electronics

Article

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

Qiu Hu [1], Weiming Cai [1,*], Shuwen Xu [2] and Shaohai Hu [3,4]

[1] School of Information Science and Engineering, NingboTech University, Ningbo 315100, China; hu_qiu@nbt.edu.cn
[2] Third Research Institute of China Electronics Technology Group Corporation, Beijing 100846, China
[3] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China; shhu@bjtu.edu.cn
[4] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China
* Correspondence: caiwm@nit.zju.edu.cn; Tel.: 0574-88130027

Abstract: Multimodal medical image fusion is a fundamental, but challenging, problem in the fields of brain science research and brain disease diagnosis, as it is challenging for sparse representation (SR)-based fusion to characterize activity levels with a single measurement and not lose effective information. In this study, the Kronecker-criterion-based SR framework was applied for medical image fusion with a patch-based activity level, integrating salient features of multiple domains. Inspired by the formation process of vision systems, the spatial saliency was characterized by textural contrast (*TC*), composed of luminance and orientation contrasts, to promote the participation of more highlighted textural information in the fusion process. As a substitute for the conventional l_1-norm-based sparse saliency, the sum of sparse salient features (*SSSF*) was used as a metric for promoting the participation of more significant coefficients in the composition of the activity level measurement. The designed activity level measurement was verified to be more conducive to maintaining the integrity and sharpness of detailed information. Various experiments on multiple groups of clinical medical images verified the effectiveness of the proposed fusion method in terms of both visual quality and objective assessment. Furthermore, this study will be helpful for the further detection and segmentation of medical images.

Keywords: multimodality medical image; image fusion; sparse representation (SR); Kronecker criterion; activity level measure

check for updates

Citation: Hu, Q.; Cai, W.; Xu, S.; Hu, S. MRI Image Fusion Based on Sparse Representation with Measurement of Patch-Based Multiple Salient Features. *Electronics* **2023**, *12*, 3058. https://doi.org/10.3390/electronics12143058

Academic Editor: Heung-Il Suk

Received: 9 June 2023
Revised: 3 July 2023
Accepted: 5 July 2023
Published: 12 July 2023

1. Introduction

Over the past several decades, a variety of information processing technologies have led to major achievements in clinical diagnosis research [1], such as image classification [2,3], image fusion [4], and image segmentation [5]. The main purpose of medical image fusion is to combine the complementary information from various sensors to construct a new image with which to assist medical experts with diagnosis. Despite the simplicity of the idea, there are 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 regarding dense tissues, but lacks soft tissue information. In contrast, magnetic resonance imaging (MRI) is more suitable for soft tissues, but is short on dense tissue information. More crucially, single imaging tends to be ineffective at characterizing the symptoms of different diseases.

To overcome these challenges, a variety of image fusion methods have been proposed. The content of the image can be either visual (i.e., color, shape, or texture) or textual (i.e., to identify datasets appearing within an image). Some new advances in the fusion field consider these two aspects simultaneously [6,7]. To further improve fusion performance, some new features, such as different image moments [8–10], have also been used in image fusion.

The mainstream directions of image fusion mainly focus on the visual content, including the spatial domain [11,12] and transform domain [13,14]. The former usually addresses the fusion issue via image blocks or pixel-wise gradient information for multi-focus fusion [15–17] and multi-exposure fusion [18–20] tasks. The latter merges the transform coefficients relevant to source images with different reconstruction algorithms to obtain a fused image, which is recognized as being effective for multimodal image fusion [21,22]. Multi-scale transform (MST)-based medical image fusion is a mainstream research direction. Dual-tree complex wavelet transform (DTCWT) [23], non-subsampled shearlet transform (NSST) [24], and non-subsampled contourlet transform (NSCT) [25] are conventional MST methods for image fusion. In recent years, some novel MST-based methods have been proposed. Xia combined sparse representation with a pulse-coupled neural network (PCNN) in the NSCT domain for medical image fusion [26]. Yin proposed a parameter-adaptive pulse-coupled neural network as part of an NSST domain (NSST-PAPCNN)-based medical image fusion strategy [27]. Dinh proposed a Kirsch compass operator with a marine-predator-algorithm-based method for medical image fusion [28].

Differently from MST, the principle of SR is more in accordance with the human visual system (HVS), and compared to the MST-based methods, SR-based methods have two main distinctions. For the first distinction, the fixed basis limits the MST-based methods to expressing significant features, while the SR-based methods are flexible and can procure more intrinsic features by means of dictionary learning. For the second distinction, the MST-based methods are sensitive to noise and misregistration with large decomposition level settings, while the SR-based methods with overlapping patch-wise modes are robust for misregistration, which guarantees the accuracy of the spatial location of tissues. Therefore, a wide range of research on SR-based medical image fusion has been conducted in recent years [29–31].

