KIPS Trans. Comp. and Comm. Sys.
Vol.11, No.6 pp.185~192
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2022.11.6.185

Multimodal Medical Image Fusion Based on Double-Layer Decomposer and Fine Structure Preservation Model

Yingmei Zhang† · Hyo Jong Lee††

ABSTRACT

Multimodal medical image fusion (MMIF) fuses two images containing different structural details generated in two different modes into a comprehensive image with saturated information, which can help doctors improve the accuracy of observation and treatment of patients' diseases. Therefore, a method based on double-layer decomposer and fine structure preservation model is proposed. Firstly, a double-layer decomposer is applied to decompose the source images into the energy layers and structure layers, which can preserve details well. Secondly, The structure layer is processed by combining the structure tensor operator (STO) and max-abs. As for the energy layers, a fine structure preservation model is proposed to guide the fusion, further improving the image quality. Finally, the fused image can be achieved by performing an addition operation between the two sub-fused images formed through the fusion rules. Experiments manifest that our method has excellent performance compared with several typical fusion methods.

Keywords : Multimodal Medical Image, Image Fusion, Double-layer Decomposer, Fine Structure Preservation Model

복층 분해기와 상세구조 보존모델에 기반한 다중모드 의료영상 융합

장 영 매† · 이 효 종††

요 약

다중모드 의료영상 융합(MMIF)은 각기 다른 특징들을 나타내는 여러 종류의 모드의 이미지를 풍부한 정보가 포함된 하나의 결과 이미지로 통합하는 것이다. 이러한 의료영상 융합은 의사가 환자의 병변을 정확하게 관찰하고 치료하는 것을 도와줄 수 있다. 이러한 목적에 영향을 받아 본 논문에서는 복층 분해기 및 미세구조 보존 모델에 기반한 새로운 방법을 제안한다. 첫째, 복층 분해기를 사용하여 소스 이미지를 미세정보 보존의 특성을 갖는 에너지 층과 구조적 층으로 분해하였다. 둘째, 구조 텐서 연산자와 max-abs를 결합하여 구조적 층을 융합한다. 에너지 층의 융합을 위해 미세구조 보존 모델을 제안하였으며 이미지 융합성능을 크게 향상시킬 수 있었다. 마지막으로, 융합규칙을 통해 형성된 두 개의 융합된 하위 이미지를 합산하여 구축하였다. 실험을 통하여 제안된 방법이 현재까지 최첨단 융합 방법들과 비교하여 우수한 성능을 나타내는 것을 검증하였다.

키워드 : 다중모드 의료영상, 이미지 융합, 복층 분해기, 미세구조 보존 모델

1. Introduction

As we know, the images obtained by a single imaging device cannot meet the needs of medical research at this stage, which makes the fusion of multimodal medical images rise rapidly and become a popular subject. To date, its research findings have been strongly linked to accurate clinical information and better treatment judgment[1]. For example, the computed tomography (CT) image focus on dense structures' tissue information (bones and implants) but cannot obtain detailed information on their corresponding position. The magnetic resonance images (MRI) observe soft tissue anatomical information from a high spatial resolution perspective but have shortcomings in monitoring human metabolic activity information. Therefore, it is urgent to explore MMIF that can integrate the useful features of two raw images

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR 2019R1D1A3A03103736) and in part by project for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Ministry of SMEs and Startups in 20 (Grant No. S3114049), and by project for 'Customized technology partner' funded Korea Ministry of SMEs and Startups in 2022 (RS-2022-00155266).
※ 이 논문은 2021년 한국정보처리학회 ACK 2021의 우수논문으로 "이중스케일분해기와 미세정보 보존모델에 기반한 다중모드 의료영상 융합연구"의 제목으로 발표된 논문을 확장한 것임.
† 준 회 원 : 전북대학교 컴퓨터공학부 박사과정
†† 종신회원 : 전북대학교 컴퓨터공학부, CAIIT 교수
Manuscript Received : December 24, 2021
Accepted : February 3, 2022
* Corresponding Author : Hyo Jong Lee(hlee@jbnu.ac.kr)

into a comprehensive image[2]. The breakthroughs in MMIF can aid biomedical research and clinical diagnosis, such as routine staging/restaging, surgical navigation, radiotherapy planning, and future health prediction[3]. Once a fusion results are obtained, doctors can input them into the medical systems like percutaneous image-guided interventions and image-guided procedures for further diagnosis and analysis, and then give patients a treatment plan or medical advice based on professional judgment to protect their lives [4].

