Neurocomputing 157 (2015) 143–152

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A new contrast based multimodal medical image fusion framework

Gaurav Bhatnagar [a,*], Q.M. Jonathan Wu [b], Zheng Liu [c]

[a] Indian Institute of Technology Jodhpur, Jodhpur, India
[b] University of Windsor, Windsor, Canada
[c] Toyota Technological Institute, Nagoya, Japan

ARTICLE INFO

Article history:
Received 21 November 2013
Received in revised form
9 January 2015
Accepted 10 January 2015
Communicated by Huiyu Zhou
Available online 19 January 2015

Keywords:
Multimodal fusion
Medical imaging
Non-subsampled contourlet transform
Entropy
Directive contrast

ABSTRACT

Medical image fusion plays an important role in clinical applications such as image-guided surgery, image-guided radiotherapy, noninvasive diagnosis, and treatment planning. The main motivation is to fuse different multimodal information into a single output. In this instance, we propose a novel framework for spatially registered multimodal medical image fusion, which is primarily based on the non-subsampled contourlet transform (NSCT). The proposed method enables the decomposition of source medical images into low- and high-frequency bands in NSCT domain. Different fusion rules are then applied to the varied frequency bands of the transformed images. Fusion coefficients are achieved by the following fusion rule: low-frequency components are fused using an activity measure based on the normalized Shannon entropy, which essentially selects low-frequency components from the focused regions with high degree of clearness. In contrast, high-frequency components are fused using the directive contrast, which essentially collects all the informative textures from the source. Integrating these fusion rules, more spatial feature and functional information can be preserved and transferred into the fused images. The performance of the proposed framework is illustrated using four groups of human brain and two clinical bone images from different sources as our experimental subjects. The experimental results and comparison with other methods show the superior performance of the framework in both subjective and objective assessment criteria.

1. Introduction

To support more accurate clinical information to physicians for better diagnosis, multimodal medical images are needed, such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and magnetic resonance angiography (MRA). Medical image fusion helps physicians to extract features from different modalities that may not be normally visible in the images. For example, the CT image can show dense structures like bones and implants with less distortion, but it cannot detect physiological changes, while the MR image can provide normal and pathological soft tissues information, but it cannot support the bones information [1]. Even a single modality can provide complementary and occasionally conflicting information due to its dependence on variable parameters. For instance, T1 weighted MR imaging gives enhanced detail of anatomical structures whereas T2 weighted MR imaging gives greater contrast between normal and abnormal tissues. Therefore, only one kind of multimodal image may not be sufficient to provide accurate clinical requirements to the physicians [1].

So far, many image fusion frameworks have been proposed in the literature [2–9] with some specific for multimodal medical image fusion [10–21]. These frameworks can be broadly classified into three categories based on the stage at which the combination mechanism takes place. This characterization includes pixel-level or sensor-level, feature-level, and decision-level fusion [2]. Among these, the most popular framework is pixel-level fusion due to the advantage of containing the originally measured quantities, easy implementation and computationally efficient [7]. Hence, in this paper, we concentrate our efforts on pixel level-fusion, and the terms image fusion or fusion are intently used for pixel level fusion throughout the paper. The well-known pixel-level frameworks are based on principal component analysis (PCA), independent component analysis (ICA), gradient pyramid (GP) filtering, etc. [22–25]. These approaches are not fully suitable for the application of medical image fusion since the features are sensitive to the human visual system existing in different scales [12]. Therefore, a multiscale or multiresolution analysis is more suitable for the fusion purposes. With the development of multiresolution analysis, wavelet transform has been identified as an ideal method for fusion. However, it is argued that wavelet decomposition is good at isolated discontinuities, but with a poor performance at edges

* Corresponding author.
E-mail addresses: goravb@iitj.ac.in (G. Bhatnagar), jwu@uwindsor.ca (Q.M.J. Wu), zheng.liu@ieee.org (Z. Liu).

http://dx.doi.org/10.1016/j.neucom.2015.01.025
0925-2312/

l textured regions. Further, it captures limited directional ormation along vertical, horizontal and diagonal directions [ ]. These issues are rectified in a recent multiscale decomposition, namely contourlet and its non-subsampled version. Contourlet is a "true" 2-D sparse representation for 2-D signals like images where sparse expansion is expressed by contour segments. As a result, it can capture 2-D geometrical structures in visual information much more effectively than the traditional multiscale methods [26]. In contrast, NSCT inherits all the advantages of contourlet transform along with shift-invariance property and effectively suppressing pseudo-Gibbs phenomena. Hereafter, some authors have proposed image fusion framework using NSCT [17–21]. Among these, most of the frameworks are implemented in multi-focus fusion. If implemented for medical imaging, the results are not of the same quality as those for the multimodal medical image fusion. The main reason is the structure of medical images. Due to fact, traditional fusion rules such as weighted average, olute maximum, spatial frequency and saliency do not effi-ently utilize prominent information present in the low- and high-frequency coefficients and result in the poor quality [21]. Therefore, two new fusion rules are proposed in this work to address these issues.

In this paper, a fully automated framework for medical image fusion is proposed in the non-subsampled contourlet transform (NSCT) domain. After the source images are decomposed by the NSCT, the coefficients of the low- and high-frequency portions are fused using two different fusion processes, which are chosen considering the physical meaning of the coefficients. Therefore, the coefficients of the low- and high-frequency bands are treated erently: the former is selected with an activity measurement cess, and the latter is selected by a contrast based process. The ed image is then obtained by taking inverse NSCT transform on the fused low- and high-frequency coefficients. Both qualitative and quantitative performance evaluations are carried out to validate the proposed framework. The final fused images are obtained by applying inverse NSCT on the fused low- and high-frequency coefficients. Extensive experiments on different multimodal CT/MRI and MR-T1/MR-T2 data-sets are carried out along with two clinical examples. Performance comparison of the proposed framework with the existing methods demonstrates the efficiency of the proposed method.

The rest of the paper is organized as follows. The NSCT is described in detail in Section 2 followed by the introduction of multimodal medical image fusion framework in Section 3. Experimental results and discussion are given in Section 4 and the concluding remarks are presented in Section 5.

2. Non-subsampled contourlet transform (NSCT)

NSCT based on the theory of countourlet transform (CT) is a kind of multi-scale and multi-direction computation framework of discrete images [26]. It can be divided into two phases luding non-subsampled pyramid (NSP) and non-subsampled ctional filter bank (NSDFB). The former phase ensures the multiscale property by using two-channel non-subsampled filter bank, producing one low-frequency and one high-frequency image at each NSP decomposition level. Subsequent NSP decomposition stages are carried out to decompose the available low-frequency component iteratively to capture the singularities in the image. As a result, NSP results in $k+1$ sub-images, which consist of one low- and k high-frequency images having the same size as the source image where k denotes the number of decomposition levels. Fig. 1 (a) shows the NSP decomposition with $k=3$ levels. The NSDFB is two-channel non-subsampled filter bank which is constructed by combining the directional fan filter banks. NSDFB allows the direction decomposition with l stages in high-frequency images from NSP at each scale and produces 2^l directional sub-images with the same size as the source image. Therefore, NSDFB offers the NSCT with the multi-direction property and provides us with more precise directional details information. A four channel NSDFB constructed with two-channel fan filter banks is illustrated in Fig. 1(b).

3. Proposed multimodal medical image fusion framework

The proposed framework realizes on a new definition of t directive contrast in NSCT domain, which takes a pair of source image denoted by A and B to generate a composite image F. The basic condition in the proposed framework is that all the source images must be registered in order to align the corresponding pixels. The definition of the directive contrast and the proposed fusion framework are described below.