However, there are still drawbacks to these SR-based methods. Firstly, they may be insufficient to handle fine details due to an over-complete dictionary, and this highly redundant dictionary will lead to visual artifacts in the fused result [32]. Secondly, the dictionary atoms are updated in column vector form, resulting in the loss of correlation and structural information. In addition to the drawbacks of SR itself, the fusion weight accuracy will inevitably be reduced with an unreasonable fusion strategy of coefficients. One issue relates to the activity level measure, which helps to recognize distinct features in the fusion process, and another issue concentrates on the integration of coefficients into the counterparts of the fused image. For the former issue, the l_1-norm mode is a conventional solution to describe the detailed information contained in sparse vectors [33]; however, the solution is insufficient to express the sparse saliency well, since detailed information that characterizes the activity level with the 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 a single measurement of the activity level, thereby further leading to the loss of detailed information. For this issue, it may reduce the contrast of the 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 at the cost of minor information loss.

Based on the above discussion, we adopted a promising signal decomposition model, known as Kronecker-criterion-based SR [34], to solve the medical image fusion problem. The main contributions of this work are illustrated as follows:

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

ficients to measure the composite activity level through the sum of differences in the adjacent areas.

The rest of this paper is organized as follows: Section 2 provides a brief description of conventional sparse representation theory and the Kronecker-criterion-based SR, i.e., the separable dictionary learning algorithm. The detailed fusion scheme is described in Section 3. The experimental results and a 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 [35]. Given $y \in R^m$ in the 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 consisting of the fidelity term and penalty term is defined as

$$\underset{D,\alpha}{\operatorname{argmin}} \frac{1}{2}\|y - D\alpha\|_2^2 + \lambda R(\alpha) \tag{1}$$

where α represents the sparse vector, $\|\bullet\|_2^2$ represents the l_2-norm, and λ represents the regularization parameter of the penalty term $R(\alpha)$. SR can be roughly divided into two categories, the greedy scheme (e.g., matching pursuit (MP) [36] and orthogonal matching pursuit (OMP) [37]) with $R(\alpha) = \|\alpha\|_0$ and the convex optimization scheme (e.g., alternating direction method of multipliers (ADMM) [38]) 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, its ability to handle fine details with an over-complete dictionary may be insufficient, since atoms (i.e., vectors) of the pre-trained over-complete dictionary are updated one-by-one with either the method of optimal directions (MOD) or k-singular value decomposition (K-SVD). This can be understood as extracting image textural information from only a one-dimensional direction; this breaks the potential correlations within the image, thus causing the obtained pre-trained dictionary to become unstructured. Meanwhile, the highly redundant dictionary is sensitive to random noise and may cause visual artifacts to appear. Therefore, there is a deviation between the 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 [34]. On the premise of ensuring the quality of image reconstruction, the penalty in l_0-norm with $R(\alpha) = \|\alpha\|_0$ and the corresponding objective function of separable dictionary learning [39] is defined as

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

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

The steps of the separable dictionary-learning algorithm include sparse coding and dictionary update. The dictionary optimization problems were found using the extensional two-dimensional OMP (2D-OMP) greedy algorithm and the ASeDiL (analytic separable dictionary learning) algorithm to obtain the sparse coefficients and the pre-trained sub-dictionaries $\{D_A, D_B\}$, respectively, using the method described by [40]. The dictionary pre-training model is shown in Figure 1.

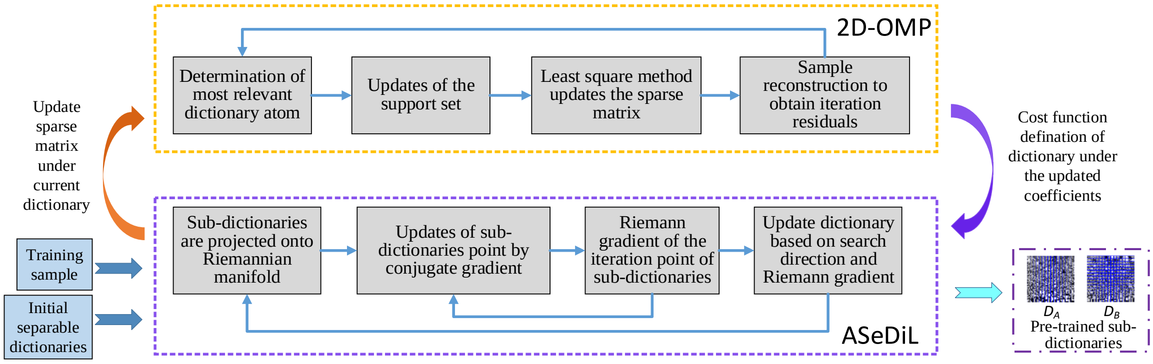

Figure 1. Dictionary pre-training model on Riemannian manifold.

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

$$\underset{S}{\text{argmin}}\|S\|_0 \text{ such that } \|D_A S D_B^T - Y\|_2^2 < \varepsilon \tag{3}$$

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

Combining the constraints with the l_2-norm of the dictionary atoms equaling 1, and with no correlation of atoms in the dictionary, the log function was employed to fit the full rank and the column irrelevance of sub-dictionaries. Then, the objective function of the dictionary update was written as

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

where ω and ψ represent the fitting parameters, and $h(D_A)$, $p(D_A)$, $h(D_B)$, $p(D_B)$ are defined as

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

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

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

With the aforementioned separable structure, the obtained sparse matrix composed of correlation coefficients becomes able characterize more textural and structural information. This can not only increase the 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 the source images.

