



A novel multi-modal medical image fusion algorithm

Xinhua Li¹ · Jing Zhao²

Received: 28 March 2020 / Accepted: 2 July 2020 / Published online: 9 July 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

In order to improve the contrast of image fusion and highlight the unique characteristics of medical images, a multi-modal medical image fusion algorithm in the framework of non-subsampled contourlet transform (NSCT) is proposed in this paper. Firstly, the computed tomography images and magnetic resonance image are decomposed into low- and high-frequency sub-bands through the NSCT of multi-scale geometric transformation; secondly, for the low-frequency sub-band, the local area standard deviation method is selected for the fusion, while for the high-frequency sub-band, an adaptive pulse coupling neural network model is constructed and the fusion rules are set by the cumulative ignition times of iterative operation in the network; finally, the fusion image is obtained through image reconstruction. Experimental results show that the fusion results of the algorithm in this paper can improve the image fusion quality significantly and it has certain advantages in both visual effects and objective evaluation indexes, which provides a more reliable basis for clinical diagnosis and treatment of diseases.

Keywords NSCT · Medical image fusion · PCNN · Computed tomography

1 Introduction

Along with the development of the imaging technology, a large number of imaging systems with different functions emerge in the medical field and the resulting images provide a reliable basis for clinical diagnosis. Based on different imaging principles and the difference of the instruments used, medical imaging is divided into anatomical and functional imaging in general. An important feature of anatomical imaging is that it can reveal the features of the lesions that are easily overlooked; while the main application of functional imaging is to show the metabolic effects of various tissues and organs in the human body, which, however, is unable to provide a more detailed imaging environment for the detailed features of tissues and organs due to its low resolution. Image fusion is to filter and organically integrate the target information, that is, effective feature points, in two

or more images based on certain rules, on the basis of which, an image with complete information in higher resolution is obtained, making it easier for observers to accept image information.

With the fusion method determined, image fusion methods can be divided into two kinds of methods based on their modes: the Spatial Domain Based and the Transform Domain Based image fusion methods. Although image fusion can be realized through both the spatial domain based and the transform domain based algorithms, by comparison, the former is less robust and susceptible to noises; while the later can also apply different fusion rules purposely to improve the fusion effect besides overcoming the above disadvantages (Du et al. 2016).

The so-called medical image fusion is that a variety of single-modal medical images are integrated comprehensively to describe the morphological structure and metabolic status of lesions in full perspective and provide a more reliable basis for the diagnosis and treatment of diseases. A common medical image fusion is the fusion of single-mode computed tomography (CT) images and magnetic resonance images (MRIs). CT imaging is based on the differences of human bodies in absorbing X-rays. It has a high resolution and can capture the human bone structure. MRI imaging is based on the nuclear magnetic resonance technology. It can capture the soft tissue information of human body but

✉ Jing Zhao
30026@sdwu.edu.cn
Xinhua Li
10013@sdwu.edu.cn

¹ Network Information Center, Shandong Women's University, Jinan, China

² School of Data and Computer Science, Shandong Women's University, Jinan, China

with a lower resolution. The complementarity of CT and MR images makes the fusion of the two possible (Li et al. 2018).

2 Related works

So far, many scholars have conducted in-depth researches in the field of medical image fusion.

Zhou et al. (2017) proposed a DTCWT-based lung cancer CT/PET image adaptive fusion algorithm in the Pilella framework. The reconstructed fusion image can better highlight the edge and texture information of the lesion. Fei et al. (2017) proposed a multi-modal medical image fusion mechanism based on sparse representation and decision mapping which not only increased the speed of the algorithm, but also improved the quality of the fusion results. Shen et al. (2013) introduced a new medical image fusion algorithm. It had a great improvement in reducing noise interferences and increasing information acquisition.

Huang et al. (2017) proposed a new medical image fusion method combined with the non-subsampling shear wave transform (NSST) and the spiked cortical model (SCM) which better suppressed the pixel distortion and preserved the information of the source image. Qiu et al. (2017) built a sparse representation model for the image based on the traditional KSVD dictionary learning algorithm in order to solve the unsatisfactory image expression ability and poor fusion effect of dictionary learning. Benjamin and Jayasree (2018) proposed an image fusion method based on cascaded PCA and translation invariant wavelet transform. Experimental results show that the fusion framework has better performance in visual and quantitative evaluations.