3.1. Directive contrast in NSCT domain

The contrast feature measures the separation between the intensity values of a pixel and its neighboring pixels. The human visual system is highly sensitive to the intensity contrast rather than the intensity value itself. Generally, the same intensity value looks like a different intensity value depending on intensity values of neighboring pixels. Therefore, local contrast is developed and is defined as [27]

$$C=\frac{L-L_B}{L_B}=\frac{L_H}{L_B} \tag{1}$$

where L is the local luminance and L_B is the luminance of the local background. Generally, L_B is regarded as local low-frequency and hence, $L-L_B=L_H$ is treated as local high-frequency. This definition is further extended as directive contrast for the multimodal image fusion. These contrast extensions take high-frequency as the pixel value in multiresolution domain. However, considering single pixel is insufficient to determine whether the pixels are from clear parts or not. Therefore, directive contrast is integrated with the sum-modified Laplacian [28] to get salient features.

In general, the larger absolute values of high-frequency coefficients correspond to the sharper brightness in the image and lead to the salient features such as edges, lines, and region boundaries. However, these are very sensitive to the noise, which can be taken as the useful information and leading to misinterpretation of the actual information in the fused images. Hence, a proper way to select high-frequency coefficients is necessary to ensure better information interpretation. The sum-modified Laplacian is integrated with the directive contrast in NSCT domain to produce accurate salient features. Mathematically, the directive contrast in NSCT domain is given by

$$D_{l,\theta}(i,j)=\begin{cases}\dfrac{\mathrm{SML}_{l,\theta}(i,j)}{I_l(i,j)} & \text{if } I_l(i,j)\neq 0\\ \mathrm{SML}_{l,\theta}(i,j) & \text{if } I_l(i,j)=0\end{cases} \tag{2}$$

where $\mathrm{SML}_{l,\theta}$ is the sum-modified Laplacian of the NSCT frequency bands at scale l and orientation θ. On the other hand, $I_l(i,j)$ is the low-frequency sub-band at the coarsest level (l). The sum-modified Laplacian is defined by the following equation:

$$\mathrm{SML}_{l,\theta}(i,j)=\sum_{x=i-m}^{i+m}\sum_{y=j-n}^{j+n}\nabla^2_{l,\theta}I(x,y) \tag{3}$$

where

$$\begin{aligned}\nabla^2_{l,\theta}I(i,j)=&\,|2I_{l,\theta}(i,j)-I_{l,\theta}(i-\mathrm{step},j)-I_{l,\theta}(i+\mathrm{step},j)|\\&+|2I_{l,\theta}(i,j)-I_{l,\theta}(i,j-\mathrm{step})-I_{l,\theta}(i,j+\mathrm{step})|\end{aligned} \tag{4}$$

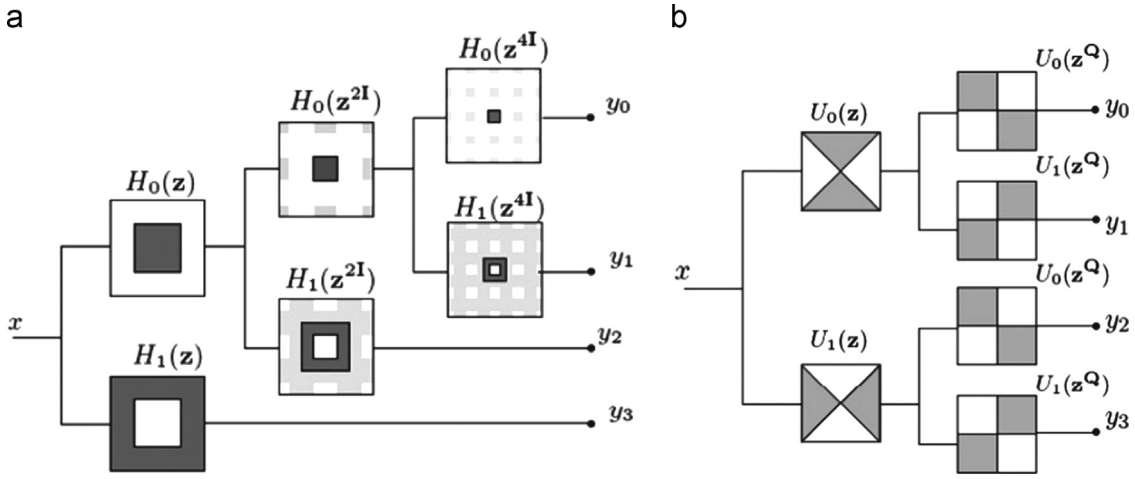

Fig. 1. (a) Three-stage non-subsampled pyramid decomposition. (b) Four-channel non-subsampled directional filter bank.

In order to accommodate possible variations in the size of texture elements, a variable spacing (step) between the pixels is used to compute SML and it is always equal to 1 [28]. The proposed directive contrast, defined by Eq. (2), not only extracts more useful features from high-frequency coefficients but also effectively deflects noise to be transferred from high-frequency coefficients to the fused coefficients.

Proposed fusion framework

In this subsection, the proposed fusion framework will be discussed in detail. Considering, two perfectly registered source images A and B, the proposed image fusion approach consists of the following steps:

1. Perform ℓ-level NSCT on the source images to obtain one low-frequency and a series of high-frequency sub-images at each level and direction θ, i.e.,

$$A : \{C^A_{\ell}, C^A_{l,\theta}\} \text{ and } B : \{C^B_{\ell}, C^B_{l,\theta}\} \tag{5}$$

where C^*_{ℓ} are the low-frequency sub-images and $C^*_{l,\theta}$ represents the high-frequency sub-images at level $l \in [1,\ell]$ in the orientation θ.

2. *Fusion of low-frequency sub-images:* The coefficients in the low-frequency sub-images represent the approximation component of the source images. The simplest way is to use the conventional averaging method to produce the composite bands. However, they cannot give the fused low-frequency component of high quality for medical image because it leads to the reduced contrast in the fused images. Therefore, a new criterion is proposed here based on the activity measurement of the low-frequency coefficients. The complete process is described as follows:
   - First, the activity measurement for the low-frequency coefficients is obtained in a region R centered at (x,y) using the normalized Shannon entropy as

$$E_A(x,y) = \frac{1}{|R|}\sum_{i,j \in R} (C^A_{\ell}(i,j))^2 \log (C^A_{\ell}(i,j))^2 \tag{6}$$

$$E_B(x,y) = \frac{1}{|R|}\sum_{i,j \in R} (C^B_{\ell}(i,j))^2 \log (C^B_{\ell}(i,j))^2 \tag{7}$$

   where $|R|$ is the size of the region, which essentially is the total number of coefficients contained in R.

   - Extract the salient information from low-frequency coefficients of each image at location (x,y) as

$$S_A(x,y) = \frac{E_A(x,y)}{E_A(x,y)+E_B(x,y)} \tag{8}$$

$$S_B(x,y) = \frac{E_B(x,y)}{E_A(x,y)+E_B(x,y)} \tag{9}$$

   - Fuse the low-frequency coefficients as

$$C^F_{\ell}(x,y) = S_A(x,y)C^A_{\ell}(x,y) + S_B(x,y)C^B_{\ell}(x,y) \tag{10}$$

   It is evident that entropy is an efficient and widely used tool to quantify the information contained in an image. The motivating factor for using entropy is that the human vision is sensitive to the area that has a sharp change in contrast. Therefore, the proposed activity measure utilizes entropy to highlight the transition between the focused (sharp) and defocused (blur) regions in the low-frequency components. According to the proposed activity measure, given in Eqs. (6) and (7), the maximum value of entropy represents the coefficients having the same frequency in the region, which further indicates the defocused (blur) regions. Hence, low-frequency fusion rules will decide the degree of clearness in the coefficients and provide an efficient way to collect coefficients from focused regions in the fused low-frequency coefficients.

3. *Fusion of high-frequency sub-images:* The coefficients in the high-frequency sub-images usually include detail components of the source image. It is noteworthy that the noise is also related to high-frequencies and may cause miscalculation of sharpness value and therefore affects the fusion performance. Therefore, a new criterion is proposed here based on directive contrast. The whole process is described as follows:
   - First, obtain the directive contrast for NSCT high-frequency sub-images at each scale and orientation using Eqs. (2)–(4), denoted by $D_{C^A_{l,\theta}}$ and $D_{C^B_{l,\theta}}$ at each level $l \in [1,\ell]$ in the direction θ.
   - Fuse the high-frequency sub-images as

$$C^F_{l,\theta}(x,y) = \begin{cases} C^A_{l,\theta}(x,y) & \text{if } D_{C^A_{l,\theta}}(x,y) \geq D_{C^B_{l,\theta}}(x,y) \\ C^B_{l,\theta}(x,y) & \text{if } D_{C^A_{l,\theta}}(x,y) < D_{C^B_{l,\theta}}(x,y) \end{cases} \tag{11}$$

4. Perform ℓ-level inverse NSCT on the fused low-frequency (C^F_{ℓ}) and high-frequency $(C^F_{l,\theta})$ subimages, to get the fused image (F).