3. Proposed Fusion Method

The framework of the proposed method is shown in Figure 2. Supposing that there are K pre-registered source images denoted by $I_k, k \in \{1, 2, \ldots, K\}$, the *r-th* overlapping image patch of the *k-th* source image I_k^r was obtained through the sliding window technique, and the corresponding sparse coefficient map S_k^r was learned through the 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 consisted of the following two steps:

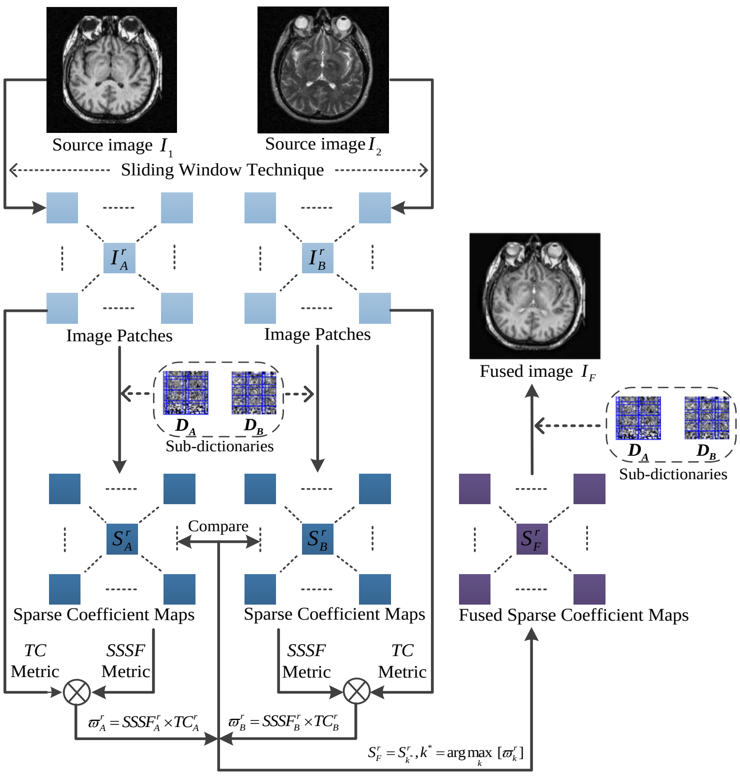

Figure 2. Overall framework of the proposed method.

3.1. *The Measurement of Activity Level for Fusion*

3.1.1. Spatial Saliency by Textural Contrast

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

First of all, the luminance contrast was defined by considering the distinctiveness of the intensity attributes between each pixel and the corresponding image patch [41]. To increase the useful dynamic ranges and to suppress high contrast effectively in the background, the *n-th* order statistic was applied as

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

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

Along with luminance contrast, the local image structure was captured by orientation contrast through a weighted structure tensor. It is worth noting that we focused on weighted gradient information rather than the gradient itself, and this highlights the main features of the source images. The weighted structure tensor [42] was able to effectively summarize the dominant orientation and the energy along this direction based on the weighted gradient field:

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

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

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

where $LSM_k^r(x,y)$ represents the local salient metric, which reflects the importance of the pixel (x,y) by computing the sum of intensity around it, and is calculated by

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

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

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

The orientation contrast related to the eigenvalues of β_1 and β_2 of this matrix is calculated as

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

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 the salient area contains luminance contrast and orientation contrast, as mentioned, we then attempted to define the texture contrast with two parts:

$$TC_k^r(x,y) = LC_k^r(x,y) \times OC_k^r(x,y) \tag{13}$$

Here, each part was smoothed by Gaussian filtering, as in [43], and TC_k^r was 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, the *SSSF* metric [44] is able to highlight more significant coefficients to participate in the composition of activity level measures through the sum of differences in adjacent areas, which is defined as

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

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

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

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

3.2. Fusion Scheme

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

$$\varpi_k^r(x,y) = SSSF_k^r(x,y) \times TC_k^r(x,y) \tag{16}$$

where ϖ_k^r is the measurement result of the source image patch I_k^r. Then, the maximum weighted activity level measure was used to achieve the fused coefficient map:

$$S_F^r(x,y) = S_{k^*}^r(x,y), k^* = \underset{k}{\operatorname{argmax}}[\varpi_k^r(x,y)] \tag{17}$$

Then, the fusion result was obtained by sparse reconstruction as

$$I_F^r = D_A S_F^r D_B^T \tag{18}$$

The final fused image I_F was constituted by the overall fused image patches.

4. Experiments

4.1. Experimental Setting

4.1.1. Source Images

In our experiments, three categories of "Acute stroke", "Hypertensive encephalopathy", and "Multiple embolic infarctions" from clinical multimodal image pairs in the Whole Brain Atlas Medical Image (WBAMI) database were used as test images, which can be downloaded at: http://www.med.harvard.edu/aanlib/home.htm, accessed on 8 June 2023. The database covers a variety of modal combination types, and it is widely applied for the verification of medical image fusion performance [22–24]. The spatial resolution was set to 256×256 for all test images. To make sure the registration was realized, we took the feature-based registration algorithm for each pair, i.e., the method of complementary Harris feature point extraction based on mutual information [45], which has strong robustness and is able to adapt to various image characteristics and variations.

