Received: 26 September 2019 | Revised: 11 December 2019 | Accepted: 15 December 2019

DOI: 10.1002/ima.22395

RESEARCH ARTICLE

WILEY

A novel region-based multimodal image fusion technique using improved dictionary learning

Bikash Meher[1] | **Sanjay Agrawal**[1] | **Rutuparna Panda**[1] | **Lingraj Dora**[2] | **Ajith Abraham**[3]

[1]Department of Electronics & Telecommunication Engineering, VSS University of Technology, Burla, India

[2]Department of Electrical and Electronics Engineering, VSS University of Technology, Burla, India

[3]MIR Labs, Washington, USA

Correspondence
Rutuparna Panda, Department of Electronics & Telecommunication Engineering, VSS University of Technology, Burla, India.
Email: r_ppanda@yahoo.co.in

Abstract

Recently, the sparse representation (SR) based algorithms have gained much attention from the researchers in the area of image fusion (IF). The building of a compact discriminative dictionary plays a vital role in the sparse-based IF techniques. In this context, an efficient multimodal IF method based on improved dictionary learning is investigated. The key contributions of this paper are: (a) An improved KSVD algorithm is suggested for the dictionary learning process, (b) to reduce the computational time, only the informative patches are selected using energy feature, and (c) a novel region-based fusion scheme is suggested for the first time for the problem on hand. The suggested technique is tested with a number of multimodal images from Harvard Medical School brain database. The results are compared with state-of-the-art multiscale transform-based methods and modified SR-based methods. Unlike earlier methods, our proposed technique generates an adaptive dictionary through selection of informative patches only. This results in a compact dictionary with improved computational efficiency. The experimental results reveal that our approach outperforms other methods. The potential application of the suggested method could be in pathological images for follow-up study and better treatment planning.

KEYWORDS

dictionary learning, image fusion, region-based fusion, sparse representation

1 | INTRODUCTION

The development in emerging medical imaging techniques has drawn attention of researchers in the field of health care. The different imaging techniques provide both complementary and redundant information. For example, computed tomography (CT) can represent a dense structure of bones and hard tissue with a smaller amount of distortion, while MRI is better in visualizing soft tissues. Similarly, positron emission tomography (PET) image provides info regarding blood flow having low spatial details. A single modality image does not give complete and correct info. So, to get both the anatomical and functional information in a single image, the different modality images need to be merged. The multimodal IF tries to combine info from manifold modality images to get a more comprehensive and precise picture of the source images. As we get a single medical image using multimodal IF, it is not only useful for disease diagnosis but also lowers the storage cost.

Numerous works have been reported so far on IF techniques.[1-3] Some of the methods are used for

multimodal medical IF.[4-6] These methods are classified into three levels: pixel, feature, and decision.[7] Most of the existing multimodal medical IF is carried out on pixel level.[8,9] Pixel-level IF possess advantages like ease of implementation, computational efficiency, and so on. But it has certain disadvantages such as poor contrast and blurring effect. To avoid these, a more convenient fusion process which follows the systematic fusion rule is the region-based IF.[10] In this technique, it takes a block of pixels in place of a sole pixel for the fusion process.

Usually, the IF methods are categorized into two types: spatial domain and transform domain.[11] In spatial domain, fusion generally takes the spatial information of pixels.[12] This method is more appropriate for the fusion of images which are taken from the same imaging sensors, that is, multifocus [12] and multi-exposure fusion.[13] In transform domain technique, the information contained in them is combined after transforming.[14] The output image is got by taking the inverse transformation of the merged coefficients. A variety of transform domain techniques have been suggested utilizing multiscale transform (MST). These are Laplacian pyramid (LP),[15] wavelet transform,[16] contourlet transform (CT),[17] non-subsampled contourlet transform (NSCT),[18] sparse representation (SR),[20] and so on. As this image representation approach is suitable for human visual perception, transform domain techniques are found to be very useful in multimodal medical IF.[21] Here, the inputs are taken from various imaging devices. But the disadvantage of transform-based techniques is the fewer inclusion of spatial info in their pixel coefficient selection. Because of which the methods fail to preserve the edge and texture info leading to distortion in the output images. Another disadvantage of the method is the absence of a wavelet kernel or overcomplete dictionary to process various images. Further, it is necessary to predetermine the number of decomposition levels.

Lewis et al[16] presented a transform-based fusion technique employing dual-tree complex wavelet transform (DT-CWT). They used it for decomposition of the source images and merge them employing the fusion rules. The drawback of this transform is its poor ability to identify the contours and edges of the image in fusion. Li et al[22] suggested an IF technique using guided filter. The filter uses an edge preserving smoothing process for the fusion. However, the implementation of the filter is complex and time consuming. Also, it produces a halo effect in the fused images. Yin et al [23] presented a multimodal medical IF scheme based on non-subsampled shearlet transform (NSST). The authors used the NSST to decompose the input images. The parameter-adaptive pulse-coupled neural network (PA-PCNN) model is used for the fusion of high-frequency components. The low-frequency components are fused using energy preservation and detail extraction. The fused image is then obtained using inverse NSST. Zhu et al[24] proposed a multimodality medical IF using phage congruency and Laplacian energy in NSCT domain. The authors decomposed the source images into high-pass and low-pass bands using NSCT domain. The high-pass bands are fused using phase congruency and low-pass bands are fused using Laplacian energy. The output image is obtained by taking the inverse transform. However, the method has higher computational cost than the NSST method. Further, it does not give good fusion results for PET-MRI images.

The SR technique is an emerging technique successfully applied in various IF usages.[24,25] It is perceived from the works that the techniques employing SR tend to give better fusion results than the traditional MST-based techniques. In SR model, the design of an overcomplete dictionary plays an important role. The dictionary can be constructed in two ways: (a) preconstructed fixed dictionary using the different analytical paradigms like wavelet, discrete cosine transform (DCT), contourlet, and so on; and (b) learned dictionary constructed using high-quality natural images.

Liu et al[26] presented an IF technique integrating MST and SR. The dictionary is built using 40 high-resolution natural images. The authors used MST method to get low-pass and high-pass coefficients. The low-pass coefficients are fused with SR-based methods. However, the selection of high-quality natural images is a difficult task. Further, a large quantity of patches is needed for dictionary learning. Yang and Li[19] proposed a multifocus IF scheme employing SR. The researchers utilized the DCT dictionary which is a preconstructed dictionary. The disadvantage of such an approach is the production of a large number of patches and dependence on the input images.

Zong and Qiu[27] presented a medical IF scheme utilizing SR and learned dictionary. The authors generated sub-dictionaries using classified patches. They used online dictionary approach and least angle regression algorithm to get the sparse coefficients of each patch. However, the authors are silent on the number of sub-dictionaries created. In Reference,[28] a joint sparsity model is introduced for IF. The major problem with the method was the use of total image patches obtained for the dictionary learning process. So, the training process took more execution time. Also, the dictionary became very much redundant. Yin[29] proposed a joint sparsity method for fusion. However, the functioning of the suggested technique strongly depended on external pre-collected dataset for the construction of the dictionary. Yang and Li[30] expanded their method by proposing an algorithm named as simultaneous orthogonal matching

pursuit (SOMP). Their technique is experimented on various types of images. However, the dictionary used in this technique is image reliant, which limits the flexibility of the method. In Reference,[32] the authors introduced an adaptive SR (ASR) scheme for IF. They used the histogram of gradient information method for dictionary learning. A group of six compact dictionaries is formed using this method. High computation complexity is the limitation of this technique. To decrease the computational complexity, Kim et al[33] suggested K-means method, PCA method for dictionary learning. The technique suggested by the authors produced a compact and informative dictionary. However, the number of clusters in the K-means technique has to be predefined. Liu et al[32] suggested convolutional sparse representation based on morphological component analysis (CS-MCA) for the medical IF at pixel level. In this method, for every input image, first the convolutional sparse representations of cartoon and texture components are obtained by CS-MCA model employing the pre-learned dictionaries. Finally, the output image is computed as the superposition of the fused cartoon and texture components.

From the above discussions, it is observed that the prevailing fusion procedures use either a prebuilt dictionary (with less preprocessing time) or a learned dictionary (with a priori information). So, the building of an overcomplete dictionary is a vital section. Inspired from the study, here a novel adaptive dictionary learning process is suggested. In medical images, all the image patches may not contain useful information. So, it is not wise to use all the image patches for the construction of a dictionary, which may affect the fusion performance. In order to get a good fused image, the informative patches are selected from all the patches to form a good dictionary.

Most of the SR-based IF approaches suggested till now are based on the pixel level. Here, the fusion is performed using the region-level approach, as it suppresses the shortcomings of pixel-based IF described earlier. An informative sampling procedure using the energy feature of the patches from the input images is applied for the dictionary learning procedure. The computation of variance of each patch helps in discarding the zero informative patches. A suitable threshold value is applied to choose the useful patches for the dictionary learning. Then the output image is found by using the region-based fusion rule. The main contributions of this work are:

1. Only the informative patches are selected using energy feature which is used as a training data for dictionary learning. The advantages of doing this is that it increases the computation efficiency.
2. An improved KSVD algorithm is suggested. The dictionary learning process is improved by modifying the sparse coding stage.
3. A region-based instead of pixel-based fusion scheme is suggested for the first time for the problem on hand.

The rest of the manuscript is arranged as follows: Section 2 explains the sparse theory. Section 3 describes the proposed scheme. Results and discussions are elaborated in Section 4. Lastly, the conclusions are drawn in Section 5.

2 | SPARSE THEORY

The SR is a tool for finding the sparsity in the natural signals. The primary idea is that a natural signal $x \in K^n$ (K is real) is approximated by a linear summation of a few atoms from an overcomplete dictionary $D \in K^{n \times m}(n < m)$, where n represents size of the signal x and m represents the size of the dictionary. So, the signal x is represented as $x \simeq D\alpha$, where $\alpha \in K^m$ represents sparse coefficient vector. SR goal is to evaluate the sparsest coefficients α that consists of least nonzero components amid all the possible solutions. This is an optimization problem and is expressed as

$$\min_{\alpha} \|\alpha\|_0 \text{ subject to } \|x - D\alpha\|_2 < \varepsilon \tag{1}$$

where ε is the error ($\varepsilon > 0$) and $\|\cdot\|_0$ is the l_0 norm of α which counts the nonzero factors. This optimization problem is solved using some pursuit algorithms. The most popular technique is the orthogonal matching pursuit (OMP).[35]

In sparse modeling, the role of the dictionary is very significant. The dictionary is constructed in two ways. The first one is the analytical methods in which the dictionary is constructed using the wavelet transform, DCT, curvelet transform, and so on. The advantage of this technique is: simple and fast in execution. Nevertheless, this dictionary is not applicable to an arbitrary signal. The second category uses machine-learning approaches. The dictionary is learned from a large number of sample patches taken from example images employing a specific training algorithm. Most commonly, the K-singular value decomposition method (KSVD)[36] is preferred to solve this type of problem. This algorithm is briefly explained below. P image patches having size ($\sqrt{n} \times \sqrt{n}$) are extracted from the example images. The image patches are arranged in a column vector in the K^n space. Then the training database $Z = \{z_i\}_{i=1}^{P}$ is constructed with

every $z_i \in K^n$. The model of dictionary learning is expressed as

$$\min_{D,\{\alpha_i\}_{i=1}^{P}} \sum_{i=1}^{P} \|\alpha_i\|_0 \text{ subject to } \|z_i\text{-}D\alpha_i\|_2 < \varepsilon \quad i = 1,2,...,P \quad (2)$$

where ε is the error ($\varepsilon > 0$), $D, \{\alpha_i\}_{i=1}^{P}$ is represented as the unknown sparse vector corresponding to $\{z_i\}_{i=1}^{P}$ and the unknown dictionary to be learned is represented as $D \in K^{n \times m}$. Conventionally, α (sparse coefficient) is taken fixed or predetermined and then D (dictionary) is updated. However, this updating is not optimal. An improved KSVD approach is suggested in this paper, where the variables D and α are updated following the rules given in the next section.