Aishwarya and Thangammal (2018) proposed an adaptive dictionary learning algorithm for multi-modal medical image fusion. By discarding zero information blocks and estimating the remaining image patches with MSF, useful information blocks were separated for dictionary learning. In this way, not only the amount of computation is small, but the fused image was in high quality as well. Lin et al. (2013) extracted the geometric flow and bandelet coefficients of a single-modal image through bandelet transform, fused and optimized the geometric flow with pulse coupling neural network (PCNN) and sparse similarity and fused the updated bandelet coefficients following the absolute maximum rule. The fusion image after the inverse transformation has excellent visual effects with good objective indicators.

In addition to the traditional fusion methods mentioned above, deep learning has achieved good results in the image processing field with its powerful feature extraction and data expression capabilities and has also been widely used in multimodal medical image fusion in recent years. For example, Liu et al. (2017) proposed a deep learning strategy for medical image fusion. The strategy uses the Laplacian

Pyramid to reconstruct the image in the fusion process after generating a weighted map of the source image using the deep network; Luo et al. (2018) used an improved CNN model to perform multi-view fusion on MR images to estimate the volume of ventriculus sinister; Liang et al. (2019) designed a multi-layer cascaded fusion network (MCFNet) to perform feature extraction, feature fusion and image reconstruction through a CNN and DECN based end-to-end network. Chen et al. (2019) proposed a multimodal deep learning fusion network, MultiFuseNet, which establishes auxiliary diagnosis of cervical dysplasia using the multimodal data from cervical screening results. Singh and Anand (2020) proposed a multimodal medical neural image fusion model based on CNN and PCA clustering. It can capture the spatial information of the images of various modes effectively, maintain their spatial consistency and suppress noises and artifacts.

Nevertheless current medical image fusion algorithms still have many problems, such as low contrast and low resolution of fused images, unclearness of small features, and the impact of image registration errors on the fusion results are also ignorable. In view of the above situations, a medical image fusion algorithm in the NSCT transform framework is proposed in this paper.

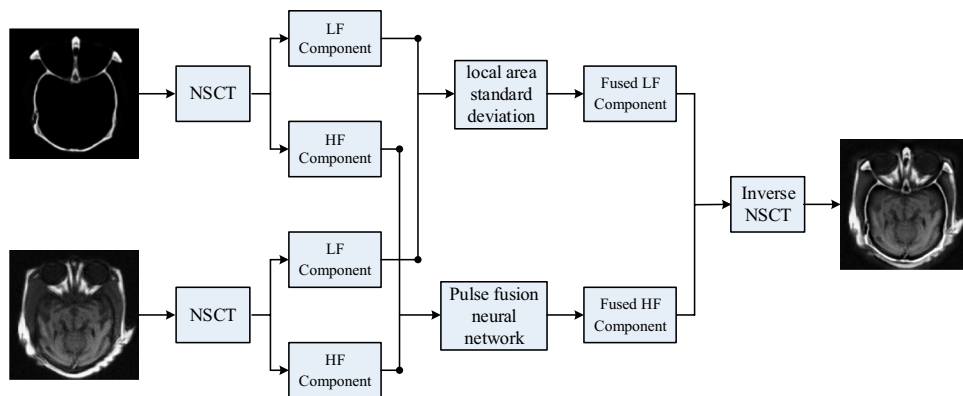
3 Framework

Firstly, the NSCT transformation is performed on the CT and MRI source images respectively to get the decomposed low-frequency sub-band coefficients and the high-frequency sub-band coefficients in all directions. Secondly, the local area standard deviation method is chosen for the fusion of low-frequency sub-band components and the processing method for the high-frequency subband is to establish an adaptive pulse coupled neural network model and input it into the pulse coupled neural network for multiple iterative calculations. The high-frequency subband coefficients are selected based on the ignition times of neuron with the maximum ignition times as a reference. Thirdly, the final fused image is achieved through inverse NSCT transform from the fused low-frequency and high-frequency subband components. The flowchart for the new algorithm proposed in the paper shown in Fig. 1.

4 NSCT theory

Although contourlet transform (Do and Vetterli 2002) is better than wavelet transform in multi-scale and multi-directionality, the up-sampling and down-sampling operations used in its processing procedure make it prone to spectral aliasing, which will be a more obvious Gibbs phenomenon if

Fig. 1 The framework for the algorithm proposed in this paper



it is displayed in the fused image. The non-subsampled contourlet transform (NSCT) (Da Cunha et al. 2006) growing up on the basis of contourlet transform has avoided similar operations during processing so that it not only inherits the excellent characteristics of contourlet transform, but also got the translation invariant characteristics. Translation invariance is applicable for image fusion processing, which not only reduces the influence of image registration errors on the fused image, but also eliminates the Gibbs phenomenon in the fusion result in the fusion process.

