Directive Contrast Based Multimodal Medical Image Fusion in NSCT Domain

Gaurav Bhatnagar, *Member, IEEE*, Q. M. Jonathan Wu, *Senior Member, IEEE*, and Zheng Liu, *Senior Member, IEEE*

***Abstract*—Multimodal medical image fusion, as a powerful tool for the clinical applications, has developed with the advent of various imaging modalities in medical imaging. The main motivation is to capture most relevant information from sources into a single output, which plays an important role in medical diagnosis. In this paper, a novel fusion framework is proposed for multimodal medical images based on non-subsampled contourlet transform (NSCT). The source medical images are first transformed by NSCT followed by combining low- and high-frequency components. Two different fusion rules based on phase congruency and directive contrast are proposed and used to fuse low- and high-frequency coefficients. Finally, the fused image is constructed by the inverse NSCT with all composite coefficients. Experimental results and comparative study show that the proposed fusion framework provides an effective way to enable more accurate analysis of multimodality images. Further, the applicability of the proposed framework is carried out by the three clinical examples of persons affected with Alzheimer, subacute stroke and recurrent tumor.**

***Index Terms*—Multimodal medical image fusion, non-subsampled contour transform, phase congruency, directive contrast.**

I. Introduction

In the recent years, medical imaging has attracted increasing attention due to its critical role in health care. However, different types of imaging techniques such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), magnetic resonance angiography (MRA), etc., provide limited information where some information is common, and some are unique. For example, X-ray and computed tomography (CT) can provide dense structures like bones and implants with less distortion, but it cannot detect physiological changes [1]. Similarly, normal and pathological soft tissue can be better visualized by MRI image whereas PET can be used to provide better information on blood flow and flood activity with low spatial resolution. As a result, the anatomical and functional medical images are needed to be combined for a compendious view. For this purpose, the multimodal medical image fusion has been identified as a promising solution which aims to integrating information from multiple modality images to obtain a more complete and accurate description of the same object. Multimodal medical image fusion not only helps in diagnosing diseases, but it also reduces the storage cost by reducing storage to a single fused image instead of multiple-source images.

So far, extensive work has been made on image fusion technique [2]–[19] with various techniques dedicated to multimodal medical image fusion [20]–[28]. These techniques have been categorized into three categories according to merging stage. These include pixel level, feature level and decision level fusion where medical image fusion usually employs the pixel level fusion due to the advantage of containing the original measured quantities, easy implementation and computationally efficiency [17]. Hence, in this paper, we concentrate our efforts to pixel level fusion, and the terms image fusion or fusion are intently used for pixel level fusion. The well-known pixel level fusion are based on principal component analysis (PCA), independent component analysis (ICA), contrast pyramid (CP), gradient pyramid (GP) filtering, etc. Since, the image features are sensitive to the human visual system exists in different scales. Therefore, these are not the highly suitable for medical image fusion [22]. Recently, with the development of multiscale decomposition, wavelet transform has been identified ideal method for image fusion. However, it is argued that wavelet decomposition is good at isolated discontinuities, but not good at edges and textured region. Further, it captures limited directional information along vertical, horizontal and diagonal direction [23]. These issues are rectified in a recent multiscale decomposition 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 traditional multiscale methods [29].

In this paper, a novel fusion framework is proposed for multimodal medical images based on non-subsampled contourlet transform. The core idea is to perform NSCT on the source images followed by the fusion of low- and high-frequency coefficients. The phase congruency and directive contourlet contrast feature are unified as the fusion rules for low- and high-frequency coefficients. The phase congruency provides a contrast- and brightness-invariant representation of low-frequency coefficients whereas directive contrast efficiently determines the frequency coefficients from the clear parts in the high-frequency. The combinations of these two can preserve more details in source images and further improve the quality of fused image. The efficiency of the proposed framework is carried out by the

Manuscript received April 15, 2012; revised September 20, 2012; accepted November 15, 2012. Date of publication February 01, 2013; date of current version July 15, 2013. This work was supported by the Canada Research Chair program, the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dimitri Van De Ville.

G. Bhatnagar and Q. M. J. Wu are with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada (e-mail: goravb@uwindsor.ca; jwu@uwindsor.ca).

Z. Liu is with the Intelligent Information Processing Laboratory, Toyota Technological Institute, Nagoya, Japan (e-mail: zheng.liu@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2244870

extensive fusion experiments on different multimodal CT/MRI dataset. Further, visual and quantitative analysis show that the proposed framework provides a better fusion outcome when compared to conventional image fusion techniques. The salient contributions of the proposed framework over existing methods can be summarized as follows.

- This paper proposes a new image fusion framework for multimodal medical images, which relies on the NSCT domain.
- Two different fusion rules are proposed for combining low- and high-frequency coefficients.
- For fusing the low-frequency coefficients, the phase congruency based model is used. The main benefit of phase congruency is that it selects and combines contrast- and brightness-invariant representation contained in the low-frequency coefficients.
- On the contrary, a new definition of directive contrast in NSCT domain is proposed and used to combine high-frequency coefficients. Using directive contrast, the most prominent texture and edge information are selected from high-frequency coefficients and combined in the fused ones.
- The definition of directive contrast is consolidated by incorporating a visual constant to the SML based definition of directive contrast which provide a richer representation of the contrast.
- Further, the proposed scheme is also extended for multispectral fusion in $l\alpha\beta$ color space which essentially rectifies the IHS color space undesirable cross-channel artifacts and produce best quality output with natural spectral features and improved the color information.

The rest of the paper is organized as follows. NSCT and phase congruency are described in Section II followed by the proposed multimodal medical image fusion framework in Section III. Experimental results and discussions are given in Section IV and the concluding remarks are described in Section V.

II. Preliminaries

This section provides the description of concepts on which the proposed framework is based. These concepts include NSCT and phase congruency and are described as follows.

A. Non-Subsampled Contourlet Transform (NSCT)

NSCT, based on the theory of CT, is a kind of multi-scale and multi-direction computation framework of the discrete images [29]. It can be divided into two stages including non-subsampled pyramid (NSP) and non-subsampled directional filter bank (NSDFB). The former stage ensures the multiscale property by using two-channel non-subsampled filter bank, and one low-frequency image and one high-frequency image can be produced at each NSP decomposition level. The subsequent NSP decomposition stages are carried out to decompose the low-frequency component available iteratively to capture the singularities in the image. As a result, NSP can result in $k+1$ sub-images, which consists 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 gives the NSP decomposition

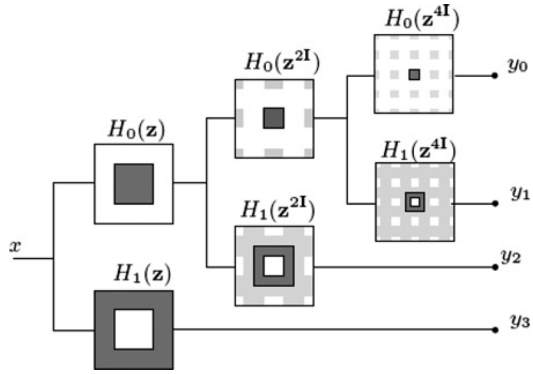

Fig. 1. Three-stage non-subsampled pyramid decomposition.

