International Journal of Bio-Inspired Computation

ISSN online: 1758-0374 - ISSN print: 1758-0366
https://www.inderscience.com/ijbic

Multi-modal MRI image fusion of the brain based on joint bilateral filter and non-subsampled shearlet transform

Changhan Meng, Mengxing Huang, Yuchun Li, Yu Zhang, Siling Feng, Yuanyuan Wu

DOI: 10.1504/IJBIC.2023.10054703

Article History:

Received:	29 January 2022
Last revised:	29 December 2022
Accepted:	28 January 2023
Published online:	04 April 2023

Multi-modal MRI image fusion of the brain based on joint bilateral filter and non-subsampled shearlet transform

Changhan Meng, Mengxing Huang and Yuchun Li

State Key Laboratory of Marine Resource Utilization in South China Sea,
College of Information Science and Technology,
Hainan University,
Haikou, China
Email: mch611@163.com
Email: huangmx09@163.com
Email: liyuchun_19@163.com

Yu Zhang*

College of Computer Science and Technology,
Hainan University,
Haikou, China
Email: yuzhang_nwpu@163.com
*Corresponding author

Siling Feng and Yuanyuan Wu

State Key Laboratory of Marine Resource Utilization in South China Sea,
College of Information Science and Technology,
Hainan University,
Haikou, China
Email: fengsiling2008@163.com
Email: wyuanyuan@163.com

Abstract: Multi-modal brain MRI image fusion is one of the hottest discussed issues in the current research of medical image processing and has a deep impact on brain science and diagnosis. In this study, a fusion algorithm based on the joint bilateral filter (JBF) and the non-subsampled shearlet transform (NSST) is proposed. First, the multi-modal brain MRI images were decomposed by NSST and JBF models to derive the high-frequency component and energy layer. Secondly, the corresponding energy layer images and high-frequency components are fused. Thirdly, the inverse NSST transform is performed on the energy layer fusion image and the high-frequency fusion image to obtain the ultimate fusion image. Finally, the algorithm was evaluated using a publicly available brain dataset. The experimental results show that the algorithm achieves good performance in terms of both subjective evaluation and objective metrics.

Keywords: magnetic resonance image fusion; multi-modal; non-subsampled shearlet transform; NSST; joint bilateral filter; JBF.

Reference to this paper should be made as follows: Meng, C., Huang, M., Li, Y., Zhang, Y., Feng, S. and Wu, Y. (2023) 'Multi-modal MRI image fusion of the brain based on joint bilateral filter and non-subsampled shearlet transform', *Int. J. Bio-Inspired Computation*, Vol. 21, No. 1, pp.26–35.

Biographical notes: Changhan Meng is currently studying as a postgraduate student at the College of Information and Communication Engineering at Hainan University. Her research interests include multi-modal fusion of medical images.

Mengxing Huang received his PhD from Northwestern Polytechnic University in 2007. Currently, he is a Professor at Hainan University and serves as the Dean of the College of Information Science and Technology. His research interests include big data and intelligent information processing, ocean information perception and fusion AI and smart service.

Yuchun Li is currently pursuing her PhD at Hainan University. Her research interests include medical image processing.

Yu Zhang received his PhD in Computer Science and Technology from the School of Computer Science, Northwestern Polytechnic University in 2015. Currently, he is an Associate Professor at Hainan University. His research interests include co-modelling and co-verification of AI-enabled CPS, medical image processing.

Siling Feng received her PhD in Computer Software and Theory from the University of Electronic Science and Technology in 2014. Currently, she is a Professor at Hainan University. Her research interests include intelligent computing, image processing and knowledge graphs.

Yuanyuan Wu received her PhD in Engineering from the Southeast University in 2010. Currently, she is an Associate Professor at Hainan University. Her research interests include complex systems analysis and intelligent control.

1 Introduction

Medical imaging technology is a way of assisting doctors in visualising the internal structures of a patient's body, with many illnesses requiring medical imaging devices to reach a diagnosis. The medical images obtained from different imaging devices carry different information about the condition and have different emphases, while certain modal images present complementary information. Computed tomography (CT) is more sensitive to compact structures, including bone, and can clearly show organ structures through its greyscale reflection of tissue density, but soft tissues cannot be shown in detail (Li et al., 2020). Magnetic resonance imaging (MRI), by contrast, shows anatomical structural information about soft tissue structures due to their high spatial resolution. Nevertheless, no information about the metabolic activity of the body can be exhibited. T1-weighted imaging in MRI precisely reflects anatomical structures to some degree, while T2-weighted imaging primarily furnishes details of tissue lesions (Li et al., 2021b). Fusing the different modal images and retaining the respective feature information can help doctors to better understand the patient's physical condition (James and Dasarathy, 2014; Huang et al., 2020) and accurately determine the diseased areas of the body. Multi-modal medical image fusion (MMIF) (Hermessi et al., 2021) is a gateway to accomplish this. Thus, multi-modal MRI image fusion has a high research value and tremendous potential (Hermessi et al., 2021).

Currently, multi-modal MRI image fusion algorithms can categorise into three main groups: sparse representation (SR) (Yang and Li, 2010; Zhu et al., 2018; Zhou et al., 2019), deep learning (DL) (Liu et al., 2017b; Zhang et al., 2020; Liu et al., 2017a), and multi-scale transform (MST) (Yin et al., 2018; Zhu et al., 2019). Among them, SR-based algorithms are mainly classified into two groups: fixed dictionaries and dictionary learning-based methods. The fixed dictionary has limited transformations to cater to image content and application diversity, and has the disadvantage of being time-consuming and computationally complex (Yang and Li, 2010). Whereas dictionary learning-based methods (Zhu et al., 2018; Zhou et al., 2019) not only need longer training time, but may also suffer from colour distortion. DL-based algorithms are a hot research topic in computing and are widely used in image fusion. According to research (Liu et al., 2017b), the convolutional neural network (CNN) fusion algorithm is the most suitable architecture for image fusion. Zhang et al. (2020) presented a general image fusion framework based on convolutional neural networks (IFCNN). They utilised two convolutional layers to extract salient features and selected fusion rules based on image type. But DL-based algorithms are not only demanding on the dataset, but also complex and time-consuming to calculate, with cumbersome parameter settings. The MST-based fusion algorithm decomposes the image to different scales and resolutions to obtain sub-band images containing different information types. Depending on the sub-band type, different fusion rules are selected for image fusion. This type of algorithm processes image information in a way that is strongly similar to the way the human visual system (HVS) (Li et al., 2017) perceives vision. Zhu et al. (2019) constructed a medical image fusion technique based on the non-subsampled contour wave transform (NSCT) and designed a local Laplace energy norm as a fusion rule for the high-frequency components. But the MST algorithm has an inherent flaw in that some useful information is inevitably lost in the process of decomposition and reconstruction, failing to achieve optimal fusion performance.

