이중스케일분해기와 미세정보 보존모델에 기반한 다중 모드 의료영상 융합연구

장영매, 이효종
전북대학교 컴퓨터공학부
zym1megan@gmail.com, hlee@jbnu.ac.kr

Multimodal Medical Image Fusion Based on Two-Scale Decomposer and Detail Preservation Model

Yingmei Zhang, Hyo Jong Lee*,
*Dept. of Computer Science, Jeonbuk National University
*Corresponding author

Abstract

The purpose of multimodal medical image fusion (MMIF) is to integrate images of different modes with different details into a result image with rich information, which is convenient for doctors to accurately diagnose and treat the diseased tissues of patients. Encouraged by this purpose, this paper proposes a novel method based on a two-scale decomposer and detail preservation model. The first step is to use the two-scale decomposer to decompose the source image into the energy layers and structure layers, which have the characteristic of detail preservation. And then, structure tensor operator and max-abs are combined to fuse the structure layers. The detail preservation model is proposed for the fusion of the energy layers, which greatly improves the image performance. The fused image is achieved by summing up the two fused sub-images obtained by the above fusion rules. Experiments demonstrate that the proposed method has superior performance compared with the state-of-the-art fusion methods.

keywords: *multimodal medical image fusion; two-scale decomposer; detail preservation model.*

1. Introduction

With the development of various imaging devices, multi-modal medical image fusion has become an important research topic for obtaining accurate clinical information that can be provided to physicians for better diagnosis [1]. Relying only on a single modality medical image is insufficient for diagnosing a patient's condition. For example, the computed tomography (CT) image can capture tissue information of dense structures, such as bones and implants, but cannot provide detailed information of soft tissue. The magnetic resonance (MR) images display soft-tissue anatomy information with high spatial resolution but fail to detect human metabolic activity information. Therefore, the fusion of medical images of different modalities into an integrated image that includes the useful feature information of the two source images is in line with the times [2].

In the past few decades, multi-scale transformation (MST), the term, has become more and more popular in the field of image fusion since MST can decompose two source images into two sub-layers, each of which contains image features from the source images. Generally speaking, the fusion rules constructed based on these image features will result in a fusion image with good performance [3]. The more typical MST algorithms are complex wavelet transforms (CVT) [4], nonsubsampled contourlet transform (NSCT) [5], and dual-tree complex wavelet transform (DTCWT) [5]. Although these methods mentioned above can extract more image features, a fact that cannot be ignored is that the fused image has artifacts in the edge area, causing image distortion. This phenomenon occurs because the spatial consistency between images is not considered into the image decomposition process. As for the fusion rule, “absolute maximum” and “averaging” were the most commonly used single rules in earlier years. Their principle is to take the maximum and average values of all pixels in a sliding window of a certain scale. Obviously, this kind of rule will definitely damage the quality of the fused image. In recent years, the first application of sparse representation (SR) to image fusion by Yang et al. [6] is very popular because of its good effect. However, time-consuming has always been the weakness of the SR method due to the real-time training of a complete dictionary.