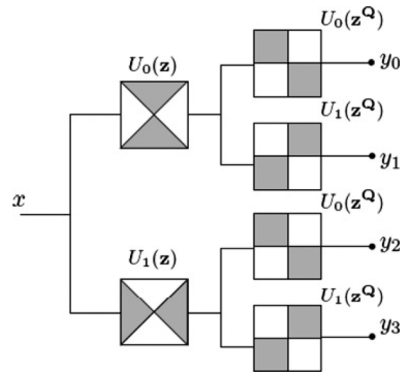

Fig. 2. Four-channel non-subsampled directional filter bank.

with $k=3$ levels. The NSDFB is two-channel non-subsampled filter banks which are 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, the 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. 2.

B. Phase Congruency

Phase congruency is a measure of feature perception in the images which provides a illumination and contrast invariant feature extraction method [30], [31]. This approach is based on the Local Energy Model, which postulates that significant features can be found at points in an image where the Fourier components are maximally in phase. Furthermore, the angle at which phase congruency occurs signifies the feature type. The phase congruency approach to feature perception has been used for feature detection. First, logarithmic Gabor filter banks at different discrete orientations are applied to the image and the local amplitude and phase at a point (x,y) are obtained. The phase congruency, $P^o_{x,y}$, is then calculated for each orientation o as shown in (1) at the bottom of the next page, where $W^o_{x,y}$ is the weight factor based on the frequency spread, $A^{o,n}_{x,y}$ and $\phi^{o,n}_{x,y}$ are the respective amplitude and phase for the scale n, $\widetilde{\phi}^o_{x,y}$ is the weighted mean phase, T is a noise threshold constant and ε is a

small constant to avoid divisions by zero. The symbol $[]_+$ denotes that the enclosed quantity is equal to itself when the value is positive, and zero otherwise. Only energy values that exceed T, the estimated noise influence and are counted in the result. The appropriate noise threshold, T is readily determined from the statistics of the filter responses to the image. For details of this phase congruency measure and its implementation see [31]. The main properties, which acted as the motivation to use phase congruency for multimodal fusion, are as follows.

- The phase congruency is invariant to different pixel intensity mappings. The images captured with different modalities have significantly different pixel mappings, even if the object is same. Therefore, a feature that is free from pixel mapping must be preferred.
- The phase congruency feature is invariant to illumination and contrast changes. The capturing environment of different modalities varies and resulted in the change of illumination and contrast. Therefore, multimodal fusion can be benefitted by an illumination and contrast invariant feature.
- The edges and corners in the images are identified by collecting frequency components of the image that are in phase. As we know, phase congruency gives the Fourier components that are maximally in phase. Therefore, phase congruency provides the improved localization of the image features, which lead to efficient fusion.

III. Proposed Multimodal Medical Image Fusion Framework

In this section, we have discussed some of the motivating factors in the design of our approach to multimodal medical image fusion. The proposed framework realizes on the directive contrast and phase congruency 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 block diagram of the proposed framework is depicted in Fig. 3 but before describing it, the definition of directive contrast is first described, which is as follows.

A. Directive Contrast in NSCT Domain

The contrast feature measures the difference of the intensity value at some pixel from the 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 [8]

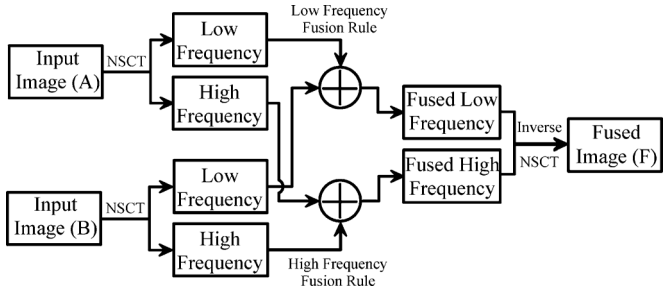

Fig. 3. Block diagram of proposed multimodal medical image fusion framework.

$$C = \frac{L - L_B}{L_B} = \frac{L_H}{L_B} \tag{2}$$

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 multimodal image fusion [10]. 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, the directive contrast is integrated with the sum-modified-Laplacian [32] to get more accurate 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, region boundaries, and so on. However, these are very sensitive to the noise and therefore, the noise will be taken as the useful information and misinterpret the actual information in the fused images. Hence, a proper way to select high-frequency coefficients is necessary to ensure better information interpretation. Hence, 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

$$D_{l,\theta}(i,j) = \begin{cases} \frac{SML_{l,\theta}(i,j)}{I_l(i,j)}, & \text{if } I_l(i,j) \neq 0 \\ SML_{l,\theta}(i,j), & \text{if } I_l(i,j) = 0 \end{cases} \tag{3}$$

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,

$$P^o_{x,y} = \frac{\sum_n W^o_{x,y} \left[A^{o,n}_{x,y} \left(\cos\left(\phi^{o,n}_{x,y} - \widetilde{\phi}^o_{x,y}\right) - \left|\sin\left(\phi^{o,n}_{x,y} - \widetilde{\phi}^o_{x,y}\right)\right| \right) - T \right]_+}{\sum_n A^{o,n}_{x,y} + \varepsilon} \tag{1}$$

$I_l(i,j)$ is the low-frequency sub-band at the coarsest level (l). The sum-modified-Laplacian is defined by following equation

$$\mathrm{SML}_{l,\theta}(i,j) = \sum_{x=i-m}^{i+m} \sum_{y=j-n}^{j+n} \nabla^2_{l,\theta} I(x,y) \tag{4}$$

where

$$\begin{aligned}\nabla^2_{l,\theta} I(i,j) = &|2I_{l,\theta}(i,j) - I_{l,\theta}(i-\mathrm{step},j) \\ &- I_{l,\theta}(i+\mathrm{step},j)| \\ &+ |2I_{l,\theta}(i,j) - I_{l,\theta}(i,j-\mathrm{step}) \\ &- I_{l,\theta}(i,j+\mathrm{step})|\end{aligned} \tag{5}$$