Non-subsampled Pyramid Filter Banks (NSPFB) and Non-subsampled Directional Filter Banks (NSDFB) are the key components of NSCT, which realize the multi-scale and multi-direction decompositions of input source images. Thus, multi-scale decomposition is in two steps: the first step is scale decomposition, which is implemented using NSPFB—after the source image is input into NSPFB, low-frequency and high-frequency subband components are obtained respectively; the second step is directional decomposition, which is implemented using NSDFB—the high-frequency subband obtained in the first step is further decomposed to obtain the subband components in each direction of the high-frequency subband. See the decomposition process and the frequency domain division of NSCT in Fig. 2.

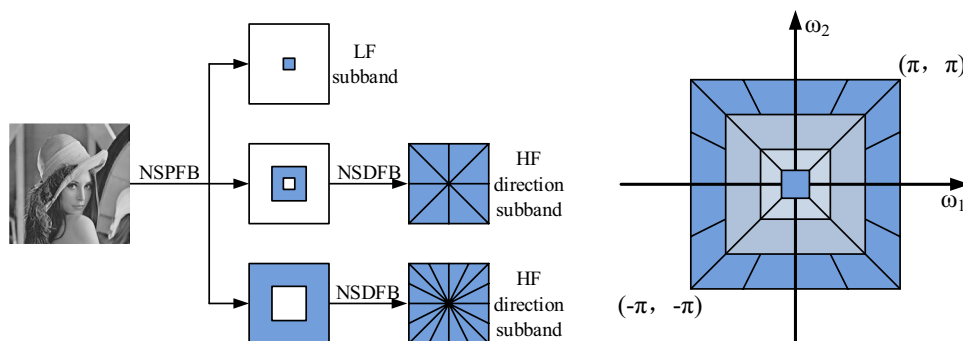
(1) Nonsubsampled Pyramid Filter Banks (NSPFB)

The 2D two-channel filter bank is used to solve the non-subsampled pyramid decomposition. For multi-scale decomposition, the low-frequency components obtained from the previous scale decomposition are input into a low-pass filter after upsampling and the high-frequency components obtained from the previous scale decomposition are inputted into a high-pass filter after upsampling for corresponding low-pass and high-pass filtering to obtain the low frequency and high frequency subband components after pyramid decomposition. In the process, the filter of the subsequent state is achieved after an interpolation operation is performed on the filter of the previous stage. A low-pass subband and a high-pass subband are obtained after the source image pass the first-level decomposition. Hence, if one subband can be achieved after the image passes a stage decomposition, what is more important is that all subband images are consistent with the source image in size and scale. See the structure of NSPFB in Fig 3, where $H_0(z), H_1(z)$ and $G_0(z), G_1(z)$ meet the Bezout identity:

$$H_0(Z)G_0(Z) + H_1(Z)G_1(Z) = 1. \tag{1}$$

(2) Nonsubsampled Directional Filter Banks (NSDFB)

Fig. 2 NSCT decomposition and frequency distribution



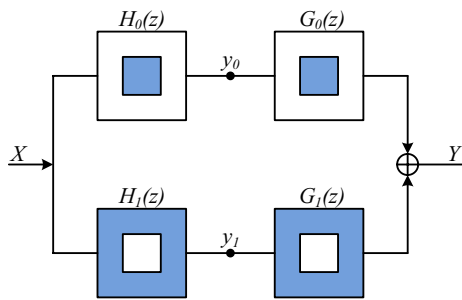


Fig. 3 Nonsubsampled Pyramid Filter Banks (NSPFB)

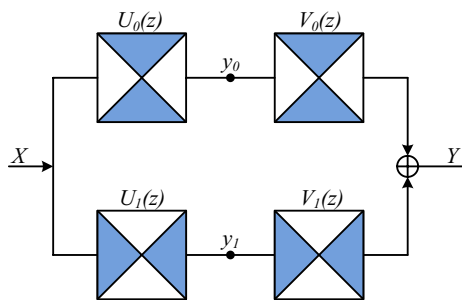


Fig. 4 Nonsubsampled Directional Filter Banks (NSDFB)

