Multi-sensor Image Fusion using Optimized and Adjustable Non-subsampled Shearlet Transform and Measurement of Fusion Performance

A

thesis Submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Amit Vishwakarma

Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
Guwahati - 781 039, India
July, 2019

Declaration

This is to certify that the thesis entitled "**Multi-sensor Image Fusion using Optimized and Adjustable Non-subsampled Shearlet Transform and Measurement of Fusion Performance**", submitted by me to the *Indian Institute of Technology Guwahati*, for the award of the degree of Doctor of Philosophy, is a bonafide work carried out by me under the supervision of Prof. M.K. Bhuyan. The content of this thesis, in full or in parts, have not been submitted to any other University or Institute for the award of any degree or diploma.

Signed: _____

Amit Vishwakarma
Department of Electronics and Electrical Engineering,
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India.

Date: _____

Dedicated
to My family, teachers and friends. . .

Certificate

This is to certify that the thesis entitled "**Multi-sensor Image Fusion using Optimized and Adjustable Non-subsampled Shearlet Transform and Measurement of Fusion Performance**", submitted by Amit Vishwakarma (146102014), a research scholar in the *Department of Electronics and Electrical Engineering*, *Indian Institute of Technology Guwahati*, for the award of the degree of Doctor of Philosophy, is a record of an original research work carried out by him under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Signed: _____

Supervisor: Prof. M.K. Bhuyan
Department of Electronics and Electrical Engineering,
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India.

Date: _____

Acknowledgements

The completion of my dissertation symbolizes the end of my five years study in Guwahati, India. The dissertation would not have been finished without the support of many people, to whom I would like to express my sincere gratitude. For Prof. M.K. Bhuyan, my advisor for over five years, I do not have sufficient words to convey my thanks to him. He kindly accepted me to be in his research group I came to Indian Institute of Technology Guwahati. Since then, he has always been supportive to me and guided me on different research projects. I have learned a lot about research process from him and I wish I can have the enthusiasm as much as he has in exploring new research findings. My sincere thanks also goes to my other committee members Prof. Prabin Kumar Bora, Prof. Kannan Karthik, Prof. Arijit Sur, and Prof. Hanumant Singh Shekhawat for their valuable inputs to this dissertation.

Besides, I would like to thank my current and senior researches in my Group, who helped me a lot in my research projects. They always discussed with me about the detail of our approach and gave me substantial suggestions. I appreciate the time working and getting along with many graduate students in research group and EEE department. It was a great experience working and discussing with them. To my good friends here at the IIT Guwahati, thank you all for enriching my life and helping me in few difficult times. Last but not least, this dissertation is dedicated to my family, for their love and support to me. Without their support and encouragement, I would not have had the chance to pursue my Ph.D. degree and expand my horizon.

Sincerely
Amit Vishwakarma

Abstract

Image fusion is a technique to assimilate information acquired from similar or dissimilar sources of images. Compared to the source images, fused images convey more information, and they have more clarity. However, distorted fused images are obtained in many cases, which only misinterpret the actual scene information. This happens due to the shortcomings of fusion rules, fusion framework, image decomposition techniques, and inaccurate source image registration.

To address these issues, we proposed few efficient fusion techniques to generate better quality fused images for different computer vision applications. In the proposed method presented in ***Chapter 2****, non-subsampled shearlet transform (NSST) is used to decompose pre-registered source images into low- and high-frequency components. These low- and high-frequency coefficients are fused by using our proposed modified weighted salience and local difference fusion rules, respectively. To enrich edge information in the fused image, Canny edge detector with scale multiplication is employed. Moreover, a metric Q_{TS} is proposed to jointly measure both texture and structural information present in the fused image. The proposed metric is formulated on the basis of local standard deviation filtering, local information entropy, and local difference filtering. Both subjective and objective results validate the proposed fusion framework and the metric Q_{TS}. The proposed method can deal with noisy source images due to the deployment of anisotropic diffusion filtering. However, proposed method is unable to produce state-of-the-art results in case of multi-focus images due to non-optimal frequency response of NSST filters and lack of adaptivity in fusion rules. Moreover, there is still a scope of formulating a new metric to measure other attributes of fused images.*

In ***Chapter 3****, an optimized non-subsampled shearlet transform (NSST) is developed, which is applied to decompose source images into low- and high-frequency bands. The low-frequency bands are fused using proposed descriptor obtained by the superposition of scale multiplied Canny edge detector features and Hessian features. The high-frequency bands are fused using unsharp masking based fusion rules. Moreover, a metric Q_E is formulated on the basis of Karhunen-Loeve transform (KLT). The information of image pixel variance for both source and fused images can be measured by using the proposed metric Q_E, and it gives an indication of the amount of variance information transferred*

from the source images to the fused image. Both subjective and objective analysis show the efficacy of the proposed fusion structure and the metric Q_E. The improvement in performance is due to the deployment of more complex optimized NSST and fusion rules. However, the multi-scale transform (MST) or multi-geometrical analysis (MGA) tools (such as NSST, curvelet, contourlet, etc.) based low-frequency band fusion some times under perform as compared to the sparse representation based low-frequency fusion.

MST and sparse representation based image fusion techniques are very popular. However, existing fusion methods can not perfectly capture significant details, texture and fine information of the source images due to few limitations of MST- and sparse representation-based techniques. Moreover, for objective evaluation of fusion algorithms, it is quite important to simultaneously measure both singularities and structural information of the source images preserved in the fused image. So, a metric which can efficiently measure both singularity and structural information of the fused image is much needed for fair comparison of existing fusion methods. To address these issues, a modified Meyer window based adjustable non-subsampled shearlet transform (ANSST) is proposed to decompose the pre-registered source images into low- and high-frequency coefficients in ***Chapter 4****. The low-frequency bands are fused using convolutional sparse representation (CSR) modeling. The high-frequency bands are fused using proposed Information Entropy, Standard Deviation, and Range (IESDR) descriptor, which considers entropy, standard deviation, and range filtering features, respectively. Moreover, a metric Q_{SS} is proposed to jointly measure both singularities and structural information present in the fused image. The proposed metric is formulated on the basis of the proposed native division and native difference filtering. To show the applicability of the proposed ANSST, the proposed ANSST is applied for image denoising. Both subjective and objective results validate the proposed fusion framework and the metric Q_{SS}. The main limitation of this method is that it is more computationally complex than the methods proposed in* ***Chapter 2*** *and* ***Chapter 3****.*

In ***Chapter 2****, we proposed local difference features to fuse high-frequency information. To show one application of our local difference features, an image mosaicking framework is proposed with the help of these features. A local difference-based modified Harris detector is proposed to increase time efficiency of image mosaicking algorithms in* ***Chapter 5****.*

Contents

List of Figures

List of Tables

1

Introduction

Image fusion is a technique to assimilate information acquired from similar or dissimilar sources of images. Compared to the source images, fused images convey more information, and they have more clarity. Image fusion techniques are employed to combine images from diverse applications, such as biomedical, multi-focus, multi-exposure, infrared imaging, etc. Medical image fusion has many clinical applications, like non-invasive diagnosis, treatment planning, image-guided surgery, and radiotherapy. The visible-infrared image fusion techniques are used in many applications, such as concealed weapon detection, runway aviation, etc.

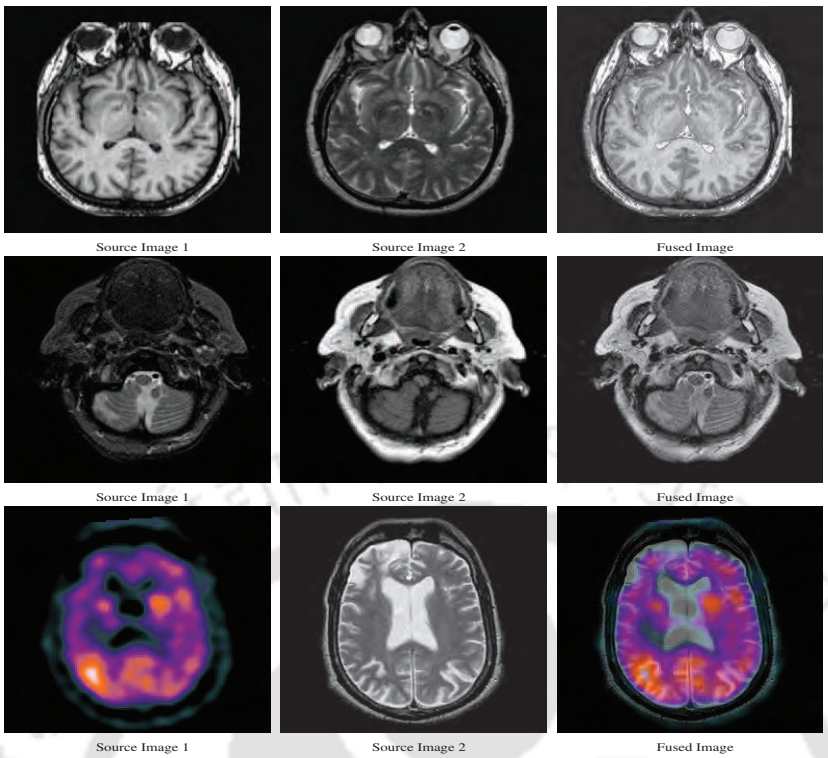


FIGURE 1.1: Examples of medical fused images [5].

1.1 Introduction

Image fusion is recognized as an effective enhancing methodology broadly included in high-quality imaging systems. Three elementary requirements are defined for optimum image fusion [1]. First, the fusion process needs to transfer prominent information of the source images into the fused image. Second, all the relevant information of the source images should be preserved after the fusion process. Third, a fused image should be free from any undesirable artifacts.

Few examples of medical image fusion are shown in Fig. 1.1. Computerized tomography (CT) images show hard tissue details, like bones, internal organs, etc. In contrast, magnetic resonance imaging (MRI) is used to show the anatomy, physiological processes, and soft tissues of a human body. Image fusion methods are applied to fuse CT and MRI images to observe hard and soft tissue details simultaneously in a single fused image. Single-photon emission computed tomography (SPECT) images are used to show blood flow information inside the human body. Hence, radiologist considers SPECT and MRI images for better diagnosis and clinical applications [2–4].

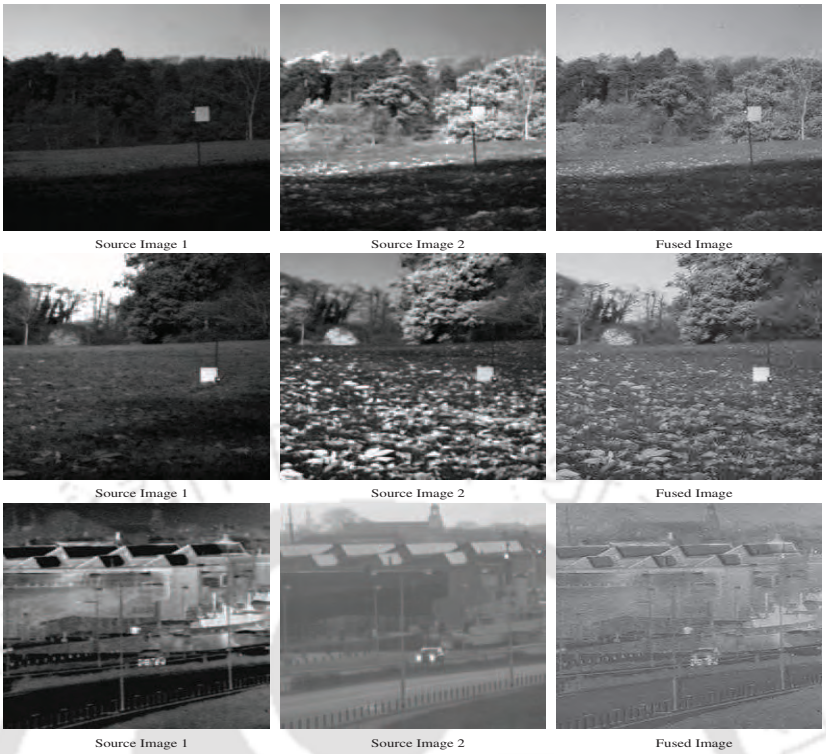


FIGURE 1.2: Examples of visible-infrared fused images [5].

Some examples of visible-infrared image fusion are shown in Fig. 1.2. The visible-infrared image fusion is another case of multi-sensor image fusion [1]. Infrared images capture thermal radiations emitted from objects, and the scene illumination has less influence on these images. However, infrared images have a less spatial resolution as compared to visible images. Thus, texture information of the visible image is partially absent in the respective infrared image. In contrast, visible images are formed due to the reflected light, and the images contain significant visible information. Fusion of complementary attributes of infrared and visible images can produce significantly high quality fused images for many important applications, such as concealed weapon detection, human presence detection etc. The fused image highlights user specific prominent image regions in contrast to the source images.

Examples of multi-focus image fusion are shown in Fig. 1.3. Multi-focus image fusion is used to obtain the single image in which all the objects are in focus. The far and near defocus regions appeared in multiple images are removed by multi-focus image fusion. The fused image contains all the regions of multiple out-of-focus images. Some multi-sensor image fusion methods are described in [6–12].

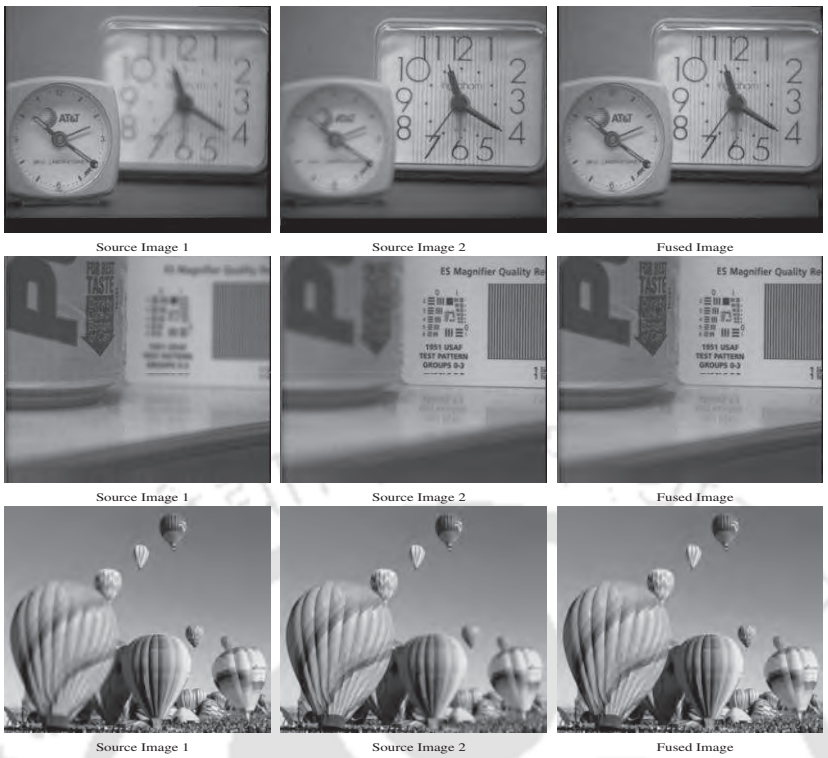


FIGURE 1.3: Examples of multi-focus fused images [5].

There is a need to perfectly register two images to minimize spatial artifacts in the fused image prior to perform image fusion. Image registration is the process of overlaying images (two or more) of the same scene taken at different times, from different viewpoints, and/or by different sensors. Some image registration algorithms are proposed in [13, 14].

This thesis introduces new techniques for fusion of single and multi-modal source images. In the following section, existing image fusion techniques are briefly discussed to mainly highlight the pros and cons of the existing methods. Finally, motivation and objectives of this research work are fixed for contributing in this domain.

1.2 Literature Survey

As explained earlier, image fusion is combined intelligibility of two or more source images of different modalities, such that the fused image gives more information for decision making. Image fusion make the data more accurate, interpretable, reliable, and less ambiguous. There are several objectives of image fusion, such as increasing the spatial

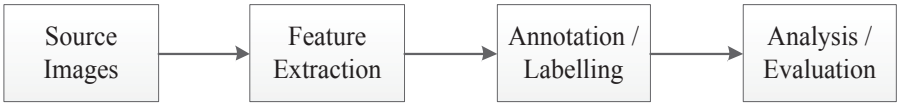


FIGURE 1.4: Basic sequential steps to perform image fusion.

resolution, boosting of certain image contents or filling of missing information in an image.

There are four basic sequential steps in general to perform image fusion as shown in Fig. 1.4. Image fusion techniques can be broadly categorized into three classes, and they are pixel-, feature-, and decision-level fusion. This classification is made on the basis of the level at which fusion is performed. Merging is performed before the feature extraction step in pixel-level fusion. In feature-level fusion, merging is performed after the feature extraction step, but prior to annotation/labelling step. However, merging is performed after the annotation/labelling step in decision-level fusion.

1.2.1 Pixel-level fusion

Pixel-level fusion techniques perform merging on lowest measured physical parameters of images. For fusion, the accurate registration of multi-modal source images is a highly important requirement as misregistration may lead to false interpretation of information. Pixel-level image fusion techniques can be implemented in either spatial domain or transform domain.

In spatial domain methods, fusion is performed directly on raw image pixels using fusion rules. Dictionary learning based fusion methods fall in the category of spatial domain methods [15]. In the transform domain methods, the image is first converted into the transform domain representation, and subsequently the transform coefficients are fused by appropriate fusion rules. Finally, inverse transform is applied to get the fused images. The quality of fused image obtain using pixel-level fusion mainly depends on the transformation technique and applied fusion rules. Most commonly applied fusion rules are choose max or min [16, 17], weighted average [18], machine learning approaches [19], window-based approaches [20], and component substitution [21].

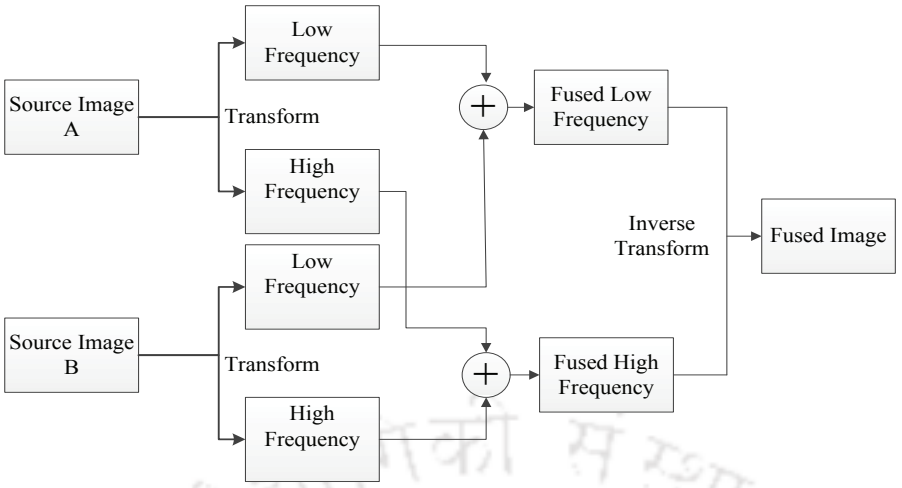


FIGURE 1.5: Basic framework of transform-based image fusion methods [5].

Multi-scale transform (MST) based fusion methods: Multi-scale transform (MST) techniques are widely used in different image analysis tasks. The MST techniques basically consist of pyramid transforms (such as Laplacian [22], gradient [23], etc.) and discrete wavelet transform [24]. The basic flowchart of transform-based fusion methods are shown in Fig. 1.5. MST methods decompose the source images into different frequency bands. These bands are fused using suitable fusion rules. The final fused image is obtained using the inverse MST operation. Few multi-geometrical analysis (MGA) tools are also applied to perform image fusion. These tools are curvelet transform [25, 26], ridgelet transform [27], contourlet transform [28, 29], non-subsampled contourlet transform (NSCT) [30], shearlet transform (ST) [31, 32], non-subsampled shearlet transform (NSST) [33], and Directionlets [34], etc. There are number of fusion strategies which are being developed to get better fusion results. Some of them are briefly discussed here.

The activity level measurement based fusion rules depend on coefficients level measure, window level measure, and regions (odd shape window) level measure. Fusion is done by observing one coefficient at a time in coefficient-based activity level measurement. However, fusion is performed by observing the group of pixels in an window and region based activity level measurements [35].

The grouping (merging) is done either after single- or multi-scale decomposition of the source images in coefficients grouping-based fusion schemes [36, 37]. The coefficients combining fusion methods employ max and weighted averaging fusion rules. Coefficients

with higher activity level are given priority in choose max scheme. The weights used in weighted averaging scheme are calculated by considering neighbours of the coefficients. The consistency verification based fusion rule ensures similarity between the neighbouring pixels of the fused image. Majority filter is widely used to perform consistency verification procedure [38].

The averaging rule is a basic and simple image fusion scheme in which the average values of the pixels intensities are assigned to the pixels of the fused image. The choose max and choose min fusion rules select high- and low-intensity pixels of the source images, respectively, which are eventually assigned to the fused image, respectively [36]. The weighted averaging based principal component analysis fusion rule is widely used to fuse two source images, where the normalized eigenvectors are assigned as weights to the source images [39].

Image fusion in transform and spatial domain: The wavelet transform is widely used for multi-scale and multi-resolution analysis of images. 2-D version of the wavelet transform is obtained by the tensor product of two 1-D wavelet transforms. It decomposes source images into one low-frequency and a number of high-frequency bands. These decomposed bands are fused using any suitable fusion rules. The wavelet transform has isotropic scaling and shift-variant nature. So, it can not efficiently represent directional details, and it also produces ringing artifacts in the fused images [40].

The inefficiency of wavelet transform to represent smooth curves in images is overcome by using curvelet transform. Therefore, the curvelet transform is widely used for image fusion tasks [41]. In discrete space, contourlet transform is also proposed to represent smooth curves in images. The designing of contourlet transform is directly done in discrete space, which results less computational complexity as compared to curvelet. That is why, it is also used for fusion of multi-modal source images [28, 42].

The effects of curvelet and contourlet transform on image fusion performance are analyzed in [43]. The experiments performed in [44] show that the filter length, shift invariance, and the level of decomposition can affect the fusion performance. In [45], contourlet transform is used to fuse remote sensed images and it was found that due to shift variant nature of contourlet transform, generated fused images are affected by

ringing artifact (also known as pseudo-Gibbs phenomenon). These ringing artifacts appear in the fused images due to misregistration error of the source images. Due to shift invariant property of NSCT, it is finally used to minimize ringing artifacts.

The ripplet transform was introduced in [46] to efficiently represent smooth directional information. However, it can not efficiently represent highly complex anisotropic structures present in the source images. The ripplet transform along with sparse representation was presented in [47] to improve the fusion performance for remote sensed images.

Another very popular geometrical tool is NSST, which was proposed in [48] to efficiently represent curve like information present in the source images. The contourlet transform consists of Laplacian pyramid followed by directional filters. The main advantage of the ST over the contourlet transform is that there are no constraints on the number of shearing directions which allows the use of a non-dyadic spacing. Moreover, in the ST, there are no restrictions on the size of the supports for shearing in contrast to contourlet transform. Finally, the inverse of discrete shearlet transform needs to be computed, which only requires a sum of the shearing filter responses instead of inverting a directional filter bank. This procedure improves the computational efficiency [31, 49]. NSST is a combination of non-subsampled Laplacian pyramid and a set of shearing filters. It is shift invariant version of the ST (which is inherently shift variant). The shift invariant property reduces the ringing artifacts in the fused images.

In the literature, image fusion methods based on ST [50] and NSST [51] have been recently proposed. In the ST [50] method, low-frequency bands are fused using regional energy, and high-frequency bands are fused by regional variance, regional average gradient, and regional spatial frequency-based fusion rules. However, the subsampled nature of ST produces ringing artifacts in the fused images. Moreover, this drawback is removed by NSST presented in [51]. In this method, structure tensor and NSST are used to fuse medical images. However, averaging rule for low-frequency fusion and weight generation procedure produce blurring in the fused images.

The wavelets and Laplacian pyramid-based methods are also applied to perform image fusion. However, the fixed basis function and fixed frequency response of analysis

and synthesis filters bring some limitations in these methods. Therefore, few data dependent approaches was introduced to perform image fusion, such as empirical mode decomposition [52, 53]. The empirical mode decomposition approach decomposes the input signal which is stationary and non-linear in nature. The generated intrinsic mode functions are used to perform image fusion. The empirical mode decomposition is used for pansharpening task which shows some improvements over earlier methods proposed in [53]. However, fusion results obtained using these methods have some artifacts due to the use of different intrinsic mode functions for different input channels.

To resolve the issues of mode mixing and misalignment of channels, multivariate empirical mode decomposition is proposed to generate a matched set of intrinsic mode functions for each input signal. In [54], satisfactory results are obtained for the fusion of remote sensed images. Another multivariate empirical mode decomposition based method is proposed to perform the fusion of multi-exposure images, multi-focus and remote sensed images [55]. The results obtained by this method were better as compared to the results obtained using principal component analysis, DWT, and NSCT based fusion methods.

The spatial domain fusion methods directly operate on the pixels values to produce the fused image. These methods also require registration of source images similar to transform domain fusion methods. The very simple spatial domain fusion method takes the average intensity values of the source images. Spatial domain methods have now become very popular due to the advancement of different edge preserving filters, such as bilateral, guided filters, etc. [56, 57].

In [58], cross bilateral filter is proposed to fuse multi-model and multi-focus source images. In osseous and vascular imagery fusion, the experiments were carried out using different edge-preserving filters, such as robust bilateral filter [59], weighted bilateral filter [59, 60], rolling guidance filter [61]. These edge preserving filters highlight spatial characteristics of the fused images. In [62], bilateral and directional filters are used to fuse multi-modal and multi-exposure source images in multi-scale representation. Moreover, a very efficient edge preserving guided filter is used to fuse multi-model source images in [63]. The detailed and base sub-images are obtained using guided filter decomposition,

and they are combined to generate the fused images. The Gaussian and bilateral filters are jointly used to fuse infrared and visible source images in [64].

Non multi-scale decomposition based fusion methods: The fusion methods which do not use multi-scale decomposition representation for image fusion fall in the category of non-multi-scale decomposition based fusion methods. The non-multi-scale decomposition based fusion methods are generally classified into five categories [55]. These five categories are pixel-level weighted averaging methods, non-linear methods, estimation theory-based methods, colour composite fusion methods, and artificial neural network-based methods [55].

Principal component analysis and adaptive weighted averaging methods belong to the category of pixel-level weighted averaging fusion [65]. In [66], a non-linear scheme was proposed to fuse visible-infrared source images. In this method, less bright images are taken into consideration. The visual perception of these images can be improved by image fusion. Some non-linear scheme based fusion methods are presented in [67, 68].

In literature, estimation theory-based methods are also used to fuse multi-model source images. These methods basically consist of two models, such as prior model and image formation model. A review on estimation theory-based image fusion method was presented in [69]. A random walk optimization and multivariate conditional Gaussian random field model based fusion approach was proposed in [70]. Some estimation theory methods, such as Markov random field are also employed to perform image fusion in [71, 72].

Artificial neural networks are popularly used to fuse multi-sensor source images. The pulse coupled neural networks [73, 74] and multi-layered perceptron neural networks [75] have been mainly used for image fusion.

The applications of pulse-coupled neural network for image fusion were discussed in [76], where the inputs to the network are spatial features of the source images. Artificial neural networks-based fusion methods are popular due to higher training capability and lower mathematical complexity. However, these methods highly depend on the quality of training data, optimal parameter tuning, and convergence of training algorithms. These factors constrain the visual qualities of the fused images. An improved artificial

neural network-based method for fusion was proposed in [77]. Few neural networks and neuro-fuzzy based medical image fusion methods were proposed in [78–80]. Some other artificial neural network-based methods which are used to fuse multi-focus and -model source images are presented in [81, 82].

Recently, sparse-based representation methods have become very popular in different computer vision applications. Therefore, sparse representation technique can be deployed for image fusion. The over complete dictionary-based methods can produce good quality fused images. Unlike traditional MST methods which have fixed basis functions, these methods learn the basis function from the input images. Moreover, the few non-zero sparse weighted coefficients ease the fusion process.

In [83], sparse representation-based multi-focus image fusion method was proposed. This method used a choose max scheme for fusion of sparse coefficients. Then, the fused image is obtained using the fused coefficients and the trained dictionary. However, the fused image obtained using this method do not have fine details of the source images. The orthogonal matching pursuit algorithm is exploited to obtain sparse coefficients in [84]. This algorithm produces special structures in sparse coefficients.

In [83, 85], the input images are partitioned into overlapping patches to obtain a sparse vector. The simultaneous orthogonal matching pursuit algorithm is used to obtain sparse coefficients from the patches of the source images in [85]. This method uses the correlation between the image pixels for extracting the sparse vector.

Some other sparse representation-based image fusion methods are presented in [86–89]. However, dictionaries based on a single mathematical model (such as K-singular value decomposition or fixed basis) cannot efficiently represent different attributes of multi-model source images. Therefore, spatial detailing and spectral dictionary are proposed to improve the visual quality of the fused images [90, 91]. In sparse representation-based image fusion methods, spatial information is lost and the computational cost also increases due to the patch-based processing.

There are some fusion methods which use DWT in combination with other tools, such as independent component analysis [92], and principal component analysis [93]. A fusion method based on sparse representation and morphological component analysis is

applied for fusion of multi-sensor source images [94]. In [95], spatial context information of the fused image is increased by using the spatial frequency of local regions of the source images. Graph cut algorithm is used to fuse medical source images for medical diagnosis in [96].

Till now, we discussed MST, non-MST, and other frequency domain-based image fusion methods. Some hybrid methods have also been proposed to combine advantages of these frequency domain methods. The joint use of sparse representation and MST for image fusion is presented in [97]. This method is computationally efficient and it can produce good quality fused images. However, blocking artifacts appear in the fused images due to the use of nonzero sparse vector, which can be minimized by deploying MST.

The method presented in [98] first decomposes the source images into low- and high-frequency components. These components are fused using the sparse representation. Subsequently, the fused image is obtained using the inverse MST. Another intensity hue saturation transform and principal component analysis based fusion method [99] was proposed to capture functional and spatial details of the source images.

The wavelet and contourlet transform are jointly used in the methods [89, 100]. In [89], sparse representation and contourlet transform are jointly used to preserve both low- and high-frequency contents of the multi-focus and multi-model source images. The image fusion framework which employs random walk was proposed to fuse multi-focus and multi-exposure source images [101]. The random walk graph-based method is also proposed to fuse multi-focus source images [102], where a fully connected graph is produced to preserve the colour and spatial information in the fused images.

MST- and MGA-based methods are good for representing high-frequency information, but inefficient to detect crucial low-frequency information. However, sparse representation based fusion methods are good for representing low-frequency information, but inefficient to represent high-frequency details [89]. The principal component analysis-based fusion methods are not suitable especially for the fusion of multi-focus source images [1–3, 39].

1.2.2 Feature-level fusion

Feature-level fusion uses more higher level processings as compared to pixel-level fusion. These methods first extract the features from the source images, then these features are fused. Unlike pixel-level fusion, feature-level fusion aims to fuse specific features instead of fusing entire source images. Image segmentation is a very crucial step in feature-level fusion. In feature-level fusion, similar regions of interest of the multi-modal source images are extracted using any suitable image segmentation technique. Then, these regions are fused with the help of image features. In feature-level fusion, we are more interested in the fusion of extracted image features as they provide localized and precise information.

This type of fusion is tricky and inefficient when the features are obtained for different source images or different feature extraction techniques. In [103], sparse feature matrix and morphological filtering-based multi-focus image fusion method is presented. This method uses sparse feature matrix for source image decomposition. The morphological filtering is used for extracting dark and bright regions of the decomposed sub-band images. The fused image is obtained using these extracted features.

In [104] three different fusion methods are proposed, and they are principal component analysis-based fusion, an edge detection-based fusion and segmentation-based fusion methods. The principal component analysis-based method fuses the features which convey significant information of the source images. The edge detection fusion method uses gradient operation to obtain the fused image. The segmentation-based fusion method preserves spectral and spatial information of the source images.

The dual-tree complex wavelet transform-based feature-level fusion method is proposed in [105]. This method uses the watershed segmentation algorithm to form a region map of interested regions. The fused image is obtained using wavelet transform. However, the fused results may be distorted when the segmentation is not accurate.

1.2.3 Decision-level fusion

Source images are processed to extract relevant information in decision-level fusion. The decision-level fusion is also known as symbol or interpretation level fusion. In this fusion,

the locally classified data from several source images are combined. The source images are annotated/labelled to apply decision rules, which gives better representation of the regions of interest. The main consideration of decision-level fusion is that high-level features of the source images can give a more robust fusion. The decision-level fusion are used in fingerprint verification, remote sensing, parallel and distributed processing systems [106].

The decision level fusion method [107] for fingerprint verification signifies the choice of the classifier. The method proposed in [107] uses the Neyman Pearson optimal rules for decision making. Some decision level fusion methods based on Bayesian interface is presented in [106].

The clear and uncertain local classifier-based decision level fusion method was proposed in [108]. The wide spectral band data, such as microwave, thermal, infrared and visible images are fused using this method. Moreover, correctness metrics are included in decision making to improve the classifier performance. Another decision level fusion method is proposed for object recognition in [109]. This method first classify the objects present in multiple source images, and then these classified objects are finally fused by using a suitable decision level fusion technique.

1.2.4 Recent trends in image fusion

The most recent and advanced image fusion techniques are briefly discussed in this section. A bilateral filter-based multi-focus image fusion method was proposed in [110] for enhancement of gradient information in the fused image. The phase coherence and detailed information are used to get sharp fused images for better visual perception. However, this method introduces some artifacts in the fused images. The image matting technique is employed to fuse multi-focus images in [111]. This method uses a morphological technique to segment the regions of the source images. These segmented regions are fed into the input of the image matting algorithm, which is used to identify the highly focus regions of the source images. However, some crucial information of the source images are lost in this method.

The exploration of singular value decomposition (SVD) in image fusion is carried out in [112] to fuse images of different resolutions and sizes. The SVD gives satisfactory performance for panchromatic and multi-spectral data. However, the high computational complexity of SVD limits its performance in some applications. To measure the sharpness of the fused images, the wavelet coefficients based sharpness metric is proposed in [113]. The wavelet transform is combined with Laplacian mixture model to fuse the source images with the help of this sharpness metric. However, the inefficient representation ability of wavelets in the case of directional information limits the fusion performance. The guided filter-based approach is proposed to decompose the source images into the base and detail layers [114] for the fusion of multi-modal and multi-focus images. The fusion of these layers is carried out using a weight optimization method. However, these generated weights introduces blurring effect in the fused images.

A 3D medical image fusion method is proposed for common clinical and diagnosis practice in [36]. The cross-scale fusion rule is applied to take care of the inconsistencies in intra-scale and inter-scale sub-images. Moreover, it also address the neighbourhood around each coefficient. However, this method introduces noise and spatial artifacts in the fused images. The dense scale invariant feature transform (SIFT) is also employed to fuse multi-exposure, multi-focus, and multi-model source images [115, 116]. It minimizes the pixel misregistration error due to the matching operation among the multi-focus source images. Dense SIFT method constitute three attributes of the source images to fuse them, such as spatial consistency, exposure quality and local contrast. However, SIFT-based methods are inefficient to preserve all the relevant information of the source images in the fused images.

The convolutional sparse representation (CSR) [117] is an improved form of sparse representation by addressing shift invariant and minimizing the misregistration artifacts in the fused images. Hence, the visual qualities of the fused images are improved by using CSR modeling in fusion framework [88]. This method outperforms most of the state-of-the-art sparse representation based fusion methods. The introduction of a convolutional neural network (CNN) in case of image fusion is presented in [118]. This method trained the CNN's on focused images patches and their blurred versions. This method produces state-of-the-art results in case of multi-focus images due to the large training data and

high model accuracy. However, this method under-performs in case of multi-model (such as medical and visible-infrared) source images.

The sparse representation-based medical image fusion method is introduced in [119]. This method employed three decision maps such as structure information map (SM), energy information map (EM), structure and energy map (SEM). The visual qualities of the medical fused images are improved as compared to previous sparse representation based methods due to the application of these maps.

1.2.5 Performance evaluation metrics

The quantitative assessment of a fused image can be carried out by two different approaches. The first approach is to compare the fusion result with a known reference image or ground truth images. However, a reference image is not always available in most of the practical applications. The second approach is a blind assessment, in which no reference image is needed, and that is why it is widely preferred in image fusion domain. In this thesis, we will use blind assessment to evaluate fusion performance.

In order to quantitatively evaluate the performance of image fusion algorithms, a number of quantitative evaluation metrics have been proposed in [120–122]. These metrics are statistically formulated. Each metric can measure some specific attributes of the fused image with respect to the source images. However, the problem is that certain fusion metrics may be suitable for some specific source images but they show altogether different performance for some other images [123, 124]. In addition, two or three quantitative metrics can not perfectly judge the fusion performance for the application like multi-focus imaging, surveillance (visible-infrared), and medical imaging, etc. [123, 124]. In this section, we will discuss about some of the popular state-of-the-art metrics which are widely used for measurement of fusion performance.

(a) Mutual information: Mutual information (MI) in image fusion represents the amount of information transferred from the source images into the fused image. If MI^{AF} and MI^{BF} represent MI between the source images A and B to the fused image F, respectively; $a \in [0, L-1]$, $b \in [0, L-1]$ and $f \in [0, L-1]$, where L is maximum

intensity level of the source images. The MI between the source images and the fuse image is given by:

$$\mathrm{MI} = \mathrm{MI}^{AF} + \mathrm{MI}^{BF} \tag{1.1}$$

where, MI^{AF} and MI^{BF} are given as:

$$\mathrm{MI}^{AF} = \sum_{f=0}^{L-1}\sum_{a=0}^{L-1} p^{AF}(a,f)\log_2\left(\frac{p^{AF}(a,f)}{p^{A}(a)p^{F}(f)}\right), \tag{1.2}$$

and

$$\mathrm{MI}^{BF} = \sum_{f=0}^{L-1}\sum_{b=0}^{L-1} p^{BF}(b,f)\log_2\left(\frac{p^{BF}(b,f)}{p^{B}(b)p^{F}(f)}\right). \tag{1.3}$$

where, $p^{A}(a)$, $p^{B}(b)$ and $p^{F}(f)$ are normalized gray level histograms of the images A, B and F, respectively. The joint gray level histogram of the source images with fused image are denoted by $p^{AF}(a,f)$ and $p^{BF}(b,f)$. The higher value of mutual information and normalized mutual information (Q_{MI}) [120] indicate better fusion results [123].