Currently, multi-scale transform (MST)-based frameworks are a widely popular application trend in image fusion. Because the framework follows the three parts of "decomposition + fusion + reconstruction", each part can independently design the corresponding decomposition (fusion or reconstruction) algorithm to enhance the overall performance of the fused image [5]. The more classic decomposition algorithms in MST are complex wavelet transforms (CVT) [6], nonsubsampled contourlet transform (NSCT) [7], and dual-tree complex wavelet transform (DTCWT) [7]. The above methods can preserve useful image information well through multi-layer decomposition, but it should be noted that there are artifacts in the edge area of the fused image, resulting in image distortion. Through analysis, we find that the lack of consideration of image spatial consistency during image decomposition leads to this situation. As for the second part (fusion), it is mainly to design appropriate fusion rules to assist in completing the main task. Common rules are "absolute maximum" and "averaging". Their principle is to operate directly on the image pixels without doing other transformation algorithms. Obviously, this rule loses image details to a certain extent. Excitedly, Yang *et al.* [8] put sparse representation (SR) into image fusion for the first time, and its effectiveness was proved by extensive research experiments. Nevertheless, the large amount of time required to train an over-complete dictionary has always been a weakness that SR methods cannot really address.

To deal with the above challenges, in this work, a MMIF based on double-layer decomposer and fine structure preservation model is proposed, which is expected to obtain fused images with high-quality.

The structure of this article is arranged below. Section 2 introduces the novel MMIF approach. The experimental analysis is discussed in Section 3. Section 4 draws conclusions.

2. The Proposed Method

2.1 General Framework

Fig. 1 outlines the general flowchart of our algorithm in detail with dashed boxes in different colors. Here, A and B refer to raw images. Firstly, the double-layer decomposer (DLD) is applied to decompose two raw images for obtaining the energy layers (E_A/E_B) and structure layers (S_A/S_B). Secondly, the structure layers are merged by combining STO [9] and max-abs, and a fine structure preservation model is proposed to guide energy layers fusion. In the end, the merged result is achieved by calculating an addition operation between two sub-fused images formed through the fusion rules in advance. The detailed description will be shown as below.

2.2 DLD

In this section, we will describe logically how to construct DLD and it actually has three steps. The image is first smoothed by utilizing a weighted average Gaussian filter to maximize the important image features of the structural layer. And then, guidance filtering(GF) is applied to recover some small-scale image features removed by the above filtering, followed by the energy layers will be obtained. In the end, the structure layer is obtained by subtracting between the raw image and the energy layer. The algorithm equations are expressed as below:

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

$$Z_j=\Sigma_{i\in N(j)}\exp(-\frac{\|j-i\|^2}{2\sigma_s^2}) \tag{2}$$

$$E(j)=\frac{1}{Z_j}\Sigma_{i\in N(j)}\exp(-\frac{\|j-i\|^2}{2\sigma_s^2}-\frac{\|GM\_R(j)-GM\_R(i)\|^2}{2\sigma_r^2})*I(i) \tag{3}$$

$$Z_j=\Sigma_{i\in N(j)}\exp(-\frac{\|j-i\|^2}{2\sigma_s^2}-\frac{\|GM\_R(j)-GM\_R(i)\|^2}{2\sigma_r^2}) \tag{4}$$

