A Survey on Multivariate Methods for Multimodal Fusion of CT/MRI Images

Riya Gupta, **Dr. C. S. Raghuvanshi**

Abstract— **In many different application domains, image fusion is widely acknowledged as a helpful technique for enhancing overall system performance. These application fields include, but are not limited to, remote sensing, traffic control, machine learning, health care applications, combat surveillance, detection of disguised munitions, and identification of faults in non-destructive testing. The amount of scholarly literature on the topic of medical image fusion has significantly increased during the last several years. Image fusion has grown in importance as a component of the commonly used image processing applications because to the large variety of capture devices that are currently available. Image fusion is the process of matching relevant information from several sensors using various mathematical models to create a single composite image. Combining data from many sensors, multiple viewpoints, and various temporal dimensions into a single image is known as image fusion. This preserves the integrity of important characteristics while improving the quality of the image. Robot vision, aerial, satellite, and medical imaging, as well as robot and vehicle navigation, are just a few of the many applications that heavily depend on this stage. This study examines many cutting-edge picture fusion methods at various levels, along with the benefits and drawbacks of each. It also explores several transform-based and spatial approaches with quality measures and their applicability across various industries. Finally, this study's results have shown several potential future application avenues, such as picture fusion applications.**

Index Terms—**Medical CT and MR image fusion, adaptive structure decomposition (ASD), MI, SSIM, EN.**

I. INTRODUCTION

Due in large part to the expansion and variety of image collecting techniques, image fusion is becoming more necessary in contemporary computer imaging systems. Image fusion is the process of using mathematical methods to combine important data from several sensors to create a complete composite image. The purpose of this composite picture is to make it more useful and useful for computer vision applications as well as human operators. Thus, the purpose of this chapter is to go over the main theories, classifications, kinds, and uses that support the idea of picture fusion.

The five basic human senses—skin, tongue, nose, ears, and eyes—collect information mostly on their own. The human brain instinctively compiles this data into a clear picture of the surroundings, which facilitates job performance and governance [1]. Such a circumstance is generally referred to as data fusion. Similarly, an independent picture of the target scene does not reliably give sufficient details about the desired area in digital image processing. To do this, two or more images of the same object must be gathered. Combining sensors of the same or distinct modalities with varying optical configurations, focal lengths, and exposure periods may result in the acquisition of thus many pictures. However, due to the limiting depth of field of each sensor, the focus on all target items often varies [2-5]. Due to this limited capacity to perceive extra aspects of the target that are present in a picture, the human brain is unable to more accurately construct and explain the composite image of the intended location. Therefore, to provide a more realistic portrayal of the targeted portion, these several obtained photos with limited and distinct information should be integrated into a single image rather than being displayed as separate source images. The use of picture fusion is required in these situations [6-7].

Images from many sensors are combined in a new field called image fusion (IF) to provide an informative image that might be utilized for decision-making [7]. The visual and analytical quality of a picture can be improved by combining many photographs. By eliminating all pertinent information from images and preventing errors in the final image, effective image fusion can retain crucial information. Following fusion, the merged picture is more suited for both human and machine perception. The first step of fusion is image registration, which involves comparing the source and reference pictures. With this kind of mapping, a comparable image is matched for further analysis based on trustworthy attributes. IF and IR are recognized as crucial instruments for the generation of important information in a wide range of fields [8]. According to the literature, the number of scholarly papers has grown dramatically since 2011, peaking at 21,672 in 2019. The growing need for low-cost, high-performance image fusion algorithms may be the cause of this rapidly expanding trend. There are several strategies to increase the efficacy of picture fusion, including recently reported techniques like sparse representation and multi-scale decomposition. Due to variations in linked images in different applications, an efficient fusion approach is necessary. For example, more and more satellites are being deployed in the field of remote sensing to take aerial photos with various temporal, spatial, and spectral resolutions. The IF is essentially a collection of image data gathered using various imaging parameters, such as spectral response, camera location, dynamic range, aperture settings, or the use of polarization filters. The pertinent information from various photos is extracted using appropriate image fusion algorithms, and this information may then be used for quality analysis, autonomous driving, traffic management, or reconnaissance.

Clinicians have been able to learn about the soft tissue, structural features, and other aspects of the human body through imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). Different sensors gather different image data for the same part, and different imaging technologies preserve distinct properties. Enhancing contrast, fusion quality, and overall perception is the aim of the fusion. The following requirements must be fulfilled by the fusion result: (a) the fused image must preserve all of the information from the original images; (b) it must not create any artificial information, like artifacts; and (c) bad states, like noise and mis-registration, must be avoided [9].

The transform domain and the spatial domain are separated in conventional medical picture fusion methods. The first studies focused on medical picture fusion methods based on spatial domains. Principal analysis and HIS are two common methods. The integrated pictures created using spatial domain technology, on the other hand, show both spectral and spatial distortion [10]. Researchers are working on the transform domain to improve fusion effects. It performs reconstruction techniques after transforming the original picture into the frequency domain or other domains to merge them. The four levels that comprise the fusion process are the signal, feature, symbol, and pixel level. These days, pixel-level transformations including pyramid transform, contour transformation, and discrete wavelet transform are commonly employed.