4. Results and discussion

Some general requirements for fusion algorithm are (1) it should be able to extract and integrate complimentary features from the input images, (2) it must not introduce artifacts or inconsistencies according to Human Visual System and (3) it should be robust and reliable. Generally, these can be evaluated subjectively and/or objectively. The former relies on human visual characteristics and the specialized knowledge of the observer, hence are vague and time-consuming but are typically accurate if performed correctly. The other one is relatively formal and is easily realized by the computer algorithms, which generally evaluates the similarity between the fused and source images. However, selecting a proper consistent criterion with the subjective assessment of the image quality is rigorous. Hence, there is a need to establish an evaluation system. Therefore, an evaluation index system is set up to evaluate the proposed fusion algorithm. These indices are determined according to the statistical parameters. The mathematical definitions of these metrics are as follows [29]:

1. *Information theory-based metrics:* The normalized mutual information metric utilizes the concept of mutual information to find the similarity between the images. Mutual information (MI) is a quantitative measure of the mutual dependence of two variables, It usually shows measurement of the information shared by two images. Mathematically, MI between two discrete random variables U and V is defined as

$$\mathrm{MI}(U,V)=\sum_{u\in U}\sum_{v\in V}p(u,v)\log_2\frac{p(u,v)}{p(u)p(v)} \tag{12}$$

where $p(u,v)$ is the joint probability distribution function of U and V whereas $p(u)$ and $p(v)$ are the marginal probability distribution functions of U and V respectively. Based on the above definition, the quality of the fused image with respect to the input images A and B can be expressed as

$$Q_{MI}=2\left[\frac{\mathrm{MI}(A,F)}{H(A)+H(F)}+\frac{\mathrm{MI}(B,F)}{H(B)+H(F)}\right] \tag{13}$$

where $H(A), H(B)$ and $H(F)$is the marginal entropy of images A, B and F respectively.

2. *Image structural similarity-based metrics:* Structural similarity (SSIM) is designed by modeling any image distortion as the combination of loss of correlation, radiometric and contrast distortion. Mathematically, SSIM between two variables U and V is defined as

$$\mathrm{SSIM}(U,V)=\frac{\sigma_{UV}}{\sigma_U\sigma_V}\frac{2\mu_U\mu_V}{\mu_U^2+\mu_V^2}\frac{2\sigma_U\sigma_V}{\sigma_U^2+\sigma_V^2} \tag{14}$$

where μ_U and μ_V are the mean intensities and σ_U, σ_V, σ_{UV} are the variances and covariance respectively. Based on the definition of SSIM, a new way to use SSIM for the image fusion assessment is proposed in [30] and is defined as

$$Q_S=\begin{cases}\lambda(w)\,\mathrm{SSIM}(A,F|w)+(1-\lambda(w))\,\mathrm{SSIM}(B,F|w),\\ \quad \text{if } \mathrm{SSIM}(A,B|w)\geq 0.75\\ \max[\mathrm{SSIM}(A,F|w),\mathrm{SSIM}(B,F|w)],\\ \quad \text{if } \mathrm{SSIM}(A,B|w)<0.75\end{cases} \tag{15}$$

where w is a sliding window of size 3×3, which moves pixel by pixel from the top-left to the bottom-right corner and $\lambda(w)$ is the local weight obtained from the local image salience. See [30] for the detailed implementation of the aforementioned metric.

3. *Image feature-based metrics*: This metric uses the edge based similarity measure to find the similarity between the images. Mathematically, $Q^{AB/F}$ is defined as

$$Q^{AB/F}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[Q_{i,j}^{AF}w_{i,j}^{x}+Q_{i,j}^{BF}w_{i,j}^{y}]}{\sum_{i=1}^{M}\sum_{j=1}^{N}[w_{i,j}^{x}+w_{i,j}^{y}]} \tag{16}$$

where A, B and F represent the input and fused images respectively. The definition of Q^{AF} and Q^{BF} are same and given as

$$Q_{i,j}^{AF}=Q_{g,i,j}^{AF}Q_{\alpha,i,j}^{AF},\quad Q_{i,j}^{BF}=Q_{g,i,j}^{BF}Q_{\alpha,i,j}^{BF} \tag{17}$$

where $Q_g^{\star F}$ and $Q_\alpha^{\star F}$ are the edge strength and orientation preservation values at location (i,j) respectively for images A and B. The dynamic range for $Q^{AB/F}$ is [0,1] and it should be as close to 1 as possible for better fusion.

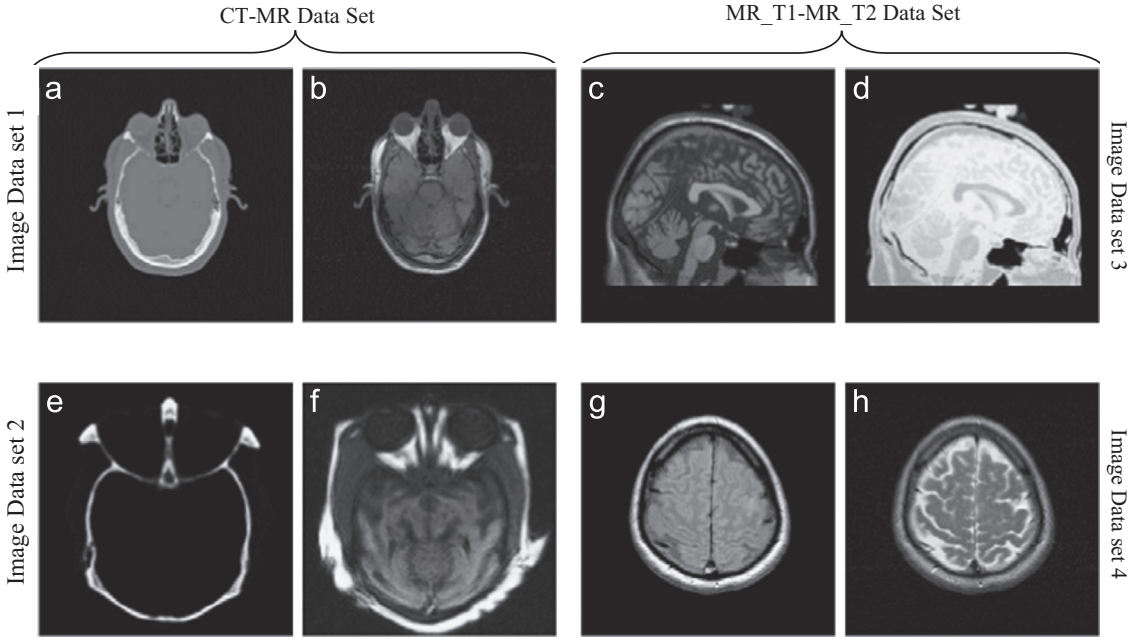

Fig. 2. Multimodal medical image data sets: (a,e) CT image, (b,f) MR image, (c,g) MR-T1 image, and (d,h) MR-T2 image.

Table 1
Evaluation indices for fused medical images.