In order to accommodate for possible variations in the size of texture elements, a variable spacing (step) between the pixels is used to compute partial derivatives to obtain SML and is always equal to 1 [32]. Further, the relationship between the contrast sensitivity threshold and background intensity is nonlinear, which makes the human visual system highly sensitive to contrast variation [33]. Hence, the above integration must be improved to provide better details by exploiting visibility of low-frequency coefficients in the above-mentioned definition. Hence, the directive contrast in NSCT domain is given as

$$D_{l,\theta}(i,j) = \begin{cases} \left(\frac{1}{I_l(i,j)}\right)^{\alpha} \frac{\mathrm{SML}_{l,\theta}(i,j)}{I_l(i,j)}, & \text{if } I_l(i,j) \neq 0 \\ \mathrm{SML}_{l,\theta}(i,j), & \text{if } I_l(i,j) = 0 \end{cases} \tag{6}$$

where α as a visual constant representing the slope of the best-fitted lines through high-contrast data, which is determined by physiological vision experiments, and it ranges from 0.6 to 0.7 [33]. The proposed definition of directive contrast, defined by (6), not only extract more useful features from high-frequency coefficients but also effectively deflect noise to be transferred from high-frequency coefficients to fused coefficients.

B. 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.,

$$A : \left\{\mathcal{C}^A_\ell, \mathcal{C}^A_{l,\theta}\right\} \text{ and } B : \left\{\mathcal{C}^B_\ell, \mathcal{C}^B_{l,\theta}\right\} \tag{7}$$

   where $\mathcal{C}^*_\ell$ are the low-frequency sub-images and $\mathcal{C}^*_{l,\theta}$ 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 methods to produce the composite bands. However, it 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 phase congruency. The complete process is described as follows.

   ► First, the features are extracted from low-frequency sub-images using the phase congruency extractor (1), denoted by $P_{\mathcal{C}^A_\ell}$ and $P_{\mathcal{C}^B_\ell}$ respectively.

   ► Fuse the low-frequency sub-images as

$$\mathcal{C}^F_\ell(x,y) = \begin{cases} \mathcal{C}^A_\ell(x,y), & \text{if } P_{\mathcal{C}^A_\ell}(x,y) > P_{\mathcal{C}^B_\ell}(x,y) \\ \mathcal{C}^B_\ell(x,y), & \text{if } P_{\mathcal{C}^A_\ell}(x,y) < P_{\mathcal{C}^B_\ell}(x,y) \\ \frac{\sum_{k\in A,B} \mathcal{C}^k_\ell(x,y)}{2}, & \text{if } P_{\mathcal{C}^A_\ell}(x,y) = P_{\mathcal{C}^B_\ell}(x,y) \end{cases} \tag{8}$$

3. *Fusion of High-frequency Sub-images:* The coefficients in the high-frequency sub-images usually include details component 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 effect the fusion performance. Therefore, a new criterion is proposed here based on directive contrast. The whole process is described as follows.

   ► First, the directive contrast for NSCT high-frequency sub-images at each scale and orientation using (3)–(5), denoted by $D_{\mathcal{C}^A_{l,\theta}}$ and $D_{\mathcal{C}^B_{l,\theta}}$ at each level $l \in [1,\ell]$ in the direction θ.

   ►Fuse the high-frequency sub-images as

$$\mathcal{C}^F_{l,\theta}(x,y) = \begin{cases} \mathcal{C}^A_{l,\theta}(x,y), & \text{if } D_{\mathcal{C}^A_{l,\theta}}(x,y) \geq D_{\mathcal{C}^B_{l,\theta}}(x,y) \\ \mathcal{C}^A_{l,\theta}(x,y), & \text{if } D_{\mathcal{C}^A_{l,\theta}}(x,y) < D_{\mathcal{C}^B_{l,\theta}}(x,y) \end{cases} \tag{9}$$

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

C. Extension to Multispectral Image Fusion

The IHS transform is a widely used multispectral image fusion methods in the research community. It works on a simple way to convert multispectral image from RGB to IHS color space. Fusion is then performed by fusing I component and source panchromatic image followed by the inverse IHS conversion to get the fused image. The IHS based process can preserve the same spatial resolution as the source panchromatic image but seriously distort the spectral (color) information in the source multispectral image. Therefore, IHS model is not a suitable for multimodal medical image fusion because a little distortion can leads to wrong diagnosis.

The aforementioned drawback can be avoided by incorporating different operations or different color-space such that undesirable cross-channel artifacts will not occur. Such a color-space is developed in [34]. This space is called $l\alpha\beta$-space and is based on the human perception research which assumes that the natural image processing is ideally done by human visual system. The RGB to $l\alpha\beta$ color space conversion can be summarized as follows. First, the RGB color space is converted to LMS cone space as

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{10}$$

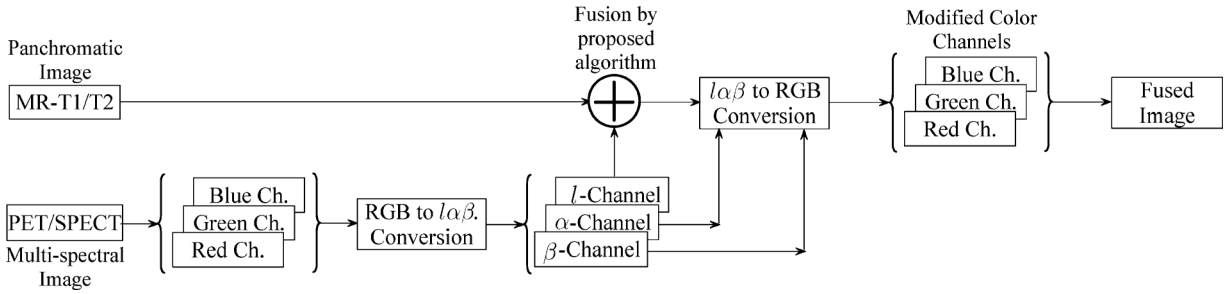

Fig. 4. Block diagram for the multispectral image fusion: synchronization of proposed fusion algorithm in $l\alpha\beta$ color space.