Inspired by the above contents, in this work, a novel

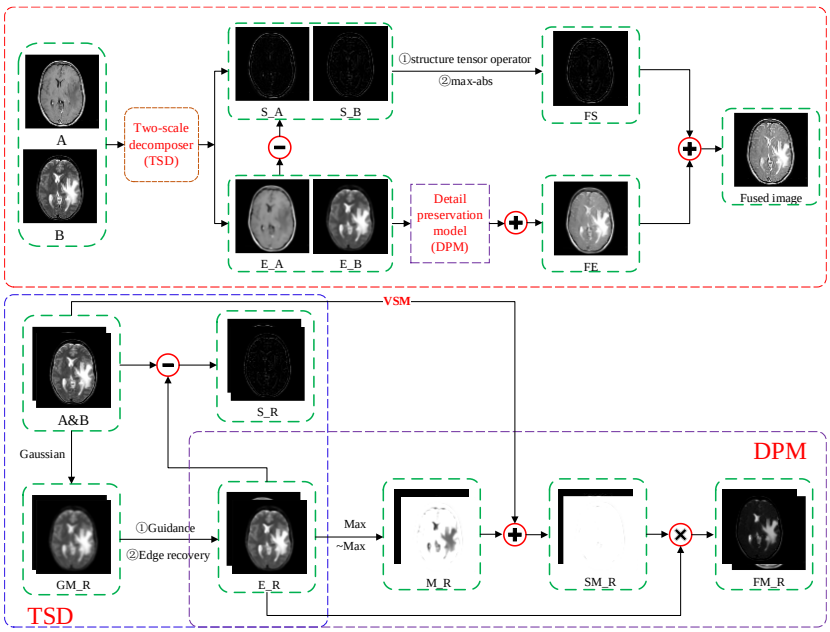


Fig. 1 The red dashed box represents the overall framework of the proposed method. The blue and purple dashed boxes show the details of TSD and DPM, respectively.

multimodal medical image fusion based on two-scale decomposer and detail preservation model is proposed, which obtain a good performance and high efficiency.

The remainder of this paper is listed as follows. Section II describes the proposed MMIF algorithm in detail. The experimental analysis is given in Section III. The conclusions are presented in Section IV.

2. The proposed method

2.1 overall framework

In this section, we will introduce the overall framework of the proposed method in detail, which are shown in Fig. 1. Assume that A and B represent the source images. The first step is to use the proposed two-scales decomposer (TSD) to decompose the source image into the energy layers(E_A/E_B) and structure layers(S_A/S_B). And then, STO and Max-abs are combined to fusing the structure layers, the detail preservation model is proposed for the fusion of the energy layers. The fused image is achieved by summing up the two fused sub-images obtained by the above fusion rule. The detailed description of the algorithm will be explained in the next section.

2.2 TSD

In this part, a TSD is proposed to guide the decomposition process. The first step is to use a weighted average Gaussian filter to smooth the image, the purpose of which is to maximize the useful features of the input image to the decomposition structure layer. Next, to recover some small-scale image features that were deleted by the Gaussian filtering operation, guidance filtering (GF) is applied, and thus the energy layers will be obtained through these two steps. Logically, the structure layer is generated by subtracting the energy layer from the source image. The algorithm of this part can be defined as:

$$GM\_R(j)=\frac{1}{Z_j}\sum_{i\in N(j)}\exp\left(-\frac{\|j-i\|^2}{2\sigma_s^2}\right)\times I(i) \tag{1}$$

$$Z_j=\sum_{i\in N(j)}\exp\left(-\frac{\|j-i\|^2}{2\sigma_s^2}\right)$$

$$E(j)=\frac{1}{Z_j}\sum_{i\in N(j)}\exp\left(-\frac{\|j-i\|^2}{2\sigma_s^2}-\frac{\|GM\_R(j)-GM\_R(i)\|^2}{2\sigma_r^2}\right)\times I(i) \tag{2}$$

$$Z_j=\sum_{i\in N(j)}\exp\left(-\frac{\|j-i\|^2}{2\sigma_s^2}-\frac{\|GM\_R(j)-GM\_R(i)\|^2}{2\sigma_r^2}\right)$$

$$S(j)=I(j)-E(j) \tag{3}$$

where I and j denote pixel coordinates, $GM\_R(*), I(*), N(j), \sigma_s$, and σ_r denote the output of the Gaussian filter, the input images, the set of neighbor pixels of i, spatial weight, and range weight, respectively. Z_j is the normalization operation. In this study, σ_s, σ_r are assigned as 2 and 0.05, respectively.