(b) Metric based on structure similarity (SSIM): The SSIM consider three criteria for modelling of image distortion, an they are loss of correlation, radiometric and contrast information. Generally, SSIM between the two variable U and V is given as [121]:

$$\mathrm{SSIM}(U,V) = \frac{\sigma_{UV}}{\sigma_U\sigma_V}\frac{2\mu_U\mu_V}{\mu_U^2+\mu_V^2}\frac{2\sigma_U\sigma_V}{\sigma_U^2+\sigma_V^2} \tag{1.4}$$

where, μ_U and μ_V are the mean intensity value of variable U and V, respectively. The σ_U, σ_V and σ_{UV} are the variances and covariance of U and V, respectively. In Eq. (1.4) the term $\frac{\sigma_{UV}}{\sigma_U\sigma_V}$ represents structure information, $\frac{2\mu_U\mu_V}{\mu_U^2+\mu_V^2}$ represents luminance information and $\frac{2\sigma_U\sigma_V}{\sigma_U^2+\sigma_V^2}$ represents contrast information. An approach is proposed in [125] to calculate the SSIM, if the ground truth image is not available, is given as:

$$Q_S = \begin{cases} \lambda(w)\mathrm{SSIM}(A,F|w) + (1-\lambda(w))\mathrm{SSIM}(B,F|w), & if \quad \mathrm{SSIM}(A,B|w) \geq 0.75 \\ \max[\mathrm{SSIM}(A,F|w), \mathrm{SSIM}(B,F|w)], & if \quad \mathrm{SSIM}(A,B|w) < 0.75 \end{cases} \tag{1.5}$$

where w is window size and $\lambda(w)$ is local weight obtained by local image salience, such as variance of the pixels of the source image. The optimum value of Q_S is 1. More detail implementation of this metric is given in [125].

(c) Edge based similarity measure: Edge transfer metric $Q^{AB/F}$ uses magnitude and orientation of the edges obtained using Sobel edge detector. The edge information transferred from source images to the fused image is given as:

$$Q^{AB/F} = \frac{\sum_{i=1}^{M_0}\sum_{j=1}^{N_0}\left[Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)\right]}{\sum_{i=1}^{M_0}\sum_{j=1}^{N_0}\left[w^A(i,j) + w^B(i,j)\right]} \tag{1.6}$$

where, $w^A(i,j)$, $w^B(i,j)$, $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$ are defined as follows:

$$w^A(i,j) = |\sqrt{S_i^A(i,j) + S_j^A(i,j)}|^L \tag{1.7}$$

$$w^B(i,j) = |\sqrt{S_i^B(i,j) + S_j^B(i,j)}|^L \tag{1.8}$$

$$Q^{AF}(i,j) = Q_a^{AF}(i,j)Q_g^{AF}(i,j) \tag{1.9}$$

$$Q^{BF}(i,j) = Q_a^{BF}(i,j)Q_g^{BF}(i,j) \tag{1.10}$$

where, $M_0 \times N_0$ is the size of the image. The $w^A(i,j)$ and $w^B(i,j)$ are gradient magnitudes of the source images A and B, respectively. For source images A and B, $S_i(i,j)$ and $S_j(i,j)$ denote Sobel edge detected images in horizontal and vertical direction, respectively. Edge preservation values of each of the source images at location (i,j) are denoted as $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$. Moreover, the edge strength and orientation preservation values at location (i,j) for each source images A and B are denoted as $Q_a^{xF}(i,j)$ and $Q_g^{xF}(i,j)$, respectively. The dynamic range of $Q^{AB/F}$ is [0,1] and it should be as close to 1 as possible for better fusion [122].

1.3 Problem Definition and Motivation

There are number of fusion strategies which are being formulated to get better fusion results. However, each fusion method has some major drawbacks, which are listed as follows.

- In case of misregistered source images, the shift-variant MST produces ringing artifacts in the fused image [89]. These ringing artifacts are also known as pseudo-Gibbs phenomenon. MGA techniques, such as NSCT [2], curvelet transform [126], and NSST [127] are employed to capture edge information of the source images. Both NSCT and NSST are shift invariant representations, and they can reduce pseudo-Gibbs phenomena. Moreover, the non-adaptive structure of most of the transform based fusion methods limits the quality of fusion results.

- The selection of decomposition level in MST and MGA methods is a crucial task. Spatial features of images cannot be extracted perfectly if less number of decomposition levels are used, while many decomposition levels produce noisy pixels and bad quality reconstruction when the source images are not perfectly registered. The performances of MST-based methods are mainly degraded due to non-perfect registration. Therefore, the effects of misregistration errors in the fusion performance and time complexity issues of the fusion framework need to be considered.

- The existing fusion methods are inefficient to perfectly seize significant low- and high-frequency information of input images because of some limitations of applied fusion rules. The traditional fusion rules are inefficient to capture all the significant attributes of the source images.

- The conventional sparse representation fusion methods suffer from few drawbacks, such as blurry representation of high-frequency information in the fused image. The sparse representation methods have limited ability in detailed information representation and they are highly sensitive to non-perfect registration.

- It is quite important to measure different attributes of the fused image to objectively measure the performance of different fusion techniques. Therefore, a metric which can efficiently measure different attributes of the fused image would be quite useful for comparison of different fusion techniques.

From the brief literature survey presented in this report, it is evident that a significant amount of work is needed for generation of high-quality fused images. Accordingly, this thesis looks into several aspects concerned with the extraction of low- and high-frequency image features, efficient metrics to measure fused image quality, addressing

of time complexity and misregistration error issues, and aims at developing suitable algorithms that can take care of some of the limitations of the existing methods.
The motivation behind this research work are given below:

(1) One of the most challenging and necessary tasks in image fusion is to preserve all the relevant information of the source images without introducing any spatial artifacts in the fused image. The source images are obtained from different imaging modalities, such as different image sensors, imaging principal, or optical directions. The resulting images have different spatial resolutions, disparity values, image details, and distortions. For example, computed tomography (CT) imaging convey information of hard tissues of the human body. Magnetic resonance (MR) imaging carries information of soft tissues of the human body. The slices of single photon emission computed tomography (SPECT) video are used to make a 3D model tissues. Hence, fusion of different medical source images gives an image which has significantly more information.

Similarly, in the case of visible-infrared image fusion, texture and spatial information of visible images are need to properly fuse with the radiation information of the infrared images. The fused result should be free from any spatial distortion. In case of multi-focus images, source images with different focus areas of the same scene need to be properly fused such that all the focus areas of the source images should appear in the fused images with minimum spatial artifacts.

(2) In image fusion literature, few metrics are proposed to measure specific information, such as edges, entropy, etc. However, more features of the source images need to be considered for judgment of visual qualities of the fused images. Hence, there is a wide scope to formulate such metrics which can simultaneously measure different attributes of the source images. Moreover, it is found in [123] that metrics reliability changes with different types of source images.

(3) In image fusion literature, very few works relating to the minimization of the effects of image registration errors for image fusion are presented. It is quite important to minimize misregistration errors so that the quality of the fused image can be improved.

(4) The problems of detailed information representation and shift-variance of sparse representation are resolved by using CSR. Hence, CSR can be used to improve image fusion performance.

(5) Image mosaicking is an image processing technique which is useful for tiling images. One important step in an image mosaicking framework is the auto-sorting algorithm, which is to be performed to minimize registration errors in the mosaic image. Another step in mosaicking is the detection of interest points for matching of the source images obtained after auto-sorting. However, in the presence of noisy and pseudo-periodic structures in the source images, the existing auto-sorting methods generally produce distortions in the final mosaic image. Secondly, most of the popular interest point detection algorithms do not specifically consider computational issues. So, these two issues need to be addressed.

This thesis introduces new methods for fusion of single and multi-modal source images, including new metrics to judge the visual qualities of the fused image. Additionally, the proposed framework is used to perform image enhancement and denoising. Finally, an efficient image mosaicking framework is developed.

1.4 Objectives

The goal of this research work is to develop fused images which contain significant information of source images. Additionally, fused images should be free from any undesirable artifacts. For this, different information or attributes (such as texture, smoothness, edges, contours, corners, etc.) of the source images needs to be captured. These information are available in low- and high-pass frequency bands of the source images. Most of these information can be captured by developing efficient fusion rules. It is also important to develop more flexible or adaptive methods for fusion as non-adaptiveness generally limit the performance of fusion performance. The source image misregistration error also limits the performance of fusion results by producing ringing artifacts. In image fusion literature, a number of metrics are proposed to measure fusion performance. However,

these metrics can not effectively measure fusion quality in some of the instances. Hence, there is a scope of proposing new metrics for comprehensive analysis of fused images.

1.5 Major Contributions

The main contributions of the thesis can be summarized as follows:

- In our proposed fusion framework, the low- and high-frequency coefficients are fused by using our proposed modified weighted salience and local difference fusion rules, respectively to preserve very important information of the source images in the fused images.
- The low-frequency bands are fused using the proposed descriptor obtained by superposition of scale multiplied Canny edge detector features and Hessian features. Moreover, the high-frequency bands are fused using fusion rules which are defined based on the basic principles of unsharp masking, and this is done to preserve relevant information of the source images in the fused images.
- The low-frequency bands are fused by CSR modeling. Moreover, the high-frequency bands are fused by our proposed information entropy, standard deviation, and range (IESDR) descriptor to preserve significant information of the source images in the fused images.
- A metric Q_{TS} is proposed to jointly measure both texture and structural information present in the fused images. To measure the variance information in the fused image, a metric Q_E is formulated on the basis of Karhunen-Loeve transform (KLT). Finally, a metric Q_{SS} is formulated to measure singularities and structural information of the fused image.
- The parameter of the Kaiser window can be adjusted to obtain optimized nonsubsampled Laplacian pyramid filters, and so this finally optimizes the NSST to obtain better fusion results.
- A modified Meyer window based adjustable nonsubsampled shearlet transform (ANSST) is proposed to make NSST more adaptive for image fusion and denoising.

- A local difference features based Harris interest point detector is proposed to produce mosaic image in a time efficient manner as compared to conventional Harris interest point detector while preserving the visual quality of mosaic images.

1.6 Summary

The prime objective of image fusion is to produce a more informative fused image as compared to the source images. A number of image fusion methods have been proposed in the last decade. This thesis discusses about existing fusion methods, their advantages and drawbacks, and applications of different fusion frameworks. The proposed fusion methods are formulated to address some of the limitations of the state-of-the-art methods.

The image fusion techniques are broadly divided into multi-scale decomposition based fusion, non-multi-scale decomposition based fusion, sparse representation based fusion, and hybrid techniques. Precise and accurate registration of source images is a very crucial task to get good quality fused images. Misregistration of source images leads to spatial artifacts, such as ringing or pseudo-Gibbs phenomenon in the fused images. The distortion and the presence of noise in the fused images are mainly due to the limitations of the fusion framework. The proper design of the fusion framework can minimize image distortion.

Image decomposition tools and fusion rules are two important aspects of a image fusion framework. Different image decomposition techniques are employed to perform image fusion. Although various image decomposition tools are being developed for image fusion, not much importance was given in the direction of development of efficient fusion rules. The fusion rules which use neighbourhood operation and pixel correlations, generally give better results. Different hybrids methods are also proposed to generate better quality fused images. These hybrid methods generally employ the tools like sparse representation, machine learning, edge-preserving filters, multi-scale and geometrical representation, etc. However, these methods also have some drawbacks, and hence, there is a scope of developing more efficient image fusion frameworks.

Another open area of research in the domain of image fusion is the development of quantitative metrics to judge visual qualities of the fused images. Almost all (except multi-focus images) image fusion algorithms do not have ground truth results to judge a fusion framework. Few metrics have been proposed in the literature to measure some specific attributes of the fused images. However, there is still a scope to formulate more versatile and efficient metrics to measure global and other attributes of the source images present in the fused images. It is also very important that the metrics can measure visual qualities of the fused images obtained using multi-modal and multi-sensor source images.

1.7 Outline of the Thesis

The thesis mainly addresses different aspects of fusion of medical, visible-infrared and multi-focus source images. In **Chapter 1**, we review previous works related to image fusion and highlight the strategies of different fusion algorithms. In this discussion, the major shortcomings of the most popular state-of-the-art methods are mentioned, and subsequently, the scope of this research work is finalized. The main objective of this research work is to develop efficient fusion algorithms to fuse multi-sensor source images. In view of this, a novel NSST-based image fusion framework is developed for fusion of medical and visible-infrared images in **Chapter 2**. For this, fusion rules are also proposed to select low- and high-frequency coefficients for fusion. In addition, a new metric Q_{TS} is proposed to measure texture and structural information of the fused images. In **Chapter 3**, we proposed an optimize NSST-based fusion framework. Additionally, a variance preservation metric Q_E is proposed, which is formulated using KLT to measure variance information of the fused images. In **Chapter 4**, a adjustable NSST (ANSST) and CSR-based image fusion framework is proposed to minimize the source image registration errors. Moreover, a metric Q_{SS} is formulated to measure singularity and structural information. In **Chapter 5**, we proposed a relatively simple and fast image mosaicking framework. In **Chapter 6**, we conclude this thesis and suggests a few directions for the future scope.

2

Non-subsampled Shearlet Transform-based Image Fusion using Modified Weighted Saliency and Local Difference

Existing image fusion methods can not efficiently capture significant edges, texture and fine details of the source images due to inefficient fusion framework. In addition, for objective evaluation of fusion algorithms, not much attention is given to simultaneously measure both texture and structural information of the source images which are preserved in the fused image. To address these issues, NSST is used to decompose pre-registered source images into low- and high-frequency components. These low- and high-frequency coefficients are fused by using our proposed modified weighted salience and local difference fusion rules, respectively. To enrich edge information in the fused image, Canny edge detector with scale multiplication is employed. Moreover, a metric Q_{TS} is proposed to jointly measure both texture and structural information present in the fused image. The proposed metric is formulated on the basis of local standard deviation filtering, local information entropy, and local difference filtering. Both subjective and objective results validate the proposed fusion framework and the metric Q_{TS}.

2.1 Introduction

Image fusion is a technique to assimilate information acquired from similar or dissimilar sources of images. Compared to the source images, fused images convey more information, and they have more clarity. Image fusion techniques are employed to combine images from diverse applications, such as biomedical, multi-focus, multi-exposure, infrared imaging, etc.

The method presented in [128] uses the concept of total variation and gradient transfer for visible-infrared image fusion. This method sharpens infrared images to emphasize texture information in the fused images. A detailed review of different fusion methods employed for fusion of visible and infrared images is presented in [129]. Image fusion significantly improves the clarity of background and foreground of a scene of multi-focus images.

In spatial domain, artificial neural network (ANN) and convolutional neural network (CNN) based image fusion methods were developed and these methods showed promising results [118, 130]. In [130], multi-focus image fusion problem was interpreted as a classification problem. In [118], original images are given as input to CNN. Hence, the network used in this case is deeper as compared to the method proposed in [130]. The main advantage of this method is the use of CNN model for jointly performing activity level measurements and fusion rules generation.

Shearlet based transform domain image fusion methods are introduced in [131, 132]. In [132], averaging and absolute value maximum selection fusion rules are employed. Due to shift variant nature of the shearlet, this method produces ringing artifacts in the fused images. The ringing artifacts can be removed by using NSST [131]. Also, pixel-level fused image sufferers from spatial distortions due to direct operations on pixels.

To deal with above-mentioned drawbacks of the existing image fusion methods, NSST is employed for fusion in our proposed method. An advantage of NSST is that there are no restrictions on the number of directions for shearing and size of the supports. On the other hand, directional filter banks needs to be constructed for contourlet. Furthermore, inversion of NSST only needs a summation of shift-invariant shearing filters rather than

inverting the directional filter banks as in the case of contourlet, which results in computationally efficient implementation. Moreover, curvelet suffers from the implementation issues in the discrete domain [31]. NSCT is a combination of non-subsampled Laplacian pyramid and non-subsampled directional filter bank. While NSST is a combination of non-subsampled Laplacian pyramid and different combinations of shift-invariant shearing filters. Due to higher complexity of NSCT over NSST, NSST is chosen in our proposed method.

In our method, anisotropic diffusion edge preserving filtering is applied on high-frequency bands to reduce noises of the fused image. The main advantage of anisotropic diffusion filtering is that it removes noises while preserving the edges. A novel fusion rule is proposed to fuse high-frequency coefficients of the transformed images, which is based on local pixel difference. Our proposed local difference rule is able to extract high-frequency information by considering absolute differences of the magnitude of the high-frequency coefficients in a neighborhood.

To fuse low-frequency bands, weighted salience method [133] is modified by including local difference features in the form of energy. Experimentally, it is observed that the local difference featuers are more suitable for fusion of high-frequency bands as compared to sum-of-modified-Laplacian (SML) [2], spatial frequency (SF) [134] and gradient features. Moreover, Canny edge detector with scale multiplication technique is used to seize prominent edge information of the source images [135].

Additionally, there is a need of an effective metric to measure preserve texture and structural features of the fused image, which are transferred from the source images. Thus, a novel metric Q_{TS} is proposed to find the amount of texture and structural information of the source images which is preserved in the fused image. The metric Q_{TS} is based on three descriptors: local standard deviation, local information entropy, and local difference.

The two texture descriptors, local information entropy and local standard deviation are used to measure both regular and random texture detail, and brightness information. The local difference features are also incorporated in our proposed metric to measure structural information of the images. Therefore, proposed metric Q_{TS} measures both

texture and structural information present in the fused image. The proposed metric can judge the quality of fused images obtained from different image fusion algorithms. Validation of metric Q_{TS} is done by finding the effects of free parameters on the proposed fusion framework and using the method presented in [124]. Experimental results show the efficacy of our proposed image fusion framework and the proposed metric Q_{TS}.

The rest of this chapter is organized as follows: The fundamental concepts of NSST, Canny edge detector with scale multiplication, anisotropic diffusion, local standard deviation filtering and local information entropy filtering are described in Section 2.2. In Section 2.3, our proposed local difference fusion rule for high-frequency coefficients, proposed modified weighted salience fusion rule for low-frequency coefficients, and the proposed fusion framework is elaborately described. The concept of the proposed texture and structural information preservation metric (Q_{TS}) is explained in Section 2.4. Also, all the experimental results are presented in this section. At last, concluding remarks are given in Section 2.5.

2.2 Preliminaries

2.2.1 Non-subsampled shearlet transform (NSST)

Shearlet transform (ST) produces a set of well-localized waveforms at various scales, locations, and orientations [31]. The number of trapezoids is higher in shearlet as compared to curvelet transform and contourlet transform. As discussed in the previous section, shift invariant NSST [31] is used to efficiently represent high- and low-frequency information of source images. Shearlet makes a set of well-localized waveforms at different scales, locations, and orientations. The number of oriented trapezoids is more in shearlet as compared to curvelet and contourlet. In NSST, the source image is decomposed into low-pass and high-pass sub-images using non-subsampled Laplacian pyramid. High-pass sub-images are fed to shift-invariant shear filters and the low-pass sub-image is further decomposed into low-pass and high-pass sub-images. The inverse NSST transform is applied by taking the sum of all shift invariant shearing filter responses at the respective levels of decomposition, and inverse non-subsampled Laplacian pyramid is finally

applied to get the source image. NSST has number of attributes, such as localization in frequency and spatial domain, optimally sparse and directional sensitivity with parabolic scaling. For dimension $n = 2$, continuous shearlet transform of a signal $f(x)$ is defined by a mapping as follows:

$$SH_{\psi}f(a, s, t) = \langle f, \psi_{ast} \rangle \tag{2.1}$$

where a is scale, s is orientation, t is location, $\psi_{ast}(x) = |\det \mathbf{M}_{as}|^{-1/2}\psi(\mathbf{M}_{as}{}^{-1}(x-t)), \mathbf{M}_{as} = \begin{pmatrix} a & -\sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix}$ for $a > 0$, $s \in R$ and $t \in R^2$. $\psi_{ast}(x)$ are called shearlets. Matrix $\mathbf{M}_{as} = \mathbf{B}_s\mathbf{A}_a$ is linked with two different matrices: anisotropic dilation done by matrix $\mathbf{A}_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ and shearing done by matrix $\mathbf{B}_s = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}$. Here, ψ is localized function with appropriate admissibility conditions, and conditions are ψ_{ast} should be uniformly smooth, localized in space and possess anisotropic moments or vanishing moments.

2.2.2 Canny edge detector with scale multiplication

Canny edge detector is very efficient in detecting edges of an image. However, canny edge detector misses sharp discontinuities as it uses less efficient detection and localization criteria. To attain better edge detection performance, Bao et. al proposed a modified Canny edge detector with scale multiplication [135]. The selection of proper scale for detection filter is very crucial for detecting valid edge pixels as large scale filters are robust to noise, but such a filter may remove fine details of an image, and vice versa. By judiciously selecting the size of the scale filter, and by applying multi-scale approach, localization accuracy can be improved.

In scale multiplied Canny edge detection method, a scale function is used, which is the product of detection filter responses obtained at two scales. In this method, single threshold technique is employed to obtain edge maps. Moreover, detection criteria results are improved by introducing non-linearity. Utilization of both detection and localization criteria improve edge detection performance of Canny edge detector. In our work, if we

apply scale multiplied Canny edge detector method to an image $\mathbf{I}(i,j)$, then the output edge image is represented as $\mathbf{SMC}_I(i,j)$.

2.2.3 Anisotropic diffusion

Anisotropic diffusion is an edge preserving noise removal technique, which smooth an image without smoothing the structural information such as edges [136]. In discrete domain, anisotropic diffusion is given as follows:

$$\mathbf{I}_{t+1}(x',y') = \mathbf{I}_t(x',y') + \lambda(c'_n d'_n + c'_s d'_s + c'_e d'_e + c'_w d'_w) \tag{2.2}$$

where, c is flux function or diffusion rate, $\mathbf{I}$ is an image, t is iteration or time or scale, x' and y' are discrete coordinates, λ is a integration or stability constant. Subscripts n', s', e' and w' indicate north, south, east and west nearest neighbour, respectively. Also d'_n, d'_s, d'_e and d'_w are difference between central pixel and the four nearest neighbours, while c'_n, c'_s, c'_e and c'_w are corresponding conduction coefficients. The values of d'_i and c'_i are evaluated at every iteration. The values of coefficients c'_i is obtained as follows:

$$c'_i = e^{-(d'_i/k)^2} \tag{2.3}$$

where, k is a conduction coefficient which control the effect of d'_i on c'_i.

2.2.4 Local standard deviation filtering and local information entropy

In local standard deviation filtering, patches of any sizes are considered and standard deviation of a local window is calculated, and finally it is assigned to the center pixel of the respective patch. At the boundary of images, mirror symmetric extension of pixels is carried to prevent distortion of boundary pixels. The standard deviation gives contrast and brightness information. The higher standard deviation indicates better contrast and brightness.

Local information entropy indicates unpredictability of a local region [137]. The regions with highly varying information tend to have a uniform distribution, which results

in higher entropy, and vice-versa. The regions having more intensity variations indicate point, edge, contours and texture details of an image.

2.3 Proposed Method

The schematic diagram of the proposed method is shown in Fig. 2.1. In our method, NSST is applied to decompose source images $\mathbf{A}$ and $\mathbf{B}$ into low-frequency bands $\mathbf{L}_A(m,n)$ and $\mathbf{L}_B(m,n)$, and high-frequency bands $\mathbf{H}^A_{k,l}(m,n)$ and $\mathbf{H}^B_{k,l}(m,n)$. Here, k and l indicate decomposed high-frequency band in k^{th} direction at l^{th} decomposition level. Applying the proposed fusion rules and Canny edge detector with scale multiplication operation, the final fused image $\mathbf{F}$ is obtained[1].

2.3.1 Proposed local differences

To capture features from high-frequency bands, difference of center pixel to its neighbourhood pixels are considered. In our method, horizontal and vertical neighbours of the pixel are also considered for extracting features. The proposed local difference features are extracted as given in Eq. (2.4). The local difference fusion rule captures information of horizontal, vertical, diagonal edges, anisotropic or curve like features and point features.

Let $\mathbf{F}(i,j)$ be an image of size $M \times N$ and window size is $w_1 \times w_2$. By applying local difference filtering, we get an image $\mathbf{LD}^F(i,j)$. In Eq. (2.4), the window size is 3×3,

[1]This work has been published in Multimedia Tools and Applications, Springer 2018 (Refer item 4 for details).

but this rule can be extended for large windows in a similar way.

$$\begin{aligned}\mathbf{LD}^F(i,j) = &\left\{\sum_{w_1,w_2=-1}^{1} |\mathbf{F}(i,j) - \mathbf{F}(i+w_1, j+w_2)|\right\} \\ &+ |\mathbf{F}(i-1,j-1) - \mathbf{F}(i-1,j)| \\ &+ |\mathbf{F}(i-1,j) - \mathbf{F}(i-1,j+1)| \\ &+ |\mathbf{F}(i-1,j+1) - \mathbf{F}(i,j+1)| \\ &+ |\mathbf{F}(i,j+1) - \mathbf{F}(i+1,j+1)| \\ &+ |\mathbf{F}(i+1,j+1) - \mathbf{F}(i+1,j)| \\ &+ |\mathbf{F}(i+1,j) - \mathbf{F}(i+1,j-1)| \\ &+ |\mathbf{F}(i+1,j-1) - \mathbf{F}(i,j-1)| \\ &+ |\mathbf{F}(i,j-1) - \mathbf{F}(i-1,j-1)|\,.\end{aligned} \tag{2.4}$$

2.3.2 Modified weighted salience

The weighted salience method is used to fuse decomposed gradient pyramid bands [133]. In this fusion method, salience regions of the source images are considered for fusion. This salience is obtained by finding energy of local regions of an image. However, this weighted salience rule basically gives preference to those pixels whose intensity values are higher. Thus, the resultant image contains most of the high-intensity pixels. However, in case of inter-modality medical image fusion and visible-infrared image fusion, variations of pixels intensity within a local region also need to be considered for better visualization.

High energy within local regions of an image indicates presence of high-intensity values, but the energy measure can not efficiently seize the variational information of the pixel values in the local regions. Hence, the salience measure is modified in our proposed method by incorporating local difference features with the energy measure. The local difference features consider changes in the pixel values in a local region. In our method, modified weighted salience-based fusion rule is applied. The modified salience measure obtained from energy and local difference features can efficiently capture texture, fine details, and structural information.

Modified weighted salience-based fusion rule is implemented in two steps: *modified salience measure* and *match measure*. The modified salience is measured at $\mathbf{L}(m,n)$ as

local energy and local difference features, where $\mathbf{L}$ is the low-frequency band of source image $\mathbf{I}$, and m, n are discrete coordinates. The modified salience pattern $\mathbf{S}_I(m,n)$ within a window of size $\hat{m} \times \hat{n}$ is given as:

$$\mathbf{S}_I(m,n) = \sum_{\hat{m},\hat{n}} (\mathbf{L}_I(m+\hat{m}, n+\hat{n})^2 + \mathbf{LD}^L(m+\hat{m}, n+\hat{n})^2) \tag{2.5}$$

If modified salience measure $\mathbf{S}_I(m,n)$ of the corresponding pixels of both low-frequency bands $\mathbf{L}_A$ and $\mathbf{L}_B$ are equal, then their average is considered. The pixel with maximum modified salience measure is considered if modified salience measures are different. Furthermore, match measure is used for taking a decision for selection or averaging. The relative amplitude of the modified salience measure is generally considered to select the pixels of the fused image. Alternatively, correlation between low-frequency bands can be used. At the sample point (m,n) within a window, local normalized correlation is given as [133]:

$$\mathbf{M}_{AB}(m,n) = \frac{\sum_{\hat{m},\hat{n}} \mathbf{L}_A(m+\hat{m}, n+\hat{n}) \mathbf{L}_B(m+\hat{m}, n+\hat{n})}{0.5(\mathbf{S}_A(m,n) + \mathbf{S}_B(m,n))} \tag{2.6}$$

$\mathbf{M}_{AB}$ has a value of 1 for the identical patterns. For remaining patterns, $\mathbf{M}_{AB}$ is less than 1. If the match at the respective sample of $\mathbf{L}_A$ and $\mathbf{L}_B$ is high (if greater than threshold α), then the average of the decomposed coefficient is taken. Otherwise, coefficients with high modified salience measure are considered if the match is lower than threshold α.

In summary, if $\mathbf{M}_{AB}(m,n) < \alpha$, then $\mathbf{W}_{min}(m,n) = 0$ for lower modified salience measure and $\mathbf{W}_{max}(m,n) = 1$ for higher modified salience measure in Eq. (2.8), otherwise if $\mathbf{M}_{AB}(m,n) \geq \alpha$ then $\mathbf{W}_{min}(m,n)$ and $\mathbf{W}_{max}(m,n)$ are obtained as:

$$\begin{aligned} \mathbf{W}_{\min}(m,n) &= (\mathbf{M}_{AB}(m,n) - \alpha)(2-2\alpha)^{-1} \\ \mathbf{W}_{\max}(m,n) &= 1 - \mathbf{W}_{\min}(m,n) \end{aligned} \tag{2.7}$$

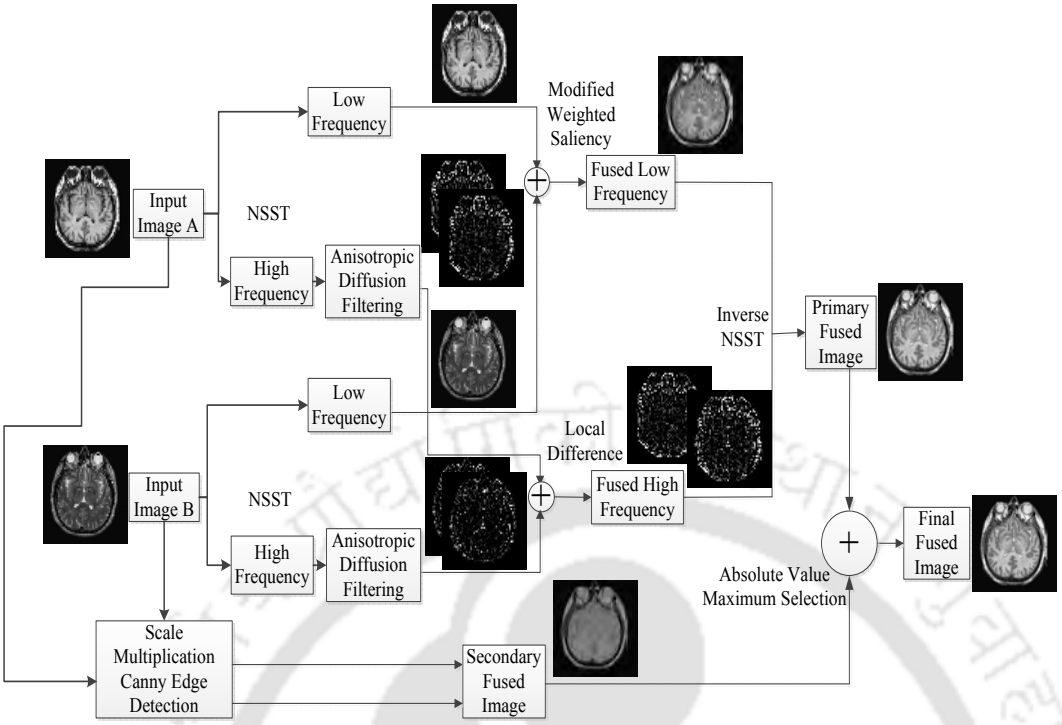


FIGURE 2.1: Schematic diagram of the proposed image fusion method.

2.3.3 Proposed fusion framework

The proposed fusion rule uses NSST[2] for image decomposition. The source images **A** and **B** are first decomposed into low- and high-frequency bands.

Low-frequency band fusion rule: In our proposed method, modified weighted salience method is used to fuse low-frequency bands. The low-pass bands having higher value of modified salience measure and larger weight $\mathbf{W}_{max}(m,n)$ are assigned to respective low-pass bands. If $\mathbf{S}_A(m,n) > \mathbf{S}_B(m,n)$, then $\mathbf{W}_A(m,n) = \mathbf{W}_{max}(m,n)$ and $\mathbf{W}_B(m,n) = \mathbf{W}_{min}(m,n)$, otherwise $\mathbf{W}_A(m,n) = \mathbf{W}_{min}(m,n)$ and $\mathbf{W}_B(m,n) = \mathbf{W}_{max}(m,n)$. At each sample position of $\mathbf{L}_A(m,n)$ and $\mathbf{L}_B(m,n)$, weights $\mathbf{W}_A(m,n)$ and $\mathbf{W}_B(m,n)$ are assigned to the low-frequency bands and combined as follows:

$$\mathbf{L}_F(m,n) = \mathbf{W}_A(m,n)\mathbf{L}_A(m,n) + \mathbf{W}_B(m,n)\mathbf{L}_B(m,n) \tag{2.8}$$

where, $\mathbf{L}_F(m,n)$ is the fused low-frequency band.

[2]The related work has been published in IEEE International Conference on Wireless Communication, Signal Processing and Networking (WiSPNET) 2017 (Refer item 4 for details).

High-frequency band fusion rule: Prior to fuse high-frequency bands, anisotropic diffusion filtering is applied to reduce noise. Then, proposed local difference-based fusion rule is used for fusion. Local difference fusion rules enhance edge features by considering more number of absolute differences in an window in contrast to gradient, sum of modified Laplacian (SML) [2] and spatial frequency (SF) [134] based fusion rules. In each high-frequency bands $\mathbf{H}^{A}_{k,l}(m,n)$ and $\mathbf{H}^{B}_{k,l}(m,n)$, Eq. (2.4) is applied, and local difference feature images $\mathbf{LD}^{A}_{k,l}(m,n)$ and $\mathbf{LD}^{B}_{k,l}(m,n)$ are obtained. Then by using Eq. (2.9), selection of fused coefficients $\mathbf{H}^{F}_{k,l}(m,n)$ are carried out as follows:

$$\mathbf{H}^{F}_{k,l}(m,n) = \begin{cases} \mathbf{H}^{A}_{k,l}(m,n), & \prod_{k,l} \mathbf{LD}^{A}_{k,l}(m,n) \geq \prod_{k,l} \mathbf{LD}^{B}_{k,l}(m,n) \\ \mathbf{H}^{B}_{k,l}(m,n), & \text{Otherwise.} \end{cases} \tag{2.9}$$

where, the product symbol is the Hadamard product [138] of all local difference feature images of decomposed high-frequency bands. The Hadamard product of all local difference images are considered as multiplication of different bands at different scales and orientation.

The previous image fusion methods consider high-frequency fusion rules based on gradient, SML and SF. The high-frequency information is not properly extracted as these operations do not consider all the neighbourhood spatial information. In most of the previous methods [4, 89], low-frequency fusion rules select higher magnitude coefficients, which can not perfectly capture sharpness and texture details. So, we proposed the fusion framework mainly to address the above mentioned issues.

In our proposed method, local difference fusion rule can effectively highlight edges, corners, and texture details as more information from the neighbourhood pixels is employed. Our modified weighted salience-based fusion rule effectively extracts information of brightness, illumination and contrast from low-frequency coefficients. In our method, the proposed energy measure can efficiently carry forward variance and structural information of the source images into the fused images as compared to the previous methods. The multi-scale representation of the images in our proposed method highlights edge, corner and fine features of the images more effectively as all the high-frequency bands

of different scales and orientations are multiplied. The main steps of our proposed algorithm for pre-registered source images are given below.

Proposed algorithm

STEP 1: The pre-registered source images are decomposed using NSST.

STEP 2: The low-frequency band is fused using modified weighted salience-based fusion method, and after applying anisotropic diffusion filtering on high-frequency bands, the high-frequency bands are fused by our proposed local difference method.

STEP 3: The primary fused image $\mathbf{F}_I(i,j)$ is obtained by taking inverse NSST.

STEP 4: Canny edge detector with scale multiplication is applied on the source images $\mathbf{A}(i,j)$ and $\mathbf{B}(i,j)$, and the feature images $\mathbf{SMC}_A(i,j)$ and $\mathbf{SMC}_B(i,j)$ are obtained. The pixels of the feature images are compared to get the pixels of the source images having prominent edge information as follows:

$$\mathbf{S}_{SMC}(i,j) = \begin{cases} \mathbf{A}(i,j), & \text{if} \quad \mathbf{SMC}_{\mathrm{A}}(i,j) > \mathbf{SMC}_{\mathrm{B}}(i,j) \\ \mathbf{B}(i,j), & \text{if} \quad \mathbf{SMC}_{\mathrm{A}}(i,j) < \mathbf{SMC}_{\mathrm{B}}(i,j) \\ \frac{\mathbf{A}(i,j)+\mathbf{B}(i,j)}{2}, & \text{if} \quad \mathbf{SMC}_{\mathrm{A}}(i,j) = \mathbf{SMC}_{\mathrm{B}}(i,j) \end{cases} \tag{2.10}$$

STEP 5: The primary fused image $\mathbf{F}_I(i,j)$ and the secondary fused image $\mathbf{S}_{SMC}(i,j)$ are fused using absolute value maximum selection fusion rule. This rule selects the pixels having higher intensity values, so that prominent details of $\mathbf{F}_I(i,j)$ and $\mathbf{S}_{SMC}(i,j)$ can be preserved. The final fused image $\mathbf{F}_F(i,j)$ is obtained as:

$$\mathbf{F}_F(i,j) = \begin{cases} \mathbf{S}_{SMC}(i,j), & \text{if} \quad \mathbf{S}_{\mathrm{SMC}}(i,j) \geq \mathbf{F}_{\mathrm{I}}(i,j) \\ \mathbf{F}_I(i,j), & \text{Otherwise.} \end{cases} \tag{2.11}$$

2.4 Experimental Results

To carry out experiments, standard images are obtained from [89, 139, 140]. The proposed method is compared with the state-of-the-art methods proposed for fusion [2–4, 24, 89, 114, 118].

2.4.1 Metrics for performance evaluation

The proposed method is tested using four fusion metrics, where three metrics are commonly used, and the fourth one is the proposed texture and structural information preservation metric Q_{TS}. Each metric measures specific aspects of the fused image with respect to the source images.