The transform domain-based method produces noise during fusion processing, but it has the advantages of good structure and minimal distortion. Consequently, denoising also hinders image fusion [11]. The suggested fusion technique almost never uses the spatial domain alone, as is evident from the papers released in the past two years. Nonetheless, a lot of new approaches combine spatial domain and transform domain approaches, as PCA-DWT [12,]. A deep learning-based method for merging medical pictures emerged as a result of the deep learning boom in 2017. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), U-Net networks, GANs, and other deep learning models have been used in medical picture registration and segmentation in recent years. However, medical image fusion has only employed CNNs and U-Net networks. A convolutional neural network, a type of neural network used for image processing, consists of three layers: a convolutional layer, a pooling layer, and a fully connected layer. Deep learning frameworks like Tensorflow, MatConvNet, and Caffe are used in medical picture fusion. The U-Net network is currently trained using the Pytorch deep learning framework.

IMAGE FUSION LEVELS

There are multiple levels of information representation where the precise fusion process can occur. Based on the level of abstraction, three distinct levels, namely pixel, feature, and decision levels, can be used to achieve image fusion. The framework for the IF process is depicted in Figure 1. Firstly, multiple images of external scenes are captured utilizing a single sensor or multiple sensors and noise/artifacts introduced during the acquisition process are removed. Then, to facilitate with the image fusion process, image registration is a procedure of mapping the source images with the aid of a reference image to match the related images pursuant to specific features. [4]

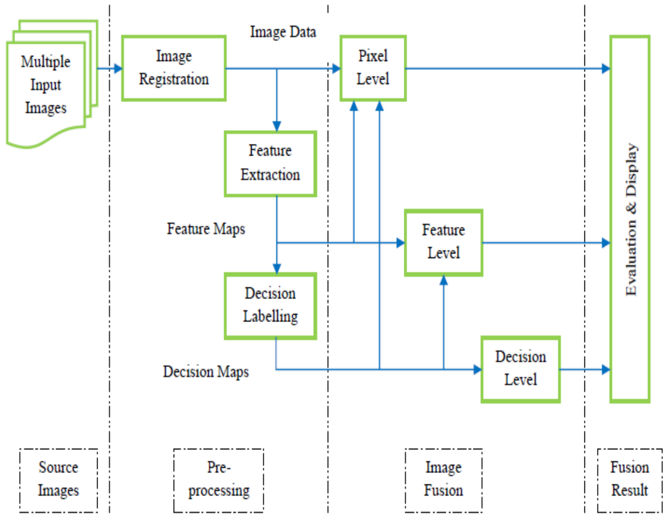

Figure 1: Generic block diagram representing the process of image fusion

PIXEL-LEVEL IMAGE FUSION (PLIF)

"Image fusion at the pixel level," or fusion at the lowest processing level, is the phrase used to describe the combining of empirically established parameters. An ensemble of pixels from each of the source pictures is utilized to create each individual pixel in the fused image. Pixel-by-pixel fusion requires sub-pixel-by-pixel perfect registration of each source picture. In order to ensure that the data from each source belongs to the same physical components in the actual world, image registration is an essential pre-processing step for pixel-fusion algorithms, as was previously mentioned [13].

FEATURE-LEVEL IMAGE FUSION (FLIF)

Using this technique, pictures are divided into many areas from which different characteristics (such edges, textures, and minute details) are extracted. As a result, the fusion depends on characteristics that are extracted that are more informative and characterize the content of an image, instead of employing pixels that are more confident. The relevant data is then combined into a single feature vector using neural networks, Support Vector Machines (SVM), and clustering techniques. Machine learning, region-based, and content-based similarity matching are the three other subcategories of the FLIF approach. The most extensively studied techniques, region-based approaches, offer the chance to adjust the segmented regions of interest (ROI) to semantic standards that are helpful in mitigating spatial distortions. Three alternative methods are accustomed to performing region-based image fusion: (1) the region partition: employed in medical image fusion applications to produce superior fusion outcomes. (2) statistical and estimation-based algorithms: primarily appropriate for multi-focus image fusion applications and (3) focus region detection and saliency map-based algorithms: is versatile and has a wide range of uses [14].

DECISION-LEVEL IMAGE FUSION (DLIF)

This is the topmost level for fusing images. The input source images are first processed independently for feature extraction and identification before being classed using local classifiers. The features are then integrated by using fuzzy logic, evidence-based reasoning, statistics, voting, heuristics, and artificial intelligence. In a newest study, decision-level fusion algorithms were divided into two categories: (1) soft fusion, where the classifier provides a number to indicate how confident it is in the judgement; and (2) hard fusion, where logical information (such as class membership) values are combined. The paramount obstacle at this level is the exigency for prior information, which is challenging to acquire because the environment and the target's features are changing. As a result, algorithms are typically difficult to implement in this highest IF level [15].