A filter bank with a fan-shaped support interval in the frequency domain is a component that implements nonsubsampled direction filtering. Similar to the operation of the above-mentioned NSPFB, the previous-stage filter is interpolated with different sampling matrices to obtain the next-stage filter, and then this filter is used to decompose the sub-band components after the previous-stage direction decomposition to obtain the next Level subband component. Therefore, the direction decomposition of the high-pass subband components on each scale can be obtained through NSDFB, and the directional subband components can be obtained, and the size of all directional subband images is consistent with the size of the source image. The structure of NSPFB is shown in Fig. 4, where $U_0(z)$, $U_1(z)$ and $V_0(z)$, $V_1(z)$ also meet the Bezout identity:

$$U_0(z)V_0(z) + U_1(z)V_1(z) = 1. \tag{2}$$

5 Fusion rules

5.1 Low-frequency subband fusion rules

At present, the fusion method selected for most low-frequency subband coefficients is a simple averaging method

generally. This method has a great impact on the contrast of fused images. Sometimes the target information in the source image cannot be completely extracted into the fused image. The local area standard deviation could represent the image gray intensity changes in local areas, while the parts with obvious gray level changes usually reflect image characteristics. So the important characteristics in images can be extracted according to the feature. This algorithm uses the local area standard deviation method as the fusion rule for low frequency subband components.

Firstly, calculate the local area standard deviation $E_j^C(m, n)$ and $E_j^M(m, n)$ of the low-frequency subband component $L_j^C(m, n)$ and $L_j^M(m, n)$, where $N_1 \times M_1$ represents the size of the neighborhood, and 3×3 or 5×5 are selected. The details are as follows:

$$E_j^C = \sqrt{\frac{\sum_{i=-(N_1-1)/2}^{i=(N_1-1)/2} \sum_{j=-(M_1-1)/2}^{j=(M_1-1)/2} [L_j^C(m+i, n+j) - \overline{L_j^C(m, n)}]^2}{N_1 \times M_1}} \tag{3}$$

$$E_j^M = \sqrt{\frac{\sum_{i=-(N_1-1)/2}^{i=(N_1-1)/2} \sum_{j=-(M_1-1)/2}^{j=(M_1-1)/2} [L_j^M(m+i, n+j) - \overline{L_j^M(m, n)}]^2}{N_1 \times M_1}} \tag{4}$$

Then, Formulas (3) and (4) are applied to select the low frequency subband coefficients after fusion. The specific procedures are to compare the difference between the standard deviation of the local areas of the two input medical images and the threshold. When the difference is larger, the coefficient of the image with a large value is taken; when the threshold is larger, the average value of the low-frequency subband coefficients of the two images are taken. The threshold is valued between 0.1 and 0.3:

$$L_j^{CM} = L_j^C(m, n) \times 0.5 + L_j^M(m, n) \times 0.5. \tag{5}$$

If $E_j^C - E_j^M > th$, then $L_j^{CM} = L_j^C(m, n)$; if $|E_j^C - E_j^M| > th$, then $L_j^{CM} = L_j^C(m, n) \times 0.5 + L_j^M(m, n) \times 0.5$; if $E_j^C - E_j^M < th$, then $L_j^{CM} = L_j^M(m, n)$. Through the above computation, the low frequency coefficient after fusion is obtained.

5.2 High-frequency subband fusion rules

To better extract the linear features such as contours, edges and textures contained in the original high-frequency subbands during the fusion of high-frequency subband coefficients, the fusion rules is designed based on the improved pulse coupled neural network (PCNN) in this paper. The specific method is that the ignition times of neurons is selected to determine the high-frequency subband coefficient after fusion.

PCNN is a network model established by simulating the mechanism of the animal cerebral cortex system. It has the characteristics of global coupling, spatial proximity and synchronous excitation. It is widely used in image denoising, pattern recognition and image segmentation and fusion fields. The PCNN model at the normal scale is highly complicated with a large number of parameters, moreover, the correlation between the results used for medical image processing and each parameter is ambiguous, therefore using the simplified PCNN model, not only the original characteristics of the model can be kept, the parameters of the model can be reduced effectively as well (Tan 2018), see the simplified PCNN model in Fig. 5.