The filter fusion algorithm is very similar to the MST-based algorithm. By giving weights to the different signal components, it decomposes the image into a structural layer containing rich gradient and small energy information and an intensity layer containing pure energy information. The filter fusion algorithm can achieve the purpose of preserving edges and smoothing noise with low computational complexity (Goyal et al., 2012). Kumar (2015) used a cross bilateral filter (CBF) to extract detailed images, then fused them by weighted averaging. But it is generally effective in medical image fusion. Qin et al. (2019) proposed a medical image fusion method based on weighted least squares filtering. The algorithm preserves the colour and edge structure well, but the texture layer without processing is directly combined with the fused sub-images, which can lead to unclear texture details. Li et al. (2021b) proposed a multi-modal medical image fusion algorithm based on joint bilateral filters and local gradient energy (JBF-LGE). The algorithm decomposed medical images into a structural layer containing rich

details and an energy layer with pure intensity information, and designed a local gradient energy (LGE) operator to fuse the structural layer. But the contour lines of bone structures in the fused images produced by this method sometimes have inconsistent brightness. Li et al. (2021a) also proposed an algorithm (JBF-SR) that decomposes the image by JBF and uses an SR-based fusion rule on the high-frequency components. The fusion rule was obtained by training with the KSVD algorithm. Dictionary training requires consideration of many parameters, such as number, content, and iterations of training data. Although the convergence speed of the KSVD algorithm has been greatly improved, the training time is still long. As for the non-subsampled shearlet transform (NSST), it has better multi-resolution property, anisotropy and translation invariance. From the perspective of approximation theory, it is an optimal approximation, a 'true' sparse representation of the image in all directions and scales, without the need for dictionary training. Based on the above-mentioned problems, this paper combines NSST and JBF techniques to propose a new multi-modal medical image fusion algorithm. The experimental results show that the proposed method outperforms some advanced fusion algorithms in terms of visual and objective evaluation.

To put it in a nutshell, the main contributions of this paper are as follows:

- An improved multi-modal medical image fusion algorithm is proposed. In the image decomposition segment, we introduce the NSST transform and JBF technique to propose a novel decomposition scheme. We decompose the source image into an energy layer containing pure energy information and a series of high-frequency sub-bands, which well preserves the texture information of the source image.
- The approach is quite versatile so that it is scalable to other types of image fusion with good fusion performance.

The remaining sections of this paper are scheduled as follows. Section 2 provides a brief overview of the joint bilateral filter (JBF) theory. Section 3 details the proposed multi-modal medical image fusion algorithm. Section 4 undertakes substantial experiments, covering its setup, discussion of experimental results, along with the extended applications. Section 5 concludes the paper as well as the subsequent work.

2 Joint bilateral filter

Bilateral filter (BF) (Tomasi and Manduchi, 2002) smoothes and denoises an image while preserving its edge information. Nonetheless, many bilateral filters can accidentally exert a blurring effect while denoising, resulting in partial loss of detail and structural information in the image. In order to overcome the shortcomings of bilateral filter, the JBF (Petschnigg et al., 2004) was created. Joint bilateral filter quite resembles bilateral filter, which can be described as a special example of a JBF.

Equation (1) is the formula for JBF representing the filtered output value of image I at pixel p.

$$J_p = \frac{1}{w_p^{jbf}} \sum_{q \in \Omega} I_q G_s(\|p-q\|) G_r\left(\left\|\tilde{I}_p - \tilde{I}_q\right\|\right) \tag{1}$$

where I denotes the source image and J is the output image after the source image I has been filtered. And $\widetilde{I}$ is the introduced guide map, normally a blurred image of its source image I. Both p and q are pixel indices and q is the pixel within the image window Ω centred on p. The size of the image window Ω is usually 3. The J_p represents the pixel value obtained by the filtered p^{th} pixel of the source image I, and I_p, I_q denote the p^{th} pixel of the source image I and the q^{th} pixel within the image window Ω centred on the p^{th} pixel of the source image I, respectively. Similarly, $\widetilde{I}_p$, $\widetilde{I}_q$ represent the p^{th} pixel of the guide image $\widetilde{I}$ and the q^{th} pixel within the image window Ω centred on the p^{th} pixel of the guide image $\widetilde{I}$, respectively. G_s means the kernel function for the spatial domain and G_r is denotes the kernel function for the pixel range domain, namely the weight distribution function, generally a Gaussian function. w_p^{jbf} indicates the normalised term and is calculated as follows:

$$w_p^{jbf} = \sum_{q \in \Omega} G_s(\|p-q\|) G_r\left(\left\|\tilde{I}_p - \tilde{I}_q\right\|\right) \tag{2}$$

JBF actually introduces a guide map and takes it as the basis for the value domain weight calculation. To be more specific, the JBF filter kernel was formed to consider both the greyscale similarity of neighbouring pixels in the guide map $\widetilde{I}$ and the geometric closeness of neighbouring pixels (Kumar, 2015). The formed filter kernel is then used to filter the input source image I.

3 The proposed algorithm

Combining JBF and NSST, a new fusion algorithm is proposed in this paper. The steps are divided into two main steps: source image decomposition, and image fusion reconstruction. Firstly, the multi-modal brain MRI images are disaggregated by NSST to gain their low-frequency and high-frequency components. Meanwhile, the source images are decomposed into the energy layer and the structure layer image by the technique of two-scale image decomposition from the JBF model. And the high-frequency component and the energy layer image which replaces the low-frequency component are fused applying the abs-max criterion. By the inverse NSST, it is available to restructive the image at last. The schematic diagram of the algorithm is displayed in Figure 1. The red dashed boxes belong to the source image decomposition step, while the blue dashed boxes represent step 2, namely image fusion reconstruction. The legend is shown at the bottom right, with different coloured arrows representing the different operational steps. The dark blue represents Gaussian blur, the yellow represents edge recovery, the light

green represents the NSST decomposition and the purple represents the NSST inverse conversion.