3 | SUGGESTED FUSION APPROACH

3.1 | Improved dictionary learning process

A compact and discriminative overcomplete dictionary used for the IF produces a better fused image. To build such a dictionary, the image patches are selected using the energy of the patches. The block diagram of the dictionary learning process is shown in Figure 1. Assume that the input images of size ($M \times N$) are registered. The patches are taken from the two input images using the sliding window procedure using a unity step length. The patch size is taken as $w \times w$ ($w = 8$). The selection of patch size is decided from the test results. This is also discussed in various sources. Let the ith patch of the input images A and B be P_{Ai} and P_{Bi}, respectively. The total number of patches obtained using the sliding window technique of patch size $w \times w$ is $(M - w + 1)(N - w + 1)$. Usually, all the image patches do not contain suitable clinical info. So, utilizing all the image patches results in a redundant dictionary and also takes more time for the learning process. For these reasons, only the useful patches are chosen from the total image patches. The local variance of the images is utilized to express the image details. The patches having variance greater than or equal to 5 are selected. It is to be noted that the threshold value of 5 is taken after an exhaustive experiment with different values ranging from 1 to 20. Let the selected patches obtained after this operation be denoted as $P_A = \{\bar{P}_A^q\}_{q=1}^{K1}$ and $P_B = \{\bar{P}_B^r\}_{r=1}^{K2}$, where *K1* and *K2* are the remaining patches after discarding the zero informative patches. For the construction of a good quality dictionary, the training data should contain informative patches as it provides a richer data representation than a dictionary constructed with a traditional procedure. In this paper, the informative patches are selected by calculating the energy of each patch. The energy (E_P) is determined using the following expression for ith informative patches.

$$E_{P_i} = \sum P_i^2 \quad (3)$$

The energy of every patch in the input images is calculated using (3) and denoted as $E_A = \{E_A^q\}_{q=1}^{K1}$ and $E_B = \{E_B^r\}_{r=1}^{K2}$. A threshold value to improvise the performance is defined below:

$$TH_v = 0.1 * \max(E_v), v \in (A,B) \quad (4)$$

where E_v of the vth input image computes the energy of the subsequent patches. The threshold value is selected such that the blocks of P_A and P_B preserve the characteristics (low and high level) of the subsequent input images. The value of 0.1 is chosen empirically which reduces the number of patches required. Appropriate rule is utilized to get the patch with sufficient energy in it. This set forms the training dataset for the dictionary learning procedure. Let the training dataset be denoted as TR, $\{TR_l | l = 1, 2, 3, ..., L\}$. The training dataset is formed by comparing the energy of every patch of P_A and

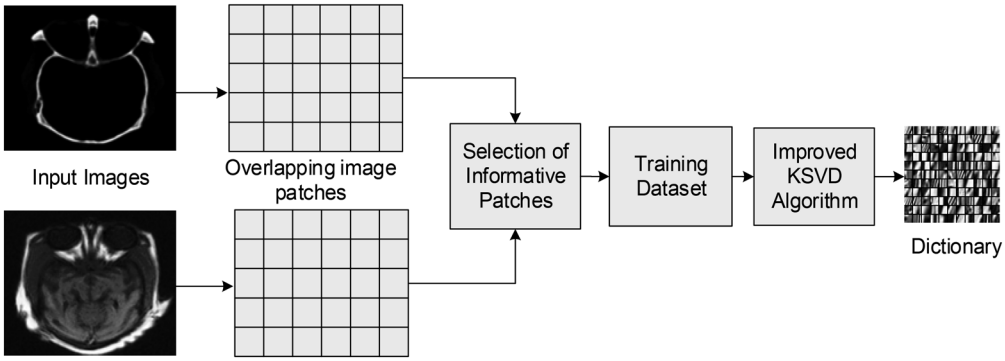

FIGURE 1 Block diagram of the dictionary learning approach

P_B with their corresponding threshold values. For example, if $E_A^q > Th_A$, select $\bar{P}_A^q$, else ignore $\bar{P}_A^q$.

$$TR = \begin{cases} select\, \bar{P}_A^q, & if\, E_A^q > Th \\ Ignore\, \bar{P}_A^q, & else. \end{cases} \tag{5}$$

Equation (5) is utilized and iterated till all the useful patches are obtained for the dictionary learning process. The input images are of size (256 × 256). The patches are of size (8 × 8) and the step size is unity. Each image produces 62 001 number of image blocks. So, the total number of patches for the dictionary learning will be 124 002. Applying the proposed dictionary learning approach, the training dataset is reduced to approximately 38 000 image blocks (ie, 31%). The advantage is that the processing time is reduced and more memory space is available. Further, the average value of every block in the training dataset is reduced to zero. Finally, to obtain an adaptive dictionary, the improved K-SVD procedure is proposed. The dictionary is formed initially by using the training data which is obtained from the image patches. The KSVD algorithm performs in two steps: (a) sparse coding and (b) dictionary update. The algorithm follows the alternation between the sparse coding stage and the dictionary update stage. In this paper we have improved the sparse coefficient update using some modification in the error tolerance. The OMP algorithm is used to estimate the sparse vector.

The algorithm to show the process of updating D and α is presented below.

1. Form the dictionary D using the training dataset TR.
2. Find the best sparse coefficient α using

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \|TR - D\alpha\|_2, \tag{6}$$

$$\text{such that } \|TR - D\alpha\|_2^2 \le \varepsilon$$

where $\varepsilon = \sqrt{\frac{1}{2}\log(m)}$ and m is the number of columns of the dictionary D. The sparse coefficient is computed utilizing the OMP[35] technique.

3. For this value of α, compute D by

$$\mathrm{D} = \underset{D}{\operatorname{argmin}} \|TR - D\alpha\|_2 + \beta\|\alpha\|_1, \tag{7}$$

$$\text{such that } a \le \|d_i\|_2^2 \le b, \qquad i = 1, \cdots, P$$

where a, b is in the range [0,1] and β represents the regularization parameter in the range [0,1]. The lower bound of the errors in the signals is taken as proposed by Lee et al.[35]

4. Iterate between steps 2 and 3 until convergence.

3.2 | Fusion method

The framework of the suggested fusion method is depicted in Figure 2. In this process, the input images are divided into overlapping patches. Using lexicographic ordering, the patches are converted to column vectors. The average value of each block is normalized to zero. To get the sparse coefficients, the OMP procedure is used with the dictionary. Then, the images are reconstructed from the sparse coefficients.

Here, the region-based approach is introduced. The following steps are employed:

1. The average image is obtained using the two registered source images.
2. The average image is partitioned into different regions using the fuzzy c-means clustering (FCM) segmentation algorithm which is new in this application. Each source image is segmented into three clusters.
3. The images A and B are divided utilizing the results of step (2).
4. The energy of each segment of the source images is computed based on the segmented images.
5. The image is reconstructed from the sparse coefficients (α_A and α_B). Each vector of the sparse matrix is reshaped to $w \times w$ patch. Then the patches are moved to their original respective position. The overlapping of patches occurs during this process. So, a simple mean operation is employed to all the overlapping patches.
6. The energy of each region of the source images is compared. The regions are selected for fusion following Equation (7a).

$$ROF_i = \begin{cases} RO\alpha_{Ai} & \mathrm{E}_i^A \le \mathrm{E}_i^B \\ RO\alpha_{Bi} & \mathrm{E}_i^A > \mathrm{E}_i^B \end{cases} \tag{7a}$$

where ROF_i is the ith region of the fused image, $RO\alpha_{Ai}$ and $RO\alpha_{Bi}$ are the regions of the image reconstructed from the sparse coefficients α_{A} and α_{B}, respectively, E_i^A and E_i^B are the energy of the i^{th} region of the source images A and B, respectively.