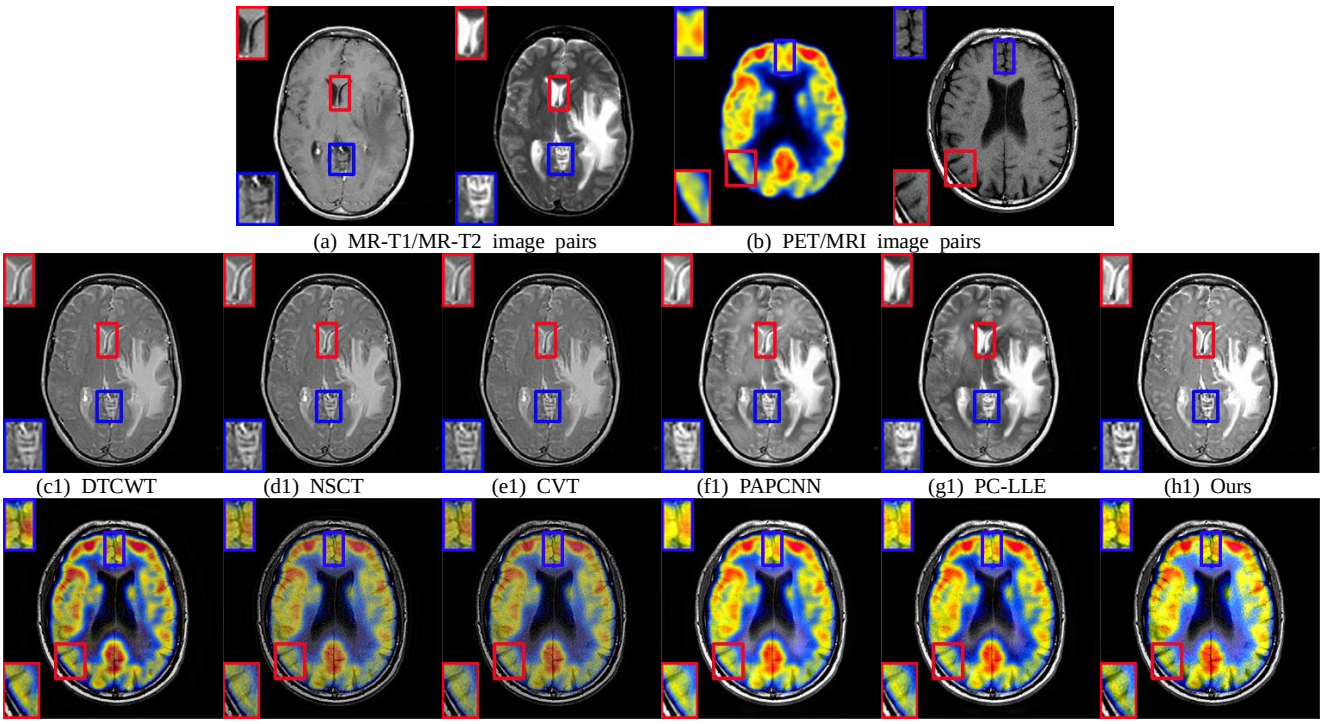
2.3 Fusion

As for the fusion part, structure tensor operator (STO) and max-abs are combined to fuse the structural layer containing rich gradient information. On the one hand, STO is proven to be able to extract image gradients effectively, on the other hand, max-abs can extract salient features. Therefore, combining these two advantages can ensure that the source image information is transferred to the fusion image to the greatest extent. The rule is designed as follows:

$$\begin{aligned} S\_R &= STO(R) \\ FS &= \max(abs(S\_A), abs(S\_B)) \end{aligned} \tag{4}$$

where $R\in\{A,B\}$, and FS denotes the fused images of the structure layer.

For the energy layer, a detail preservation model is proposed, which can keep the details well. We first take the



(a) MR-T1/MR-T2 image pairs (b) PET/MRI image pairs

(c1) DTCWT (d1) NSCT (e1) CVT (f1) PAPCNN (g1) PC-LLE (h1) Ours

(c2) DTCWT (d2) NSCT (e2) CVT (f2) PAPCNN (g2) PC-LLE (h2) Ours

Fig. 2. The fused images by using six different fusion methods with two typical image pairs. (a) is MR-T1/MR-T2 image pairs, (b) PET/MRI image pairs, (c1)-(h1) are the fused results by DTCWT, NSCT, CVT, PAPCNN, PC-LLE and ours on the MR-T1/MR-T2 image pairs. (c2)-(h2) are the fused results on the PET/MRI image pairs.