The basic units of PCNN are neuron. Each neuron is composed of three components: a receiving domain, a modulation domain and a pulse generating domain. Multiple neurons are interconnected to form feedback. The receiving domain receives the external input signal, and then the obtained signal is multiplied in the modulation domain and the final output pulse is achieved in the pulse generation domain. The basic process is as follows: the receiving domain receives signals from the feedback domain and the link domain; these signals enter the modulation domain through Channel L and F. The L_{ij} in Channel L is multiplied with the F_{ij} in Channel F after being multiplied with the link strength and pluses the offset to obtain the U_{ij} , and enters the generation domain; the pulse generation domain includes a pulse generator and a comparator, which determines whether

a high-level pulse (i.e., ignition) is generated by setting a threshold. The mathematical expression for the simplified PCNN is as follows:

$$F_{ij}[n] = I_{ij}[n] \tag{6}$$

$$L_{ij}[n] = e^{-\alpha_L} L_{ij}[n - 1] + V_L \sum_{pq} W_{ij,pq} Y_{ij}[n - 1] \tag{7}$$

$$U_{ij}[n] = F_{ij}[n](1 + \beta L_{ij}[n]). \tag{8}$$

If $U_{ij}[n] \geq \theta_{ij}[n]$, then $Y_{ij}[n] = 1$; if $U_{ij}[n] < \theta_{ij}[n]$, then $Y_{ij}[n] = 0$, so

$$\theta_{ij}[n] = e^{-\alpha_\theta} \theta_{ij}[n - 1] + V_\theta Y_{ij}[n] \tag{9}$$

where F_{ij} represents feedback input, L_{ij} is link input, I_{ij} represents external stimulus, W represents weight coefficient, U_{ij} represents internal activity term, θ_{ij} represents dynamic threshold, Y_{ij} is pulse output; α_L, α_θ are attenuation coefficients; V_L, V_θ are the link input magnification factor and the threshold magnification factor respectively; β represents the internal active link strength factor; and ij is the neuron position.

When a pulse coupled neural network (PCNN) is operating on an image, one neuron corresponds to one pixel. The external stimulus of the neuron I_{ij} is the gray value of the pixel. The output signals of the coupling and feedback inputs of two subsystems are modulated and multiplied to obtain the internal activity item U_{ij} , which is then compared with the dynamic threshold θ_{ij} : if $U_{ij} > \theta_{ij}$, the neuron triggers ignition, $Y_{ij} = 1$; otherwise, $Y_{ij} = 0$, then, a 2D matrix is output by the number of ignitions. The value of each element in the matrix represents the number of ignitions of the neurons in the PCNN network at the pixel. The Lena image is input into the above network model. The ignition maps output after 50 and 100 iterations are shown in Fig. 6 respectively.

The original high-frequency subband coefficients are $H_{j,l,k}^C(m, n)$ and $H_{j,l,k}^M(m, n)$ respectively. The fusion rule for designing high-frequency subbands based on the intensity of neuron ignition is as follows:

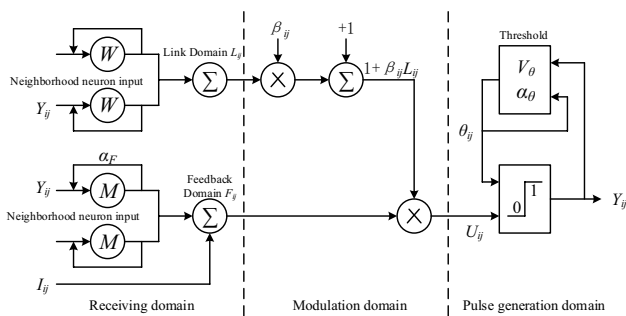
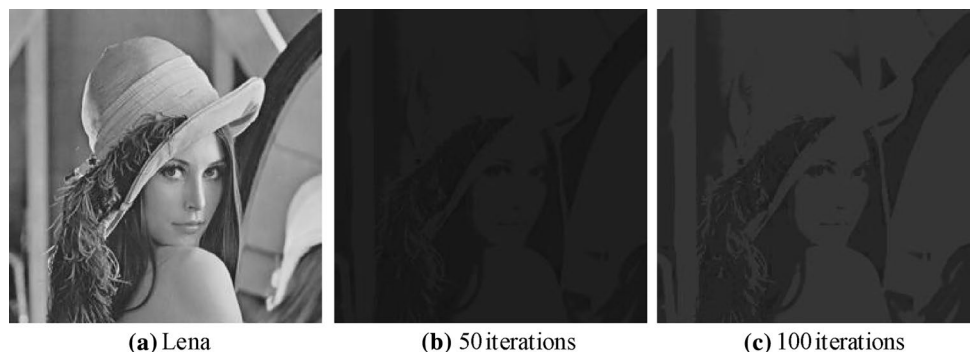


Fig. 5 Simplified PCNN model

Fig. 6 Lena original image and ignition map



If $\zeta_j^C(m, n) > \zeta_j^M(m, n)$, then $H_{j,l,k}^{CM}(m, n) = H_{j,l,k}^C(m, n)$; if $\zeta_j^C(m, n) < \zeta_j^M(m, n)$, then $H_{j,l,k}^{CM}(m, n) = H_{j,l,k}^M(m, n)$, where $\zeta_j^C(m, n)$ and $\zeta_j^M(m, n)$ are the ignition matrices of the original high-frequency subbands respectively, and $H_{j,l,k}^{CM}(m, n)$ is the fused high-frequency subband coefficient.