Table 1: Brief comparison of different levels of image fusion

Characteristics	Pixel Level	Feature Level	Decision Level
Nature of sensory data	Multiple images	Features extracted from images	Symbol to represent a Decision
Level of data interpretation	Low	Medium	High
Registration level: Spatial Temporal	High Medium	Medium Medium	Low Low
Registration method: Spatial Temporal	Shared optics or coalignment of sensors Synchronization	Geometrical modifications Synchronization	If applicable, Symbol's spatial characteristics Symbol's Temporal characteristics
Fusion method	Image estimation or pixel attribute combination	Geometrical or temporal correspondence, and feature attribute combination	Logical and statistical inference
Advantages	• Executable in both the spatial and transform domains • More computationally efficient • Preserves information content more accurately • Easy to implement	• Improve image quality (information content and image contrast) • Extract newer features • Less responsive to noise	• Strongly localized features are better handled i.e., allows explicit handling of localization uncertainty due to mis-registration error • Need prior knowledge • Reduced processing Requirement
Disadvantages	• High registration requirement • Sensitive to noise	• Difficult to apply due to potential heterogeneity in Retrieved characteristics from several modalities • Inter-image variability can lead to poor fusion performance	• Algorithms are complex to implement • Require good set of features to increase accuracy and reliability of fusion process

Table 1 provides a brief comparison amongst various levels at which image fusion is carried out.

II. FUSION METHODS

The deep learning-based fusion strategy, transform domain-based fusion method, and spatial domain-based fusion approach are all covered in this chapter. As seen in Figure 2,

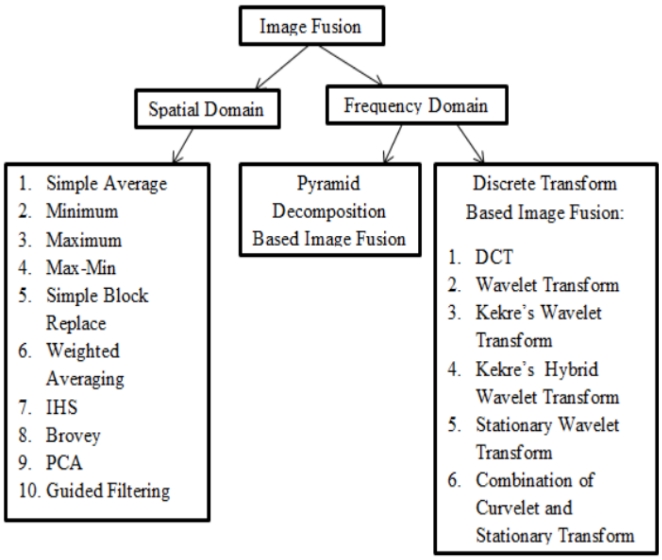

Figure 2: Image Fusion Techniques

Transform Domain

In the transform area, which has been the subject of significant research in recent years, the multiscale transform (MST) theory is the foundation for most medical picture fusion algorithms. Decomposition, fusion, and reconstruction are the three steps of the MST-based fusion technique. To acquire the low-frequency and high-frequency coefficients, the transform domain-based medical image fusion method transforms the source image from the time domain to the frequency domain or other domains. Discrete wavelet, nonsubsampled shearlet, and nonsubsampled contourlet transforms are three of the most often utilized transformations in medical image fusion systems [17].

Fusion Based on Nonsubsampled Contourlet Transform (NSCT): Do et al. [17] introduced the multiscale contourlet transform. It offers advantages in smoothness processing and is suitable for generating multidirectional and multiresolution situations. However, because it lacks translation invariance and is prone to creating pseudo-Gibbs phenomenon (artefact) around the singular point of the reconstructed picture, which causes image distortion, it is not the ideal option for image fusion. For this reason, several scholars have carried out more in-depth study. Cunha et al. [18] introduced a multiscale decomposition technique called nonsubsampled contourlet transform, which is an enhancement of contourlet transform. Translation invariance and spectral aliasing avoidance are two features of NSCT. Because the original image's structural information is retained throughout the deconstruction and reconstruction processes, more direction information may be recovered. In recent years, one of the most popular techniques in the transform field for medical picture fusion has been the nonsubsampled contourlet transform. The NSPFB and NSDFB filters calculate the multiscale and multidirection decompositions to produce subband pictures with a range of scales and directions after NSCT first breaks down the source image to create the coarse and detailed layers.

Fusion Method Based on Discrete Wavelet Transform (DWT): The discrete wavelet transform's capacity to generate a stable output from a variety of input frequency signals and its good placement in the time and frequency domains maintain the particular information in the image. As a result, when multimodal medical image fusion algorithms were first being developed, the discrete wavelet transform (DWT) was the technique that was most frequently utilized. The discrete wavelet transform is a quantitative and visual fusion technique that overcomes the drawbacks of principle component analysis. DWT-based fusion algorithms are most frequently used for MRI and PET image fusion, while they may be used for a variety of applications [21, 22]. By recovering the intensity component from the PET image, the IHS transform lessens color distortion and preserves more anatomical details. After the source image has been preprocessed and improved, the intensity component of the PET image is extracted using the IHS transform, which reduces color distortion and maintains more anatomical information. High- and low-frequency subbands are obtained by applying the DWT transform to the intensity components of MRI and PET. The inverse DWT transform is employed to create the fused picture after the high- and low-frequency subbands are fused using different fusion criteria [23].

Image Fusion Based on Deep Learning

Deep learning is still in its infancy in the realm of medical image fusion research. Krizhevsky et al. [24] suggested the convolutional neural network (CNN) as a popular deep learning model. Deep learning is commonly utilized for medical picture segmentation and registration, as opposed to medical image fusion. The fusion rules and activity level measurement (feature extraction) are flaws in medical image fusion systems based on spatial domain and transform domain that need artificial design, and their link is very weak. In order to overcome the aforementioned issues, Liu et al. employed CNN to image fusion for the first time in 2017. Their results in the transform and spatial domains were promising. In medical image segmentation, the U-Net network model is widely used. Despite being a relatively new issue, medical image fusion research technology has progressed from 2D to 3D [25] and has shown encouraging results in the segmentation of medical images.