Images	Indices	PCA	Gradient [5]	Wavelet [10]	Contourlet [12]	NSCT-1 [20]	NSCT-2 [21]	Proposed
DS 1	Q_{MI}	**2.6001**	1.3869	1.5420	1.6025	1.8028	1.8499	1.8503
	Q_S	0.5133	0.4297	0.4187	0.4277	0.4651	0.4703	**0.4725**
	$Q^{AB/F}$	0.6092	0.6637	0.5175	0.6485	0.6652	**0.6814**	0.6772
DS 2	Q_{MI}	1.3646	1.2793	1.2740	1.3244	1.4497	1.4759	**1.4899**
	Q_S	0.7985	0.7173	0.7737	0.7928	0.7837	0.8108	**0.8142**
	$Q^{AB/F}$	0.3686	0.5769	0.5090	0.5923	0.5678	0.5853	**0.5931**
DS 3	Q_{MI}	3.6627	2.5205	3.0773	3.8314	3.9161	3.9133	**3.9493**
	Q_S	0.6760	0.6508	0.6585	0.6674	0.6561	0.6892	**0.6950**
	$Q^{AB/F}$	0.6645	0.6942	0.6176	0.6816	0.6841	0.6961	**0.6992**
DS 4	Q_{MI}	1.4915	1.2983	1.3349	1.3413	1.5502	1.5085	**1.5170**
	Q_S	0.8414	0.8130	0.8259	0.8303	0.8743	0.8976	**0.8989**
	$Q^{AB/F}$	0.3434	0.5613	0.4635	0.5689	0.5699	**0.5724**	0.5691

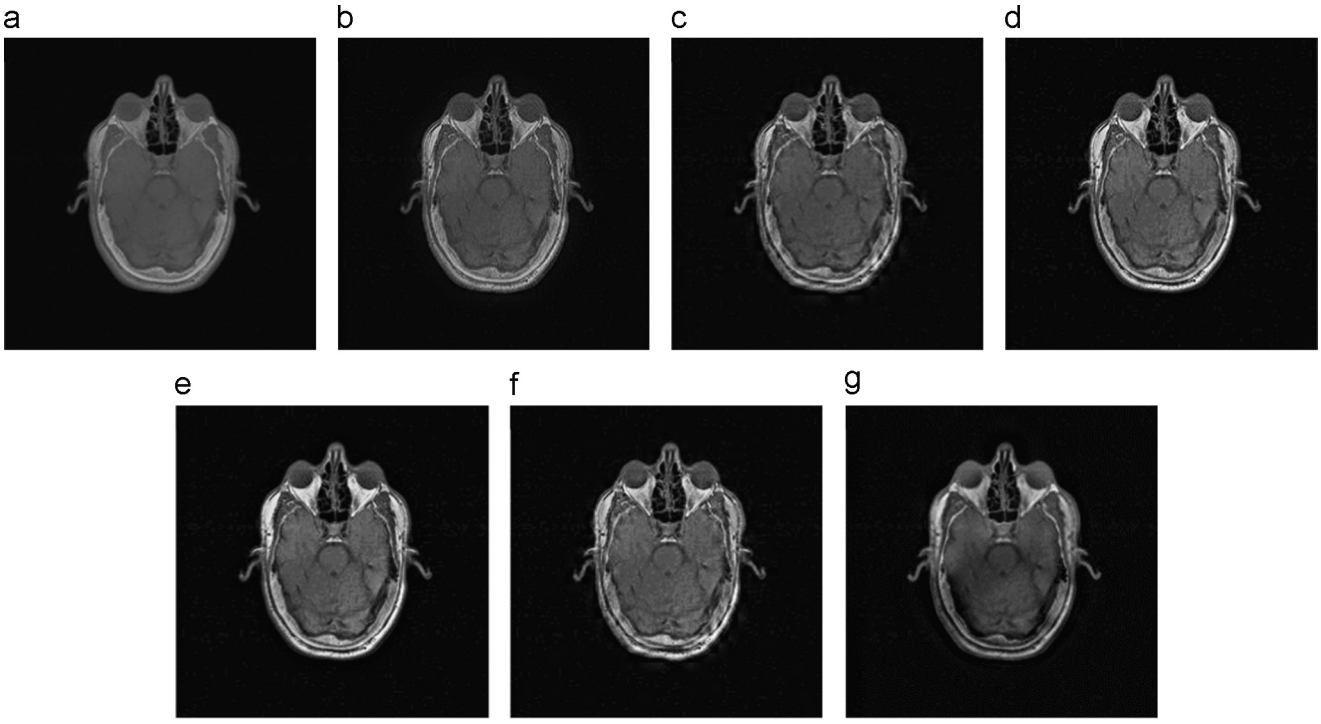

Fig. 3. The multimodal medical image fusion results for data set 1: fused images from (a) PCA [22], (b) Gradient Pyramid [5], (c) wavelet [10], (d) contourlet [12], (e) non-subsampled contourlet [20], (f) non-subsampled contourlet [21], and (g) proposed technique.

To evaluate the performance of the proposed image fusion approach, four different data sets of human brain are considered (see Fig. 2). These images are characterized in two different groups: (1) CT–MR and (2) MR-T1–MR-T2. The images in Fig. 2(a,e) and (b,f) are CT and MR images whereas Fig. 2(c,g) and (d,h) T1-weighted MR image (MR-T1) and T2-weighted MR image (MR-T2), respectively. The corresponding pixels of two input images have been perfectly co-aligned. All images have the same size of 256×256 pixel, with 256-level gray scale. The proposed medical fusion technique is applied to these image sets.

It can be seen that due to various imaging principles and environments, the source images with different modalities contain complementary information. For all these image groups, results of the proposed fusion framework are compared with the traditional PCA [22], Gradient Pyramid [5], wavelet [10], contourlet [12] and non-subsampled contourlet [20,21] based methods. The comparison of statistical parameters for the fused images according to different fusion algorithms is shown in Table 1 and visually in Figs. 3–6. From figure and table, it is clear that the proposed algorithm not only preserves spectral information but also improves the spatial detail information over all existing algorithms, which can be easily observed by the obtained maximum values of evaluation indices (highlighted in bold in the Table 1).

The PCA algorithm gives baseline results. For all experimental images, PCA based methods give poor results relative to other algorithms. This was expected because this method has no scale selectivity. This limitation is rectified in pyramid and multiresolution based algorithms but on the cost of quality i.e., the contrast of the fused image is reduced which is greater in pyramid based algorithms and comparatively lesser in multiresolution based algorithms. Among multiresolution based algorithms, the proposed algorithm based on NSCT performs better. The main reason behind the better performance is the use of directive contrast which essentially takes the advantage of both contrast and

