J. Appl. & Pure Math. Vol. **5***(2023), No. 5 - 6, pp. 389 - 406*
https://doi.org/10.23091/japm.2023.389

MOSAICFUSION: MERGING MODALITIES WITH PARTIAL DIFFERENTIAL EQUATION AND DISCRETE COSINE TRANSFORMATION

GARGI TRIVEDI, RAJESH SANGHAVI*

Abstract. In the pursuit of enhancing image fusion techniques, this research presents a novel approach for fusing multimodal images, specifically infrared (IR) and visible (VIS) images, utilizing a combination of partial differential equations (PDE) and discrete cosine transformation (DCT). The proposed method seeks to leverage the thermal and structural information provided by IR imaging and the fine-grained details offered by VIS imaging create composite images that are superior in quality and informativeness. Through a meticulous fusion process, which involves PDE-guided fusion, DCT component selection, and weighted combination, the methodology aims to strike a balance that optimally preserves essential features and minimizes artifacts. Rigorous evaluations, both objective and subjective, are conducted to validate the effectiveness of the approach. This research contributes to the ongoing advancement of multimodal image fusion, addressing applications in fields like medical imaging, surveillance, and remote sensing, where the marriage of IR and VIS data is of paramount importance.

AMS Mathematics Subject Classification: 54H30, 68U10, 65M06.
Key words and phrases : Infrared images(IR), visible images(VI), partial differential equation(PDE), discrete cosine transformation (DCT), image fusion.

1. Introduction

The fusion of infrared (IR) and visible (VIS) images has emerged as a powerful and transformative technique in the realm of image processing, computer vision, and remote sensing[1, 2]. The synergy between these two distinct imaging modalities, each preprocessing its unique set of advantages and limitations, opens doors to a wide range of applications with significant societal and technological implications. Infrared imaging captures the thermal radiation emitted

Received August 23, 2023. Revised September 8, 2023. Accepted September 15, 2023.
*Corresponding author.

by objects, enabling the detection of temperature variations and objects in total darkness or adverse weather conditions. Visible imaging, on the other hand, records the electromagnetic spectrum within the range of human vision and is adept at capturing fine details, colors, and textures. By combining these modalities, multi-modal image fusion aims to create composite images that inherit the strengths of both while mitigating their respective weaknesses[3]. The motivation behind fusing IR and VIS images is manifold. In the domain of surveillance and security, the fusion enhances object detection and recognition, enabling the identification of threats, intruders, or objects of interest even in low-light conditions[4]. In medical imaging, multi-modal fusion can aid in the diagnosis of diseases by providing complementary information for tissue characterization and assessment[5].

The fusion process itself is a multifaceted challenge, involving the integration of pixel-level information from IR and VIS sources. It necessitates a careful balance between retaining the essential structural and thermal details from the IR modality and the fine-grained visual cues from the VIS modality. Achieving this balance while minimizing artifacts and noise is a non-trivial task, leading to the development of innovative fusion methods and algorithms.

This research endeavors to contribute to the evolving field of multi-modal image fusion by proposing a fusion methodology that harnesses the potential of differential equations and discrete transformation to optimally combine IR and VIS images. The proposed approach aims to create fused images that enhance the overall quality, informativeness, and applicability of the composite data, making strides towards improved decision-making and understanding in domains where IR and VIS data intersect In this paper, we present the details of our fusion methodology, including the principles, parameters, and processes involved. We also provide comprehensive evaluations, both objective and subjective, to validate the efficacy of our approach. The outcomes of this research not only have the potential to advance the state of the art in multi-modal image fusion but also stand to benefit numerous fields where the marriage of IR and VIS imaging is a vital component in addressing contemporary challenges.

2. Litrature Review

Multi-modal image fusion, particularly the fusion of infrared (IR) and visible (VIS) images, has emerged as a transformative technique with far-reaching applications. This review provides an overview of the key developments and methodologies in IR and VIS image fusion which offers an attractive proposition due to the complementary nature of these two modalities. IR imaging excels in capturing thermal radiation, making it effective in low-light conditions and adverse weather. On the other hand, VIS imaging offers high-resolution, color-rich details. By combining these modalities, multi-modal fusion aims to create composite images that harness the strengths of both, surpassing the limitations of individual modalities[6, 7].

Early approaches to multi-modal image fusion involved traditional methods, including simple pixel-level averaging and weighted averaging. While these techniques provided fundamental insights into fusion principles, they often struggled to preserve vital details and eliminate noise. As technology advanced, spatial domain fusion methods gained prominence. Techniques such as wavelet transform and Laplacian pyramids allowed for the extraction and combination of information at different scales, improving fusion results and detail preservation[8, 9].In the quest for more robust fusion methods, frequency domain approaches came to the fore. Discrete Fourier transform (DFT)[10] and discrete cosine transformation (DCT) have been employed to analyze and fuse images in the frequency domain[11, 12]. These methods excel in the selection of dominant frequency components and the subsequent combination of information.

Recent years have witnessed a shift towards advanced fusion techniques, including the integration of deep learning-based approaches. Convolutional neural networks (CNNs) and generative adversarial networks (GANs) have shown promise in automating the fusion process, learning representations, and optimizing fusion performance[13, 14]. Applications of IR and VIS image fusion are extensive. In the medical field, fusion has found applications in disease diagnosis, tissue characterization, and surgical navigation, ultimately leading to more accurate clinical assessments and improved patient outcomes. In remote sensing and surveillance, the fusion enhances object detection and recognition, contributing to improved situational awareness and decision-making in complex environments challenges in multimodal image fusion include the selection of an optimal fusion method, the choice of fusion parameters, and the fine-tuning of fusion algorithms. Future research directions encompass the integration of real-time processing, the development of domain-specific fusion methods, and user-centric optimization for diverse applications. This research contributes to the advancement of multi-modal image fusion by proposing a novel methodology that harnesses differential equations and discrete cosine transformation to optimally combine IR and VIS images. The goal is to create fused images that enhance overall quality, informativeness, and applicability, thereby addressing contemporary challenges and fostering improved decision-making and understanding in domains where IR and VIS data converge.