The data in LMS cone space show a great deal of skew and this can be eliminated by converting LMS cone space channels to logarithmic color space, i.e.,

$$\Gamma = \lg L, \quad \Omega = \lg M, \quad \Psi = \lg S \tag{11}$$

The logarithmic color space is further transformed in three orthogonal color-space $(l\alpha\beta)$ as

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \Gamma \\ \Omega \\ \Psi \end{bmatrix} \tag{12}$$

In $l\alpha\beta$ color space, l represents an achromatic channel whereas α and β are chromatic yellow-blue and red-green channels and these channels are symmetrical and compact. The inversion, $l\alpha\beta$ to RGB space, is done by the following inverse operations.

$$\begin{bmatrix} \Gamma \\ \Omega \\ \Psi \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} \tag{13}$$

and

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.4679 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} 10^{\Gamma} \\ 10^{\Omega} \\ 10^{\Psi} \end{bmatrix} \tag{14}$$

The proposed fusion algorithm can easily be extended for the multispectral images by utilizing proposed fusion rules in $l\alpha\beta$ color space (see Fig. 4). The core idea is to transform multispectral image from RGB color space to the $l\alpha\beta$ color space using the process given above. Now, the panchromatic image and the achromatic channel (l) of the multispectral image are fused using proposed fusion algorithm followed by the inverse $l\alpha\beta$ to RGB conversion to get the final fused image.

IV. Results and Discussions

Some general requirements for fusion algorithm are: (1) it should be able to extract complimentary features from 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 or objectively. The former relies on human visual characteristics and the specialized knowledge of the observer, hence vague, time-consuming and poor-repeatable but are typically accurate if performed correctly. The other one is relatively formal and easily realized by the computer algorithms, which generally evaluate 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 create an evaluation system. Therefore, first an evaluation index system is established to evaluate the proposed fusion algorithm. These indices are determined according to the statistical parameters.

A. Evaluation Index System

1) *Normalized Mutual Information:* 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

$$\mathrm{MI}(U,V) = \sum_{u\in U}\sum_{v\in V} p(u,v)\log_2\frac{p(u,v)}{p(u)p(v)} \tag{15}$$

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

$$Q_{\mathrm{MI}} = 2\left[\frac{\mathrm{MI}(A,F)}{H(A)+H(F)} + \frac{\mathrm{MI}(B,F)}{H(B)+H(F)}\right] \tag{16}$$

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

2) *Structural Similarity based Metric:* 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

$$\mathrm{SSIM}(U,V) = \frac{\sigma_{UV}}{\sigma_U\sigma_V}\frac{2\mu_U\mu_V}{\mu_U^2+\mu_V^2}\frac{2\sigma_U\sigma_V}{\sigma_U^2+\sigma_V^2} \tag{17}$$

where μ_U, μ_V are mean intensity 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 [35] and is defined as

$$Q_S = \begin{cases} \lambda(w)\mathrm{SSIM}(A,F|w) + (1-\lambda(w))\mathrm{SSIM}(B,F|w), \\ \qquad\qquad \text{if } \mathrm{SSIM}(A,B|w) \geq 0.75 \\ \max\left[\mathrm{SSIM}(A,F|w), \mathrm{SSIM}(B,F|w)\right], \\ \qquad\qquad \text{if } \mathrm{SSIM}(A,B|w) < 0.75 \end{cases} \tag{18}$$

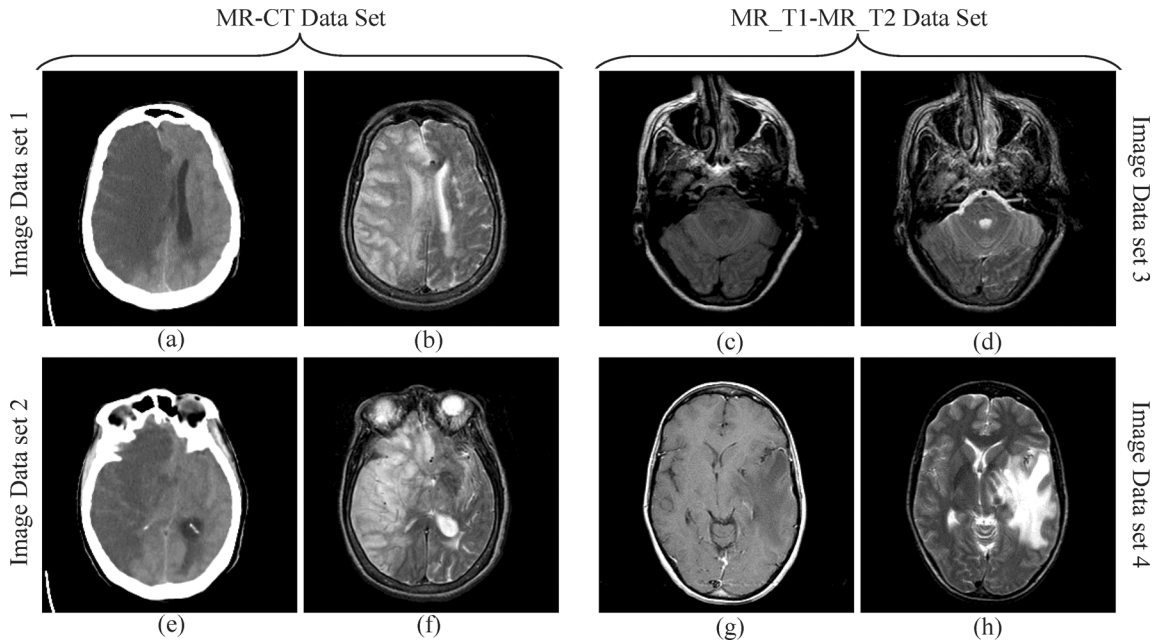

Fig. 5. Multimodal medical image data sets: (a), (e) CT image (b), (f) MRI image (c), (g) MR-T1 image (d), (h) MR-T2 image.

where w is a sliding window of size, 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 [35] for the detailed implementation of the aforementioned metric.

3) *Edge Based Similarity Measure*: The edge based similarity measure gives the similarity between the edges transferred in the fusion process. Mathematically, $Q^{AB/F}$ is defined as

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[Q_{i,j}^{AF}w_{i,j}^{x} + Q_{i,j}^{BF}w_{i,j}^{y}\right]}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[w_{i,j}^{x} + w_{i,j}^{y}\right]} \tag{19}$$

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

$$\begin{aligned} Q_{i,j}^{AF} &= Q_{g,i,j}^{AF}Q_{\alpha,i,j}^{AF}, \\ Q_{i,j}^{BF} &= Q_{g,i,j}^{BF}Q_{\alpha,i,j}^{BF} \end{aligned} \tag{20}$$

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.