Normalized mutual information (Q_{MI}) represents the amount of information transferred from source images into the fused image [123, 141]. Structure similarity based index (Q_S) is a measure of structure, luminance, and contrast information in fused image [125]. Edge-based similarity measure ($Q^{AB/F}$) measures edge information transferred from the source images to the fused image after the fusion process [122].

Proposed texture and structural similarity metric (Q_{TS}): For measuring texture and structural similarity between original and fused images, a novel texture and structural information similarity measure is proposed. In this, texture information is obtained using local information entropy; brightness and structural information is obtained using local standard deviation and local difference method given in Eq. (2.4). Local information entropy indicate texture information of an image. The local high textured image regions give higher entropy values as compared to the plain image regions. The proposed metric Q_{TS} combines entropy, standard deviation and local difference features, and it effectively measures how much texture and structural information of the source images are preserved in the fused image.

Let us take two source images **A**, **B** and a fused image **F**. Local standard deviation filtering and local information entropy are applied to get $\mathbf{S}(m,n)$ and $\mathbf{R}(m,n)$ features for each of the pixels of the images as given in Eqs. (2.12) and (2.13). For images of size

$M \times N$ $(1 \leq m \leq M$ and $1 \leq n \leq N)$ and the window $\mathbf{P}$ of size $x \times y$, the square of local standard deviation filtering applied on image $\mathbf{A}$, and it is given by:

$$\mathbf{S}_A^2(m,n) = \left[Z \sum_{x,y \in \mathbf{P}} \left[\mathbf{A}(m+x, n+y) - Z \sum_{x,y \in \mathbf{P}} \mathbf{A}(m+x, n+y) \right]^2 \right] \tag{2.12}$$

where, $Z = \frac{1}{xy-1}$. Local information entropy of an image $\mathbf{A}$ is given as:

$$\mathbf{R}_A(m,n) = \sum p(\mathbf{A}(m,n)) \log_2 p(\mathbf{A}(m,n)) \tag{2.13}$$

where, $p(\mathbf{A}(m,n))$ is the probability occurrence of the coefficient $\mathbf{A}(m,n)$. Relative local standard deviation and local information entropy images $\mathbf{S}^{AF}(m,n)$ and $\mathbf{R}^{AF}(m,n)$ of the input image $\mathbf{A}$ with respect to the fused image $\mathbf{F}$ are obtained as follows:

$$\mathbf{S}^{AF}(m,n) = \begin{cases} \frac{\mathbf{S}_F(m,n)}{\mathbf{S}_A(m,n)}, & \text{if} \quad \mathbf{S}_A(m,n) > \mathbf{S}_F(m,n) \\ \frac{\mathbf{S}_A(m,n)}{\mathbf{S}_F(m,n)}, & \text{otherwise} \end{cases} \tag{2.14}$$

$$\mathbf{R}^{AF}(m,n) = \begin{cases} \frac{\mathbf{R}_F(m,n)}{\mathbf{R}_A(m,n)}, & \text{if} \quad \mathbf{R}_A(m,n) > \mathbf{R}_F(m,n) \\ \frac{\mathbf{R_A}(m,n)}{\mathbf{R_F}(m,n)}, & \text{otherwise} \end{cases} \tag{2.15}$$

$$\mathbf{LD}^{AF}(m,n) = \begin{cases} \frac{\mathbf{LD}^F(m,n)}{\mathbf{LD}^A(m,n)}, & \text{if} \quad \mathbf{LD}^A(m,n) > \mathbf{LD}^F(m,n) \\ \frac{\mathbf{LD}^A(m,n)}{\mathbf{LD}^F(m,n)}, & \text{otherwise} \end{cases} \tag{2.16}$$

where, $\mathbf{LD}^A(m,n)$ and $\mathbf{LD}^F(m,n)$ are obtained by applying local difference filtering given in Eq. (2.4) on image $\mathbf{A}$ and $\mathbf{B}$. Texture and structural preservation values are given as:

$$\mathbf{Q}^{AF}(m,n) = \mathbf{Q}_S^{AF}(m,n) \mathbf{Q}_R^{AF}(m,n) \mathbf{Q}_{LD}^{AF}(m,n) \tag{2.17}$$

where,

$$\mathbf{Q}_S^{AF}(m,n) = \left(1 + e^{-\left(\mathbf{S}^{AF}(m,n)\right)}\right)^{-1} \tag{2.18}$$

$$\mathbf{Q}_R^{AF}(m,n) = \left(1 + e^{-\left(\mathbf{R}^{AF}(m,n)\right)}\right)^{-1} \tag{2.19}$$

$$\mathbf{Q}_{LD}^{AF}(m,n) = \left(1 + e^{-\left(\mathbf{LD}^{AF}(m,n)\right)}\right)^{-1} \tag{2.20}$$

$\mathbf{S}^{AF}(m,n)$, $\mathbf{R}^{AF}(m,n)$ and $\mathbf{LD}^{AF}(m,n)$ are used to derive preservation values.

TABLE 2.1: The objective image fusion performance metrics for subjective similarity

Metric	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{TS}
relevance (r)	0.648	0.742	0.723	**0.769**
Correct Ranking (CR)	61.7%	69.2%	68.35%	**71.95%**

$\mathbf{Q}_S^{AF}(m,n)$, $\mathbf{Q}_R^{AF}(m,n)$ and $\mathbf{Q}_{LD}^{AF}(m,n)$ indicate perceptual loss of information in the fused image $\mathbf{F}$, which measure how much texture and structural detail of the source images is present in the fused image. $\mathbf{Q}^{AF}(m,n)$ lies between 0 to 1. If $\mathbf{Q}^{AF}(m,n) = 0$, then it corresponds to complete loss of texture and structural information, which is transferred from $\mathbf{A}$ to $\mathbf{F}$ at location (m,n). While $\mathbf{Q}^{AF}(m,n) = 1$ indicates fusion from $\mathbf{A}$ to $\mathbf{F}$ with no loss of texture and structural information. Using $\mathbf{Q}^{AF}(m,n)$ and $\mathbf{Q}^{BF}(m,n)$, our proposed normalized weighted performance metric (Q_{TS}) for $M \times N$ size images is given as follows:

$$Q_{TS} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}\left[\mathbf{Q}^{AF}(m,n)\mathbf{W}^{A}(m,n)+\mathbf{Q}^{BF}(m,n)\mathbf{W}^{B}(m,n)\right]}{\sum_{n=1}^{N}\sum_{m=1}^{M}\left[\mathbf{W}^{A}(m,n)+\mathbf{W}^{B}(m,n)\right]} \tag{2.21}$$

where, $\mathbf{W}^A(m,n)$ and $\mathbf{W}^B(m,n)$ are weights assigned to $\mathbf{Q}^{AF}(m,n)$ and $\mathbf{Q}^{BF}(m,n)$, respectively. The weights are chosen as $\mathbf{W}^A(m,n) = \mathbf{S}_A(m,n)$ and $\mathbf{W}^B(m,n) = \mathbf{S}_B(m,n)$. For optimal fusion results, metric Q_{MI} should be high, values of Q_S, $Q^{AB/F}$ and Q_{TS} should be close to 1.

Validation of proposed Q_{TS} metric: The proposed texture and structural similarity metric Q_{TS} is validated using the Correct Ranking (CR) and subjective relevance (r) measures [124]. The dataset-1 shown in Fig. 2.2 is used to perform this validation task. For validation, the r measure should be close to 1 and CR should be close to 100%. The r and CR values in Table 2.1 show that the best performance is obtained for Q_{TS}. Hence, this analysis shows that our proposed metric Q_{TS} can be used to effectively compare different image fusion algorithms. However, this metric is not suitable for measuring the performance of source images having low-texture and low-contrast information such as infrared images, night-vision scenario images, etc.

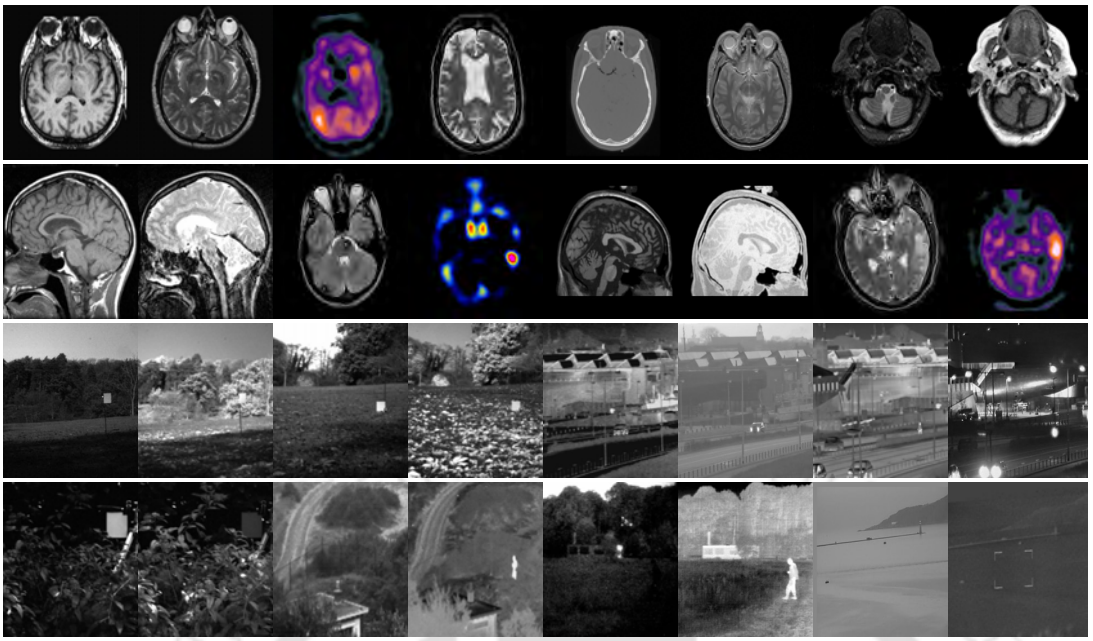


FIGURE 2.2: Image dataset-1 of eight pairs of medical and eight pairs of visible-infrared images.

2.4.2 Discussion on free parameters

Performance of the proposed fusion method is evaluated for different parameter settings. The dataset-1 of 16 pairs of source images shown in Fig. 2.2 is used for analysis of free parameters, and the average values of metrics, *i.e.*, Q_{MI}, Q_S, $Q^{AB/F}$ and Q_{TS} are calculated. The default setting of parameters are $\lambda = 1/7$, $k = 30$, $t = 15$, $\alpha = 10$ and Canny edge detector with scale multiplication $scale = 3$. In Fig. 2.3, the parameter k is varied and the other parameters are set to default values. In a similar way, variations with respect to t, λ, α and $scale$ are evaluated. The best parameters value of k is 1.18 with 2 iterations, $\lambda = 0.14$ and $\alpha =$ -0.7 are chosen. This parameter setting is finally used to generate fused images by our proposed method.

2.4.3 Experimentation and analysis

The presented work uses three datasets to perform comparative analysis. Dataset-1 shown in Fig. 2.2 is employed to find optimum values of different parameters used in our algorithm. Finally, it is also used for comparative studies of different fusion methods and different non-subsampled Laplacian pyramid filters of NSST. Dataset-1 (Fig. 2.2)

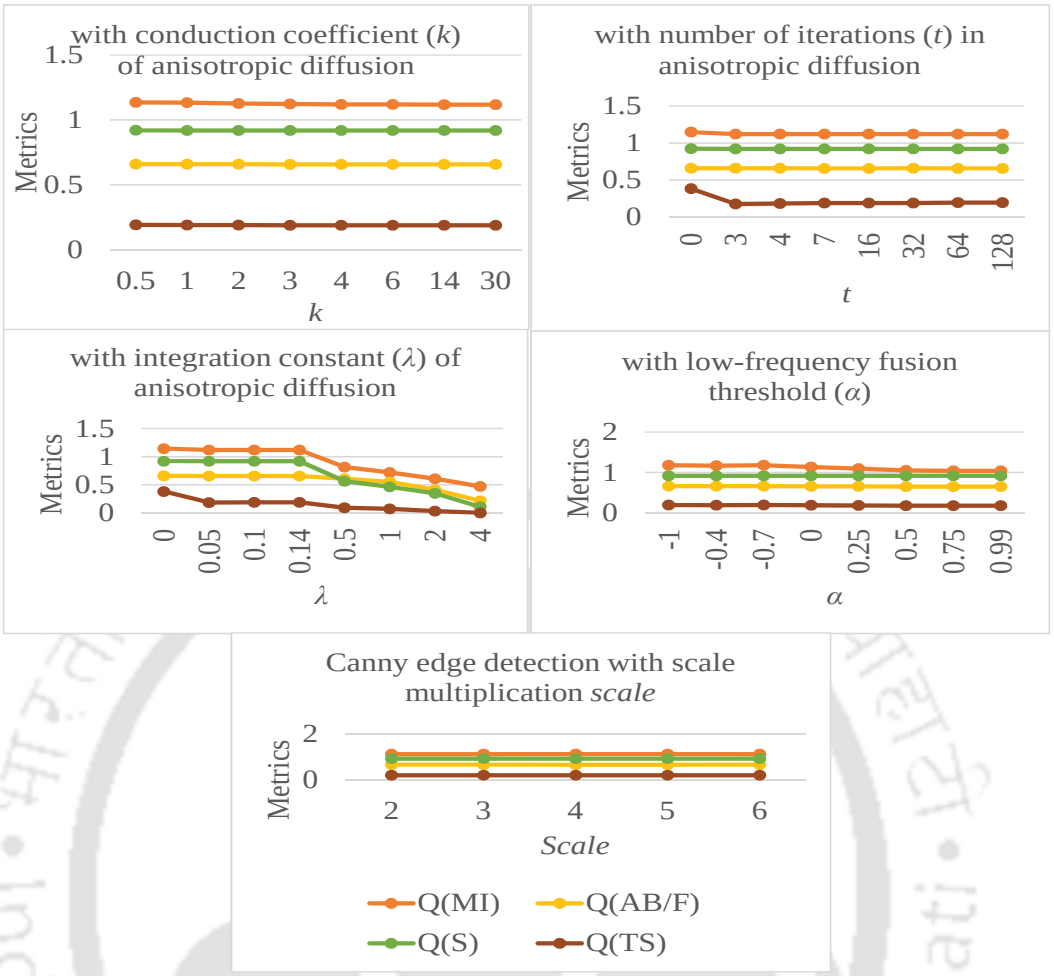


FIGURE 2.3: Performance of proposed method with the variations of free parameters, k, t, λ, α, and *scale*.

consists of 8 pairs of medical images and 8 pairs of visible-infrared images. Dataset-2 consists of 4 pairs of natural colour images as shown in Fig. 2.8 and it is used for comparative study of different colour spaces. Dataset-3 consists of 4 gray scale ground truth images as shown in Fig. 2.10 and it is used to compare different image enhancement methods, such as Gradient, SML, SF, and LD. After analyzing different non-subsampled Laplacian pyramid filters metrics responses, the *pyr* non-subsampled Laplacian pyramid filters is chosen for NSST, and it is applied for image fusion and image enhancement. For fusion of colour images, YCbCr model is used in our proposed method.

The state-of-the-art methods are NSCT and phase congruency (NSCT SML) [2], stationary wavelet transform and NSCT (SWT NSCT) [3], Type-2 fuzzy approach (NSCT fuzzy) [4], discrete wavelet transform (DWT) [24], multi-scale transform and sparse representation (MST SR) [89], guided filtering (GFF) [114], and convolutional neural network

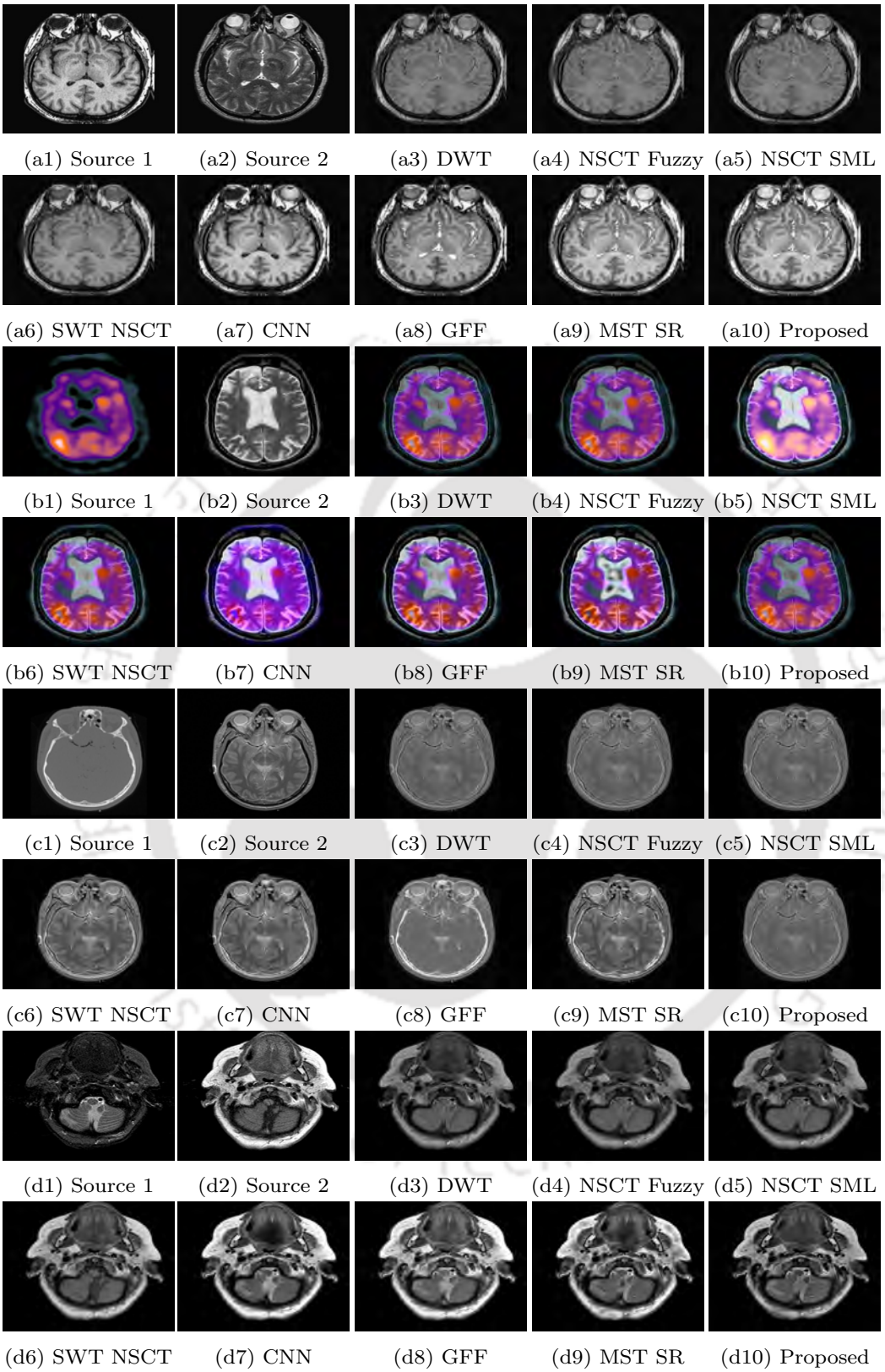


FIGURE 2.4: Source and fused medical images obtained by using different fusion methods.

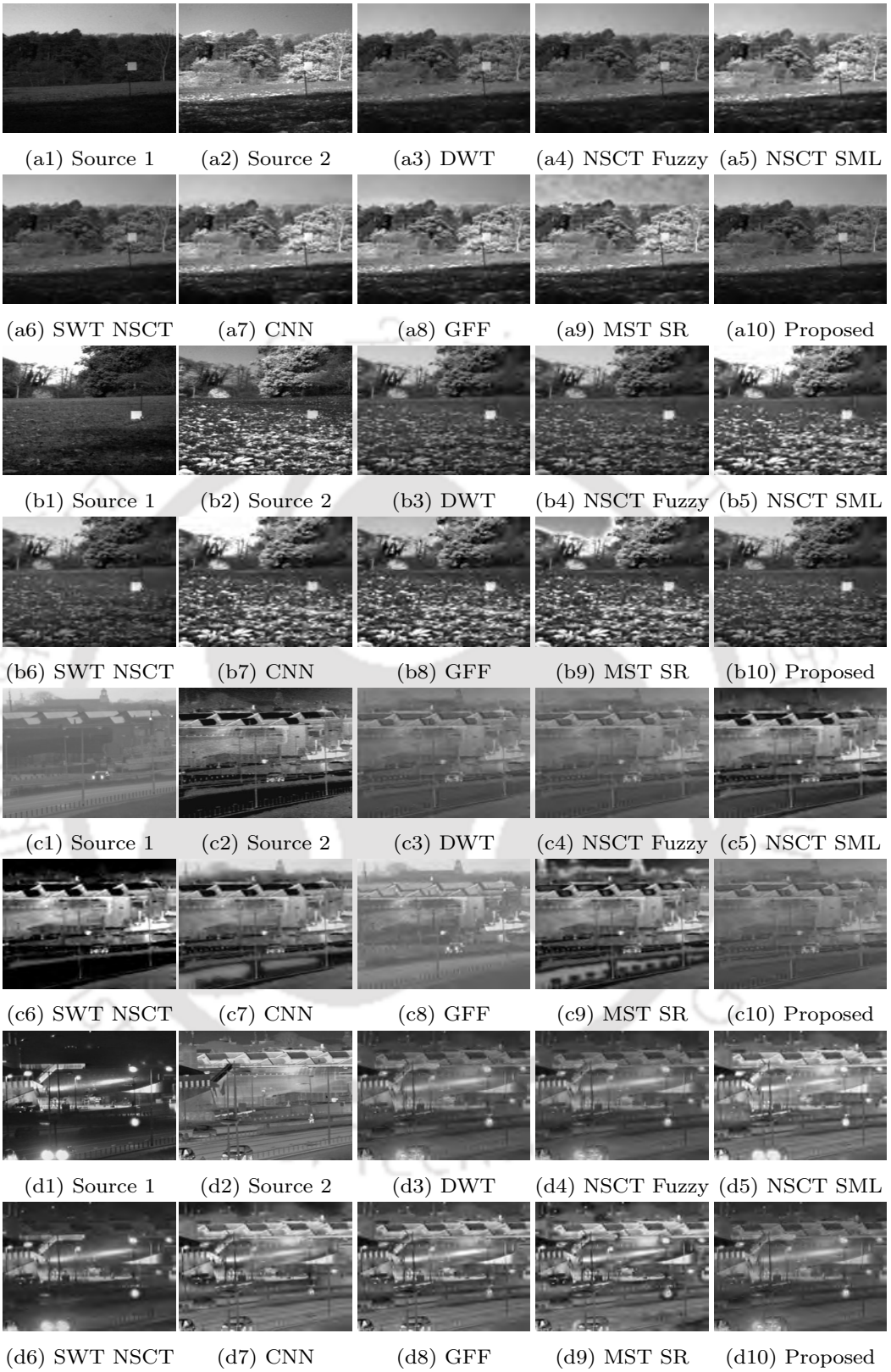


FIGURE 2.5: Source and fused visible-infrared images obtained by using different fusion methods.

TABLE 2.2: Evaluation metrics for fused images of dataset 1.

Methods	Metics/Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{TS}
DWT	Medical	0.6341(0)	0.3634(0)	0.7468(0)	0.1255(0)
	Visible-Infrared	0.4302(0)	0.4509(0)	0.7207(0)	0.1722(0)
NSCT Fuzzy	Medical	0.6539(0)	0.3302(0)	0.7399(0)	0.1690(0)
	Visible-Infrared	0.4519(0)	0.4096(0)	0.7068(0)	0.1055(0)
NSCT SML	Medical	0.6461(0)	0.3790(0)	0.7474(0)	0.2454(0)
	Visible-Infrared	0.4435(0)	0.4786(0)	0.7334(0)	0.1530(0)
SWT NSCT	Medical	0.7054(0)	0.4646(0)	0.8217(3)	0.2244(0)
	Visible-Infrared	0.5562(0)	0.3591(0)	0.8208(1)	0.1111(0)
CNN	Medical	0.9981(2)	0.6087(3)	0.8202(2)	0.5127*(6)
	Visible-Infrared	0.8145(2)	0.6674(2)	0.8215(3)	0.3937(2)
GFF	Medical	0.6887(0)	0.6075(2)	0.8109(0)	0.4048(0)
	Visible-Infrared	0.5780(0)	0.6696*(4)	0.8211(1)	0.3686(0)
MST SR	Medical	0.8229(0)	0.5920(0)	0.8210(0)	0.3581(0)
	Visible-Infrared	0.7137(0)	0.6476(0)	0.7997(0)	0.3505(0)
Proposed	Medical	1.0753*(6)	0.6098*(3)	0.8287*(4)	0.4077(2)
	Visible-Infrared	0.9802*(4)	0.6036(2)	0.8295*(3)	0.4167*(6)

The numbers in the parentheses indicate number of image pairs that this approach outperforms other approaches. The results with * are significantly better than the others.

(CNN) [118].

For comparison, neural network (NN) based feature-level fusion method [142] and SVM based decision-level fusion method [143] are considered. Also, the proposed method is compared with the method which uses NSST [131]. For comparative study, the original parameter settings are maintained for all the previous methods used for comparison. The proposed and the other methods are tuned to optimum parameter settings to produce best possible results.

In our method, 2-level of NSST decomposition is used to save memory, and to reduce computational time and noise. Increase in decomposition level also enhances noise in the fused image. The modified weighted salience method is used to fuse low-frequency bands, and local difference fusion rule is applied to fuse high-frequency bands. Inverse NSST is applied to get the primary fused image. A 3×3 window is employed in Eqs. (2.5) and (2.6). Secondary and final fused image is obtained by using Eqs. (2.10) and (2.11),

TABLE 2.3: Metric values for medical images shown in Fig. 2.4

Methods	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{TS}	Run time (in seconds)
DWT	(a1)-(a2)	0.6664	0.3721	0.7557	0.296	0.65
	(b1)-(b2)	0.7471	0.4534	0.7192	0.2279	0.57
	(c1)-(c2)	0.4671	0.4269	0.71	0.1222	0.64
	(d1)-(d2)	0.6618	0.4	0.8062	0.1807	0.61
NSCT Fuzzy	(a1)-(a2)	0.6643	0.3401	0.7457	0.1753	2.11
	(b1)-(b2)	0.7675	0.4443	0.733	0.1655	2.18
	(c1)-(c2)	0.4843	0.3994	0.6907	0.1135	2.07
	(d1)-(d2)	0.679	0.3726	0.8057	0.1475	2.13
NSCT SML	(a1)-(a2)	0.6694	0.3784	0.7574	0.2204	2.13
	(b1)-(b2)	0.7667	0.4539	0.723	0.1663	2.14
	(c1)-(c2)	0.478	0.4447	0.7254	0.1151	2.14
	(d1)-(d2)	0.6704	0.4109	0.8056	0.1655	2.09
SWT NSCT	(a1)-(a2)	0.7272	0.463	0.9677*	0.1683	1.64
	(b1)-(b2)	0.8522	0.5407	0.937	0.1524	1.61
	(c1)-(c2)	0.5077	0.4542	0.8348	0.113	1.54
	(d1)-(d2)	0.7795	0.589	0.9014	0.1723	1.60
CNN	(a1)-(a2)	1.1372	0.6283	0.9611	0.6085*	37.54
	(b1)-(b2)	1.3426*	0.6567	0.9152	0.6222	36.41
	(c1)-(c2)	0.7703	0.5973*	0.8358	0.3515	36.44
	(d1)-(d2)	0.9239	0.6513	0.8867	0.4701*	37.41
GFF	(a1)-(a2)	0.6938	0.6057	0.8189	0.4188	0.34
	(b1)-(b2)	0.643	0.6942*	0.677	0.4084	0.36
	(c1)-(c2)	0.5068	0.5905	0.8172	0.2471	0.37
	(d1)-(d2)	0.6389	0.5892	0.8076	0.3082	0.33
MST SR	(a1)-(a2)	0.7674	0.5742	0.8176	0.3149	11.8
	(b1)-(b2)	0.8168	0.6414	0.6568	0.2742	12.3
	(c1)-(c2)	0.5448	0.5632	0.7326	0.2369	11.88
	(d1)-(d2)	0.7557	0.5796	0.8693	0.1335	11.24
Proposed	(a1)-(a2)	1.163*	0.6647*	0.8356	0.4547	2.15
	(b1)-(b2)	1.3007	0.6923	0.9583*	0.648*	2.01
	(c1)-(c2)	1.0113*	0.5807	0.8484*	0.3542*	2.07
	(d1)-(d2)	1.1186*	0.6574*	0.9189*	0.2884	2.02

The results with * are significantly better than the others.

TABLE 2.4: Metric values for visible infrared images shown in Fig. 2.5

Methods	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{TS}	Run time (in seconds)
DWT	(a1)-(a2)	0.7486	0.568	0.8432	0.1935	0.67
	(b1)-(b2)	0.4072	0.5467	0.8356	0.2008	0.51
	(c1)-(c2)	0.3786	0.4429	0.7155	0.1352	0.57
	(d1)-(d2)	0.3866	0.4463	0.7156	0.1594	0.68
NSCT Fuzzy	(a1)-(a2)	0.7768	0.5442	0.8252	0.1982	2.03
	(b1)-(b2)	0.4236	0.5182	0.8111	0.1384	2.11
	(c1)-(c2)	0.4072	0.3781	0.6904	0.1625	2.14
	(d1)-(d2)	0.4002	0.3982	0.7005	0.1029	2.08
NSCT SML	(a1)-(a2)	0.7672	0.6004	0.8493	0.1751	2.16
	(b1)-(b2)	0.4189	0.5841	0.845	0.1836	2.18
	(c1)-(c2)	0.3936	0.4637	0.7212	0.1129	2.11
	(d1)-(d2)	0.3946	0.4665	0.7184	0.1405	2.18
SWT NSCT	(a1)-(a2)	0.7793	0.4402	0.7636	0.1897	1.57
	(b1)-(b2)	0.8245	0.5486	0.7672	0.2145	1.68
	(c1)-(c2)	0.41	0.3728	0.7044	0.1291	1.59
	(d1)-(d2)	0.4564	0.3832	0.8516	0.136	1.64
CNN	(a1)-(a2)	1.0144	0.7355	0.9756*	0.4641	37.88
	(b1)-(b2)	0.9562*	0.757	0.946	0.492*	37.21
	(c1)-(c2)	0.8407	0.7245	0.8239	0.4165	38.24
	(d1)-(d2)	0.6467	0.6529	0.8027	0.4024	36.28
GFF	(a1)-(a2)	0.8431	0.749	0.9718	0.4275	0.32
	(b1)-(b2)	0.647	0.7599*	0.9566	0.4658	0.39
	(c1)-(c2)	0.6415	0.705	0.9175	0.4148	0.32
	(d1)-(d2)	0.3807	0.6648	0.8004	0.3663	0.39
MSTSR	(a1)-(a2)	0.9276	0.7272	0.9352	0.1749	12.7
	(b1)-(b2)	0.8201	0.7515	0.932	0.1262	11.2
	(c1)-(c2)	0.7629	0.6752	0.817	0.3788	12.7
	(d1)-(d2)	0.5442	0.6368	0.7148	0.3534	11.36
Proposed	(a1)-(a2)	1.1648*	0.7593*	0.9655	0.5041*	2.07
	(b1)-(b2)	0.9145	0.7269	0.9691*	0.4478	2.08
	(c1)-(c2)	1.0143*	0.7265*	0.9212*	0.4451*	2.01
	(d1)-(d2)	1.0273*	0.6687*	0.8596*	0.4701*	2.07

The results with * are significantly better than the others.

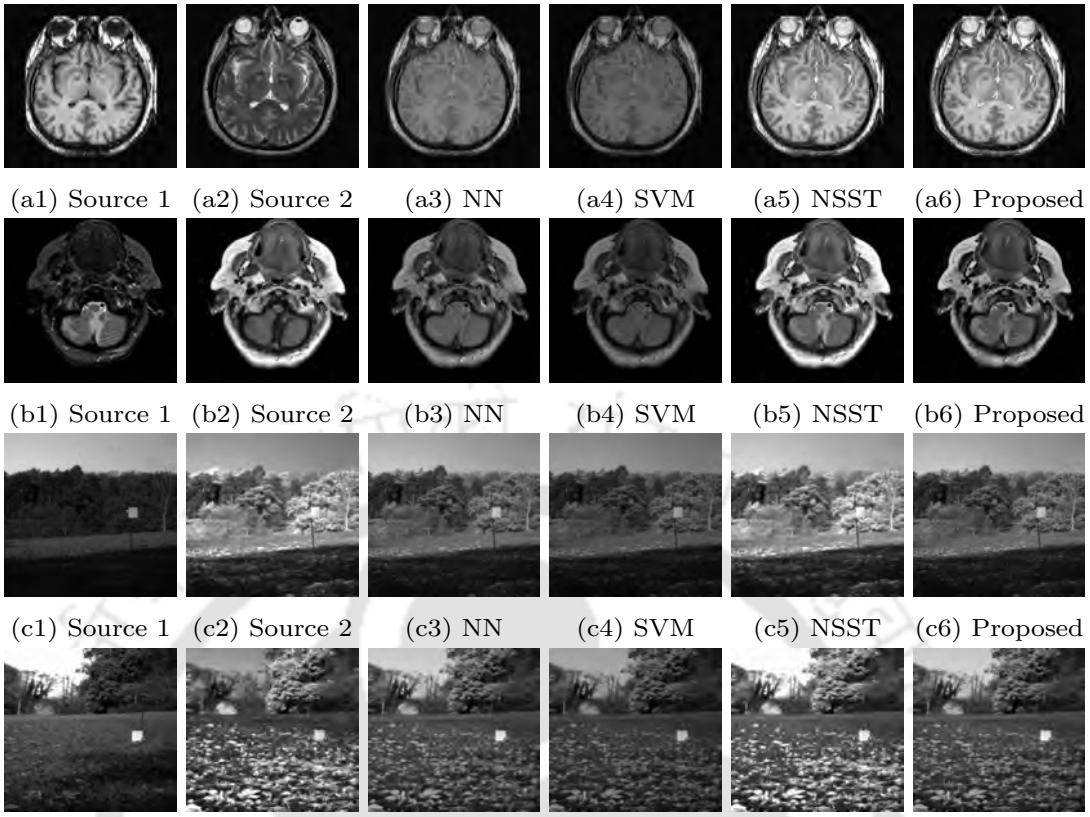


FIGURE 2.6: Source and fused images obtained by using different fusion methods.

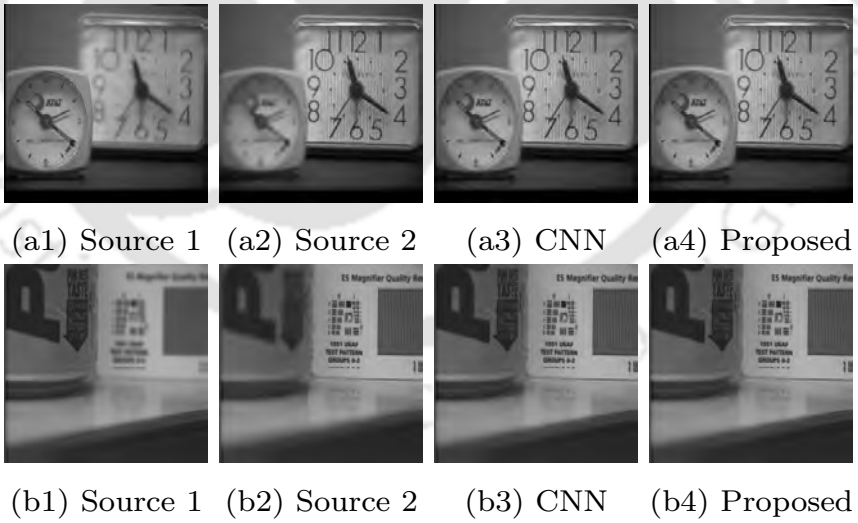


FIGURE 2.7: Subjective comparison of proposed and CNN method for 2 pairs of multi-focus images.

respectively. The secondary fused image is used to get maximum edge information from the source images and it is fused with primary fused image. For proposed metric Q_{TS}, the size of the window is 3×3 for local standard deviation (Eq. (2.12)), local information

TABLE 2.5: Performance metrics for fused images of dataset 1.

Methods	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{TS}	Run time (in seconds)
NN	(a1-a2)	0.7096	0.6547	0.5840	0.1654	7.82
	(b1-b2)	0.8425	0.6011	0.5840	0.2654	7.95
	(c1-c2)	0.6947	0.5944	0.8572	0.1953	7.55
	(d1-d2)	0.6227	0.5905	0.7628	0.1295	7.45
SVM	(a1-a2)	0.6941	0.5821	0.5273	0.1055	4.35
	(b1-b2)	0.7468	0.6748	0.5642	0.1672	4.12
	(c1-c2)	0.6266	0.5735	0.6426	0.1052	4.25
	(d1-d2)	0.5823	0.5524	0.6358	0.1685	4.85
NSST	(a1-a2)	0.8601	0.6062	0.6761	0.3835	2.06
	(b1-b2)	0.8975	0.6778*	0.6843	0.4258*	2.11
	(c1-c2)	0.8317	0.5702	0.8119	0.4011	2.05
	(d1-d2)	0.8875	0.5527	0.8236	0.4578*	2.09
Proposed	(a1-a2)	1.1631*	0.6647*	0.8356*	0.4547*	2.15
	(b1-b2)	1.1186*	0.6574	0.9189*	0.2884	2.02
	(c1-c2)	1.1648*	0.7593*	0.9655*	0.5041*	2.07
	(d1-d2)	0.9145*	0.7269*	0.9691*	0.4478	2.08

The results with * are significantly better than the others.

entropy (Eq. (2.13)) and local difference (Eq. (2.4)).

Our proposed method is able to retain most prominent information of both the source images as seen in Figs. 2.4(a10)-(d10) and 2.5(a10)-(d10). This claim is further justified with the help of objective comparison as shown in Tables 2.2, 2.3 and 2.4. The results obtained using the proposed and CNN [118] method are very competitive, but still our method can give slightly better results.