3. Fusion Approaches in Multi-modal Image Fusion

In multi-modal image fusion, various approaches are employed to combine information from different sources. These approaches can be broadly categorized into spatial domain fusion, frequency domain fusion, and hybrid fusion methods. Each approach has its advantages and is selected based on the specific requirements of the fusion task.

3.1. Spatial Domain Fusion. Spatial domain fusion involves traditional methods based on pixel-level operations. One of the fundamental techniques

in this domain is the Non-Subsampled Shearlet Transform (NSST). Spatial domain fusion methods focus on capturing spatial structures and relationships within images. The Non-Subsampled Shearlet Transform (NSST) is a powerful tool for spatial domain fusion, particularly when the preservation of spatial structures, textures, and directional information is crucial[15, 16, 17].

NSST represents an image I as a linear combination of shearlet coefficients at various scales and orientations. Each shearlet coefficient encapsulates specific spatial frequencies and directions within the image. Mathematically, NSST can be expressed as:

$$I(x,y) = \sum_{j=1}^{J} \sum_{l=1}^{L_j} \sum_{k=1}^{K_j} s_j(k,l) \cdot \psi_{j,k,l}(x,y) \tag{1}$$

In equation (1), $I(x,y)$ represents the pixel value of the image at coordinates (x,y), J denotes the number of scales, L_j is the count of shearlets at scale j, K_j is the count of shearlets at each scale j, $s_j(k,l)$ denotes the shearlet coefficients, $\psi_{j,k,l}(x,y)$ is the shearlet basis function.

In the context of spatial domain fusion, NSST can be applied to the input images to extract their spatial and directional components. These components can then be fused to create a new image that encapsulates essential features from each modality. The fusion process using NSST is defined as:

$$F_{\text{NSST}}(x,y) = \sum_{j=1}^{J} \sum_{l=1}^{L_j} \sum_{k=1}^{K_j} \alpha_{j,k,l} \cdot s_{j,k,l} \tag{2}$$

Where in equation (2), $F_{\text{NSST}}(x,y)$ represents the pixel value of the fused image, $\alpha_{j,k,l}$ denotes fusion weights determined based on the shearlet coefficients, $s_{j,k,l}$ are shearlet coefficients obtained from the NSST of the input images.

3.2. Frequency Domain Fusion. utilize techniques like the Discrete Wavelet Transform (DWT) to analyze the frequency components of images. DWT is a powerful tool in this context, as it decomposes an image into wavelet coefficients at different scales and positions, enabling selective fusion based on frequency characteristics.DWT is a widely adopted technique in frequency domain fusion. It decomposes an image into wavelet coefficients at multiple scales and positions[18, 19, 20, 21, 22]. The DWT of an image can be mathematically represented as.

$$DWT(I) = LL_J \oplus LH_J \oplus HL_J \oplus HH_J \oplus \ldots \oplus LH_1 \oplus HL_1 \oplus HH_1 \tag{3}$$

Here in equation (3), $DWT(I)$ represents the DWT coefficients of the input image I, LL_J is the approximation at the finest scale, LH_J, HL_J, and HH_J represent horizontal, vertical, and diagonal details at the finest scale, Subscripts from J to 1 indicate different scales, with J being the finest scale.

In frequency domain fusion, DWT coefficients are utilized to capture frequency information from the input images. The fusion process combines coefficients from various scales and orientations to create the fused image which is expressed below.

$$F_{\mathrm{DWT}}(x,y) = \sum_{j=1}^{J} \sum_{d\in\{LL,LH,HL,HH\}} \alpha_{j,d} \cdot DWT(I_{j,d})(x,y) \tag{4}$$

In equation (4),$F_{\mathrm{DWT}}(x,y)$ represents the pixel value of the fused image, $I_{j,d}(x,y)$ denotes the DWT coefficient at scale j and direction d, $\alpha_{j,d}$ are fusion weights determined based on the DWT coefficients.

3.3. Hybrid Fusion. Hybrid fusion methods combine both spatial and frequency domain techniques to exploit the strengths of both domains. These methods leverage the complementary nature of spatial and frequency information[23]. The fusion equation for hybrid fusion can be defined as:

$$F_{\mathrm{hybrid}}(x,y) = \alpha \cdot F_{\mathrm{spatial}}(x,y) + (1-\alpha) \cdot F_{\mathrm{frequency}}(x,y) \tag{5}$$

Here, $F_{\mathrm{hybrid}}(x,y)$ represents the fused pixel value, $F_{\mathrm{spatial}}(x,y)$ and $F_{\mathrm{frequency}}(x,y)$ are the results obtained from spatial and frequency domain fusion, and α is the fusion weight. Hybrid fusion methods aim to combine the advantages of both domains to enhance the overall fusion quality.

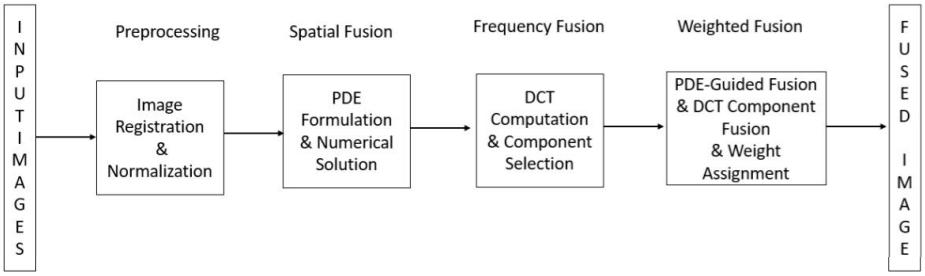

FIGURE 1. Process Flow of Mosaic Fusion

4. Proposed Hybrid Fusion Method(Mosaic Fusion)

This section provides a comprehensive overview of the proposed method (Mosaic Fusion) intended to overcome the challenges inherent in multi-modal image fusion. section offers a detailed, step-by-step explanation of the entire fusion process, starting from the preprocessing to the weighted fusion of results based on differential equations and Discrete Cosine Transform (DCT), which will be clarified using mathematical equations, algorithmic descriptions, and insights into the fundamental principles. The resultant fused image consolidates information from the input modalities, effectively addressing the research problem and

enriching our understanding of the subject matter. This introductory overview lays the groundwork for a comprehensive explanation of the entire fusion process, highlighting the significance of each step.