6 Experimental results and analysis

After the algorithm proposed in this paper is determined as a feasible research direction, in order to explore its effectiveness, four sets of CT and MRI gray scale experimental images which have received registration check are used for fusion operation. The experimental platform adopts an Intel Core i7-5500U CPU, a PC with 4GB of memory, a Win 10 system and a simulation environment of Matlab 2013b. Since the algorithm explored in this paper is based on the transform domain, the Contralet Transform (CL) algorithm, the Continuous Wavelet Transform (CWT) algorithm and the algorithm named MSGA in Shen et al. (2013) are the best choice for fusion effect comparison. After that, due to the diversity of visual senses and medical conditions, the subjective and objective two-way evaluation methods are chosen to verify the superiority of the effects of the fused experimental images. In order to verify the above algorithms, four sets of CT and MRI gray scale images which have received strict registration are selected in the experiment to compare the fusion results. The experimental platform adopts an Intel Core i7-5500U CPU, a PC with 4GB of memory, a Win 10 system and a simulation environment of Matlab 2013b.

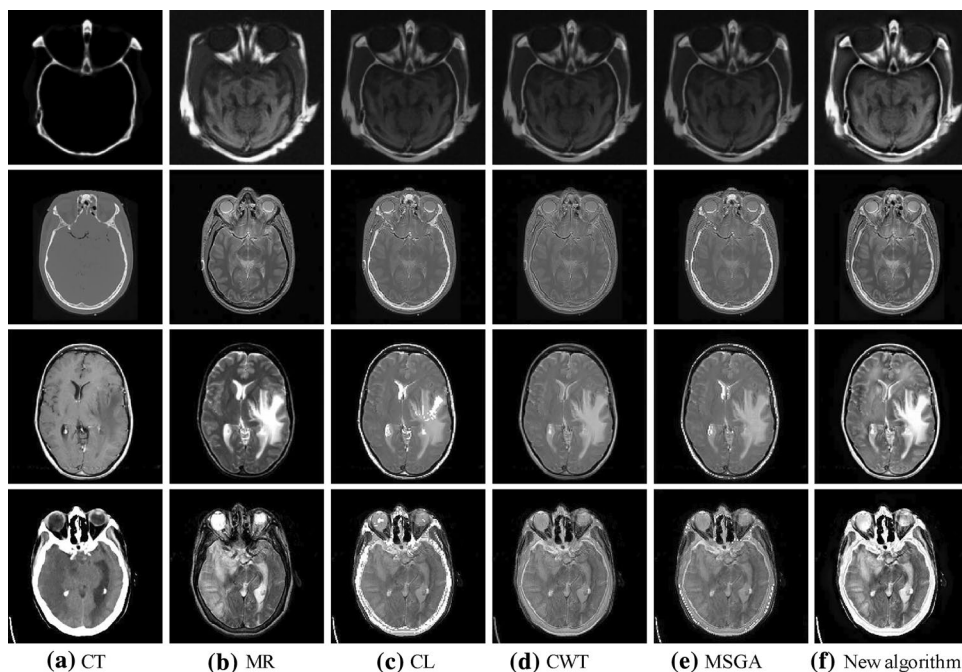
Since the Contralet Transform (CL) algorithm, the Continuous Wavelet Transform (CWT) algorithm and MSGA (Shen et al. 2013) are all transform domain, and the MSGA (Shen et al. 2013) has good effect in medical image fusion. So they are chosen to compare the fusion effects and the subjective and objective evaluations are used to evaluate the quality of the fused image comprehensively.

6.1 Subjective evaluation

The subjective evaluation is to make a qualitative evaluation of the quality of each image after fusion based on the visual effects of the observer. It is a very intuitive method to judge the fusion effect. The results of the algorithm proposed in this paper and other fusion methods are shown in Fig. 7. The images in each row are a set of images with the same CT/MR source images with their fusion images for comparison and the images in each column are a set of images in the same type, where (a) are CT images mainly displaying skeletal structure information, (b) are MR images focusing on displaying the detailed images of the soft tissues inside the organ, (c)–(e) are fusion images of the contrast algorithms and (f) are the fusion images of the algorithm proposed in this paper.