7. All the regions obtained are integrated to get the fused image.

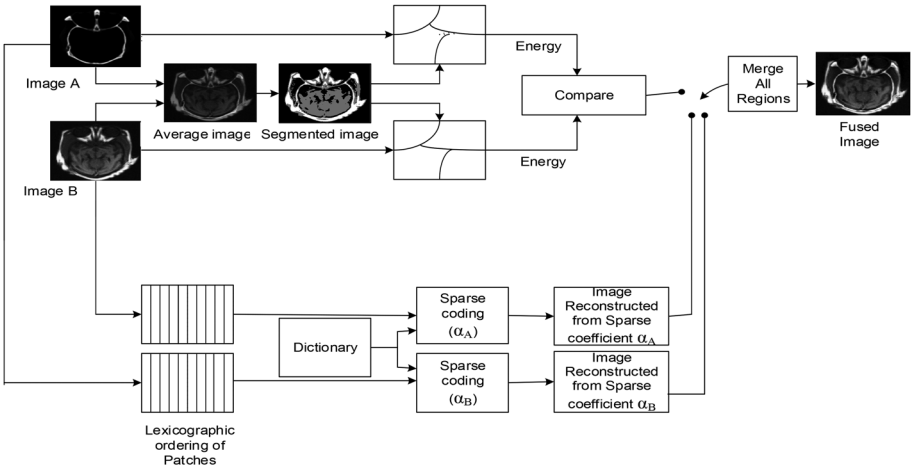

FIGURE 2 The framework of the proposed scheme

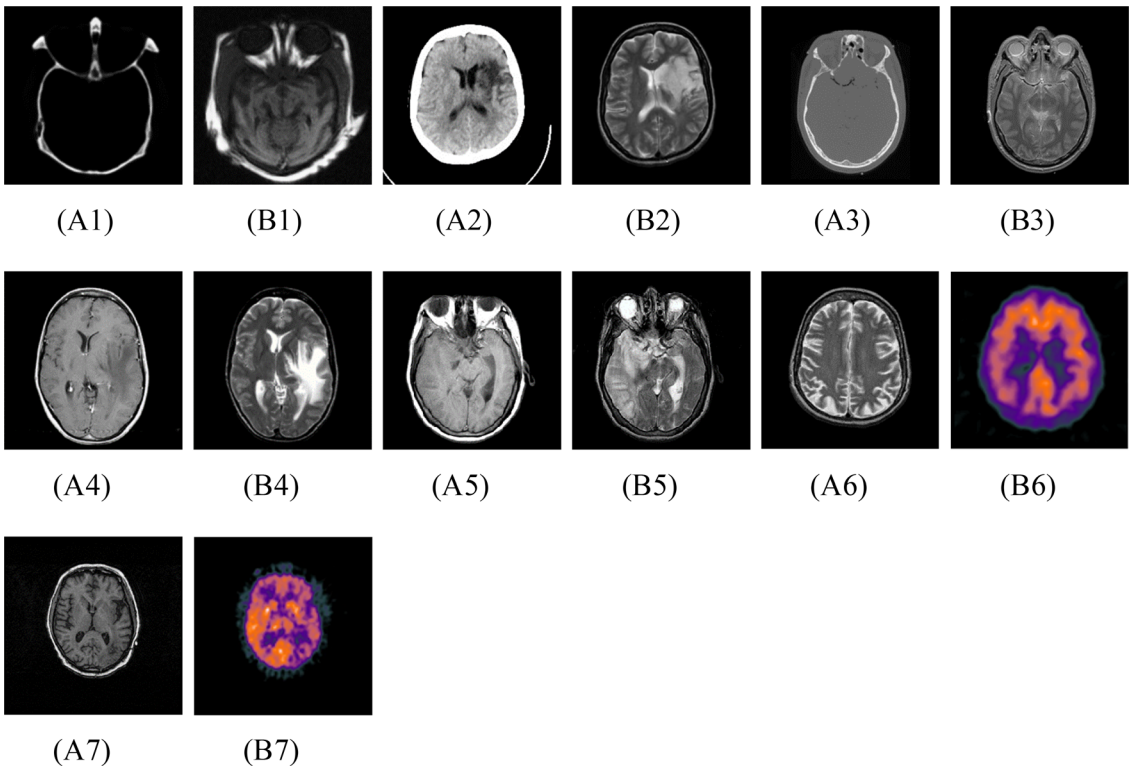

FIGURE 3 Input multimodal medical images: {((A1), (B1)), ((A2), (B2))} group 1 (CT and MRI); {((A3), (B3)), ((A4), (B4)), ((A5), (B5))} group 2 (MR-T1 and MR-T2); ((A6), (B6)) group 3 (MRI and SPECT); and ((A7), (B7)) group 4 (MRI and PET) [Color figure can be viewed at wileyonlinelibrary.com]

4 | RESULTS AND DISCUSSIONS

To validate the suggested method, benchmark image datasets are used. In this paper, we have used the different multimodality image pairs like CT and MRI, T1-weighted MRI (MR-T1) and T2-weighted MRI (MR-T2), MRI and PET, and MRI and SPECT (see Figure 3). The input images comprise two groups of CT and MRI images, three groups of MR-T1 and MR-T2 images, one group of MRI and SPECT images, and one group of MRI and PET images. All the input images are taken from www.med.harvard.edu/aanlib/home.[37] The input images are of the same size, that is, 256×256 pixels. The image patch size is taken as 8×8. The dictionary size is fixed as 256. The values of a and b are set as 0.4 and 1, respectively. The parameter β is fixed at 0.02 and the number of iterations for improved KSVD algorithm is set to 50. The experiments are executed in core i5 processor having 8 GB RAM using MATLAB simulating environment.

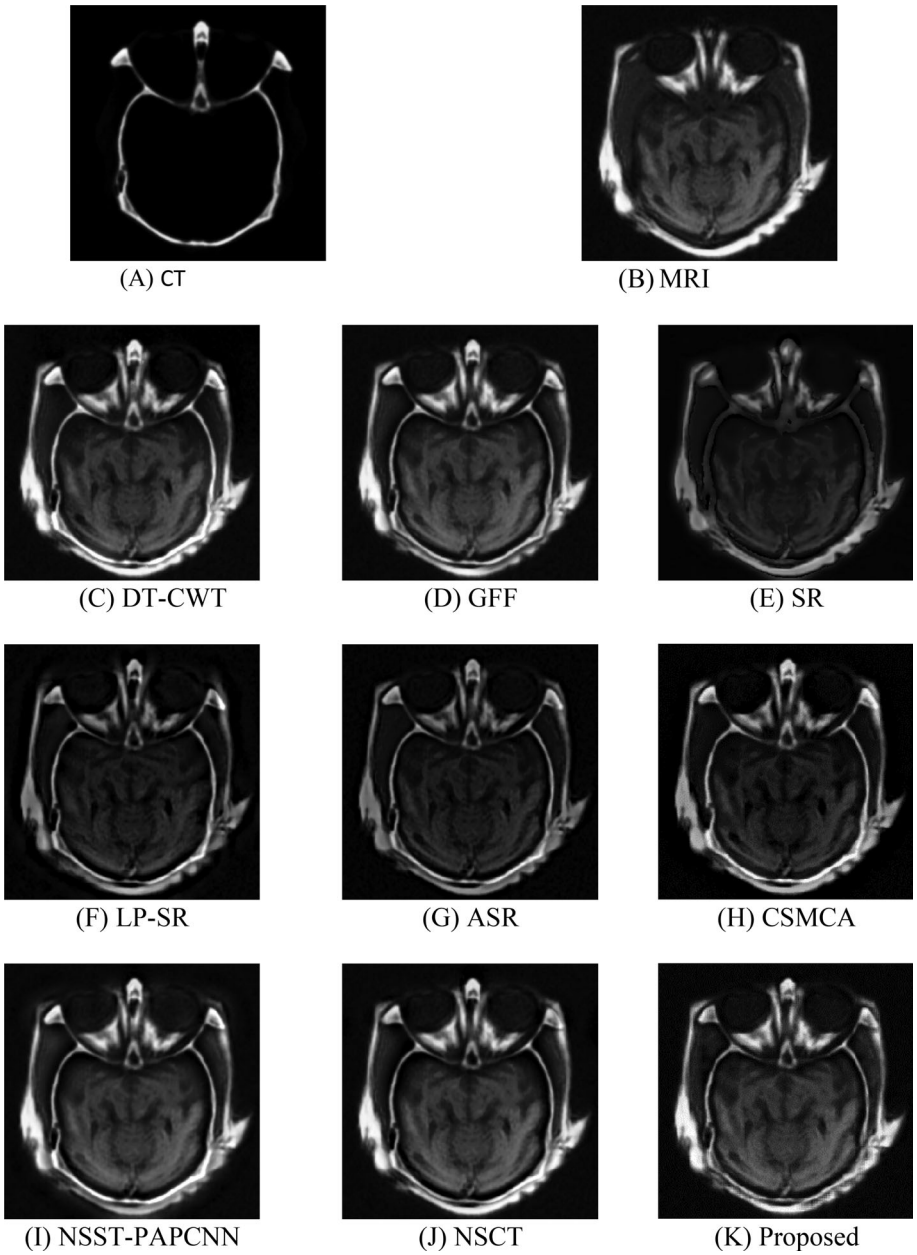

FIGURE 4 Results of fusion for group 1 images. A, CT; B, MRI; C to K, fused images

Different metrics for the assessment of the quality of the fused image are used. These are Petrovic metric ($Q^{RS/F}$),[38] Piella's metrics,[39] entropy (E),[15] mutual information (MI)[40], feature mutual information (FMI),[41] and visual information fidelity for fusion (VIFF).[42] The fusion metrics are briefly explained below:

Petrovic ($Q^{RS/F}$): It calculates the quantity of the edge information conveyed from the input images to the output. It is expressed as

$$Q^{RS/F} = \frac{\sum_{y=1}^{Y}\sum_{z=1}^{Z} Q^{RF}(y,z)w_R(y,z) + Q^{SF}(y,z)w_S(y,z)}{\sum_{y=1}^{Y}\sum_{z=1}^{Z}(w_R(y,z)+w_S(y,z))} \quad (8)$$

where $Q^{RF}(y,z) = \left[Q_g^{RF}(y,z)Q_\alpha^{SF}(y,z)\right]^{1/2}$; $Q_g^{RF}(y,z)$ is the edge strength and $Q_\alpha^{RF}(y,z)$ is the orientation preservation values at the location (y,z); Y and Z represent the size of the images; w_R and w_S represent the weight factors of $Q^{RF}(y, z)$ and $Q^{SF}(y, z)$, respectively. The Q^{SF} is similarly defined as Q^{RF}. The values of $Q^{RS/F}$ vary in the range [0,1]. The $Q^{RS/F}$ value should be nearer to 1.

4.1 | Entropy (E)

It calculates the quantity of information contained in the fused image. The output image comprising a large amount of information has high entropy value. It is stated as

TABLE 1 Objective evaluation performance comparison of group 1 (CT and MRI) images with MST-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	6.2725	2.1933	0.6062	0.9040	0.3488	0.6138	0.5111	0.6832
GFF	6.7971	2.5411	**0.8003**	0.9033	0.4865	0.9220	0.6932	0.8575
NSST-PAPCNN	**6.9552**	2.0553	0.7664	0.9037	0.5566	0.9967	**0.9921**	0.9931
NSCT	6.3275	2.5711	0.7843	**0.9089**	**0.5694**	0.9968	0.9919	0.9930
Proposed	6.8606	**3.0177**	0.7987	0.8801	0.5503	**0.9969**	0.9801	**0.9932**

Note: The bold values indicates the best values.

TABLE 2 Objective evaluation performance comparison of group 1 (CT and MRI) images with SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	6.1928	2.1068	0.5986	0.8920	0.3576	0.5393	0.4369	0.6499
DTCWT-SR	6.7317	2.0873	0.6293	0.9224	0.3837	0.7372	0.5253	0.7214
CVT-SR	6.7921	2.0785	0.5838	**0.9227**	0.3786	0.7003	0.4439	0.6701
NSCT-SR	6.5892	2.1145	0.7461	0.9185	0.4850	0.8388	0.6371	0.8201
SR	6.2388	2.6375	0.7452	0.8971	0.4267	0.9160	0.6220	0.8367
LP-SR	6.8601	2.3236	0.7672	0.9084	**0.4912**	0.8319	0.7009	0.8383
ASR	6.1778	2.7118	0.7747	0.9028	0.3744	0.7237	0.5935	0.7526
CSMCA	6.3275	2.0353	0.7643	0.9089	0.4752	0.6670	0.6592	0.8030
Proposed	**6.8606**	**3.0177**	**0.7987**	0.8801	0.4603	**0.9969**	**0.9701**	**0.9932**

Note: The bold values indicates the best values.

$$E = -\sum_{r=0}^{L-1} P_r \log P_r \tag{9}$$

where P_r is the probability value of the rth gray level in an image.