An artificial neural network with supervised learning that is trainable and multistage feedforward is called a CNN. Convolution is a complex procedure. The first parameter in a convolutional network is frequently referred to as an input, the second as a kernel function, and the end product as a feature map. Three key architectural concepts in CNN are sparse representations, also known as sparse weights, parameter sharing, and isomorphic representations. Conventional neural networks manage link interactions by matrix multiplication. Since every input unit has an output unit, a sizable amount of storage is needed. The neurons are only linked to a few neurons close to the previous stage because of the convolutional network's sparse representation, and the local convolution operation is used, which lowers storage requirements and improves processing performance. The nonuniqueness of weights in traditional networks is eliminated by CNN's parameter sharing. Because its weights remain constant, the CNN stage is simpler to store than earlier stages. Automatic encoders are fully linked in the traditional sense of the word. Despite U-Net's local connection structure, the source picture and vector output are not always spatially aligned. The visual impact of the fusion image is increased when the source image and vector output are aligned in space. The U-Net is a complete convolution network with contraction and extension pathways [26]. In-depth learning requires a large number of samples for training, however U-Net, which is built on a full convolutional neural network, can train a small number of instances with data improvement. This advantage only addresses the problem of a limited sample size of medical image data.

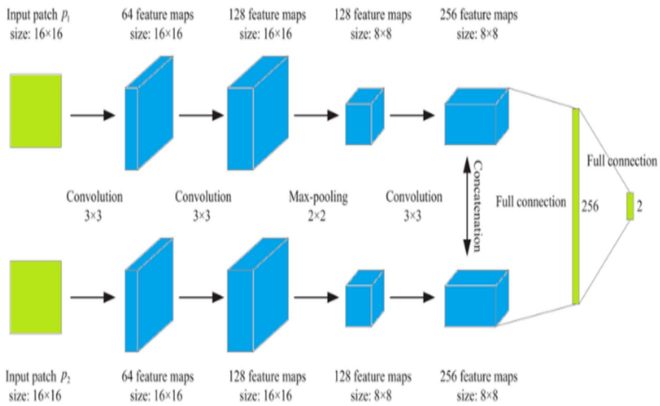

Figure 2: Schematic diagram based on CNN fusion algorithm

Convolutional Neural Networks (CNN) for Image Fusion. The fusion technique outlined in [27] is inappropriate for medical picture fusion as the intensity of medical pictures vary at the same location. Yu et al. proposed the first CNN-based technique for merging medical pictures. This technique generates a weight map using the Siamese network. The CNN model compares patch similarity using three models, including the Siamese network. Since the two weight branches of the original picture are similar, the techniques for feature extraction and activity level measurement are the same. This has several advantages over pseudosiamese and 2-channel models, and the siamese model's popularity in fusion applications is also influenced by how simple it is to train. Once the weight map is obtained, the pyramid transform is used for multiscale decomposition, and the Gaussian pyramid decomposition is used to make the fusion process more akin to human visual perception. Moreover, the decomposed coefficients are adaptively modified using the localized similarity-based fusion approach. The method offers an enhanced fusion approach by combining the CNN model with the traditional pyramid-based and similarity-based fusion methods. The algorithm is displayed in Figure 2.

This is primarily due to three factors: (a) the requirement for a substantial quantity of annotated training set data; (b) the drawn-out training procedure; and (c) the complexity of the convergence problem and the requirement for frequent overfitting corrections. Liang et al. [28] claim that the MCFNet network approach refers to different kinds of medical image histograms and transforms 1.2 million natural photographs in ILSVRC 2013 ImageNet into medical images with the same texture distribution or intensity as training data sets. Medical image data sets and reconstructed data sets are quite comparable. 256 256 photographs are randomly chosen from the modified images and trained using medical images in order to minimize overfitting. Future studies will keep concentrating on improving this strategy's loss function.

Spatial Domain

In the early stages of research, medical image fusion technology based on spatial domain is a hot issue. It has a straightforward fusion method, and the merged picture may be produced by simply applying the fusion rules to the pixels of the original image. The Brovey method, the maximum selection method, the minimum selection method, the saturation method of hue intensity, the high-pass filtering method, the principal component analysis method, and the average method are examples of spatial domain fusion approaches. The quantity of research in the medical image fusion method's spatial domain has significantly decreased in recent years due to spectral and spatial distortion in the fused picture of the spatial domain. In order to create novel research techniques, researchers frequently use spatial domain fusion processes as part of the transformation domain [29].

A quick introduction to the highly useful IHS technique will be given here.