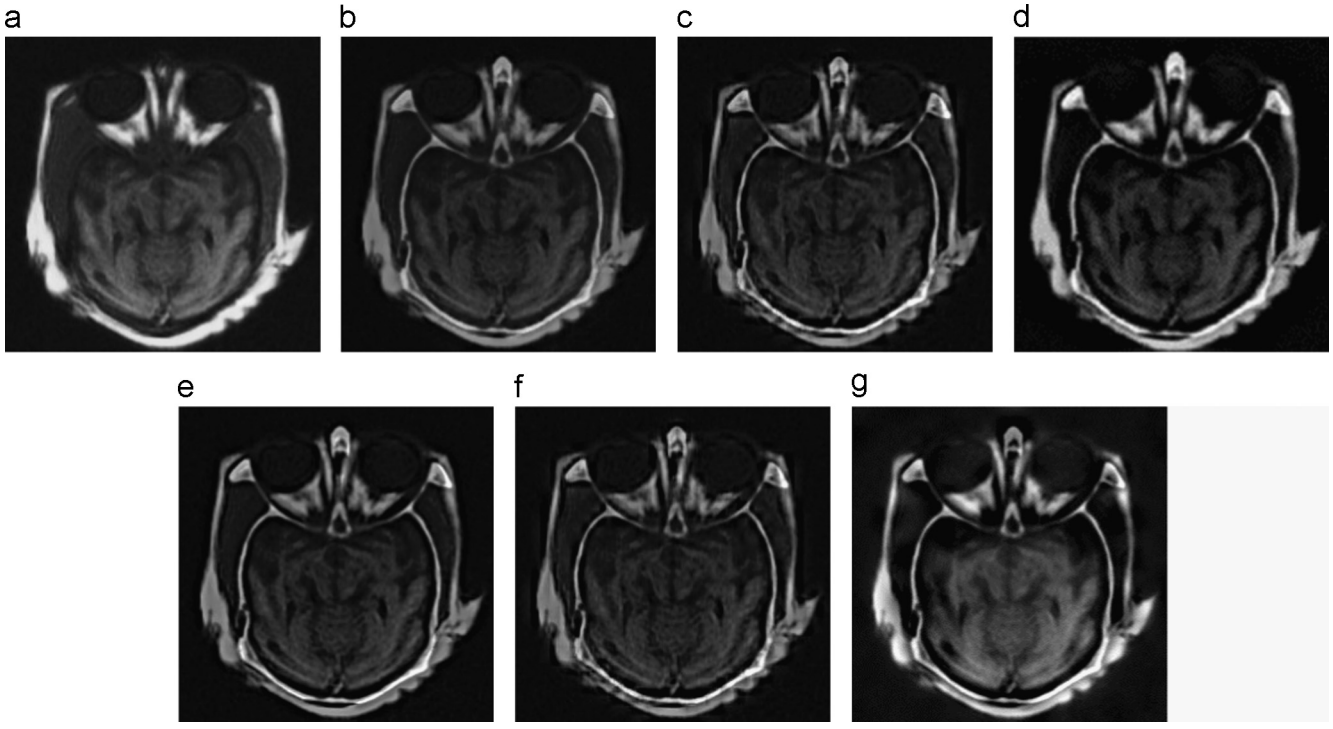

Fig. 4. The multimodal medical image fusion results for data set 2: fused images from (a) PCA [22], (b) Gradient Pyramid [5], (c) wavelet [10], (d) contourlet [12], (e) non-subsampled contourlet [20], (f) non-subsampled contourlet [21], and (g) proposed technique.

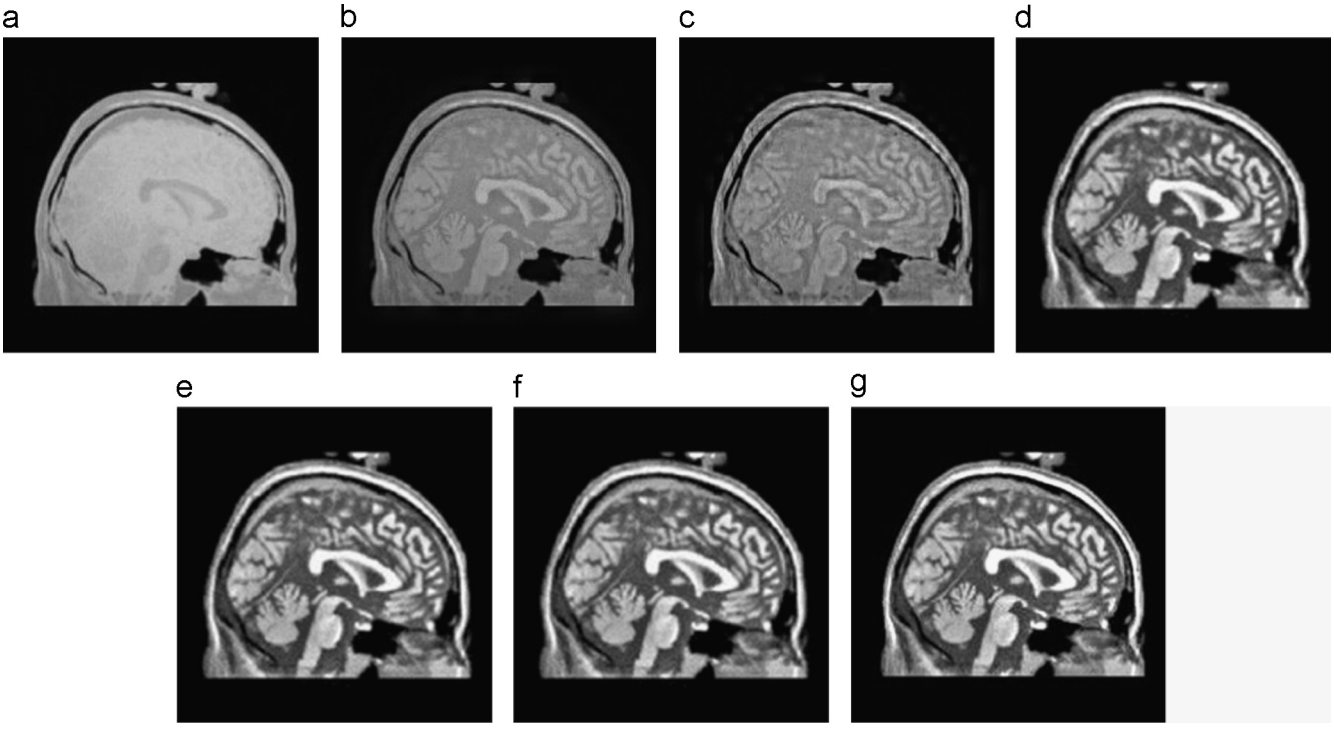

Fig. 5. The multimodal medical image fusion results for data set 3: Fused images from (a) PCA [22], (b) Gradient Pyramid [5], (c) wavelet [10], (d) contourlet [12], (e) non-subsampled contourlet [20], (f) non-subsampled contourlet [21], and (g) proposed technique.

visibility. Further, the shift-invariance property of NSCT produces clearer and more natural fused image than other multiresolution based fused results. This is also justified by the fact that shift-invariant decomposition overcomes pseudo-Gibbs phenomena successfully and improves the quality of the fused image around edges. Further, the effect of different region sizes (used in the fusion of low-frequency components) on the fused images is given in Table 2. It is clear from the table that there is a negligible increment in the indices, which essentially proves that the quality of the fused image is persistent with the change in the region size.

Finally, a comparative study on computational complexities has been done to explore efficiency of the proposed work. In this study, the execution times of various fusion methods for the used data sets are considered. The final timing for execution is depicted

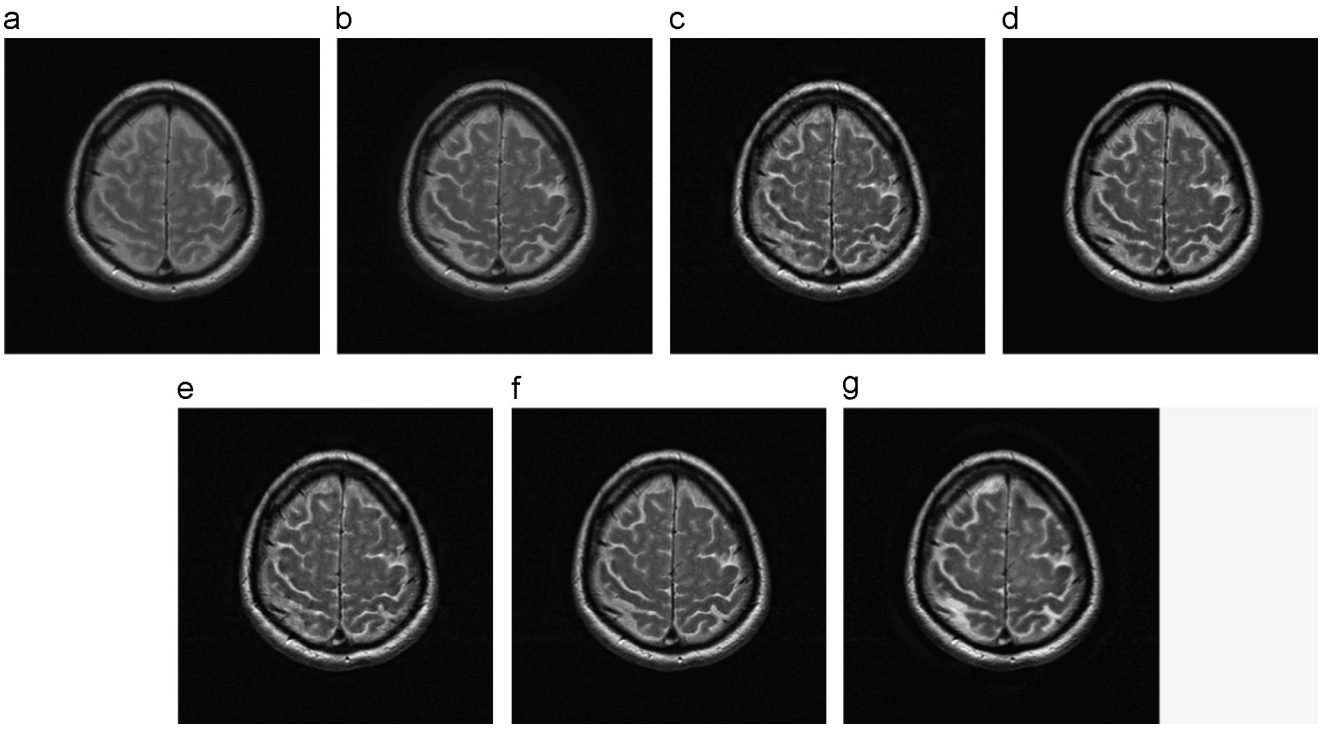

Fig. 6. The multimodal medical image fusion results for data set 4: fused images from (a) PCA [22], (b) Gradient Pyramid [5], (c) wavelet [10], (d) contourlet [12], (e) non-subsampled contourlet [20], (f) non-subsampled contourlet [21], and (g) proposed technique.