B. Experiments on CT/MRI Image Fusion

To evaluate the performance of the proposed image fusion approach, four different datasets of human brain are considered (see Fig. 5). These images are characterized in two different groups 1) CT-MRI and 2) MR-T1-MR-T2. The images in Figs. 5(a),(e) and (b),(f) are CT and MRI images whereas Fig. 5(c,g) and (d),(h) T1-weighted MR image (MR-T1) and T2-weighted MR image (MR-T2). 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 principle and environment, the source images with different modality contain complementary information. For all these image groups, results of proposed fusion framework are compared with the traditional PCA (MS rule), Contrast Pyramid [5], Gradient Pyramid [6], wavelet [20], contourlet [22] and non-subsampled contourlet (NSCT-1 [11] and NSCT-2 [12]) based methods. {In order to do a fair comparison, the same experimental images are used for all existing methods. The level of decomposition is set to 3 for all the pyramid, wavelet and countourlet based methods, including proposed. For wavelet based method [20], images are decomposed using the 'db4' wavelet since it has used frequently in the existing wavelet based methods. For implementing NSCT, maximally flat filters and diamond maxflat filters are used as pyramidal and directional filters respectively.

The comparison of statistical parameters for fused images according to different fusion algorithms are shown in Table I and visually in Fig. 6. From figure and table, it is clear that the proposed algorithms not only preserve spectral information but also improve the spatial detail information than the existing algorithms (highlighted by red arrows), which can also be justified by the obtained maximum values of evaluation indices (see Table I). 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 therefore it cannot captures prominent information localized in different scales. This limitation is rectified in pyramid and multiresolution based algorithms but on the cost of quality i.e., the contrast of the fuse image is reduced which is greater in pyramid based algorithms and comparatively less in multiresolution based algorithms.

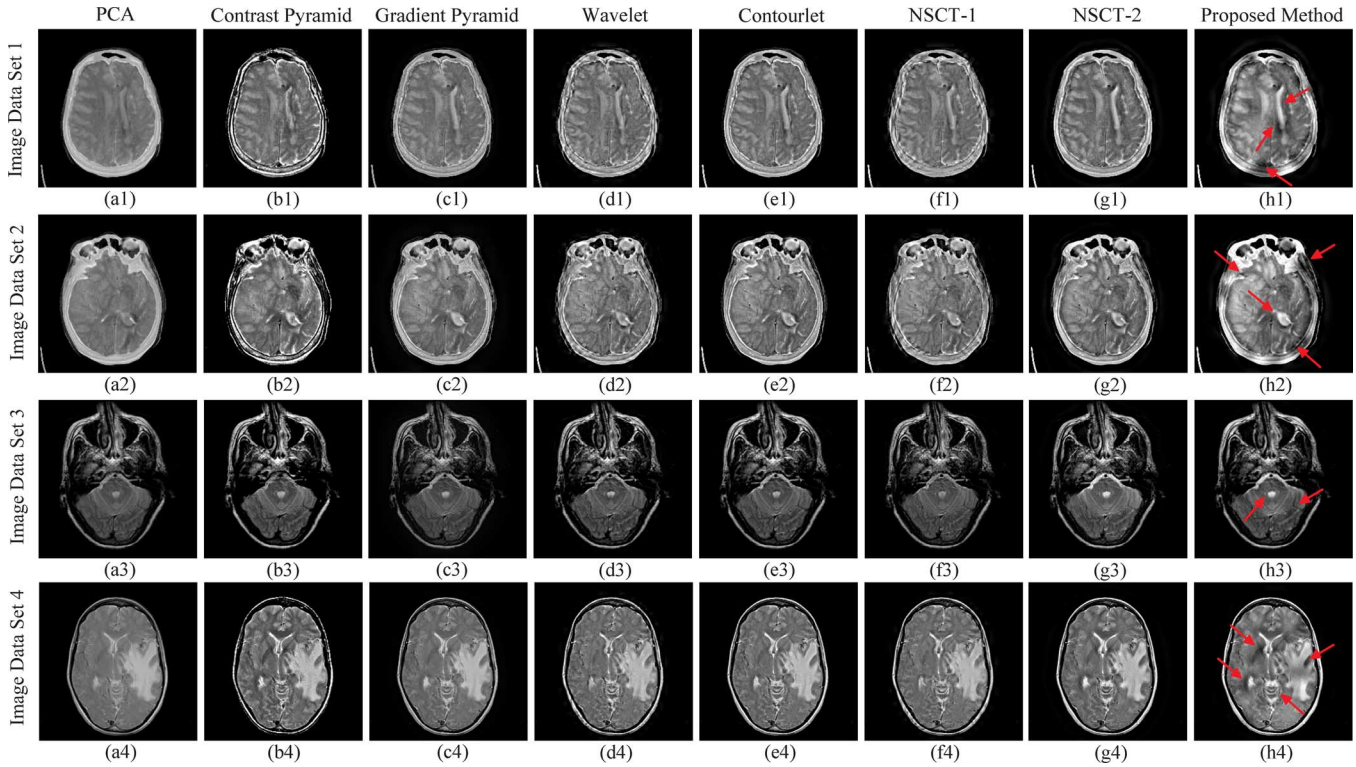

Fig. 6. The multimodal medical image fusion results of different fusion algorithms: Fused images from (a1), (a2), (a3), (a4) PCA based technique; (b1), (b2), (b3), (b4) Contrast Pyramid based technique; (c1), (c2), (c3), (c4) Gradient Pyramid based technique; (d1), (d2), (d3), (d4) Wavelet based technique; (e1), (e2), (e3), (e4) Contourlet based technique; (f1), (f2), (f3), (f4) NSCT based technique 1; (g1), (g2), (g3), (g4) NSCT based technique 2; (h1), (h2), (h3), (h4) proposed technique.

TABLE I
EVALUATION INDICES FOR FUSED MEDICAL IMAGES

Images Modalities	Indices	PCA	Contrast [5]	Gradient [6]	Wavelet [20]	Contourlet [22]	NSCT-1 [11]	NSCT-2 [12]	Proposed
Image Dataset 1 (MRI and CT)	Q_{MI}	**1.5645**	1.0372	0.9417	0.8812	1.0380	1.0367	1.0471	1.0813
	Q_S	0.8415	0.8059	0.7495	0.7551	0.7968	0.7965	0.8166	**0.8726**
	$Q^{AB/F}$	0.5226	0.6863	0.7055	0.6669	0.7424	0.7457	0.7538	**0.7560**
Image Dataset 2 (MRI and CT)	Q_{MI}	0.9436	0.9412	0.8466	0.8340	0.9463	0.9537	0.9517	**0.9681**
	Q_S	0.7357	0.7545	0.6253	0.7247	0.7663	0.7695	0.7727	**0.7795**
	$Q^{AB/F}$	0.5545	0.5922	0.7235	0.7038	0.7776	0.7781	0.7796	**0.7825**
Image Dataset 3 (MR-T1 and MR-T2)	Q_{MI}	1.0259	1.1723	0.8688	1.0286	1.1405	1.7824	1.1816	**1.1865**
	Q_S	0.9054	0.9068	0.8083	0.9043	0.9259	0.9262	0.9401	**0.9527**
	$Q^{AB/F}$	0.6730	0.5909	0.6655	0.6403	0.6916	0.6924	0.6958	**0.6991**
Image Dataset 4 (MR-T1 and MR-T2)	Q_{MI}	**1.4094**	1.0672	0.9556	0.9535	1.0684	1.6949	1.0694	1.0695
	Q_S	0.7945	0.8389	0.7815	0.7497	0.8112	0.8107	0.8112	**0.8117**
	$Q^{AB/F}$	0.4408	0.4301	0.5636	0.5326	0.6780	0.6780	**0.6785**	0.6783

