BME Horizon

Science Exploration

Open Access Research Article

A multi-modal medical image fusion framework based on two-scale decomposition and saliency detection

Harmanpreet Kaur[1,*], Renu Vig[2], Naresh Kumar[1]

[1]Department of Electronics and Communication Engineering, University Institute of Engineering and Technology, Panjab University, Chandigarh 160014, India.
[2]Vice Chancellor, Panjab University, Chandigarh 160014, India.

***Correspondence to:** Harmanpreet Kaur, Department of Electronics and Communication Engineering, University Institute of Engineering and Technology, Panjab University, Chandigarh 160014, India. E-mail: harman.phdpub@gmail.com

Received: July 05, 2024 **Accepted:** August 14, 2024 **Published:** November 05, 2024

Cite this article: Kaur H, Vig R, Kumar N. A multi-modal medical image fusion framework based on two-scale decomposition and saliency detection. BME Horiz. 2024;2:122. https://doi.org/10.70401/bmeh.2024.122

Abstract

Due to imaging sensors' constraints, it is difficult to acquire a medical image that subsequently incorporates functional metabolic data and anatomical tissue characteristics. An important tool to aid in diagnostic procedures and intra-operative management is multimodal medical image fusion, an efficient method to combine the supplemental information gathered from numerous modalities. This research proposes a unique fusion approach for multimodal medical images that makes use of mean filter and maximum symmetric surround saliency detection. First, the images to be fused are decomposed into base layer and a series of detail layers via mean filter. While the guided filtering based fusion rule is used to merge the detail layers, for the fusion of base layer, a linear summation is employed. The fused base and detail images are subsequently combined to construct the final fused image. The proposed medical image fusion method performed more effectively than existing cutting-edge approaches in the context of both visual appearance and quantitative assessment, according to careful evaluation on a widely used database. The proposed method overcomes the challenge of cutting-edge image fusion methods in maintaining substantial edges and energy information on multi-modal medical images.

Keywords: Multi-modal image fusion, saliency detection, pixel-level, multi-scale decomposition, image reconstruction

1. Introduction

In image-based application fields like defense surveillance, remote sensing, medical imaging, and computer vision, image fusion is widely acknowledged as a useful tool for enhancing overall system performance. Medical images have become increasingly crucial in clinical diagnosis because of advancements in imaging technology. The benefits and limits of various medical imaging modalities vary. For instance, images from computed tomography (CT) reveal detailed structural information about bones and implants. However, diagnosis of soft tissue tumours is challenging. Although magnetic resonance imaging (MRI) provides detailed structural information on organs, it is unable to accurately depict calcification or lesions in bone tissues or identify human metabolic processes. While MR-T1yields improved anatomical specifics, the contrast between diseased and normal tissues is strengthened by MR-T2. Whilst Single Photon Emission Computed Tomography (SPECT) imaging uses nuclear imaging techniques to analyze blood flow in tissues and organs, Positron Emission Tomography (PET) images offer excellent contrast for precisely mapping tumours. However, the resolution of PET and SPECT pictures is poor. Each imaging modality inherently has its unique traits and built-in limitations because of the corresponding different process. As a result, medical images captured using a single modality cannot completely and accurately depict the information contained in medical tissues, which causes an intricate diagnosis procedure and erroneous diagnoses. Over the past few years, this encouraged researchers to combine this supplementary data into a single fused image. In contrast to individual source medical images, the fused image is more accurate and detailed. The goal of medical image fusion is to increase the utilization of medical images, aid physicians in comprehending the information contained in images, and facilitate more thorough, accurate, and efficient illness diagnosis and treatment. In addition to precisely implementing the localization and characterization of the lesion, it may also lower the overall expenses of patient-related database storage[1,2].

Image fusion process can be classified into four paramount classes depending upon the input data: multi-view, multi-sensor,

multi-focus, and multi-modal fusion. Further, three steps can be taken to achieve it: (a) pixel/signal level (lowest level of fusion), (b) feature/objective level (medium level of fusion) and (c) decision/symbolic level (highest level of fusion). Fusion at pixel-level aims to merge the information existing in the original input images pixel by pixel. In feature level image fusion, the conspicuous features of an image (like edges, formation, length, segments, and directions) extricated from the input source images are fused to create a better descriptive image. Decision level image fusion utilizes decision scores to bring about an ultimate fusion decision. This paper is grounded on pixel-level image fusion for multimodal medical images.

Numerous image fusion approaches have been put forth in recent years and these approaches can be broadly divided into two groups: transform domain (TD) and spatial domain (SD) approaches[3,4]. Depending upon pixel intensities, the spatial domain techniques directly operate with pixels or areas in those domains. Minimal computing complexity, excellent real-time functionality, and a superior signal-to-noise ratio of amalgamated images are benefits of SD approaches. Despite this, it could result in edge blur, reduced contrast, and decreased sharpness. In gray scale space, using different multi-scale decomposition (MSD) transforms is another widely employed method for pixel-level image fusion. Pyramid transform and wavelet transform falls under this category. In the above MSD-based fusion approaches, each source image undergoes MSD transformations first in order to produce several sub-bands that include the decomposed information of various frequencies or alignments. Following that, according to a set of fusion rules, the associated sub-bands from each source images are fused. In the final stage, inverse MSD transforms will yield the fused images. Image fusion has also seen success with the use of other, more sophisticated MSD transforms such the dual-tree complex wavelet, curvelet transform, contourlet transform, shearlet transform. Since linear filters are utilized in the decomposition stage of the pyramid transform, halo artifacts might be introduced into the fused images. The up sampling and down sampling processes employed by conventional and advanced wavelet transforms during the transform process could blur out certain contours in the fused images. For the purpose to tackle the aforementioned problems, non-sub sampled transforms namely non-subsampled contourlet transform or non-subsampled shearlet transform are added to MMIF methods since they improve the shift invariance and directional selectivity features, which are crucial for image fusion but suffer from high computing complexity.

By linearly merging redundant dictionaries and sparse coefficients, sparse representation (SR) offers a succinct and flexible representation of images. As a result, it has gained substantial attention in image fusion. Whilst SR-based approaches have a solid statistical theoretical foundation, the coefficient selection and dictionary learning model have a major impact on the final result since poor coefficient selection might lead to unsatisfactory outcomes. Additionally, the majority of SR-based approaches require a lot of time. There has been a lot of focus on several edge-preserving-filter (EPF) based techniques recently like weighted least squares filter, guided image filter (GIF), cross bilateral filter, anisotropic diffusion filter (ADF) etc. When an image is fused using EPF-based approaches, the edge information is preserved, which successfully solves the issue where edge information could be compromised via the decomposition and reconstruction with MST-based techniques. But their efficacy depends up on the development of decomposition algorithms and fusion rules. Recently, the visual saliency detection (VSD) has gained popularity in the field of image fusion[5-7].

Researchers have tried to merge SD-based, TD-based, SR-based, EPF-based and VSD-based methods in order to benefit from these approaches while simultaneously overcoming their drawbacks (including complexity, memory space needs, and implementation challenges)[4]. These techniques can improve upon the drawbacks associated with conventional image fusion approaches to deliver sufficient fusion performance. With the goal of preserving energy and important edge information in medical images and to effectively use saliency extraction's capacity for image processing, an improved medical image fusion approach based on multi-scale image decomposition via mean filter and saliency detection via maximum symmetric surround saliency is proposed to overcome these shortcomings. The remainder of this paper is structured as follows. Section 2 provides a brief explanation of the relevant theory, and Section 3 presents the proposed fusion approach. Section 4 reports experimental findings with pertinent analysis and conclusions are provided in Section 5.