4.1. Image preprocessing. Before the fusion process, the multi-modal images undergo prepossessing to ensure consistency and quality. The following steps are performed:

4.1.1. Image Registration. If the input images are not aligned spatially, a registration process is applied to bring them into a common coordinate system. Image registration is essential to ensure that corresponding features in both modalities are accurately aligned. Two input images are denoted as $I_1(x, y)$ and $I_2(x, y)$ where x and y are the spatial coordinates. The goal is to find a transformation function T that maps the coordinates of I_2 to match those of I_1 so that corresponding features are accurately aligned. The registration process is expressed as:

$$I_2'(x, y) = I_2\big(T(x, y)\big) \tag{6}$$

where ,$I_2'(x, y)$ represents the registered version of original unregistered image $I_2(x, y)$, and $T(x, y)$ is given by

$$T(x, y) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

in this θ represents the rotation angle,Δx and Δy the translation parameters.

4.1.2. Normalization. This step ensures that pixel intensities have similar ranges, making them more amenable to fusion by Scaling the pixel values to a common range, such as [0, 1] is a crucial step in image processing which is important for accurate fusion and the preservation of image features across different modalities. Mathematically, the normalization of each pixel value in an image $I(x, y)$ is shown below.

$$I_{\text{normalized}}(x, y) = \frac{I(x, y) - \min(I)}{\max(I) - \min(I)} \tag{7}$$

Here, $I_{\text{normalized}}(x, y)$ represents the normalized pixel value at coordinates (x, y) in the image I. $\min(I)$ and $\max(I)$ are the minimum and maximum pixel values in the image I.

4.2. Partial Differential Equation - Based Fusion. The fusion problem is represented as a set of partial differential equations [24, 25, 26]. These equations capture the relationships between pixel intensities from different modalities, considering spatial variations and structural information as shown below.

$$\frac{\partial U}{\partial t} = \nabla \cdot \big(c(x, y, t)\nabla U\big) \tag{8}$$

Here, U represents the fused image, ∇ is the gradient operator, and $c(x, y, t)$ is the diffusion coefficient. To solve this equation numerically using finite differences, it can discretize it in both time (t) and space (x, y) by forming the basis for the numerical solution of the differential equations using finite differences[27]. First, it divides the continuous time interval into discrete time steps. Let Δt be the time step. The time derivative is then approximated as:

$$\frac{\partial U}{\partial t} \approx \frac{U^{n+1} - U^n}{\Delta t} \tag{9}$$

Here, U^{n+1} represents the value of the fused image at the next time step, and U^n represents the current time step.

Secondly, it divides the spatial domain (the pixel grid) into discrete grid points. Let Δx and Δy be the grid spacing in the x and y directions, respectively. The spatial derivatives are approximated using finite differences in the x direction and the y direction: respectively are shown in the following equation.

$$\frac{\partial}{\partial x} \approx \frac{U(i+1, j) - U(i-1, j)}{2\Delta x} and \frac{\partial}{\partial y} \approx \frac{U(i, j+1) - U(i, j-1)}{2\Delta y} \tag{10}$$

From equation (8) and equation (9), a differential equation can be written in its discrete form by Combining Temporal and Spatial Discretizations as

$$\frac{U^{n+1} - U^n}{\Delta t} = \nabla \cdot (c(x, y, t)\nabla U) \tag{11}$$

$$U^{n+1} = U^n + \Delta t \nabla \cdot (c(x, y, t)\nabla U) \tag{12}$$

The above equation shown in represents the iterative update to obtain the fused image at the next time step, denoted as U^{n+1} which is solved for by iteratively updating U^n using the discrete spatial representation and the partial differential equation governing the fusion process.

4.2.1. Discrete Cosine Transform (DCT). The numerical solution of the differential equations guides the fusion process. It helps in preserving essential features while reducing noise and artifacts. The fusion is performed iteratively to optimize the results.

The Discrete Cosine Transform (DCT) is a fundamental mathematical tool widely used in digital image processing. Known for its efficient energy compaction characteristics, the DCT effectively concentrates most of its larger coefficients in the low-frequency range [28, 29, 30]. The DCT is formally defined by [29] as follows:

$$Img\,(p, q) = \alpha\,(p)\,\alpha\,(q) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} Img\,(x, y) \cos\left(\frac{\pi\ (2x+1)p}{2m}\right) \cos\left(\frac{\pi\ (2y+1)q}{2n}\right) \tag{13}$$

Here, the discrete frequency variables $p \in [0, m-1]$ and $q \in [0, n-1]$ represent

the indices of the (x, y) pixels within the image. The coefficients $\alpha(p)$ and $\alpha(q)$ are defined as:

$$\alpha(p) = \begin{cases} \frac{1}{\sqrt{m}}, & p = 0 \\ \sqrt{\frac{2}{m}}, & 1 \leq p \leq m-1 \end{cases} \tag{14}$$

$$\alpha(q) = \begin{cases} \frac{1}{\sqrt{n}}, & q = 0 \\ \sqrt{\frac{2}{n}}, & 1 \leq q \leq n-1 \end{cases} \tag{15}$$

The basis functions of the DCT used as a foundation, represent the weights applied to each function. Visualizing these basis functions, a typical matrix size of 8×8 employs 64 basic functions, as shown in Figure 1.