Table I The objective indexes on the image pairs from Fig. 2. Red: maximum values; Blue: second largest values.

Methods	MR-T1/MR-T2 image pairs						PET/MRI image pairs					
	CVT	DTCWT	NSCT	PAPCNN	PC-LLE	Ours	CVT	DTCWT	NSCT	PAPCNN	PC-LLE	Ours
Q^{ABF}	0.4498	0.6138	**0.6600**	0.6053	0.6570	**0.6813**	0.4532	0.6095	0.6630	0.6515	**0.6842**	**0.7767**
SSIM	0.4988	0.7379	**0.8016**	0.7285	0.7637	**0.8169**	0.4823	0.6862	0.6431	0.6869	**0.7543**	**0.9009**
AG	7.6474	7.3823	7.5654	**8.7383**	8.6799	**9.5689**	9.1202	8.9614	9.8729	10.2941	**10.3887**	**10.8902**
Q^{CB}	0.4953	0.6212	0.6556	0.6637	**0.6739**	**0.6756**	0.4559	0.5850	0.5699	0.5949	**0.6246**	**0.6852**

Max-abs of the energy layers to get the intensity maps(M_R) and then add them to the visual saliency map [8] corresponding to the input images to get a supplementary map (SM_R) to ensure a better structure. The rule is shown as:

$$SM\_R = \sum_{X \in R} (M\_X + V\_X)$$
$$FE = SM\_R \times E\_R \quad (5)$$

where FE denotes the fused energy layer.

Following the fusion rule, the final fusion image(F) is obtained through addition operation between FS and FE:

$$F = FS + FE \quad (6)$$

3. Experiments

To illustrate the effectiveness of the proposed method, in this section, five mainstream methods are compared with the proposed algorithm, together with four recognized objective indicators are introduced. They are CVT [4], NSCT [5], DTCWT [5], phase congruency and local laplacian energy (PC-LLE) [2], and adaptive pulse coupled neural network in nonsubsampled shearlet transform domain (PAPCNN) [1], respectively. Six metrics including edge reservation (QABF) [9], structure similarity-based (SSIM) [10], average gradient (AG) [11], and human vision perception-based metrics (QCB) [12]. The higher the indexes value, the better the algorithm performance. All the experiments are conducted in MATLAB R2017b. The dataset is downloaded from the Whole Brain Atlas database of Harvard Medical School, http://www.med.harvard.edu/aanlib/ home.html.

Fig. 2 shows the MR-T1/MR-T2 and PET/MRI images fused results produced by six different methods. From these fusion results, we can observe that the contrast of the results produced by DTCWT, NSCT and CVT are reduced a lot. While the contrast of PAPCNN and PC-LLE are significantly enhanced, but the detail texture of the organ tissue is missing compared to the proposed method (see the blue zoomed box). Moreover, our result has no artifacts, which can be proved from the red magnified area of Fig. 2(h1). As for the PET/MRI image pairs, a conclusion similar to Fig. 2(a) achieves that the image performance obtained by our method is better than other five since more details are extracted from the original images, which can be verified from the image features in the red and blue magnified area.