Among multiresolution based algorithms, the algorithms based on NSCT performs better. This is due to the fact that NSCT is an multi-scale geometric analysis tool which utilizes the geometric regularity in the image and provide a asymptotic optimal representation in the terms of better localization, multi-direction and shift invariance. 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. If the NSCT based methods have been compared then it can be observed that the performance of the proposed method is better than existing NSCT based methods [11], [12]. The algorithm in [11] gives poor results with respect to other NSCT methods. This algorithm uses a directional vector, obtained from high frequency sub-bands, to fuse low-frequency sub-bands. This directional vector essentially defines the clarity factor and is used to collect pixels from blur and clear regions. This algorithm performs somewhat good in the case of multifocus images but the performance degraded when it is applied to the medical images. This is because this algorithm is not able to utilize prominent information present in the low-frequency efficiently and results in the poor quality. Here, it is important to mention that the method in [11] still perform better than other multiresolution based algorithms. The performance of the proposed and the method in [12] is close to each other, providing the good quality fused images compared to others. However, looking carefully at the results, clearly the output from [12] suffer considerably from less contrast and less visibility in the corpus callosum,

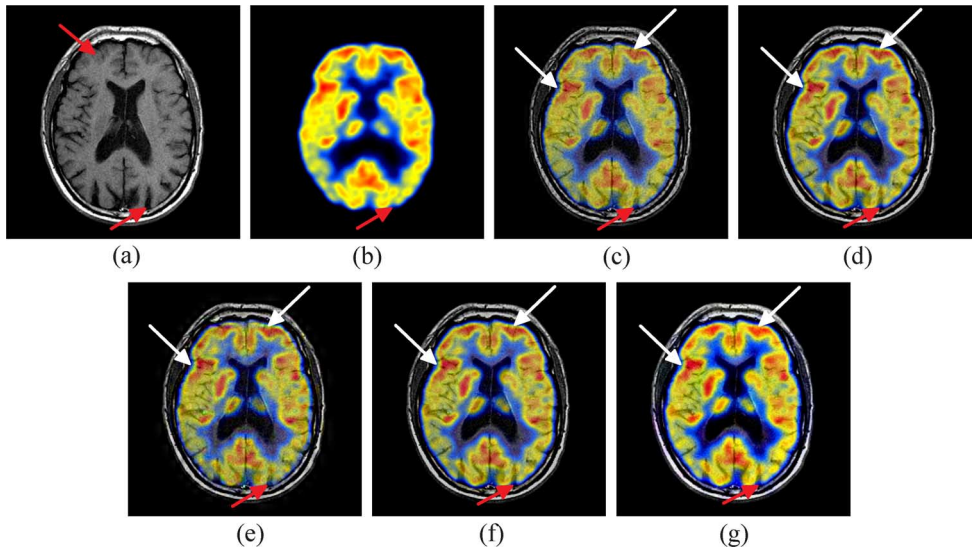

Fig. 7. Brain images of the man affected with Alzheimer: (a) MRI image (b) PET image; Fused images by (c) wavelet (d) contourlet (e) NSCT-1 (f) NSCT-2 (g) proposed method.

septum pellucidum, sulcus, flax cerebri and other structures which is quiet better in the output of the proposed method (see the regions highlighted by red arrows in the Fig. 6). The main reason behind the better performance is the proposed fusion rules for low- and high-frequency coefficients which extract all prominent information from the images and provide more natural output with increased visual quality. Therefore, it can be concluded from Fig. 6 and Table I that both the visual and statistical evaluation proves the superiority of the proposed method over existing methods.

C. Clinical Examples on PET/MRI and SPECT/MRI Image Fusion

Despite the great success of the MRI-CT fusion, its role in neuroscience is considered to be limited compared with the potential of PET-MRI and SPECT/MRI fusion. PET can provide functional eloquent brain areas such as motor or speech regions by using specific activation tasks. On the other hand, single-photon emission computed tomography (SPECT) images reveal the metabolic change that has significant clinical values. Therefore, in modern era PET/MRI and SPECT/MRI fusion are analyzed over MRI-CT fusion for the better diagnosis in different diseases. In order to demonstrate the practical value of the proposed scheme in medical imaging, three clinical cases are considered where PET/MRI and SPECT/MRI medical modalities are used. These includes the case of Alzheimer, subacute stroke and brain tumor respectively. The images have been downloaded from the Harvard university site (http://www.med.harvard.edu/AANLIB/home.html).

The first case is of a 70 year-old man who began experiencing difficulty with memory about 9 months prior to imaging. He had a history of atrial fibrillation and was taking warfarin. He had become lost on several occasions, and had difficulty orienting himself in unfamiliar circumstances. This man is affected by the diseases namely Alzheimer (highlighted by red arrows). Fig. 7(a)–(b) shows the MRI and PET images the person. MRI image showed a globally widened hemispheric sulci, which is more prominent in parietal lobes. Regional cerebral metabolism is markedly abnormal, with hypometabolism in anterior temporal and posterior parietal regions. These changes are bilateral, but the right hemisphere is slightly more affected than the left, and the posterior cingulate is relatively spared.

Fig. 8 shows the subacute stroke case of a 65 year old man who suddenly experienced tingling in the left hand and arm, and on examination had a syndrome of left neglect: he failed to explore the left half of space, and extinguished both left tactile and left visual stimuli when presented on both sides simultaneously. The MRI study revealed that the frontal pole in the old infract is replaced with the high signal of cerebrospinal fluid left after liquifaction necrosis (highlighted by red arrow). The beginning of new symptoms corresponds to the right parietal infarction with hyperperfusion. There is a subtle abnormality in the MRI image and a luxury hyperperfusion in the SPECT image (highlighted by red arrow).

Fig. 9 shows the recurrent tumor case of a 51 year old woman sought medical attention because of gradually increasing right hemiparesis (weakness) and hemianopia (visual loss). At craniotomy, left parietal anaplastic astrocytoma was found. A right frontal lesion was biopsied. The evolution of high tumor Thallium uptake, indicating astrocytoma recurrence is revealed by the SPECT study, which is pointed by an red arrow in the SPECT image whereas a large region of mixed signal on MRI image gives the signs of the possibility of active tumor (encircled).