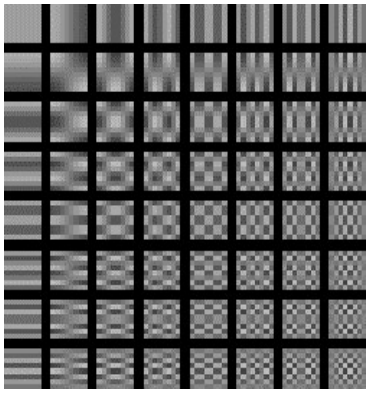

FIGURE 2. The 64 Basis Functions of an 8×8 Matrix.

In an $n \times n$ matrix representing a two-dimensional image or signal, frequencies progress vertically from top to bottom and horizontally from left to right. The basis function located at the top left, distinguished by a constant value, is commonly referred to as the DC (Discrete Cosine) basis function. Similarly, the DCT coefficient $Img(0, 0)$ is often termed the DC coefficient.

4.2.2. Frequency Domain Transformation and Component Selection. Utilizing Equation (12), the process of fusing the dominant frequency components selected from both modalities involves employing the Discrete Cosine Transform (DCT). The outcome of this fusion, termed as U_f, emerges from the amalgamation of these dominant components, thereby expressing images in the context of spatial frequency components. The procedure entails the selection of these dominant frequency components through a comparison of each coefficient's magnitude with a predetermined threshold. Coefficients with magnitudes exceeding the threshold are deemed dominant and contribute to the fusion process.

For each Modality:

$$|(u,v)|_D = \begin{cases} |(u,v)|_F, & \text{if } |(u,v)|_F \geq T \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

The matrices $[|(u,v)|_D]$ now contain the dominant frequency components for each modality. These components are retained for the fusion process, while the less informative components (those below the threshold) are set to zero.

4.3. Weighted Fusion and Weighting Strategies. The selected DCT coefficients from Modality 1, $|(u,v)|_{D1}$, and Modality 2,$|(u,v)|_{D2}$, are combined using a weighted approach:

$$|(u,v)|_F = w_1 \cdot |(u,v)|_{D1} + w_2 \cdot |(u,v)|_{D2} \tag{17}$$

Here, $U_f(u,v)$ represents the fused image, and w_1 and w_2 are the weights that determine the contribution of Modality 1(IR) and Modality 2(VIS) to the fused image.

A weighted fusion scheme is applied to combine the results from the PDE-based fusion and the DCT-based component selection. The fused DCT coefficients, denoted as $F_{\text{fused}}(u,v)$, are obtained using the following weighted combination:

$$F_{\text{fused}}(u,v) = \alpha \cdot F_{\text{PDE}}(u,v) + (1-\alpha) \cdot F_{\text{DCT}}(u,v) \tag{18}$$

Where, $F_{\text{fused}}(u,v)$ is the fused DCT coefficients,$F_{\text{PDE}}(u,v)$ is the DCT coefficients from the differential equation-based fusion, $F_{\text{DCT}}(u,v)$ is the DCT coefficients from DCT component selection, α is the fusion weight.

5. Experiments and Results

In our experimental evaluations, we extensively tested our hybrid image fusion technique(Mosaic Fusion) using MATLAB R2021a on a system powered by an Intel i7 processor, operating at 144 Hz, and with 16 GB of RAM. The purpose of these simulations was to thoroughly assess the efficacy of our proposed approach. It's important to note that all images used in these experiments were in grayscale, and the source images can be referenced in [31, 32].

Our assessment involved a comparative analysis of our Mosaic method against established fusion techniques, including Non-Subsampled Shearlet Transform (NSST), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) [34], and Fourth-Order Partial Differential Equations (FPDE) [33]. We conducted both quantitative and qualitative analyses to discern and compare the strengths and limitations of these fusion methods. A series of experiments were conducted to evaluate our proposed methodology for multimodal image fusion, integrating partial differential equations (PDE) and discrete cosine transformation (DCT). This section outlines the critical components and setup used in these experimental assessments."

5.1. Parametric Settings. This section delineates the critical parameters and configurations employed in our approach to multimodal image fusion. These settings play a pivotal role in shaping the behavior and performance of the fusion algorithm, ensuring optimal outcomes and reproducibility for future reference.

To achieve spatial alignment of input modalities, we performed image registration using a rigid transformation model, allowing a maximum registration error of 0.5 pixels. To ensure pixel intensity consistency across modalities, we applied min-max scaling normalization, thereby transforming pixel intensities within the input modalities into the range [0, 1]. This normalization process aimed to standardize intensities for effective fusion. Employing a first-order partial differential equation (PDE) as per Equation (8) for the fusion process, we utilized a finite differences numerical scheme to solve these PDEs efficiently. A time step of $\Delta t = 0.01$ seconds was selected to maintain numerical stability. For the Discrete Cosine Transform (DCT) configuration, a block size of 8×8 pixels was found to be effective in capturing relevant frequency components. We applied a threshold ($T = 0.1$) to select DCT components, retaining those with magnitudes surpassing this value for fusion. The fusion process adopted a weighted sum approach, assigning a weight of 0.6 to the PDE-guided fusion result and a combined weight of 0.4 to the DCT components. This weighting scheme was determined through a sensitivity analysis aimed at enhancing fusion performance.

5.2. Objective Evaluation. In the realm of image processing and quality assessment, four widely utilized metrics for evaluating the quality of fused images are the Normalized Mutual Information (NMI), Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), and visual assessments by experts.