$$S(j)=I(j)-E(j) \tag{5}$$

where $R\in A,B$, i, j are the pixel positions. $GM\_R(*)$

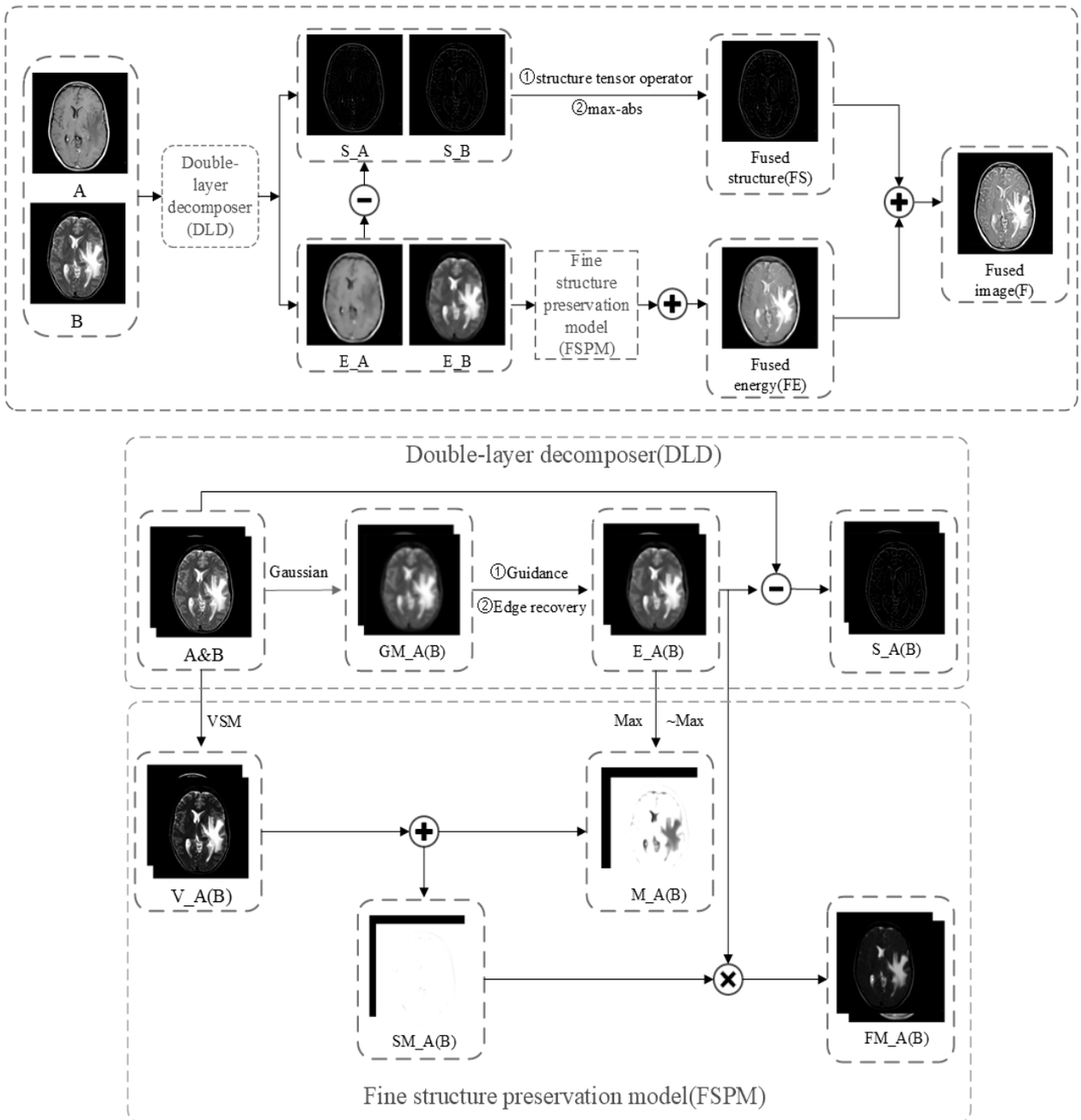

Fig. 1. The First Row(red) is the Overall flowchart. The Second(blue) and Third(orange) Row are DLD and FSPM Description, Respectively

signify the result of the Gaussian filtering. *I(i)* mean the raw images, and *N(i)* represent the set of neighbor pixels of *I*. σ_s show the spatial weight and it is designated as 2. σ_r denote the range weight and its value is equal to 0.05. Z_j indicate normalization.