2. Related Theory

2.1 Saliency detection-based methods

Visual attention refers to the brain function that decides which portion of the vast amount of sensory data is currently of most interest. Over the past few decades, psychologists, neurobiologists, and computational neuroscientists have conducted extensive research on visual attention and greatly benefited from one another. Systems for computational visual attention are designed to find Regions of Interest (ROI) in images based on theories of the human visual system. Computational saliency models convert the source image into a saliency map, where local signal strength reflects the saliency of the local image. A key factor in differentiating a model is whether it relies on bottom-up or top-down effects as presented in Table 1. Salient ROI are those that draw our attention in a bottom-up manner. For a feature to be salient, it must be sufficiently discriminative compared to those in the immediate vicinity. Attention systems with practical applications are becoming more and more sought after in computer vision and graphics, cognitive systems, and mobile robotics like image compression, video summarization, progressive image transmission, image segmentation, object recognition, content aware image scaling and image quality assessment. To explain eye movements in both basic and complicated scenes, attention models combine heuristics, cognitive, and neurological data, as well as methods from machine

learning and computer vision[8].

Table 1. Classification of factors that drive visual attention.

Bottom-Up	Top-Down
Exogenous process	Endogenous process
Stimulus driven	Goal oriented/task specific
Based on characteristics of a visual scene	Influenced by cognitive processes. (Knowledge, expectations, and existing goals)
Fast, involuntary, and feed-forward	Slow, task driven, voluntary, and closed loop

A modality will be considerably better if it draws attention to irregularities that are medically significant (salient). Reading medical images can be difficult because sometimes important imaging results are not apparent, or they may be too delicate for the human eye. Therefore, having a saliency model that is capable of accurately predicting the intrinsic saliency of target areas as exhibited by the modality would be helpful for assessing new technologies. For instance, radiologists must quickly assess a huge number of clinical images, thus they must use an effective approach to focus their visual attention. The majority of computational models for visual attention, however, were created in the context of natural environments, and their application to medical imaging has not been extensively explored. Extensive research has been conducted on saliency-based methods for their high efficiency, visual fidelity, and low computational complexity for fusion applications[9]. These techniques focus on identifying and highlighting salient objects while ignoring their surroundings, resulting in improved image fusion and the preservation of salient areas' integrity. After reviewing the literature, it is concluded that the researchers have rarely employed saliency detection techniques in medical imaging because there are no saliency models developed for medical image fusion. A brief overview of some of the conventional salient detection methods is presented in Table 2. As a result, the purpose of the current paper research work is to examine the performance of existing maximum symmetric surround saliency (MSSS) detection technique for MMIF which could enhance radiologists' training and diagnostic accuracy.

Table 2. Bottom-up saliency models.

Model	Year	Features	Domain	Description
Saliency-based visual attention model[10]	1998	Spatial, static, natural stimuli, free-viewing, space-based	Bio-inspired hierarchical approach	This fundamental approach employs three different feature channels–colour, intensity, and orientation–to generate feature maps that contain a centre-surround structure.
Attention based on information maximization[11]	2005	Spatio-temporal, static, synthetic stimuli, free-viewing, space-based	Purely computational model	For this approach, images are subjected to ICA for the identification of appropriate basis. These images are then projected onto this basis to generate the saliency maps by computing Shannon's self-information measure.
Graph-based visual saliency[12]	2007	Spatial, static, natural stimuli, free-viewing, space-based	Depends on both biological/computational models.	This method takes various aspects into consideration at different levels and creates a network that links all grid positions for each feature map.
Spectral residual approach[13]	2007	Spatio-temporal, static, synthetic stimuli, free-viewing, space-based	Purely computational model	By scrutinizing the log-spectrum of an image, this approach extracts the spectral residual without making any presumptions about the preliminary awareness of significant objects in an image.
Frequency-tuned salient region detection[14]	2009	Spatial, static, natural stimuli, free-viewing, space-based	Purely computational model	In this method, the image is treated as one entity that surrounds each pixel. By adjusting the low and high-frequency cut-offs, attention is equally drawn to the significant objects, and distinct borders are formed in the image.

Context-aware saliency detection[15]	2010	Spatial, static, natural stimuli, free-viewing, space-based	Purely computational model	This model has been adapted based on four principles of human visual attention with the aim of identifying areas of interest in images.
Maximum symmetric surround saliency[16]	2010	Spatial, static, natural stimuli, free-viewing, space-based	Purely computational model	This computational method efficiently uses colour and luminance features to create full-resolution saliency maps that effectively suppress the background. It is easy to implement and provides a reliable solution.

ICA: independent component analysis.

2.2 Maximum symmetric surround saliency

This technique[17] is built around the idea that depending on an object's placement in the image, the information about its size can be inferred. This method makes use of the surrounding regions (sub-images) that are proportionate with respect to the pixel whose saliency has to be determined. As a result, the center-surround bandwidth varies according to how far the pixel is from the border of the image. Subsequently, the value of MSSS at the specified pixel $S(x, y)$ for an input image of width w and height h is calculated as shown below:

$$S(x,y) = \|I_\mu(x,y) - I_g(x,y)\| \tag{1}$$

where $S(x, y)$ represents value of pixel saliency at location (x, y), $|.|$ is the norm value, $I_\mu(x, y)$ indicates mean of all sub-image pixel vectors, and $I_g(x, y)$ is the equivalent image pixel vector in the source image that has been filtered via a Gaussian distribution, such that

$$I_\mu(x,y) = \frac{1}{A}\sum_{m=x-x_o}^{x+x_o}\sum_{n=y-y_o}^{y+y_o} I(m,n) \tag{2}$$

where area A and offsets x_o, y_o of the sub-image is calculated as follows:

$$A = (2x_o + 1)(2y_o + 1) \tag{3}$$

$$x_o = \text{minimum}(x, w - x) \tag{4}$$

$$y_o = \text{minimum}(y, h - y) \tag{5}$$

Hence, MSSS is a saliency region detection technique that improves upon and preserves the benefits of full resolution saliency maps with clearly defined edges. Further, it takes advantage of colour and brightness properties, is easy to develop, and uses computing effectively. It is disadvantageous; however, in case the salient object is partially occluded by the image boundaries, it will be perceived as background and will be less noticeable.

2.3 Edge preserving filter

GIF[18] has been widely applied to image processing as an edge-preserving smoothing filter, which not only smoothes the input image but also preserves the edge information. It computes the output like other linear transform invariant filters but uses a second image to filter the input image for guidance purpose. Second image may be the same input image or a translated version of it or a totally different image. Assume that a source image is I, guidance image is I_g, and the guided filter (GF) output is I_o. For a pixel (i, j) centred at point l, the output is given as

$$I_o(i,j) = a_l I_g(i,j) + b_l, \forall (i,j) \in w_l \tag{6}$$

where a_l and b_l are linear constant coefficients in window w_l and w_l is a 2-D window with centre l and radius r. Here, the coefficients a_l and b_l are determined by decreasing the square variance between output image $I_o(i, j)$ and source image $I(i, j)$. The fitting function is given as

$$V(a_l, b_l) = \sum_{(i,j)\in w_l}\left\{[I_o(i,j) - I(i,j)]^2 + \delta a_l^2\right\} \tag{7}$$

where δ is a regularization parameter that controls smoothness and prevents it from being too large. The optimal solution of linear constant coefficients can be acquired by the linear regression method and should minimize the expected fitting function V(a_l, b_l). The solution is given by

$$a_l = \frac{\frac{1}{|\tilde{w}|}\sum_{(i,j)\in w_l}\left\{I_g(i,j)I(i,j) - \mu_l \bar{I}_l(i,j)\right\}}{\sigma_l^2 + \delta} \tag{8}$$

$$b_l = \bar{I}_l(i,j) + a_l\mu_l \tag{9}$$

where μ_l and σ_l^2 are mean and variance of guided image respectively with window w_l, $|\tilde{w}|$ denote pixel number in w_l and $\bar{I}_l$ is the mean pixel value of the source image to be filtered in i.e., $\bar{I}_l(i,j) = \frac{1}{w}\sum_{(i,j)\in w_l} I(i,j)$. Once a_l and b_l are obtained, the filter output $Io(i, j)$ can be solved using Eq. 6. Because pixel (i, j) is contained in all the local windows of window w_l that are centered on the pixel l, the GF output will change with varying w_l. To solve this problem, all possible averages of a_l and b_l are considered to acquire the final GF output by