It can be seen from the comparison in the first set of images that the five fusion algorithms can all fuse the images properly, but compared with the algorithm proposed in this paper, the other images are slightly poorer in the overall contrast, especially the parts at the contours of fusion images and the peaks on both sides of the fusion images from the CT images; while the fusion image obtained by the algorithm

Fig. 7 Comparisons of four sets of images in fusion



proposed in this paper not only can highlight the contour information from the CT, but also can retain the characterization of the detailed information in the cavity. In the second set of images, the fusion effect of the CWT method is general, while the CL method, MSGA (Shen et al. 2013) and the algorithm proposed in this paper are more obvious in characterizing the contour information and the algorithm proposed in this paper characterizes the intra-cavity tissue more closely to the MR source image with a better fusion effect. Similar to the second sets of images, in the third set of images, all the methods in contrast are unable to reflect the part of the contour information from the CT image or display the intracranial gray matter perfectly except the algorithm proposed in this paper, which can synthesize the salient features of the two source images and fully represent the details of the images. In the fourth set of images, the fusion effects of the CWT method and the method MSGA (Shen et al. 2013) are poorer. The white contours and eye contours in the source images can hardly be reflected completely. Although the white contour can be seen in the fusion result of the CL method, the contour line in it is intermittent. The fusion effect is average. Only the method proposed in this paper can characterize the tissue structure information in the cavity in detail while highlighting the contour lines properly. The fusion is good.

6.2 Objective evaluation

Due to the diversity of medical conditions, subjective evaluation will also reflect its diversity. Under the joint action of the human visual sensitivity and the diversity, subjective evaluation is incomplete and has a certain error rate. In this case, the experimental results of the fused images are analyzed by certain objective evaluation criteria. In terms of the characteristics of medical image fusion, five indicators including information entropy (IE), mutual information (MI), average gradient (AvG), peak signal-to-noise ratio (PSNR) and the correlation coefficients (CC) are used to identify the quality of the fusion images multi-directionally (Du et al. 2016). IE characterizes the amount of information contained in the fused image. The larger the IE, the larger the amount of information. MI represents the amount of information transferred from the source image to the fused image. The larger the MI, the better the fusion effect. AvG indicates the gradient information of image, which reflects specifically the details and textures contained in the image. The larger the AvG, the higher the sharpness of the details. CC characterizes the linear correlation between fused and source images. The larger the CC, the higher the correlation between two images and the better the fusion effect. The specific index values are shown in Table 1.

It can be seen from the above table that the algorithm proposed in this paper has certain advantages on all the five

Table 1 Objective evaluation indicators for four sets of images

Image sets	MI	IE	AvG	CC
The 1st set				
CL	2.4931	6.0236	5.7570	0.8130
CWT	2.3576	6.0522	5.9721	0.8135
MSGA (Shen et al. 2013)	2.1259	6.0489	5.8625	0.8126
New algorithm	2.6112	6.9123	9.7769	0.9197
The 2nd set				
CL	1.3777	4.9918	9.0394	0.8894
CWT	1.3750	4.9316	7.6113	0.9618
MSGA (Shen et al. 2013)	1.1800	4.9894	9.7463	0.8883
New algorithm	1.3862	5.2948	9.5840	0.9122
The 3rd set				
CL	1.2638	4.4412	9.8074	0.9041
CWT	1.2222	4.4083	6.9423	0.9372
MSGA (Shen et al. 2013)	1.1972	4.7527	10.3962	0.9122
New algorithm	1.2481	5.1924	10.9830	0.9369
The 4th set				
CL	0.9141	4.6538	13.6224	0.7539
CWT	0.9370	4.6740	10.1941	0.8734
MSGA (Shen et al. 2013)	0.8611	4.9140	12.9502	0.8490
New algorithm	0.9351	5.1777	14.2581	0.8998

indicators such as IE, MI, AvG and CC, especially in IE and AvG. The algorithm proposed in this paper has a large lead over in the four sets of images, 0.8887 and 4.0640 respectively at the maximum, showing obvious advantages, which indicates that the fusion image of the algorithm contains rich information and a lot of details. This is more consistent with the purpose of medical image fusion of extracting significant detailed features for accurate diagnosis and treatment. Seeing from the correlation coefficient CC and the mutual information MI, the linear correlation between the images fused using the algorithm proposed in this paper and the source images are all about 90%, the mutual information has a slight advantage over other methods and the evaluation results are generally consistent with the subjective evaluation, indicating that the fusion effect of the algorithm proposed in this paper is more efficient and feasible than other methods.