2.3 Fusion

In the fusion process, we simultaneously employ STO and max-abs as a new strategy to merge the structural layer to preserve the gradient information inherited from the raw images. There are two main reasons: i)STO is a very effective tool in extracting structure gradients of images; ii)max-abs is sufficient to extract salient features. Therefore, this paper comprehensively utilizes these two operators into a fusion strategy, which can maximize the feature transfer from the raw image to the fused image. The fusion strategy is composed as below:

$$S\_R = STO(R) \tag{6}$$

$$FS = \max(abs(S\_A),\ abs(S\_B)) \tag{7}$$

where FS denotes the fused images of the structure layer.

Next, a fine detail preservation model (FSPM) is put forward for the energy layer, which has the merit of well-preserved details. The intensity maps (M_Rs) are obtained by operating Max-abs on the energy layers. Then M_Rs are added to the visual saliency map (VSM) [10] to get two supplementary maps (SM_Rs), which th purpose is to ensure a better structure. The rule appears as:

$$SM\_R = \sum_{X \in R}(M\_X + V\_X) \tag{8}$$

$$FE = SM\_R \times E\_R \tag{9}$$

where FE denotes the output result of energy layer.

At the end of the algorithm, the fused image(F) can be obtained by adding FS to FE:

$$F = FS + FE \tag{10}$$

3. Experiments

To verify the fusion performance of our method, six mainstream methods, including our method, are compared with each other, and six publicly recognized objective indicators are introduced. The six compared methods include CVT[6], NSCT[7], DTCWT [7], phase congruency and local Laplacian energy (PC-LLE) [2], and adaptive pulse coupled neural network (PAPCNN) [1], respectively. Six metrics comprise edge-based (QABF) [11], structure similarity-based (SSIM) [12], average gradient-based (AG) [13], and human vision perception-based (QCB) [14], the sum of the correlations of differences (SCD) [15] and spatial frequency (SF) [16]. The common convention is that the higher the metrics value, the better the method expressiveness. Meanwhile, all experiments are completed in the MATLAB R2017b environment, and the dataset was downloaded from the Whole Brain Atlas database of Harvard Medical School [17].

Fig. 2 shows several fusion results obtained by six compared methods on four different types of image datasets (MR-T1/MR-T2, CT/MRI, PET/MRI, and SPECT/MRI). From Fig. 2(a1)-(a6) obtained from acting on MR-T1/MR-T2 image pairs, we can infer that the brightness of the fused results generated by DTCWT, NSCT and CVT is dim, which means that very little salient information inherited from the infrared raw image is preserved in these methods, resulting in poor image quality. Although the brightness of PAPCNN and PC-LLE is greatly improved, the structure information of important tissues is missing to varying degrees compared with our method(see magnified blue box). Furthermore, our method produces a fused result with no artifacts, and this phenomenon can be derived from the upper left area of Fig. 2(a6). For the fused results on CT/MRI image pairs, the overall brightness of the first three methods is dim, but this situation is improved in the latter three methods, such as in PAPCNN and PC-LLE, clear and bright appearance of the soft tissue can be observed. However, the morphological structure of soft tissue is sharpened resulting in artifacts and loss of details. Our method can preserve the brightness and soft tissue texture details simultaneously and make the fused image look more in line with the human visual system, which is conducive to doctors' treatment of pathological locations and can be proved by Fig. 2(b6). As for the image pairs of PET/MRI and SPECT/MRI, a conclusion can be drawn that the fused result produced by ours is superior to several other methods because more soft tissue structure details are extracted from the original images and have a better visual contrast.

