	0.3271	0.3270	<b>0.3319</b>
	$Q_W$	0.5633	0.5791	0.5873	0.6035	0.6061	<b>0.6090</b>
	$Q_{NMI}$	0.7466	0.6693	0.7097	0.8554	0.8357	<b>0.8827</b>
<b>Hypertensive encephalopathy</b> (28 pairs of CT/MR-Gad, CT/MR-T2)	$Q^{AB/F}$	0.5062	0.5343	0.5840	0.6242	0.6248	<b>0.6290</b>
	$Q_S$	0.6974	0.6699	0.7165	0.7144	0.7163	<b>0.7211</b>
	$Q_U$	0.6283	0.4506	0.5825	0.6395	0.6413	<b>0.6474</b>
	$Q_0$	0.3152	0.3051	0.3130	0.3540	<b>0.3563</b>	0.3541
	$Q_W$	0.5984	0.6254	0.6419	0.6607	0.6671	<b>0.6736</b>
	$Q_{NMI}$	0.6883	0.6240	0.6680	0.7091	0.7040	<b>0.7464</b>
<b>Multiple embolic infarctions</b> (60 pairs of CT/MR -PD, CT/MR-T1, CT/MR-T2)	$Q^{AB/F}$	0.4584	0.5140	0.5545	<b>0.5850</b>	0.5784	0.5840
	$Q_S$	0.6893	0.6785	0.7002	0.6939	0.6952	<b>0.7016</b>
	$Q_U$	0.6146	0.4438	0.6146	0.6331	0.6343	<b>0.6412</b>
	$Q_0$	0.3211	0.3158	0.3111	0.3449	0.3458	<b>0.3488</b>
	$Q_W$	0.5562	0.5851	0.5842	0.5962	0.5977	<b>0.5994</b>
	$Q_{NMI}$	0.6951	0.6327	0.6536	0.7204	0.7095	<b>0.7575</b>

## 5. Conclusions

In this paper, a multi-modal medical image fusion method with Kronecker-criterion-based SR is proposed. The main contribution of the proposed method is summarized in three parts. Firstly, a novel activity level measure integrates spatial saliency and transform saliency is proposed to represent more abundant texture structure features. Secondly, inspired by the formation process of vision system, the spatial saliency is characterized by textural contrast that consists of luminance contrast and orientation contrast to promote more highlighted texture information to participate in fusion process. Thirdly, as a substitution of the conventional  $l_1$ -norm based sparse saliency, the sum of sparse salient features metric characterizes the transform saliency to promote more significant coefficients participate in the composition of activity level measure. Experimental results of different clinical medical image categories demonstrate the effectiveness of the proposed method. Extensive experiments demonstrate the state-of-the-art performance of the proposed method in terms of visual perception and objective assessment. Taking into account the influence of computational efficiency, some measures can be try to get a more compact and adaptive dictionary, such as taking source images as training sample and testing sample simultaneously, and some feature selection rules can be used to exclude unfeatured image patches.

**Supplementary Materials:** The Whole Brain Atlas Medical Image (WBAMI) Database information can be downloaded at: <http://www.med.harvard.edu/aanlib/home.htm>.

**Author Contributions:** All authors have read and agreed to the published version of the manuscript. The individual contributions are provided as: Conceptualization, Shaohai Hu and Qiu Hu; methodology, validation, data curation, writing—original draft preparation, writing—review and editing, Qiu Hu; resources, Qiu Hu, Weiming Cai and Shaohai Hu; investigation, Weiming Cai, Shuwen Xu and Shaohai Hu.

**Funding:** This research received no external funding.

**Data Availability Statement:** There were no new data created.

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their insightful comments and suggestions, which have greatly improved this paper. This work was supported by the Natural Science Foundation of China [Grant No. 32073028, 62172030, 62002208].

**Conflicts of Interest:** The authors declare that there is no conflict of interests regarding the publication of this paper.