4.2 | Mutual information (MI)

It computes the quantity of info conveyed from the input images to the output image. The mathematical expression of MI is

$$\mathrm{MI}_F^{RS} = \mathrm{MI}_{FR}(f,a) + \mathrm{MI}_{FS}(f,b) \tag{10}$$

where

$$\mathrm{MI}_{FR}(f,a) = \sum_{f,a} p_{FR}(f,a) \log_2 \frac{p_{FR}(f,a)}{p_F(f) p_R(a)}$$

$$MI_{FS}(f,b) = \sum_{f,b} p_{FS}(f,b) \log_2 \frac{p_{FS}(f,b)}{p_F(f) p_S(b)}$$

where $\mathrm{MI}_{FR}(f, a)$ and $\mathrm{MI}_{FS}(f, b)$ are the normalized MI value between the source images and the output image; p_R, p_S, and p_F represent the gray-level histograms of the input images and the output image, respectively. $p_{FR}(f, a)$ and $p_{FS}(f, b)$ are the normalized joint histograms between the output and the input images. *MI* value should be high.

4.3 | FMI

The FMI metric calculates the mutual information available in all the image characteristics. The most suitable feature is selected for making this procedure more adaptable. The gradient map is very suitable and often utilized as a feature of an image. This map comprises information such as texture, gradients or edge strength, and orientations. The value of FMI varies in the range [0,1] and it should be high.

4.4 | VIFF

The VIFF metric calculates how much distortion is present between the input and the output image. It is based on the consistency of the human visual system.

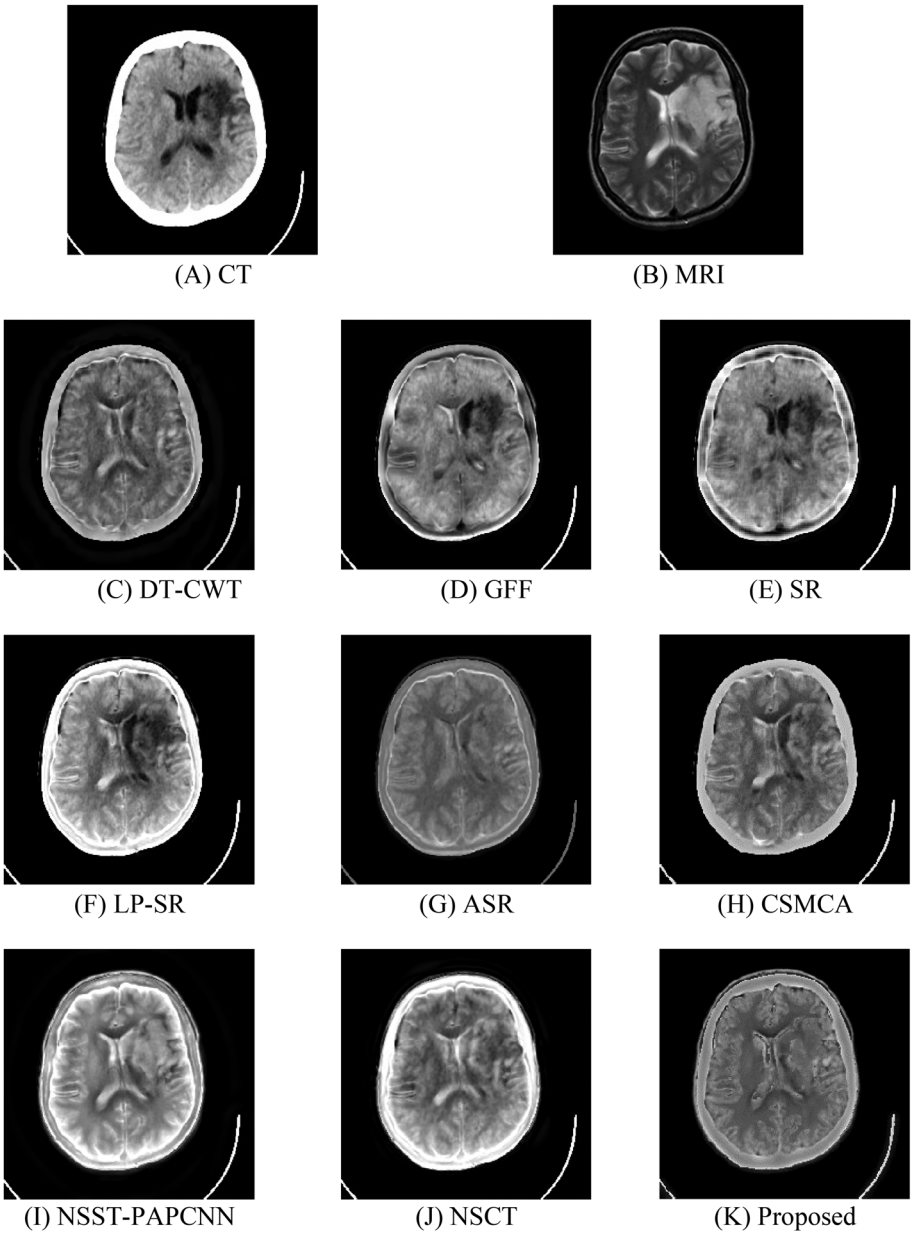

FIGURE 5 Results of fusion for group 1 images. A, CT; B, MRI; C to K, fused images

TABLE 3 Objective evaluation performance comparison of group 1 (CT and MRI) images with MST-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	**5.1557**	2.3611	0.5856	0.7849	0.2338	0.9754	0.9118	0.9615
GFF	4.4949	2.1760	**0.6448**	0.8872	0.2614	0.9935	0.9890	0.9990
NSST-PAPCNN	5.0525	2.1355	0.6446	**0.9091**	0.4118	0.9928	0.9876	0.9991
NSCT	4.7033	2.3510	0.5873	0.9025	**0.6279**	0.9922	**0.9922**	0.9993
Proposed	4.6548	**2.9936**	0.6418	0.8951	0.4835	**0.9965**	0.9869	**0.9995**

Note: The bold values indicates the best values.

TABLE 4 Objective evaluation performance comparison of group 1 (CT and MRI) images with SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	4.3812	2.3727	0.6223	0.8812	0.3480	0.9959	0.9842	0.9958
DTCWT-SR	4.5488	2.9886	0.5782	0.8363	0.2379	0.9944	0.9681	0.9926
CVT-SR	4.5037	2.3530	0.5910	0.6893	0.1625	0.9767	0.8940	0.9569
NSCT-SR	4.5701	2.7701	0.6613	0.8257	0.1835	0.9883	0.9635	0.9834
SR	4.5480	2.5573	0.6304	0.8393	**0.5477**	0.9851	0.9553	0.9964
LP-SR	4.5865	**3.2189**	**0.6822**	0.8092	0.3985	0.9958	0.9588	0.9966
ASR	4.5531	2.9340	0.6637	0.8094	0.5432	0.9815	0.9578	0.9970
CSMCA	3.9945	2.3958	0.5575	**0.9218**	0.3586	0.7768	0.5594	0.7106
Proposed	**4.6548**	2.9936	0.6418	0.8951	0.4835	**0.9965**	**0.9869**	**0.9995**

Note: The bold values indicates the best values.

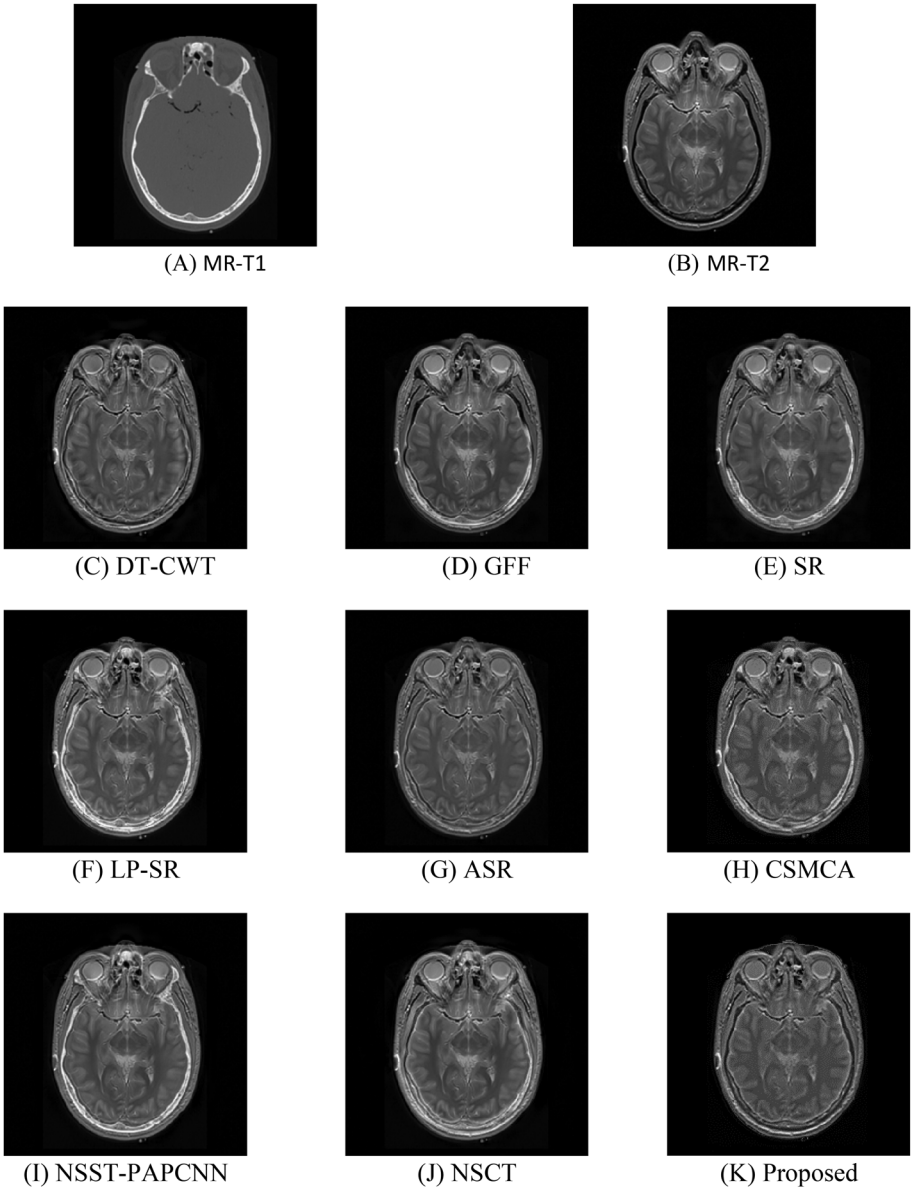

FIGURE 6 Results of fusion for group 2 images. A, MR-T1; B, MR-T2; C to K, fused images

TABLE 5 Objective evaluation performance comparison of group 2 (MR-T1 and MR-T2) images with non-SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	5.2133	2.3992	0.6954	0.8748	0.4579	0.8108	0.3593	0.6518
GFF	5.0910	2.7838	0.7304	**0.8838**	0.4816	0.8941	0.4192	0.6801
NSST-PAPCNN	5.3330	2.0684	0.7091	0.8836	0.6433	0.9992	**0.9919**	0.9961
NSCT	5.3827	2.3601	0.6970	0.8674	**0.6898**	0.9992	0.9908	0.9962
Proposed	**5.6935**	**3.1077**	**0.7637**	0.8642	0.4288	**0.9992**	0.9899	**0.9963**

Note: The bold values indicates the best values.