Munsell, an American scientist, proposed the IHS model, which explains the characteristics of the human visual system. It has two characteristics:

(1) The hue and saturation components have a high correlation with people's perception of color, but the intensity component has nothing to do with the image's color information. Consequently, this method is frequently used by researchers to address the color issue in image fusion, particularly when merging PET/SPECT pictures with color information. Chen [7] proposed a new technique for combining MRI and PET by combining the IHS model with the Log-Gabor transform and utilizing IHS to dissect the PET picture. The three basic characteristics of hue (H), saturation (S), and intensity (I) are obtained using this approach (I). The Log-Gabor transform, which is made up of the logarithmic transformation of the Gabor filter, is used to break down the intensity components of MRI and PET images into high-frequency and low-frequency subbands. A novel approach based on two-level fusion of visibility measurement and a weighted average rule is used for the fusing of low-frequency subbands, while a unique technique is used for the fusion of high-frequency subbands. A fused image is created by inversely HISing the original hue and saturation components as well as the inverse Log-Gabor transformed component. It may successfully lessen color distortion while preserving the original image's details and structures. In terms of visual perception, this method performs better than the previous IHS+FT method. Haddadpour et al. [30] suggested a new fusion technique that combines the IHS with the two-dimensional Hilbert transform. When merging high- and low-frequency subbands, the technique introduces the concept of BEMD. Bidirectional empirical mode decomposition (BEMD) is a kind of empirical mode decomposition that is extended by empirical mode decomposition. Its envelope surface makes it widely employed in biomedicine. The method achieves better contrast and color intensity than the PCA and wavelet algorithms without any noticeable distortion. Information entropy (EN) is rather low, which is one drawback. Figure 3 illustrates the IHS domain fusion method, which integrates MRI and PET data.

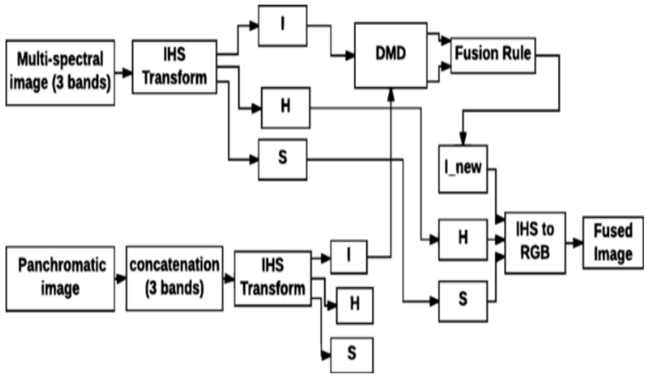

Figure 3: Framework diagram based on the IHS domain fusion method

III. The Multimodal Fusion Approach

The brain's soft tissue anatomy may be determined via magnetic resonance imaging, or MRI, but its function cannot. The image is incredibly clear and artifact-free due to the abundance of protons in the neurological system, fat, soft tissue, and articular cartilage defects. Its great spatial resolution, lack of radiation harm to the human body, and plenty of information make it a valuable tool for clinical diagnosis. The MRI image of the bone appears fuzzy due to the extremely low proton density in the bone. The CT image is referred to as computed tomography imaging. An X-ray is used to scan the human body. Because bone tissue absorbs density at a higher rate than soft tissue, bone tissue CT images are remarkably clear [31]. CT pictures show less cartilage information, or anatomical information, because X-rays have a low absorption rate and poor permeability in soft tissue. Emission of Single-Photons Computed Tomography, or SPECT, is a type of functional imaging that displays blood flow through veins and arteries as well as the metabolism of human tissues and organs. It provides information on both benign and malignant tumors and is frequently used to diagnose a range of tumor diseases. Conversely, SPECT has weak positioning ability and limited resolution. Positron Emission Tomography, or PET imaging, provides reliable information about blood flow and can be used to pinpoint the patient's lesion location. When positrons and electrons in the tissue collide, photons are produced. PET is used to measure the number of photons in the brain, producing a color image of brain function information that can be used to detect tumors. Its sensitivity is high, but it is challenging to get precise information about the position of the brain structure; the lack of soft tissue and bone boundary resolution results in very low spatial resolution and a high likelihood of spatial distortion.

Examples of imaging method fusions used in medical picture fusion include MRI and PET, MRI and CT, MRI and SPECT, CT and PET, CT and SPECT, SPECT and PET, and MRI-T1 and MRI-T2. The diagnosis of liver metastases, Alzheimer's disease, and brain tumors can be made with MRI/PET fusion images; tinnitus patients can benefit from MRI/SPECT fusion images for lesion localization and vertebral bone metastases; lung cancer can be better diagnosed with CT/PET fusion image energy; abdominal research can be done with SPECT/PET; and vascular blood flow can be diagnosed with ultrasound/MRI. A few hot fusion techniques will be the main topic of the sections that follow.

MRI and CT Fusion: To make up for the absence of information in a single imaging, the benefits of clear soft tissue information in MRI images and clear bone information in CT images are merged. A guided filtering-based MRI and CT fusion approach (GF) was proposed by Na et al. [32]. The edge degree and clarity problems are fixed in the fused image by extracting feature information while preserving the edge information of the original image. [33] suggests a method for fusing Frei-Chen operators that is based on the NSST domain. Visual inspection of the fusion products clearly improves their contrast and structural similarity. Additionally, quantitative evaluation is an improvement above existing approaches. It is challenging to choose the membership degree based on the intuitionistic fuzzy inference fusion process. The fuzzy-PCNN rule, which was also proposed by Mishra et al., uses multiple membership functions to create fuzzy membership from specific parts of high-frequency coefficients. The high-frequency coefficient is fused using the L2 norm set operation, and the fused image's SF, EN, and SD have higher values. Following the fusing of MRI/CT images in the NSST domain, Singh et al. introduced a novel fusion technique that favorably affects visual quality and quantitative indicators by utilizing the ripple transform and NSST transform cascade. Other methods for MRI/CT image fusion include contour transforms based on non-subsampling, multiscale, and multiresolution techniques.