Normalized Mutual Information (NMI) can be expressed by considering two random variables X and Y, and their joint distribution $p(X, Y)$. The mutual information (MI) between X and Y is given by,

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x) \cdot p(y)}{p(x,y)} \right) \tag{19}$$

Where, - $I(X;Y)$ is the mutual information between X and Y,$p(x,y)$ is the joint probability distribution of X and Y ,$p(x)$ and $p(y)$ are the marginal probability distributions of X and Y, respectively.The individual entropies of X and Y are given by,

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{20}$$

$$H(Y) = -\sum_{y \in Y} p(y) \log p(y) \tag{21}$$

The NMI between X and Y is calculated as the normalized mutual information, taking into account their individual entropies:

$$NMI(X;Y) = \frac{I(X;Y)}{\sqrt{H(X) \cdot H(Y)}} \tag{22}$$

NMI ranges from 0 to 1, where 0 indicates no mutual information (no similarity) between X and Y, and 1 represents perfect mutual information (perfect similarity). In the context of image fusion, NMI can be used to evaluate the similarity between the fused image and the input modalities. A higher NMI value indicates a stronger agreement or similarity between the fused image and the original images, suggesting that the fusion process has effectively preserved the information from the input modalities.

Root Mean Square Error (RMSE) is a measure of the average magnitude of the errors between the pixels of the fused image and a reference image. It quantifies the square root of the average of the squared differences between the corresponding pixels of the two images. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_{\text{fused}}(i,j) - I_{\text{ref}}(i,j))^2} \tag{23}$$

Where, $I_{\text{fused}}(i,j)$ is the pixel value in the fused image at coordinates (i,j), $I_{\text{ref}}(i,j)$ is the pixel value in the reference image at the same coordinates, M and N are the dimensions of the images.

RMSE is a measure of the discrepancy between the fused image and the reference image, where lower RMSE values indicate higher similarity.

Peak Signal-to-Noise Ratio (PSNR) measures the ratio of the peak signal power to the power of the noise, expressed in decibels (dB). PSNR is defined as,

$$PSNR = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{RMSE^2} \right) \tag{24}$$

Where, MAX is the maximum possible pixel value (e.g., 255 for 8-bit images), $RMSE$ is the Root Mean Square Error.

PSNR quantifies the quality of the fused image by considering the magnitude of errors in relation to the maximum possible pixel value. Higher PSNR values indicate higher image quality and fidelity.

Table 1 presents a comparative performance evaluation of various fusion methods applied to multi-modal images across different assessment metrics and diverse image scenes. The assessment metrics include Normalized Mutual Information (NMI), Root Mean Squared Error (RMSE), Peak Signal-to-Noise Ratio (PSNR) in decibels, and Visual Assessment on a scale from 1 to 10. For the 'Soldier with Jeep' image, the Fourth-Order Partial Differential Equations (FPDE) method outperformed others, achieving the highest NMI score (0.725) and the second-highest Visual Assessment score (8.0), signifying superior information preservation and visual fidelity. Conversely, the Mosaic Approach showcased relatively lower performance in NMI (0.782) but excelled in RMSE

Table 1. Performance Comparison of Multi-Modal Images with Five Fusion Methods

Methods	DCT	DWT	NSST	FPDE	Mosaic Approach
NMI Score	0.572	0.521	0.621	0.725	0.682
RMSE Score	0.074	0.063	0.057	0.065	0.031
PSNR (dB)	60.947	60.122	60.930	60.470	60.824
Visual Assessment (Scale 1- 10)	6.0	6.8	7.2	8.0	8.5
Soldier with Jeep					
NMI Score	0.377	0.157	0.877	0.852	0.796
RMSE Score	0.093	0.104	0.125	0.011	0.045
PSNR (dB)	57.178	58.432	58.439	57.616	57.924
Visual Assessment (Scale 1- 10)	6.8	7.3	6.8	8.5	8.4
Tank					
NMI Score	0.457	0.357	0.852	0.843	0.956
RMSE Score	0.073	0.112	0.067	0.0782	0.057
PSNR (dB)	56.878	58.457	59.679	58.553	58.610
Visual Assessment (Scale 1- 10)	5	6.3	6.8	7.5	8.0
Gun					

(0.045) and Visual Assessment score (8.5), highlighting the accuracy of pixel-wise predictions.In the 'Tank' image evaluation, the Non-Subsampled Shearlet Transform (NSST) method exhibited the highest NMI score (0.877), indicating strong preservation of mutual information between the modalities. However, the Mosaic Approach scored highest in Visual Assessment (8.0), suggesting better performance in maintaining visual quality. For the 'Gun' image, the Mosaic method demonstrated relatively higher NMI scores (0.956) than all the other methods, also achieving a remarkable Visual Assessment score (8.0) and higher PSNR (58.610). The results emphasize the varying capabilities of each fusion method across different scenes, with each method displaying distinct strengths and weaknesses. This underscores the importance of selecting the appropriate fusion method based on specific image characteristics and intended application.

5.2.1. Visual Quality Comparison. Visual quality comparisons were made using the "Soldier with Jeep," "Man with Gun," and "Tank" datasets, each consisting of IR and VI source images and fused images generated by various fusion methods. The visual comparison results are depicted in Figures 3 to 5.