## References

1. H. Hermessi, O. Murali, E. Zagrouba, Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Processing*, 2021, 183, pp. 108036.
2. S. Unar, W. Xingyuan, Z. Chuan, Visual and textual information fusion using kernel method for content based image retrieval. *Information Fusion*, 2018, 44, pp. 176-187.
3. S. Unar, X. Wang, C. Wang, Y. Wang, A decisive content based image retrieval approach for feature fusion in visual and textual images. *Knowledge-Based Systems*, 2019, 179, pp. 8-20.
4. Z. Xia, X. Wang, W. Zhou, R. Li, C. Wang, C. Zhang, Color medical image lossless watermarking using chaotic system and accurate quaternion polar harmonic transforms. *Signal Processing*, 2019, 157, pp. 108-118.
5. C. Wang, X. Wang, Z. Xia, C. Zhang, Ternary radial harmonic Fourier moments based robust stereo image zero-watermarking algorithm. *Information Sciences*, 2019, 470, pp. 109-120.
6. C. Wang, X. Wang, Z. Xia, B. Ma, Y. Shi, Image description with polar harmonic Fourier moments. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30 (12), pp. 4440-4452.
7. S. Liu, J. Zhao, M. Shi, Medical image fusion based on improved sum-modified-laplacian. *Int J Imaging Syst Technol*, 2015, 25 (3), pp. 206-212.
8. S. Liu, J. Zhao, M. Shi, Medical image fusion based on rolling guidance filter and spiking cortical model. *Comput Math Methods Med*, 2015, 2015, pp. 06-214.
9. S. Sneha, R. S. Anand, Ripplet domain fusion approach for CT and MR medical image information. *Biomedical Signal Processing and Control*, 2018, 46, pp. 281-292.
10. S. N. Talbar, S. S. Chavan, A. Pawar, Non-subsampled complex wavelet transform based medical image fusion. *Proceedings of the Future Technologies Conference*, 2018, pp. 548-556.
11. W. Zhao, H. Yang, J. Wang, X. Pan, Z. Cao, Region- and pixel-level multi-focus image fusion through convolutional neural networks. *Mobile Networks and Applications*, 2021, 26(4), pp. 40-56.
12. X. Jin, R. Nie, X. Zhang, D. Zhou, K. He, Multi-focus image fusion combining focus-region-level partition and pulse-coupled neural network. *Soft Computing*, 2019, 23 (13), pp. 4685-4699.
13. X. Qiu, M. Li, L. Zhang, X. Yuan, Guided filter-based multi-focus image fusion through focus region detection. *Signal Processing: Image Communication*, 2019, 72 (12), pp. 35-46.
14. H. Li, K. Ma, H. Yong, L. Zhang, Fast multi-scale structural patch decomposition for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 2020, 29, pp. 5805-5816.
15. K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, L. Zhang, Robust multi-exposure image fusion: a structural patch decomposition approach. *IEEE Transactions on Image Processing*, 2017, 26(5), pp. 2519-2532.
16. H. Li, T. N. Chan, X. Qi, W. Xie, Detail-preserving multi-exposure fusion with edge-preserving structural patch decomposition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 99, pp. 1-1.
17. R. R. Nair, T. Singh, MAMIF: multimodal adaptive medical image fusion based on B-spline registration and non-subsampled shearlet transform. *Multimedia Tools and Applications*, 2021, 80 (4), pp. 19079-19105.
18. W. Kong, Q. Miao, Y. Lei, Multimodal sensor medical image fusion based on local difference in non-subsampled domain. *IEEE Transactions on Instrumentation and Measurement*, 2019, 68 (4), pp. 938-951.
19. K. Padmavathi, M. V. Karki, M. Bhat, Medical image fusion of different modalities using dual tree complex wavelet transform with PCA. *International Conference on Circuits, Controls, Communications and Computing*, 2016, pp. 1-5.
20. X. Xi, X. Luo, Z. Zhang, Q. You, X. Wu, Multimodal medical volumetric image fusion based on multi-feature in 3-D shearlet transform. *International Smart Cities Conference*, 2017, pp. 1-6.
21. F. Shabanzade, H. Ghassemian, Multimodal image fusion via sparse representation and clustering-based dictionary learning algorithm in nonsubsampling contourlet domain. *8th International Symposium on Telecommunications*, 2016, pp. 472-477.
22. J. Xia, Y. Chen, A. Chen, Y. Chen, Medical image fusion based on sparse representation and pcnn in nsct domain. *Comput. Math. Methods Med.*, 2018, 5(2018), pp. 1-12.
23. M. Yin, X. N. Liu, Y. Liu, X. Chen, Medical image fusion with parameter-adaptive pulse coupled-neural network in nonsubsampling shearlet transform domain. *IEEE Trans. Instrum. Meas.*, 2018, 68(1), pp. 49-64.
24. P. H. Dinh, A novel approach based on three-scale image decomposition and marine predators algorithm for multi-modal medical image fusion. *Biomedical Signal Processing and Control*, 2021, 67(2), pp. 1-14.
25. C. Pei, K. Fan, W. Wang, Two-scale multimodal medical image fusion based on guided filtering and sparse representation. *IEEE Access*, 2020, 8, pp. 140216-140233.
26. H. R. Shahdoosti, A. Mehrabi, Multimodal image fusion using sparse representation classification in tetralet domain, *Digital Signal Processing* 79 (2018) 9-22.
27. T. Ling, X. Yu, Medical image fusion based on fast finite shearlet transform and sparse representation. *Computational and Mathematical Methods in Medicine*, 2019, 2019, pp. 3503267.
28. M. Elad, I. Yavneh, A plurality of sparse representations is better than the sparsest one alone. *IEEE Transactions on Information Theory*, 2009, 55(10), pp. 4701-4714.