3.1 Source image decomposition

In this paper, a new decomposition scheme is proposed by combining JBF and NSST techniques. The decomposition scheme can be specifically divided into two parts, JBF-based dual-scale decomposition and NSST multi-scale decomposition. The algorithm first decomposes the source image into an energy layer containing pure energy information and a structure layer including rich gradient and small energy information by JBF-based dual-scale decomposition. Also, the source image is decomposed into a low-frequency sub-band and a series of high-frequency sub-bands (48) using the NSST transform. Finally, the energy layer and high-frequency sub-bands are retained and subjected to the NSST inversion transform to obtain the final fused image.

3.1.1 The dual-scale decomposition based on JBF

Medical images are mostly greyscale images, unlike other types of images that have greater colour information. What is more, they are mainly single content information, without as much content information as other types of images. In summary, the main information in medical images is luminance and detail, while the dual-scale decomposition based on JBF is under the assumption that "greyscale and gradient information are two of the key features of medical images", which is very suitable for the characteristics of MRI images.

The dual-scale decomposition based on JBF is applied to the dual-scale component separation technique (Li et al., 2021b), which is an excellent method for performing basis-detail information separation. The detailed process of its action is shown in Figure 2.

It starts with smoothing and Gaussian filtering of the source image. The smoothing process is calculated as follows.

$$\boldsymbol{R_\sigma} = G_\sigma * \boldsymbol{I} \tag{3}$$

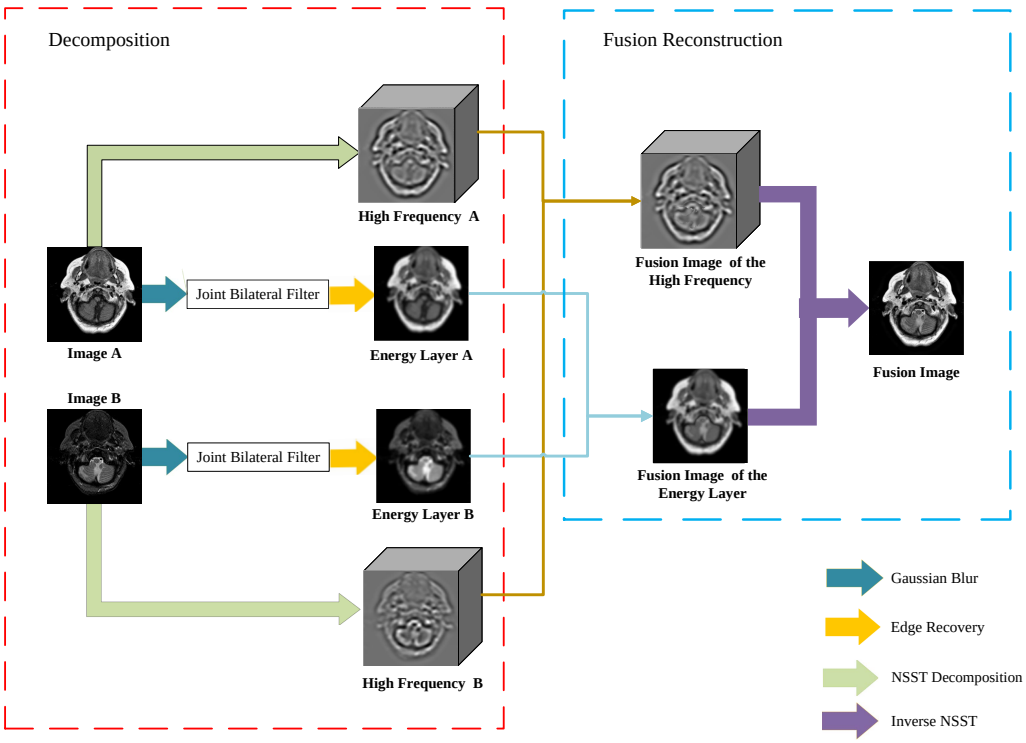
$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{4}$$

where $\boldsymbol{I}$ is the input source image, G_σ represents a Gaussian filter template with variance σ^2, and $\boldsymbol{R}_\sigma$ is the output of the smoothing process with standard deviation σ. The σ^2 denotes the scale (Lei et al., 2019), and (x,y) indicates the position of the ranks of the pixels. Through this measure, the smoothed image will not contain information on structural scales smaller than σ. Subsequently, a weighted average Gaussian filter is employed to obtain the global blurred image, computed as follows:

$$\boldsymbol{G}(j) = \frac{1}{\mathbf{Z}_j}\sum_{i\in N(j)}\exp\left(-\frac{\|j-i\|^2}{2\sigma_s^2}\right)\boldsymbol{I}(i) \tag{5}$$

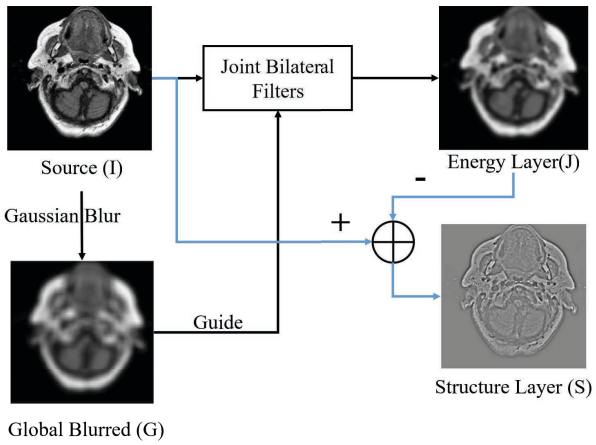
$$\mathbf{Z}_j = \sum_{i\in \boldsymbol{N}(j)}\exp\left(-\frac{\|j-i\|^2}{2\sigma_S^2}\right) \tag{6}$$

Figure 1 The overview of the proposed algorithm (see online version for colours)



where $\boldsymbol{I}$ and $\boldsymbol{G}$ denote the input image and the filtered output global blurred image respectively, and $\mathbf{Z}_j$ denotes normalisation, while σ_S is the smoothing parameter. The parameter i, j is the pixel index and $N(j)$ represents the local area of the j^{th} pixel. Whereas i is a pixel within $N(j)$. $\boldsymbol{G}(j)$ denotes the pixel value of the output after the j^{th} pixel of the input image is filtered.