In this comparative analysis, we contrast our proposed approach against the aforementioned four techniques, evaluating their respective assessment indicators. Table 1 summarizes the optimal results for the five assessment metrics obtained by these methods, highlighting the effectiveness of our proposed technique in handling source images across three distinct datasets. Our approach consistently outperforms the others across multiple parameters, producing fused images with reduced artifacts and noise, while enhancing detail and texture characteristics. The visual appeal of the composite images is notably improved. In essence, our proposed method significantly enhances human visual perception by effectively extracting and preserving fine details and texture elements from the source images at an optimal scale. These assessment metrics further ensure

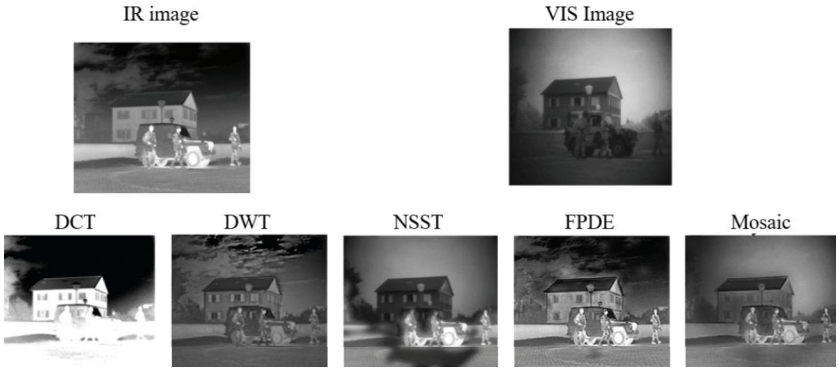

FIGURE 3. Solider with Jeep

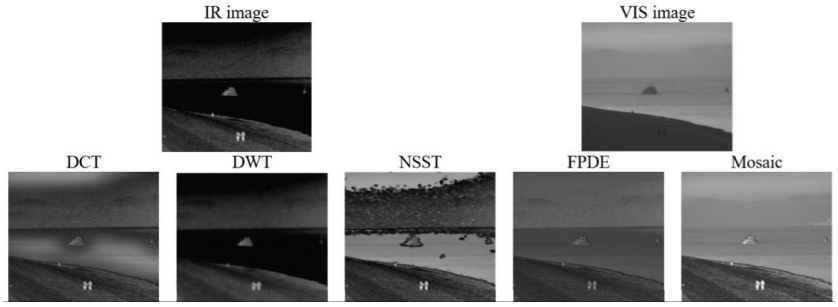

FIGURE 4. Tank

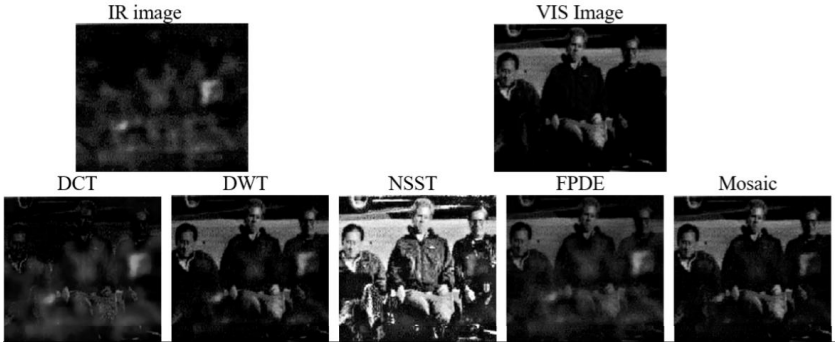

FIGURE 5. Man with gun

the suitability of our fusion results for detailed observation and detection purposes. Figure 3 vividly illustrates that our developed method produces merged

images with improved discernibility, contrast, and brightness. Our technique adeptly retains essential edges and crucial data while minimizing fusion artifacts and loss. Notably, our approach effectively eliminates block effects and artifacts while successfully restoring original textures. Compared to the other four methods, NSSCT images appear considerably blurry, lacking texture and sharpness. Similarly, some areas in the three combined DWT images exhibit a loss of texture and sharpness. FPDE-fused images, particularly in Figures 4 and 5, exhibit increased artificial noise and indistinct features, often appearing darker and lacking certain prominent elements. The Mosaic approach generates images with exaggerated infrared features and excessive brightness, leading to an unnatural viewing experience, as evident in Figure 3. Trustingly, our fusion approach expertly retains a greater amount of detail and texture from the source images compared to the other methods, resulting in clearer images that are well-suited for human vision.

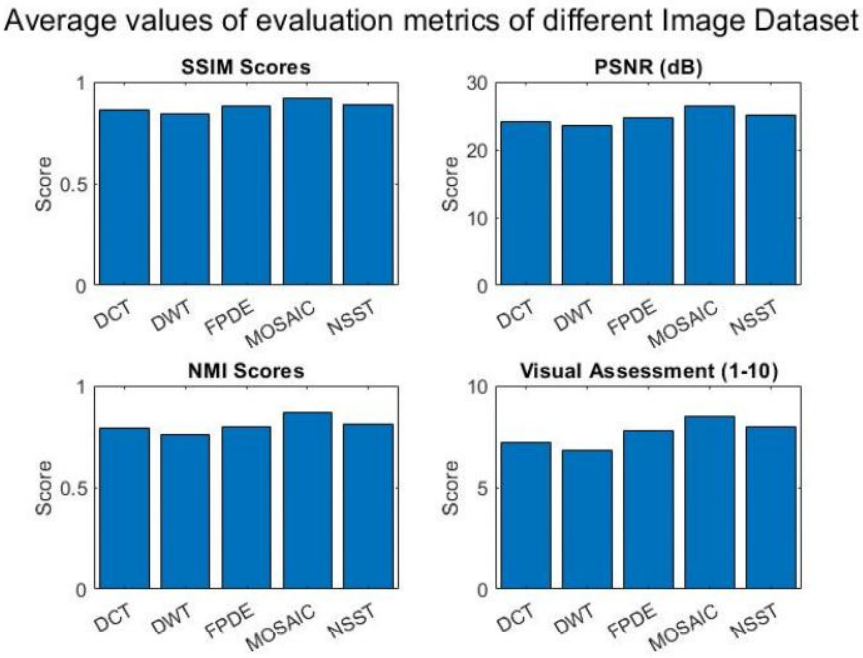

Figure 6. Average Evaluation Scores of Different Fusion Methods

Figure 6 illustrates the average evaluation scores of five different fusion methods—Mosaic, DCT, DWT, NSST, and FPDE—across fifteen pairs of images. The metrics evaluated in this chart include SSIM (Structural Similarity Index),

PSNR (Peak Signal-to-Noise Ratio) in decibels, NMI (Normalized Mutual Information), and Visual Assessment scores. Each method's performance is represented by these metrics, where higher scores indicate better results. This comparison provides insights into the varying performance levels of each method across multiple evaluation criteria.