29. Y. Liu, X. Chen, K. Ward Rabab, Z. Wang, Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*, 2016, 23(12), pp. 1882–1886.
30. M. Ghassemi, Z. Shakeri, A. D. Sarwate, W. U. Bajwa, Learning mixtures of separable dictionaries for tensor data: analysis and algorithms. *IEEE Transactions on Signal Processing*, 2019, 68, pp. 33–48.
31. B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, 381(6583), pp. 607–609.
32. B. L. Sturm, M. G. Christensen, Cyclic matching pursuits with multiscale time-frequency dictionaries. *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, 2011.
33. K. Schnass, Average performance of orthogonal matching pursuit (OMP) for sparse approximation. *IEEE Signal Processing Letters*, 2018, 26(10), pp. 1566–1567.
34. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 2010, 3(1), pp. 1–122.
35. Q. Hu, S. Hu, F. Zhang, Multi-modality medical image fusion based on separable dictionary learning and Gabor filtering. *Signal Processing: Image Communication*, 2020, 83, pp. 115758.
36. X. Hou, L. Zhang, Saliency detection: a spectral residual approach. *IEEE Conference on Computer Vision & Pattern Recognition*, 2007, pp. 1–8.
37. S. Pan, J. Zhang, X. Liu, X. Guo. Complementary Harris feature point extraction method based on mutual information. *Signal Processing*, 2017, 130, pp. 132–139.
38. V. Petrović, Subjective tests for image fusion evaluation and objective metric validation. *Information Fusion*, 2007, 8(2), pp. 208–216.
39. C. Yang, J. Zhang, X. Wang, X. Liu, A novel similarity based quality metric for image fusion. *Inf. Fusion*, 2008, 9(2), pp. 156–160.
40. Z. Wang, A. C. Bovik, A universal image quality index. *IEEE Signal Processing Letters*, 2002, 9(3), pp. 81–84.
41. G. Piella, H. Heijmans, A new quality metric for image fusion. *International Conference on Image Processing*, 2003, 3, pp. 173–176.
42. G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion. *Electronics Letters*, 2002, 38(7), pp. 313–315.
43. X. Li, X. Guo, P. Han, X. Wang, T. Luo, Laplacian re-decomposition for multimodal medical image fusion. *IEEE Transactions on Instrumentation and Measurement*, 2020, 99, pp. 1–1.
44. W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, A. K. Jaiswal, Multimodal medical image fusion algorithm in the era of big data. *Neural Computing and Applications*, 2020, 3, pp. 1–21.
45. Y. Liu, X. Chen, R. Ward, Z. J. Wang, Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Processing Letters*, 2019, 26(3), pp. 485–489.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.