In order to correspond to subjective effects, the objective metrics are also measured and the results are shown in Table 1. From the table, one fact can be stated that our method achieves the maximum value among the three types of image pairs(MR-T1/MR-T2, PET/MRI and SPECT/MRI), which means that different aspects of medical information are highlighted. For example, the best indicators are reflected in the MR-T1/MR-T2 image pair showing that the physicians can clearly observe the diseased tissue presented in the anatomy structure, which is helpful for prescribing medicine to cure them. And in the PET/MRI and SPECT/MRI image pairs, the best metrics mean that doctors can make better health

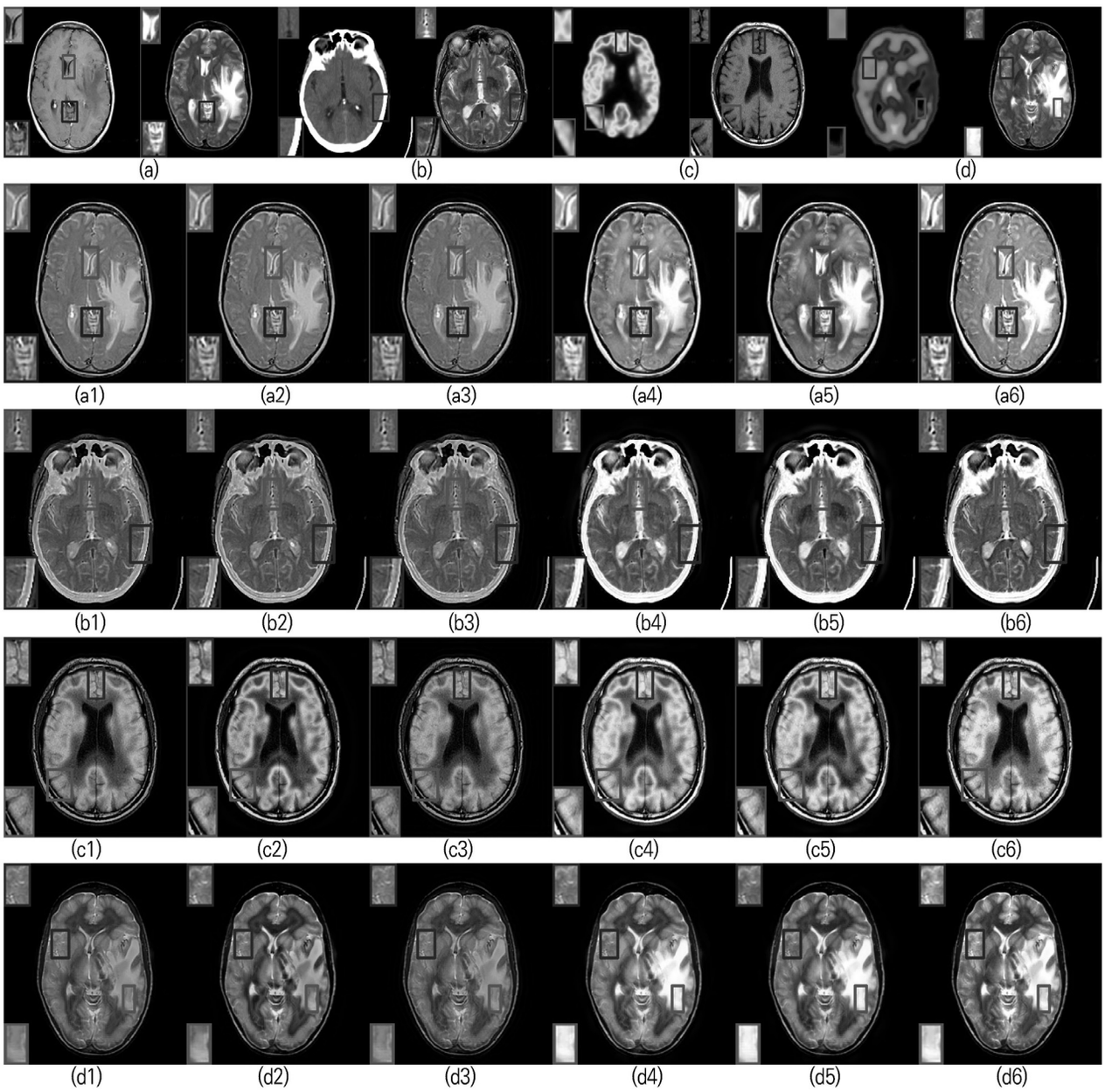

Fig. 2. Fused Images Generated on Four Typical Images Under Six Different Fusion Methods. (a) MR-T1/MR-T2 Image Pairs (b) CT/MRI Image Pairs (c) PET/MRI Image Pairs (d) SPECT/MRI Image Pairs. (a1)-(a6) are Fused Results by DTCWT, NSCT, CVT, PAPCNN, PC-LLE and Ours on the (a) MR-T1/MR-T2 Images. (b1)-(b6) are Fused Results on the (b) CT/MRI Images. (c1)-(c6) are Fused Results on the (c) PET/MRI Images. (d1)-(d6) are Fused Results on the (d) SPECT/MRI Images. Note that the Red and Blue Boxes on the Upper Left and Lower Left are Magnified Areas to Better Observe the Subjective Effect.

recommendations to patients or the ordinary people based on anatomical comparison images of soft tissue structures that reflect cellular bioactivity. Although two indicators (SSIM and SF) in CT/MRI image pairs are the second largest, the difference between their values and the maximum is only 0.0051 and 0.3375, respectively.