4.1.2. Objective Evaluation Metrics

Since there are limitations of a single objective metric in terms of reflecting the fusion result accurately, the six popular objective metrics, namely, the Xydeas–Petrovic index [46], the structural similarity index Q_S [47], the universal image quality index Q_U [48], the overall image quality index Q_0 [48], the weighted fusion quality index Q_W [49], and the mutual information index Q_{NMI} [50], were adopted to evaluate the fusion performance in this study. The higher the scores of the above metrics, the better the fusion result of the corresponding fusion method. A classification of these metrics is shown in Table 1.

Table 1. Classification of different objective assessment metrics.

Category	Metric	Symbol	Description
Textural-feature-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 in the fused image.
Comprehensive-evaluation-based metrics	Weighted fusion quality index	Q_W	It values the structural similarity by addressing the 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.

4.1.3. Methods for Comparison

Since it was inspired by the transform-domain-based method [33] to design the activity level measure and fusion scheme in our method, to carry out a fair and clear comparison, the conventional l_1-norm-based scheme [33] and the sum of sparse salient features (*SSSF*)-based scheme were used to verify the advantages of the proposed activity level measurement. Meanwhile, in each of the medical image fusion categories, some newly published representative medical image fusion methods, such as LRD [51], NSST-MSMG-PCNN [52], and CSMCA [53], were used for comparison with the proposed method. The competitors adopted the default parameters in the corresponding literature.

4.1.4. Algorithm Parameter Setting

For the proposed method, to obtain the pre-trained sub-dictionaries, we chose 10^4 patches with sizes of 8×8 from different uncorrupted images to be included in the training dataset. Furthermore, the training patches were normalized with a zero mean and unit l_2-norm, and the initial sub-dictionaries were obtained by the MATLAB function *randn* with normalized columns. Following the experimental setup detailed in [40], the spatial size of the sliding window was set to 8×8, the patch-wise step size was set to 1 to keep the shift invariant of SR, the two Kronecker-criterion-based separable dictionaries were set to the same size of 8×16, and the tolerance of the reconstruction error ε was set to 0.01.

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

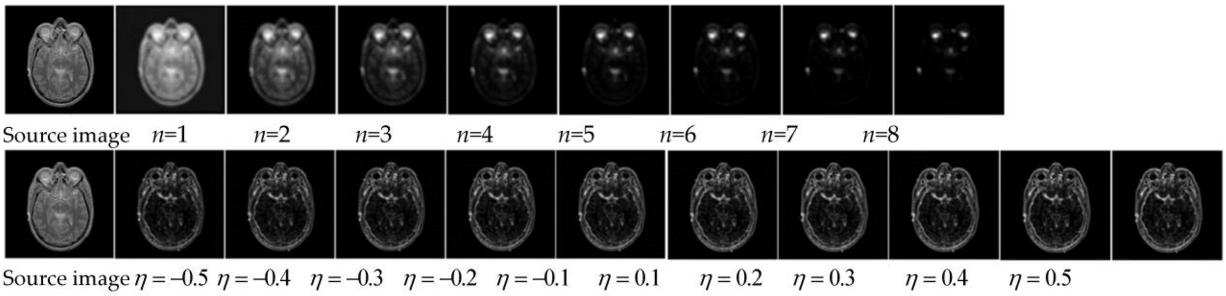

Figure 3. Parameter setting through quantitative experiments: The first line indicates the effect of n on luminance contrast, and the second line indicates the effect of η on orientation contrast.

4.2. Comparison to Other Fusion Methods

The subjective visual and objective metrics were used to evaluate the proposed method. The comparison experiment contained 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

In the experiments using multimodal medical image fusion, CT and MRI image fusion were the most common. This is because the information provided by CT and MRI images can act as a good supplement, while the multimodal combination category can be expanded to other types of fusion, such as the method used in this paper. Figure 4 shows the nine randomly selected groups of multimodal-fused medical images in subjective visual experiments. The first three groups belong to "Acute stroke", the second three groups belong to "Hypertensive encephalopathy", and the last three groups belong to "Multiple embolic infarctions". To more intuitively reflect the superiority of the proposed method, one group showing typical fusion was selected from each of the three WBAMI categories as an example with which to conduct a detailed analysis on the amplification of representative regions, as shown in Figures 5–7.

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

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

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 appeared when using the LRD method (see in Figure 7c). NSST-MSMG-PCNN, CSMCA, l_1-norm, and *SSSF* caused losses of luminance, and all led to partial reductions in detail (see in Figure 7d–g). In contrast, the proposed method was obviously superior to its competitors in terms of luminance and detail retention (see in Figure 7h).