The objective indexes are also measured and are shown in

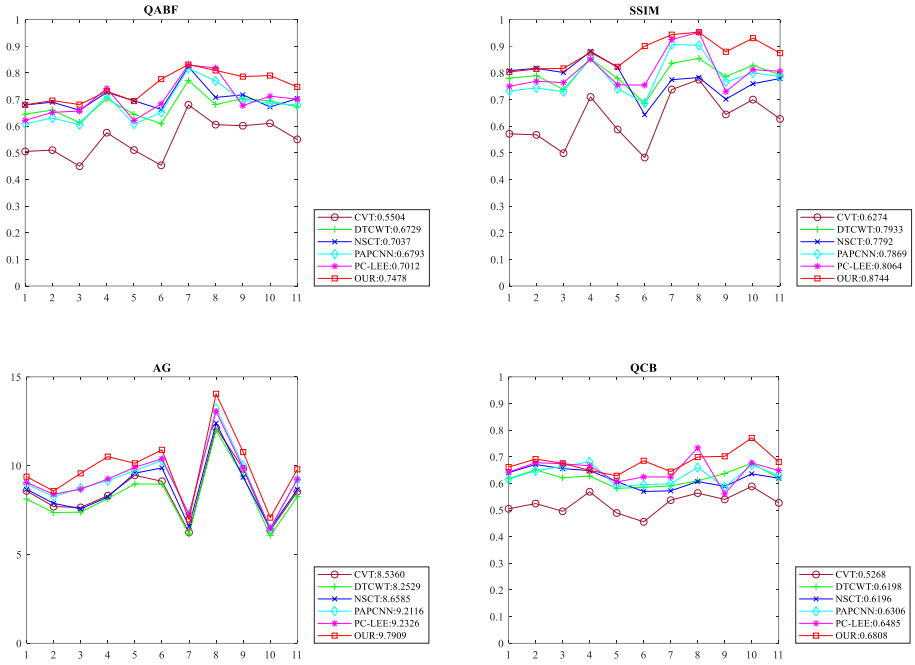


Fig. 3 The quantitative index diagram of 10 groups of images. The last column, that is, the eleventh column, is the average value of the first 10 groups of images.

Table I. From the results, our method achieves the maximum value of all indicators. At the same time, in order to further demonstrate the effectiveness of our method, we performed an average operation on 10 sets of images, as shown in Fig. 3. As can be seen from the figure, our results have achieved superior results in the vast majority of images including the average value (the last column).

4. Conclusion

In this paper, a novel multimodal medical image fusion is proposed. The input images are decomposed by the TSD method into the structure layers and energy layers. Two fusion rules are designed based on STO and DPM to fusing the corresponding image layers, respectively. The subjective results and objective indicators observed in the experiment show that our method has a better image performance compared to several mainstream comparison methods.

Acknowledgment

This research was supported by Basic Science Research Program through the NRF of Korea funded by the Ministry of Education (GR 2019R1D1A3A03103736).

REFERENCES

[1] M. Yin, X. Liu, Y. Liu, X. Chen, “Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain”, *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49-64, 2019.

[2] Z. Zhu, M. Zheng, G. Qi, D. Wang, Y. Xiang, “A Phase Congruency and Local Laplacian Energy Based Multi-Modality Medical Image Fusion Method in NSCT Domain”, IEEE Access, vol. 7, pp. 20811–20824, 2019.

[3] S. Li, X. Kang, L. Fang, J. Hu, H. Yin, “Pixel-level image fusion: a survey of the state of the art”, *Inf. Fusion*, vol. 33, pp. 100–112, 2017.

[4] R. Singh, A. Khare, “Fusion of multimodal medical images using Daubechies complex wavelet transform-a multiresolution approach”, *Inf. Fusion*, vol. 19, pp. 49–60, 2014.

[5] Y. Liu, S. Liu, Z. Wang, “A General Framework for Image Fusion Based on Multi-scale Transform and Sparse Representation”, *Inf. Fusion*, vol. 24, no. 1, pp. 147-164, 2015.

[6] B. Yang, S. Li, “Multifocus image fusion and restoration with sparse representation”, *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884-892, 2010.

[7] V. Prasath, R. Pelapur, G. Seetharaman, K. Palaniappan, “Multiscale structure tensor for improved feature extraction and image regularization”, *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6198–6210, 2019.

[8] Y. Zhai and M. Shah, “Visual attention detection in video sequences using spatiotemporal cues”, in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, pp. 815-824, 2006.

[9] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure", *Electronics Letters*, vol. 36, no. 4, pp. 308-309, 2000.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004.

[11] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, “Detail preserved fusion of visible and infrared images using regional saliency extraction and multiscale image decomposition,” *Opt. Commun.*, vol. 341, pp. 199–209, Apr. 2015.

[12] Y. Han, Y. Cai, Y. Cao, and X. Xu, “A new image fusion performance metric based on visual information fidelity,” *Inf. Fusion*, vol. 14, no. 2, pp. 127–135, Apr. 2013.