IV. Main Applications In Diverse Domains

Image Fusion is widely used in various application areas. The technical areas involve analysis of images and videos which falls under the fusion application category. Image fusion aims to reduce the corpus of data, with decreased redundancy and uncertainty. Image fusion increases confidence levels with superior visual outputs and with even more reinforced conclusions for further processing. Image fusion is more useful for further machine processing [8]. The application areas are expanding with the introduction of new acquisition devices for scientific research. Image fusion has extensively been used in various fields such as computer graphics, robotics, situation awareness, Surveillance, target tracking, intelligence gathering, person authentication, remote sensing and satellite imaging. In addition image fusion has also been incorporated in biometrics based applications as, face recognition, biometric audio-visual speech synchrony, speech recognition, and other application areas as video indexing, multi-sensor fusion, information retrieval, data mining and machine learning [22].

Remote Sensing Applications: In addition to the above mentioned modalities, it incorporates a number of IF techniques that have shown promise in IF applications. These methods include light and range detection, the intermediate resolution image spectroradiometer, and the Synthetic Aperture Radar. Byun et al. [27] suggested the area-based IF method for combining panchromatic, multispectral, and synthetic aperture radar images. Landsat and intermediate resolution imaging spectroradiometer data are combined to generate synthetic Landsat imagery using a high spatial approach and temporal data fusion. Furthermore, a mixture of spectrum information has recently been used to study the synthesis of air-bone hyper-spectral and Light Detection and Ranging (LiDAR) data. Earth imaging satellites such as Quickbird, Worldview-2, and IKONOS have contributed several datasets for pansharpening applications.

Applications in the Medical Domain: A collection of brain pictures from registered CT and MRI scans has been made available by Harvard Medical School. Figure 4 illustrates the use of IF in medical diagnosis by combining CT and MRI images. The CT scan captures bone structures with great spatial resolutions, whereas the MRI captures soft tissue structures such as the brain, eyes, and heart [28]. The accuracy and therapeutic value of CT and MRI scans can be improved by combining them with IF techniques. The most difficult task in this field is also completed like follows:

(1) There aren't any IF techniques for medical emergencies.
(2) Estimation of objective image fusion effectiveness
(3) Mis-registration

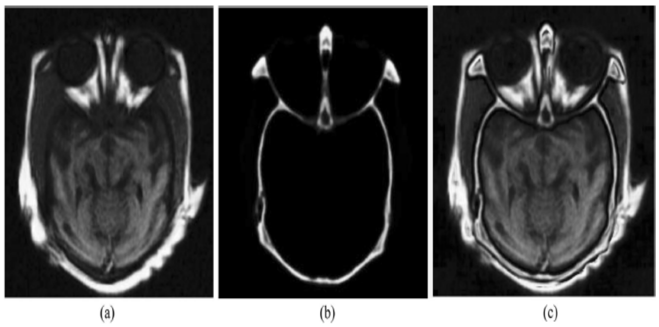

Figure 4: Examples of IF in medical diagnosis domain. **a)** MRI **b)** CT **c)** Fused image

The radiological imaging techniques have refined multi-folds in recent past. The same organ, tissue or tumor is captured using different radiological devices, under varied conditions. Availability of different modalities, helps to capture complimentary information, from organs, leading to disclosure of additional critical content to fill the missing gaps in clinical diagnostic analysis for superior diagnosis. It is very important to extract all the possible information. This acquisition of clinical content from multiple perspectives and medical modalities adds up to the bulk of clinical data. Medical image fusion technique unites multiple modalities into a single composite image. It brings about patient data which may be acquired from single or multiple modalities in order to device a single fused image, rich in clinical content, enhanced in texture, features and anatomical details. The purpose aims to reduce the amount of data to be processed, preserving all possible diagnostic content without loss or artifacts. This is turn reduces redundancy and uncertainty in medical analysis and direct towards higher degree of diagnostic confidence to keep the quality intact. The fused image must be superior in comparison to the original; else the fusion shall be deemed to be irrelevant with respect to domain performance or diagnosis. Image fusion is a complex mechanism and becomes even more critical using neurological images. Pictorially the fusion components are described in figure 5.

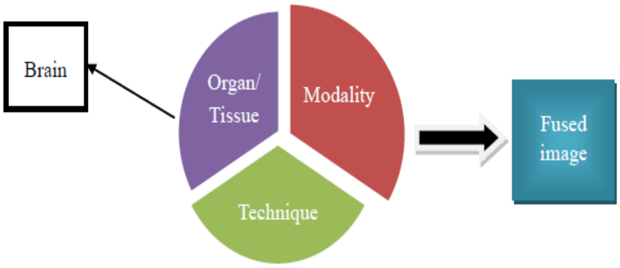

Figure 5: Image Fusion Components

Surveillance Domain Applications: Figure 6 illustrates the fusion of visible and infrared pictures as an illustration of IF in the surveillance sector. It can "see in the dark" even when there is no light because of its high temperature, which makes it sensitive to things. The limited spatial resolution of infrared pictures can be solved by combining visible and infrared images using a fusion approach. Additionally, merging visible and infrared photos has solved a difficulty in military reconnaissance, picture dehazing, and facial identification. The primary obstacles in this field are as follows:

(1) Computing efficiency
(2) Imperfect environmental conditions

Figure 6: Examples of IF in surveillance domain. **a)** Visible image **b)** Infrared image **c)** Fused image

Photography Domain Applications: Figure 7 illustrates the merging of multi-focus photos as an illustration of image fusion (IF) in the photography business. Since the cameras have smaller depths than conventional cameras, it is impossible to focus on everything that is at different distances from the camera in a single image. It is impossible to get every object in focus inside a single camera picture for every potential distance between the objects because of the camera's limited depths [29–30]. Utilizing the multi-focus IF approach, which combines many images of the same subject captured at various focus points to create a single fully in-focus image, is the answer to this issue. These are some of the numerous troubles that this domain is presently facing:

(1) Effect of moving target objects
(2) Relevance in consumer electronics

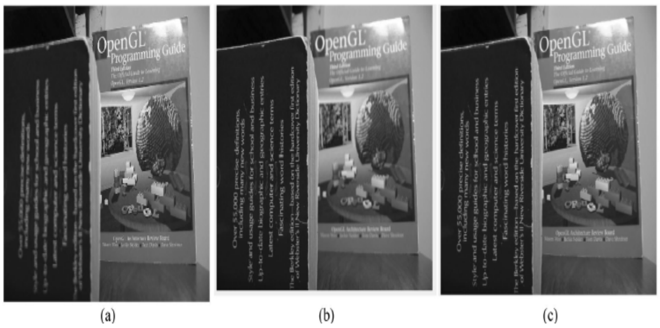

Figure 7: Examples of IF in photography domain. **a)** Back-focus Image **b)** Fore-focus image **c)** Fused image

V. PERFORMANCE EVALUATION METRICS

There are a number of effectiveness assessment indicators that are anticipated to be utilized in order to evaluate the efficacy of the various approaches to IF strategies.

The mean square error (MSE) is a statistical measure that is used to quantify the degree of inaccuracy as well as the actual disparity between the ideal and expected outcomes,

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(A_{ij} - B_{ij}\right)^2$$

The Structural Similarity Index Metric (SSIM) is a metric that determines how structurally analogous two or more images are to one another. It is produced by carrying out radiometric measurements and mimicking any contrast distortion that may occur. There is a combination of luminance image distortion, contrast distortion, loss correlation, and structural distortion that occurs between the source images and the final image. Each of these elements contributes to the overall distortion,

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}$$

A calculation that determines the ratio of peak power to noise value power is based on the peak signal to noise ratio, also known as PSNR,

$$PSNR = 10log_{10}\left\{\frac{r^2}{MSE}\right\}$$

The mutual information offers the information amount detail of the source images, which are then combined to generate the final image from the mutual information. The most effective application of the IF method is represented by the highest Mutual Information index,

$$MI_{AF} = \sum_{a,f} P_{A,F}(a,f)log\left[\frac{P_{AF}(a,f)}{P_A(a)P_F(f)}\right]$$

Entropy, often known as EN, is a measure that is utilized to assess the information content of an image. It also generates sensitive noise inside the image. An image that contains a significant amount of information has a low cross entropy,

$$EN = -\sum_{l=0}^{L-1} p_l log_2 p_l$$

VI. CONCLUSION

From the spatial domain to the transform domain to the deep learning level, the process of merging medical pictures has progressed. The rapid expansion of this sector reflects the significant need for computer-assisted clinical diagnostics. Numerous fusion procedures have been proposed by different researchers, and each offers a distinct set of benefits in terms of several significant assessment metrics. In contrast, there are around thirty different kinds of assessment indices that may be applied to the fusion of medical images. In summary, the medical image fusion method and other image fusion techniques on medical image fusion research in recent years are covered in this article. Additionally, it integrates the benefits of several approaches and the fusion effect with the recently introduced fusion technique. The research platform and data sets related to the case of various imaging fusion techniques and the statistical research trend are also explored in this work. Deep learning research in medical image fusion is the trend that will arise in the future, according to the part that came before this one.

References

[1] R. S. Blum, Z. Xue, and Z. Zhang, "An Overview of Image," Multisensor image fusion ans its applications, 2006.

[2] F. E. Z. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," Egyptian Informatics Journal, vol. 17, no. 1, pp. 99–124, 2016. https://doi.org/10.1016/j.eij.2015.09.002.

[3] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," Inf. Fusion, vol. 19, no. 1, pp. 4–19, 2014. https://doi.org/10.1016/j.inffus.2013.12.002.

[4] H. Hermessi, O. Mourali, and E. Zagrouba, "Multimodal medical image fusion review: Theoretical background and recent advances," Signal Processing, vol. 183, pp. 1-28, 2021. https://doi.org/10.1016/j.sigpro.2021.108036.

[5] M.A. Azam, K.B. Khan, S. Salahuddin, E. Rehman, S.A. Khan, M.A. Khan, S. Kadry, A.H. Gandomi, "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," Comput. Biol. Med., vol. 144, 2022. https://doi.org/10.1016/j.compbiomed.2022.105253.

[6] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," Inf. Fusion, vol. 33, pp. 100–112, 2017. https://doi.org/10.1016/j.inffus.2016.05.004.

[7] A. Kahol, G. Bhatnagar, "Ensemble-Inspired Multi-focus Image Fusion Framework," Lecture Notes on Data Engineering and Communications Technologies, vol 175, 2023. https://doi.org/10.1007/978-981-99-3432-4_21.