Here, the results are compared with the best four algorithms obtained with the earlier analysis, i.e., Guihong *et al.* [20], Yang *et al.* [22], Zhang and Guo [11] and Chai *et al.* [12]. From Figs. 7–9, it can be observed that all the fusion algorithms have fairly good spatial information but the spectral distortions are somewhat high in the existing algorithms, i.e., spectral information is lost in the case of existing algorithms which is greater in the case of [20] and comparatively lesser in [11], [12], [22]. The color information is also distorted in the existing algorithms

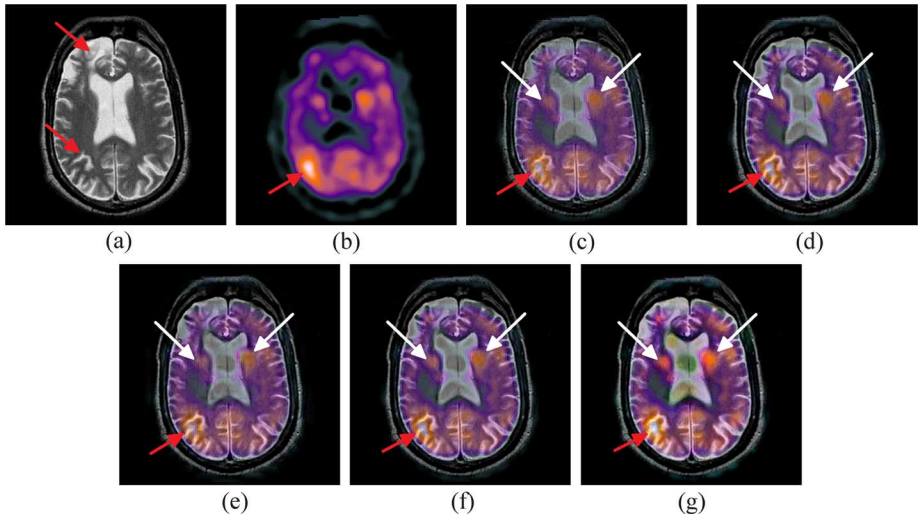

Fig. 8. Brain images of the man affected with Subacute Stroke: (a) MRI image (b) SPECT image; Fused images by (c) wavelet (d) contourlet (e) NSCT-1 (f) NSCT-2 (g) proposed method.

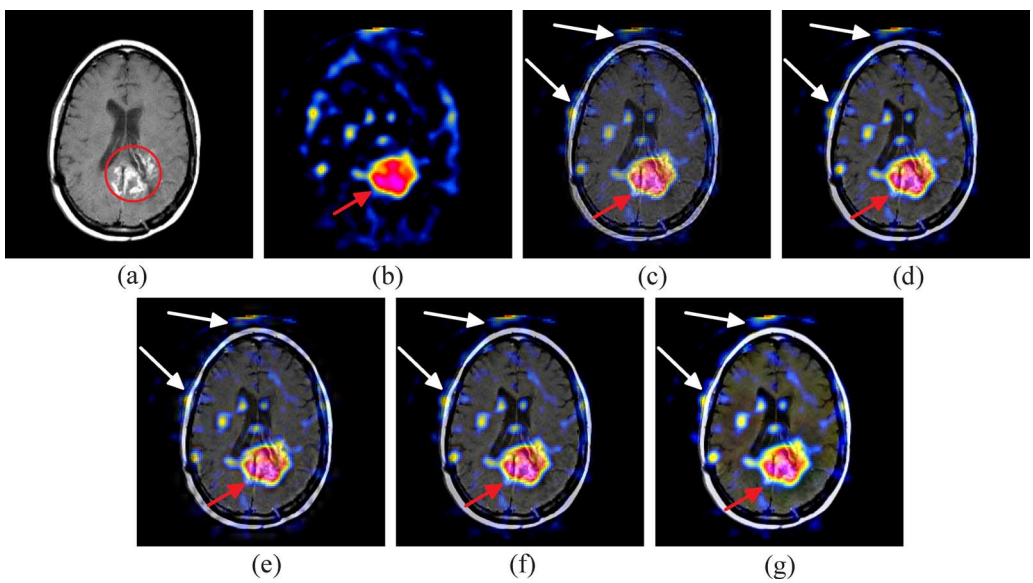

Fig. 9. Brain images of the man with Recurrent Tumor: (a) MRI image (b) SPECT image; Fused images by (c) wavelet (d) contourlet (e) NSCT-1 (f) NSCT-2 (g) proposed method.

TABLE II
EVALUATION INDICES FOR FUSED MEDICAL IMAGES IN CLINICAL EXPERIMENTS

Disease	Indices	Wavelet [20]	Contourlet [22]	NSCT-1 [11]	NSCT-2 [12]	Proposed
Alzheimer	Q_{MI}	1.3249	1.4444	1.3425	1.4834	**1.5017**
	Q_S	0.6972	0.7521	0.7537	0.7576	**0.7972**
	$Q^{AB/F}$	0.5165	0.6505	0.6535	0.6548	**0.6722**
Subacute Stroke	Q_{MI}	1.2232	1.2788	1.2816	1.2853	**1.3740**
	Q_S	0.8427	0.8664	0.8675	0.8842	**0.8907**
	$Q^{AB/F}$	0.5066	0.6086	0.6133	0.6126	**0.6278**
Recurrent Tumor	Q_{MI}	0.9903	1.0112	1.7984	1.8116	**1.9809**
	Q_S	0.7845	0.7988	0.8047	0.8175	**0.8248**
	$Q^{AB/F}$	0.5440	0.6778	0.6791	0.6819	**0.6875**

(shows with the white arrows). On the contrary, the color information is least distorted and the spatial details are as clearer as the original MRI image, and the spectral features are also natural. This fact is also justified from the Table II, where the proposed method are with higher evaluation indices among all methods. These evaluation indices are defined for gray-scale images and the PET image is a color image. Therefore, these metrics are evaluated with each color channel in turn and then take the average of all values as the final result. Therefore, the proposed method not only preserve the crucial features exist in both original images but also improves the color information when compared to existing methods.