$$I_o(i,j) = \bar{a}(i,j)I_g(i,j) + \bar{b}(i,j) \tag{10}$$

where $\bar{a}(i,j)$ and $\bar{b}(i,j)$ are average coefficients given by

$$\bar{a}(i,j) = \frac{1}{|\tilde{w}|}\sum_{l\in w(i,j)} a_l \tag{11}$$

$$\bar{b}(i,j) = \frac{1}{|\tilde{w}|}\sum_{l\in w(i,j)} b_l \tag{12}$$

3. Proposed Methodology

Figure 1 depicts the proposed method's overall fusion diagram, which is primarily made up of following components: image enhancement, decomposition and reconstruction, saliency detection, weight map construction and sub-images fusion. First, the source images are enhanced using adaptive histogram equalization (AHE). Then, mean filter based two-scale decomposition is utilized to decompose the enhanced images to obtain a base layer and detail layers (sub-images). The use of mean filter eliminates the up-sampling and down-sampling processes (as achieved in traditional approaches) while still achieving image decomposition. Further, base layers are combined based on conventional linear summation rule, aiming to provide good contrast and overall look for the fused image. Additionally, the saliency detection scheme and guided image filtering is employed to integrate the detail layers, making the fused image look more realistic and appropriate for human visual perception. Finally, the sub-images are integrated using linear summation fusion rule and the final fused image is reconstructed. The proposed approach not only achieves image decomposition but is relatively simple, less time-consuming, uses less memory and computational resources and inherits the benefits of traditional image fusion methods introduced before. To begin with, assume two preregistered source images of width w and height h.

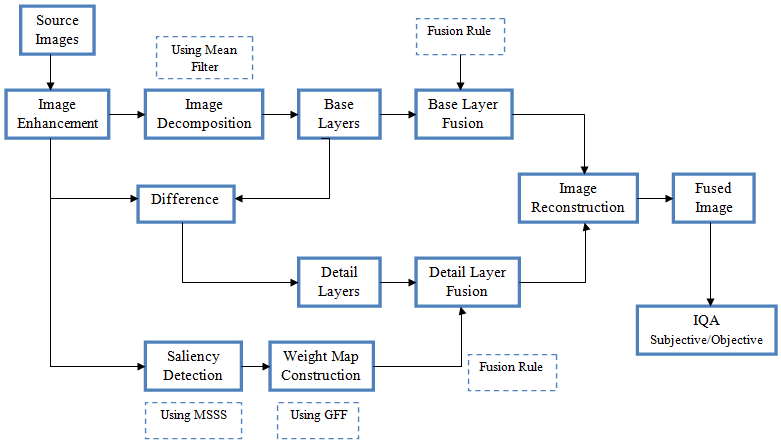

Figure 1. Proposed multi-modal medical image fusion framework. IQA: image quality assessment; GFF: guided filtering-based fusion; MSSS: maximum symmetric surround saliency.

3.1 Image enhancement

Medical images carry the pertinent information that doctors need to diagnose for planning the suitable treatment. The diagnostic process of medical images requires strong ability to visualize the minor abnormality and accepts no possibility for error in diagnosing the disease as it will strongly affect the life of the patient. Therefore, image enhancement is done to recover the lost contrast and improve the visual quality of the image, thereby helping clinicians for making accurate decisions. Hence, it is the technique of modifying an image appearance to improve the visibility of interesting features while maintaining the intrinsic information content same[18]. In this work, AHE has been employed to sharpen the definition of edges in each area of an input image and raise local contrast at an 8-tile depth and a clip limit of 0.004 working at full dynamic range. The original source images $I_1(x, y)$ and $I_2(x, y)$ are shown in Figure 2a,b, Figure 3a,b, and Figure 4a,b and the corresponding enhanced images $I'_1(x, y)$ and $I'_2(x, y)$ are shown in Figure 2c,d, Figure 3c,d, and Figure 4c,d.

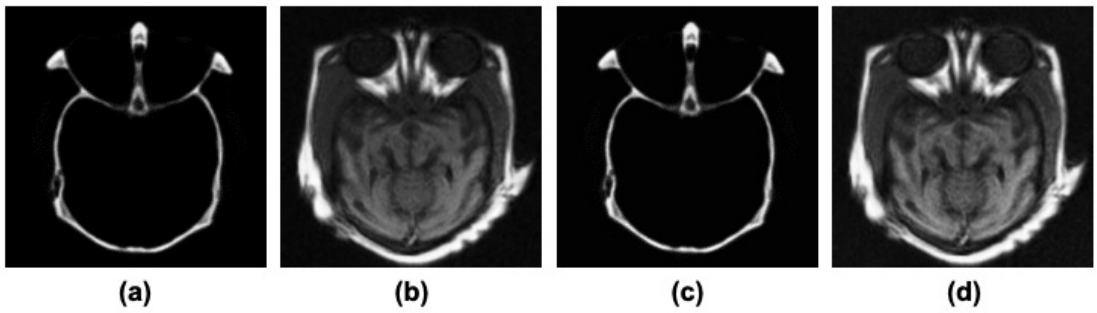

Figure 2. Multimodal source images for Dataset 1 (a) CT; (b) MRI; (c) Enhanced image for CT; (d) Enhanced image for MRI. MRI: magnetic resonance imaging; CT: computed tomography.

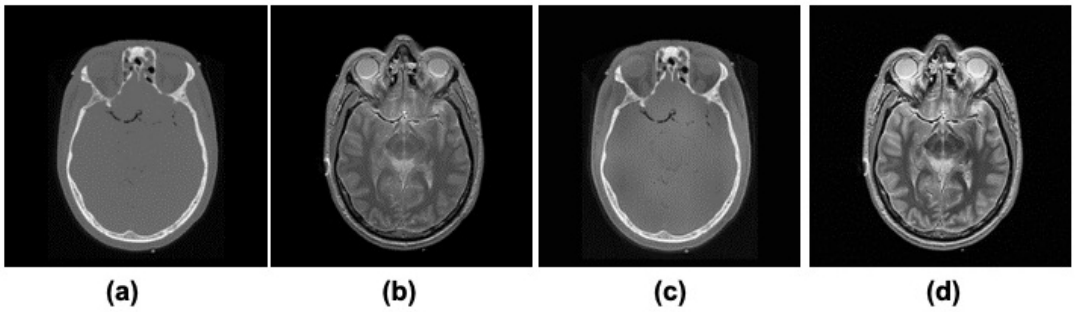

Figure 3. Multimodal source images for Dataset 2 (a) CT; (b) MRI; (c) Enhanced image for CT; (d) Enhanced image for MRI. MRI: magnetic resonance imaging; CT: computed tomography.

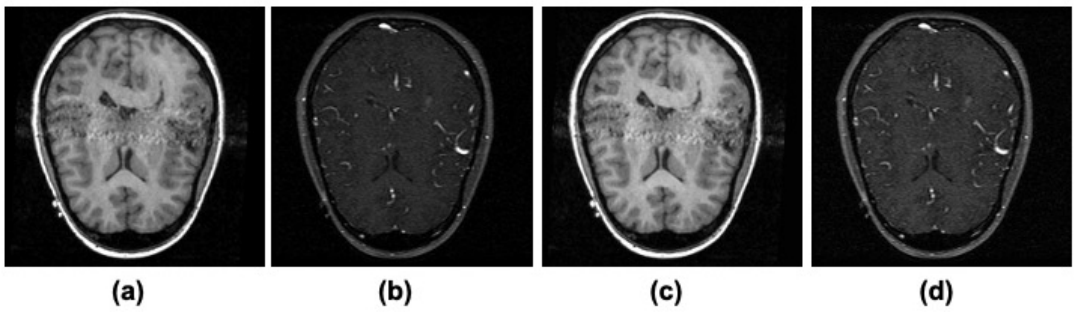

Figure 4. Multimodal source images for Dataset 3 (a) MR-TI; (b) MR-T2; (c) Enhanced image for MR-T1; (d) Enhanced image for MR-T2.