Figure 2 The schematic diagram of the dual-scale decomposition based on JBF (see online version for colours)



It is essential to mention that the global blurred image cannot be used directly as an energy layer, but also needs to undergo an edge recovery operation. In the edge recovery, the JBF is introduced to keep the large-scale structure of the image from being destroyed during the smoothing process. Together with equations (1) and (2), the result after JBF is calculated by the following equation.

$$\boldsymbol{J}(j) = \frac{1}{\mathbf{Z}_j} \sum_{i \in N(j)} g_d(i-j) g_r(i-j) \boldsymbol{I}(i) \tag{7}$$

$$\mathbf{Z}_j = \sum_{i \in N(j)} \exp\left(-\frac{\|j-i\|^2}{2\sigma_S^2} - \frac{\|\boldsymbol{G}(j) - \boldsymbol{G}(i)\|^2}{2\sigma_r^2} \right) \tag{8}$$

where $\boldsymbol{J}$ denotes the result image after the input image has been filtered by the JBF. The parameter j is also the pixel index, so $\boldsymbol{J}(j)$ represents the j^{th} pixel value after JBF of the incoming image. The σ_S and σ_r are smoothing parameters. The σ_S is used to adjust the weight value of pixels with larger spatial distance, and the σ_r is the standard deviation of the similarity factor controlling the greyscale range, which is used to adjust the weight value of pixels with larger pixel differences. Their size determines the filtering effect of the BF. And g_d and g_r denote the spatial distance function and the intensity range function respectively. According to the theory of JBF, they set weights based on the distance between pixels (i.e., geometric similarity) and intensity differences (i.e., greyscale similarity), respectively, calculated as follows.

$$g_d(i-j) = \exp\left(-\frac{\|j-i\|^2}{2\sigma_s^2} \right) \tag{9}$$

$$g_r(i-j) = \exp\left(-\frac{\|\boldsymbol{G}(j) - \boldsymbol{G}(i)\|^2}{2\sigma_r^2} \right) \tag{10}$$

With the above calculation, the edge recovery effect, i.e., edge preservation, can be achieved. Ultimately, the relationship between the energy layer, the structure layer and the source image can be shown by equation (11).

$$\boldsymbol{S} = \boldsymbol{I} - \boldsymbol{J} \tag{11}$$

where $\boldsymbol{S}$ is the structural layer image, $\boldsymbol{I}$ denotes the source image and $\boldsymbol{J}$ is the energy layer image, i.e., the outcome of the JBF. The structure layer contains primarily details of the source image, i.e., small-scale information (Li et al., 2021b), including textures, details, small edges, etc.

3.1.2 The NSST multi-scale decomposition

As shown in Figure 1, with the proposed algorithm, both source images A and B are also decomposed by NSST. There are four levels of decomposition and the final decomposition results in one low-frequency component and 48 high-frequency components at different scales, where the first level of decomposition produces eight sub-bands, the second level produces eight sub-bands, and the third level produces 16 sub-bands and the fourth level produces 16 sub-bands.

Compared to the low-frequency images obtained by NSST decomposition, the energy layer from JBF decomposition contains more information and to some extent includes some of the structural features. The high-frequency component, on the other hand, usually contains rich edge and texture information. Thus, the proposed algorithm substitutes the low-frequency image with the energy layer image as a way to improve the fusion performance.

3.2 Image fusion reconstruction

After the above decomposition, we can obtain the new low-frequency component and the NSST high-frequency component. Since the NSST component has weak structure and high intensity, we take the absolute value of the pixel as the pixel's activity and use the Abs-Max criterion for NSST component fusion, which is calculated as follows.

$$FE = AE. * map + \widetilde{map}. * BE \tag{12}$$

$$\text{map}(x,y) = \begin{cases} 1, & |AE(x,y)| > |BE(x,y)| \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where FE denotes the fused image of the energy layer, AE and BE denote the energy layers of the two source images respectively, and the map indicates the fusion coefficient.

For the NSST high-frequency component, as a rule, a larger absolute pixel value means more active pixels in the image, so the high-frequency coefficients are often used for image fusion with the absolute maximum rule. Such is also

the case with the proposed algorithm. The calculation is as follows:

$$FH(i) = \begin{cases} AH(i), & AH(i) \geq BH(i) \\ BH(i), & \text{otherwise} \end{cases} \tag{14}$$

where FH denotes the fused image of high-frequency sub-bands. i represents the scale of high-frequency sub-bands and the value range is $[1-48]$. Namely, $FH(i)$ denotes the fused image of high-frequency sub-bands with scale i. The AH and BH denote the high-frequency sub-bands of the two source images, respectively. Finally, the proposed algorithm performs the inverse NSST transform of the new low-frequency component and the high-frequency component, thus reconstructing the image to obtain the final fused image.

4 Experiment

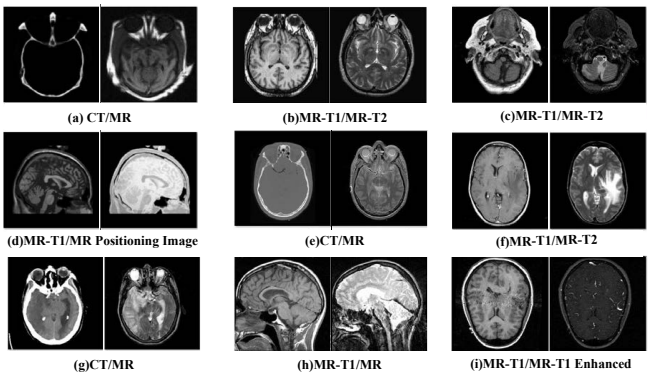
4.1 Experimental dataset

To verify the efficiency of the algorithm, 407 pairs of multi-modal medical brain image pairs are selected for the experiments in this paper. The image dataset was obtained from The Whole Brain Atlas[1], a medical brain image dataset established by Harvard Medical School (Dagley et al., 2017). Figure 3 shows nine pairs of representative multi-modal medical images of the brain. These source images were all from different imaging devices and contained modality types such as CT, MR-T1, MR-T1 enhanced, MR-T2, and MR positioning image. We also randomly selected 398 pairs of images in The Whole Brain Atlas dataset. There were 159 pairs of CT and MR fusion types and 239 pairs of MR-T1 and MR-T2. The images belonging to MR-T1 and MR-T2 fusion types were from normal brains, while the images belonging to CT and MR fusion types were from patients with cerebrovascular diseases (acute stroke, hypertensive encephalopathy, multiple embolic infarction, etc.) or neoplastic diseases (meningioma, metastatic bronchogenic carcinoma, etc.) or infectious diseases (cerebral toxoplasmosis, etc.). In addition, to explain the generality of the algorithm, two pairs from each of the IR-visible image dataset[2] and the multi-focus image dataset[3] were selected to prove the fusion capability of the proposed algorithm.