Table 2
Evaluation indices for proposed fusion framework using different region sizes.

Images	Indices	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11
DS 1	Q_{MI}	1.8503	1.8589	1.9621	2.0780	2.1931
	Q_S	0.4725	0.4786	0.4980	0.5045	0.5051
	$Q^{AB/F}$	0.6772	0.6794	0.6804	0.6804	0.6808
DS 2	Q_{MI}	1.4899	1.4969	1.5102	1.5572	1.5597
	Q_S	0.8142	0.8193	0.8199	0.8203	0.8206
	$Q^{AB/F}$	0.5931	0.5974	0.5992	0.6042	0.6074
DS 3	Q_{MI}	3.9493	3.9495	3.9495	3.9496	3.9498
	Q_S	0.6950	0.6956	0.6963	0.6963	0.6964
	$Q^{AB/F}$	0.6992	0.6994	0.6997	0.6998	0.6999
DS 4	Q_{MI}	1.5170	1.5196	1.5197	1.5197	1.5208
	Q_S	0.8989	0.8993	0.8994	0.8996	0.8996
	$Q^{AB/F}$	0.6992	0.5703	0.5735	0.5735	0.5757

Table 3
Execution time (in seconds) comparison.

Images	PCA	Gradient [5]	Wavelet [10]	Contourlet [12]	NSCT-1 [20]	NSCT-2 [21]	Proposed
DS 1	0.0328	0.1249	0.0780	1.9682	2.2220	2.2245	2.2198
DS 2	0.0311	0.1267	0.0766	1.9649	2.2235	2.2236	2.2232
DS 3	0.0333	0.1262	0.0759	1.9522	2.2198	2.2189	2.2194
DS 4	0.0320	0.1268	0.0783	1.9632	2.2114	2.2122	2.2119

in Table 3. The execution times are found by running all the algorithms in MATLAB on a terminal with Intel Core 2 Duo-2.6 GHz processor with 4 GB RAM. The code for each fusion method is executed 8 times and the average time for each data set is presented in the table. The execution time for NSCT based techniques is comparatively high when compared to the other methods, which is quiet obvious since NSCT has no fast algorithm. In contrast, the execution time of the proposed method is comparable with that of other NSCT methods. Although the proposed framework needs more execution time but it obtains much clearer and more natural fused image.

5. Clinical applications

In order to demonstrate the practical value of the proposed scheme in medical imaging, two clinical cases are considered where X-ray and bone scan medical modalities are used. An X-ray is primarily used to determine if there is damage to bones such as a fracture or dislocated joint and locate a foreign object (i.e., bullet). A bone scan, on the other hand, is a nuclear scanning test by which certain bone abnormalities can be identified. It is generally used to diagnose a number of bone related conditions, including, bone inflammation (bone pain due to a fracture), light fractures that may not be visible in X-ray, damage bone detection (due to certain infections) and cancer of the bone. The result of bone scan can be viewed in 'hot' and 'cold' spots. Hot spots appear darker and denote an area of high tracer uptake whereas cold spots appear light and indicate the area of less tracer uptake. Hot spots usually indicate the affected area in the bones and the perfect location can be obtained by combining X-rays with the bone scans.

The first case is of a 59-year-old female who is experiencing the right calcaneal discomfort from 12 months. The X-ray was performed for the possibility of fracture and no fracture was found (see Fig. 7(a)). After her complains of continued pain, a bone scan was performed where a focus of increased tracer activity has been identified, as depicted in Fig. 7(b). The X-ray is then fused with the bone scan image that identified the location to the medial part of the anterior portion of

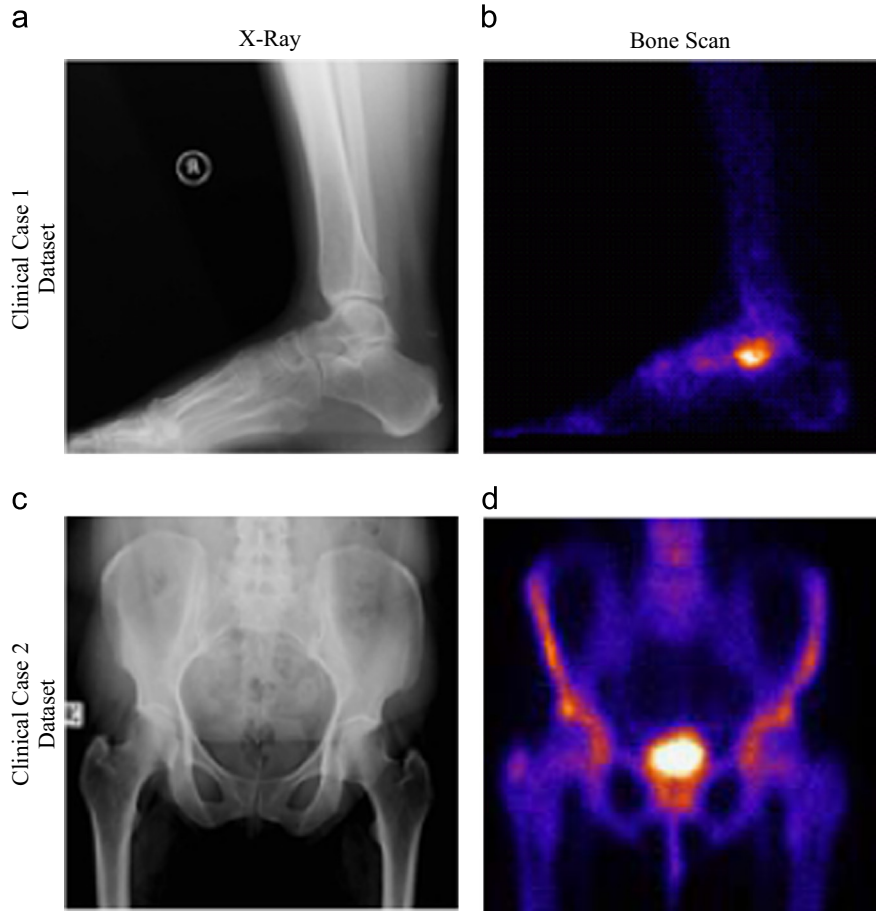

Fig. 7. Clinical Case 1: (a) X-ray image and (b) bone scan image; Clinical Case 2: (c) X-ray image and (b) bone scan image.