In Fig. 2.4, (a1), (a2), (b2), (c2), (d1) and (d2) are MR source images; (b1) is a SPECT image; and (c1) is a CT image. In Fig. 2.5, source images (a1), (b1), (c1) and (d1) are visible images; and (a2), (b2), (c2) and (d2) are infrared images. In Figs. 2.4 and 2.5, it is seen that DWT, NSCT Fuzzy and NSCT methods can not preserve brightness information due to inefficient fusion rules. Also, DWT based method produces ringing artifacts. The SWT NSCT and MST SR based fused images are shown in Figs. 2.4 and 2.5, and they contain noisy pixels, and fine details are suppressed. The fused images obtained using CNN, GFF, and MST SR methods produce bright fused images,

FIGURE 2.8: Comparison of different colour spaces by proposed image fusion technique in dataset-2.

and hence prominent information of one of the source images having low brightness are marginalized. The undesirable blurring is observed for GFF based fused images in Figs. 2.4 and 2.5. The CNN and MST SR methods produce spatial artifacts. We also observed that the fused images produced by our method preserve more prominent information as compared to other state-of-the-art methods.

In Tables 2.3 and 2.4, evaluation metric values for medical and visible-infrared images are shown. The performance of proposed method is quantitatively better than the exiting methods. Additionally, run times for different methods are shown in Tables 2.3 and 2.4. The proposed metric Q_{TS} gives very reliable results as observed from the results shown in Figs. 2.4 and 2.5, and in Tables 2.3 and 2.4. In Figs. 2.4(d9), 2.5(a9) and 2.5(b9), the fused images obtained by using MST SR method suffers from spatial (texture and structural) distortions. Hence, the lowest Q_{TS} metric values are obtained for these fused images.

In Fig. 2.6 and Table 2.5, we showed comparisons of the proposed method with the methods proposed in [131, 142, 143]. These experiments were performed on the dataset-1 of Fig. 2.2, and the results for four pairs of source images are shown in Fig. 2.6 and

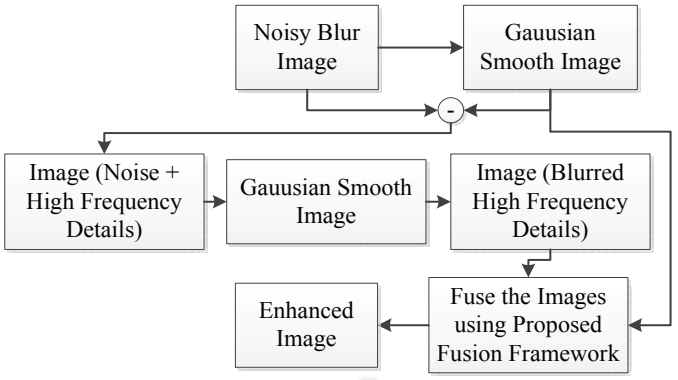


FIGURE 2.9: Image enhancement by proposed fusion method.

FIGURE 2.10: Qualitative comparison of the proposed local difference and other high-frequency coefficients selection rules, where experiments were performed in dataset-3.

Table 2.5. In Fig. 2.6, the fused images obtained by using NN and SVM methods are blurred, and some significant information are missing. NSST based fused images have high contrast and the prominent information are also missing. However, fused images obtained by using the proposed method preserves most of the prominent information of the source images. The quantitative comparisons as shown in Table 2.5, indicates that the proposed method also outperforms NN [142], SVM [143] and NSST [131] based methods. In Table 2.5, except for the metric $Q^{AB/F}$, the proposed method produces

TABLE 2.6: Objective comparison of proposed and CNN method is presented for Fig. 2.7.

Method	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{TS}	Run time (in seconds)
CNN	(a1-a2)	1.2434*	0.7509*	0.8982*	0.7003*	33.51
	(b1-b2)	1.3016*	0.7596*	0.8983	0.7342*	33.03
Proposed	(a1-a2)	1.2058	0.7464	0.8975	0.4879	2.15
	(b1-b2)	1.2811	0.7591	0.9019*	0.5493	2.16

The results with * are significantly better than the others.

high metric values for medical and visible-infrared images for most of the cases.

Subjective and objective comparison of the proposed and the CNN method for 2 pairs of multi-focus images are shown in Fig. 2.7 and Table 2.6, respectively. In subjective analysis, it is difficult to judge the performance of both the methods. Quantitatively, CNN gives better results than the proposed method for the multi-focus images. However, proposed method gives better results for medical and visible-infrared images.

DWT [24] fusion method is less computationally complex as compared to other techniques, and this methods are highly efficient to detect point discontinuities present in the images. The main drawback is the presence of ringing artifacts in the fused images due to shift variant property of the transformation. If the source images are not perfectly registered, then shift variant transformation alters transformed coefficients, which results distorted fused image. Moreover, DWT method is inefficient to approximate edge information. NSCT fuzzy [4] approach suffers from low contrast. Moreover, this method is complex due to NSCT and complex fusion rules.

NSCT SML [2] method is unable to capture prominent information of the source images, such as texture details. In addition, NSCT and phase congruency increase computational complexity. SWT NSCT [3] method suffers from low contrast as it uses principal component analysis for fusion. However, SWT is able to represent texture information more efficiently than NSCT. In CNN- based method [118], it is difficult to find optimal parameter settings and appropriate network architecture for fusion of medical and visible-infrared images.

GFF [114] is a regression based weighted fusion method, but crucial information is lost due to multiple iterations and inefficient weight generation process. MST SR [89] method

TABLE 2.7: Quantitative evaluation of different colour spaces of dataset-2.

Metrics	HSV	NTSC (YIQ)	YCbCr	Lab
Q_{MI}	0.9914(1)	0.9958(0)	1.0103(0)	1.0413*(3)
$Q^{AB/F}$	0.7106*(2)	0.7032(0)	0.70697(2)	0.6973(0)
Q_S	0.8430(0)	0.8518(0)	0.8567(1)	0.8706*(3)
Q_{TS}	0.4927(1)	0.4693(0)	0.5009*(3)	0.3847(0)

The numbers in the parentheses indicate number of image pairs that this approach outperforms other approaches. The results with * are significantly better than the others.

fails to capture fine information present in the source images due to inefficient dictionary. The fused image obtained using this method contains noisy pixels, as absolute value maximum selection fusion rule is employed without using any noise reduction techniques.

Our proposed method selects high- and low-frequency components from the source images more efficiently via local difference and modified weighted salience fusion rules, respectively. Additionally, edge information is boosted using Canny edge detection with scale multiplication technique.

2.4.4 Comparative study of different colour spaces for proposed fusion method

For gray-scale image fusion, only the luminance component of the colour is considered. For colour image fusion, colour components of any one of the source images is used, and the fused image is obtained by performing inverse colour space model transformation. However, problem arises when we arbitrarily choose a colour space model, and as seen in Fig. 2.8, the results are not plausible. Fused images obtained using Lab and NTSC (YIQ) colour model contain false colours. Experimental results given in Fig. 2.8 and Table 2.7 show that YCbCr and HSV can give superior image fusion performance. Dataset-2 (shown in Fig. 2.8) is used to perform this experiment.

Any colour model can be transformed into a different absolute colour model, and vice-versa. Some colour models have gamut constraints, which should be analyzed before colour space transformation. Moreover, rounding error can also cause inaccurate results.

TABLE 2.8: Comparison of image enhancement performance for dataset-3.

Metrics	Gradient	SML [2]	SF [134]	Proposed
MAE	11.77615(0)	11.77525(1)	11.774775(0)	11.7732*(3)
PSNR	-25.2787(0)	-25.2775(1)	-25.2771(1)	-25.2756*(2)
SSIM	0.61656(0)	0.61673(1)	0.6165(0)	0.61693*(3)

The numbers in the parentheses indicate number of image pairs that this approach outperforms other approaches. The results with * are significantly better than the others.

TABLE 2.9: Average metrics values for different non-subsampled Laplacian pyramid filters for dataset-1.

Index	*9-7*	*maxflat*	*pyr*	*pyrexc*
Q_{MI}	1.091325	1.092225	1.094025*	1.094025*
$Q^{AB/F}$	0.60725	0.607475*	0.6074	0.6074
Q_S	0.897975	0.898025*	0.897975	0.897975
Q_{TS}	0.390375	0.389925	0.392075*	0.39205

The results with * are significantly better than the others

2.4.5 Image enhancement using proposed fusion framework

The proposed fusion method can also be employed to enhance visual quality of an image as shown in Fig. 2.9. First, Gaussian smoothing is performed on the noisy blurred source image, and the resultant image is subtracted from the noisy blurred source image. This subtraction produces an image having noises and high-frequency information. This image is again fed into Gaussian smoothing operation (relatively small size kernel) which removes remaining noisy pixels with slightly more blurring. Subsequently, the two images obtained just after Gaussian smoothing operation are fused using the proposed fusion framework. The resultant fused image is relatively noise free and less blurred.

This framework is implemented by our proposed fusion method by using gradient, SML, SF, and proposed local difference (LD) rules to select high-frequency coefficients. Dataset-3 (ground truth images of Fig. 2.10) is used to perform this comparative study. From visual comparison of the images in Fig. 2.10, it is difficult to find the best image enhancement method. However, the results of Table 2.8, which are presented in terms of mean absolute error (MAE), peak signal to noise ratio (PSNR), and structure similarity

TABLE 2.10: Average run time of state-of-the-art techniques on dataset-1

Method	Time in seconds
DWT [24]	0.6
NSCT Fuzzy [4]	2.12
NSCT SML [2]	2.11
SWT NSCT [3]	1.61
CNN [118]	37.75
GFF [114]	0.34
MST SR [89]	11.2
NN [142]	7.57
SVM [143]	4.15
NSST [131]	2.05
Proposed	2.04

measure (SSIM) clearly indicate that the proposed local difference rule captures high-frequency coefficients more efficiently as compared to the gradient, SML and SF based fusion rules. This comparative study shows that the proposed fusion method can be effectively employed for image enhancement. Also, our proposed local difference fusion rule can be employed for the applications where high-frequency coefficients need to be selected. The large local difference values of high-frequency coefficients indicate sharper brightness and contrast in an image, which corresponds to edges, points, border regions etc.

The negative values of PSNR (in dB) indicates that the enhanced images are not close to the ground truth images, which makes the ratio of square of peak image pixel value to mean square error to become low. In this case, mean square error of ground truth and enhanced images is greater than square of peak image pixel value of ground truth images. However, our proposed local difference technique produces higher values of PSNR compared to other techniques.

2.4.6 Comparison of different non-subsampled Laplacian pyramid filters

Comparative study of different non-subsampled Laplacian pyramid filters [144] are carried out using proposed fusion framework. The *9-7* filters are obtained from *9-7* 1-D

prototypes filters. The *maxflat* filters are derived from 1-D filters using maximally flat mapping function with 4 vanishing moments. The *pyr* filters are derived from 1-D filters using maximally flat mapping function with 2 vanishing moments. The *pyrexc* filters are same as *pyr*, but two high-pass filters are exchanged for *pyrexc*. In Table 2.9, average values of all the quantitative metrics for all the source images of dataset-1 (Fig. 2.2) are presented. From this comparison, it is found that *pyr* filter is the best choice for multi-sensor fusion. Moreover, in general, the response of *pyr* filter depends on the spectrum of the multi-sensor images.

2.4.7 Computational time efficiency analysis

The comparison of computational time efficiencies of the different state-of-the-art fusion methods is given in Table 2.10 for dataset-1 (Fig. 2.2). Our proposed method is implemented in MATLAB R2017a platform on a computer with a processor speed of 3.20 GHz, 64-bit operating system and 8 GB RAM. As shown in Table 2.10, CNN and MST SR methods are very time-consuming as they need optimal parameter setting and sparse coding. The NN and SVM also have high time complexity. The time complexity of NSST, NSCT Fuzzy, NSCT SML and the proposed method is almost same. NSCT, Fuzzy rules and phase congruency techniques are computationally complex. Our method is more complex than GFF and DWT-based methods as these methods uses relatively less complex fusion rules and framework.

Order of computational complexity of curvelet transform is $O(N^2 \log N)$ [145], and $O(N^2)$ and $O(2^{\log_2(1/2(L/2-1))}LN)$ for contourlet transform and shearlet transform, respectively [28, 146], where $N \times N$ is size of the image and L is the total number of directions at the finest scale. The computational time complexity of SWT NSCT is lower than NSCT Fuzzy, NSCT SML and the proposed method as SWT NSCT method employed comparatively simple fusion rules, such as absolute value maximum selection rule. Also, fusion frameworks of NSCT Fuzzy and NSCT SML methods are less complex as compared to the proposed method. The proposed method employs relatively complex fusion rules, anisotropic diffusion filtering, scale multiplication-based Canny edge detection and the concept of generation of primary and secondary fused images. The

complexity issue of the proposed method is sacrificed to some extent in view of improving the quality of the fused images.

2.5 Conclusion

Image fusion is a very challenging research area in the sense that the prominent details of the source images need to be preserved in the fused image. In this chapter, we proposed an NSST based fusion method for multi-sensor images. For this, novel fusion rules are proposed to select low-frequency and high-frequency coefficients for fusion. Deployment of canny edge detector with scale multiplication in the fusion framework enhances the edge information in the fused image. Also, a metric Q_{TS} is proposed to measure texture and structural information in the fused images. Experimental results show that the fusion performance evaluation implemented by the proposed and others objective measures strongly agree with the subjective analyses. Experimental results explicitly show that our metric Q_{TS} is perceptually significant and essential to judge the performance of fusion algorithms.

The proposed method is tested on different colour spaces, and it is found that YCbCr is an efficient model for colour image fusion. Moreover, the proposed fusion method can be employed for image enhancement. The proposed method can eliminate some of the noises present in the source images as our method uses window-based fusion rules. We believe that our technique can be efficiently implemented for real-time applications using a more efficient programming approach.

The proposed method can deal with noisy source images due to the deployment of anisotropic diffusion filtering. However, the proposed method is unable to produce state-of-the-art results in case of multi-focus images due to non-optimal frequency response of NSST filters and lack of adaptivity in fusion rules. Moreover, there is still a scope of formulating a new generalize metric to measure other attributes of fused images.

3

An Optimized Non-subsampled Shearlet Transform-based Image Fusion using Hessian Features and Unsharp Masking

Existing image fusion approaches are not so efficient to seize significant edges, texture and fine features of the source images due to ineffective and non-adaptive fusion structure. Also for objective evaluation of fusion algorithms, there is a need of a metric to measure source image features which are preserved in the fused image. To address these issues, an optimized non-subsampled shearlet transform (NSST) is developed, which is applied to decompose the source images into low- and high-frequency bands. The low-frequency bands are fused using proposed descriptor obtained from superposition of scale multiplied Canny edge detector features and Hessian features. The high-frequency bands are fused using unsharp masking based fusion rule. Moreover, a metric Q_E is formulated on the basis of Karhunen-Loeve transform (KLT). The information of image pixel variance for both source and fused images can be measured by using the proposed metric Q_E, and it gives an indication of the amount of variance information transferred from the source images to the fused image. Both subjective and objective analysis show the efficacy of the proposed fusion structure and the metric Q_E.

3.1 Introduction

The implementation complexity of curvelet and NSCT is higher as compared to NSST. Due to that, NSST is employed in our proposed method. More details on computational complexities of these transforms are given in Section 2.4.7.

The proposed method utilizes optimized NSST with an objective to get a high-quality fused image. Optimization of NSST is implemented by designing 2-D filters of non-subsampled Laplacian pyramid using optimized prototype low-pass filter and quadrature mirror filter (QMF) [147, 148]. In this method, a new fusion rule is implemented to fuse high-frequency shearlet coefficients, which is based on unsharp masking. Unsharp masking based rules are quite effective in capturing high-frequency information. In this rule, high-frequency coefficients are enhanced using an unsharp masking method, then based on these enhanced coefficients, fusion is performed on high-frequency coefficients. Additionally, for fusion of low-frequency coefficients, a feature descriptor is formulated by superposition of scale multiplied Canny edge detector features [135] and Hessian features [149]. This descriptor is used to boost important attributes of the source images, such as point, lines, edges etc in the fused low-frequency band.

To judge the quality of the fused image, there is a requirement of an effective metric to measure overall feature information, such as texture, points, lines, edges, corners, smoothness etc of the fused image, which are transferred from source images to the fused image. Therefore, a novel metric Q_E is proposed to measure overall feature information of the source images which is preserved in the fused image. The proposed metric Q_E is based on KLT [150].

The proposed metric Q_E measures the variance of data of the fused image with respect to the variances of data of the source images. Furthermore, Q_E can measure the quality of fused images obtained from different image fusion algorithms. Experimental results validate the effectiveness of our proposed image fusion framework and the proposed metric Q_E. Some very popular image fusion algorithms and their applications are available in [151–156].

The rest of this chapter is organized as follows: The fundamental concepts of NSST, scale multiplied Canny edge detector features, Hessian features, and unsharp masking are described in Section 3.2. In Section 3.3, our proposed optimize NSST, proposed feature descriptor for low-frequency band fusion and unsharp masking for high-frequency band fusion are elaborately described, and the proposed fusion structure is illustrated. The design of the proposed variance preservation metric (Q_E) is explained in Section 3.4. Also, all the experimental results are presented in this section. At last, concluding remarks are given in Section 3.5.

3.2 Preliminaries

3.2.1 Non-subsampled shearlet transform (NSST)

Shearlet transform (ST) produces a set of well-localized waveforms at various scales, locations, and orientations [31]. In NSST, the source image is decomposed into low-pass and high-pass sub-images using non-subsampled Laplacian pyramid. High-pass sub-images are fed to shift-invariant shear filters and the low-pass sub-image is further decomposed into low-pass and high-pass sub-images. The inverse NSST transform is applied by taking the sum of all shift invariant shearing filter responses at the respective levels of decomposition, and inverse non-subsampled Laplacian pyramid is finally applied to get the source image. The detailed explanation of NSST is given in Section 2.2.1.

3.2.2 Optimization of protoype filters

The motivation behind optimized NSST is the concept of optimized filter bank and transmultiplexer theory [147, 148]. A filter bank divides the spectrum of a input signal into some smaller non-overlapping frequency bands. One simple way to design the filter bank system is to design a low-pass prototype filter $H_0(z)$, then design QMF as:

$$H_1(z) = H_0(-z), F_0(z) = H_0(z), F_1(z) = -H_1(z). \tag{3.1}$$

where, analysis low- and high-pass filters are $H_0(z)$ and $H_1(z)$, respectively in z-domain. $F_0(z)$ and $F_1(z)$ are synthesis low- and high-pass filters in frequency domain, respectively. The 2-D extension of these filters can be implemented using the method proposed by Tay and Kingsbury [157].

Lin and Vaidyanathan proposed an approach to optimize prototype filter by Kaiser window, and the objective function is constructed based on the attributes of Nyquist filter [147]. The filter bank has perfect reconstruction property if the polyphase components of the prototype satisfy pairwise power complementary conditions [148]. Suppose, the prototype filter $P(e^{j\omega})$ has a linear phase charateristics. The criteria for approximate reconstruction is given in terms of $P(e^{j\omega})$, which is given by [147]:

$$P(e^{j\omega}) \approx 0, \quad \text{for} \quad |\omega| > \pi/M. \tag{3.2}$$

$$T(e^{j\omega}) \approx 1, \quad \text{where} \quad T(e^{j\omega}) = \sum_{k=0}^{2M-1} |P(e^{j(\omega - k\pi/M)}|^2. \tag{3.3}$$

The filter $G(e^{j\omega})$ is given by $G(e^{j\omega}) = |P(e^{j\omega})|^2$. The condition in Eq. (3.3) means that $G(e^{j\omega})$ is a approximately Nyquist $(2M)$ filter. Hence, the objective function [147] is given as:

$$\phi_{new} = \max_{n, n \neq 0} |g(2Mn)|. \tag{3.4}$$

In this optimization technique, we have direct control over stopband attenuation of the prototype filter, which ensures control over stopband attenuation of all the analysis and synthesis filters of non-subsampled Laplacian pyramid. In our proposed method, we only use the class of filters which are designed using Kaiser window.

3.2.3 Scale multiplied Canny edge detector

Canny edge detector is very efficient in detecting edges of an image. More detailed on this is presented in Section 2.2.2.

3.2.4 Hessian features

Hessian feature detector is an affine invariant detector [149]. Hessian features show local curvature of a function of some variables. These features are obtained from the Hessian matrix. The Hessian feature descriptor of any image highlights corners, small ridges, and fine details present on the images. The Hessian feature descriptor is obtained by applying normalizing scaling factor to 1st and 2nd derivatives. Then, taking element wise operation, descriptor of original image size is obtained. In proposed method, we apply Hessian feature detector on an image $\mathbf{I}(i,j)$, and the output descriptor image of highlighted features is represented as $\mathbf{HEF}_I(i,j)$, and obtained as:

$$\mathbf{HEF}_I(i,j) = \nabla^2\mathbf{I}_{xx}(i,j) \circ \nabla^2\mathbf{I}_{yy}(i,j) - \nabla^2\mathbf{I}_{xy}(i,j)^2 \tag{3.5}$$

where, $\nabla^2\mathbf{I}_{xx}(i,j)$, $\nabla^2\mathbf{I}_{yy}(i,j)$ and $\nabla^2\mathbf{I}_{xy}(i,j)$ are second order derivative in x, y and xy directions, respectively. Moreover, the operation "$\circ$" show the element wise multiplication, and (i,j) are the spatial coordinates.

3.2.5 Unsharp masking

The unsharp masking [158] is the technique to sharpen the image by adding high-frequency information to that image. Unsharp masking operation mainly highlights edge, corners, and fine details of image. These features are highly useful to produce the high-quality fused image. In our method, unsharp masking is employed to fuse high-frequency coefficients. The unsharp masking for an image $\mathbf{I}(i,j)$ is carried out as follows:

$$\mathbf{UM}_I(i,j) = \mathbf{I}(i,j) + \gamma(\mathbf{I}(i,j) - \overline{\mathbf{I}(i,j)}) \tag{3.6}$$

where, $\mathbf{UM}_I(i,j)$ is unsharp masking descriptor map, which is used to select high-frequency coefficients, γ is gain, and $\overline{\mathbf{I}(i,j)}$ is Gaussian blurred version of the image $\mathbf{I}(i,j)$.

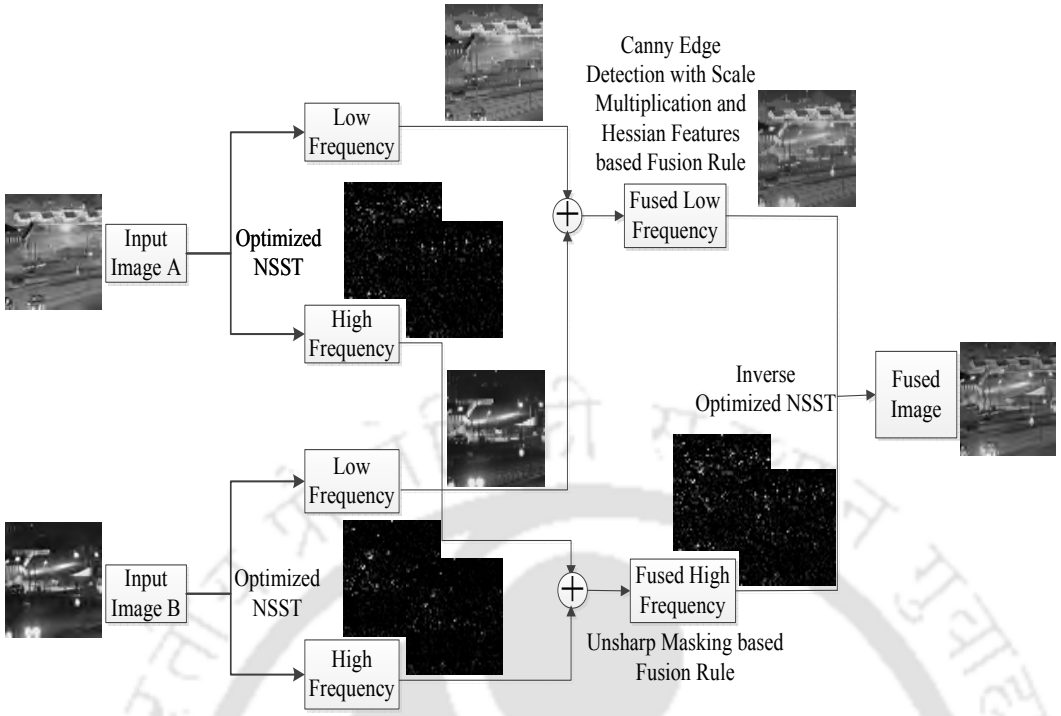


FIGURE 3.1: Schematic diagram of the proposed image fusion framework.

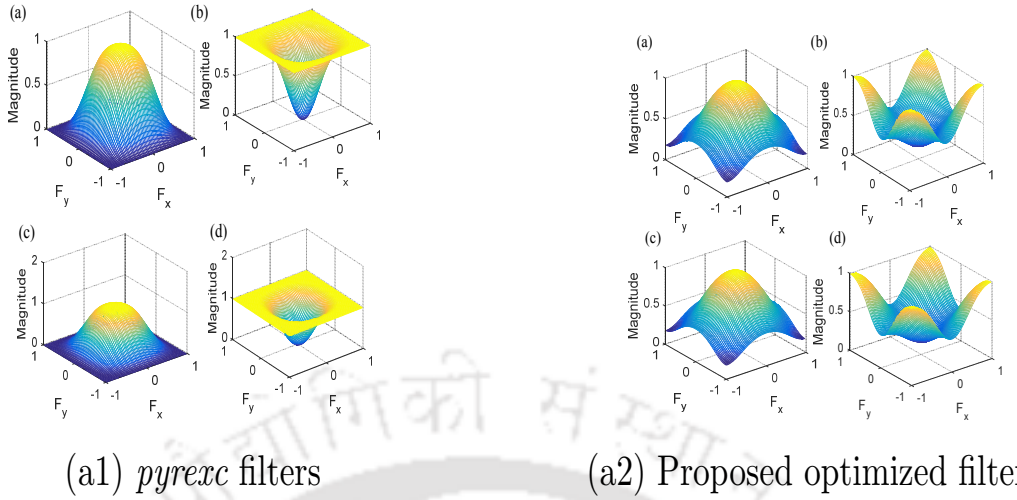
3.3 Proposed Method

The schematic diagram of the proposed method is shown in Fig. 3.1. In our method, optimized NSST is applied to decompose the source images **A** and **B** into low-frequency bands $\mathbf{L}_A(i,j)$ and $\mathbf{L}_B(i,j)$, and high-frequency bands $\mathbf{H}_A^{k,l}(i,j)$ and $\mathbf{H}_B^{k,l}(i,j)$. Here, k and l indicate the decomposed high-frequency band in k^{th} direction at the l^{th} decomposition level. After applying the proposed fusion rules, the final fused image **F** is obtained[3].

3.3.1 Proposed optimized non-subsampled shearlet transform

In our proposed technique, an optimized low-pass prototype filter is designed using Kaiser window with adjustable parameter β. Subsequently, QMF having transfer function in the form of a delay are designed. Hence, these QMF is alias-free filter bank, and they are also known as perfect reconstructed filter bank. Then using the method given in [157], 2-D filters of non-subsampled Laplacian pyramid are designed. The reason behind the

[3]This work has been published in Journal of Visual Communication and Image Representation, Elsevier 2018 (Refer item 4 for details).



(a1) *pyrexc* filters (a2) Proposed optimized filters

FIGURE 3.2: Frequency responses of (a) analysis low-pass filter (b) analysis high-pass filter (c) synthesis low-pass filter (d) synthesis high-pass filter for both *pyrexc* filters and the proposed filters.

selection of this method is that the conventional non-separable filters are designed with less implementation complexity. Hence, proposed filters perform better as compared to separable filters [157].

The proposed method optimizes non-subsampled Laplacian pyramid, and so this optimizes the NSST. The parameter β of the Kaiser window can be adjusted to obtain best fusion results. Parameter β can be varied to control the trade-off between the stopband attenuation and the mainlobe width of the filters. A high value of β results in higher stopband attenuation and wider mainlobe width, and vice versa. Fig. 3.2(a1) shows the frequency response of *pyrexc* filters for non-subsampled Laplacian pyramid, and Fig. 3.2(a2) shows the frequency response of the proposed filters for non-subsampled Laplacian pyramid. Our proposed filters consider more low-frequency information, as low-frequency information has more visual impacts than the high-frequency information.

3.3.2 Proposed feature descriptor

The proposed feature descriptor is formulated on the basis of superposition of scale multiplied Canny edge detector and Hessian features. The proposed descriptor contains the highlighted features related to the edges provided by scale multiplied Canny edge detector and corners with fine information provided by Hessian. The superimposed form

FIGURE 3.3: Response of scale multiplied canny edge detector (SMC) and Hessian (HEF).

of the proposed descriptor $\mathbf{D}$ of $\mathbf{L}_A$ is obtained as follows:

$$\mathbf{D}_{L_A}(i,j) = (1 + \exp(-\mathbf{SMC}_{L_A}(i,j) - \mathbf{HEF}_{L_A}(i,j)))^{-\alpha} \tag{3.7}$$

where, $\mathbf{L}_A$ is decomposed low-frequency band of source image $\mathbf{A}$, α is a constant whose higher values average the source pixels, and (i, j) are spatial coordinates of the image. Fig. 3.3 shows the intuition behind the use of the proposed descriptor, where SMC features highlight the edge information of low-frequency images. On the other hand, HEF features highlights corner, points, contour information of low-frequency images. So, the features captured using SMC and HEF are used to generate high-quality fused images. Hence, the proposed descriptor gives equal importance to both SMC and HEF features. The sigmoid function is used as a mapping function to fix the range of the superimposed features between [0, 1]. The pixels which are not highlighted by SMC and HEF are fused by third averaging condition as given in Eq. (3.8).

3.3.3 Proposed fusion framework

The proposed fusion rule uses optimized NSST for image decomposition. The source images $\mathbf{A}$ and $\mathbf{B}$ are first decomposed into low- and high-frequency bands[4].

Low-frequency band fusion rule: In the proposed method, scale multiplied Canny edge detector and Hessian features are used to form a descriptor, which is used to fuse low-frequency sub-images. The low-pass sub-images convey crucial information in the region where texture, edge, corners and variational information are dominant. Using this

[4]The related work has been published in Proceedings of 2nd International Conference on Computer Vision & Image Processing, Singapore, LNCS, Springer 2018 (Refer item 4 for details).

descriptor given in Eq. (3.7), low-frequency fused coefficients are obtained as follows:

$$\mathbf{L}_F(i,j) = \begin{cases} \mathbf{L}_A(i,j), & \text{if} \quad \mathbf{D}_{L_A}(i,j) > \mathbf{D}_{L_B}(i,j) \\ \mathbf{L}_B(i,j), & \text{if} \quad \mathbf{D}_{L_A}(i,j) < \mathbf{D}_{L_B}(i,j) \\ 0.5(\mathbf{L}_A(i,j) + \mathbf{L}_B(i,j)), & \text{Otherwise.} \end{cases} \tag{3.8}$$

where, $\mathbf{L}_F(m,n)$ is the fused low-frequency sub-image.

High-frequency band fusion rule: For fusing high-frequency bands, unsharp masking-based fusion rule is proposed. Unsharp masking-based fusion rules enhance edge, texture and fine features by enhancing high-frequency information of images. It is simple and less complex compared to gradient, sum-of-modified-Laplacian [159] and spatial frequency [134] based fusion rules. In each high-frequency bands $\mathbf{H}_A^{k,l}(i,j)$ and $\mathbf{H}_B^{k,l}(i,j)$, Eq. (3.6) is applied, and unsharpened mask feature images $\mathbf{UM}_A^{k,l}(i,j)$ and $\mathbf{UM}_B^{k,l}(i,j)$ are obtained. Then by using Eq. (3.9), selection of fused coefficients $\mathbf{H}_F^{k,l}(i,j)$ are carried out as:

$$\mathbf{H}_F^{k,l}(i,j) = \begin{cases} \mathbf{H}_A^{k,l}(i,j), & \text{if} \quad \prod_{k,l} \mathbf{UM}_A^{k,l}(i,j) \geq \prod_{k,l} \mathbf{UM}_B^{k,l}(i,j) \\ \mathbf{H}_B^{k,l}(i,j), & \text{Otherwise.} \end{cases} \tag{3.9}$$

where, the product symbol is the Hadamard product [138] of all unsharp masked feature images of decomposed high-frequency sub-images. The Hadamard product of all unsharp masked images are considered as multiplication of different sub-images at different scales and orientation. The high-frequency bands of the source images obtained by using NSST decomposition consist of edge, contour, corner and fine details. The multi-scale representation can efficiently extract edges, contours, and corner information of an image [22, 100, 135]. Due to the non-subsampled nature of the proposed optimized NSST, the edges or corners in the decomposed high-frequency bands occupy the same spatial positions as that of the original image. So, multiplication of all unsharp masking features of high-frequency decomposed bands amplifies prominent high-frequency information [100], and our proposed multiplication procedure is implemented as given in Eq. (3.9).

The main steps of our proposed algorithm are given below. First, the source images which need to be fused are perfectly registered, and subsequently the steps given in the

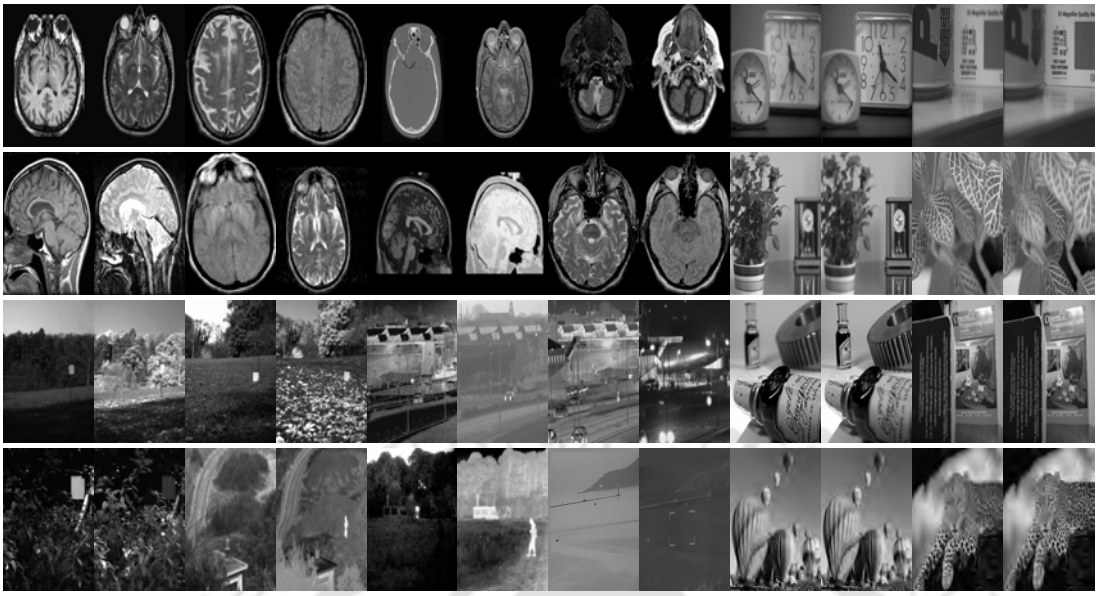


FIGURE 3.4: Image dataset-1 of eight pairs of each medical, visible-infrared, and multi-focus images.

following algorithm are applied.

Proposed algorithm

STEP 1: Design low-pass prototype filter using Kaiser window with filter length N, adjustable parameter β and passband edge frequency $\omega_p = 0.5/\pi$ radian/sample.

STEP 2: Iterate for ω_p to find the local minima, which gives minimum value of defined objective function as given in Eq. (3.4). Constrained optimization is achieved by putting constrained on $\omega_p \in [0.3/\pi, 0.5/\pi]$.

STEP 3: Design analysis and synthesis filters of QMF using Eq. (3.1), and subsequently 2-D filters are designed to construct non-subsampled Laplacian pyramid.

STEP 4: The pre-registered source images are decomposed using adjustable NSST.

STEP 5: The low-frequency band is fused using proposed superposition form of scale multiplied Canny edge detector features and Hessian features as given in Eq. (3.8), and the high-frequency bands are fused using proposed unsharp masking features as given in Eq. (3.9).

STEP 6: The fused image $\mathbf{F}_I(i, j)$ is obtained by taking inverse optimize NSST.

FIGURE 3.5: Performance of proposed method with the variations of free parameters. Vertical axis shows the metric values.

3.4 Experimental Results

To carry out experiments, the standard images are downloaded from the link provided in [89, 140]. The proposed method is compared with the following state-of-the-art methods [1, 24, 33, 88, 89, 114, 118, 132, 160]. Analysis of image fusion techniques is carried out both subjectively and objectively. Four metrics are used to compare the performance of the proposed algorithm with the state-of-the-art methods.