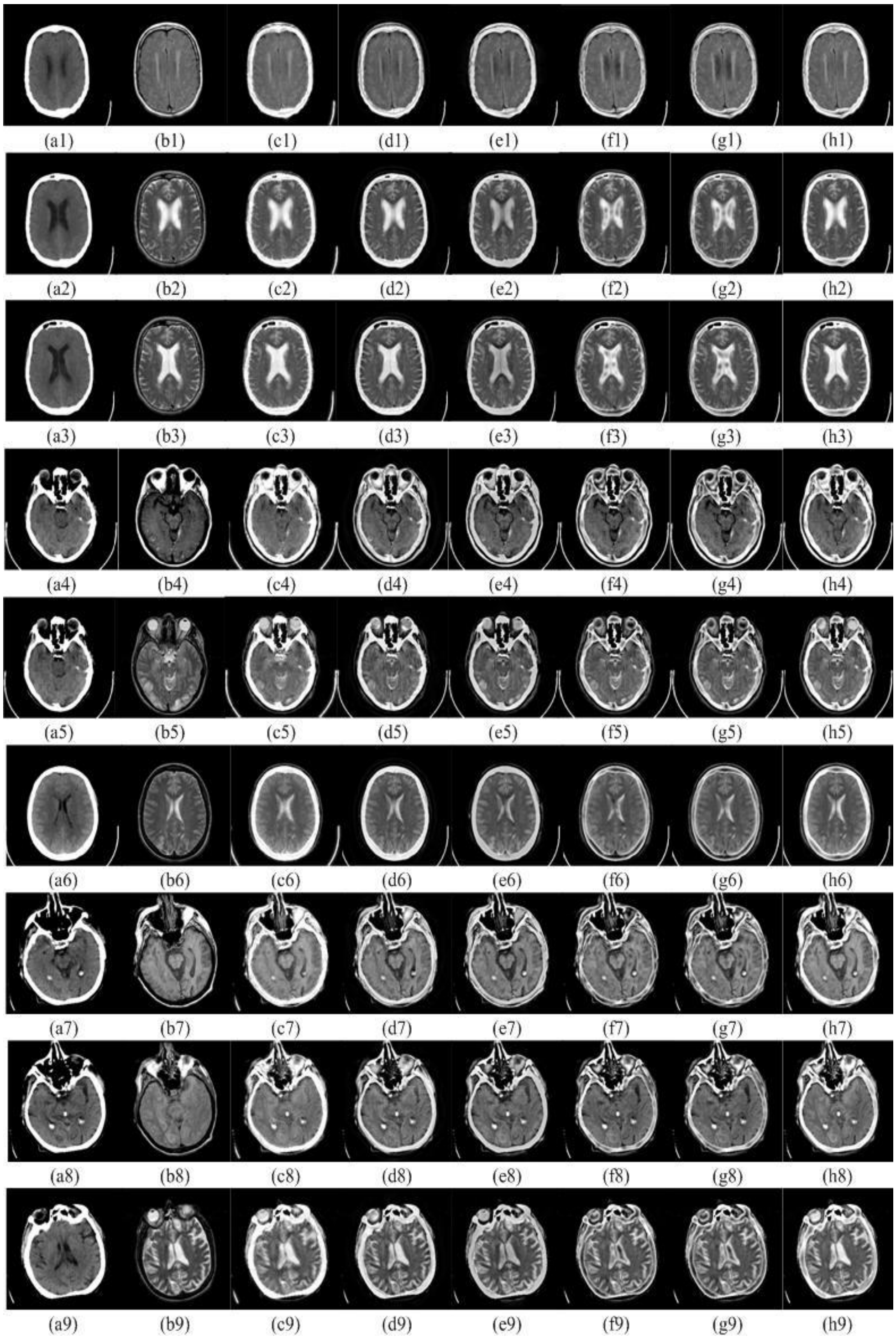

Figure 4. Source images and the corresponding fusion results with nine pairs of CT/MRI images: (**a1**,**b1**) image group 1 (CT and MR-PD); (**a2**,**b2**) image group 2 (CT and MR-T2); (**a3**,**b3**) image group 3 (CT and MR-T2); (**a4**,**b4**) image group 4 (CT and MR-Gad); (**a5**,**b5**) image 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**) of the LRD-based method; fused images (**d1**–**d9**) of the NSST-MSMG-PCNN-based method; fused images (**e1**–**e9**) of the CSMCA-based method; fused images (**f1**–**f9**) of the l_1-norm-based method; fused images (**g1**–**g9**) of the *SSSF*-based method; and fused images (**h1**–**h9**) using the proposed method.

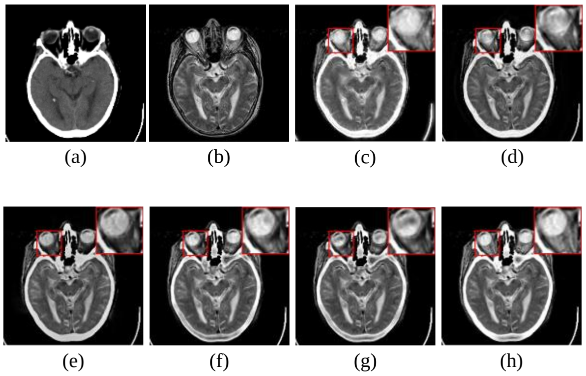

Figure 5. The CT/MR-T2 image pair from the "Acute stroke" category and the corresponding fusion results achieved using different methods: (**a**,**b**) represent the CT image and the MR-T2 image, respectively; (**c**) is the fusion result of LRD; (**d**) is the fusion result of NSST-MSMG-PCNN; (**e**)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