3.2 Image decomposition

In contrast to the conventional IF methods, in order to more effectively utilize the valuable information contained in the original image, the source image is decomposed at level 2 into sub-images: base layer and detail layers using mean filter, thereby eliminating the procedures of up-sampling and down-sampling. The low-frequency information, or usually the comprehensive information of the image, is contained in the base layer i.e., the base image is a smoothed rendition of the original image, which approximates the

original image and comprises the primary information. The base layers for two enhanced source images $I'_1(x, y)$ and $I'_2(x, y)$ are obtained as

$$B_1(x,y) = I'_1(x,y) * C(x,y) \tag{13}$$

$$B_2(x,y) = I'_2(x,y) * C(x,y) \tag{14}$$

where x, y defines pixel location, $C(x, y)$ is the mean filter of square window size w and * indicates the convolution operation. The window size of 3 for decomposition level 1 and 6 for decomposition level 2 is considered in this work. On the other hand, the high frequency information of the image, which often includes texture details, are a boundary and so on of the image, is mostly contained in the detail layer. Two sets of detail layers ($D_{11}(x, y)$, $D_{12}(x, y)$ and $D_{21}(x, y)$, $D_{22}(x, y)$) originate when the base layers obtained in Eq. 1 and Eq. 2 are subtracted from the original enhanced source images $I'_1(x, y)$ and $I'_2(x, y)$ respectively.

3.3 Saliency detection and weight map construction

To maintain the saliency regions of the Human Visual System (HVS) in the fused image, the saliency map $S(x, y)$ of a gray scale image $I(x, y)$ is obtained using MSSS detection introduced in Section 2.2. The MSSS values $S_1(x, y)$, $S_2(x, y)$ for the source images $I_1(x, y)$, $I_2(x, y)$ respectively are expressed as

$$S_1(x,y) = \|I_{\mu 1}(x,y) - I_f(x,y)\| \tag{15}$$

$$S_2(x,y) = \|I_{\mu 2}(x,y) - I_f(x,y)\| \tag{16}$$

Where $\|.\|$ is the norm value, $I_{\mu 1}(x, y)$, $I_{\mu 2}(x, y)$ are the mean values of source images pixel vectors and $I_f(x, y)$ is the Gaussian filtered pixel vector of the source images. The change emphasizes salient details like lines and edges that are more important than their surrounding areas. The saliency maps obtained above provide the degree of saliency of detailed information. Then, corresponding weight maps are produced by comparing the saliency maps' pixel maxima as shown:

$$P_1(x,y) = \begin{cases} 1, S_1(x,y) \geq S_2(x,y) \\ 0, \qquad \text{otherwise} \end{cases} \tag{17}$$

$$P_2(x,y) = \begin{cases} 1, S_2(x,y) \geq S_1(x,y) \\ 0, \qquad \text{otherwise} \end{cases} \tag{18}$$

The weight map calculates the magnitude of focus of every single pixel in input images that is related to the structure of the image's objects. The object boundaries in the source images are not perfectly matched with the computed saliency weight maps thereby producing artefacts in the fused image. As optimization-based methods are relatively inefficient for the generation of weight maps, in this paper, GIF is chosen as a weight map construction method because of its great computational effectiveness and excellent edge preservation capabilities. Also, the computational time of GIF is autonomous to the filter size[17]. The resulting weight maps $W_1(x, y)$, $W_2(x, y)$ of the detail layers are defined as

$$W_1(x,y) = \text{GIF}\left(P_1(x,y), I'_1(x,y)\right) \tag{19}$$

$$W_2(x,y) = \text{GIF}\left(P_2(x,y), I'_2(x,y)\right) \tag{20}$$

3.4 Fusion rules (for base layer and detail layer fusion)

With the objective to retrieve and transmit as much feature and detail information as possible into the amalgamated image, the proposed fusion approach adopts different rules for base images and detail images. The base images $B_1(x, y)$, $B_2(x, y)$ are added to obtain the resulting base image $B_F(x, y)$.

$$B_F(x,y) = B_1(x,y) + B_2(x,y) \tag{21}$$

For detail images fusion, firstly, maximum fusion rule is employed to obtain the detail layers $D_1(x, y)$ and $D_2(x, y)$, represented as

$$D_1(x,y) = \max\left(D_{11}(x,y), D_{12}(x,y)\right) \tag{22}$$

$$D_2(x,y) = \max\left(D_{21}(x,y), D_{22}(x,y)\right) \tag{23}$$

Saliency weight maps generated in Eq. 19 and Eq. 20 are then used to combine the important information from the detail layers into an amalgamated image $D_F(x, y)$. The resultant detail layer is given as

$$D_F(x,y) = W_1(x,y)D_1(x,y) + W_2(x,y)D_2(x,y) \tag{24}$$

3.5 Image reconstruction and fused image

Then, the resultant fused image $I_F(x, y)$ is reconstructed via combining/super-imposing the fused base layer $B_F(x, y)$ and the fused detail layer $D_F(x, y)$. The linear summation of the two layers is done as follows:

$$I_F(x,y) = B_F(x,y) + D_F(x,y) \tag{25}$$

3.6 Image quality assessment

Lastly, the quality of the fused image is evaluated both subjectively (qualitative) and objectively (quantitative) in order to verify the effectiveness of the proposed approach.

The implementation process of the proposed pixel level MMIF utilizing two-scale image decomposition via saliency detection is summarized in Table 3.

Table 3. Algorithm for proposed MMIF framework using hybrid approach.

ALGORITHM: THE PROPOSED MMIF FRAMEWORK	
Input	*Pre-registered multi-modal medical images $I_1(x,y)$ and $I_2(x,y)$*
Output	*Fused image $I_F(x,y)$*
Steps	
1	*Image enhancement using AHE to get $I'_1(x,y)$* and *$I'_2(x,y)$*.
2	*Mean filter-based image decomposition to obtain sub-images: base layers $B_1(x,y)$ and $B_2(x,y)$ and two sets of detail layers {$D_{11}(x,y)$, $D_{12}(x,y)$}* and {$D_{21}(x,y)$, $D_{22}(x,y)$}.
3	*Obtain saliency maps $S_1(x,y)$* and *$S_2(x,y)$ using MSSS detection.*
4	*Produce weight maps $P_1(x,y)$* and *$P_2(x,y)$ by comparing the saliency maps' pixel maxima.*
5	*Construct optimized weight maps $W_1(x,y)$* and *$W_2(x,y)$ via guided image filtering.*
6	*Generate fused images $B_F(x,y)$ utilizing addition fusion rule for base images.*
7	*Generate fused images $D_F(x,y)$ utilizing weighted average fusion rule for detail images.*
8	*Image reconstruction using linear summation rule to obtain final fused image $I_F(x,y)$.*
9	*Image Quality Assessment: Subjective and Objective.*

MMIF: multimodal medical image fusion; AHE: adaptive histogram equalization; MSSS: maximum symmetric surround saliency.

4. Experimental Set-up

The efficiency of the proposed approach was examined using a public dataset made up of 3 sets of medical images. The first two sets comprise of CT/MRI and the third set is of MR-T1/MR-T2 images. With a spatial resolution of 256*256, all clinical images come from the human brain. The Harvard Medical School database is used to get publicly available datasets[19]. The proposed approach is assessed and contrasted with six cutting-edge fusion methods to demonstrate its positive aspects on a publicly available Harvard dataset. Moreover, the parameter settings remain unchanged from the time they were published.

• Discrete Cosine Transform (DCT)[20].

• Structure-aware image fusion (SAIF)[21].

• Fourth-order partial differential equations (FPDE)[22].

• Synchronized-anisotropic diffusion model (S-ADE)[23].

• Anisotropic diffusion and Karhunen-Loeve transform (ADF)[24].

• Weighted-least-square-based scheme (VSMWLS)[25].

• Weight-optimised anisotropic diffusion filtering (WOADF)[26].

4.1 Image quality assessment (IQA) metrics

For analysing the performances of the fusion approaches, objective image quality assessment metrics are used in addition to the subjective evaluation. The goal of IQA metrics is to determine whether the merged image's attributes are consistent with human vision. Typically, there are three types of objective image quality metrics: full-reference, reduced-reference, and no-reference[27]. Following metrics are used for the experimentation purpose.