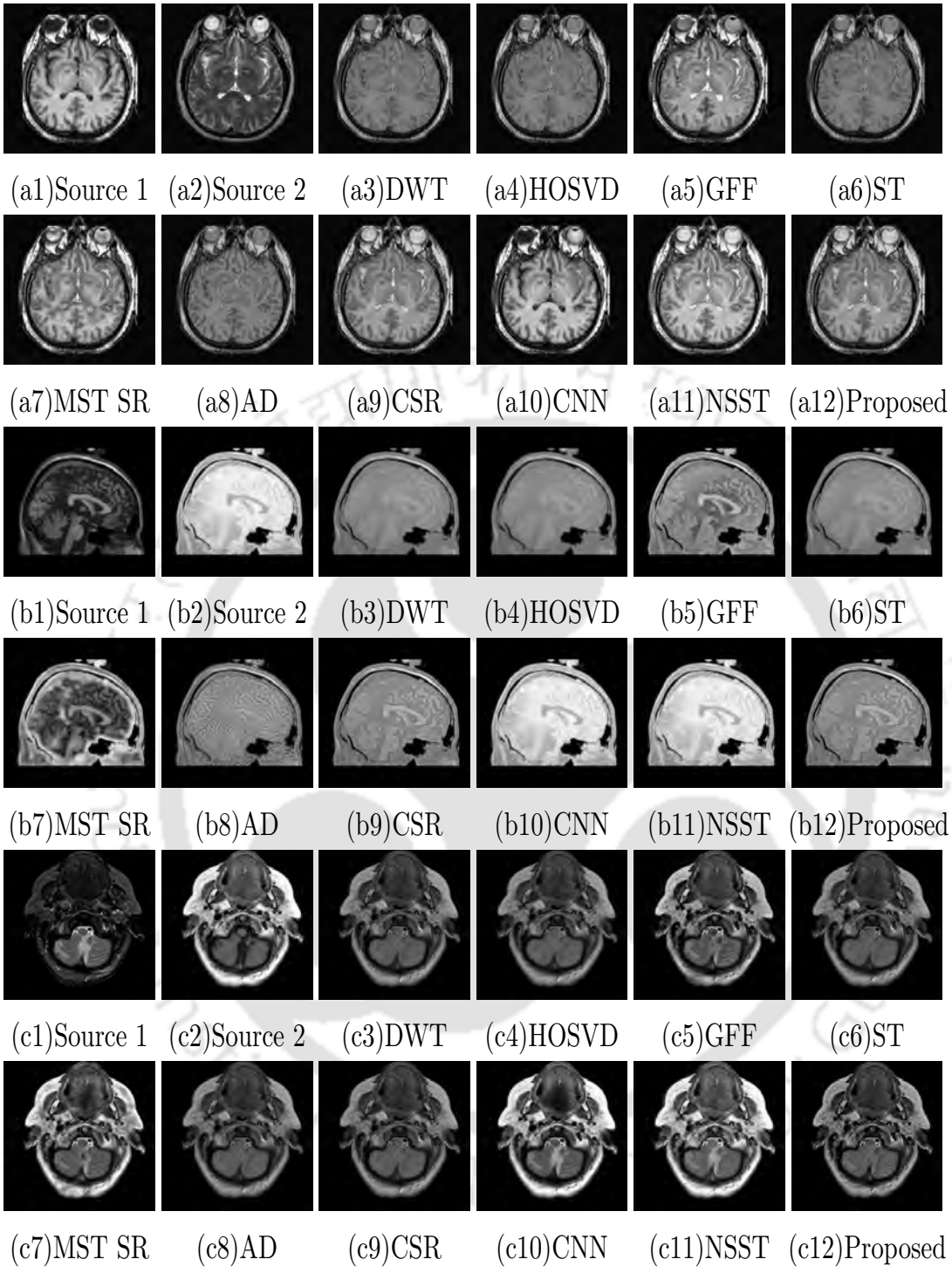


FIGURE 3.6: Medical source and the corresponding fused images obtained by using different image fusion methods.

3.4.1 Metrics for performance evaluation

The proposed method is tested using four fusion metrics (such as Q_{MI}, Q_S, $Q^{AB/F}$ and Q_E), where three metrics are commonly used and the fourth one is the proposed

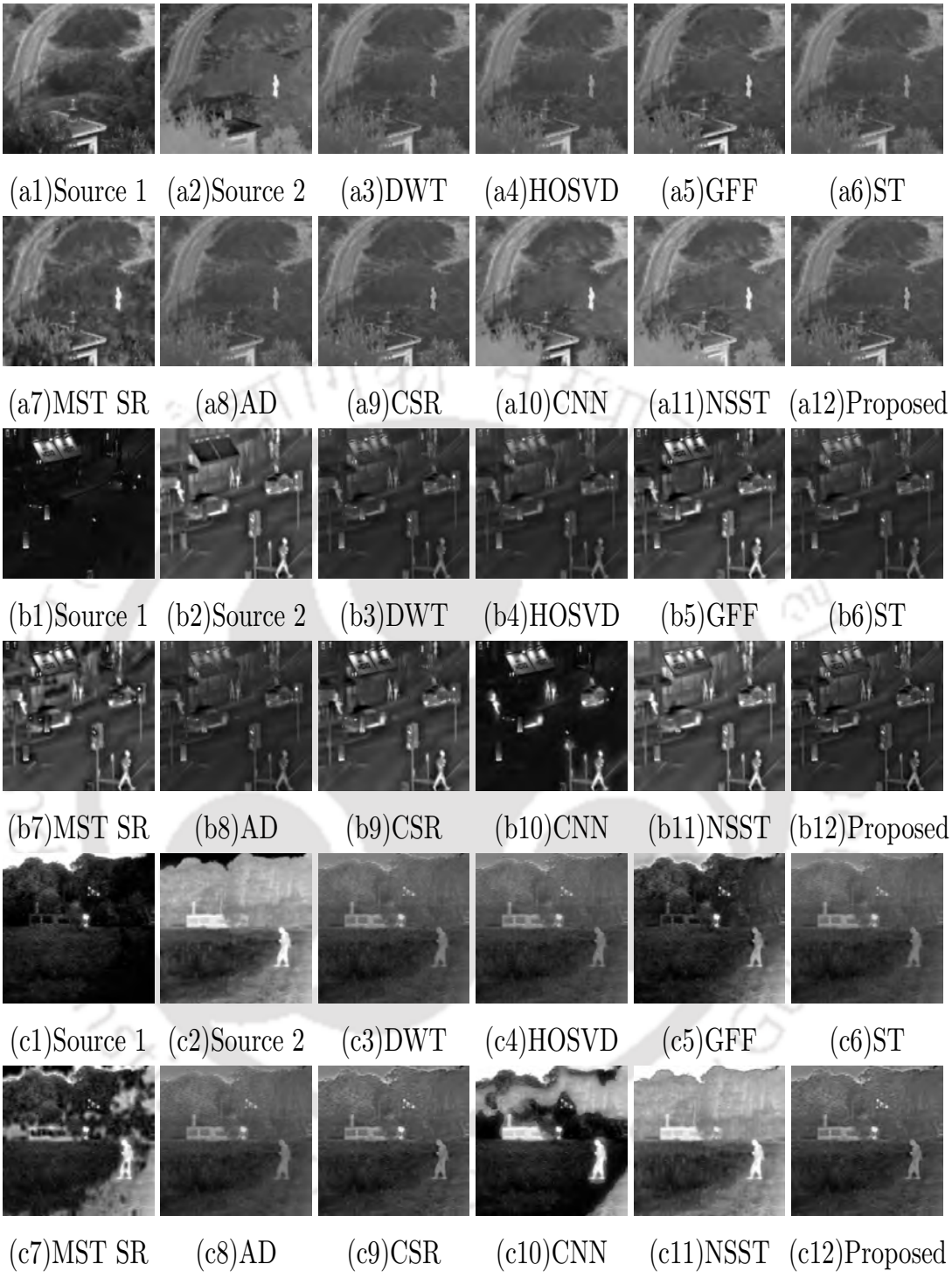


FIGURE 3.7: Visible-infrared source and the corresponding fused images obtained by different image fusion algorithms.

variance similarity metric Q_E. Details about these evaluation metrics can be found in Sections 1.2.5 and 2.4.1. Each metric analyzes specific aspects of the fused image with respect to the source images.

FIGURE 3.8: Multi-focus source and the corresponding fused images obtained by using different image fusion methods.

Proposed variance similarity metric (Q_E): KLT or principal component analysis is a theoretically optimum method for data representation. The KLT selects a basis set on the basis of the statistics of an image by estimating the covariance matrix of the image.

To formulate the metric Q_E, the following steps are taken: First, an image $\mathbf{X}$ of size $M \times N$ is considered, and the column mean $\mathbf{m}_{\mathrm{X}}$ (M dimensional) is computed by considering columns of image $\mathbf{X}$ as vectors. Let $\mathbf{x}$ is a set of column vectors of image $\mathbf{X}$. Then, the covariance matrix $\mathbf{C}_{\mathrm{X}}$ ($M \times M$ dimensional) is computed.

The non-zero non-diagonal elements of the covariance matrix $\mathbf{C}_{\mathrm{X}}$ indicate pixels dependency. The elements of $\mathbf{C}_{\mathrm{X}}$ can be nonzero or zero in principal diagonal. To ease the metric formulation step, we apply singular value decomposition (SVD) on $\mathbf{C}_{\mathrm{X}}$ to obtain stretch matrix $\mathbf{\Sigma}$ ($M \times M$ dimensional), which contains singular values of $\mathbf{C}_{\mathrm{X}}$. SVD operation generates factorization in the form of $\mathbf{U\Sigma V}^T$. In general, $\mathbf{U}$ is a real or complex unitary matrix, $\mathbf{\Sigma}$ is a diagonal (stretch) matrix with non-negative real numbers (singular values of $\mathbf{C}_{\mathrm{X}}$) on the diagonal, and $\mathbf{V}$ is a real or complex unitary matrix.

The singular values which are arranged in the descending order represent the strength of any regions of an image $\mathbf{X}$, *i.e.*, pixels have some similarities, such as intensity similarity, texture similarity, etc. These singular values represent the variance (stretch) information of the image $\mathbf{X}$. Now, we arrange singular values of $\mathbf{C}_{\mathrm{X}}$ in a vector $\mathbf{s}_{\mathrm{X}}$ (M dimensional) through the following diagonal ($diag$) operation:

$$\mathbf{s}_{\mathrm{X}} = diag(\mathbf{\Sigma}). \tag{3.10}$$

The vector $\mathbf{s}_{\mathrm{X}}$ gives variance (stretch) information of the image $\mathbf{X}$ in M orthogonal directions. For measuring variance information and pixels relationship of an image, the proposed metric Q_E is formulated. The higher values of Q_E indicates more variance information, *i.e.*, better fused image quality.

Let us take two source images $\mathbf{A}$, $\mathbf{B}$ and a fused image $\mathbf{F}$. Since, all three images $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{F}$ have equal size and let take the number of rows in each image is M. For m = 1,2,3,...,M, relative variance preservation vectors $\mathbf{s}^{AF}(m,1)$ and $\mathbf{s}^{BF}(m,1)$ of the source images $\mathbf{A}$ and $\mathbf{B}$ with respect to the fused image $\mathbf{F}$ are obtained as follows:

$$\mathbf{s}^{AF}(m,1) = \begin{cases} \mathbf{s}_F(m,1)/\mathbf{s}_A(m,1), & \text{if} \quad \mathbf{s}_A(m,1) > \mathbf{s}_F(m,1) \\ \mathbf{s}_A(m,1)/\mathbf{s}_F(m,1), & \text{otherwise.} \end{cases} \tag{3.11}$$

$$\mathbf{s}^{BF}(m,1) = \begin{cases} \mathbf{s}_F(m,1)/\mathbf{s}_B(m,1), & \text{if} \quad \mathbf{s}_B(m,1) > \mathbf{s}_F(m,1) \\ \mathbf{s}_B(m,1)/\mathbf{s}_F(m,1), & \text{otherwise.} \end{cases} \tag{3.12}$$

where, $\mathbf{S}_A(m,1)$, $\mathbf{S}_B(m,1)$ and $\mathbf{S}_F(m,1)$ are obtained by applying Eq. (3.10) to the images $\mathbf{A}, \mathbf{B}$ and $\mathbf{F}$, respectively. $\mathbf{s}^{AF}(m,1)$ and $\mathbf{s}^{BF}(m,1)$ are used to derive preservation values as:

$$\mathbf{q}^{AF}(m,1) = \left(1 + e^{-(\mathbf{s}^{AF}(m,1))}\right)^{-1} \tag{3.13}$$

$$\mathbf{q}^{BF}(m,1) = \left(1 + e^{-(\mathbf{s}^{BF}(m,1))}\right)^{-1} \tag{3.14}$$

$\mathbf{q}^{AF}(m,1)$ and $\mathbf{q}^{BF}(m,1)$ indicate perceptual loss of information in the fused image $\mathbf{F}$, which measure how much variance of the source images is present in the fused image. Sigmoid function is used to represent variance information preserve values. If $\mathbf{q}^{AF}(m,1) = 0$, then it corresponds to complete loss of variance information transferred from $\mathbf{A}$ to $\mathbf{F}$. While $\mathbf{q}^{AF}(m,1) = 1$ indicates fusion from $\mathbf{A}$ to $\mathbf{F}$ with no loss of texture information. Using $\mathbf{q}^{AF}(m,1)$ and $\mathbf{q}^{BF}(m,1)$, our proposed normalized weighted performance metric (Q_E) for $P \times Q$ size images is given as follows:

$$Q_E = \frac{\sum_{m=1}^{M} \left[\mathbf{q}^{AF}(m,1)\mathbf{s}^{AF}(m,1) + \mathbf{q}^{BF}(m,1)\mathbf{s}^{BF}(m,1)\right]}{\sum_{m=1}^{M} \left[\mathbf{s}^{AF}(m,1) + \mathbf{s}^{BF}(m,1)\right]}. \tag{3.15}$$

Metric Q_E lies between 0 and 1. For optimal fusion results, metric Q_{MI} should be high, values of Q_S, $Q^{AB/F}$ and Q_E should be close to 1.

3.4.2 Discussion on free parameters

Performance of the proposed fusion method is tested for different parameter settings. The dataset-1 of Fig. 3.4 is used to perform this task and average values of metrics, *i.e.*, Q_{MI}, Q_S, $Q^{AB/F}$ and Q_E are plotted. The default setting of the parameters are set as: Filter length $N = 9$, Kaiser window adjustable parameter $\beta = 5$, low-frequency fusion descriptor $\alpha = 10$, Canny edge detector with scale multiplication $scale = 4$, and unsharp masking gain $\gamma = 2$. In Fig. 3.5(a), N is varied and the other parameters are

set to default values. In a similar way, variations with respect to β, α, γ and *scale* are estimated.

The best suitable value of $N = 9$, with Kaiser window adjustable parameter $\beta = 5$, sigmoid parameter of low-frequency fusion $\alpha = 10$, $\gamma = 2$ and $scale = 4$. This parameter setting is finally used to generate fused images by our proposed method. It is observed from the experimental analysis shown in Fig. 3.5 that the proposed metric Q_E behaves like other three standard metrics. The proposed technique takes filter length of the non-subsampled Laplacian pyramid filters as 9×9, which is obtained from the 1-D QMF filters of length $N = 9$ with Kaiser window adjustable parameter $\beta = 5$.

3.4.3 Experimentation and analysis

The source images are pre-registered prior to fusion. Comparative analysis is performed with the following state-of-the-art methods: discrete wavelet transform (DWT) [24], higher order singular value decomposition (HOSVD) [160], guided filtering (GFF) [114], multi-scale transform and sparse representation (MST SR) [89], anisotropic diffusion (AD) [1], convolutional neural network (CNN) [118], convolutional sparse representation (CSR) [88], shearlets transform (ST) [132] and NSST [33].

For optimized non-subsampled Laplacian pyramid filters, filter lengths N of the prototype filter is set at 9 and adjustable parameter β of Kaiser window is set at 5. These values are selected on the basis of the experimental analysis illustrated in Fig. 3.5. First, a prototype filter is designed using the proposed method given in [147]. Then, the real-valued QMF filters are designed with the help of a prototype filter by using Eq. (3.1). Then by using the mapping method given in [157], 2-D filters of length 9×9 are designed. The four 2-D filters are subsequently used as decomposition and reconstruction filters of non-subsampled Laplacian pyramid. Two-level of optimized NSST decomposition is used to save memory and computational complexity. Two-level decomposition produces one low-frequency and four high-frequency bands. However, the number of noisy pixels and misregistration error increase with the number of decomposition level. This creates spatial artifacts in the fused image [89].

TABLE 3.1: Performance of image fusion methods using dataset-1 shown in Fig. 3.4.

Method	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_E
DWT	Medical	0.5918(0)	0.3085(0)	0.6603(0)	0.3736(0)
	Infrared	0.2536(0)	0.2793(0)	0.6079(0)	0.1567(0)
	Multi-focus	0.9839(0)	0.6109(0)	0.8377(0)	0.8959(1)
HOSVD	Medical	0.5839(0)	0.2790(0)	0.6465(0)	0.3763(0)
	Infrared	0.2397(0)	0.2699(0)	0.5970(0)	0.1589(0)
	Multi-focus	0.9557(0)	0.5982(0)	0.8348(0)	0.8921(1)
GFF	Medical	0.5585(0)	0.5172(1)	0.6678(0)	0.5575(0)
	Infrared	0.2034(0)	0.4989(2)	0.6454(0)	0.7123(3)
	Multi-focus	1.1613(1)	0.6726(0)	0.8222(0)	0.8860(0)
ST	Medical	0.7005(0)	0.4155(0)	0.5673(0)	0.4678(0)
	Infrared	0.3610(0)	0.3873(0)	0.5470(0)	0.2507(0)
	Multi-focus	1.0968(0)	0.6740(1)	0.8702(0)	0.9957(0)
MST SR	Medical	0.6614(0)	0.4887(0)	0.6794(0)	0.5749(0)
	Infrared	0.3452(0)	0.4314(0)	0.5942(0)	0.6289(0)
	Multi-focus	1.1614(1)	0.6704(0)	0.8245(0)	0.8852(0)
AD	Medical	0.5705(0)	0.2992(0)	0.6963(0)	0.3695(0)
	Infrared	0.2595(0)	0.3149(0)	0.6014(0)	0.1493(0)
	Multi-focus	1.0054(0)	0.6067(0)	0.8469(2)	0.8901(0)
CSR	Medical	0.8438(1)	0.5524(1)	0.7063(1)	0.5494(0)
	Infrared	0.5329(0)	0.4917(1)	0.5828(0)	0.5087(0)
	Multi-focus	1.0304(0)	0.6274(0)	0.8221(0)	0.8866(0)
CNN	Medical	**1.0345**(4)	0.5667(3)	**0.8048**(3)	0.7478(2)
	Infrared	0.5233(2)	0.4063(0)	0.6417(2)	0.6964(2)
	Multi-focus	**1.2114**(5)	0.6758(2)	0.8217(0)	0.8854(0)
NSST	Medical	0.9278(1)	0.5421(0)	0.7174(2)	0.7311(1)
	Infrared	0.5145(2)	0.5044(1)	0.6211(0)	0.6847(0)
	Multi-focus	1.0265(0)	0.6731(1)	0.8472(3)	0.9119(2)
Proposed	Medical	0.7173(2)	**0.5668**(3)	0.6878(2)	**0.8105**(5)
	Infrared	**0.5358**(4)	**0.5291**(4)	**0.7096**(6)	**0.7186**(3)
	Multi-focus	1.1520(1)	**0.6813**(4)	**0.8475**(3)	**0.9212**(4)

As explained earlier, the proposed superposition form obtained from the scale multiplied Canny edge detector features and Hessian features is used to fuse low-frequency band, and the proposed unsharp masking based rule is used to fuse high-frequency bands. The sigmoid function parameter $\alpha = 10$ in low-frequency coefficients fusion is taken to

get nearly equal contributions from the corresponding source pixels. In our method, the kernel size is selected as 3×3. Inverse optimized NSST is applied to get the fused image.

The source images contain partially complementary information. Hence, the resultant fused image can be used for better visualization of the scene, and the prominent information of the source images should present in the fused image. Fig. 3.6 shows medical image fusion, Fig. 3.7 shows visible-infrared image fusion, and finally Fig. 3.8 shows multi-focus image fusion, where the fused images are obtained by different image fusion algorithms. The objective analysis of the proposed method is carried out with the help of different evaluation metrics, and the results are shown in Table 3.1, and both subjective and objective analysis show the efficacy of our proposed fusion framework.

DWT-based fusion method is less computationally complex, and this method is highly efficient to detect point discontinuity in the images. The main drawback of this method is the presence of ringing artifacts in the fused images due to sampling rate conversion operations. If the source images are misregistered, then the transformed coefficients are changed due to sampling operation, which results sub-optimal fusion. Moreover, DWT method is inefficient to approximate curve-like information. HOSVD method produces artificial edges and blurred fused image. GFF based image fusion technique uses smoothed weight maps for image fusion. As the GFF method reasonably distribute the weights all over the image, the fusion result of GFF is comparable to MST SR. However, GFF can not always assign perfect weights in the texture or smooth regions of the source images. So, the texture information in the fused image is distorted, and the fused image suffers from blurring artifacts. ST method [132] produces ringing artifacts in the fused image due to the subsampled nature of the shearlet. Moreover, averaging fusion rule for low-frequency bands produces blurry fused images. The absolute value maximum selection fusion rule for high-frequency bands selects noisy pixels, which degrades the quality of the fused image.

MST SR based image fusion techniques utilize the advantages of both MST and sparse representation based techniques. Both MST and sparse representation based methods produce blurred fused images. Moreover, MST SR method unable to produce high-quality fusion results when the source images have strong edges [88]. AD based

method suffer from blurring and low contrast. CSR modelling for image fusion is not efficient mainly due to inefficient representation of fine information of source images during dictionary formation. The CNN method based fused image suffers from high brightness and other spatial distortions such as missing some details in the fused image. NSST method [33] is developed to fuse multi-focus images. Due to its fusion rules, it is unsuitable for fusion of medical and visible-infrared images. The fused images obtained by this method have high brightness and spatial artifacts. Moreover, NSST method based multi-focus fused images contain less high-frequency information due to the restrictions imposed by fusion rules. Table 3.1 shows the efficacy of the proposed metric Q_E, as this metric can also efficiently judge the quality of the fused images.

The proposed metric Q_E is suitable for evaluation of fused images when the source images are partially different and they have complementary detailed information. For example, in Table 3.1, the quantitative comparisons between CNN and the proposed method show very close results. CNN produces higher values of Q_{MI} and Q_S for medical images. However, in subjective comparisons, CNN produces low-quality fused images. Hence, the metric Q_E shows lower values for CNN based method, which is well justified. For multi-focus images, CNN produces higher value of Q_{MI}, which is as par the subjective results and the proposed metric Q_E. The metric Q_E also shows high values for CNN as compared to $Q^{AB/F}$ and Q_S.

In Fig. 3.9, source images are contaminated with Gaussian noise with zero mean and variance of 0.01. This is done to show the effectiveness of the proposed metric Q_E in noisy source images. In Fig. 3.9, GFF, CSR and proposed methods produce high-quality fused images for noisy medical source images as compared to other methods. Therefore, Q_E is higher for GFF, CSR and proposed methods as shown in Table 3.2. However, Q_E shows lower values for MST SR and CNN methods as these methods produce spatially distorted fused images for noisy visible-infrared source images. Subjective analyses as illustrated in Fig. 3.9 demonstrated that CSR and the proposed method produce high-quality fused images as compared to other methods. Therefore, Q_E is higher for both CSR and the proposed methods as per Table 3.2. However, Q_E shows lower values for MST SR method, as this method produce spatial artifacts in the fused images.

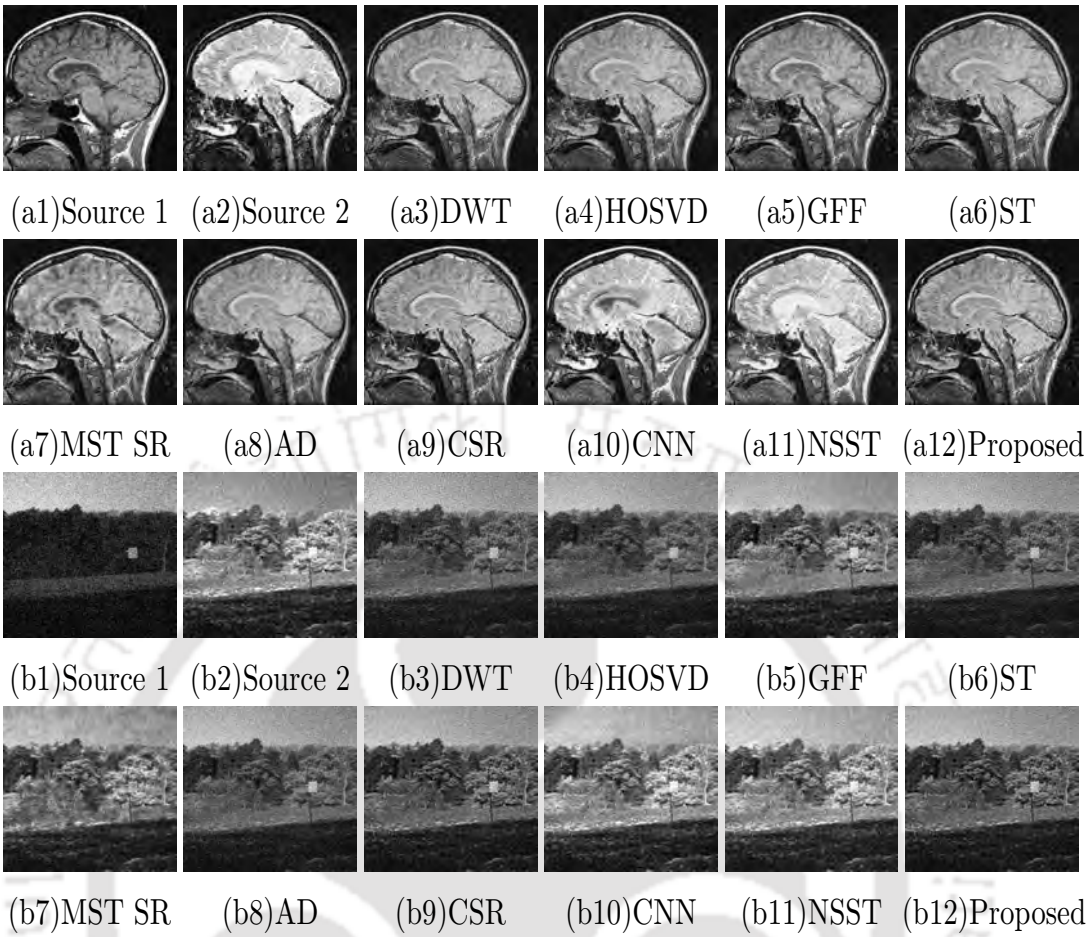


FIGURE 3.9: Noisy source and the corresponding fused images obtained by using different image fusion methods.

The reason behind the superior performance of our proposed filters and the overall fusion framework is due to the application of our proposed optimized NSST and efficient fusion rules. In our proposed fusion framework, the coefficients of the filters are adjusted by iterating passband edge frequency ω_p for achieving best fusion results. Furthermore, our proposed unsharp masking based fusion rule for high-frequency coefficients is quite efficient in preserving information of the source images. The high-frequency information of the source images is efficiently captured by unsharp masking based feature map. The lower value of sigmoid function parameter α used in low-frequency fusion descriptor degrades the quality of the fused image by selecting relatively maximum intensity values. This maximum intensity selection setting underestimates low intensity pixels, specially in the fusion of medical and visible-infrared images. The higher values of α such as 10, allows contributions (by averaging operation) from the source pixels, which is desired in optimal fusion.

TABLE 3.2: Performance of image fusion methods in noisy source images for Fig. 3.9.

Methods	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_E
	Medical	0.3768	0.3246	0.6161	0.9331
DWT	Infrared	0.354	0.31	0.6678	0.7047
	Medical	0.3709	0.3164	0.595	0.9329
HOSVD	Infrared	0.3452	0.2131	0.4989	0.5033
	Medical	0.403	0.4461	0.6584	0.7908
GFF	Infrared	0.4583	0.5141	**0.8447**	0.8745
	Medical	0.3897	0.3507	0.6294	0.9206
ST	Infrared	0.3754	0.3447	0.6948	0.7187
	Medical	0.3835	0.415	0.6062	0.7775
MST SR	Infrared	0.3768	0.4265	0.7832	0.4886
	Medical	0.4579	0.3262	0.6562	0.8329
AD	Infrared	0.3971	0.2996	0.615	0.4937
	Medical	0.4278	0.4423	0.5931	0.9346
CSR	Infrared	0.4211	0.4849	0.7753	0.8767
	Medical	**0.581**	0.431	0.7104	0.7885
CNN	Infrared	0.4662	**0.6051**	0.8248	0.8689
	Medical	0.4344	0.3959	0.6295	0.9331
NSST	Infrared	0.4157	0.3558	0.7688	0.8206
	Medical	0.5294	**0.4562**	**0.7224**	**0.936**
Proposed	Infrared	**0.4694**	0.371	0.7463	**0.8821**

To produce high-quality fused images, it is important to preserve both significant low- and high-frequency information. In our proposed fusion method, equal importance is given to both low- and high-frequency information. For high-frequency information, the unsharp masking based fusion rule amplifies high-frequency coefficients using the gain parameter (γ). The gain parameter can be adjusted to select high-frequency coefficients. Moreover, the visual quality of the fused image is further improved by multiplying all the unsharp masking feature images of different scales and orientations. For low-frequency information, the low-frequency bands are fused by using the proposed descriptor, which is obtained from scale multiplied canny edge detector and Hessian features. In Fig. 3.3, SMC and HEF feature are shown for low-frequency band images. In our proposed method, combined SMC and HEF features are used to fuse low-frequency bands to retain low-frequency information.

TABLE 3.3: The objective image fusion performance metrics for subjective similarity

Metric	Q_{MI}	$Q^{AB/F}$	Q_S	Q_E
relevance (r)	0.741	0.815	0.811	**0.856**
Correct Ranking (CR)	65.3%	74.4%	75.1%	**77.2%**

The performance of the proposed method is objectively compared, and the results are shown in Tables 3.1 and 3.2, where in Table 3.1 the numbers within the parentheses indicate number of image pairs that a particular approach outperforms other approaches. The proposed method is efficient to seize significant information as compared to the previous methods due to the following reasons. In optimize NSST, non-subsampled Laplacian pyramid filter coefficients are adjusted to obtain high-quality fused images. In previous fusion methods, the concept of optimized NSST is not explored. The main objective of our proposed optimization is to make non-subsampled Laplacian pyramid filters frequency response close to the ideal filter frequency responses. Moreover, significant information, such as edges, corners, contours, fine and variational details of the source images are captured using the proposed low-frequency descriptor and the unsharp masking technique.

The deployment of scale multiplied canny edge detector in our proposed method highlights the valid edges in the low-frequency bands. The Hessian features highlight contours, corner and fine details of the low-frequency bands. The unsharp masking based high-frequency fusion rule selects the edges, corners and contours of the source images. The best values of the parameters of our proposed algorithm are experimentally selected as shown in Fig. 3.5, and the corresponding parameter settings make the proposed method adaptive and efficient. Compared to the existing methods, more effective tuning parameters, such as filter length (N), Kaiser window adjustable parameter (β), low-frequency fusion sigmoid parameter (α) used in our algorithm improve the performance of the proposed method. The optimized NSST is proposed to represent edge, texture and detailed information more efficiently. Additionally, our proposed fusion rules can effectively utilizes all these information to fuse images.

Validation of proposed Q_E metric: The proposed variance similarity metric Q_E is validated using the Correct Ranking (CR) and subjective relevance (r) measures [124].

TABLE 3.4: Comparison between the proposed filter and other non-subsampled Laplacian pyramid filters using dataset-1 shown in Fig. 3.4.

Metrics	*maxflat*	*9-7*	*pyr*	*pyrexc*	Proposed
Q_{MI}	0.730(3)	0.729(4)	0.728(0)	0.728(2)	**0.743**(15)
$Q^{AB/F}$	0.535(3)	0.535(0)	0.537(0)	0.537(0)	**0.588**(21)
Q_S	0.733(0)	0.735(0)	0.739(0)	0.749(6)	**0.761**(18)
Q_E	0.709(2)	0.710(0)	0.712(3)	0.712(2)	**0.837**(17)

The dataset-1 shown in Fig. 3.4 is used to perform this validation task. For validation, the r measure should be close to 1 and CR should be close to 100%. The r and CR values in Table 3.3 show that the best performance is obtained for Q_E. The performance of $Q^{AB/F}$ and Q_S are almost same, whereas Q_{MI} score gives lowest values of r and CR. Hence, this analysis indicates that our proposed metric Q_E can be used to effectively compare different image fusion algorithms. Moreover, this metric (Q_E) can be used for measuring the performance of source images having low-texture and low-contrast such as infrared images, night-vision scenario images, etc. However, this metric can not jointly measure texture and structural information in contrast to our previous proposed metric Q_{TS} (as mentioned in Section 2.4.1).

3.4.4 Comparison of various family of non-subsampled Laplacian pyramid filters

The next set of experiments is carried out on dataset-1 for comparison of different non-subsampled Laplacian pyramid filters. Four metrics defined earlier are used to compare the performance of proposed filters with the existing non-subsampled Laplacian pyramid filters. Table 3.4 shows the metric responses for different filters, and the proposed filter outperforms other existing filters for non-subsampled Laplacian pyramid. The brief regarding non-subsampled Laplacian pyramid filters can be found in Section 2.4.6

3.4.5 Computational efficiency analysis

The comparison of computational efficiencies of the different state-of-the-art fusion methods is given in Table 3.5. Our proposed method is implemented in MATLAB R2017a

TABLE 3.5: Average run time of different image fusion methods for dataset-1.

Method	DWT	HOSVD	GFF	MST SR
Time/second	0.56	0.73	0.33	17.32
Method	AD	CSR	CNN	Proposed
Time/second	0.66	37.34	35.57	**0.43**

platform on a computer with a processor speed of 3.20 GHz, 64-bit operating system and 8 GB RAM. As shown in Table 3.5, CSR, CNN and MST SR are very time consuming methods, as NSCT, sparse coding and learning are computationally complex. The HOSVD, AD and DWT methods have almost similar and moderate time complexity. Our method is having relatively similar time complexity with GFF method. Moreover, proposed and GFF methods uses relatively less complex fusion rules and framework. However, our method can produce better fused images than the standard methods in most of the instances.

3.5 Conclusion

Image fusion is a very challenging research area as the main challenge is to preserve prominent details of the source images in the fused image. In this chapter, we proposed an optimized NSST based fusion method for multi-sensor and multi-modal images. For this, fusion rules are proposed to select high- and low-frequency coefficients by employing unsharp masking, and superposition of scale multiplied Canny edge detector features and Hessian features. Deployment of scale multiplied Canny edge detector features and Hessian features in the fusion mechanism enhances the information in the fused images. Also, a metric Q_E is proposed to measure variance information in the fused images.

Experiments show that the fusion performance evaluations by the proposed and others objective measures agree with the subjective analyses. Experimental results explicitly show that our metric Q_E is perceptually significant and essential to judge the performance of fusion algorithms. The proposed method gives quite reliable and consistent results for different modalities of images. We believe that our technique can be implemented in a real-time application using a more efficient programming approach and architecture.

The proposed method is able to produce state-of-the-art results in case of multi-focus images. The proposed method has highest time efficiency as compared to the methods proposed in **Chapter 2** and **Chapter 4**. MST and MGA tools (such as NSST, curvelet, contourlet, etc.)-based low-frequency band fusion generally under perform as compared to sparse representation-based low-frequency fusion [89]. However, computational complexity of fusion algorithms increases due to the joint use of sparse representation with either MST or MGA tools.

4

Image Fusion using Adjustable Non-subsampled Shearlet Transform

Multi-scale geometrical transform (MST) and sparse representation based image fusion techniques are widely used in many applications. However, existing fusion methods are inefficient to perfectly seize significant details and texture information of input images because of some limitations of MST- and sparse representation based techniques mainly, ringing artifacts due to misregistration error. Moreover, it is quite important to measure singularities and structural information of the fused image to objectively measure the performance of different fusion techniques. So, a metric which can efficiently measure singularity and structural information of the fused image would be quite useful for comparison of different fusion techniques. To address these issues, a modified Meyer window based adjustable non-subsampled shearlet transform (ANSST) is proposed for decomposition of pre-registered input images into low- and high-frequency coefficients. The low-frequency bands are fused by convolutional sparse representation (CSR) modelling. The high-frequency bands are fused by our proposed Information Entropy, Standard Deviation, and Range (IESDR) descriptor, which consider entropy, standard deviation and range filtering features. Moreover, a metric Q_{SS} is formulated to measure singularities and structural information of the fused image. The Q_{SS} metric is developed by using the concept of the proposed native division and native difference filtering. To show the applicability of the proposed image framework, the proposed ANSST is applied for image denoising. Experimental results verify the effectiveness of proposed fusion method and the metric Q_{SS}.

4.1 Introduction

The proposed method uses two signal processing tools, namely NSST and sparse representation to perform image fusion. However, NSST or sparse representation can generate reasonably good fused images, but fusion techniques based on these tools suffers from few drawbacks.

The fused images generated only using MST method have low contrast especially in the case of multimodal image fusion. Moreover, selection of decomposition level in MST methods is a crucial task. Spatial features of images can not be extracted perfectly if less number of decomposition levels are used. While many decomposition levels produce noisy pixels and bad quality reconstruction when the source images are not perfectly registered. The performance of MST based methods is mainly degraded due to non-perfect registration. In the proposed technique, CSR method is combined with the MST to address the above-mentioned problems. Also, max rule is incorporated in our method to preserve image contrast.

The conventional sparse representation based fusion methods [89, 161–164] suffer from few drawbacks, such as blurry representation of high-frequency information in the fused image. The sparse representation methods have limited ability in detailed information representation and they are highly sensitive to non-perfect registration. The problems of detailed information representation and shift-variance of sparse representation are resolved by using CSR.

CSR is a shift-invariant image representation technique, which can give better performance even if the images are misregistered. However, due to the limitation of dictionary training algorithms, CSR and sparse representation-based method are inefficient to represent fine details of input images. Hence, in our proposed method, MST (ANSST) method is combined with the CSR to represent fine details of the input images in the fused image efficiently.

The main contributions of the proposed work are as follows. First, ANSST is reformulated by incorporating the parameter ξ, which provides enhanced versatility to obtain better fusion and denoising performance. The proposed modified Meyer window

is employed to bring adaptability of the NSST coefficients, which can be used to adjust the frequency response of ANSST filters. Second, IESDR descriptor is formulated to fuse high-frequency bands, which combines texture, edges, contours, fine information of source images. The deployment of IESDR descriptor significantly improves the quality of the fused images. Third, ANSST and CSR is jointly used in our framework to address the limitations of ANSST and CSR. Finally, the metric Q_{SS} is formulated to measure singularities and structural details of the fused image. The metric Q_{SS} is formulated based on proposed native division and native difference filtering. Experimental results indicate the effectiveness of the proposed image fusion structure and the proposed metric Q_{SS}. Validation of the proposed metric Q_{SS} is performed by analyzing the impacts of free parameters on the proposed fusion structure and also by using the method proposed in [124].

4.2 Preliminaries

4.2.1 Non-subsampled shearlet transform (NSST)

Shearlet transform (ST) produces a set of well-localized waveforms at various scales, locations, and orientations [31]. In NSST, the source image is decomposed into low-pass and high-pass sub-images using non-subsampled Laplacian pyramid. High-pass sub-images are fed to shift-invariant shear filters and the low-pass sub-image is further decomposed into low-pass and high-pass sub-images. The inverse NSST transform is applied by taking the sum of all shift invariant shearing filter responses at the respective levels of decomposition, and inverse non-subsampled Laplacian pyramid is finally applied to get the source image. The detailed explanation of NSST is given in Section 2.2.1.