TABLE 6 Objective evaluation performance comparison of group 2 (MR-T1 and MR-T2) images with SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	5.2020	2.4601	0.5826	0.8548	0.5053	0.6909	0.3419	0.5721
DT-CWT-SR	5.2849	2.3971	0.5383	0.8567	0.5161	0.6533	0.3608	0.4901
CVT-SR	**5.7932**	2.3485	0.5047	0.8479	0.5208	0.4027	0.3344	0.4728
NSCT-SR	5.2909	2.3485	0.6253	0.8628	0.6030	0.6857	0.4319	0.6744
SR	5.2748	2.6713	0.6982	0.8795	0.5029	0.8830	0.4059	0.6772
LP-SR	5.1415	2.7705	0.7457	0.8827	**0.6073**	0.8416	0.4219	0.6919
ASR	5.1067	2.6482	0.6923	0.8693	0.4845	0.8833	0.3748	0.6777
CSMCA	5.0043	2.1742	0.7250	**0.8926**	0.5402	0.8616	0.4477	0.6954
Proposed	5.6935	**3.1077**	**0.7637**	0.8642	0.4288	**0.9992**	**0.9899**	**0.9963**

Note: The bold values indicates the best values.

The VIFF is mathematically expressed as

$$\mathrm{VIFF}(I_{S_1},I_{S_2},I_F)=\frac{\mathrm{VIF}(I_{S_1},I_F)+\mathrm{VIF}(I_{S_2},I_F)}{2} \tag{11}$$

where

$$\mathrm{VIF}(I_S,I_F)=\frac{\mathrm{VIND}_{l,c}(I_S,I_F)}{\mathrm{VID}_{l,c}(I_S,I_F)}$$
$$=\frac{\frac{1}{2}\log_2\left(\frac{\left|s_{l,c}^2C_U+\sigma_N^2I\right|}{\left|\sigma_N^2I\right|}\right)}{\frac{1}{2}\log_2\left(\frac{\left|g_{l,c}^2s_{l,c}^2C_U+\left(\sigma_{V_{c,l}}^2+\sigma_N^2\right)I\right|}{\left(\sigma_{V_{l,c}}^2+\sigma_N^2\right)I}\right)}$$

where c,l denotes the cth block and lth sub-band of images, respectively, $\mathrm{VIND}_{c,\ l}(I_S, I_F)$ is the visual info without distortion for the source I_S and output I_F. $\mathrm{VID}_{c,\ l}(I_S, I_F)$ is the visual info with distortion for the source I_S and output I_F, $g_{l,c}$, and $s_{l,c}$ are the scalar gain and random +ve scalar, respectively, C_U is the variance of Gaussian vector random field $U_{l,c}$, σ_N^2 is the covariance of noise N, $\sigma_{V_{l,c}}^2$ is the stationary additive zero Gaussian noise. A high value of VIFF is preferred.

4.5 | Piella's metric (Q, Q_W, Q_E)

It calculates the salient information obtainable in the output image utilizing the local measurement such as correlation coefficient, mean luminance, edge information, and contrast. The dynamic range of Piella's metric is [0 1]. The values close to 1 are considered as better fusion performance. It is expressed as

$$Q(a,b,f)=\frac{1}{|W|}\sum_{w\in W}(\lambda(w)Q_0(a,f|w)+(1-\lambda(w))Q_0(b,f|w)), \tag{12}$$

where $\lambda(w)$ is the local weight and computed as $\lambda(w)=\frac{s(a|w)}{s(a|w)+s(b|w)}$ .The $s(a|w)$ and $s(b|w)$ are the local saliency of the two input images a and b, respectively. The value of Q_0 is defined by Wang and Bovik.[43]

The weighted fusion quality index is expressed as

$$Q_W(a,b,f)=\sum_{w\in W}c(w)(\lambda(w)Q_0(a,b,f|w)+(1-\lambda(w))Q_0(b,f|w)), \tag{13}$$

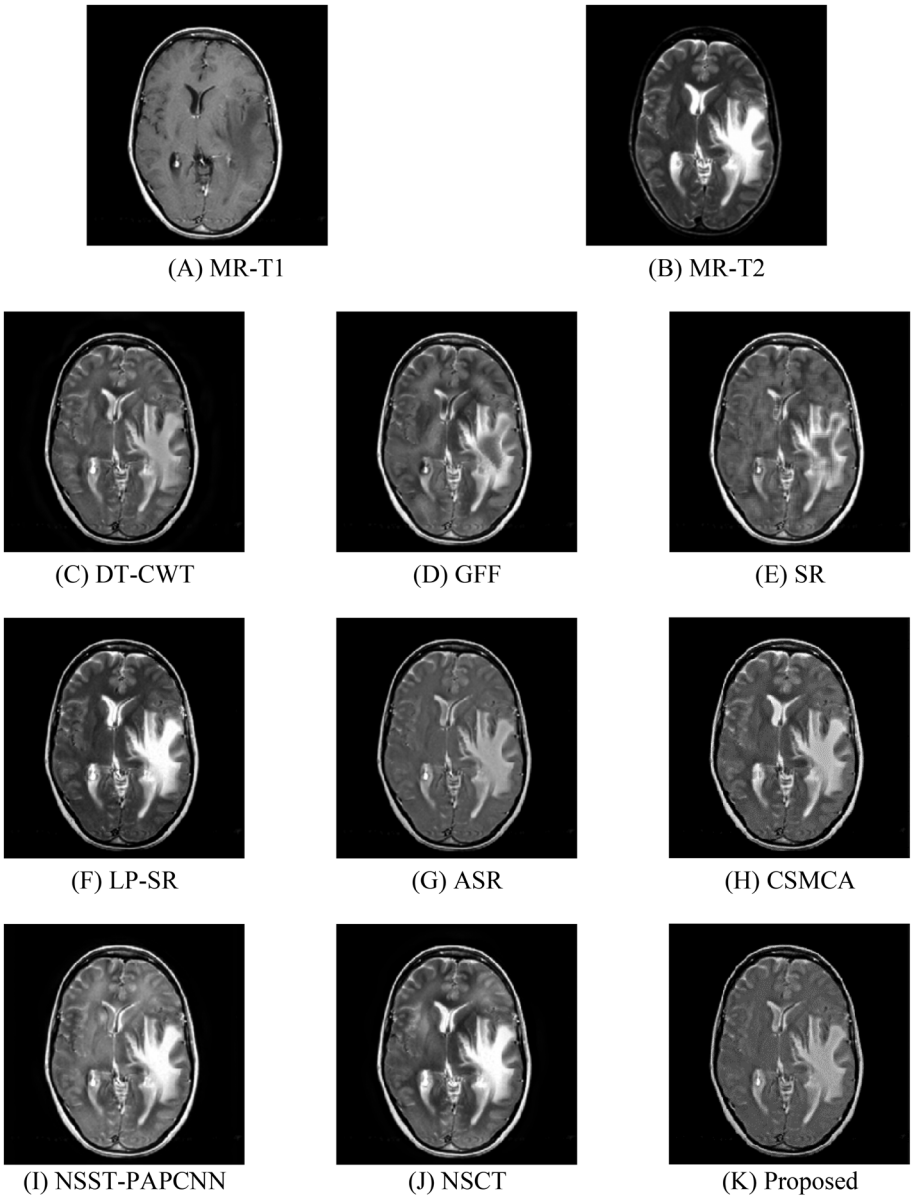

FIGURE 7 Results of fusion for group 2 images. A, MR-T1; B, MR-T2; C to K, fused images

TABLE 7 Objective evaluation performance comparison of group 2 (MR-T1 and MR-T2) images with non-SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	4.6626	2.3611	0.7247	0.9042	0.4842	0.7096	0.6228	0.7596
GFF	4.4081	2.1760	0.7211	**0.9048**	0.4961	0.8497	0.6920	0.8477
NSST-PAPCNN	5.0599	2.1357	0.7215	0.8967	**0.6859**	0.9979	0.9978	0.9997
NSCT	**5.1026**	2.3406	**0.7484**	0.9023	0.6248	0.9983	**0.9980**	0.9997
Proposed	4.8304	**2.8360**	0.6916	0.8916	0.5476	**0.9985**	0.9944	**0.9998**

Note: The bold values indicates the best values.

where $c(w) = C(w)/\left(\sum_{w' \in W} C(w')\right)$ and $C(w)$ is denoted as $C(w) = s(a|w) + s(b|w)$.

The edge-dependent fusion quality is expressed as

$$Q_E(a,b,f) = Q_W(a,b,f) \cdot Q_W(a',b',f')^{\alpha}, \tag{14}$$

where α is a factor which determines the impact of the edge images in comparison to the source images.

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	4.4736	2.4606	0.5727	0.9341	0.4304	0.6768	0.6168	0.7004
DTCWT-SR	4.0368	2.3646	0.6111	**0.9481**	0.4810	0.6614	0.7246	0.7819
CVT-SR	4.577	2.364	0.5811	0.9395	0.4531	0.5628	0.5485	0.6060
NSCT-SR	4.8770	2.4377	0.6578	0.9408	0.4408	0.6845	0.7761	0.8424
SR	4.5306	2.5573	0.7048	0.9058	0.4842	0.8351	0.6858	0.8469
LP-SR	4.3361	2.5189	**0.7622**	0.9054	0.5123	0.6356	0.6037	0.7185
ASR	**4.9560**	2.5272	0.6861	0.8981	0.4860	0.8668	0.6748	0.6777
CSMCA	4.3905	2.3719	0.7294	0.9027	0.4940	0.8722	0.7000	0.8444
Proposed	4.8304	**2.8360**	0.6916	0.8916	**0.5476**	**0.9985**	**0.9944**	**0.9998**

Note: The bold values indicates the best values.