To make the proposed method more convincing, in this paper, an average on 10 groups of image pairs are calculated and the indicator trend is displayed in Fig. 3. As can be inferred from the above figure, our algorithm obtains excellent fusion performance on most image pairs. Especially for the average value (the last column), the six indicators have achieved the maximum values, which also show that our method has good robustness.

Table 1. The Objective Metrics from the Four Image Pairs in Fig. 2. Bold Font Represents the Best Values.

(a)	CVT	DTCWT	NSCT	PA PCNN	PC LLE	Ours	(b)	CVT	DTCWT	NSCT	PA PCNN	PC LLE	Ours
QABF	0.4498	0.6138	0.6600	0.6053	0.6570	**0.6813**	QABF	0.5053	0.6454	0.6797	0.6094	0.6221	**0.6802**
SSIM	0.4988	0.7379	0.8016	0.7285	0.7637	**0.8169**	SSIM	0.5713	0.7504	**0.8087**	0.7324	0.7500	0.8036
AG	7.6474	7.3823	7.5654	8.7383	8.6799	**9.5689**	AG	8.5797	8.1135	8.6685	8.9685	9.0516	**9.3796**
QCB	0.4953	0.6212	0.6556	0.6637	0.6739	**0.6756**	QCB	0.5047	0.6178	0.6404	0.6155	0.6428	**0.6615**
SCD	0.8315	0.8447	0.8679	1.4059	1.0573	**1.5868**	SCD	1.0053	1.0062	1.0312	1.4013	1.4091	**1.4519**
SF	24.6022	24.4773	25.0487	28.2320	28.4243	**30.7470**	SF	27.2055	26.9191	28.2567	**31.5830**	31.1560	31.2455

(c)	CVT	DTCWT	NSCT	PA PCNN	PC LLE	Ours	(d)	CVT	DTCWT	NSCT	PA PCNN	PC LLE	Ours
QABF	0.4532	0.6095	0.6630	0.6515	0.6842	**0.7767**	QABF	0.6109	0.6951	0.6724	0.6869	0.7127	**0.7902**
SSIM	0.4823	0.6862	0.6431	0.6869	0.7543	**0.9009**	SSIM	0.6995	0.8286	0.7596	0.7999	0.8122	**0.9307**
AG	9.1202	8.9614	9.8729	10.2941	10.3887	**10.8902**	AG	6.3547	6.0691	6.4587	6.4545	6.4777	**7.0733**
QCB	0.4559	0.5850	0.5699	0.5949	0.6246	**0.6852**	QCB	0.5887	0.6760	0.6360	0.6698	0.6771	**0.7708**
SCD	1.0625	1.0793	1.1884	1.5697	1.4597	**1.6787**	SCD	0.9843	0.9904	1.0689	1.5368	1.4199	**1.6372**
SF	25.0240	25.1049	28.4792	29.5234	29.6329	**30.3336**	SF	18.1652	17.8338	19.4110	19.3636	19.4903	**21.8165**