[8] F. Shabanzade and H. Ghassemian, "Multimodal image fusion via sparse representation and clustering-based dictionary learning algorithm in nonsubsampled contourlet domain," in 2016 8th International Symposium on Telecommunications (IST), Tehran, Iran, September 2016.

[9] Sharma, R. K., Mittal, A., & Agrawal, V. (2012). A design of hybrid elliptical air hole ring chalcogenide As2Se3 glass PCF: application to lower zero dispersion. International Journal of Engineering Research and Technology, 1(3).

[10] Mahima, N. B. Padmavathi, and M. V. Karki, "Feature extraction using DPSO for medical image fusion based on NSCT," in 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 265–269, Bangalore, India, May 2017.

[11] H. H. Inbarani, A. T. Azar, and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," Computer Methods and Programs in Biomedicine, vol. 113, no. 1, pp. 175–185, 2014.

[12] S. Singh, D. Gupta, R. S. Anand, and V. Kumar, "Nonsubsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network," Biomedical Signal Processing and Control, vol. 18, pp. 91–101, 2015.

[13] B. Zitova, J. Flusser, "Image registration methods: a survey," Image and Vision Comp., vol. 21, pp. 977–1000, 2003. https://doi.org/10.1016/S0262-8856(03)00137-9.

[14] S. Singh, H. Singh, G. Bueno, O. Deniz, S. Singh, H. Monga, P.N. Hrisheekesha, A. Pedraza, "A review of image fusion: Methods, applications and performance metrics," Digital Signal Processing, vol. 137, 15 June 2023, 04020. https://doi.org/10.1016/j.dsp.2023.104020.

[15] S. Singh, N. Mittal, H. Singh, "Review of Various Image Fusion Algorithms and Image Fusion Performance Metric," Archives of Computational Methods in Engineering, vol. 28, pp. 3645–3659, 2021. https://doi.org/10.1007/s11831-020-09518-x.

[16] Y. Xiong, Y. Wu, Y. Wang, and Y. Wang, "A medical image fusion method based on SIST and adaptive PCNN," in 2017 29th Chinese Control And Decision Conference (CCDC), pp. 5189–5194, Chongqing, China, May 2017.

[17] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 1, pp. 49–64, 2019.

[18] H. Ouerghi, O. Mourali, and E. Zagrouba, "Non-subsampled shearlet transform based MRI and PET brain image fusion using simplified pulse coupled neural network and weight local features in YIQ colour space," IET Image Processing, vol. 12, no. 10, pp. 1873–1880, 2018.

[19] Sharma, R. K., & Sharma, S. (2014). Design of HPCF with nearly zero flattened Chromatic Dispersion. International Journal of Engineering and Applied Sciences, 1(2).

[20] Maruthi R, Lakshmi I (2017) Multi-focus image fusion methods–a survey. Comput Eng 19(4):9–25

[21] Meher B, Agrawal S, Panda R, Abraham A (2019) A survey on region based image fusion methods. Inf Fus 1(48):119–132

[22] Liu Z, Chai Y, Yin H, Zhou J, Zhu Z (2017) A novel multi-focus image fusion approach based on image decomposition. Inf Fus 1(35):102–116

[23] James AP, Dasarathy BV (2014) Medical image fusion: a survey of the state of the art. Inf Fus 1(19):4–19

[24] Madkour M, Benhaddou D, Tao C (2016) Temporal data representation, normalization, extraction, and reasoning: a review from clinical domain. Comput Methods Programs Biomed 1(128):52–68

[25] Bai L, Xu C, Wang C (2015) A review of fusion methods of multi-spectral image. Optik-Int J Light Electron Optics 126(24):4804–4807

[26] Bavachan B, Krishnan DP (2014) A survey on image fusion techniques. IJRCCT 3(3):049–052

[27] Song L, Lin Y, Feng W, Zhao M (2009) A novel automatic weighted image fusion algorithm. In: 2009. ISA 2009.

International Workshop on Intelligent Systems and Applications, p 1–4

[28] Singh N, Tanwar, P (2012) Image fusion using improved contourlet transform technique. Int J Recent Technol Eng (IJRTE), vol 1, no. 2

[29] He K, Sun J, Tang X (2010) Guided image filtering. European conference on computer vision. Springer, Berlin, pp 1–14

[30] Harris JR, Murray R, Hirose T (1990) IHS transform for the integration of radar imagery with other remotely sensed data. Photogramm Eng Remote Sens 56(12):1631–1641

[31] Smith LI (2002) A tutorial on principal components analysis. Statistics 51(1):52

[32] Li S, Kang X, Hu J (2013) Image fusion with guided filtering. IEEE Trans Image Process 22(7):2864–2875.

[33] J. Du, W. Li, B. Xiao, and Q. Nawaz, ``Union Laplacian pyramid with multiple features for medical image fusion,'' Neurocomputing, vol. 194, pp. 326339, Jun. 2016.

Riya Gupta, M.Tech Scholar, Department of Computer Science & Engineering, Faculty of Engineering & Technology, Rama University, Kanpur, (U.P.) India.

Dr. C. S. Raghuvanshi, Professor, Department of Computer Science & Engineering, Faculty of Engineering & Technology, Rama University, Kanpur, (U.P.) India.