TABLE 8 Objective evaluation performance comparison of group 4 (MR-T1 and MR-T2) images with SR-based methods

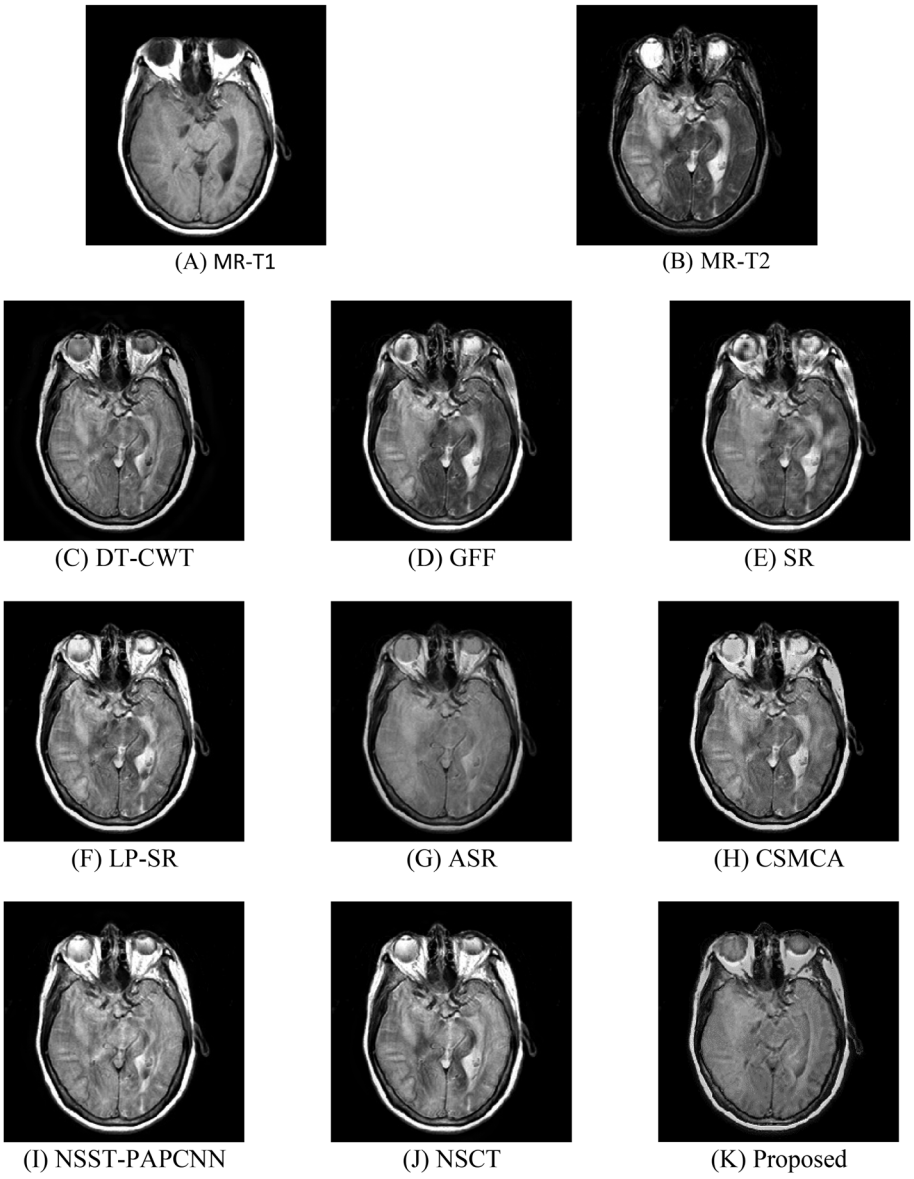

FIGURE 8 Results of fusion for group 2 images. A, MR-T1; B, MR-T2; C to K, fused images

TABLE 9 Objective evaluation performance comparison of group 2 (MR-T1 and MR-T2) images with non-SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	5.0133	2.3939	0.5116	0.7616	0.3661	0.9851	0.9255	0.9740
GFF	4.9577	2.1966	0.5345	**0.8906**	0.5006	0.9966	**0.9982**	0.9996
NSST-PAPCNN	5.0197	2.1661	0.7141	0.8709	**0.8177**	0.9974	0.9968	0.9991
NSCT	5.0364	2.5068	**0.7147**	0.8773	0.7892	0.9977	0.9968	0.9996
Proposed	**5.0691**	**3.1974**	0.5851	0.8413	0.4516	**0.9982**	0.9937	**0.9998**

Note: The bold values indicates the best values.

TABLE 10 Objective evaluation performance comparison of group 2 (MR-T1 and MR-T2) images with SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	4.7424	3.0821	0.4767	0.8525	0.4537	0.9982	0.9784	0.9981
DTCWT-SR	5.1109	3.1434	0.5578	0.8623	0.4042	0.9974	0.9720	0.9972
CVT-SR	**5.4896**	2.4273	0.4839	0.7076	0.2026	0.9814	0.9076	0.9638
NSCT-SR	5.0167	2.8395	0.5608	0.8184	0.2890	0.9929	0.9652	0.9915
SR	4.9612	2.5142	0.6040	0.7509	**0.6065**	0.9784	0.9267	0.9913
LP-SR	4.7100	3.1588	0.5613	0.8525	0.4537	0.9981	0.9784	0.9981
ASR	4.8413	2.5008	0.6047	0.7749	0.6151	0.9784	0.9653	0.9962
CSMCA	4.6821	2.3276	**0.7010**	**0.8782**	0.6044	0.8386	0.6486	0.8124
Proposed	5.0691	**3.1974**	0.5851	0.8413	0.4516	**0.9982**	**0.9937**	**0.9998**

Note: The bold values indicates the best values.

To show the potency of the suggested technique, our results are compared with DT-CWT,[16] GFF,[21] NSST-PAPCNN,[22] NSCT,[23] LP-SR,[26] DWT-SR,[26] DTCWT-SR,[26] CVT-SR,[26] NSCT-SR,[26] SR,[27] ASR[32] and CSMCA[34] methods. The quantitative results are shown in two groups. The first group compares the proposed method with the MST-based methods. The second group compares our method with the variations of SR-based methods. It is to be noted that the results shown are obtained from our implementations of the methods used for comparison. In this paper, the decomposition level is taken four in MST method. The qualitative results are shown for MST-based and SR-based methods only. For CSMCA method, we have only taken the gray images as inputs. It is observed that similar results are obtained with other SR-based methods also.

From Figure 4, it is observed that the fused image obtained in (D) and (F) have better visual quality but the local information preservation is not so good. The GFF, LP-SR, ASR, and CSMCA approaches fail to represent the original info of input images and non-natural traces are introduced in the output image. The fused images found in (I) and (J) are very similar to the fused image of the suggested method. The tissue is clearly visible in the fused image got from the proposed method. Compared to the source images, the output image found with the suggested technique retains more tissue information content. From Table 1, it is seen that the indices MI, Q, and Q_W of the suggested technique are high as compared to the other methods. Nevertheless, the rest of the values show a comparable result. Further, it is seen that the GFF method outperforms in terms of $Q^{RS/F}$. From Table 2, it is observed that the proposed method leads other methods in terms of E, $Q^{RS/F}$, Q, Q_W, Q_E, and MI. The CVT-SR method shows high value in terms of FMI. The VIFF value is high in LP-SR method. As the proposed technique uses the region-based fusion and the dictionary is learned using the improved KSVD approach, the overall performance may be better as compared to the other methods.

It is observed from Figure 5D–F that the middle part of the fused image looks darker as compared to the rest of the methods. The Figure 5F,I,J appears brighter. The bone part is visually more prominent. The soft tissue as well as the bone structure is visually sharp. Figure 5K retains both the tissue and bone structure information. It looks visually clear. From Table 3, it is seen that the proposed method gives higher MI, Q, and Q_W values. Further, the $Q^{RS/F}$ and Q_E values of the proposed technique are close to the values obtained from the other methods. In Table 4, it is seen that better E, Q, Q_E, and Q_W values are obtained using the proposed technique. Nevertheless, the other indices are close to the values found from the

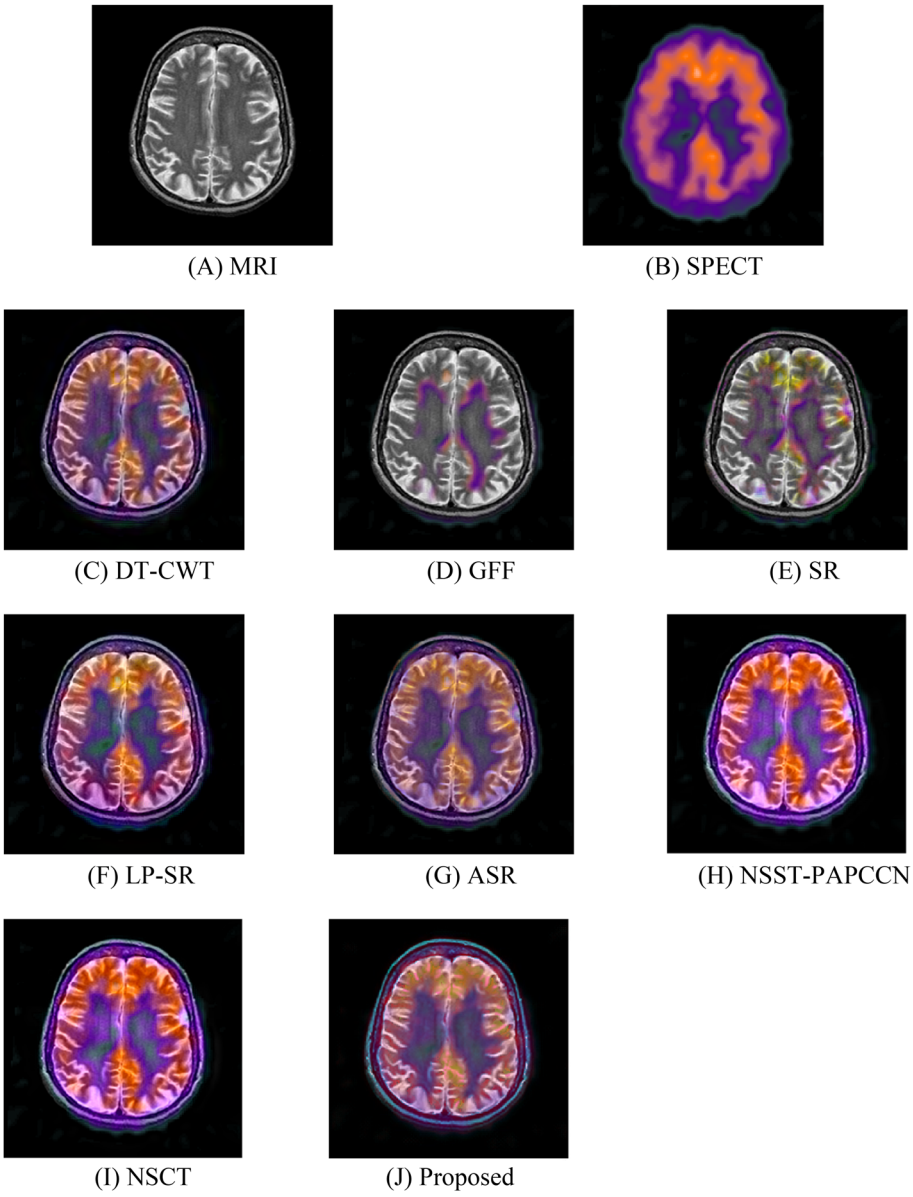

FIGURE 9 Results of fusion for group 3 images. A, MRI; B, SPECT; C to J, fused images [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 11 Objective evaluation performance comparison of group 3 (MRI and SPECT) images with non-SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	5.2707	2.3363	0.6360	0.8369	0.2651	0.9894	0.9347	0.9874
GFF	4.9763	2.1553	0.3693	0.8129	0.1284	0.9345	0.9054	0.9439
NSST-PAPCNN	5.2230	2.4750	0.7331	0.8348	0.7059	0.9971	0.9581	0.9985
NSCT	5.1396	2.7762	**0.7442**	0.8588	**0.7679**	0.9970	0.9793	0.9989
Proposed	**5.2770**	**2.9426**	0.6044	**0.8848**	0.4842	**0.9991**	**0.9967**	**0.9999**

Note: The bold values indicates the best values.

other methods. However, the LP-SR leads the MI and $Q^{RS/F}$ values. The SR method outperforms in terms of VIFF values, whereas the CSMCA method shows high value in terms of FMI.