4.2 Fusion evaluation metrics

The objective evaluation metrics used in this paper are information entropy (EN) (Meher et al., 2019), standard deviation (SD) (Meher et al., 2019), feature mutual information (FMI) (Cvejic et al., 2006), gradient-based metric ($Q_{AB/F}$) (Xydeas and Vladimir, 2000; Shen et al., 2020), and the sum of the correlations of differences (SCD) (Aslantas and Bendes, 2015). These five indicators are evaluated quantitatively. Where the calculated value is higher, the integration performance is superior.

Figure 3 Nine pairs of multi-modal medical images



4.3 Ablation experiment

To verify the rationality of the proposed algorithm we performed two ablation experiments.

The first ablation experiment verified the role of NSST and JBF in the algorithm. The experiment first kept the fusion rules consistent and compared the proposed algorithm with the pure NSST algorithm and the JBF algorithm for objective evaluation of the metrics, as shown in Table 1. From Table 1, it can be seen that the proposed algorithm performs best on the metric $Q_{AB/F}$, with all other metrics falling in between the JBF and NSST algorithms. The EN, SD, and SCD values of the proposed algorithm are slightly worse than those of the JBF algorithm, but much better than those of the NSST algorithm. Similarly, the FMI values are slightly lower than those of the NSST algorithm, but much better than those of the JBF algorithm. The experimental results demonstrate that the proposed algorithm combines the advantages of both NSST and JBF, and both of them are improved within the proposed algorithm.

Table 1 Quantitative results of ablation experiment 1

Method	*EN*	*SD*	*FMI*	*SCD*	$Q_{AB/F}$
Proposed	0.7867	0.3018	0.8824	1.4216	*0.5902*
JBF	*0.8283*	*0.3080*	0.8599	*1.4929*	0.4881
NSST	0.7810	0.2979	*0.8833*	1.3841	0.5830

The second ablation experiment demonstrated the reasonableness of replacing the low-frequency component of the NSST with the energy layer obtained from the JBF decomposition. A consistent fusion rule is used for the experiments, and Algorithm 1 is the proposed algorithm. It replaces the low-frequency component of NSST with the energy layer obtained by JBF decomposition, and finally the energy layer obtained by JBF decomposition with the high-frequency component of NSST is inverted by NSST. Algorithm 2 replaces the high-frequency components of NSST with the structural layer obtained by JBF decomposition, and finally the structural layer obtained by JBF decomposition and the low-frequency components of NSST is performed as an inverse operation

to obtain the final fused image. Likewise, we use the fusion metric to measure the fusion performance as shown in Table 2. As can be seen from Table 2, the proposed algorithm (Algorithm 1) outperforms in all evaluation metrics, especially in the metrics SD, $Q_{AB/F}$ and SCD, with values much better than the comparison algorithm (Algorithm 2). In other words, the fusion of the energy layer and high-frequency components was superior to that of the structural layer and low-frequency components. Therefore, it is more reasonable to replace the low-frequency component of NSST with the energy layer obtained by JBF decomposition.

Table 2 Quantitative results of ablation experiment 2

Method	*EN*	*SD*	*FMI*	*SCD*	$Q_{AB/F}$
Algorithm 1	*0.7867*	*0.3018*	*0.8824*	*1.4216*	*0.5902*
Algorithm 2	0.7750	0.2823	0.8780	1.1180	0.5452

4.4 Experimental results and discussion

In this paper, the proposed algorithm is compared with four other algorithms with better fusion results, namely the IFCNN (Zhang et al., 2020), JBF-LGE (Li et al., 2021b), NSCT (Zhu et al., 2019) and CBF (Kumar, 2015) algorithms.

4.4.1 Subjective evaluation

It is well-known that the most straightforward way to judge the fusion performance of a fusion algorithm is to directly evaluate the fusion results visually through the human visual system. Figures 4 and 5 show the fused images obtained by different algorithms for four pairs of multi-modal medical images. The source images in Figure 4 are CT and MR, and in Figure 5 are MR-T1 and MR-T2. The red rectangular box zooms in on some areas to visualise the detailed information in the image. Comparing Figures 4 and 5, it can be seen that the fused images of the CBF and IFCNN algorithms have lower overall brightness and less energy information than the other fused images, and the fused images of the NSCT algorithm have less soft body texture information and blurrier edges than the proposed algorithm. Comparing the detail enlarged area (red rectangular box) in the second row of Figure 5, it can be seen that the JBF-LGE algorithm fused image retains less texture information than the proposed algorithm.

4.4.2 Objective evaluation indicators

Experiments were selected from 398 pairs of medical brain images from The Whole Brain Atlas for fusion performance evaluation experiments. The dataset is divided into two main categories, CT and MR, MR-T1 and MR-T2. These evaluation metrics are all positive, with higher values associated with better fusion performance, and the specific objective evaluation metric data is shown in Table 3. Each fusion metric in the table is the average of 398 pairs of source images, and the values shown in italic are the best scores under that metric. A box-and-whisker diagram is also drawn based on the objective evaluation metrics, as shown in Figure 6. The box-and-whisker diagram consists of a 'box' and a 'whisker', representing the dispersion of the data. The 'box' has a straight line representing the middle of the sample, with the top and bottom bounds representing the 75% and 25% values respectively. The two 'whiskers' are the maximum and minimum values of the data, and outlier points are generally plotted separately, with a '+'. In addition, the mean values of the objective evaluation indicators are marked with a diamond.

Table 3 Mean values of objective evaluation metrics for 398 fused images

Method	*IFCNN*	*JBF-LGE*	*NSCT*	*CBF*	*Proposed*
EN	0.8091	*0.8668*	0.8398	0.7543	0.8597
SD	0.2990	0.3316	0.3255	0.2772	*0.3318*
SCD	1.1639	1.2776	1.1873	0.8007	*1.3016*
$Q_{AB/F}$	0.5700	0.5436	0.5609	*0.5711*	0.5662

As can be seen from Table 3 and Figure 6, the proposed algorithm in terms of the standard deviation (SD) and the sum of the correlations of differences (SCD), is much better than the three compared algorithms, IFCNN, NSCT and CBF, and slightly superior to the JBF-LGE algorithm. In terms of information entropy (EN) and gradient-based metric ($Q_{AB/F}$), the proposed algorithm ranked second and was very slightly different from the best metric. On the whole, the proposed algorithm appears to perform well on the objective evaluation metrics and has an advantage over all three of the selected comparison algorithms.