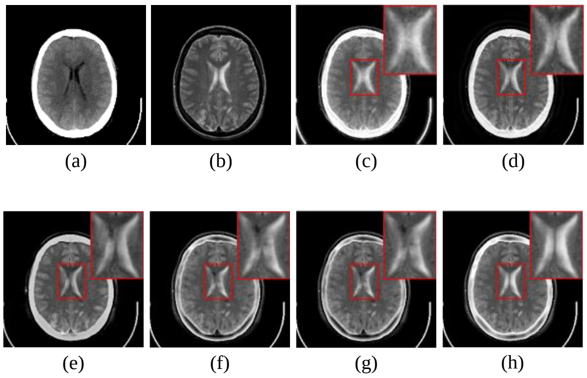

Figure 6. The CT/MR-T2 image pair from the "Hypertensive encephalopathy" category and the corresponding fusion results with different methods: (**a**,**b**) are the CT image and MR-T2 image, respectively; (**c**) is the fusion result of LRD; (**d**) is the fusion result of NSST-MSMG-PCNN; (**e**) is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

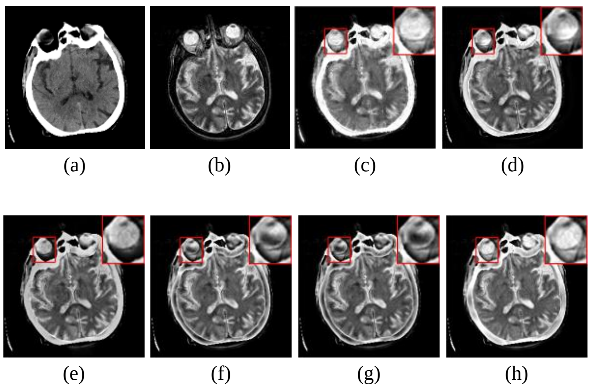

Figure 7. The CT/MR-T2 image pair from the "Multiple embolic infarctions" category and the corresponding fusion results with different methods: (**a**,**b**) are the CT image and MR-T2 images, respectively; (**c**) is the fusion result of LRD; (**d**) is the fusion result of NSST-MSMG-PCNN; (**e**) is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

Through these subjective comparison experiments, it was difficult to contain the completed information for SR-based image fusion with a single measurement of the activity level, such as the CSMCA, l_1-norm, or SSSF. Moreover, the proposed method was not only able to retain luminance and detail information from the source images, but also performed better in terms of robustness to artificial interference, keeping the fused edges more distinct. Therefore, the proposed method offered better subjective visual performance than the competitors.

4.2.2. Objective Quality Evaluation

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

Table 2. Objective assessment of different fusion methods.

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

Furthermore, without changing the fusion framework, the ablation experiments were carried out to verify the universal advantages of the proposed method over the l_1-norm- and SSSF-based schemes, which only considered the transform domain situation of the activity level measurement. Through the six commonly used fusion metrics, the Q_{NMI} metric of the proposed method had the most obvious advantage over the two experiments in which ablation was competed. This indicates that the proposed new activity level measure plays a significant role in the retention of the textural information of the source images. Furthermore, it is worth noting that the *SSSF*-based scheme showed slightly significant superiority over the l_1-norm-based scheme in all test examples; this reveals the justification for using SSSF as a substitute for l_1-norm in order to participate in the construction of the activity level measure in the transform domain.

5. Conclusions

In this study, a multi-modal medical image fusion method with Kronecker-criterion-based SR was proposed. The main contribution of the proposed method is summarized in three parts. Firstly, a novel activity level measure integrating spatial saliency and transform saliency was proposed to represent more abundant textural structure features. Secondly, inspired by the formation process of the vision system, the spatial saliency was characterized by the textural contrast consisting of luminance contrast and orientation contrast to induce more highlighted textural information to participate in the fusion process. Thirdly, as a substitution for conventional l_1-norm-based sparse saliency, the sum of the sparse salient features metric characterizes the transform saliency to promote more significant coefficients and to participate in the composition of the activity level measure. The experimental results of different clinical medical image categories demonstrated the effectiveness of the proposed method. Extensive experiments have demonstrated 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 attempt to obtain a more compact and adaptive dictionary, such as by taking source images as the training sample and testing samples simultaneously, and some feature selection rules can be used to exclude unfeatured image patches.

Author Contributions: The individual contributions are provided as: Conceptualization, S.H. and Q.H.; methodology, validation, data curation, writing—original draft preparation, writing—review and editing, Q.H.; resources, Q.H., W.C. and S.H.; investigation, W.C., S.X. and S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of China [Grant No. 32073028, 62172030, 62002208], the Ningbo Youth Science and Technology Innovation Leading Talent Project [Grant No. 2023QL004].

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.