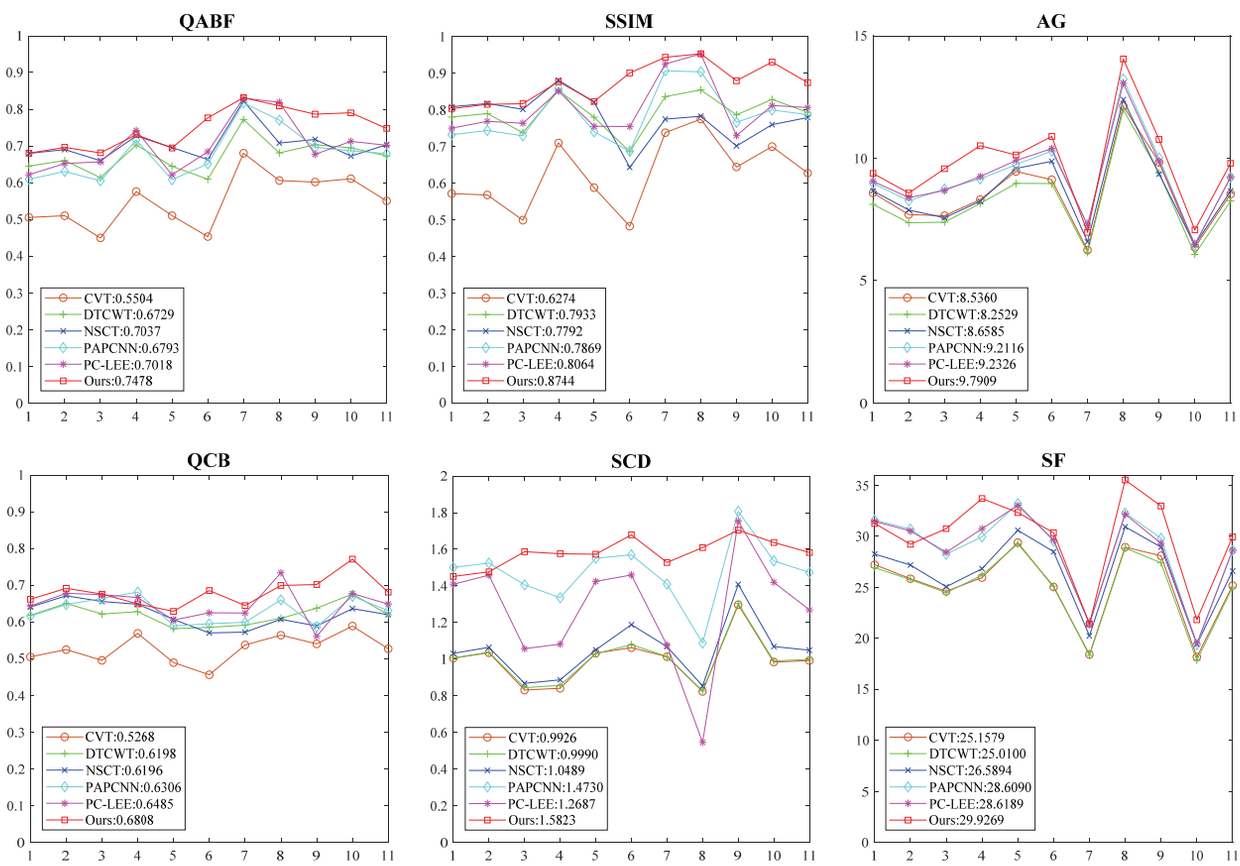

Fig. 3. Trend of quantitative indicators. The eleventh column is the average of the first 10 sets of image pairs.