4.5 Other types of image fusion

The proposed algorithm has a high generalisation capability that can be extended to other image fusion types. To demonstrate its generalisation capability, we applied the proposed algorithm to multi-focus image fusion and infrared and visible (IR-VIS) image fusion types. The specific fusion results are shown in Figures 7 and 8. In Figures 7, Figures 7(a) and 7(b) are the infrared image and the visible image, respectively. Figures 7(c)–7(g) are the fused images obtained from IFCNN, JBF-LGE, NSCT, CBF and the proposed algorithms, respectively. But in Figure 8, Figures 8(a) and 8(b) are the source images with different focal points, respectively.

Figure 4 Fusion results of different methods on a pair of CT and MR images (see online version for colours)

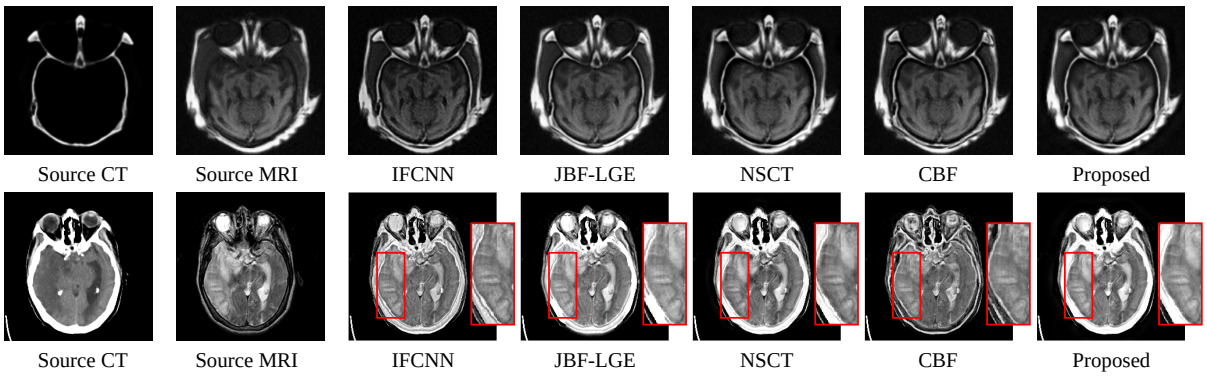


Figure 5 Fusion results of different methods on a pair of MR-T1 and MR-T2 images (see online version for colours)

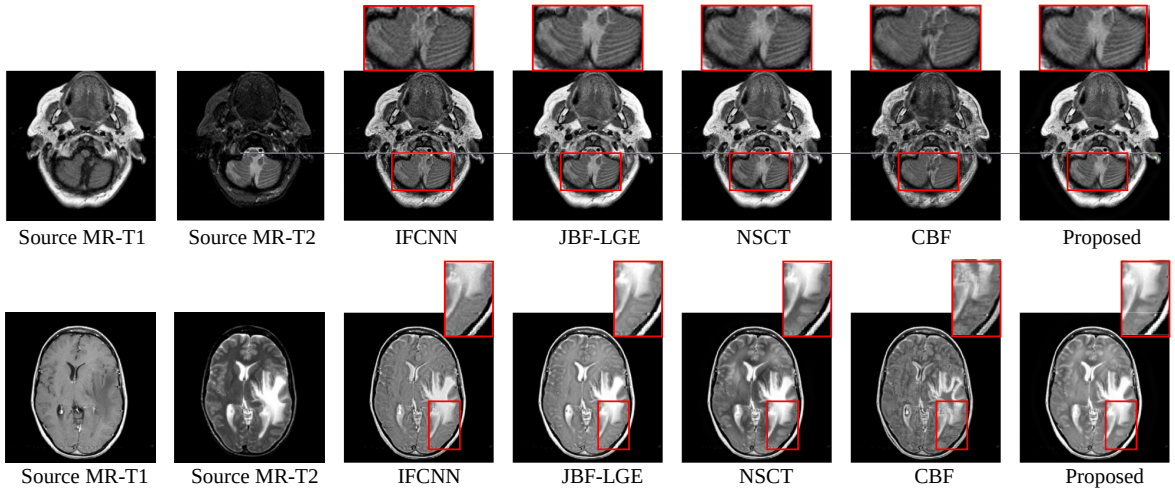


Figure 6 Box-and-whisker diagram of objective evaluation indicators for 398 pairs of the source image (see online version for colours)