V. Conclusion

In this paper, a novel image fusion framework is proposed for multi-modal medical images, which is based on non-subsampled contourlet transform and directive contrast. For fusion, two different rules are used by which more information can be preserved in the fused image with 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 experiment, two groups of CT/MRI 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 can enhance the details of the fused image, and can improve the visual effect with much less information distortion than its competitors. These statistical assessment findings agree with the visual assessment. Further, in order to show the practical applicability of the proposed method, three clinical example are also considered which includes analysis of diseased person's brain with alzheimer, subacute stroke and recurrent tumor. A MATLAB implementation of the proposed algorithm is available online at sites.google.com/site/goravdma/Home/code/project1.

References

[1] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proc. IEEE*, vol. 91, no. 10, pp. 1699–1721, Oct. 2003.

[2] G. Bhatnagar, Q. M. J. Wu, and B. Raman, "Real time human visual system based framework for image fusion," in *Proc. Int. Conf. Signal and Image Processing*, Trois-Rivieres, Quebec, Canada, 2010, pp. 71–78.

[3] A. Cardinali and G. P. Nason, "A statistical multiscale approach to image segmentation and fusion," in *Proc. Int. Conf. Information Fusion*, Philadelphia, PA, USA, 2005, pp. 475–482.

[4] P. S. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis," *Photogrammetric Eng. Remote Sens.*, vol. 55, pp. 339–348, 1989.

[5] A. Toet, L. V. Ruyven, and J. Velaton, "Merging thermal and visual images by a contrast pyramid," *Opt. Eng.*, vol. 28, no. 7, pp. 789–792, 1989.

[6] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.

[7] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.

[8] A. Toet, "Hierarchical image fusion," *Mach. Vision Appl.*, vol. 3, no. 1, pp. 1–11, 1990.

[9] X. Qu, J. Yan, H. Xiao, and Z. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1508–1514, 2008.

[10] G. Bhatnagar and B. Raman, "A new image fusion technique based on directive contrast," *Electron. Lett. Comput. Vision Image Anal.*, vol. 8, no. 2, pp. 18–38, 2009.

[11] Q. Zhang and B. L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, 2009.

[12] Y. Chai, H. Li, and X. Zhang, "Multifocus image fusion based on features contrast of multiscale products in nonsubsampled contourlet transform domain," *Optik*, vol. 123, pp. 569–581, 2012.

[13] G. Bhatnagar and Q. M. J. Wu, "An image fusion framework based on human visual system in framelet domain," *Int. J. Wavelets, Multires., Inf. Process.*, vol. 10, no. 1, pp. 12500021–30, 2012.

[14] S. Yang, M. Wang, L. Jiao, R. Wu, and Z. Wang, "Image fusion based on a new contourlet packet," *Inf. Fusion*, vol. 11, no. 2, pp. 78–84, 2010.

[15] Q. Miao, C. Shi, P. Xu, M. Yang, and Y. Shi, "A novel algorithm of image fusion using shearlets," *Opt. Commun.*, vol. 284, no. 6, pp. 1540–1547, 2011.

[16] S. Li, B. Yang, and J. Hu, "Performance comparison of different multiresolution transforms for image fusion," *Inf. Fusion*, vol. 12, no. 2, pp. 74–84, 2011.

[17] R. Redondo, F. Sroubek, S. Fischer, and G. Cristobal, "Multifocus image fusion using the log-Gabor transform and a multisize windows technique," *Inf. Fusion*, vol. 10, no. 2, pp. 163–171, 2009.

[18] S. Yang, M. Wang, Y. Lu, W. Qi, and L. Jiao, "Fusion of multiparametric SAR images based on SW-nonsubsampled contourlet and PCNN," *Signal Process.*, vol. 89, no. 12, pp. 2596–2608, 2009.

[19] Y. Chai, H. Li, and X. Zhang, "Multifocus image fusion based on features contrast of multiscale products in nonsubsampled contourlet transform domain," *Optik—Int. J. Light Electron Opt.*, vol. 123, no. 7, pp. 569–581, 2012.

[20] Q. Guihong, Z. Dali, and Y. Pingfan, "Medical image fusion by wavelet transform modulus maxima," *Opt. Express*, vol. 9, pp. 184–190, 2001.

[21] V. Barra and J. Y. Boire, "A general framework for the fusion of anatomical and functional medical images," *Neuro Image*, vol. 13, no. 3, pp. 410–424, 2001.

[22] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, pp. 203–211, 2008.

[23] F. E. Ali, I. M. El-Dokany, A. A. Saad, and F. E. Abd El-Samie, "Curvelet fusion of MR and CT images," *Progr. Electromagn. Res. C*, vol. 3, pp. 215–224, 2008.

[24] N. Boussion, M. Hatt, F. Lamare, C. C. L. Rest, and D. Visvikis, "Contrast enhancement in emission tomography by way of synergistic PET/CT image combination," *Comput. Meth. Programs Biomed.*, vol. 90, no. 3, pp. 191–201, 2008.

[25] S. Daneshvar and H. Ghassemian, "MRI and PET image fusion by combining IHS and retina-inspired models," *Inf. Fusion*, vol. 11, no. 2, pp. 114–123, 2010.

[26] Y. Yang, D. S. Park, S. Huang, and N. Rao, "Medical image fusion via an effective wavelet-based approach," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 44-1–44-13, 2010.

[27] S. Das, M. Chowdhury, and M. K. Kundu, "Medical image fusion based on ripplet transform type-I," *Progr. Electromagn. Res. B*, vol. 30, pp. 355–370, 2011.

[28] T. Li and Y. Wang, "Biological image fusion using a NSCT based variable-weight method," *Inf. Fusion*, vol. 12, no. 2, pp. 85–92, 2011.

[29] A. L. da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[30] P. Kovesi, "Image features from phase congruency," *Videre: J. Comput. Vision Res.*, vol. 1, no. 3, pp. 2–26, 1999.

[31] P. Kovesi, "Phase congruency: A low-level image invariant," *Psychol. Res. Psychologische Forschung*, vol. 64, no. 2, pp. 136–148, 2000.

[32] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 493–500, 2007.

[33] A. B. Watson, "Efficiency of a model human image code," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2401–2417, 1987.

[34] D. L. Ruderman, T. W. Cronin, and C. C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *J. Opt. Soc. Amer. A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[35] C. Yang, J. Zhang, X. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, pp. 156–160, 2008.

Gaurav Bhatnagar (M'10) is with Computer Vision and Sensing Systems Laboratory, Department of Electrical and Computer Engineering at the University of Windsor, Windsor, ON, Canada. He received his Ph.D. degree in interdisciplinary (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 in 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 40 journal articles and conference proceedings, and has contributed to two books in his area of interest.

Q. M. Jonathan Wu (M'92–SM'09) 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 ten 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 (M'03–SM'05) 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, where he currently holds a research officer position. He holds an adjunct professorship at the University of Ottawa. 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 member of SPIE.