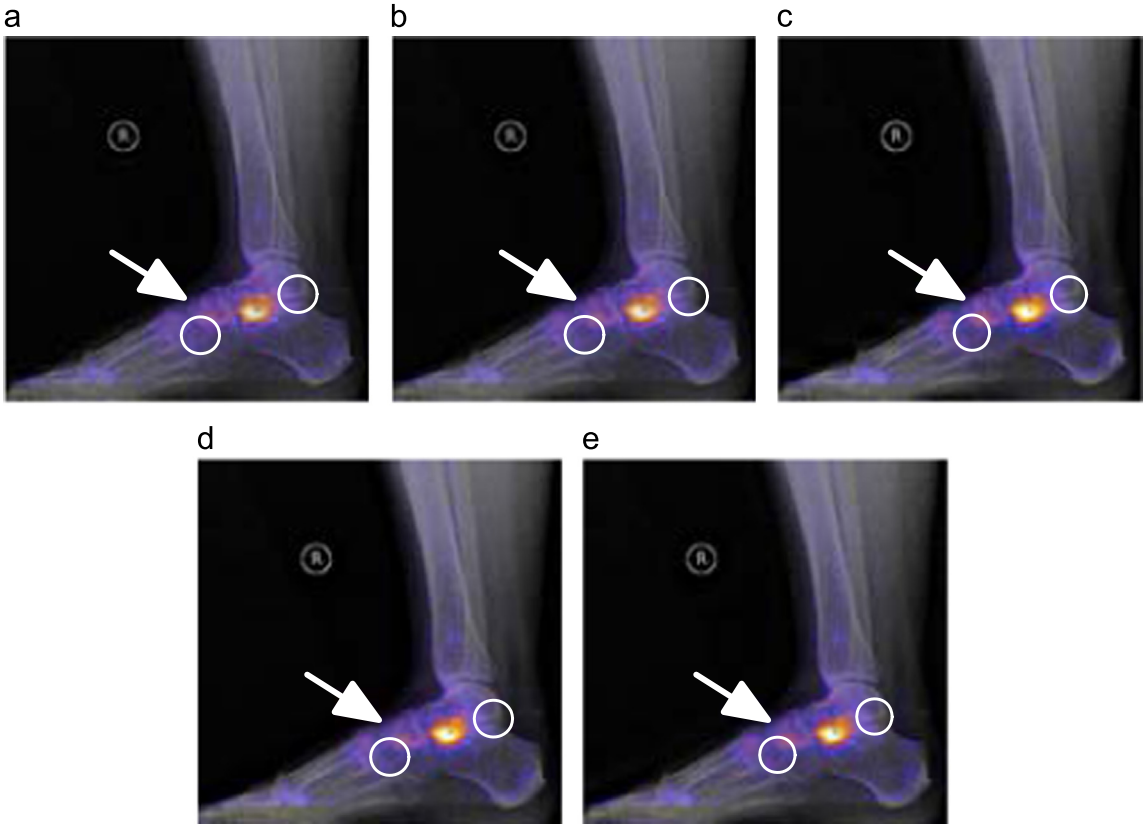

Fig. 8. Clinical Case 1: fused image from (a) Guihong et al. [10], (b) Yang et al. [12], (c) Das et al. [20], (d) Yang et al. [21], and (e) proposed technique.

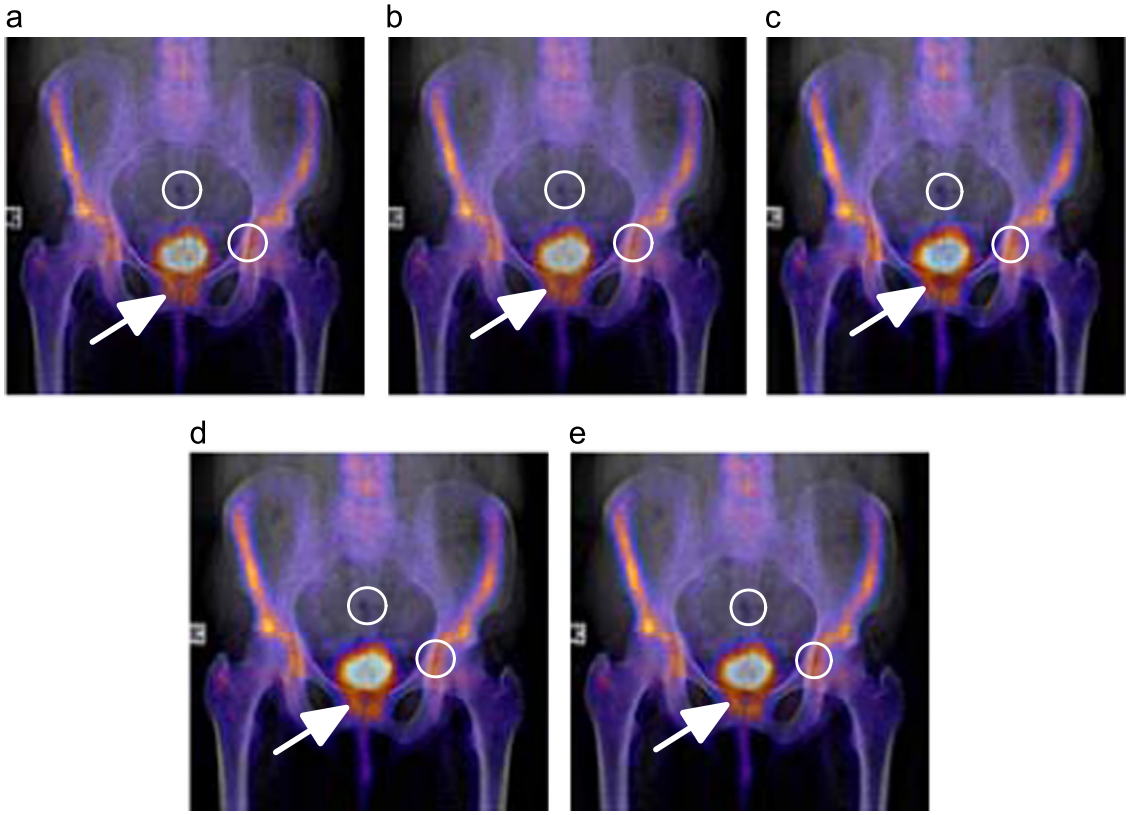

Fig. 9. Clinical Case 2: fused image from (a) Guihong et al. [10], (b) Yang et al. [12], (c) Das et al. [20], (d) Yang et al. [21], and (e) proposed technique.

Table 4
Evaluation indices for fused images in clinical applications.

Images	Indices	Wavelet [10]	Contourlet [12]	NSCT-1 [20]	NSCT-2 [21]	Proposed
Clinical Case 1	Q_{MI}	32.8357	24.7964	40.7922	41.9057	**42.1005**
	Q_S	0.7949	0.8032	0.8091	0.8147	**0.8232**
	$Q^{AB/F}$	0.5601	0.5668	0.5677	**0.5769**	0.5749
Clinical Case 2	Q_{MI}	12.8196	13.2336	15.0318	15.2769	**15.5797**
	Q_S	0.7387	0.7478	0.7478	0.7498	**0.7584**
	$Q^{AB/F}$	0.5435	0.5936	0.5961	0.5953	**0.5961**

the subtalar joint (Fig. 8). This abnormality is unlikely to represent residual activity associated with a fracture as it does not spread significantly into either adjacent talus or calcaneum.

The second case is of a 57-year-old female with ongoing right hip, groin and buttock pain after a fall several months earlier. Again the X-ray was normal but the results of bone scan were remarkable. Delayed images showed a mild focal uptake in the right greater trochanter in keeping with avulsion injury/enthesopathy at the insertion of the gluteus muscles, see Fig. 7(d). After fusion, mild increased uptake is noted at the superior aspect of both hip joints corresponding to marginal osteophytes seen on X-ray (Fig. 9).

Here, the results are compared with the best four algorithms obtained from the earlier analysis, i.e., Guihong et al. [10], Yang et al. [12], Das et al. [20] and Yang et al. [21]. From Figs. 8 and 9, it can be observed that all the fusion algorithms have fairly good spatial information. However, existing methods have somewhat high spectral distortions which are more in the case of [10] and comparatively less in [21]. The color information is also distorted in the existing algorithms (shown with the white arrows). On the contrary, the color information is least distorted and the spatial details are much clearer as the original X-ray image, and the spectral features are also natural (shown with the white circles) in the fused images by the proposed method. The same observation can be realized from Table 4, where the quantitative comparison of the results is depicted for the clinical applications (highlighted in bold). Therefore, the proposed method not only preserves the crucial features existing in both original images but also improves the color information when compared to the existing methods.