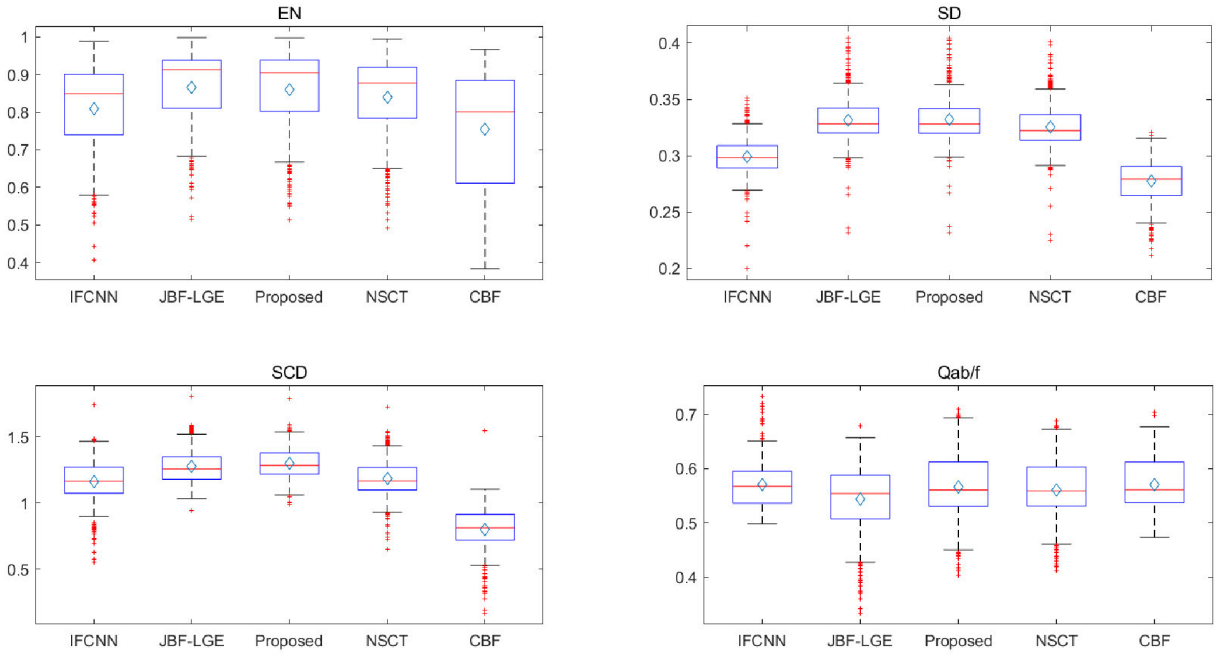


Figure 7 IR-VIS fusion images under different algorithms

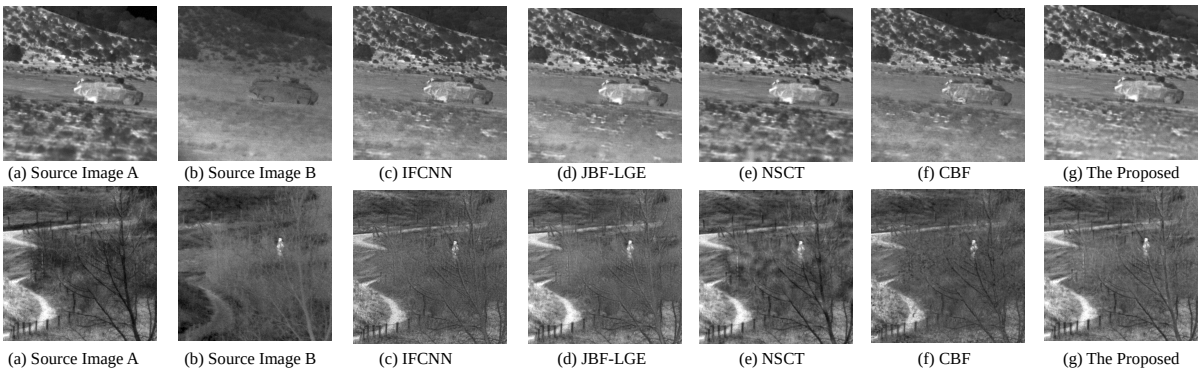
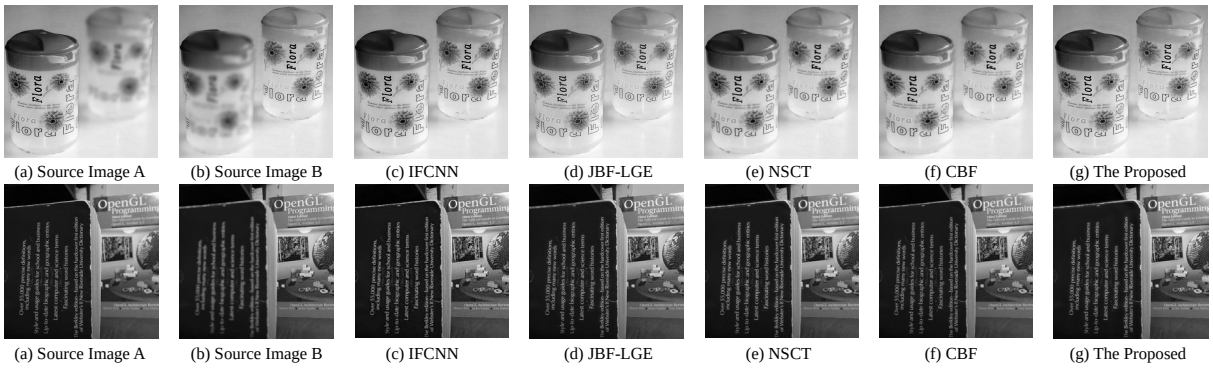


Figure 8 Multi-focus fusion images with different algorithms



It can be seen from Figures 7 and 8 that the proposed algorithm performs well on other fusion types, successfully retaining the feature information about the source image pair, especially on the type of fusion of IR and VIS, which is more in line with the human visual system. From the fused results of the IR-VIS images, we can observe that the fused images successfully retain the feature information of both images. Comparing the first pair of the multi-source images, Figures 7(c) to 7(g) reveal that the proposed algorithm retains more information on the lawn structure than the JBF-LGE algorithm, the CBF algorithm and the IFCNN algorithm, but slightly less than the NSCT algorithm. Comparing the second pair of the multi-source images, especially the tree structure information, the proposed algorithm has a better fusion effect. Overall, the proposed algorithm preserves the structural information of the image better than the other four compared algorithms in terms of visual effect and has greater fusion performance. Judging from the fusion results of the multi-focus images, we can observe that the fused images successfully retain the clear image information of both images. Comparing the images in Figures 8(c) to 8(f) indicates that the fusion effect of the proposed algorithm is almost the same as the other four algorithms, making full use of the clear areas with different focal points.

5 Conclusions

MRI image fusion assists doctors in their diagnosis and treatment, playing a positive role in understanding the patient's internal physical condition and accurately determining the area of the lesion. Because of its high research significance, it is a current research hotspot in the field of imaging.

A new improved algorithm for multi-modal MRI image fusion based on JBF and NSST is proposed in this paper. In this algorithm, we first perform NSST decomposition of the original image, in which the low-frequency components are replaced by the energy layer obtained from the JBF-based dual-scale decomposition scheme, enriching the low-frequency information and thus enhancing the fusion performance. Three typical fusion algorithms were chosen for the experiments in this paper, including two categories of algorithms based on MST and deep learning. In our experiments, we evaluated the fusion performance qualitatively through direct observation of the fused images by the human eye, as well as a quantitative comparison using five commonly used metrics. As experimental results show, the method outperforms some advanced fusion algorithms in terms of subjective and objective evaluation. Moreover, the method has great generality so that it can be applied to other image fusion problems, such as multi-focus image fusion and infrared and visible image fusion. Alongside the great fusion performance and universality is the long running time of the algorithm, especially for high-frequency fusion, which needs to be further optimised by studying the fusion rules of the algorithm.