• Average gradient (AG)

It evaluates the spatial resolution of the fused image i.e. measured the degree of clarity of the fused image. Higher the value of AG, better the resolution of the fused image. It is represented as

$$\mathrm{AG} = \frac{1}{(M-1)(N-1)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[\frac{1}{2} \left(\left(\frac{\partial I_F(m,n)}{\partial m} \right)^2 + \left(\frac{\partial I_F(m,n)}{\partial n} \right)^2 \right) \right]^{\frac{1}{2}} \tag{26}$$

where $I_F(m, n)$ is the intensity of the fused image at position (m, n).

• Standard deviation (SD)

It measures contrast of a fused image i.e. used to determine how grey levels in an image are widely spread from the mean value. Higher the value of SD, better the contrast of an image.

$$\mathrm{SD} = \left[\frac{1}{M * N} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(I_F(m,n) - \mu \right)^2 \right]^{\frac{1}{2}} \tag{27}$$

where M^*N is size of image, $I_F(m, n)$ is intensity of fused image at positions (m, n), and μ is mean intensity of the image.

• Peak signal to noise ratio (PSNR)

It is used to determine the level of error/distortion between source images with the fused image. Higher the value of PSNR more is the degree of similarity between two signals. It is represented as

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{I_p}{\mathrm{RMSE}} \right) \tag{28}$$

where I_p is the dynamic range of allowable image pixel intensities.

• Normalized mutual information (Q_{NMI})

Mutual Information (MI) adds two un-normalized quantities and show biasedness towards the input source image with higher entropy value. Q_{NMI} measures the amount of information obtained in the fused image from both source images.

$$Q_{NMI} = 2 \left(\frac{\mathrm{MI}\left(I_F, I_A\right)}{H\left(I_F\right) + H\left(I_A\right)} + \frac{\mathrm{MI}\left(I_F, I_B\right)}{H\left(I_F\right) + H\left(I_B\right)} \right) \tag{29}$$

where (I_A), $H(I_B)$, $H(I_F)$ are marginal entropies of source images and the fused image respectively. Higher the value of NMI, better the quality of the fused image.

• Phase congruency based (Absolute feature) (Q_P)

It is based on phase congruency measurement and its principal moments which provide an absolute measurement of image features. Assuming L blocks in an image, the average over the entire image is obtained as

$$P'_{\mathrm{ab}} = \frac{1}{L} \sum_{l=1}^{L} P'_{\mathrm{ab}}(l) \tag{30}$$

where P'_{ab} is the product of three correlation coefficients and is given as $P'_{\mathrm{ab}} = (P_p)^a (P_{ma})^b (P_{mi})^c$. Here a, b, c are exponential parameter which can be adjusted, *ma* and *mi* are the maximum and minimum moments, p is the phase congruency measurement.

• Fusion quality index (Q_c)

This metric is based on Universal Image Quality Index. Fusion quality index measure how well the salient information from source images is transferred to the fused image without distortions.

$$Q_C(A,B,F) = \frac{1}{w}\sum_{w\in W}[s(w)Q(A,(F\mid w)) + (1-s(w)Q(B,(F\mid w))] \tag{31}$$

$$s(w) = \frac{s(A\mid w)}{s(A\mid w)+s(B\mid w)} \tag{32}$$

where $s(A|w)$, $s(B|w)$ are local saliencies of source images A and B in window w or variance of images A and B within window w. It depends upon contrast, sharpness or entropy. $s(w)$ indicates importance of source image A over B and $s(w) \in [0, 1]$. $Q(A, (F|w))$, $Q(B, (F|w))$ is the quality index over a window for source images A, B and the fused image F, w is the analysis window and w is the family of all windows. In contrast to HVS, this quality index treats different quality measurements within each window equally.

Another variant known as weighted fusion quality index gives more weight to those windows where saliency of input mages is higher. It is then defined as

$$Q_w(A,B,F) = \sum_{w\in W}\delta(w)[s(w)Q(A,(F\mid w)) + (1-s(w)Q(B,(F\mid w))] \tag{33}$$

where $\delta(w)$ is the normalized version of $\delta'(w)$ and $\delta'(w) = \max(s(A|w), s(B|w))$. To take into account the relevance of edge information, the metric is modified to edge dependant fusion quality index. It is computed with the edge images (A', B', F') instead of (A, B, F) and is given as

$$Q_E(A,B,F) = Q_w(A,B,F) * Q_w\left(A',B',F'\right)^{\alpha} \tag{34}$$

• Fusion similarity metric (Q_S)

Fusion quality index metric does not give clear idea of how similar individual source images are to the fused image. Also it is defined for only window size 8*8. Fusion similarity metric measures similarity between blocks of pixels in the source image and the fused image within the spatial position. It is represented as

$$Q_s(A,B,F) = \sum_{w\in W}[\mathrm{sim}(A,B,(F\mid w))Q(A,(F\mid w)) + (1-\mathrm{sim}(A,B,(F\mid w))Q(B,(F\mid w))] \tag{35}$$

where

$$\sin(A,B,(F\mid w)) = f(x) = \begin{cases} 0 & ,\frac{\sigma_{\mathrm{AF}}}{\sigma_{\mathrm{AF}}+\sigma_{\mathrm{BF}}}0 \\ \frac{\sigma_{\mathrm{AF}}}{\sigma_{\mathrm{AF}}+\sigma_{\mathrm{BF}}} & ,\frac{\sigma_{\mathrm{AF}}}{\sigma_{\mathrm{AF}}+\sigma_{\mathrm{BF}}} \geq 1 \\ 1 & ,\frac{\sigma_{\mathrm{AF}}}{\sigma_{\mathrm{AF}}+\sigma_{\mathrm{BF}}} \geq 1 \end{cases} \tag{36}$$

$$\sigma_{\mathrm{AF}} = \frac{1}{M-1}\sum_{m=0}^{M}(x_m-\bar{x})(y_m-\bar{y}) \tag{37}$$

Higher value indicates more similarity between source images and the fused image in spatial domain.

• Structural similarity index (SSIM) based (Q_Y)

The local structural similarity between input source images is used as a match measure based on how various operations are applied to the evaluations of various local areas. The quality of the composite image is determined by how near the $Q_Y(A, B, F)$ value is to 1.

$$Q_Y(A,B,(F\mid w)) = \begin{Bmatrix} s(w)\,\mathrm{SSIM}(A,(F\mid w)) + (1-s(w))\,\mathrm{SSIM}(B,(F\mid w)) & ,\text{ if } \mathrm{SSIM}(A,B\mid w) > 0.75 \\ \max\{\mathrm{SSIM}(A,(F\mid w)),\mathrm{SSIM}(B,(F\mid w))\} & ,\text{ if } \mathrm{SSIM}(A,B\mid w) < 0.75 \end{Bmatrix} \tag{38}$$

For pixels whose SSIM(A, $B|w$) is equal to or larger than a given threshold which means that the source images contain redundant information at this pixel, the weighted average of SSIM(A, $F|w$) and SSIM(B, $F|w$) is taken as the local quality, otherwise, the larger one of the two is taken. The overall fusion metric is calculated as

$$Q_Y(A,B,F) = \frac{1}{W}\sum_{w\in W}[Q_Y(A,B,(F\mid w))] \tag{39}$$

5. Results and Discussion

Any fusion algorithm's goal is to incorporate the necessary information from both source images into the final image. It is not possible to evaluate a fused image solely by looking at the fused image or by examining fusion metrics. It should be evaluated both qualitatively (subjectively) and quantitatively (objectively) utilizing visual display and fusion metrics. This subsection presents both qualitative and quantitative analysis of several methods.

5.1 Subjective evaluation

Three sets of multimodal medical images were chosen as the source images for this subsection. These sets include dataset 1 (CT/MRI), dataset 2 (CT/MRI), dataset 3 (MR-T1/MR-T2), which are displayed in Figure 2, Figure 3 and Figure 4.