4.2.2 Convolutional sparse representation (CSR) based fusion

CSR is an alternative method for finding sparse representation of an entire image rather than considering local patches. In CSR, any image $\mathbf{I}$ is modelled as the sum over a set of convolutions among dictionary filters $\mathbf{d}_o$ and sparse coefficient maps $\mathbf{x}_o$. The modelling

is done as follows:

$$\underset{\{\mathbf{x}_o\}}{\arg\min}\ \frac{1}{2}\left\|\sum_o \mathbf{d}_o * \mathbf{x}_o - \mathbf{I}\right\|_2^2 + \lambda \sum_o \|\mathbf{x}_o\|_1 \tag{4.1}$$

where, " $*$ " and "λ" indicate the convolution operator and regularization parameter, respectively. CSR is designed to become shift invariant [117].

Let $\{\mathbf{L}_A(m,n), \mathbf{L}_B(m,n)\}$ are low-frequency sub-images obtained after NSST decomposition of source images $\{\mathbf{A}, \mathbf{B}\}$, and a set of dictionary filters $\mathbf{d}_o$, $o \in \{1, ..., O\}$ are learned by method given in [117]. The base layer $\mathbf{L}_{b_A}(m,n)$ of source image $\mathbf{A}$ from the low-frequency image $\mathbf{L}_A(m,n)$ is obtained by solving the following optimization problem:

$$\underset{\mathbf{L}_{b_A}(m,n)}{\arg\min} \|\mathbf{L}_A(m,n) - \mathbf{L}_{b_A}(m,n)\|_F^2 + \eta\left(\|\mathbf{g}_x * \mathbf{L}_{b_A}(m,n)\|_F^2 + \|\mathbf{g}_y * \mathbf{L}_{b_A}(m,n)\|_F^2\right) \tag{4.2}$$

where, $\eta = 5$, horizontal and vertical gradient operators are indicated by $\mathbf{g}_x = \begin{bmatrix} -1 & 1 \end{bmatrix}$ and $\mathbf{g}_y = \begin{bmatrix} -1 & 1 \end{bmatrix}^T$, respectively. After getting base layer $\mathbf{L}_{b_A}(m,n)$, the detailed layer $\mathbf{L}_{d_A}(m,n)$ are obtained as follows:

$$\mathbf{L}_{d_A}(m,n) = \mathbf{L}_A(m,n) - \mathbf{L}_{b_A}(m,n) \tag{4.3}$$

Now, the detailed layers $\{\mathbf{L}_{d_A}(m,n), \mathbf{L}_{d_B}(m,n)\}$ and base layers $\{\mathbf{L}_{b_A}(m,n), \mathbf{L}_{b_B}(m,n)\}$ are fused using CSR model. Here 'm,n' are index of spatial locations. The base layers $\{\mathbf{L}_{b_A}(m,n), \mathbf{L}_{b_B}(m,n)\}$ are fused using a max rule for multi-focus images. In case of medical and visible-infrared images, an averaging rule is applied [88].

4.3 Proposed Method

The block diagram of the proposed method is given in Fig. 4.1. In our proposed method, ANSST is proposed to decompose the pre-registered source images $\{\mathbf{A}, \mathbf{B}\}$ into low-frequency sub-images $\{\mathbf{L}_A(m,n), \mathbf{L}_B(m,n)\}$ and high-frequency sub-images $\{\mathbf{H}_{k,l}^A(m,n), \mathbf{H}_{k,l}^B(m,n)\}$. Here, k and l indicate the decomposed high-frequency sub-images in k^{th} direction at l^{th} decomposition level. The CSR model is used to further decompose

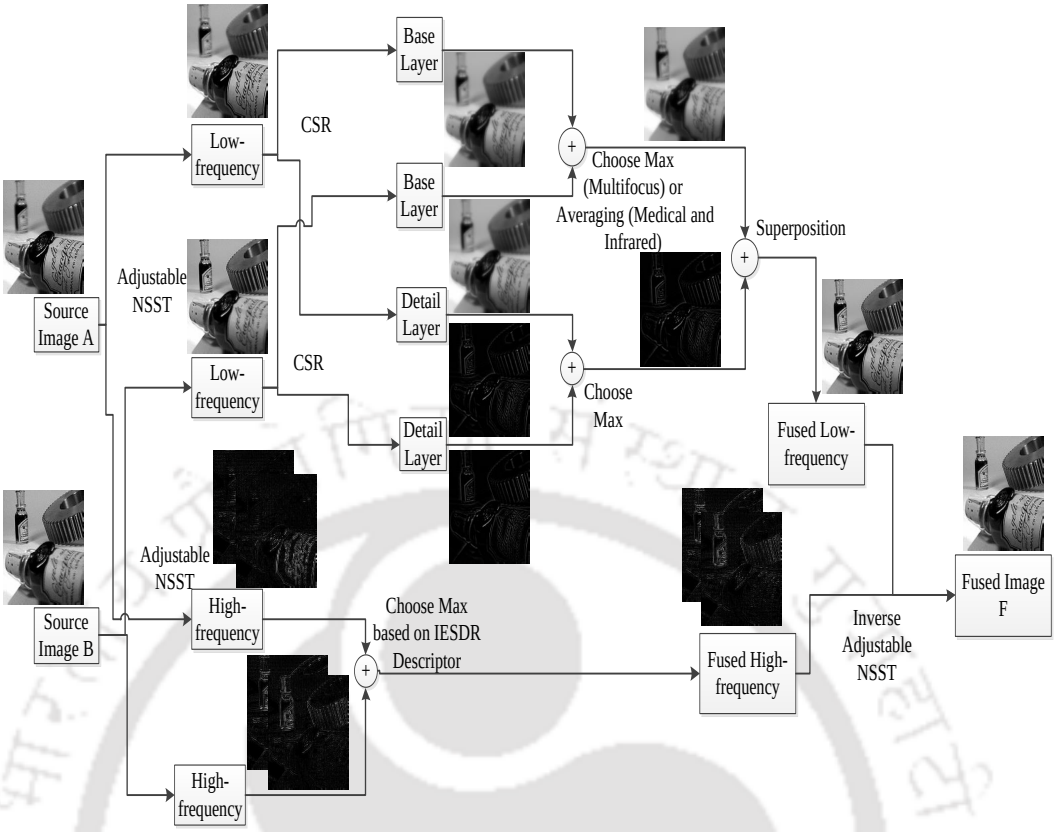


FIGURE 4.1: Schematic framework of the proposed image fusion method.

low-frequency sub-images into base layers $\{\mathbf{L}_{b_A}(m,n), \mathbf{L}_{b_B}(m,n)\}$, and detailed layers $\{\mathbf{L}_{d_A}(m,n), \mathbf{L}_{d_B}(m,n)\}$. The high-frequency sub-images $\{\mathbf{H}^A_{k,l}(m,n), \mathbf{H}^B_{k,l}(m,n)\}$ are obtained after ANSST decomposition. On these decomposed bands, depending upon the modalities of source images, fusion rules are applied to get the fused bands. Finally, the fused image $\mathbf{F}$ is obtained by applying inverse ANSST operation using $\mathbf{L}_F(m,n)$ and $\mathbf{H}^F_{k,l}(m,n)$. The detail description of the Fig. 4.1 has been presented in Section 4.3.6[5].

4.3.1 Modified Meyer window

In our proposed method, the traditional Meyer window function [31] is modified to get adaptability in shearing filters of NSST. The NSST with modified Meyer window is

[5]This work has been published in IEEE Transaction in Instruementation and Measurement 2018 (Refer item 4 for details).

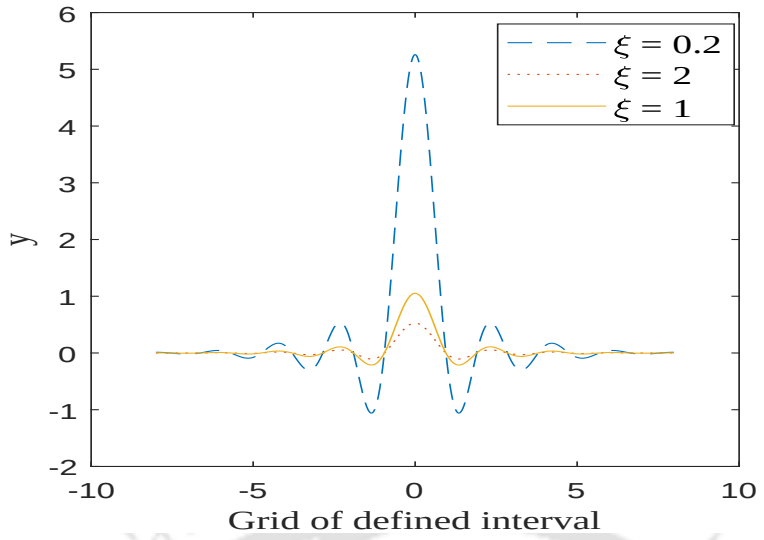


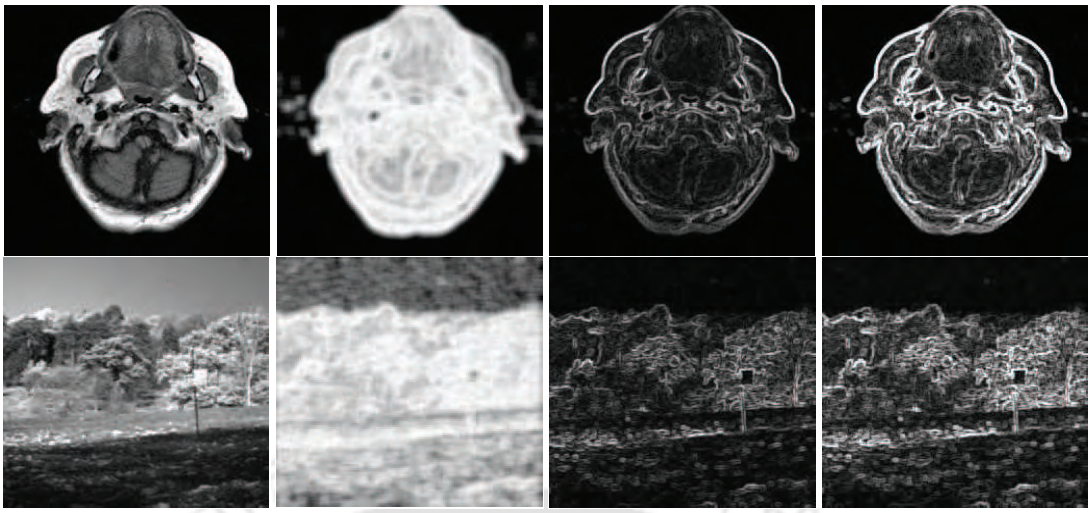
FIGURE 4.2: Comparison of Meyer window function with different values of ξ.

termed as ANSST. The Meyer window is modified as follows:

$$y = \begin{cases} 1/\xi, & 1/6 < x < 5/6 \\ \cos^2\left(\frac{\pi z}{2}\right)/\xi, & 5/6 \leq x \leq 7/6 \quad or \quad -1/6 \leq x \leq 5/6 \\ 0, & otherwise \end{cases} \tag{4.4}$$

where, y is windowed signal, $z = w^4(35 - 84w + 70w^2 - 20w^3)$ and $w = 3\left|x - \frac{1}{2}\right| - 1$.

where, ξ is a parameter used to give extra degree of freedom for overall ANSST system as compared to conventional one. Moreover, it is also used to shape the passband and stopband response of shearing filters. With parameter ξ, the response of Meyer window can be adjusted to search more appropriate coefficients of shearing filters. This is a heuristic approach to adjust the shearing filter coefficients by modifying the Meyer window. Fig. 4.2 shows the comparison of Meyer window functions with different values of ξ. The modified Meyer window has to satisfy the three main properties of Meyer window. First, the adjustable Meyer window is also compactly supported in time domain under normalization operation. Second, adjustable Meyer window is also uniformly smooth under normalization operation. Third, the vanishing moments of adjustable Meyer window is preserved under normalization operation. In our method, $\xi \in [0.1 \quad 10]$ and typical value $\xi = 0.2$ is considered. Reason for choosing this value will be justified in the experimental and analysis section.



Original image Entropy Standard deviation Range

FIGURE 4.3: Results of IESDR descriptor features.

4.3.2 IESDR descriptor

IESDR descriptor is obtained by superimposing information of entropy, standard deviation, and range filtering features. This descriptor is proposed to efficiently capture prominent high-frequency coefficients corresponding to edges and corners from the decomposed coefficients. The standard deviation of an image gives image intensity variation information. The entropy measures complexity and randomness of an image region. The range filtering operation can extract texture information present in the source images. Hence, individual use of any one of these features can not capture all the significant high-frequency information. Therefore, combination of these three features as shown in Fig. 4.3 highlights the above mentioned (edges, corners, intensity variation, complexity, randomness, texture) significant high-frequency information of an image. That is why, all these features are judiciously combined to propose our IESDR descriptor.

Let, windowed information entropy (WIE) filtering [165], windowed standard deviation (WSD) filtering [166] and windowed range (WR) filtering [167] are applied on $\mathbf{H}^{A}_{k,l}(m,n)$, and we obtained $\mathbf{WIE}^{A}_{k,l}(m,n)$, $\mathbf{WSD}^{A}_{k,l}(m,n)$ and $\mathbf{WR}^{A}_{k,l}(m,n)$, respectively. Based on these filtering responses, the superimposed IESDR descriptor is formulated as follows:

$$\mathbf{IESDR}^{A}_{k,l}(\mathrm{m},\mathrm{n}) = \{1+\exp\{-\mathbf{WIE}^{A}_{k,l}(\mathrm{m},\mathrm{n})-\mathbf{WSD}^{A}_{k,l}(\mathrm{m},\mathrm{n})-\mathbf{WR}^{A}_{k,l}(\mathrm{m},\mathrm{n})\}\}^{-\alpha} \quad (4.5)$$

TABLE 4.1: Comparison of left difference and central difference using images of Fig. 4.5

Metrics	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{SS}
Left Difference	**0.7806**	0.6273	0.8562	0.3281
Central Difference	0.7763	**0.6359**	**0.8661**	**0.3447**

where, α is sigmoid high-frequency fusion or IESDR descriptor parameter.

4.3.3 Native differences

To get the features from high-frequency sub-images, difference of center pixel to its neighbourhood pixels are taken into consideration. The proposed native difference features are extracted as given in Eq. 4.6. Let $\mathbf{F}(i,j)$ be an image of size $M \times N$ and size of window is $w_1 \times w_2$. By applying native difference filtering, an image $\mathbf{S}_F(i,j)$ is obtained. In Eq. 4.6, the window size is 3×3.

$$\mathbf{S}_F(i,j) = \sum_{w_1,w_2=-1}^{1} |\mathbf{F}(i,j) - \mathbf{F}(i+w_1, j+w_2)| \tag{4.6}$$

4.3.4 Application of central difference as prior in optimization

For comparison of left or right difference used in Eq. 4.2 with the central difference, objective and subjective comparisons are carried out. The central difference kernels defined as $\mathbf{g}_x = [-1 \quad 0 \quad 1]$ and $\mathbf{g}_y = [-1 \quad 0 \quad 1]^T$ are used for Eq. 4.2. The central difference performs better than the left difference [88], and the results are given in Fig. 4.4 and Table 4.1. Dataset-1 shown in Fig. 4.5 is used for this comparative analysis and average metrics values are presented in Table 4.1. The regions of the images which have some visual distinctions are marked by arrows. In Fig. 4.4, the fused images obtained using the central difference have comparatively less edge distortions and the contrast is preserved in these images. These visual differences are evident at higher zooming. Based on these analyses, central difference is employed in our proposed method. The reason is that the error produced by performing right or left difference is of the first order of the interval. While error produced by central difference operation is of the second order of the interval.

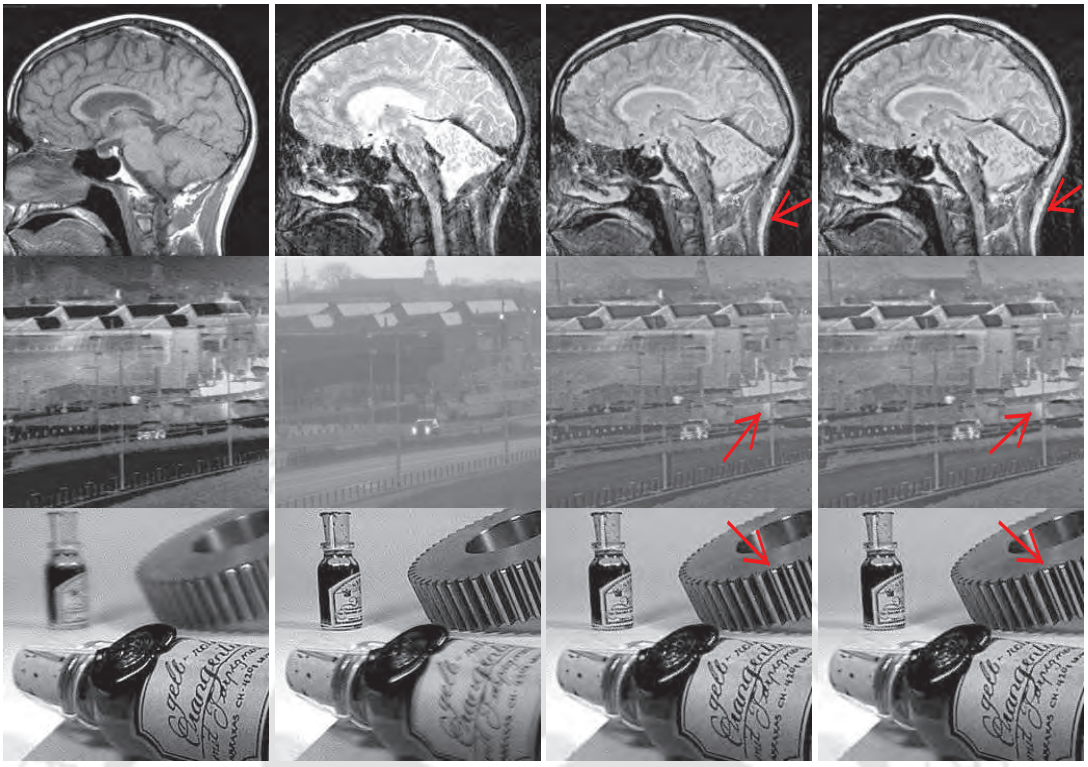


FIGURE 4.4: Comparison between left difference and central difference.

4.3.5 Local division filtering

Let $\mathbf{A}(m,n)$ represent one of the source image and $\mathbf{F}(m,n)$ represent the fused image. The native division features $\mathbf{R}_A(m,n)$ in a square window $\mathbf{P}$ for the any image $\mathbf{A}(m,n)$ is calculated as:

$$\mathbf{R}_A(m,n) = \min_{x,y \in \mathbf{P}} \frac{\min\{\mathbf{A}(m+x,n+y),\mathbf{A}(m,n)\}+\kappa}{\max\{\mathbf{A}(m+x,n+y),\mathbf{A}(m,n)\}+\kappa} \tag{4.7}$$

where, "x,y" belong to the window $\mathbf{P}$ of size 3×3, and κ is fixed at 0.001 to avoid dividing by 0. For example, native division filtering on a small region of an image is

implemented as follows:

$$\text{Native division filtering on} \left\{ \begin{bmatrix} 1 & 2 & 4 \\ 2 & 3 & 1 \\ 4 & 2 & 9 \end{bmatrix} \right\} = \min \left\{ 10^4 \begin{bmatrix} 52.25 & 124.18 & 58.82 \\ 143.79 & 176.47 & 66.44 \\ 68.62 & 113.28 & 26.14 \end{bmatrix} \right\} = 26.14 \times 10^4 \tag{4.8}$$

$$\text{Native division filtering on} \left\{ \begin{bmatrix} 8 & 7 & 9 \\ 7 & 8 & 7 \\ 8 & 7 & 1 \end{bmatrix} \right\} = \min \left\{ 10^4 \begin{bmatrix} 107.84 & 177.55 & 95.86 \\ 181.37 & 255.44 & 148.84 \\ 107.84 & 152.66 & 16.10 \end{bmatrix} \right\} = 16.10 \times 10^4 \tag{4.9}$$

Compared to the neighbourhood values, a low value is obtained for 9 in Eq. 4.8, which is 26.14×10^4. Similarly, a low value is obtained for 1 in Eq. 4.9, which is 16.10×10^4.

4.3.6 The Proposed fusion framework

The complete block diagram of the proposed method is shown in Fig. 4.1. The pre-registered source images $\{\mathbf{A}, \mathbf{B}\}$ are first decomposed into low-frequency sub-images $\{\mathbf{L}_A(m,n), \mathbf{L}_B(m,n)\}$ and high-frequency sub-images $\{\mathbf{H}^A_{k,l}(m,n), \mathbf{H}^B_{k,l}(m,n)\}$ by using ANSST.

Fusion rule for low-frequency band: In our proposed method, CSR based fusion method [88, 117] is employed to fuse low-frequency sub-images. The low-frequency sub-images $\{\mathbf{L}_A(m,n), \mathbf{L}_B(m,n)\}$ are further decomposed using CSR model into base layers $\{\mathbf{L}_{b_A}(m,n), \mathbf{L}_{b_B}(m,n)\}$ and detail layers $\{\mathbf{L}_{d_A}(m,n), \mathbf{L}_{d_B}(m,n)\}$.

Detail layer fusion

For fusion of detail layers $\{\mathbf{L}_{d_A}(m,n), \mathbf{L}_{d_B}(m,n)\}$, the CSR model is used. For fusion of detail layers, the CSR model is solved as given in Eq. 4.1 in the following form:

$$\underset{\{\mathbf{C}_{L_{d_A},o}\}}{\arg\min} \frac{1}{2}\left\|\sum_{o=1}^{O}\mathbf{d}_o * \mathbf{C}_{L_{d_A},o} - \mathbf{L}_{d_A}\right\|_2^2 + \lambda\sum_{0=1}^{O}\left\|\mathbf{C}_{L_{d_A},o}\right\|_1 \tag{4.10}$$

where, $\mathbf{C}_{L_{d_A},o}$ is sparse coefficients maps, $\mathbf{C}_{L_{d_A},1:O(m,n)}$ is a O dimensional vector having value $\mathbf{C}_{L_{d_A},o}$ at position (m,n). l_1-norm of $\mathbf{C}_{L_{d_A},1:O(m,n)}$ is applied as activity level measure as follows:

$$\overline{\mathbf{G}}_{L_{d_A}}(m,n) = \frac{1}{9}\sum_{w_1,w_2=-1}^{1}\left\|\mathbf{C}_{L_{d_A},1:O}(m+w_1, n+w_2)\right\|_1 \tag{4.11}$$

where, $\overline{\mathbf{G}}_{L_{d_A}}(m,n)$ is activity level map of detail layer $\{\mathbf{L}_{d_A}(m,n)\}$, $w_1 \times w_2$ is window size. The choose max fusion criteria is used to obtain fused coefficient maps, as follows:

$$\mathbf{C}_{L_{d_F},o}(m,n) = \begin{cases} \mathbf{C}_{L_{d_A}}(m,n), & \overline{\mathbf{G}}_{L_{d_A}}(m,n) \geq \overline{\mathbf{G}}_{L_{d_B}}(m,n) \\ \mathbf{C}_{L_{d_B}}(m,n), & \text{otherwise.} \end{cases} \tag{4.12}$$

For each detail layers $\{\mathbf{L}_{d_A}(m,n), \mathbf{L}_{d_B}(m,n)\}$, the sparse coefficients maps $\left\{\mathbf{C}_{L_{d_A},o}, \mathbf{C}_{L_{d_B},o}\right\}, o \in \{1,...,O\}$ are obtained using CSR modelling. Then, the fused detail layer is obtained as follows:

$$\mathbf{L}_{d_F}(m,n) = \sum_{o=1}^{O}\mathbf{d}_o * \mathbf{C}_{L_{d_F},o} \tag{4.13}$$

where, $\mathbf{d}_o$ is dictionary element, $\mathbf{C}_{L_{d_F},o}$ is fused coefficients maps obtained after choose max technique, and " $*$ " is the convolution operator.

Base layer fusion

The base layers are fused based on the modality of images. In case of multi-focus images, the fusion rule will be 'choose max', which is given as follows:

$$\mathbf{L}_{b_F}(m,n) = \begin{cases} \mathbf{L}_{b_A}(m,n), & \overline{\mathbf{G}}_{L_{d_A}}(m,n) \geq \overline{\mathbf{G}}_{L_{d_B}}(m,n) \\ \mathbf{L}_{b_B}(m,n), & \text{otherwise.} \end{cases} \tag{4.14}$$

where, $\{\overline{\mathbf{G}}_{L_{d_A}}, \overline{\mathbf{G}}_{L_{d_B}}\}$ are activity level maps of detail layers $\{\mathbf{L}_{d_A}(m,n), \mathbf{L}_{d_B}(m,n)\}$. For inter-modality images of medical and visible-infrared, the fusion rule will be 'averaging', which is given as follows:

$$\mathbf{L}_{b_F}(m,n) = \frac{\mathbf{L}_{b_A}(m,n) + \mathbf{L}_{b_B}(m,n)}{2} \tag{4.15}$$

The overall low-frequency fused band image is obtained using the superposition given as follows:

$$\mathbf{L}_F(m,n) = \mathbf{L}_{d_F}(m,n) + \mathbf{L}_{b_F}(m,n) \tag{4.16}$$

Fusion rule for high-frequency band: For fusing high-frequency bands, IESDR descriptor based fusion rule is proposed. In each high-frequency bands $\{\mathbf{H}^A_{k,l}(m,n), \mathbf{H}^B_{k,l}(m,n)\}$, Eq. 4.5 is applied, and high-frequency content feature images $\{\mathbf{IESDR}^A_{k,l}(m,n), \mathbf{IESDR}^B_{k,l}(m,n)\}$ are obtained. Now using Eq. 4.17, fused coefficients $\mathbf{H}^F_{k,l}(m,n)$ are selected as:

$$\mathbf{H}^F_{k,l}(m,n) = \begin{cases} \mathbf{H}^A_{k,l}(m,n), & \mathbf{IESDR}^A_{k,l}(m,n) \geq \mathbf{IESDR}^B_{k,l}(m,n) \\ \mathbf{H}^B_{k,l}(m,n), & \text{otherwise.} \end{cases} \tag{4.17}$$

First, the source images are registered. After that, the remaining steps are summarized in the following algorithm.

4.4 Experimental Results

The experimental section is divided into five parts under the subsections. In Section 4.4.1 and Section 4.4.2, image fusion evaluation metrics and the effect of free parameters on

Algorithm 1 Proposed algorithm

Input: The pre-registered source images $\{\mathbf{A}, \mathbf{B}\}$.

Output: Fused image $\mathbf{F}(m, n)$.

1: Design the ANSST using adjustable Meyer window as given in Eq. 4.4 with $\xi = 0.2$. The pre-registered source images $\{\mathbf{A}, \mathbf{B}\}$ are decomposed using ANSST.

2: The ANSST decomposed low-frequency sub-images $\{\mathbf{L}_A(m, n), \mathbf{L}_B(m, n)\}$, and they are fused using CSR modelling, where base layers $\{\mathbf{L}_{b_A}(m, n), \mathbf{L}_{b_B}(m, n)\}$ are obtained by using Eq. 4.2, and they are fused using choose max scheme for multi-focus images given in Eq. 4.14 and averaging for inter-modality images given in Eq. 4.15. The generated fused base layer is represented by $\{\mathbf{L}_{b_F}(m, n)\}$.

3: The detailed layers $\{\mathbf{L}_{d_A}(m, n), \mathbf{L}_{d_B}(m, n)\}$ are obtained by using Eq. 4.3, and they are fused using activity level measurement given in Eq. 4.13, which will give fused detailed layer $\{\mathbf{L}_{d_F}(m, n)\}$.

4: The low-frequency fused image $\{\mathbf{L}_F(m, n)\}$ is obtained by superposition of fused base $\{\mathbf{L}_{b_F}(m, n)\}$ and fused detailed layer $\{\mathbf{L}_{d_F}(m, n)\}$ as given in Eq. 4.16.

5: At last, ANSST decomposed high-frequency sub-images $\{\mathbf{H}^A_{k,l}(m, n), \mathbf{H}^B_{k,l}(m, n)\}$ are fused using our proposed IESDR descriptor as given in Eqs. 4.5 and 4.17. Then, fused image $\mathbf{F}(m, n)$ is obtained by inverse ANSST.

the proposed image fusion framework are discussed, respectively. In Section 4.4.3, the performance of different image fusion methods are compared. Additionally, the computational efficiencies of these methods are compared and it is tabulated in Section 4.4.4. This chapter mainly considers the issues related to development of an efficient image fusion method. However, the proposed ANSST can also be employed for image denoising. So, one denoising framework along with some experimental results are also included only as a case study in Section 4.4.5.

4.4.1 Performance evaluation metrics

The proposed technique is tested with the help of four metrics, where three metrics Q_{MI} [2], $Q^{AB/F}$ [122] and Q_S [125] are widely employed, and the fourth one is our proposed singularity and structural information preservation metric Q_{SS}. For optimal

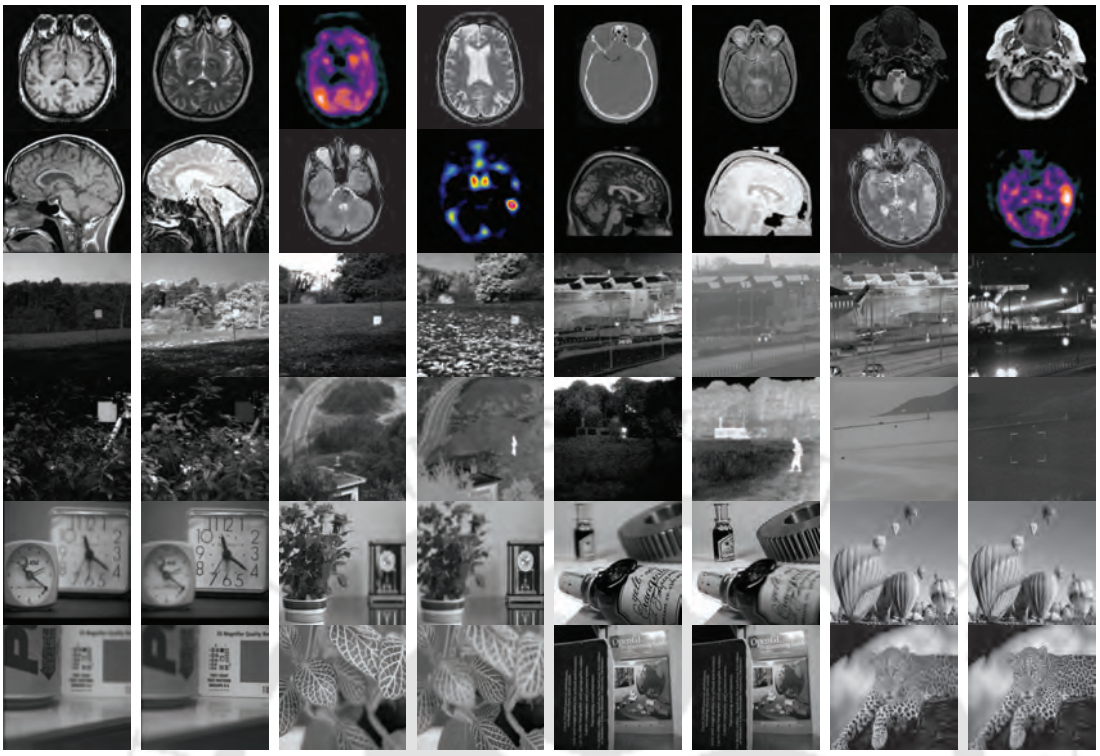


FIGURE 4.5: Image Dataset-1 of eight pairs of medical, eight pairs of visible-infrared and eight pairs of multi-focus images.

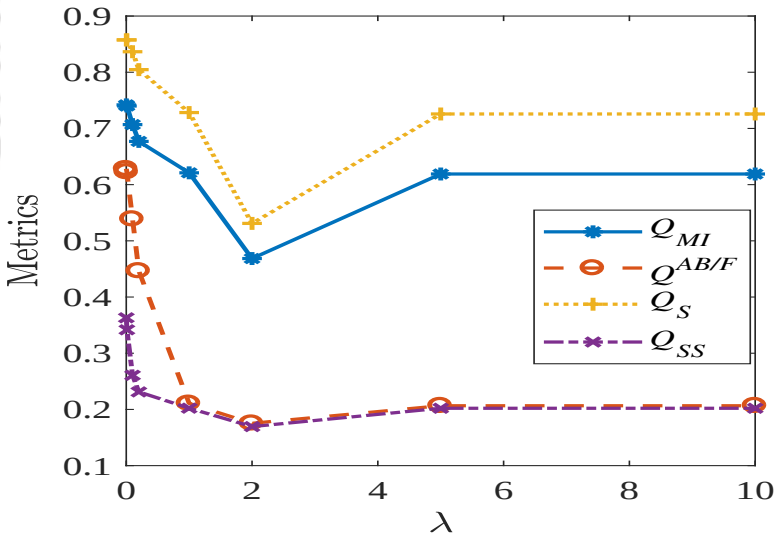


FIGURE 4.6: Metrics variations with CSR modelling parameter λ for $\alpha = 15$ and $\xi = 0.2$.

fusion performance, the metric Q_{MI} should be high, and values of Q_S, $Q^{AB/F}$ and Q_{SS} should be close to 1. Details about these evaluation metrics can be found in Sections 1.2.5 and 2.4.1.

TABLE 4.2: The objective image fusion performance metrics for subjective similarity using dataset-1.

Metric	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{SS}
relevance (r)	0.746	0.812	0.862	**0.884**
Correct Ranking (CR)	65.8%	74.9%	73.4%	**79.4%**

TABLE 4.3: Fusion performance for variations of sigmoid high-frequency fusion parameter (α) by keeping $\xi = 0.2$ and $\lambda = 0.01$.

α	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{SS}
1	0.741891	0.628578	0.68491	**0.363261**
2	0.741957	0.628455	0.68472	0.363198
4	0.742564	0.627214	0.684167	0.36257
7	0.741854	0.627561	0.684438	0.36157
10	0.742238	0.627861	0.684464	0.362568
15	0.742462	**0.629572**	**0.685385**	0.362249
25	**0.742821**	0.627629	0.684216	0.361473
35	0.741784	0.627238	0.68423	0.362562

TABLE 4.4: Fusion perofrmance for variations of Meyer window adjustable parameter (ξ) by keeping $\lambda = 0.01$ and $\alpha = 15$.

ξ	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{SS}
0.1	0.741846	0.62873	0.857024	0.362204
0.2	**0.746903**	0.62850	**0.857230**	0.363084
1	0.746284	0.62862	0.857184	0.3696
2	0.741604	0.62837	0.856607	0.369806
5	0.74669	0.62857	0.856811	0.36902
10	0.746173	**0.62921**	0.857005	0.369609
20	0.746203	0.628353	0.857030	0.369745
30	0.746684	0.628237	0.857173	**0.370648**

Proposed singularity and structural similarity metric (Q_{SS}): To measure singularity and structural information preserved in the fused image, we proposed a new metric Q_{SS}. The earlier proposed metrics do not simultaneously convey these two information. For deriving this image singularity and structural similarity measure, native division and native difference features are simultaneously used. Let us consider two input images $\{\mathbf{A}, \mathbf{B}\}$ and fused image $\mathbf{F}$. Native difference filtering and native division filtering are applied in our method to get $\mathbf{S}(m, n)$ and $\mathbf{R}(m, n)$ for each of the pixels of the image by using Eqs. 4.6 and 4.7, respectively.

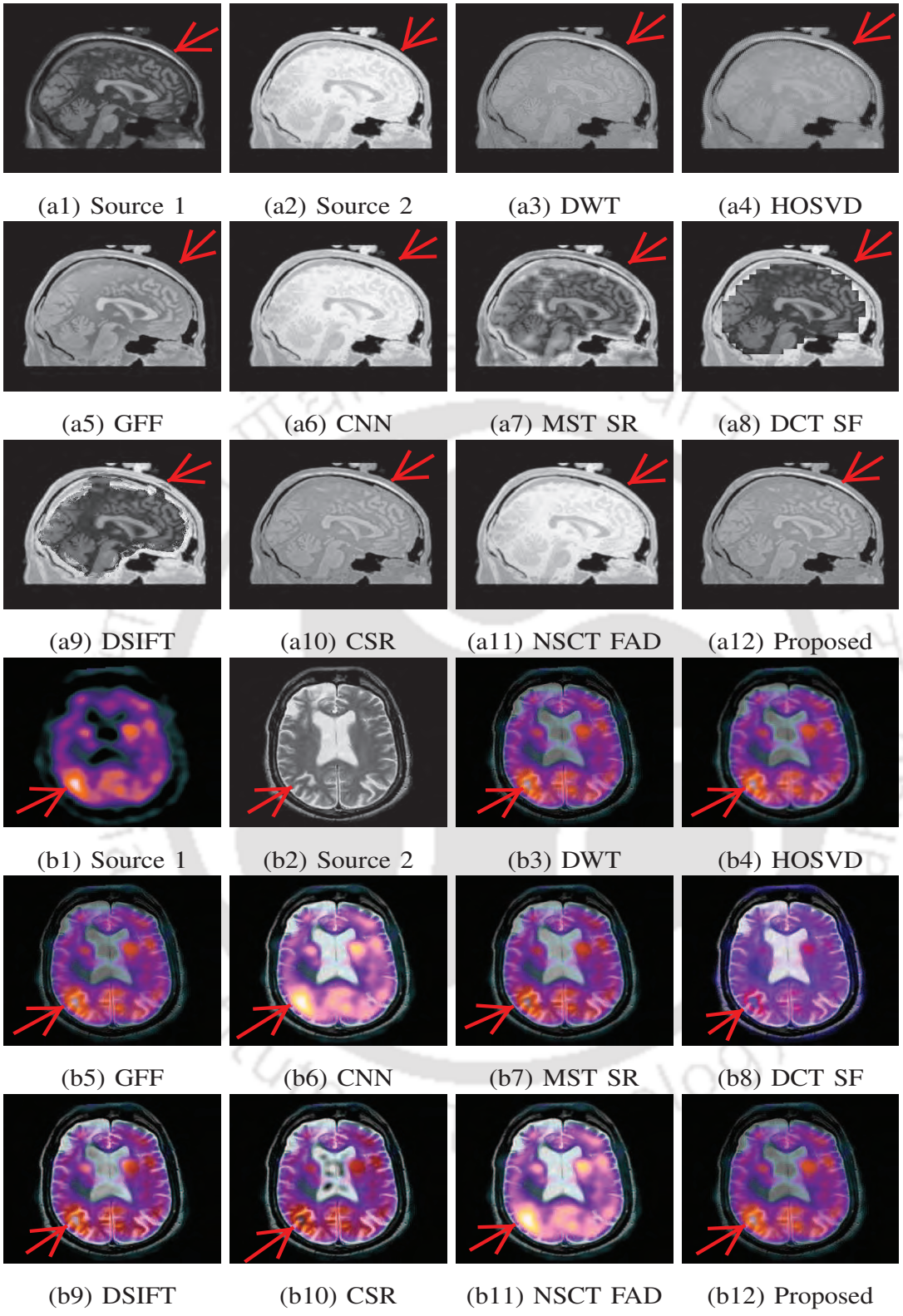


FIGURE 4.7: Medical fused images obtained by using different methods.