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

References

1. Ahmad, S.; Khan, S.; Ajmi, M.F.A.; Dutta, A.K.; Dang, L.M.; Joshi, G.P.; Moon, H. Deep learning enabled disease diagnosis for decure internet of medical things. *Comput. Mater. Contin.* **2022**, *73*, 965–979.
2. Haq, A.; Li, J.; Agbley, B.; Khan, A.; Khan, I.; Uddin, M.; Khan, S. IIMFCBM: Intelligent integrated model for feature extraction and classification of brain tumors using MRI clinical imaging data in IoT-Healthcare. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5004–5012. [CrossRef] [PubMed]
3. Haq, A.; Li, J.; Khan, S.; Alshara, M.; Alotaibi, R.; Mawuli, C. DACBT: Deep learning approach for classification of brain tumors using MRI data in IoT healthcare environment. *Sci. Rep.* **2022**, *12*, 15331. [CrossRef] [PubMed]
4. Hermessi, H.; Mourali, O.; Zagrouba, E. Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Process.* **2021**, *183*, 108036. [CrossRef]

5. Yousef, R.; Khan, S.; Gupta, G.; Siddiqui, T.; Albahlal, B.; Alajlan, S.; Haq, M. U-net-based models towards optimal MR brain image segmentation. *Diagnostics* **2023**, *13*, 1624. [CrossRef]
6. Unar, S.; Xingyuan, W.; Chuan, Z. Visual and textual information fusion using kernel method for content based image retrieval. *Inf. Fusion* **2018**, *44*, 176–187. [CrossRef]
7. Unar, S.; Wang, X.; Wang, C.; Wang, Y. A decisive content based image retrieval approach for feature fusion in visual and textual images. *Knowl. Based Syst.* **2019**, *179*, 8–20. [CrossRef]
8. Xia, Z.; Wang, X.; Zhou, W.; Li, R.; Wang, C.; Zhang, C. Color medical image lossless watermarking using chaotic system and accurate quaternion polar harmonic transforms. *Signal Process.* **2019**, *157*, 108–118. [CrossRef]
9. Wang, C.; Wang, X.; Xia, Z.; Zhang, C. Ternary radial harmonic Fourier moments based robust stereo image zero-watermarking algorithm. *Inf. Sci.* **2019**, *470*, 109–120. [CrossRef]
10. Wang, C.; Wang, X.; Xia, Z.; Ma, B.; Shi, Y. Image description with polar harmonic Fourier moments. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 4440–4452. [CrossRef]
11. Liu, S.; Zhao, J.; Shi, M. Medical image fusion based on improved sum-modified-laplacian. *Int. J. Imaging Syst. Technol.* **2015**, *25*, 206–212. [CrossRef]
12. Liu, S.; Zhao, J.; Shi, M. Medical image fusion based on rolling guidance filter and spiking cortical model. *Comput. Math. Methods Med.* **2015**, *2015*, 156043.
13. Sneha, S.; Anand, R.S. Ripplet domain fusion approach for CT and MR medical image information. *Biomed. Signal Process. Control* **2018**, *46*, 281–292.
14. Talbar, S.N.; Chavan, S.S.; Pawar, A. Non-subsampled complex wavelet transform based medical image fusion. In Proceedings of the Future Technologies Conference, Vancouver, BC, Canada, 15–16 November 2018; pp. 548–556.
15. Zhao, W.; Yang, H.; Wang, J.; Pan, X.; Cao, Z. Region- and pixel-level multi-focus image fusion through convolutional neural networks. *Mob. Netw. Appl.* **2021**, *26*, 40–56. [CrossRef]
16. Jin, X.; Nie, R.; Zhang, X.; He, D.Z.K. Multi-focus image fusion combining focus-region-level partition and pulse-coupled neural network. *Soft Comput.* **2019**, *23*, 4685–4699.
17. Qiu, X.; Li, M.; Yuan, L.Z.X. Guided filter-based multi-focus image fusion through focus region detection. *Signal Process. Image Commun.* **2019**, *72*, 35–46. [CrossRef]
18. Li, H.; Ma, K.; Yong, H.; Zhang, L. Fast multi-scale structural patch decomposition for multi-exposure image fusion. *IEEE Trans. Image Process.* **2020**, *29*, 5805–5816. [CrossRef]
19. Ma, K.; Li, H.; Yong, H.; Wang, Z.; Meng, D.; Zhang, L. Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Trans. Image Process.* **2017**, *26*, 2519–2532. [CrossRef]
20. Li, H.; Chan, T.N.; Qi, X.; Xie, W. Detail-preserving multi-exposure fusion with edge-preserving structural patch decomposition. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *99*, 4293–4304. [CrossRef]
21. Nair, R.R.; Singh, T. MAMIF: Multimodal adaptive medical image fusion based on B-spline registration and non-subsampled shearlet transform. *Multimed. Tools Appl.* **2021**, *80*, 19079–19105. [CrossRef]
22. Kong, W.; Miao, Q.; Lei, Y. Multimodal sensor medical image fusion based on local difference in non-subsampled domain. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 938–951. [CrossRef]
23. Padmavathi, K.; Karki, M.V.; Bhat, M. Medical image fusion of different modalities using dual tree complex wavelet transform with PCA. In Proceedings of the International Conference on Circuits, Controls, Communications and Computing, Bangalore, India, 4–6 October 2016; pp. 1–5.
24. Xi, X.; Luo, X.; Zhang, Z.; You, Q.; Wu, X. Multimodal medical volumetric image fusion based on multi-feature in 3-D shearlet transform. In Proceedings of the International Smart Cities Conference, Wuxi, China, 14–17 September 2017; pp. 1–6.
25. Shabanzade, F.; Ghassemian, H. Multimodal image fusion via sparse representation and clustering-based dictionary learning algorithm in nonsubsampled contourlet domain. In Proceedings of the 8th International Symposium on Telecommunications, Tehran, Iran, 27–28 September 2016; pp. 472–477.
26. Xia, J.; Chen, Y.; Chen, A.; Chen, Y. Medical image fusion based on sparse representation and pcnn in nsct domain. *Comput. Math. Methods Med.* **2018**, *5*, 2806047. [CrossRef]
27. Yin, M.; Liu, X.N.; Liu, Y.; Chen, X. Medical image fusion with parameter-adaptive pulse coupled-neural network in nonsubsampled shearlet transform domain. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 49–64. [CrossRef]
28. Dinh, P.H. A novel approach based on three-scale image decomposition and marine predators algorithm for multi-modal medical image fusion. *Biomed. Signal Process. Control* **2021**, *67*, 102536. [CrossRef]
29. Pei, C.; Fan, K.; Wang, W. Two-scale multimodal medical image fusion based on guided filtering and sparse representation. *IEEE Access* **2020**, *8*, 140216–140233. [CrossRef]
30. Shahdoosti, H.R.; Mehrabi, A. Multimodal image fusion using sparse representation classification in tetrolet domain. *Digit. Signal Process.* **2018**, *79*, 9–22. [CrossRef]
31. Ling, T.; Yu, X. Medical image fusion based on fast finite shearlet transform and sparse representation. *Comput. Math. Methods Med.* **2019**, *2019*, 3503267.
32. Elad, M.; Yavneh, I. A plurality of sparse representations is better than the sparsest one alone. *IEEE Trans. Inf. Theory* **2009**, *55*, 4701–4714. [CrossRef]