Figure 6F,I,J,K seems brighter than other methods. The top portion of the output image of the suggested technique is visibly clear as compared to other methods. Figure 6C looks darker than the rest of the images. The

TABLE 12 Objective evaluation performance comparison of group 3 (MRI and SPECT) images with SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	5.0382	2.7864	0.5438	0.8291	0.2690	0.9895	0.9336	0.9856
DTCWT-SR	5.3117	2.7122	0.4914	0.8347	0.1990	0.9870	0.9195	0.9804
CVT-SR	**5.3604**	2.4003	0.3393	0.7485	0.1434	0.9774	0.8771	0.9642
NSCT-SR	5.2372	2.8186	0.4853	0.8527	0.2486	0.9887	0.9260	0.9822
SR	4.6211	2.6354	0.5348	0.8411	0.3291	0.9724	0.9226	0.9930
LP-SR	4.9768	2.3515	0.5829	0.8840	0.3948	0.9933	0.9428	0.9901
ASR	4.8446	2.4291	**0.7765**	0.8537	0.3253	0.9743	0.9610	0.9977
Proposed	5.2770	**2.9426**	0.6044	**0.8848**	**0.4842**	**0.9991**	**0.9967**	**0.9999**

Note: The bold values indicates the best values.

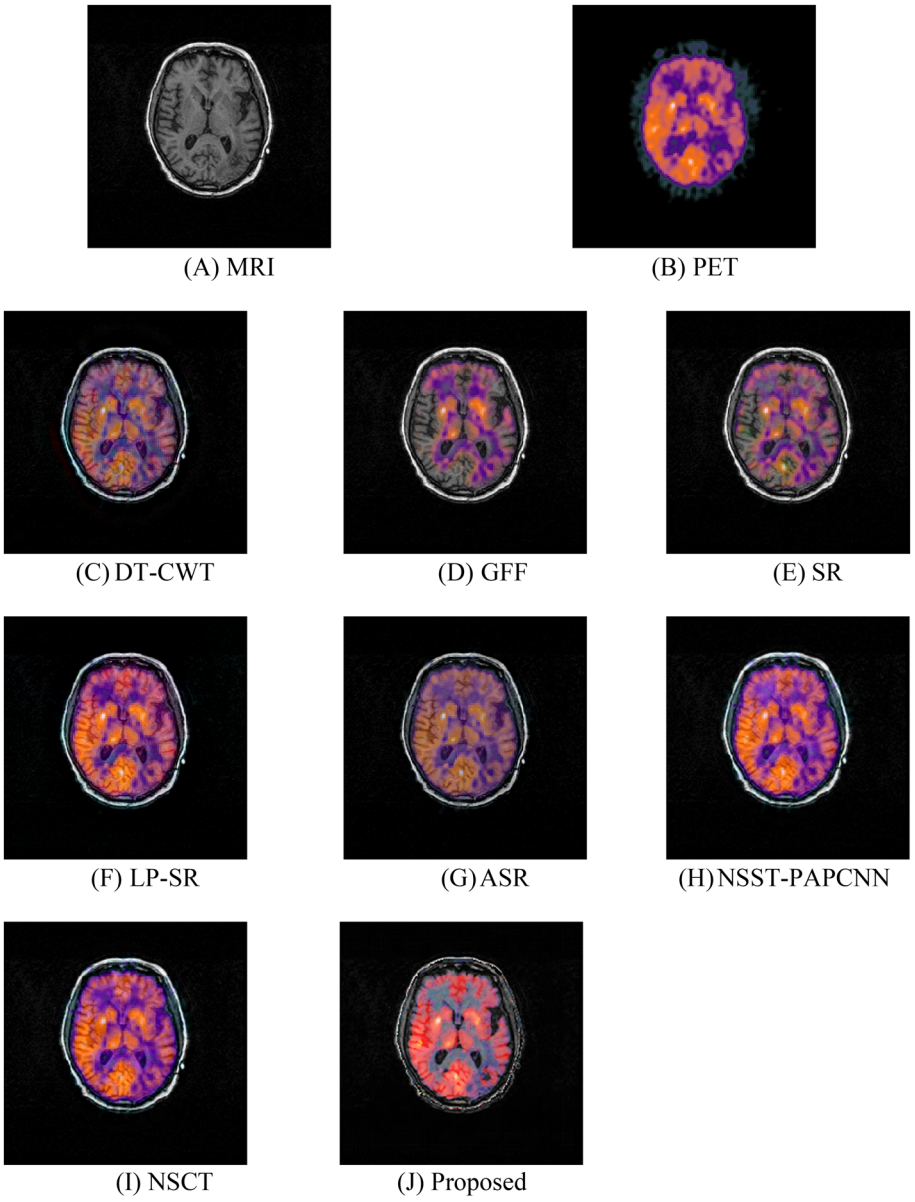

FIGURE 10 Results of fusion for group 4 images. A, MRI; B, PET; C to J, fused images [Color figure can be viewed at wileyonlinelibrary.com]

quantitative comparison in Table 5 shows that the E, MI, $Q^{RS/F}$, Q, and Q_W values of the proposed method outperform other methods. Nevertheless, the Q_E value of the suggested scheme is close to the values obtained with the NSST-PAPCNN and NSCT methods. The proposed method also shows a comparable value in terms of FMI.

TABLE 13 Objective evaluation performance comparison of group 4 (MRI and PET) images with non-SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DT-CWT	4.6463	2.3458	0.6811	0.6904	0.2993	0.9893	0.9326	0.9660
GFF	5.1088	2.0497	0.4662	0.8187	0.2106	0.9386	0.8632	0.9306
NSST-PAPCNN	5.2808	**2.5771**	0.6072	**0.9455**	0.6816	0.9989	0.9953	**0.9999**
NSCT	5.1934	2.3622	0.6401	0.8793	**0.7063**	0.9844	**0.9996**	0.9994
Proposed	**5.3624**	2.4596	**0.6904**	0.8502	0.4834	**0.9990**	0.9926	0.9684

Note: The bold values indicates the best values.

TABLE 14 Objective evaluation performance comparison of group 4 (MRI and PET) images with SR-based methods

Method	E	MI	$Q^{RS/F}$	FMI	VIFF	Q	Q_E	Q_W
DWT-SR	4.9892	2.7921	0.7124	0.6556	0.2434	0.9963	0.9772	0.9962
DTCWT-SR	4.8515	2.4964	0.5879	0.6357	0.2377	0.9911	0.9543	0.9898
CVT-SR	4.8417	2.2546	0.5328	0.5567	0.2590	0.9854	0.9282	0.9828
NSCT-SR	5.1234	**3.0375**	0.7300	0.7517	0.2517	0.9980	0.9824	0.9980
SR	5.2044	2.7123	07850	0.7234	0.3287	0.9254	0.7314	0.9257
LP-SR	4.3255	2.3636	**0.8051**	0.6016	0.2930	0.9789	0.9193	0.9728
ASR	4.5729	2.5643	0.7373	0.6062	0.3062	0.9209	0.7473	0.9211
Proposed	**5.3624**	2.4596	0.6904	**0.8502**	**0.4834**	**0.9990**	**0.9926**	**0.9984**

Note: The bold values indicates the best values.

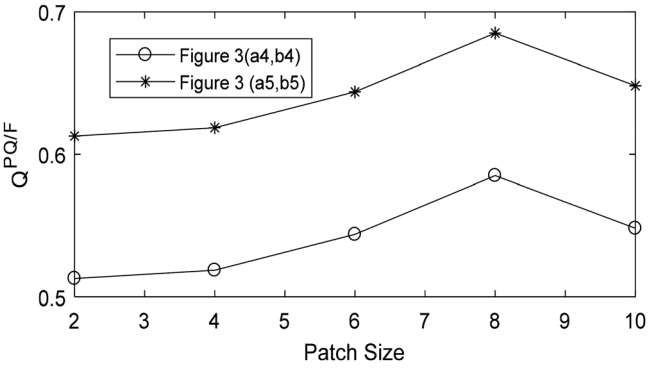

FIGURE 11 Test results of the suggested technique with different patch size

From Table 6, it is seen the suggested technique shows best results in terms of MI, $Q^{RS/F}$, Q, Q_W, and Q_E values. The CVT-SR method shows the best value in terms of E and the CSMCA method outperforms in terms of FMI. The MST-based method shows good results in terms of some indices. The reason may be the sparse technique is applied to the low-pass band where the patches are suitably represented by the OMP method.

It is seen from Figure 7D,E that the soft tissue is not visible properly, that is, some dark spots are present. The fused image in Figure 7E,J has good contrast and the tissue portion looks brighter. The fused image of the proposed technique shows a comparable result with the other methods. The lower part of the fused image is visibly sharp as compared to the other methods. Figure 7G,J,K are almost similar. From Table 7, it is observed that, the proposed technique gives the best results in terms of MI, Q, and Q_W. The NSCT method leads in terms of E, $Q^{RS/F}$, and Q_E. The GFF shows high FMI value and the NSCT-PAPCNN lead in terms of VIFF. From Table 8, it is seen that the proposed technique gives best results in terms of MI, VIFF, Q, Q_E, and Q_W. The LP-SR method leads in terms of $Q^{RS/F}$. It is seen that the E value obtained with ASR technique is the best, whereas the DTCWT-SR shows the best value in terms of FMI.

From Figure 8D,E, it is seen that the brightness is not so good as compared to other fused images. The upper part looks blurry. Figure 8F,H,I,J looks brighter. The tissue and bone structure are visually good. The proposed technique shows detail preservation of the original images. Figure 8C,G retains most of the visual information. It is observed from Table 9 that the metrics E, MI, Q, and Q_W of the proposed method gives the best result as compared to other methods. The GFF method outperforms in terms of FMI and Q_E, whereas the NSCT-PAPCNN shows high value in terms of VIFF. The reason may be the ability of the filter in preserving the edges. The FMI and Q_E value of the proposed technique produces comparable results with other methods. From Table 10, it is observed that the proposed technique outperforms in terms of MI, Q, Q_E, and Q_W. The CSMCA

method shows high value in terms of $Q^{RS/F}$ and FMI. The E value of CVT-SR method is high whereas the SR outperforms in terms of VIFF.