Dataset 1 (CT/MRI): As was addressed in Section 1, MRI can catch soft tissues found in the brain, whereas CT can only capture hard tissues. However, utilizing the fusion process, it is vital to combine all the relevant information from various images into a single image for improved illness identification and therapy. Figure 5 gives the experimental results on the fusion of first set of medical images. The FPDE and ADF-based fused images (Figure 5c,e), experience highest contrast reduction and ADF suffer loss of information. In comparison to the approaches described above, DCT (Figure 5a), SAIF (Figure 5b), and S-ADE (Figure 5d) perform exceptionally well in terms of subjective perceptions of visual impact. Further, a detailed analysis reveals that SAIF and VSMWLS (Figure 5f) contribute some artifacts into the fused images, easily resulting in medical mishaps. On the other hand, WOADF method (Figure 5g) outperforms above method experimentally in regards to detail retention and visual effect but the fused image produced by the proposed approach has incredibly sharp edges and features as evident in Figure 5h.

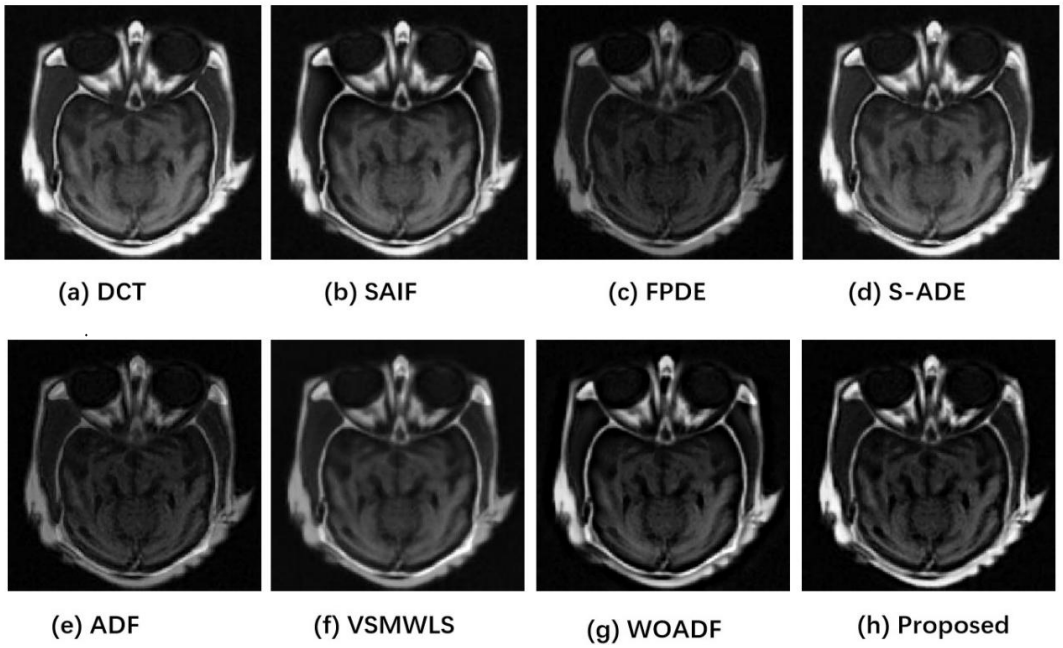

Figure 5. Visual comparison of fused images for dataset 1. DCT: discrete cosine transform; SAIF: structure-aware image fusion; FPDE: fourth-order partial differential equations; S-ADE: synchronized-anisotropic diffusion model; ADF: anisotropic diffusion and karhunen-loeve transform; VSMWLS: weighted-least-square-based scheme; WOADF: weight-optimised anisotropic diffusion filtering.

Dataset 2(CT/MRI): Figure 2 shows another set of CT and MRI images entailing a significant amount of complimentary information and Figure 6 displays the experimental results on the second set of medical images. Like dataset 1, the FPDE and ADF-based fused images contrast is reduced in dataset 2, and the blocking effect manifests in the relevant regions. As seen in Figure 6, the structure and detailed information in MRI are lost in the fused results provided by the DCT technique due to their poor definition i.e., it is unable to incorporate all the complementary information of the input images properly as there is some detail loss and poor fusion performance. The approaches of SAIF (Figure 6b), S-ADE (Figure 6d) and VSMWLS (Figure 6f) have better feature quality and contrast compared to DCT (Figure 6a), FPDE (Figure 6c), and ADF (Figure 6e) methods and combine certain vital information along with providing results that are pleasing to viewers. Further, in contrast to the fusion results of above methods, the image of the human brain's inner details is more distinct and in focus for WOADF (Figure 6g). Nonetheless, Figure 6 shows that though the contrast level of the fused image via the proposed method (Figure 6h) is not ideal, the fusion results provide visually more details compared to the source images.

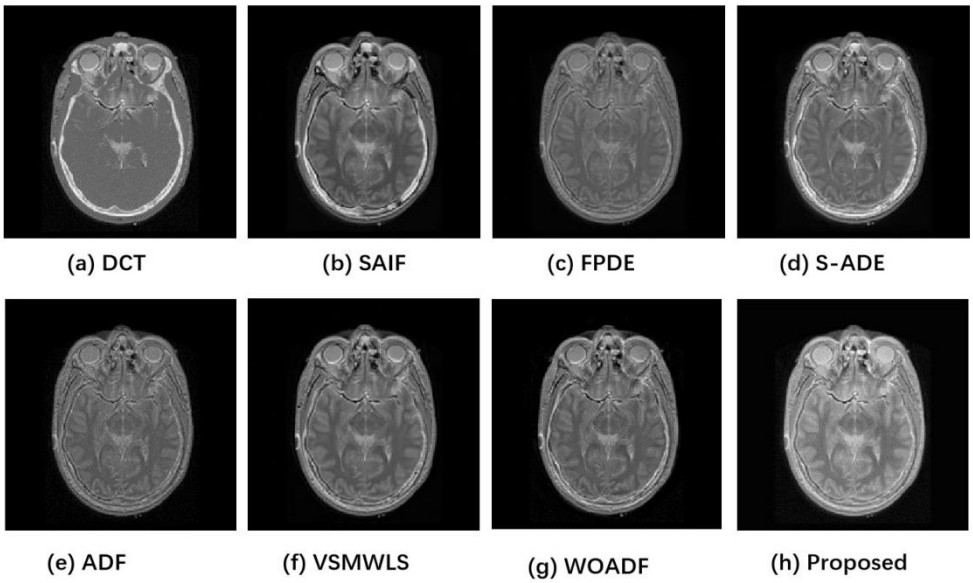

Figure 6. Visual comparison of fused images for dataset 2. DCT: discrete cosine transform; SAIF: structure-aware image fusion; FPDE: fourth-order partial differential equations; S-ADE: synchronized-anisotropic diffusion model; ADF: anisotropic diffusion and karhunen-loeve transform; VSMWLS: weighted-least-square-based scheme; WOADF: weight-optimised anisotropic diffusion filtering.

Dataset 3 (MR-T1/MR-T2): T1 and T2-weighted MR images serve as a key diagnostic tool in characterizing tissue abnormalities, and it can be observed from Figure 4 that both provide complementary information. T1-weighted images are better at displaying typical soft-tissue structures and fat distribution, which can be useful in confirming the presence of a mass containing fat. Besides, T2-weighted images show fluid and related abnormalities, including tumors, inflammation, and trauma, far more effectively. It can be observed from Figure 7c,e, the retention of edge structure and tissue information was constrained by FPDE and ADF, which resulted in a loss of energy and contrast in the fusion results. The approaches of SAIF (Figure 7b), S-ADE (Figure 7d), and VSMWLS (Figure 7f) are superior in contrast to those of other related methods but there is not enough clarity in the texture of the images in S-ADE and VSMWLS. Also, the images are evidently blurred. A clear observation suggests WOADF (Figure 7g) to be suitable among above mentioned approaches for preserving energy and retaining detailed information. However, from Figure 7h, it is evident that the features in the fused image via the proposed method have been improved with a decent blend of supplemental data from an individual image utilized as an input.