7 Conclusion

A medical image fusion algorithm based on the NSCT framework is presented in this paper. It decomposes the CT and MR source images into high-frequency and low-frequency subband components through the NSCT in multi-scale geometric transformation. For the low frequency subband component, the corresponding fusion rule is selected using the local area standard deviation method; for the high frequency subband component, a pulse coupled neural

network is established to choose the high frequency subband coefficient taking the maximum ignition times of neurons as the rule; after that the fused image is obtained through the reconstruction of NSCT image. The experimental results show that the fusion results of the algorithm proposed in this paper not only highlight the contour information in the CT image properly, but also characterizes the tissue structure information in the MR image in detail and improves the contrast of the fused image as well; in addition, it also has advantages in comparison with other fusion methods in terms of the objective evaluation indicators including information entropy (IE), mutual information (MI), average gradient (AvG), peak signal-to-noise ratio (PSNR) and the correlation coefficients (CC). This algorithm can improve the image fusion quality significantly and it has certain advantages in both visual effects and objective evaluation indexes, which provides a more reliable basis for clinical diagnosis and treatment of diseases.

Acknowledgements This work was supported by Project of the Shandong Medicine and Health Science Technology Development Plan under Grant 2017WSB04071, Shandong Province Science and Technology Development Plan Project under Grant 2014GSF118086.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

References

- Aishwarya N, Thangammal CB (2018) A novel multimodal medical image fusion using sparse representation and modified spatial frequency. *Int J Imaging Syst Technol* 28(3):175–185
- Benjamin JR, Jayasree T (2018) Improved medical image fusion based on cascaded PCA and shift invariant wavelet transforms. *Int J Comput Assist Radiol Surg* 13(2):229–240
- Chen T, Ma X, Ying X (2019) Multi-modal fusion learning for cervical dysplasia diagnosis. In: *Proceedings of international symposium on biomedical imaging*, pp 1505–1509
- Da Cunha AL, Zhou J, Do MN (2006) The nonsubsampling contourlet transform: theory, design, and applications. *IEEE Trans Image Process* 15(10):3089–3101
- Do MN, Vetterli M (2002) Contourlets: a directional multiresolution image representation. In: *Proceedings of international conference on image processing*, vol 1, pp 357–360
- Du J, Li W, Lu K (2016) An overview of multi-modal medical image fusion. *Neurocomputing* 215(26):3–20
- Fei Y, Wei G, Zongxi S (2017) Medical image fusion based on feature extraction and sparse representation. *Int J Biomed Imaging* 2017:1–11
- Huang Z, Ding M, Zhang X (2017) Medical image fusion based on non-subsampling shearlet transform and spiking cortical model. *J Med Imaging Health Inform* 7(1):229–234
- Li Y, Zuo Z, Jin G, Ke S (2018) The research progress of the wavelet transform based PET/CT image fusion algorithm. *China Med Imaging Technol* 34(08):1267–1270
- Liang X, Hu P, Zhang L, Sun J, Yin G (2019) MCFNet: multi-layer concatenation fusion network for medical images fusion. *IEEE Sens J* 19(16):7107–7119
- Lin Z, Yan J, Yuan Y (2013) The sparse representation and PCNN based multi-modal image fusion. *J Shandong Univ (Eng Sci Ed)* 43(4):13–17
- Liu Y, Chen X, Cheng J, Peng H (2017) A medical image fusion method based on convolutional neural networks. In: *Proceedings of 20th international conference on information fusion*, pp 1–7
- Luo G, Dong S, Wang K, Zuo W, Cao S, Zhang H (2018) Multi-views fusion CNN for left ventricular volumes estimation on cardiac MR images. *IEEE Trans Biomed Eng* 65(9):1924–1934
- Qiu H, Li H, Yu Z (2017) The sparse representation based medical image fusion. *Sens Microsyst* 10(1):330–341
- Shen Y, Dang J, Wang Y (2013) A new multi-scale geometric analysis based image fusion method. *Optoelectron Laser* 12:2446–2451
- Singh S, Anand RS (2020) Multimodal medical image fusion using hybrid layer decomposition with CNN-based feature mapping and structural clustering. *IEEE Trans Instrum Meas* 69(6):3855–3865
- Tan Z (2018) A study on the pulse coupled neural network based image segmentation and fusion method. Shenzhen University, Shenzhen
- Zhou T, Lu H, Wei X (2017) A Piella framework and DT-CWT based lung cancer PET/CT adaptive fusion algorithm. *J Univ Sci Technol China* 01:13–20

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.