The images have size $M \times N$, where $1 \leq m \leq M$ and $1 \leq n \leq N$. Considering source image $\mathbf{A}$, we apply Eqs. 4.6 and 4.7 to obtain $\mathbf{S}_A(m,n)$ and $\mathbf{R}_A(m,n)$. Preservation

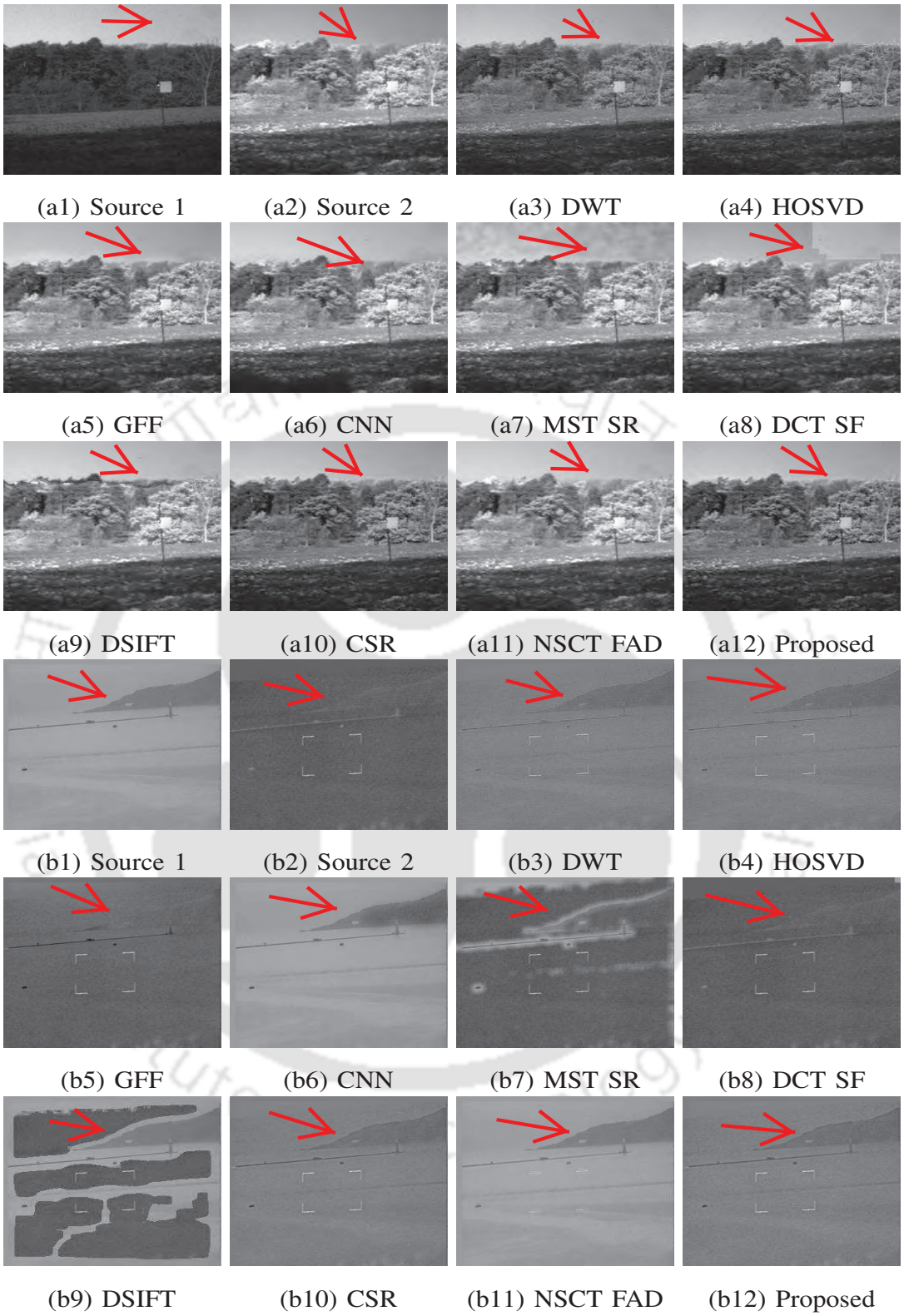


FIGURE 4.8: Visible-infrared fused images obtained by using different methods.

values $\mathbf{S}^{AF}(m,n)$ and $\mathbf{R}^{AF}(m,n)$ of input image $\mathbf{A}$ with respect to fused image $\mathbf{F}$ are

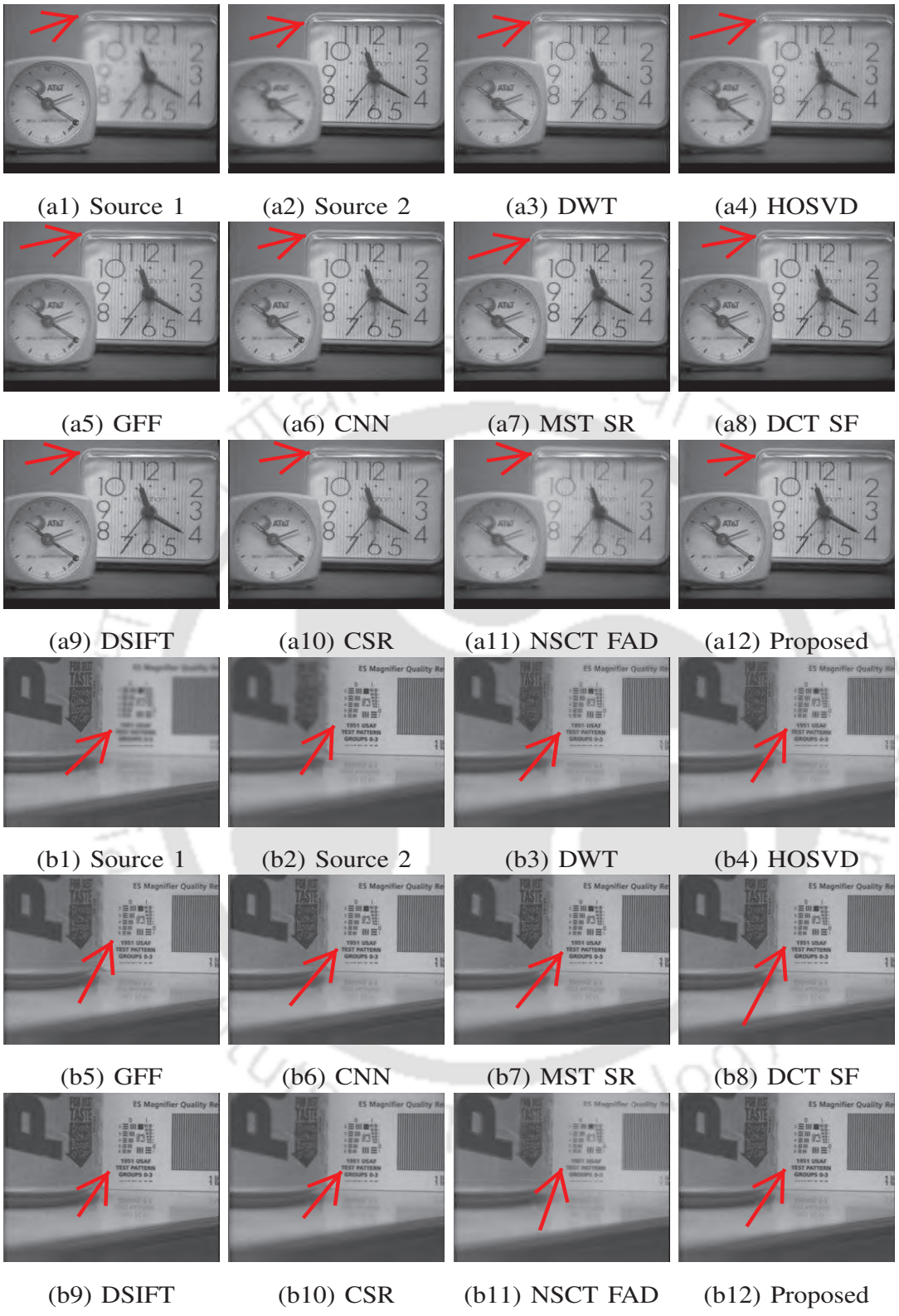


FIGURE 4.9: Multi-focus fused images obtained by using different methods.

calculated as:

$$\mathbf{S}^{AF}(m,n) = \begin{cases} \mathbf{S}_F(m,n)/\mathbf{S}_A(m,n), & \text{if} \quad \mathbf{S}_A(m,n) > \mathbf{S}_F(m,n) \\ \mathbf{S}_A(m,n)/\mathbf{S}_F(m,n), & \text{otherwise} \end{cases} \tag{4.18}$$

TABLE 4.5: Objective comparisons of different image fusion methods

Method	Images	Q_{MI}	$Q^{AB/F}$	Q_S	Q_{SS}
DWT	Medical	0.691(0)	0.408(0)	0.760(0)	0.186(0)
	Visiblle-Infrared	0.557(0)	0.574(0)	0.852(0)	0.380(0)
	Multi-focus	0.897(0)	0.646(0)	0.872(2)	0.407(0)
HOSVD	Medical	0.683(0)	0.380(0)	0.746(0)	0.197(0)
	Visiblle-Infrared	0.546(0)	0.590(0)	0.850(0)	0.390(0)
	Multi-focus	0.875(0)	0.599(0)	0.846(0)	0.364(0)
GFF	Medical	0.658(0)	0.617(0)	0.767(0)	0.246(0)
	Visiblle-Infrared	0.579(0)	0.692(2)	0.942(0)	0.471(0)
	Multi-focus	0.885(1)	0.734(0)	0.859(0)	0.501(0)
CNN	Medical	0.734(1)	0.666(0)	0.944(3)	0.332(2)
	Visiblle-Infrared	0.495(0)	0.648(0)	0.933(1)	0.449(1)
	Multi-focus	0.897(0)	0.742(1)	0.863(0)	0.553(0)
MST SR	Medical	0.761(0)	0.588(0)	0.779(0)	0.240(0)
	Visiblle-Infrared	0.541(0)	0.673(0)	0.892(0)	0.466(0)
	Multi-focus	0.901(0)	0.723(0)	0.863(0)	0.506(0)
DCT SF	Medical	0.778(3)	0.664(2)	0.875(1)	**0.339(4)**
	Visiblle-Infrared	**0.558(3)**	0.688(1)	0.970(3)	0.448(3)
	Multi-focus	**0.925(4)**	0.742(1)	0.866(0)	**0.574(5)**
DSIFT	Medical	0.723(0)	**0.679(3)**	0.854(1)	0.324(0)
	Visiblle-Infrared	0.497(0)	0.688(0)	0.825(0)	0.529(0)
	Multi-focus	0.846(0)	0.743(2)	0.864(0)	0.565(1)
CSR	Medical	0.677(0)	0.572(0)	0.776(0)	0.176(0)
	Visiblle-Infrared	0.510(0)	0.673(0)	0.875(0)	0.301(0)
	Multi-focus	0.935(1)	0.698(0)	0.864(0)	0.361(0)
NSCT FAD	Medical	0.647(0)	0.640(0)	0.776(0)	0.281(0)
	Visiblle-Infrared	0.588(2)	0.625(0)	0.960(0)	0.423(0)
	Multi-focus	0.624(0)	0.561(0)	0.860(0)	0.391(0)
Proposed	Medical	**0.787(4)**	0.676(3)	**0.945(3)**	0.319(2)
	Visiblle-Infrared	0.547(3)	**0.699(5)**	**0.977(4)**	**0.478(4)**
	Multi-focus	0.915(2)	**0.748(4)**	**0.886(6)**	0.552(2)

$$\mathbf{R}^{AF}(m,n) = \begin{cases} \mathbf{R}_F(m,n)/\mathbf{R}_A(m,n), & \text{if} \quad \mathbf{R}_A(m,n) > \mathbf{R}_F(m,n) \\ \mathbf{R}_A(m,n)/\mathbf{R}_F(m,n), & \text{otherwise} \end{cases} \tag{4.19}$$

$\mathbf{S}^{AF}(m,n)$ and $\mathbf{R}^{AF}(m,n)$ measure perceptual loss of information. The singularity and structural information of the source images which are transferred to the fused image are

obtained as:

$$\mathbf{Q}_S^{AF}(m,n) = \left(1 + e^{-(\mathbf{S}^{AF}(m,n))}\right)^{-1} \tag{4.20}$$

$$\mathbf{Q}_R^{AF}(m,n) = \left(1 + e^{-(\mathbf{R}^{AF}(m,n))}\right)^{-1} \tag{4.21}$$

Sigmoid function is used to represent singularity and structural information preserve values. Singularity and structural information preservation values are obtained as:

$$\mathbf{Q}^{AF}(m,n) = \mathbf{Q}_S^{AF}(m,n)\mathbf{Q}_R^{AF}(m,n) \tag{4.22}$$

where, $0 \leq \mathbf{Q}^{AF}(m,n) \leq 1$. If $\mathbf{Q}^{AF}(m,n) = 0$, then it corresponds to complete loss of singularities and structural details transferred from $\mathbf{A}$ to $\mathbf{F}$ at position (m,n). On the other hand, $\mathbf{Q}^{AF}(m,n) = 1$ indicates fusion from $\mathbf{A}$ to $\mathbf{F}$ with no loss of singularities and structural information. The singularity and structural similarity metric is given as:

$$Q_{SS} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}\left[\mathbf{Q}^{AF}(m,n)\mathbf{W}^{A}(m,n) + \mathbf{Q}^{BF}(m,n)\mathbf{W}^{B}(m,n)\right]}{\sum_{i=1}^{N}\sum_{j=1}^{M}\left[\mathbf{W}^{A}(m,n) + \mathbf{W}^{B}(m,n)\right]} \tag{4.23}$$

where, $\mathbf{W}^A(m,n)$ and $\mathbf{W}^B(m,n)$ are the weights assigned to $\mathbf{Q}^{AF}(m,n)$ and $\mathbf{Q}^{BF}(m,n)$, respectively. While $\mathbf{W}^A(m,n) = [\mathbf{S}_A(m,n)]$ and $\mathbf{W}^B(m,n) = [\mathbf{S}_B(m,n)]$, where $0 \leq Q_{SS} \leq 1$.

Validation of proposed Q_{SS} ***metric***: The proposed singularity and structural similarity measure Q_{SS} is validated by utilizing the subjective relevance (r) and Correct Ranking (CR) metrics [124]. The validation is done for the Dataset-1. For validation, the metrics r should be close to 1 and CR should be almost 100%. The r and CR values in Table 4.2 indicate that the metric Q_{SS} outperforms other three metrics ($Q^{AB/F}$, Q_{MI}, and Q_S). Hence, the metric Q_{SS} can also be effectively applied to compare different image fusion techniques. This metric (Q_{SS}) can be used for jointly measuring the fusion performance of source images having singularity and structural information in contrast to our earlier proposed metrics Q_{TS} and Q_E (mentioned in Sections 2.4.1 and 3.4.1, respectively).

TABLE 4.6: Average run time and deviation (in seconds) of different fusion techniques using dataset-1.

Methods	DWT	HOSVD	GFF	CNN
Time	0.701±0.04	0.88±0.24	0.52±0.22	37.29±3.82
Methods	MST SR	DCT SF	DSIFT	CSR
Time	13.86±2.91	1.07±0.08	4.5±0.31	39.35±4.85
Methods	NSCT FAD	Proposed		
Time	1.43±0.47	**43.66±4.58**		

TABLE 4.7: Average denoising performance comparison with MSPB using dataset-1 and dataset-2.

Methods	NSST ξ=1	ANSST ξ=0.5	ANSST ξ=2	MSPB
MSE	70.35	**69.12**	69.80	70.07
MAE$\times 10^4$	99.10	97.11	**95.91**	97.85

4.4.2 Discussion on free parameters

To perform our experiments, the images are obtained from [89, 140, 168]. The Dataset-1 is used for analyzing the effects of free parameters and average values of the metrics, i.e., Q_{MI}, Q_S, $Q^{AB/F}$ and Q_{SS} are shown in Fig. 4.6 and Tables 4.3-4.4. The default setting of the parameters are set at $\xi = 0.2$, $\alpha = 15$ and $\lambda = 0.01$. In Fig. 4.6, the CSR modelling parameter λ is varied and other parameters are set to their default values. Similarly, fusion performances with respect to IESDR descriptor parameter α and Meyer window adjustable parameter ξ are presented in Tables 4.3-4.4. The best suitable values are $\xi = 0.2$, $\alpha = 15$, and $\lambda = 0.01$.

4.4.3 Experimentation and analysis

Dataset-1 (Fig. 4.5) is used to perform comparative analysis of various fusion algorithms. Dataset-1 and Dataset-2 (only ground truth images shown in Fig. 4.11) is used for showing comparative results for image denoising. The *maxflat* non-subsampled Laplacian pyramid filters is chosen for the ANSST. For the proposed technique, the size of each dictionary filters is set to 8×8. The dictionary filters are learned with 50 good quality natural images using learning method given in [88]. In our work, the number of dictionary filters is set to 128 [117]. In Table 4.5, the numbers within the parentheses indicate

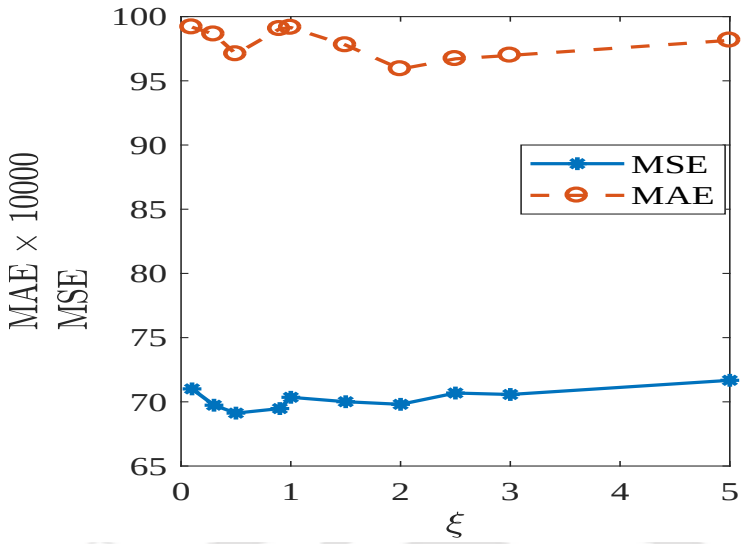


FIGURE 4.10: Average results of denoising performance with ξ.

FIGURE 4.11: Denoising performance of the proposed method. Dataset-2 contains only ground truth images of Fig. 4.11.

number of image pairs that our proposed method clearly outperforms other state-of-the-art methods.

For color image fusion (Fig. 4.5), YCbCr color model is applied. The state-of-the-art

methods such as discrete wavelet transform (DWT) [24], multi-scale transform and sparse representation (MST SR) [89], guided filtering (GFF) [114], convolutional neural network (CNN) [118], higher order singular value decomposition (HOSVD) [160], discrete cosine transform with spatial frequency (DCT SF) [134], dense scale invariant feature transform (DSIFT) [169], NSCT focussed area detection (NSCT FAD) [170], and CSR [88] are used for comparative analysis.

It is observed from Figs. 4.7-4.9 that the proposed method can preserve prominent information of both the source images. The objective comparisons shown in Table 4.5 also ascertain this observation. In Table 4.5, the average of metrics values are presented which are calculated using Dataset-1 (Fig. 4.5).

Discussion on image fusion results: The DWT (Figs. 4.7-4.9 (a3, b3)) method [24] produce blurring, ringing artifacts, and low contrast in the fused images. DWT-based fusion method is less computationally complex. The main limitation of this method is the presence of ringing distortions in the output images because of shift variant property of DWT, and inefficiency of DWT to approximate edge information. HOSVD (Figs. 4.7-4.9 (a4, b4)) method [160] produces some unwanted edges, low contrast, and blurring in the fused image. GFF method [114] produces blurring artifacts in the fused image as observed in Fig. 4.9 (a5, b5). GFF is a regression based fusion technique which is less computationally complex, but it can not preserve crucial information of the source images because of the number of iterations and inefficient weight creation technique.

The CNN (Figs. 4.7-4.8 (a6, b6)) method [118] produce spatially distorted and high contrast fused images in case of medical and visible-infrared images. The CNN method produces higher metrics values, but in visual comparison, performance is unsatisfactory due to difficulty in attaining optimal parameter settings for fusion of images of different modalities. The MST SR (Figs. 4.7-4.8 (a7, b7)) method [89] produce high contrast and spatial distortion in the fused images in case of medical and visible-infrared images. MST SR method also can not detect crucial and fine details of the input images due to inefficient dictionary learning algorithm. Moreover, fused image generated by this method contain noisy pixels due to absolute value maximum selection fusion criteria.

The DCT SF (Figs. 4.7-4.8 (a8, b8)) and DSIFT (Fig. 4.7 (a9, b9)) methods [134, 169] produce competitive results, but these methods work well in case of multi-focus image fusion. However, the fused images obtained by these methods in general suffer from spatial distortions such as too much smoothing and ringing artifacts. The CSR method [88] also produces good quality fused images, but some loss of fine information and ringing artifacts appear in the fused images (Figs. 4.7-4.9 (a10, b10)). The NSCT FAD (Figs. 4.7-4.8 (a11, b11)) method [170] also produce too much contrast in the fused images. The NSCT FAD method can not perfectly preserve fine information of the source images due to inefficient fusion rules, which results high contrast and blurring in the fused image.

Our proposed method clearly outperforms state-of-the-art techniques in visual (subjective) and objective comparisons due to four reasons: First, dictionary design is an important step of CSR modelling. However, due to the limitation of dictionary design algorithms, these dictionaries can not perfectly represent high-frequency information present in images. This limitation is removed due to the application of ANSST, which has better ability to capture high-frequency contents. Second, when the source images are misregistered, ANSST and other transform-based methods produce undesirable artifacts. The misregistration error, which is not completely removed by ANSST is further removed by CSR in our method. CSR is shift invariant representation of an image, which reduces misregistration errors. Third, the number of ANSST decomposition levels can even be fixed at two or three levels as CSR can preserve low-frequency contents more efficiently than the high-frequency contents. The low-frequency information representation ability of CSR is better than ANSST and NSST. Moreover, the overall system complexity of the proposed system would be less as only few decomposition levels of ANSST are employed. Finally, different fusion rules are deployed for different types of source images (medical, visible-infrared, and multi-focus images), and it is observed that the quality of the fused images obtained by our proposed method is superior in contrast to the state-of-the-art methods. In the proposed method, fusion rules for different types of images are formulated based on different attributes of source images, such as intensity profile, contrast, texture and variance information.

4.4.4 Computational efficiency analysis

The comparative study of average computational efficiencies or average running time of proposed and various state-of-the-art fusion techniques are presented with deviation time in Table 4.6 for Dataset-1 (Fig. 4.5). All the methods are implemented in MATLAB R2017a platform on a computer having processing speed of 3.20 GHz, and RAM of 8 GB. As presented in Table 4.6, proposed, CNN and MST SR techniques are computationally complex due to sparse coding, and the time needed to obtain best parameter settings. Moreover, the proposed method takes all the local and global information into consideration to perform fusion, and that is why the computational time is more. The time complexity of DSIFT, DCT SF, and NSCT FAD are moderate. While DWT, HOSVD and GFF methods are relatively fast as compared to other techniques. The GFF method has the lowest running time because it has computational complexity of linear order. However, our method can give best fusion results than the other well established methods.

The computational complexity of DCT SF method is lower than the proposed method as our method employs a relatively simple fusion framework. DCT SF method employs spatial frequency non-overlapped block based fusion rule by employing DCT coefficients, whereas our proposed method adopts NSST and CSR based framework to produce fused images. The complexity issue of the proposed method is sacrificed a little bit with an objective to improve the quality of the fused images.

4.4.5 Application of proposed ANSST in image denoising

To show wide applicability of our proposed ANSST, image denoising experiments are also implemented. The parameter setting for this is kept as same as given in [31]. The time domain soft threshold are applied to denoise the noisy images. Gaussian white noises with zero mean are added to the ground truth images. Figs. 4.10-4.11 show the performance of the proposed ANSST in contrast to other methods. In Fig. 4.10, average values of Mean Square Error (MSE) and Mean Absolute Error (MAE) are plotted, which are calculated for Dataset-1 (Fig. 4.5) and Dataset-2 (ground truth images of Fig. 4.11). MSE and MAE are employed for analysis of denoising performance. Both MSE and

MAE are computed between the ground truth images and the enhance denoised images. In Fig. 4.10, MAE values are multiplied by 10000 for easy comparison of MAE and MSE values. The qualitative results are also shown in Fig. 4.11 for few images.

Discussion on image denoising results: As shown in Fig. 4.10, the lowest MSE and MAE are obtained for $\xi = 0.5$ and 2, respectively. For $\xi = 0.5$ and 2, our method gives better results as compared to conventional NSST with $\xi = 1$. Moreover, it is observed that increasing or decreasing the value of ξ outside the interval [0.4, 2] produces higher MAE and MSE. In this denoising experiment, coefficients of ANSST shearing filters are adjusted to achieve better performance by varying ξ heuristically, within the interval [0.4, 2]. The Meyer window used in our method loses its morphology/characteristics for the ξ value outside the above mentioned range.

The ST and NSST is widely used for image denoising tasks [31]. The denoising experiments presented in [31] shows that NSST-based denoising results are better as compared to Wavelet-based denoising and Contourlet-based denoising. From the experimental results of proposed method, it is observed that the proposed ANSST-based denoising gives better results as compared to NSST-based denoising presented in [31]. Hence, the proposed ANSST can be effectively employed for image denoising instead of wavelets, Contourlet, ST and NSST. In recent years, many image denoising methods have been published. In Table 4.7, the proposed method is compared with the recent multi-scale patch-based image restoration (MSPB) method [171]. The proposed ANNST-based denoising performs better than NSST-based and MSPB denoising methods. The spatial artifacts are less in case of ANSST as compared to NSST and MSPB. However, MSPB denoising method performs better than NSST-based denoising with $\xi = 1$. In Fig. 4.11, the visual differences are best visible at higher zooming.

4.5 Conclusion

In this chapter, we proposed an ANSST and CSR based fusion technique for multi-sensor source images. In view of this, fusion rules are develop to fuse low- and high-frequency bands. Additionally, a metric Q_{SS} is proposed to determine both singularities

and structural details present in the fused images. Experimental outcomes indicate that the fusion performance evaluated by the proposed and other objective measures matches the qualitative analyses. Experimental outcomes explicitly indicate that the proposed metric Q_{SS} can judge the performance of different fusion techniques. The window based fusion rules for both low- and high-frequency bands are employed to consider relationship between the neighbouring pixels. That is why, the window based fusion rules are slightly robust to noises of the input images. Additionally, the proposed ANSST can be employed for denoising of images. However, the proposed fusion approach is more computational complex compared to state-of-the-art methods. In future, we will work further to resolve this issue.

The proposed method significantly minimize the misregistration errors and it can give better fusion performance. However, this method is computationally complex (lowest time efficiency) as compared to our previous two fusion methods presented in **Chapter 2** and **Chapter 3**.

5

Image Mosaicking using Improved Auto-sorting Algorithm and Local Difference-based Harris Features

*In **Chapter 2**, we proposed local difference features to fuse high-frequency information. To show one application of our local difference features, an image mosaicking framework is proposed with the help of these features. Image mosaicking is an image processing technique which is useful for tiling images. One important step in an image mosaicking framework is the auto-sorting algorithm, which is to be performed to minimize registration errors in the mosaic image. Another step in mosaicking is the detection of interest points for matching of the source images obtained after auto-sorting. However, in the presence of noisy and pseudo-periodic structures in the source images, the existing auto-sorting methods generally produce distortions in the final mosaic image. Secondly, most of the popular interest point detection algorithms do not specifically consider computational issues. So, this work mainly addresses the above-mentioned problems which are generally encountered during image mosaicking. The problem of image auto-sorting can be partially solved by adopting a phase correlation strategy. In our method, the sorting procedure is further improved by deploying the structural similarity index (SSIM) measure (as mentioned in Sect 1.2.5) instead of using the phase correlation. The issue of high time complexity of conventional corner detectors is reduced by using our earlier proposed local difference operation (as mentioned in Section 2.3.1) in place of standard Sobel edge detector. Experimental results show the efficacy of the proposed method.*

5.1 Introduction

A mosaic image has a wide range of view as compared to source images captured by number of cameras. Mosaic images are created by overlapping common regions of the source images. The goal is to find the homography or 3×3 camera matrix for each of the source images to form a mosaic image. The mosaic image formation procedure consists of finding the best seams in overlapping regions of input images, and blending them with the seams to reduce blending artifacts. Mosaic images are widely used in applications such as video surveillance, mapping, resolution improvement, motion detection, medical imaging, etc.

Image mosaicking algorithms can be divided into blending- and registration-based algorithms [172]. Blending-based methods are employed either to perform transition smoothening or to obtain optimal seam [172]. Transition smoothening based mosaicking methods includes feathering-, pyramid- and gradient-based mosaicking methods.

Registration based mosaicking can be performed in both spatial and frequency domains. The spatial domain methods include area- and feature-based mosaicking techniques. The area-based mosaicking methods consist of normalized cross-correlation and mutual information-based techniques [173]. The feature-based mosaicking methods are low-level feature- and contour-based methods [174, 175]. The low-level features based methods generally employ corner or feature detectors like Harris corner detector [176], features accelerated segment test (FAST) corner detector [177], scale-invariant feature transform (SIFT) [178] or speeded up robust features (SURF) [179]. These methods aim to produce affine transformation and photometric invariant features for image matching. The proposed method is a spatial domain approach, where Harris corner based low-level features [180, 181] are used for mosaicking.

SIFT-based mosaicking methods are affected by improper correspondence, intensity discrepancy and registration errors between the source images. SURF-based methods are affected due to the variations in colour and illumination, and affine transformations.

Contour-based mosaicking methods generally have very high computational complexity [172]. Frequency domain based image mosaicking methods are sensitive to noises and they require a high overlapping area between the images.

Image sorting is one of the pre-processing steps which needs to be performed to minimize registration errors in the final mosaic image. In image sorting, large overlapping between the source images provide more information and matching points for the estimation of positional details. Hence, mosaic images can be initially formed by the source images which have large overlapping areas. So, it is easy to find more reliable matching points to minimize registration errors.

A first comprehensive multi-step method [182] is proposed to solve the image sorting problem. This method finds the camera positions and image clusters (*i.e.*, images of the same scene) using the information of image patches and multi-view image matching. This method can spatially sort unordered images only if there is a large overlapping between unordered set of images. Large overlapping ensures significant correspondences. Moreover, this process is highly computationally complex in case of high-resolution images due to patch-based processing.

To circumvent this problem, the phase correlation-based method was proposed [183], which is somewhat simple and less computational complex. In this method, phase correlation measure first checks whether any pair of images contain an overlapping region. Subsequently, the degree of overlapping is estimated. It is found by measuring the value of the principle peak. The value of the principle peak is higher if the pair of images have large overlapping, and vice-versa. However, the principle peak become ambiguous when the overlapping images have noises and/or pseudo-periodic structures with unknown transformations, and hence, it is difficult to do perfect image registration for this case.

In our method, the sorting procedure is improved by deploying the structural similarity index (SSIM) measure instead of using phase correlation. The SSIM index-based sorting is also fast and simple. It is more robust to noise, and/or pseudo-periodic structures with unknown transformations due to local operations on images [121].

Harris corner detector is widely used in image mosaicking to find interest points for image matching. The second moment matrix used in this detector uses a Sobel edge detector to find the derivatives in horizontal and vertical directions. However, the Sobel edge detector have some drawbacks. This detector only considers the difference of pixels in horizontal and vertical directions without considering diagonal pixel relationship. The use of difference operation within a 3×3 size window makes it rotation invariant only if the rotation aligns with the exact horizontal and vertical directions. Additionally, Sobel edge detector can not perfectly match noisy source images.

In our method, the above mentioned issues of Sobel operator in the corner detector are almost tackled by using our earlier proposed local difference [184] operation in the Harris corner detector. The proposed method is a spatial domain approach, where Harris corner based low-level features [180, 181] are used for mosaicking. Some methods related to image mosaicking are available in [185–190]

5.2 Proposed method

As explained earlier, it is important to do sorting of the source images before matching for image mosaicking. Also, the corner/interest points detection schemes employed for mosaicking algorithms should be efficient. These two aspects are considered in our proposed mosaicking scheme, and these two cases are separately discussed as follows[6].

5.2.1 Auto-sorting of source images

Image mosaicking is an iterative process where source images are stitched one after another. The visual quality of a mosaic image depends on the order in which source images are stitched. The source images having higher similarity need to be stitched first, and subsequently, other less similar images are considered. We proposed to employ SSIM index (mentioned in Sect 1.2.5) to select the order of the source images for mosaicking. SSIM is a global measure of luminance (l), contrast (c) and structural (s) information in images [121]. The SSIM between two images $\mathbf{X}$ and $\mathbf{Y}$ is given as:

[6]This work is under review in Computers and Electrical Engineering, Elsevier (Refer item 4 for details).

$$SSIM(\mathbf{X},\mathbf{Y}) = [l(\mathbf{X},\mathbf{Y})]^{\alpha}.[c(\mathbf{X},\mathbf{Y})]^{\beta}.[s(\mathbf{X},\mathbf{Y})]^{\gamma} \tag{5.1}$$

where, $\alpha > 0, \beta > 0$ and $\gamma > 0$ are parameters used to modify the importance of the three components. In order to simplify the expression, we select $\alpha = \beta = \gamma = 1$. The simplified expression of SSIM is given as follows:

$$SSIM(\mathbf{X},\mathbf{Y}) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \tag{5.2}$$

where, μ_X, μ_Y are the local scalar means of images $\mathbf{X}$ and $\mathbf{Y}$, σ_X and σ_Y are the local variances of images $\mathbf{X}$ and $\mathbf{Y}$, σ_{XY} is local covariance between $\mathbf{X}$ and $\mathbf{Y}$. The constants $C_1 = C_2 = 0.01$.

For auto-sorting of source images, the SSIM between all the source images are first found out. Subsequently, the pair which has highest SSIM is first selected to get a mosaic image. Suppose, two images selected by this procedure are I_2 and I_5. Next, the third image is searched. The third image (say I_6) has the highest sum of SSIMs with I_2 and I_5. So, the source image I_6 is stitched with the first mosaic image. This procedure is iteratively followed to get the final mosaic image.

5.2.2 Local difference feature for reducing computational complexity

The local difference features are used to extract high-frequency features such as edges, corners, contours, etc. of an image. These features describe the relationship between the center pixel and its neighbouring pixels. This relation is obtained using difference operations defined in Eq. (2.4). These features are rotation invariant for a selected window. Also, they are more robust to noises as more neighbourhood pixels are considered for calculation in contrast to Sobel edge detector. Moreover, the feature extraction process is comparatively less computationally intensive as only addition and subtraction operations are needed for feature extraction. The more detail on local difference features can be found in Section 2.3.1.