33. Liu, Y.; Chen, X.; Rabab, K.W.; Wang, Z. Image fusion with convolutional sparse representation. *IEEE Signal Process. Lett.* **2016**, *23*, 1882–1886. [CrossRef]
34. Ghassemi, M.; Shakeri, Z.; Sarwate, A.D.; Bajwa, W.U. Learning mixtures of separable dictionaries for tensor data: Analysis and algorithms. *IEEE Trans. Signal Process.* **2019**, *68*, 33–48. [CrossRef]
35. Olshausen, B.A.; Field, D.J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **1996**, *381*, 607–609. [CrossRef] [PubMed]
36. Sturm, B.L.; Christensen, M.G. Cyclic matching pursuits with multiscale time-frequency dictionaries. In Proceedings of the Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 7–10 November 2011.
37. Schnass, K. Average performance of orthogonal matching pursuit (OMP) for sparse approximation. *IEEE Signal Process. Lett.* **2018**, *26*, 1566–1567. [CrossRef]
38. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2010**, *3*, 1–122. [CrossRef]
39. Hawe, S.; Seibert, M.; Kleinsteuber, M. Separable dictionary learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 438–445.
40. Hu, Q.; Hu, S.; Zhang, F. Multi-modality medical image fusion based on separable dictionary learning and Gabor filtering. *Signal Process. Image Commun.* **2020**, *83*, 115758. [CrossRef]
41. Kim, W.; Kim, C. Spatiotemporal saliency detection using textural contrast and its applications. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *24*, 646–659.
42. Zhou, Z.; Li, S.; Wang, B. Multi-scale weighted gradient-based fusion for multi-focus images. *Inf. Fusion* **2014**, *20*, 60–72. [CrossRef]
43. Hou, X.; Zhang, L. Saliency detection: A spectral residual approach. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Minneapolis, MI, USA, 17–22 June 2007; pp. 1–8.
44. Liu, X.; Mei, W.; Du, H. Structure tensor and nonsubsampled shearlet transform based algorithm for CT and MRI image fusion. *Neurocomputing* **2017**, *235*, 131–139. [CrossRef]
45. Pan, S.; Zhang, J.; Liu, X.; Guo, X. Complementary Harris feature point extraction method based on mutual information. *Signal Process.* **2017**, *130*, 132–139.
46. Petrović, V. Subjective tests for image fusion evaluation and objective metric validation. *Inf. Fusion* **2007**, *8*, 208–216. [CrossRef]
47. Yang, C.; Zhang, J.; Wang, X.; Liu, X. A novel similarity based quality metric for image fusion. *Inf. Fusion* **2008**, *9*, 156–160. [CrossRef]
48. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
49. Piella, G.; Heijmans, H. A new quality metric for image fusion. *Int. Conf. Image Process.* **2003**, *3*, 173–176.
50. Qu, G.; Zhang, D.; Yan, P. Information measure for performance of image fusion. *Electron. Lett.* **2002**, *38*, 313–315. [CrossRef]
51. Li, X.; Guo, X.; Han, P.; Wang, X.; Luo, T. Laplacian re-decomposition for multimodal medical image fusion. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6880–6890. [CrossRef]
52. Tan, W.; Tiwari, P.; Pandey, H.M.; Moreira, C.; Jaiswal, A.K. Multimodal medical image fusion algorithm in the era of big data. *Neural Comput. Appl.* **2020**, *3*, 1–21. [CrossRef]
53. Liu, Y.; Chen, X.; Ward, R.; Wang, Z.J. Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Process. Lett.* **2019**, *26*, 485–489. [CrossRef]