From Figure 9, it is seen that the soft tissue info is not well defined in (D). The remaining images are almost appearing the same. The spatial information is not clear in Figure 9E,G. The brightness near the edges is not sharp in Figure 9C. In contrast, the output image of the suggested technique retains both the spatial and spectral information and visually looks more prominent than the other fused images. From Table 11, it is observed that the quantitative indices E, MI, FMI, Q, Q_E, and Q_W for the suggested technique are superior as compared to the other methods. It is observed from Table 12 that the proposed technique shows better performance in terms of MI, FMI, VIFF, Q, Q_E, and Q_W. The value of the rest of the metrics of the proposed method is very close to the other methods. The ASR method leads in terms of $Q^{RS/F}$, whereas the CVT-SR shows high E value.

Figure 10E,G shows clear tissue information. The soft tissue in Figure 10D,E is appearing dark. All the fused images have good color content but the structural info is absent in the fused images. The suggested approach shows good visual appearance and it preserves more color information of the soft tissue. From Table 13, it is observed that the metrics E,$Q^{RS/F}$and Q of the proposed method show the best results. The NSCT-PAPCNN outperforms in terms of MI, FMI, and Q_W. In Table 14, it is seen that the indices E, FMI, VIFF, Q, Q_E, and Q_W show the best result using the suggested approach. The MI of the proposed method produces comparable value with the other methods. The LP-SR shows high $Q^{RS/F}$ value. As the proposed technique uses the adaptive dictionary for the sparse representation and the region-based fusion rule, it exhibits good fusion results.

The patch size used for the suggested method is investigated and the test results are shown in Figure 11. The effect of different patch size on $Q^{PQ/F}$ is shown. It is observed that the performance of the proposed technique improves as the patch size increases. However, it decreases for patch size of 10. Hence, we conclude to set the patch size to 8×8.

5 | CONCLUSION

A multimodal IF technique is suggested using the improved dictionary learning and the region-based fusion. An adaptive dictionary is constructed with the selected informative patches only that increase the compactness together with the computational efficiency. The proposed technique does not require the prior information of the training image patches, which is desirable. The experimental results reveal that the proposed procedure successfully integrates the complementary info from the input images to generate the output. The output images successfully preserve the texture, edge, and color information of the input. Both the qualitative and quantitative outcomes show that our technique competes with state-of-the-art IF techniques. This work may set the path for further research in the area of IF. In future, the experiments may be carried out using different medical images such as X-ray and ultrasound.

ORCID

Rutuparna Panda https://orcid.org/0000-0002-8676-0144

REFERENCES

1. Srivastava R, Khare A. Multifocus noisy image fusion using contourlet transform. *Imag Sci J*. 2015;63(7):408-422. https://doi.org/10.1179/1743131X15Y.0000000025.
2. Li S, Kang X, Fang L, Hu J, Yin H. Pixel-level image fusion: a survey of the state of the art. *Inf Fusion*. 2017;33:100-112.
3. Singh S, Anand RS. Multimodal neurological image fusion based on adaptive biological inspired neural model in nonsubsampled Shearlet domain. *Int J Imag Syst Technol*. 2019;29(1):50-64.
4. Kong W, Miao Q, Lei Y. Multimodal sensor medical image fusion based on local difference in non-subsampled domain. *IEEE Trans Instrum Meas*. 2018;68(4):938-951. https://doi.org/10.1109/TIM.2018.2865046.
5. Liu S, Zhao J, Shi M. Medical image fusion based on improved sum-modified-Laplacian. *Int J Imag Syst Technol*. 2015;25(3):206-212.
6. Bavirisetti DP, Kollu V, Gang X, Dhuli R. Fusion of MRI and CT images using guided image filter and image statistics. *Int J Imag Syst Technol*. 2017;27(3):227-237.
7. Ardeshir Goshtasby A, Nikolov S. Image fusion: advances in the state of the art. *Inf Fusion*. 2007;8:114-118.
8. Prakash O, Park CM, Khare A, Jeon M, Gwak J. Multiscale fusion of multimodal medical images using lifting scheme based biorthogonal wavelet transform. *Optik*. 2019;182:995-1014. https://doi.org/10.1016/j.ijleo.2018.12.028.
9. Liu Z, Yin H, Chai Y, Yang SX. A novel approach for multimodal medical image fusion. *Expert Systems with Applications*. 2014;41(16):7425-7435. https://doi.org/10.1016/j.eswa.2014.05.043.
10. Meher B, Agrawal S, Panda R, Abraham A. A survey on region based image fusion methods. *Inf Fusion*. 2019;48:119-132.
11. Stathaki T. *Image Fusion: Algorithms and Applications*. London, England: Academic Press; 2011.
12. Li S, Yang B. Multifocus image fusion using region segmentation and spatial frequency. *Image Vis Comput*. 2008;26(7):971-979.
13. Ma K, Li H, Yong H, Wang Z, Meng D, Zhang L. Robust multi-exposure image fusion: a structural patch decomposition approach. *IEEE Trans Image Process*. 2017;26(5):2519-2532. https://doi.org/10.1109/TIP.2017.2671921.
14. Piella G. A general framework for multiresolution image fusion: from pixels to regions. *Inf Fusion*. 2003;4(4):259-280.

15. Wang W, Chang F. A multi-focus image fusion method based on Laplacian pyramid. *J Comput*. 2011;6(12):2559-2566. https://doi.org/10.4304/jcp.6.12.2559-2566.
16. Lewis JJ, O'Callaghan RJ, Nikolov SG, Bull DR, Canagarajah N. Pixel- and region-based image fusion with complex wavelets. *Inf Fusion*. 2007;8(2 spec. iss):119-130.
17. Yang S, Wang M, Jiao L, Wu R, Wang Z. Image fusion based on a new contourlet packet. *Inf Fusion*. 2010;11(2):78-84. https://doi.org/10.1016/j.inffus.2009.05.001.
18. Zhang Q, Guo B l. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Process*. 2009;89(7):1334-1346. https://doi.org/10.1016/j.sigpro.2009.01.012.
19. Nejati, M., Samavi, S., Shirani, S. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*. 2015;25:72-84.
20. Bhatnagar G, Wu QMJ, Liu Z. Directive contrast based multimodal medical image fusion in NSCT domain. *IEEE Trans Multimedia*. 2013;15(5):1014-1024. https://doi.org/10.1109/TMM.2013.2244870.
21. Li S, Kang X, Hu J. Image fusion with guided filtering. *IEEE Trans Image Process*. 2013;22(7):2864-2875. https://doi.org/10.1109/TIP.2013.2244222.
22. Yin M, Liu X, Liu Y, Chen X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Trans Instrum Meas*. 2018;68(1):49-64.
23. Zhu Z, Zheng M, Qi G, Wang D, Xiang Y. A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain. *IEEE Access*. 2019;7:20811-20824.
24. Zhu Z, Yin H, Chai Y, Li Y, Qi G. A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inf Sci (Ny)*. 2018;432:516-529. https://doi.org/10.1016/j.ins.2017.09.010.
25. Aishwarya N, Bennila Thangammal C. A novel multimodal medical image fusion using sparse representation and modified spatial frequency. *Int J Imag Syst Technol*. 2018;28(3):175-185. https://doi.org/10.1002/ima.22268.
26. Liu Y, Liu S, Wang Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf Fusion*. 2015;24:147-164. https://doi.org/10.1016/j.inffus.2014.09.004.
27. Yang B, Li S. Multifocus image fusion and restoration with sparse representation. *IEEE Trans Instrum Meas*. 2009;59(4):884-892.
28. Zong JJ, Qiu TS. Medical image fusion based on sparse representation of classified image patches. *Biomed Signal Process Control*. 2017;34:195-205.
29. Yu N, Qiu T, Bi F, Wang A. Image features extraction and fusion based on joint sparse representation. *IEEE J Sel Top Signal Process*. 2011;5(5):1074-1082.
30. Yin H. Multimodal image fusion with joint sparsity model. *Opt Eng*. 2011;50(6):67007. https://doi.org/10.1117/1.3584840.
31. Yang B, Li S. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Inf Fusion*. 2012;13(1):10-19. https://doi.org/10.1016/j.inffus.2010.04.001.
32. Liu Y, Chen X, Ward RK, Wang ZJ. Medical image fusion via convolutional Sparsity based morphological component analysis. *IEEE Signal Process Lett*. 2019;26(3):485-489.
33. Liu Y, Wang Z. Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Process*. 2014;9(5):347-357. https://doi.org/10.1049/iet-ipr.2014.0311.
34. Kim M, Han DK, Ko H. Joint patch clustering-based dictionary learning for multimodal image fusion. *Inf Fusion*. 2016;27:198-214. https://doi.org/10.1016/j.inffus.2015.03.003.
35. Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. *Adv Neural Inf Process Syst*; 2007:801-808.
36. Tropp JA, Gilbert AC. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans Inf Theory*. 2007;53(12):4655-4666. https://doi.org/10.1109/TIT.2007.909108.
37. Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process*. 2006;54(11):4311-4322. https://doi.org/10.1109/TSP.2006.881199.
38. www.med.harvard.edu/aanlib/home, accessed in August 2019.
39. Petrović, V., and Xydeas, C. (2005). Objective image fusion performance characterisation. *Proc IEEE Int Conf Comput Vis*, Ii (January), 1866–1871.
40. Piella, G., & Heijmans, H. (2003). A new quality metric for image fusion, In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, Vol. 3, Pp. III-173. IEEE, 2003. https://doi.org/10.1109/icip.2003.1247209
41. Qu G, Zhang D, Yan P. Information measure for performance of image fusion. *Electron Lett*. 2002;38(7):313-315. https://doi.org/10.1049/el:20020212.
42. Haghighat MBA, Aghagolzadeh A, Seyedarabi H. A non-reference image fusion metric based on mutual information of image features. *Comp Elect Eng*. 2011;37(5):744-756. https://doi.org/10.1016/j.compeleceng.2011.07.012.
43. Han Y, Cai Y, Cao Y, Xu X. A new image fusion performance metric based on visual information fidelity. *Inf Fusion*. 2013;14(2):127-135.
44. Wang Z, Bovik AC. A universal image quality index. *IEEE Signal Process Lett*. 2002;9(3):81-84.

How to cite this article: Meher B, Agrawal S, Panda R, Dora L, Abraham A. A novel region-based multimodal image fusion technique using improved dictionary learning. *Int J Imaging Syst Technol*. 2020;30:558–576. https://doi.org/10.1002/ima.22395