4. Conclusion

In this paper, a novel MMIF is proposed. The DLD decomposes the two raw images into the corresponding sub-images, that is S_A/S_B or E_A/E_B. And then, novel fusion rules based on STO and FSPM are employed to merging the above obtained sub-images, respectively. From the figures and table presented in both subjective results and objective metrics, our method achieves high-quality fusion results compared to several mainstream methods. Specifically, our method can clearly observe the pathological changes of organs or soft tissues, efficiently reflect the biological activity and metabolic activity level of cells, and accurately capture the dense structural information of various tissues in the human body, which plays an important role in promoting the development of medical science. In future research, we will discuss the algorithm's complexity and extend it images captured by other imaging devices to observe its robustness.

References

[1] M. Yin, X. Liu, Y. Liu, and X. Chen, "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, 2019.

[2] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local laplacian energy based multimodality medical image fusion method in NSCT domain," *IEEE Access*, Vol.7, pp.20811-20824, 2019.

[3] L. Nie, L. Zhang, L. Meng, X. Song, X. Chang, and X. Li, "Modeling disease progression via multisource multitask learners: A case study with alzheimer's disease," *IEEE Transactions on Neural Networks and Learning Systems*, Vol.28, No.7, pp.1508-1519, Jul. 2017.

[4] H. Hermessi, O. Mourali, and E. Zagrouba, "Multimodal medical image fusion review: Theoretical background and recent advances," *Signal Processing*, Vol.183, pp.108036, 2021.

[5] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Information Fusion*, Vol.33, pp.100-112, 2017.

[6] R. Singh and A. Khare, "Fusion of multimodal medical images using Daubechies complex wavelet transform-a multiresolution approach," *Information Fusion*, Vol.19, pp.49-60, 2014.

[7] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, Vol.24, No.1, pp.147-164, 2015.

[8] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, Vol.59, No.4, pp.884-892, 2010.

[9] V. Prasath, R. Pelapur, G. Seetharaman, and K. Palaniappan, "Multiscale structure tensor for improved feature extraction and image regularization," *IEEE Transactions on Image Processing*, Vol.28, No.12, pp.6198-6210, 2019.

[10] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceeding of the 14th ACM International Conference on Multimedia*, pp. 815-824, 2006.

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

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, Vol.13, No.4, pp.600-612, 2004.

[13] 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," *Optics Communications*, Vol.341, pp.199-209, Apr. 2015.

[14] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," Information Fusion, Vol.14, No.2, pp.127-135, Apr. 2013.

[15] V. Aslantas and E. Bendes, "A new image quality metric for image fusion: the sum of the correlations of differences," *Aeu-international Journal of Electronics and Communications*, Vol.69, No.12, pp.1890-1896, 2015.

[16] S. Li and B. Yang, "Multifocus image fusion by combining curvelet and wavelet transform," *Pattern Recognition Letter*, Vol.29, No.9, pp.1295-1301, Jul. 2008.

[17] E. D. Vidoni, "The whole brain atlas: www. med. harvard. edu/aanlib," *Journal of Neurologic Physical Therapy*, Vol.36, No.2, pp.108, Jun. 2012, doi: 10.1097/NPT.0b013e3182563795.

Yingmei Zhang

https://orcid.org/0000-0001-6383-3825
e-mail : zym1megan@gmail.com
She is currently pursuing a Ph.D. degree in computer science and engineering with Jeonbuk National University, Jeonju, South Korea. Her research interests include multi-modal image processing, artificial intelligence, and pattern recognition.

Hyo Jong Lee

https://orcid.org/0000-0003-2581-5268
e-mail : hlee@jbnu.ac.kr
He received a Ph.D. degree in computer science from the University of Utah in 1991. He has been a professor at Jeonbuk National Univ. since 1991. His research interests include computer graphics, image processing, parallel processing, and artificial intelligence.