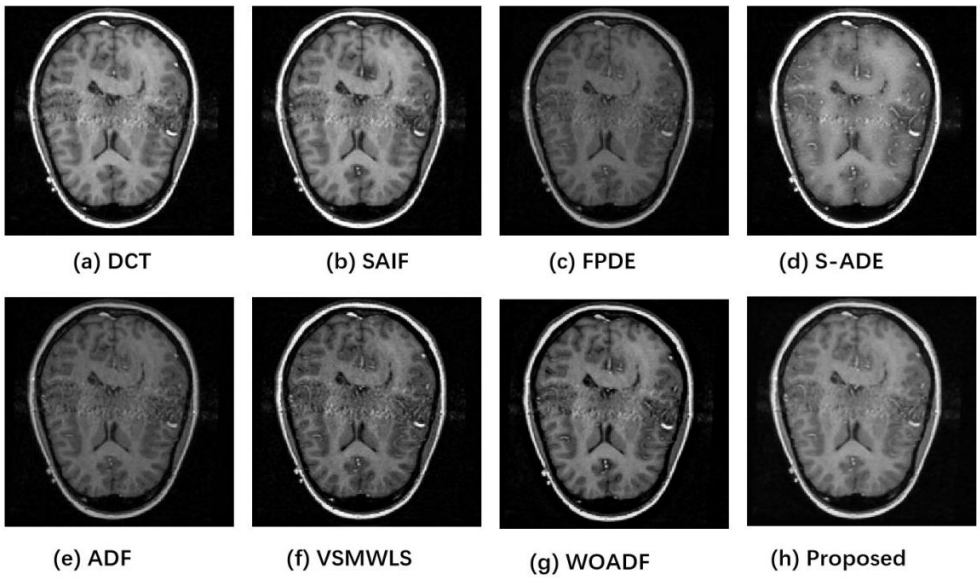

Figure 7. Visual comparison of fused images for dataset 3. DCT: discrete cosine transform; SAIF: structure-aware image fusion; FPDE: fourth-order partial differential equations; S-ADE: synchronized-anisotropic diffusion model; ADF: anisotropic diffusion and karhunen-loeve transform; VSMWLS: weighted-least-square-based scheme; WOADF: weight-optimised anisotropic diffusion filtering.

Thus, the fused images produced by the proposed approach in all the datasets is high in information content with less noise and artifacts suggesting potential benefits of integrating two-scale decomposition via mean filter with saliency detection method and guided filtering.

5.2 Objective evaluation

The human visual system is a prerequisite for subjective evaluation. Despite the ease of this kind of evaluation, it is expensive and time-consuming to conduct the numerous trials necessary to determine a compelling subjective assessment score. For unbiased quality measurement of the fused outcomes, a reliable evaluation model called as metric (as discussed in Section 4.1) is needed. The results of the objective evaluation metrics for the respective fused images in Figure 5, Figure 6 and Figure 7 are listed in Table 4. The best results under the eight metrics are bolded, second best are highlighted in red, and third best are underlined. The quantitative assessment statistics support the findings of the subjective assessment examination, showing that the proposed approach is substantially superior relative to the seven advanced comparable medical image fusion methods.

Table 4. IQA metrics of fused images for all datasets.

IQA Metrics	Image Fusion Methods							
	DCT	SAIF	FPDE	S-ADE	ADF	VSMWLS	WOADF	Proposed
	Dataset 1							
AG	6.573	6.397	4.145	6.46	4.419	6.391	6.59	**6.629**
SD	59.191	54.99	34.034	58.995	34.541	52.818	59.283	**59.362**
PSNR	55.321	55.503	56.358	55.359	56.337	55.87	56.350	**56.421**
Q_{NMI}	0.577	0.685	0.599	1.049	0.514	0.526	0.987	**1.116**
Q_P	0.575	0.574	0.501	0.587	0.457	0.541	0.571	**0.599**
Q_S	0.926	0.92	0.727	0.924	0.712	0.757	0.928	**0.967**
Q_C	0.862	0.85	0.673	0.856	0.659	0.685	0.867	**0.895**
Q_Y	0.97	0.934	0.714	0.961	0.711	0.751	0.963	**0.983**
	Dataset 2							
AG	4.335	**5.545**	3.982	5.425	4.514	5.415	5.475	5.487
SD	50.952	56.125	51.792	**61.082**	52.143	55.256	56.058	56.366
PSNR	58.389	58.385	59.431	58.276	59.326	59.042	**59.434**	59.313
Q_{NMI}	1.105	1.069	0.84	0.899	0.803	0.812	1.109	**1.112**
Q_P	0.307	0.583	0.431	0.359	0.423	0.478	0.537	**0.595**
Q_S	0.811	0.896	0.871	0.875	0.866	0.893	0.897	**0.899**
Q_C	0.79	0.752	0.659	0.676	0.567	0.735	0.742	**0.792**
Q_Y	0.935	0.941	0.819	0.853	0.762	0.884	0.942	**0.948**
	Dataset 3							
AG	8.42	8.33	5.553	7.984	5.813	8.23	**8.48**	8.348
SD	69.147	68.893	45.525	68.938	45.452	56.312	**69.222**	68.991
PSNR	56.581	56.616	57.031	56.217	57.013	56.753	57.007	**57.187**
Q_{NMI}	1.165	1.081	0.876	0.748	0.862	0.713	1.173	**1.182**
Q_P	0.801	0.753	0.596	0.242	0.663	0.49	0.818	**0.833**
Q_S	0.838	0.836	0.804	0.734	0.799	0.807	0.835	**0.842**
Q_C	0.853	0.889	0.711	0.559	0.759	0.718	0.851	**0.899**

Q_Y	0.957	0.969	0.771	0.783	0.812	0.797	0.958	**0.979**
	Average Computational Time							
Time	0.274	0.137	0.616	12.37	0.532	0.538	0.496	0.412

IQA: image quality assessment; DCT: discrete cosine transform; SAIF: structure-aware image fusion; FPDE: fourth-order partial differential equations; S-ADE: synchronized-anisotropic diffusion model; ADF: anisotropic diffusion and Karhunen-Loeve transform; VSMWLS: weighted-least-square-based scheme; WOADF: weight-optimised anisotropic diffusion filtering; AG: average gradient; SD: standard deviation; PSNR: peak signal to noise ratio; Q_{NMI}: normalized mutual information**)**; Q_p: phase congruency based; Q_s: fusion similarity metric; *Qc*: fusion quality index; Q_y: structural similarity index (SSIM) based.

AG reflects the enhancement of textural transformation features, the contrast amid individual pixels, and image fusion quality. In Table 4, the highest value (for dataset1) and second highest values (for dataset 2 and 3) of the proposed method for AG indicate higher image clarity and greater reflection of fine details and texture in the fused image. While PSNR is used to assess the degree of image distortion in an image and SD calculates the contrast of the fused image. The proposed approach showed highest values of SD followed by WOADF and DCT for dataset 1, and second highest values for dataset 2 and dataset 3. It can be further verified that the proposed approach achieved superior results for PSNR in all datasets. Therefore, the proposed approach's image fusion has a superior ability to display the contrast of minute features. Additionally, the image's clarity is the best and fewer distortions introduced during the fusion process. Furthermore, the proposed approach exhibits comparatively higher metrics for Q_{NMI}, Q_P, Q_S, Q_C and Q_Y showing increased mutual information interaction, phase congruency, sufficient edge information, high resemblance and structural information retention respectively suggesting that it preserves more substantial data pertaining to the energy, details, and structure of the medical image. As a result, the proposed approach yields a fused image with a greater amount of data, higher quality, better edge preservation, along with assessment results that are consistent with those of human visual perception. Further, less information is lost during the fusion process, and the anti-noise performance is improved.

It can be concluded that even though not all IQA measures are among the most effective for all datasets utilized for experimentation, the fusion impact is the best overall (with the majority of the metrics being listed in the top three). In general, the MMIF method that has been proposed exhibits superior efficacy, allowing it to fuse more details in the final fused image.