6. Conclusion and Future scope

In conclusion, this study has presented a novel approach that departs from the conventional separation of feature extraction, fusion, and reconstruction stages in image processing. Instead, our proposed strategy offers an integrated and optimized framework where these three stages are considered simultaneously. This holistic approach not only enhances the interconnectedness of processes but also bolsters the overall method's stability. Moving forward, there are several promising avenues for future research in this domain. Firstly, the development of more advanced loss functions tailored to specific applications could further improve fusion quality and information transfer. Additionally, exploring the integration of emerging technologies like deep reinforcement learning or generative adversarial networks (GANs) into the proposed framework may yield even more impressive results. Furthermore, investigations into real-time implementation and scalability to handle large datasets or high-resolution images could expand the practical applicability of this approach. Lastly, there is potential for exploring applications beyond image fusion, such as in video processing, medical imaging, or remote sensing, where this integrated approach may offer substantial benefits. In conclusion, while this study has presented a comprehensive framework, the field of image processing continues to evolve, offering numerous exciting prospects for future research and development. These possibilities hold the potential to further advance the capabilities of image fusion techniques and their broader applications.

Conflicts of interest : The authors declare no conflict of interest.

References

1. Y. Chen, L. Cheng, H. Wu, F. Mo, and Z. Chen, *Infrared and visible image fusion based on iterative differential thermal information filter*, Optics and Lasers in Engineering **148** (2022), 106776.
2. G. Xiao, D.P. Bavirisetti, G. Liu and X. Zhang, *Image Fusion*, Springer, Singapore, 2020. Doi: 10.1007/978-981-15-4867-3
3. G.J. Trivedi and R. Sanghvi, *Novel Approach to Multi-Modal Image using Fusing Modified Convolutional Layers*, Journal of Innovative Image Processing **5(3)** (2023), 229–253. Doi: 10.36548/jiip.2023.3.002
4. G. Zhang, R. Nie, J. Cao, L. Chen and Y. Zhu, *FDGNet: A pair feature difference guided network for multimodal medical image fusion*, Biomedical Signal Processing and Control **81** (2023), 104545. Doi: 10.1016/j.bspc.2022.104545

5. H. Kaur, D. Koundal and V. Kadyan, *Image Fusion Techniques: A Survey*, Archives of Computational Methods in Engineering **28(7)** (2021), 4425–4447. Doi: 10.1007/s11831-021-09540-7
6. B. Meher, S. Agrawal, R. Panda, L. Dora, and A. Abraham, *Visible and infrared image fusion using an efficient adaptive transition region extraction technique*, Engineering Science and Technology, an International Journal **29** (2022), 101037. Doi: 10.1016/j.jestch.2021.06.017.
7. S. Hao, T. He, B. An, X. Ma, H. Wen, and F. WangS. Hao, T. He, B. An, X. Ma, H. Wen, and F. Wang, *VDFEFuse: A novel fusion approach to infrared and visible images*, Infrared Physics & Technology **121** (2022), 2713–22. Doi: 10.1016/j.infrared.2022.104048
8. V. Kamarthi, D. Satyanarayana, and G.P.M. Ninjappa, *Multimodal Medical Image Fusion Based on Intuitionistic Fuzzy Sets andWeighted Activity Measure in NSST Domain*, Current Signal Transduction Therapy **17(2)** (2022). Doi: 10.2174/1574362417666220405151738
9. W. Tang, F. He, Y. Liu, and Y. Duan, *MATR: Multimodal Medical Image Fusion via Multiscale Adaptive Transformer*, IEEE Transactions on Image Processing (2022), 5134–5149. Doi: 10.1109/tip.2022.319328
10. Y. Li and S. Jiang, *Multi-Focus Image Fusion Using Geometric Algebra Based Discrete Fourier Transform*, IEEE Access **8** (2020), 60019-60028. Doi: 10.1109/ACCESS.2020.2981814
11. H. Jiang, P. Maharjan, Z. Li, and G. York, *DCT-Based Residual Network for NIR Image Colorization*, IEEE International Conference on Image Processing (ICIP) 2022. Doi: 10.1109/icip46576.2022.9897373
12. V.P.S. Naidu, and B. Elias, *A novel image fusion technique using DCT-based Laplacian pyramid*, International Journal of Inventive Engineering and Sciences (IJIES) (2013), 2319-9598.
13. G.J. Trivedi and R. Sanghvi, *Medical Image Fusion Using CNN with Automated Pooling*, Indian Journal Of Science And Technology **15(42)** (2023), 2267–2274. Doi: 10.17485/ijst/v15i42.1812
14. S. Liu, L. Qu, Q. Qiao, M. Wang and Z. Song, *Wavelet-based self-supervised learning for multi-scene image fusion*, Neural Computing and Applications **34(18)** (2022), 15689–15704. Doi: 10.1007/s00521-022-07242-0
15. W. Kong and Y. Lei, *A Technique for image fusion between gray-scale visual light and infrared images based on NSST and improved RF*, Optik **124(23)** (2013), 6423–6431. Doi: 10.1016/j.ijleo.2013.05.038
16. S. Budhiraja, S. Agrawal, and B. S. Sohi *Performance Analysis of Multi-scale Transforms for Saliency-Based Infrared and Visible Image Fusion*, Proceedings of the International Conference on Data Science and Applications, 2021. Doi: 10.1007/978-981-16-5120-5_60.
17. G.J. Trivedi and R. Sanghvi, *Optimizing Image Fusion Using Modified Principal Component Analysis Algorithm and Adaptive Weighting Scheme*, International Journal of Advanced Networking and Applications **15(1)** (2023), 5769-5774. Doi: 10.35444/IJANA.2023.15103
18. N. You, L. Han, D. Zhu and W. Song, *Research on Image Denoising in Edge Detection Based on Wavelet Transform*, Applied Sciences **13(3)** (2023), 1837. Doi: 10.3390/app13031837
19. M.M. Iqbal Ch, M.M. Riaz, N. Iltaf, A. Ghafoor and S.S. Ali M.M. Iqbal Ch, M.M. Riaz, N. Iltaf, A. Ghafoor and S.S. Ali, *A multi-focus image fusion using high-level DWT components and guided filter*, Multimedia Tools and Applications **79** (2020), 12817–12828. Doi: 10.1016/j.cviu.2022.103619
20. V. Rajinikanth, S.C. Satapathy, N. Dey and R. Vijayarajan, *DWT-PCA Image Fusion Technique to Improve Segmentation Accuracy in Brain Tumor Analysis*, Lecture Notes in Electrical Engineering (2018), 453–62. Doi: 10.1007/978-981-10-7329-8_46