Acknowledgements

This work was supported in part by Major Science and Technology Project of Haikou (Grant #: 2020-009), in part by Key R&D Project of Hainan province (Grant #: ZDYF2021SHFZ243), in part by National Natural Science Foundation of China (Grant #: 82260362), in part by the National Key R&D Program of China (No. 2021ZD0111000).

References

Aslantas, V. and Bendes, E. (2015) 'A new image quality metric for image fusion: the sum of the correlations of differences', *AEUE – International Journal of Electronics and Communications*, Vol. 69, No. 12, pp.1890–1896.

Cvejic, N., Canagarajah, C.N. and Bull, D.R. (2006) 'Image fusion metric based on mutual information and Tsallis entropy', *Electronics Letters*, Vol. 42, No. 11, pp.626–627.

Dagley, A., LaPoint, M., Huijbers, W., Hedden, T., McLaren, D.G., Chatwal, J.P., Papp, K.V., Amariglio, R.E., Blacker, D., Rentz, D.M., Johnson, K.A., Sperling, R.A. and Schultz, A.P. (2017) 'Harvard aging brain study: dataset and accessibility', *NeuroImage*, Vol. 144, pp.255–258, Data Sharing Part II.

Goyal, A., Bijalwan, A. and Chowdhury, M.K. (2012) 'A comprehensive review of image smoothing techniques', *International Journal of Advanced Research in Computer Engineering & Technology*, Vol. 1, No. 4, pp.315–319.

Hermessi, H., Mourali, O. and Zagrouba, E. (2021) 'Multimodal medical image fusion review: theoretical background and recent advances', *Signal Processing*, June, Vol. 183, p.108036, ScienceDirect.

Huang, J., Le, Z., Ma, Y., Fan, F., Zhang, H. and Yang, L. (2020) 'MGMDcGAN: medical image fusion using multi-generator multi-discriminator conditional generative adversarial network', *IEEE Access*, Vol. PP, No. 99, p.1.

James, A.P. and Dasarathy, B.V. (2014) 'Medical image fusion: a survey of the state of the art', *Information Fusion*, September, Vol. 19, pp.4–19.

Kumar, B.S. (2015) 'Image fusion based on pixel significance using cross bilateral filter', *Signal, Image and Video Processing*, Vol. 9, No. 5, pp.1193–1204.

Lei, F., Li, H., Gao, Y. and Zhang, Y. (2019) 'A color image segmentation method based on region salient color and fuzzy c-means algorithm', *Circuits Systems and Signal Processing*, February, Vol. 39, pp.586–610.

Li, B., Peng, H. and Wang, J. (2020) 'A novel fusion method based on dynamic threshold neural p systems and nonsubsampled contourlet transform for multi-modality medical images', *Signal Processing*, January, Vol. 178, p.107793.

Li, S., Kang, X., Fang, L., Hu, J. and Yin, H. (2017) 'Pixel-level image fusion: a survey of the state of the art', *Information Fusion*, January, Vol. 33, pp.100–112.

Li, X., Zhou, F. and Tan, H. (2021a) 'Joint image fusion and deniosing via three-layer decomposition and sparse representation', *Knowledge-Based Systems*, Vol. 224, No. 1, p.107087.

Li, X., Zhou, F., Tan, H., Zhang, W. and Zhao, C. (2021b) 'Multimodal medical image fusion based on joint bilateral filter and local gradient energy', *Information Sciences*, August, Vol. 569, pp.302–325.

Liu, Y., Chen, X., Cheng, J. and Peng, H. (2017a) 'A medical image fusion method based on convolutional neural networks', *2017 20th International Conference on Information Fusion (Fusion)*.

Liu, Y., Chen, X., Peng, H. and Wang, Z. (2017b) 'Multi-focus image fusion with a deep convolutional neural network', *Information Fusion*, July, Vol. 36, pp.191–207.

Meher, B., Agrawal, S., Panda, R. and Abraham, A. (2019) 'A survey on region based image fusion methods', *Information Fusion*, August, Vol. 48, pp.119–132.

Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H. and Toyama, K. (2004) 'Digital photography with flash and no-flash image pairs', *ACM Transactions on Graphics*, Vol. 23, No. 3, pp.664–672.

Qin, J., Shen, X., Chen, H., Lv, Y. and Zhang, X. (2019) 'A fusion algorithm for medical structural and functional images based on adaptive image decomposition', *Multimedia Tools and Applications*, Vol. 78, No. 22, pp.32605–32629.

Shen, D., Zareapoor, M. and Yang, J. (2020) 'Infrared and visible image fusion via global variable consensus', *Image and Vision Computing*, December, Vol. 104, p.104037.

Tomasi, C. and Manduchi, R. (2002) 'Bilateral filtering for gray and color images', *International Conference on Computer Vision*.

Xydeas, C. and Vladimir, P. (2000) 'Objective image fusion performance measure', *Military Technical Courier*, Vol. 56, No. 4, pp.181–193.

Yang, B. and Li, S. (2010) 'Multifocus image fusion and restoration with sparse representation', *IEEE Transactions on Instrumentation and Measurement*, Vol. 59, No. 4, pp.884–892.

Yin, M., Liu, X., Liu, Y. and Chen, X. (2018) 'Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain', *IEEE Transactions on Instrumentation and Measurement*, Vol. 68, No. 1, pp.49–64.

Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X. and Zhang, L. (2020) 'IFCNN: a general image fusion framework based on convolutional neural network', *Information Fusion*, February, Vol. 54, pp.99–118.

Zhou, F., Li, X., Zhou, M., Chen, Y. and Tan, H. (2019) 'A new dictionary construction based multimodal medical image fusion framework', *Entropy*, Vol. 21, No. 3, p.267.

Zhu, Z., Yin, H., Chai, Y., Li, Y. and Qi, G. (2018) 'A novel multi-modality image fusion method based on image decomposition and sparse representation', *Information Sciences*, March, Vol. 432, pp.516–529.

Zhu, Z., Zheng, M., Qi, G., Wang, D. and Xiang, Y. (2019) 'A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain', *IEEE Access*, February, Vol. 7, pp.20811–20824.

Notes

1 http://www.med.harvard.edu/aanlib/home.html.

2 https://sites.google.com/view/durgaprasadbavirisetti/datasets.

3 https://github.com/sametaymaz/Multi-focus-Image-Fusion-Dataset.