5.3 Computation time

For real-time operations, an IF approach that requires less calculation time and generates more aesthetically acceptable images with higher IQA metric values is preferred. Table 4 displays the average running time of the proposed approach. With a running duration of less than one second, the proposed approach, DCT, SAIF, FPDE, ADF, VSMWLS and WOADF exhibit great fusion efficiency. The SAIF method has the quickest execution time followed by DCT and the proposed approach. While S-ADE requires a lot of time, the average running time of FPDE, ADF, VSMWLS and WOADF is acceptable for experimental simulations. The experimentation has been carried out on HP notebook Intel(R) Core (TM) i5-6200U CPU @ 2.30GHz 2.40 GHz processor and 8.00 GB internal RAM. Furthermore, both from the stand points of subjective visual impacts and objective assessment metrics, the proposed approach surpasses the other seven approaches in terms of performance. In addition, it minimized the dependence on MST, has fewer computational resources and memory space, decreased algorithm complexity resulting in less computational time compared to advanced image fusion methods.

6. Conclusion and Future Directions

Using MMIF, clinicians may more easily diagnose and analyze patients' cases by combining the vital details from two images into one. Though the spatial inconsistency issues faced by the conventional multi-scale transform-based approaches can be solved using a GF-based approach, a new difficulty arises in that some information will be missed. The proposed approach combines the advantages of two-scale image decomposition and GF. While GFs can effectively minimize artifacts by optimizing the weighting maps, two-scale decomposition can maintain the fine details of multi-modal medical images. Furthermore, the proposed approach has been evaluated using three datasets of medical images. According to visual effect and objective evaluation, the proposed approach works significantly better than other widely used algorithms and would be useful for identifying targets, recognition, and various other tasks involving image processing since it could effectively maintain and improve regions and specifics related to HVS interests. This will help radiologists and other human observers find and assess medically pertinent abnormalities among normal anatomy and physiology and ensure healthy lives for patients.

It is important to note that this work just offers an exploratory effort intended to demonstrate the enormous prospective of saliency detection for medical image fusion. The subsequent study on this subject could focus on the areas listed below in a greater depth. Extension to different image fusion applications: In addition to MMIF, other types of image fusion tasks, like visible-infrared image fusion and multi-exposure image fusion, can also greatly benefit from the method. Establishing more advanced fusion approaches: The image processing methods used in this work are rather straightforward, so more complex methods or approaches can be created to boost the fusion quality. By constructing more efficient fusion techniques, there is unambiguously a lot of space to advance in this direction.

Furthermore, the endurance and expansion of the proposed approach may be intriguing topics. It has two different aspects. First, it's

important to investigate whether the proposed approach is universally applicable or not. Second, the source images acquired in normal circumstances may not always be sufficiently clear and may be distorted by noise, which has a detrimental impact on the diagnosis of diseases and the formulation of treatment regimens. As a result, evaluating the feasibility of the proposed approach is a crucial research task. In conclusion, future study will focus mostly on how to improve and optimize the efficacy of the proposed approach.

Declarations

Authors contribution

Kaur H: Methodology, writing-original draft, review & editing.
Vig R, Kumar N: Investigation, validation.
All authors approved the final version of the manuscript.

Conflicts of interest

Not applicable.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data and materials could be obtained from the corresponding author.

Funding

None.

References

1. James AP, Dasarathy BV. Medical image fusion: A survey of the state of the art. Inf Fusion. 2014;19(1):4-19. [DOI]
2. El-Gamal FEZA, Elmogy M, Atwan A. Current trends in medical image registration and fusion. Egypt Inform J. 2016;17(1):99-124. [DOI]
3. Hermessi H, Mourali O, Zagrouba E. Multimodal medical image fusion review: Theoretical background and recent advances. Signal Process. 2021;183:1-28. [DOI]
4. 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. [DOI]
5. Ma J, Zhou Z, Wang B, Zong H. Infrared and visible image fusion based on visual saliency map and weighted leastsquareoptimization. Infrared Phys Technol. 2017;82:8-17. [DOI]
6. Bavirisetti DP, Dhuli R. Two-scale image fusion of visible and infrared images using saliency detection. Infrared Phys Technol. 2016;76:52-64. [DOI]
7. Jiang Q, Jin X, Chen G, Lee SJ, Cui X, Yao S, et al. Two-scale decomposition-based multifocus image fusion framework combined with image morphology and fuzzy set theory. Inf Sci. 2020;541:442-474. [DOI]
8. Borji A, Itti L. State-of-the-Art inVisual Attention Modeling. IEEE Trans Pattern Anal Mach Intell. 2013;35(1):185-207. [DOI]
9. Wen G, Rodriguez-Niño B, Pecen FY, Vining DJ, Garg N, Markey MK. Comparative study of computational visual attention models on two-dimensional medical images. J Med Imag. 2017;4(2):025503. [DOI]
10. Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Trans Pattern Anal Mach Intell. 1998;20(11):1254-1259. [DOI]
11. Bruce N, Tsotsos J. Attention based on information maximization. J Vis. 2017;7(9):950. [DOI]
12. Harel J, Koch C, Perona P. Graph-based visual saliency. Adv Neural Inf Process Syst. 2007;19:545-552. [DOI]
13. Hou X, Zhang L. Saliency detection: A spectral residual approach. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition; 2007 Jun 17-22; Minneapolis, USA. New York: IEEE; 2007. p. 1-8. [DOI]

14. Achanta R, Hemami S, Estrada F, Süsstrunk S. Frequency-tuned salient region detection. In: Conference on Computer Vision and Pattern Recognition (CVPR). 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009 Jun 20-25; Miami, USA. New York: IEEE; 2009. p. 1597-1604. [DOI]
15. Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2010 Jun 13-18; San Francisco, CA. New York: IEEE; 2010. p. 2376-2383. [DOI]
16. Achanta R, Süsstrunk S. Saliency detection using maximum symmetric surround. In: Conference on Computer Vision and Pattern Recognition (CVPR). 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010 Jun 12-18; San Francisco, USA. New York: IEEE; 2010. p. 2653-2656. [DOI]
17. Li S, Kang X, Hu J. Image fusion with guided filtering. IEEE Trans Image Process. 2013;22(7):2864-2875. [DOI]
18. Wang Y, Wu Q, Castleman KR. Image enhancement. In: Merchant FA, Castleman KR, editors. Microscope image processing. 2nd ed. Amsterdam: Elsevier; 2023. p. 55-74. [DOI]
19. Harvard Medical School. Whole Brain Atlas [Internet]. Boston (MA): Harvard Medical School. Available from: http://www.med.harvard.edu/aanlib/.
20. Wang M, Shang X. A Fast Image Fusion with Discrete Cosine Transform. IEEE Signal Process Lett. 2020;27:990-994. [DOI]
21. Li W, Xie Y, Zhou H, Han Y, Zhan K. Structure-aware image fusion. Optik. 2018;172:1-11. [DOI]
22. Bavirisetti DP, Xiao G, Liu G. Multi-sensor image fusion based on fourth order partial differential equations. In: International Conference on Information Fusion. 2017 20th International Conference on Information Fusion (Fusion); 2017 Jul 10-13; Xi'an, China. New York: IEEE; 2017. p. 1-9. [DOI]
23. Zhu R, Li X, Zhang X, Ma M. MRI and CT medical image fusion based on synchronized-anisotropic diffusion model. IEEE Access. 2020;8:91336-91350. [DOI]
24. Bavirisetti DP, Dhuli R. Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loevetransform. IEEE Sens J. 2015;16(1):203-209. [DOI]
25. Ma J, Zhou Z, Wang B, Zong H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Phys Technol. 2017;82:8-17. [DOI]
26. Vasu GT, Palanisamy P. CT and MRI multi-modal medical image fusion using weight-optimized anisotropic diffusion filtering. Soft Comput. 2023;27:9105-9117. [DOI]
27. Liu Z, Blasch E, Xue Z, Zhao J, Laganiere R, Wu W. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE Trans Pattern Anal Mach Intell. 2012;34(1):94-109. [DOI]

Publisher's Note