21. S. Singh, H. Singh, A. Gehlot, J. kaur, and Gagandeep, *IR and visible image fusion using DWT and bilateral filter*, Microsystem Technologies (2022). Doi: 10.1007/s00542-022-05315-7
22. G.J. Trivedi and R. Sanghvi, *FUSESHARP: A Multi-Image Focus Fusion method using Discrete Wavelet Transform and Unsharp Masking*, Appl. Math. & Informatics **41(5)** (2023), 1115-1128. https://doi.org/10.14317/jami.2023.1115
23. G.J. Trivedi and R. Sanghvi, *A New Approach For Multimodal Medical Image Fusion Using PDE-Based Technique*, Suranaree J. Sci. Technol. **30(4)** (2023), 030132(1-7). https://doi.org/10.55766/sujst-2023-04-e0843
24. G.T. Vasu, and P. Palanisamy, *Gradient-based multi-focus image fusion using foreground and background pattern recognition with weighted anisotropic diffusion filter*, Signal, Image, and Video Processing (2023). Doi: 10.1007/s11760-022-02470-2
25. D.P. Bavirisetti and R. Dhuli, *Fusion of Infrared and Visible Sensor Images Based on Anisotropic Diffusion and Karhunen-Loeve Transform*, IEEE Sensors Journal **16(01)** (2016). Doi: 10.1109/jsen.2015.2478655
26. G.J. Trivedi, R. Sanghvi, V. Shah and J. Sharma, *On solution of non-instantaneous impulsive Hilfer fractional integro-differential evolution system*, Mathematica Applicanda **51(1)** (2023), 3-20. Doi: 10.14708/ma.v50i2.7168.
27. T. Sharma, S. Pathak, G.J. Trivedi and R. Sanghvi, *AFlow Modelling in Porous Medium Applying Numerical Techniques: A Comparative Analysis*, Recent Research Reviews Journal **2(2)** (2023), 288-304. https://doi.org/10.36548/rrrj.2023.2.004
28. M. Wang and X. Shang, *IA Fast Image Fusion With Discrete Cosine Transform*, IEEE Signal Processing Letters **27** (2023), 990-994. Doi: 10.1109/LSP.2020.2999788
29. D. Jiang and J. Kim, *Image Retrieval Method Based on Image Feature Fusion and Discrete Cosine Transform*, Applied Sciences **11(12)** (2021), 5701. Doi: 10.3390/app11125701
30. E. Gul and A.N. Toprak, *Contourlet and discrete cosine transform based quality guaranteed robust image watermarking method using artificial bee colony algorithm*, Expert Systems with Applications **212** (2023), 118730. Doi: 10.1016/j.eswa.2022.118730
31. A. Toet, *The TNO Multiband Image Data Collection*, Data in Brief **15** (2017), 249–251. Doi: 10.1016/j.dib.2017.09.038
32. X. Zhang, P. Ye, and G. Xiao *VIFB: A Visible and Infrared Image Fusion Benchmark*, Dataset taken from website GitHub Cited on 2023. https://github.com/xingchenzhang/VIFB
33. D.P. Bavirisetti, G. Xiao, and G. Liu *Multi-sensor image fusion based on fourth-order partial differential equations*, 2017 20th International Conference on Information Fusion (Fusion) 2017. Doi: 10.23919/icif.2017.8009719
34. R.C. Gonzalez, R.E. Woods, and S.L. Eddins *Digital image processing using MATLAB*, Knoxville, Gatesmark Publishing, 2020.

Gargi J. Trivedi is pursing her Ph.D.(Mathematics) from Charutar Vidya Mandal University, Vallabh Vidyanagar, Gujarat, India. She is working as a Teaching Assistant at the Department of Applied Mathematics, Faculty of Technology and Engineering, The Maharaja Sayajirao University of Baroda, Vadodara. Her main academic interests are Dynamical systems, Image processing, Machine learning, and image analysis.

Department of Applied Science & Humanities, G. H. Patel College of Engineering & Technology, Charutar Vidya Mandal University, Vallabh Vidhyanagar-388120, India.
e-mail: gargi1488@gmail.com

Rajesh C. Sanghvi holds a Ph.D. degree from Charotar University of Science and Technology. He earned his B.Sc. and M.Sc. degrees in Mathematics from Sardar Patel University, V V Nagar, Gujarat. He also holds a B.Sc. degree in Computer Science from Sardar Patel University, V V Nagar, Gujarat. Since 1998, he has worked as an assistant professor at G.

H. Patel College of Engineering and Technology, V. V. Nagar. His scientific interests include image processing, machine learning, fuzzy logic, optimization, and evolutionary algorithms.

Department of Applied Science & Humanities, G. H. Patel College of Engineering & Technology, Charutar Vidya Mandal University, Vallabh Vidhyanagar-388120, India.
e-mail: rajeshsanghvi@gcet.ac.in