6. Conclusions

In this paper, a novel image fusion framework is proposed for multi-modal medical images, which is based on non-subsampled contour transform and directive contrast. Two different rules are implemented to preserve more information in the fused image with an improved quality. The low-frequency bands are fused by considering phase congruency whereas directive contrast is adopted as the fusion measurement for high-frequency bands. In our experiments, two groups of CT/MR and two groups of MR-T1/MR-T2 images are fused using conventional fusion algorithms and the proposed framework. The visual and statistical comparisons

demonstrate that the proposed algorithm better preserves the image details and significantly improves the image visual effects than the other fusion methods with very less information distortion. Further, the practical value of the proposed framework in medical imaging has been validated by the two clinical cases.

References

[1] V. Barra, J.Y. Boire, A general framework for the fusion of anatomical and functional medical images, NeuroImage 13 (3) (2001) 410–424.
[2] S.T. Shivappa, B.D. Rao, M.M. Trivedi, An iterative decoding algorithm for fusion of multimodal information, EURASIP J. Adv. Signal Process. 2008 (2008) 478396-1–478396-10.
[3] A. Cardinali, G.P. Nason, A statistical multiscale approach to image segmentation and fusion, in: Proceedings of International Conference on Information Fusion, Philadelphia, PA, USA, 2005, pp. 475–482.
[4] P.S. Chavez, A.Y. Kwarteng, Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis, Photogramm. Eng. Remote Sens. 55 (1989) 339–348.
[5] V.S. Petrovic, C.S. Xydeas, Gradient-based multiresolution image fusion, IEEE Trans. Image Process. 13 (2) (2004) 228–237.
[6] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graph Models Image Process. 57 (3) (1995) 235–245.
[7] R. Redondo, F. Sroubek, S. Fischer, G. Cristobal, Multifocus image fusion using the log-Gabor transform and a Multisize Windows technique, Inf. Fusion 10 (2) (2009) 163–171.
[8] Y. Chai, H. Li, Z. Li, Multifocus image fusion scheme using focused region detection and multiresolution, Opt. Commun. 284 (2011) 4376–4389.
[9] Y. Chai, H. Li, X. Zhang, Multifocus image fusion based on features contrast of multiscale products in nonsubsampled contourlet transform domain, Optik 123 (2012) 569–581.
[10] Q. Guihong, Z. Dali, Y. Pingfan, Medical image fusion by wavelet transform modulus maxima, Opt. Express 9 (2001) 184–190.
[11] V. Barra, J.Y. Boire, A general framework for the fusion of anatomical and functional medical images, NeuroImage 13 (3) (2001) 410–424.
[12] L. Yang, B.L. Guo, W. Ni, Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform, Neurocomputing 72 (2008) 203–211.
[13] F.E. Ali, I.M. El-Dokany, A.A. Saad, F.E. Abd El-Samie, Curvelet fusion of MR and CT images, Prog. Electromagn.Res. C 3 (2008) 215–224.
[14] S. Daneshvar, H. Ghassemian, MRI and PET image fusion by combining IHS and retina-inspired models, Inf. Fusion 11 (2) (2010) 114–123.
[15] Y. Yang, D.S. Park, S. Huang, N. Rao, Medical image fusion via an effective wavelet-based approach, EURASIP J. Adv. Signal Process. 2010 (2010) 44-1–44-13.
[16] T. Li, Y. Wang, Biological image fusion using a NSCT based variable-weight method, Inf. Fusion 12 (2) (2011) 85–92.
[17] Q. Zhang, B.L. Guo, Multi-focus image fusion using the nonsubsampled contourlet transform, Signal Process. 89 (2009) 1334–1346.
[18] B. Yang, S.T. Li, F.M. Sun, Image fusion using nonsubsampled contourlet transform, in: Proceedings of International Conference on Image and Graphics, vol. 405, 2007, pp. 719–724.
[19] Y. Wu, C. Wu, S. Wu, Fusion of remote sensing images based on nonsubsampled contourlet transform and region segmentation, J. Shanghai Jiaotong Univ. (Science) 16 (2011) 722–727.
[20] S. Das, M.K. Kundu, NSCT-based multimodal medical image fusion using pulse-coupled neural network and modified spatial frequency, Med. Biol. Eng. Comput. 50 (2012) 1105–1114.
[21] Yong Yang, Song Tong, Shuying Huang, Pan Lin, Log-Gabor energy based multimodal medical image fusion in NSCT domain, Comput. Math. Methods Med. 2014 (2014) 835481:1–12.
[22] O. Rockinger, T. Fechner, Pixel-level image fusion: the case of image sequences, Proc. SPIE 3374 (1998) 378–388.
[23] C. He, Q. Liu, H. Li, H. Wang, Multimodal medical image fusion based on IHS and PCA, Proc. Eng. 7 (2010) 280–285.
[24] T. Wan, C. Zhu, Z. Qin, Multifocus image fusion based on robust principal component analysis, Pattern Recognit. Lett. 34 (9) (2013) 1001–1008.
[25] H. Zhao, Z. Shang, Y.Y. Tang, B. Fang, Multi-focus image fusion based on the neighbor distance, Pattern Recognit. 46 (3) (2013) 1002–1011.
[26] A.L. daCunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: theory, design and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.
[27] P. Whittle, Increments and decrements: luminance discrimination, Vis. Res. 26 (1986) 1677–1691.
[28] W. Huang, Z. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recognit. Lett. 28 (4) (2007) 493–500.
[29] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganière, W. Wu, Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study, IEEE Trans. Pattern Anal. Mach. Intell. 34 (1) (2012) 94–109.
[30] C. Yang, J. Zhang, X. Wang, X. Liu, A novel similarity based quality metric for image fusion, Inf. Fus. 9 (2008) 156–160.

Gaurav Bhatnagar is an Assistant Professor at Indian Institute of Technology Jodhpur, India, since September 2013. He was a postdoctoral research associate in the Department of Electrical and Computer Engineering, and a member of Computer Vision and Sensing Systems Laboratory, at University of Windsor, ON, Canada, from October 2009 to August 2013. He received his Ph.D. degree in interdisciplinary area of Applied Mathematics and Computer Science and M.Sc. degree in Applied Mathematics from the Indian Institute of Technology Roorkee, India, in 2010 and 2005, respectively. His research interests lie on the broad area of wavelet and fractional transform theory, with applications in digital watermarking, encryption techniques, biometrics and mathematical image analysis. He has co-authored more than 50 journal articles and conference proceedings, and has contributed to two books in his area of interest.

Q.M. Jonathan Wu received the Ph.D. degree in Electrical Engineering from the University of Wales, Swansea, U.K., in 1990. He was with the National Research Council of Canada for 10 years from 1995, where he became a Senior Research Officer and a Group Leader. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He has published more than 250 peer-reviewed papers in computer vision, image processing, intelligent systems, robotics, and integrated microsystems. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.

Dr. Wu holds the Tier 1 Canada Research Chair in Automotive Sensors and Information Systems. He is an Associate Editor for the IEEE Transactions on Systems, Man, and Cybernetics Part A, and the International Journal of Robotics and Automation. He has served on technical program committees and international advisory committees for many prestigious conferences.

Zheng Liu received the doctorate in engineering from Kyoto University, Japan, in 2000. From 2000 to 2001, he was a Research Fellow with the Control and Instrumentation Division of Nanyang Technological University, Singapore. He then joined the Institute for Aerospace Research (IAR), National Research Council Canada, Ottawa, as a governmental laboratory visiting fellow, in 2001. After being with IAR for five years, he transferred to the NRC Institute for Research in Construction. He then moved to Toyota Technological Institute at Nagoya where he currently holds a Professor position. He also holds an Adjunct and a Visiting Professorship at the University of Ottawa and Fuzhou University respectively. His research interests include image/data fusion, computer vision, pattern recognition, sensor/sensor network, structural health monitoring, and nondestructive inspection and evaluation. He co-chairs the IEEE IMS TC-36. He is a senior member of the IEEE and a member of SPIE.