Now, using the relation defined in Eq. (2.4), we modified the relation of Harris corner response metric [176]. Let, the change $\mathbf{E}$ obtained using a shift (x, y) be defined as

follows:

$$\mathbf{E}(u,v) = \sum_{x,y} \mathbf{W}(x,y)[\mathbf{I}(x+u, y+v) - \mathbf{I}(x,y)]^2 \tag{5.3}$$

where, $\mathbf{W}$ is a Gaussian window. Now, the second order Taylor series expansion of $\mathbf{E}$ is performed and its derivative at point (0,0) is found out as follows:

$$\begin{aligned} \mathbf{E}(u,v) \approx \mathbf{E}(0,0) + \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \mathbf{E}_u(0,0) \\ \mathbf{E}_v(0,0) \end{bmatrix} + \\ \frac{1}{2} \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \mathbf{E}_{uu}(0,0) & \mathbf{E}_{uv}(0,0) \\ \mathbf{E}_{uv}(0,0) & \mathbf{E}_{vv}(0,0) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned} \tag{5.4}$$

This indicates change in appearance for the shift (u,v). At point (0,0) few terms of the above equation become:

$$\begin{aligned} \mathbf{E}_u(0,0) &= 0 \\ \mathbf{E}_v(0,0) &= 0 \\ \mathbf{E}(0,0) &= 0 \end{aligned} \tag{5.5}$$

$$\begin{aligned} \mathbf{E}_{uu}(0,0) &= \sum_{x,y} 2\mathbf{W}(x,y)\mathbf{I}_x(x,y)\mathbf{I}_x(x,y) \\ \mathbf{E}_{vv}(0,0) &= \sum_{x,y} 2\mathbf{W}(x,y)\mathbf{I}_y(x,y)\mathbf{I}_y(x,y) \\ \mathbf{E}_{uv}(0,0) &= \sum_{x,y} 2\mathbf{W}(x,y)\mathbf{I}_x(x,y)\mathbf{I}_y(x,y) \end{aligned} \tag{5.6}$$

where, $\mathbf{I}_x$ and $\mathbf{I}_y$ are derivatives of image $\mathbf{I}$ in x and y directions, respectively. $\mathbf{E}_u$ and $\mathbf{E}_v$ are derivatives in u and v directions, respectively. $\mathbf{E}_{uu}$ and $\mathbf{E}_{vv}$ are second derivatives in u and v directions, respectively. $\mathbf{E}_{uv}$ is second derivative in u and v directions. Now $\mathbf{E}(u,v)$ approximately becomes:

$$\mathbf{E}(u,v) \approx \begin{bmatrix} u & v \end{bmatrix} \mathbf{M} \begin{bmatrix} u \\ v \end{bmatrix} \tag{5.7}$$

where, $\mathbf{M}$ is second moment matrix, and it is given as:

$$\mathbf{M} = \sum_{x,y} \mathbf{W}(x,y) \begin{bmatrix} \mathbf{I}_x^2 & \mathbf{I}_x\mathbf{I}_y \\ \mathbf{I}_x\mathbf{I}_y & \mathbf{I}_y^2 \end{bmatrix} \tag{5.8}$$

Now, the local difference operation is substituted in place of derivative operation as follows:

$$\mathbf{I}_x^2 = \mathbf{I}_x\mathbf{I}_y = \mathbf{I}_y^2 = \left[\mathbf{LD}^I(x,y)\right]^2 = A \tag{5.9}$$

The proposed $\mathbf{M}$ is given as follows:

$$\mathbf{M} = \sum_{x,y} \mathbf{W}(x,y) \begin{bmatrix} A & A \\ A & A \end{bmatrix} \tag{5.10}$$

The Harris local difference-based features is given as follows:

$$\begin{aligned} &\mathrm{Tr}(\mathbf{M}) = A + A \\ &\mathrm{Det}(\mathbf{M}) = 0 \\ &R = -k\mathrm{Tr}(\mathbf{M})^2 \\ &|R| = k\mathrm{Tr}(\mathbf{M})^2 \end{aligned} \tag{5.11}$$

where, k is sensitivity factor $0 < k < 0.25$. High values of $|R|$ indicate interest points.

After the extraction of local difference-based Harris features points from both the images, the sum of squared difference measure is used to find the distance between the corresponding feature descriptors. The aim of a descriptor is to produce unique and robust information of each interest point. Then, geometric transformation parameters for each of the images are calculated to perform image warping on a global reference frame. The global reference frame contains all the warp images[7]. Reliable matching points (inliers) are extracted using the RANSAC algorithm, which are used to form the mosaic image. The unreliable matching points (outliers) are discarded.

5.3 Experimental Results

Our proposed method is implemented to generate mosaic images from the offset removed source images. As mentioned in the existing literature [172, 183, 191–196], it is important to normalize pixel values of different source images before mosaicking. Let $\mathbf{I}_i$ represent

[7]The related work is published in Proceedings of 3rd International Conference on Computer Vision and Image Processing, LNCS, Springer 2019 (Refer item 4 for details).

i number of source images. To consider this issue, any one source image is selected, and other source images are normalized with the help of the selected one. For this normalization, the following simple offsetting principle is adopted.

$$\mathbf{I}_i = \mathbf{I}_i + (m_1 - m_i) \tag{5.12}$$

where, m_1 is the mean of selected gray scale source image $\mathbf{I}_1$ and m_i is the mean of i^{th} gray scale source image.

The value of sensitivity factor k defined in Eq. (5.11) is chosen as 0.06. The performance of our method is evaluated for dataset-1 of indoor scenes, dataset-2, dataset-3, dataset-4 and some other images. The dataset-1, dataset-2, dataset-3, dataset-4 and other images are publicly available [168, 197–201].

Fig. 5.1 shows the performance of image offset removal procedure. The left side images are less bright as compread to right side images. In Fig. 5.2, qualitative illustration of the proposed SSIM-based image sorting algorithm is presented. SSIM between image pairs 'a1-b1', 'b1-c1', and 'c1-a1' are obtained as 0.16, 0.41, and 0.02, respectively. In addition, SSIM between image pairs 'a2-b2', 'b2-c2', and 'c2-a2' are obtained as 0.21, 0.24, and 0.09, respectively. Therefore, if we want to perform mosaicking on a image set 'a1-b1-c1', then we first make a mosaic from the source images 'b1' and 'c1', and subsequently, the third image 'a1' is mosaicked with the first mosaic image. The same procedure is also applied for the image set 'a2-b2-c2'.

Fig. 5.3 shows the local difference features obtained by applying Eq. (2.4) on images 'a1' and 'a2'. The proposed local difference operation highlights the prominent high-frequency features of images 'a1' and 'a2', and these features are shown in images 'b1' and 'b2', respectively.

The datasets shown in Figs. 5.4-5.7 are used to qualitatively evaluate our proposed mosaicking method. In Fig. 5.8, the mosaicking results obtained by using the phase correlation based sorting method are shown [183]. In Fig. 5.8, seam and boundary distortions are more easily visible. Fig. 5.9 and Fig. 5.10 shows the mosaic image obtained

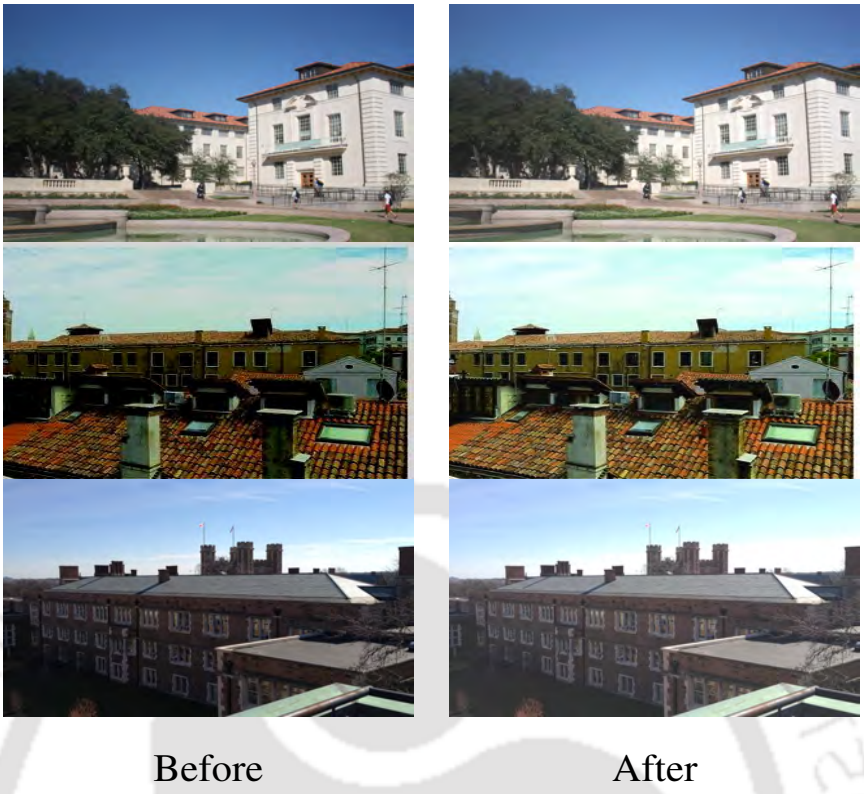


FIGURE 5.1: Qualitative performance of the offset removal procedure.

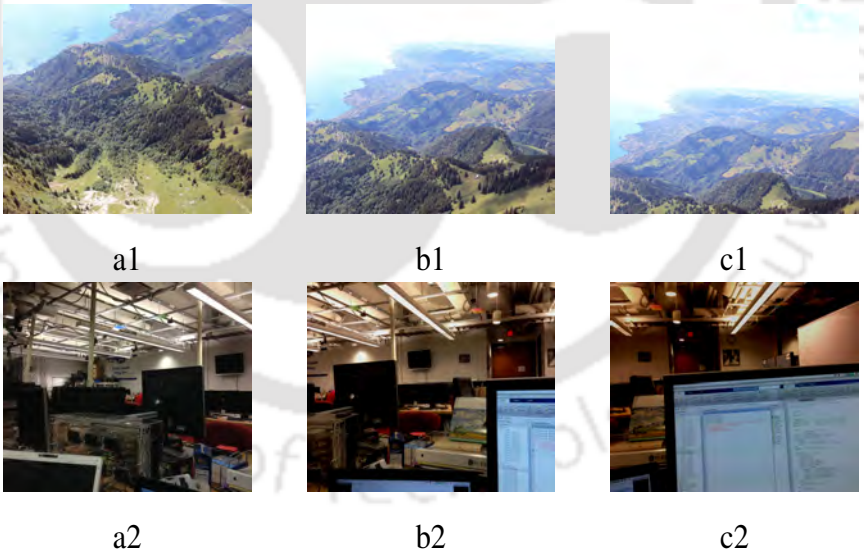


FIGURE 5.2: Images used to show the effectiveness of SSIM. SSIM between image pairs 'a1-b1', 'b1-c1', and 'c1-a1' are 0.16, 0.41, and 0.02, respectively. In addition, SSIM between image pairs 'a2-b2', 'b2-c2', and 'c2-a2' are 0.21, 0.24, and 0.09, respectively.

using the images of dataset-1, dataset-2, dataset-3 and dataset-4 with our proposed auto-sorting technique. In this evaluation, conventional Harris corner detector is employed. Figs. 5.9-5.10 (b),(d) show the mosaic image obtained by our proposed method, where

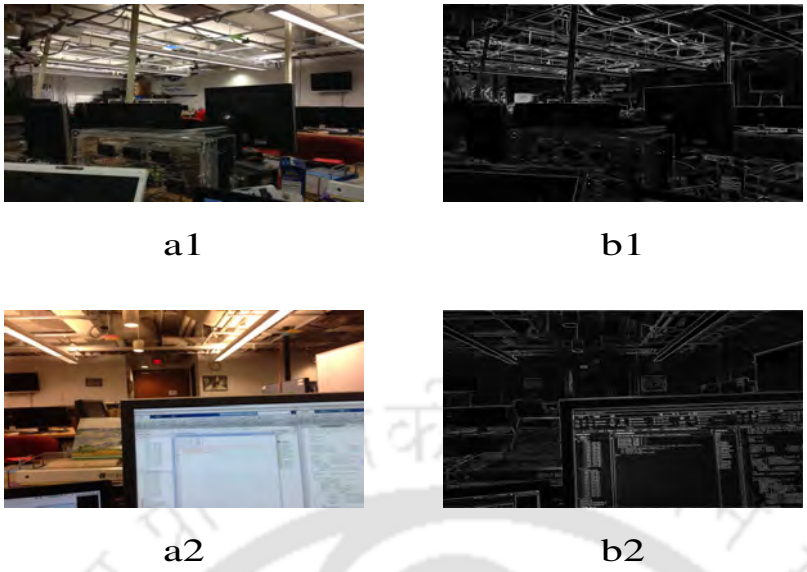


FIGURE 5.3: Source images 'a1' and 'a2' and the output of our proposed local difference operation are 'b1' and 'b2' images, respectively.

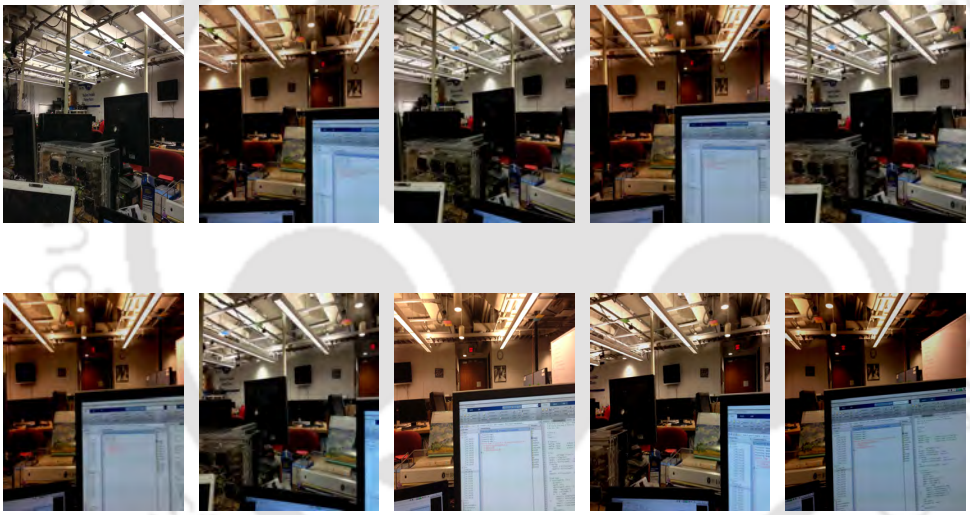


FIGURE 5.4: Image Dataset-1

FIGURE 5.5: Image Dataset-2

the proposed auto-sorting and local difference-based Harris features are employed. The improvement in mosaicking is observed in Figs. 5.9-5.10 (b),(d). In Figs. 5.9-5.10 (a),(c)

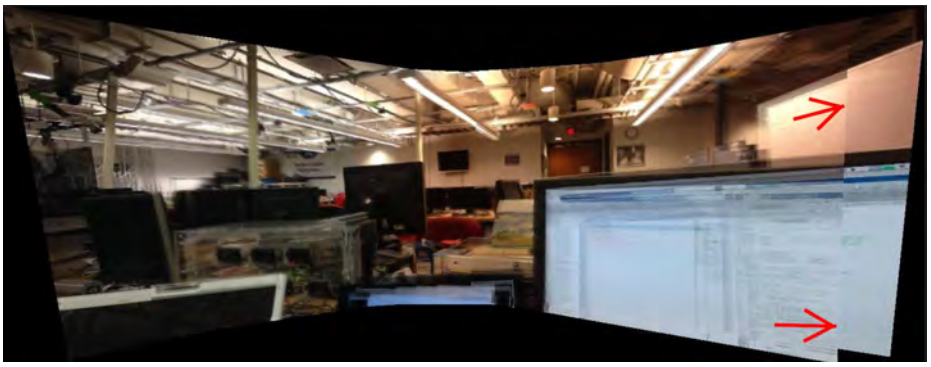
FIGURE 5.6: Image Dataset-3

FIGURE 5.7: Image Dataset-4

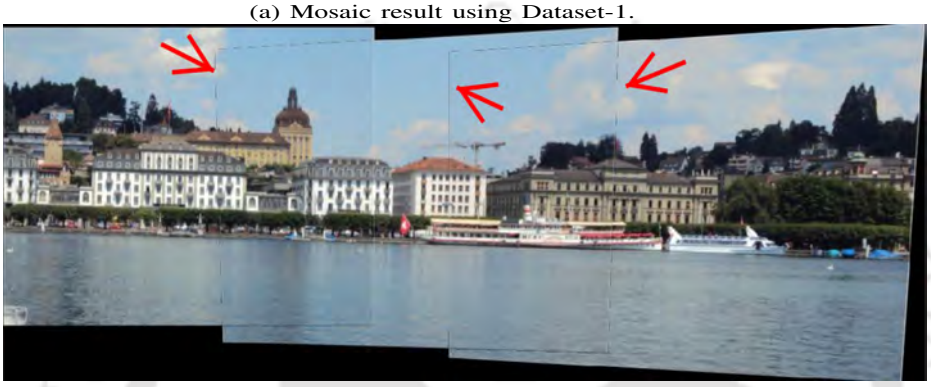
seam distortions are observed (highlighted using arrows). The seam distortions are seen in Figs. 5.9-5.10 (a),(c), while this is comparatively less prominent in the mosaic image obtained by our method (Figs. 5.9-5.10 (b),(d)).

The average time complexity (calculated for 5 different datasets [168, 197–201]) for different mosaicking algorithms are shown in Table 5.1. For all the methods presented in Table 5.1, RANSAC algorithm is used to obtain reliable matching points. These reliable points are then used to find homography or transformation matrix between the pair of images. All these methods are implemented in MATLAB R2017a platform in a PC having processor speed of 3.20 GHz, 64-bit operating system, and 8 GB RAM. For implementation of mosaicking methods, following steps are followed:

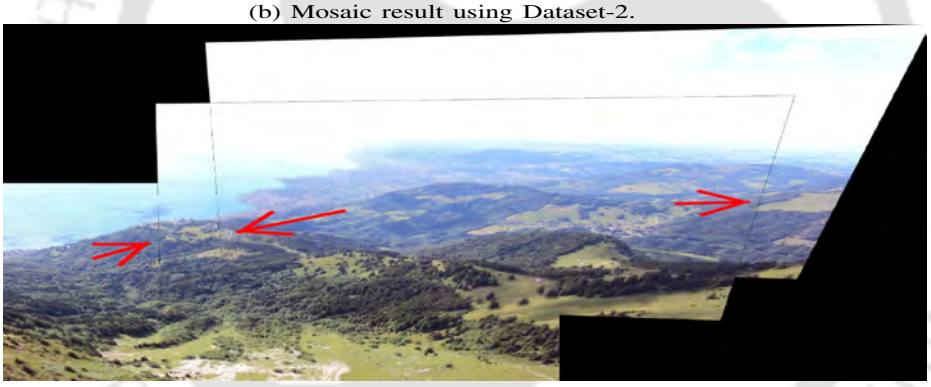
1. First, the interest points in all the source images are found out.
2. Then, a discriptor is formed in a local neighbourhood of an interest point.
3. The matching operation is performed.
4. The reliable matching points (inlier) are found out using the RANSAC algorithm.
5. The homography for each source image are obtained with the help of the selected matching points.



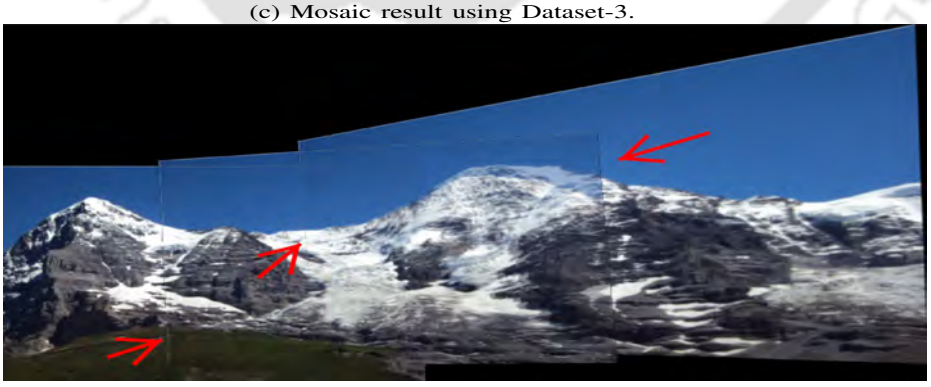
(a) Mosaic result using Dataset-1.



(b) Mosaic result using Dataset-2.



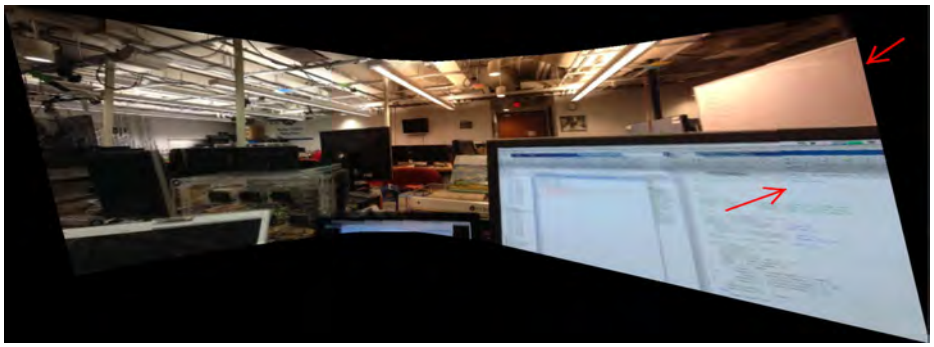
(c) Mosaic result using Dataset-3.



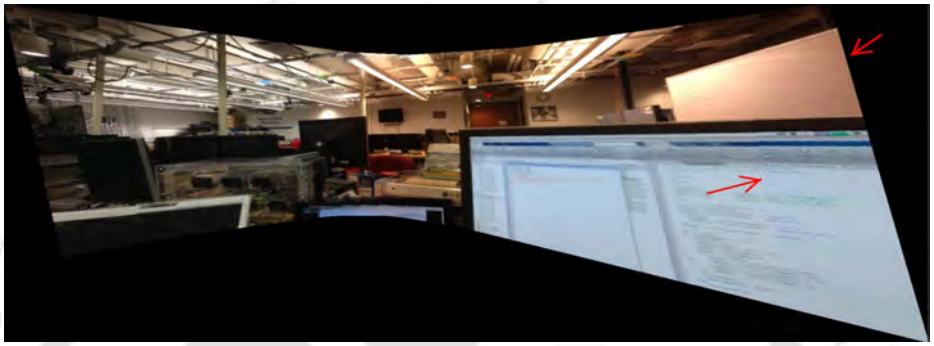
(d) Mosaic result using Dataset-4.

FIGURE 5.8: Mosaic image obtained using phase correlation sorting method.

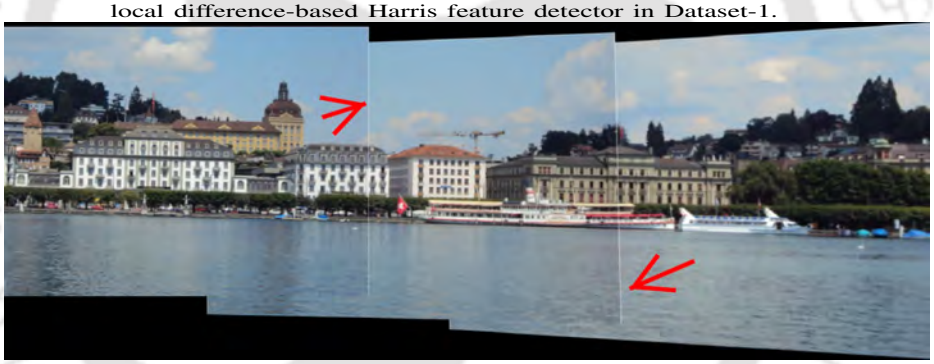
6. At last, all the images are projected on to a plane to generate the mosaic image.



(a) Mosaicking result obtained using the proposed auto-sorting and conventional Harris corner detector in Dataset-1.



(b) Mosaicking result obtained using proposed auto-sorting and proposed local difference-based Harris feature detector in Dataset-1.



(c) Mosaicking result obtained using proposed auto-sorting and conventional Harris corner detector in Dataset-2.

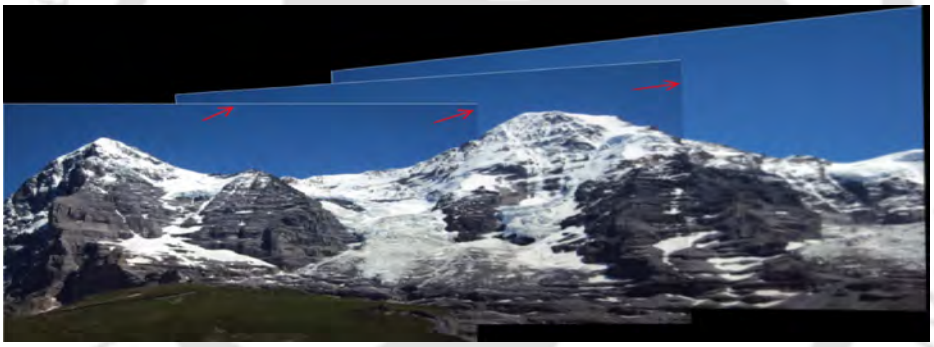
(d) Mosaicking result obtained using proposed auto-sorting and proposed local difference-based Harris feature detector in Dataset-2.

FIGURE 5.9: Mosaic image comparison using dataset-1 and dataset-2.

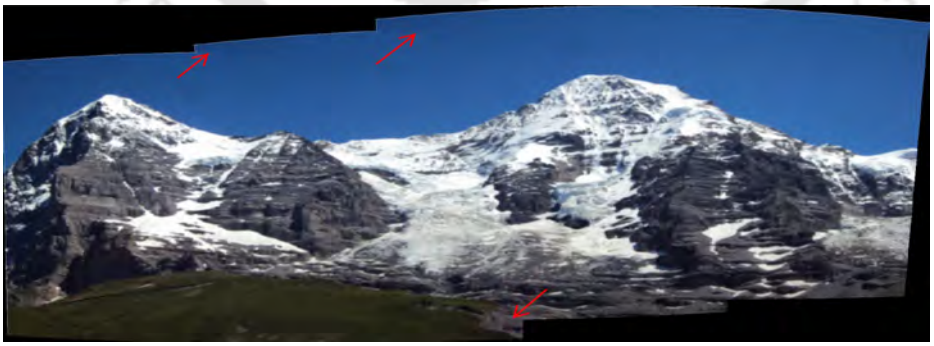
The image offset removal scheme is implemented as a pre-processing step. The mosaicking methods presented in Table 5.1 mainly differ in the steps of finding interest

(a) Mosaicking result obtained using the proposed auto-sorting and conventional Harris corner detector in Dataset-3.

(b) Mosaicking result obtained using proposed auto-sorting and proposed local difference-based Harris feature detector in Dataset-3.



(c) Mosaicking result obtained using proposed auto-sorting and conventional Harris corner detector in Dataset-4.



(d) Mosaicking result obtained using proposed auto-sorting and proposed local difference-based Harris feature detector in Dataset-4.

FIGURE 5.10: Mosaic image comparison using dataset-3 and dataset-4.

points, forming descriptor, and matching operation. The computational time for image auto-sorting and offset removal are not considered for comparison of time complexities.

In Table 5.1, the Harris method [176] uses the Harris corner detector to find the interest points and it uses 11×11 size neighbourhood to form the descriptor. The matching operation is performed using a sum of squared difference measure.

In SURF method [179], the determinant of Hessian blob detector is used to find interest points. The sum of Haar wavelet responses is used as a descriptor. Moreover, the matching operation is performed by using the sign of Laplacian (*i.e.*, the trace of Hessian matrix). In binary robust invariant scalable keypoints (BRISK) method [202], scale-space FAST-based detector [177] is used to find interest points. The assembly of a bit-string is used to make a descriptor for each of the interest points. The matching operation between two descriptors is computed by Hamming distance.

In SIFT method [178], interest points are chosen from the maxima or minima of multi-scale difference of Gaussian. The descriptor is obtained by considering pixels around a radius of the interest point location. These pixels are blurred and resampled in local image orientation planes. Moreover, the matching operation is performed by calculating Euclidean distance between two descriptors.

In Harris SIFT method, interest points are extracted using the Harris corner detector. The descriptor and matching operation is formulated as per the SIFT method. In our proposed local difference-based Harris feature detector, the interest points are obtained by employing local difference operation in place of Sobel edge detector as shown in Eq. (5.9). The 11×11 size neighbourhood around each of the interest points is used to form a descriptor. Finally, the matching operation is performed using a sum of squared difference measure.

As shown in Table 5.1, the Harris SIFT and SIFT methods are more computational complex. SURF, BRISK, Harris, and local difference-based Harris features have almost same time requirement. The proposed local difference-based Harris features have lowest execution time requirement. Also, the proposed method can generate better quality mosaicking images as compared to conventional Harris corner detector method.

TABLE 5.1: Average time complexity of different image mosaicking algorithms

Method	Seconds
Harris	35.18
SURF	46.36
BRISK	46.16
Harris SIFT	1120.64
SIFT	468.08
Proposed local difference-based Harris features	**26.35**

5.4 Conclusion

In **Chapter 2**, we proposed local difference features to fuse high-frequency information. To show the wide applicability of our local difference features, an image mosaicking framework is proposed with the help of these features. In the last few decades, the generation of mosaic images and their use in various computer vision and graphics applications have been active fields of research. The primary aim of image mosaicking is to enhance field of view and image resolution. The mosaicking image quality mainly depends on the visual quality of the source images. Also, the quality of the final mosaic image depends on the order in which the source images are overlapped to produce the mosaic image. For matching of different source images, the relevant interest points of these images need to be selected. The proposed method discusses the issues related to image auto-sorting and detection of Harris interest points in different source images. In our method, the existing phase correlation-based image auto-sorting method is improved by using the SSIM measure. Moreover, the problem of high time complexity of conventional Harris corner detectors is minimized by employing our earlier proposed local difference features. Both subjective and objective results validate our proposed techniques. We believe that our method can be efficiently implemented for real-time applications using a more efficient programming approach.

6

Conclusion and Future Work

There is a humbling amount of past work on multi-sensor image fusion. This dissertation described our proposed image fusion techniques for medical, multi-focus and visible-infrared images. In our methods, efficient fusion rules are formulated to boost information in the fused images. Our proposed local difference fusion rule is able to extract high-frequency information by considering absolute differences of the magnitude of the high-frequency coefficients in a neighborhood. An optimized and adjustable non-subsampled shearlet transform is proposed to give adaptivity in overall fusion framework. Also, three fusion performance metrics are proposed to judge performance of different fusion algorithms. At last, to show one application of our earlier proposed local difference features, these features are employed for image mosaicking. This chapter reflects on these contributions, concludes and discusses future works for image fusion. We hope that the image fusion platform developers find our contributions useful, and benefit from our proposed techniques.

6.1 Conclusion

Image fusion can be applied to improve the visual quality of an image so that more distinctive features can be obtained for computer vision tasks. Image fusion provides comprehensive information which advances the analysis and characterization of the image content in many different areas. However, it is almost impracticable to show all the visual information in a single fused image. Therefore, a more practical aim that is generally admitted is that a fused image should preserve visually prominent information comprised in each source image, while not producing noticeable artifacts. This thesis focused on multi-sensor image fusion problems in transform domain. We proposed efficient and effective fusion algorithms and provided numerical experiments on real-world data to demonstrate the performance improvement of our proposed methods.

We reviewed different types of fusion frameworks, described their advantages, and drawbacks in **Chapter 1**. To address some of the drawbacks of the existing algorithms, image fusion methods are proposed for multi-sensor images, such as medical, multi-focus and visible-infrared.

In **Chapter 2**, we proposed an NSST based fusion method for multi-sensor images. For this, novel fusion rules are proposed to select low-frequency and high-frequency coefficients for fusion. Deployment of canny edge detector with scale multiplication in the fusion framework enhances the edge information in the fused image. Also, a metric Q_{TS} is proposed to measure texture and structural information in the fused images. Experimental results show that the fusion performance evaluation implemented by the proposed and others objective measures strongly agree with the subjective analyzes. Experimental results explicitly show that our metric Q_{TS} is perceptually significant and essential to judge the performance of fusion algorithms. The proposed method is tested on different color spaces, and it is found that YCbCr is an efficient model for color image fusion. Moreover, the proposed fusion method can be employed for image enhancement. The proposed method can eliminate some of the noises present in the source images as our method uses window-based fusion rules and anisotropic diffusion filtering.

In **Chapter 3**, we proposed an optimized NSST based fusion method for multi-sensor and multi-modal images. For this, fusion rules are proposed to select high- and low-frequency coefficients by employing unsharp masking, and superposition of scale multiplied Canny edge detector features and Hessian features. Deployment of scale multiplied Canny edge detector features and Hessian features in the fusion mechanism enhances the information in the fused images. Also, a metric Q_E is proposed to measure variance information in the fused images. Experiments show that the fusion performance evaluations by the proposed and others objective measures agree with the subjective analyses. Experimental results explicitly show that our metric Q_E is perceptually significant and essential to judge the performance of fusion algorithms. The proposed method gives quite reliable and consistent results for different modalities of images.

In **Chapter 4**, we proposed ANSST- and CSR-based fusion techniques for multisensor source images. In view of this, fusion rules are develop to fuse low- and high-frequency bands. In addition, a metric Q_{SS} is proposed to determine both singularities and structural details present in the fused images. Experimental outcomes indicate that the fusion performance evaluated by the proposed and other objective measures matches the qualitative analyses. Experimental outcomes explicitly indicate that the proposed metric Q_{SS} can judge the performance of different fusion techniques. The window-based fusion rules for both low- and high-frequency bands are employed to consider the relationship between the neighboring pixels. That is why, the window-based fusion rules are slightly robust to noises of the input images. In addition, the proposed ANSST can be employed for denoising of images. However, the proposed fusion approach is more computational complex as compared to state-of-the-art methods.

In **Chapter 5**, we proposed a local difference based modified Harris detector to increase the time efficiency of image mosaicking algorithms. Moreover, SSIM based image auto-sorting algorithm is proposed to minimize registration errors in mosaic images.

6.2 Future Works

Several image representation techniques (such as sparse representation, MST, MGA, neural networks, etc.) have their distinct advantages and some drawbacks. By the combining of different image representation techniques, the advantages of different methods can be utilize to produce better quality fused images.

Recently, deep learning (such as CNN's) and advanced sparse representation techniques (such as CSR, multi-scale patch processing) has gained importance in different research areas. These techniques can also be applied for obtaining better fusion performance. There is enough research scope in the area of deep learning-based image fusion frameworks for medical, visible-infrared, remote sensing and multi-exposure images. Another important research direction may be the application of optimized NSST, ANSST, and CSR modeling for image restoration.

The proposed fusion rules which employ the IESDR descriptor is able to preserve crucial information related to image entropy, texture, contrast in the fused images. However, there is a scope to include other attributes of the source images in the descriptor to get better fusion performance.

Another recent direction could be the development of a more versatile quantitative metric which can efficiently judge different types of source images. One possible solution may be to pass information of the source images to the metric for auto adjustment of its parameters for judging the fusion performance.

Bibliography

[1] D. P. Bavirisetti and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 203–209, 2016.

[2] G. Bhatnagar, Q. J. Wu, and Z. Liu, "Directive contrast based multimodal medical image fusion in NSCT domain," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1014–1024, 2013.

[3] V. Bhateja, H. Patel, A. Krishn, A. Sahu, and A. Lay-Ekuakille, "Multimodal medical image sensor fusion framework using cascade of wavelet and contourlet transform domains," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 6783–6790, 2015.

[4] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multimodal sensor medical image fusion based on type-2 fuzzy logic in NSCT domain," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3735–3745, 2016.

[5] M. K. Bhuyan, *Computer Vision and Image Processing Fundamentals and Applications (1st Edition)*. CRC Press, 2019.

[6] N. Aishwarya and C. Bennila Thangammal, "An image fusion framework using novel dictionary based sparse representation," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 21 869–21 888, Nov 2017.

[7] F. Feng, Q. Ran, and W. Li, "Multi-level fusion of graph based discriminant analysis for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22 959–22 977, Nov 2017.

[8] P. Geng, M. Huang, S. Liu, J. Feng, and P. Bao, "Multifocus image fusion method of ripplet transform based on cycle spinning," *Multimedia Tools and Applications*, vol. 75, no. 17, pp. 10 583–10 593, Sep 2016.

[9] X. Ji and G. Zhang, "Image fusion method of SAR and infrared image based on curvelet transform with adaptive weighting," *Multimedia Tools and Applications*, vol. 76, no. 17, pp. 17 633–17 649, Sep 2017.

[10] M. Kanmani and V. Narasimhan, "An optimal weighted averaging fusion strategy for thermal and visible images using dual tree discrete wavelet transform and self tunning particle swarm optimization," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 20 989–21 010, Oct 2017.

[11] T. Wenjing, G. Fei, D. Renren, S. Yujuan, and L. Ping, "Face recognition based on the fusion of wavelet packet sub-images and Fisher linear discriminant," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22 725–22 740, Nov 2017.

[12] X. Zhang, X. Li, and Y. Feng, "Image fusion based on simultaneous empirical wavelet transform," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 8175–8193, Mar 2017.

[13] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognition*, vol. 48, no. 3, pp. 772 – 784, 2015.

[14] J. Ma, J. Jiang, C. Liu, and Y. Li, "Feature guided Gaussian mixture model with semi-supervised EM and local geometric constraint for retinal image registration," *Information Sciences*, vol. 417, pp. 128 – 142, 2017.

[15] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Transactions on Cybernetics*, vol. 44, no. 7, pp. 1001–1013, July 2014.

[16] A. Meraoumia, S. Chitroub, and A. Bouridane, "Fusion of finger-knuckle-print and palmprint for an efficient multi-biometric system of person recognition," in *Proceedings of IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.

[17] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2828–2838, Oct 2006.

[18] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2, pp. 119 – 130, 2007, Special Issue on Image Fusion: Advances in the State of the Art.

[19] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 4, pp. 640–646, July 2011.

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

[21] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295–309, Jan 2011.

[22] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

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

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

[25] S. Li and B. Yang, "Multifocus image fusion by combining curvelet and wavelet transform," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1295 – 1301, 2008.

[26] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, June 2002.

[27] T. Chen, J. Zhang, and Y. Zhang, "Remote sensing image fusion based on ridgelet transform," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, vol. 2, July 2005, pp. 1150–1153.

[28] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, Dec 2005.

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

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

[31] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.

[32] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM Journal on Mathematical Analysis*, vol. 39, no. 1, pp. 298–318, 2007.

[33] G. Guorong, X. Luping, and F. Dongzhu, "Multi-focus image fusion based on non-subsampled shearlet transform," *IET Image Processing*, vol. 7, no. 6, pp. 633–639, August 2013.

[34] X. Zhou, W. Wang, and R.-a. Liu, "Compressive sensing image fusion algorithm based on directionlets," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, p. 19, Feb 2014.

[35] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proceedings of the Fourth International Conference on Computer Vision*, May 1993, pp. 173–182.

[36] R. Shen, I. Cheng, and A. Basu, "Cross-scale coefficient selection for volumetric medical image fusion," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1069–1079, April 2013.

[37] Z. Xue, R. S. Blum, and Y. Li, "Fusion of visual and IR images for concealed weapon detection," in *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, vol. 2, July 2002, pp. 1198–1205.

[38] H. Li, B. S. Manjunath, and S. K. Mitra, "Multi-sensor image fusion using the wavelet transform," in *Proceedings of First International Conference on Image Processing*, vol. 1, Nov 1994, pp. 51–55.

[39] H. R. Shahdoosti and H. Ghassemian, "Combining the spectral PCA and spatial PCA fusion methods by an optimal filter," *Information Fusion*, vol. 27, pp. 150–160, 2016.

[40] P. Ganasala and V. Kumar, "CT and MR image fusion scheme in nonsubsampled contourlet transform domain," *Journal of Digital Imaging*, vol. 27, no. 3, pp. 407–418, Jun 2014.

[41] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Information Fusion*, vol. 8, no. 2, pp. 143 – 156, 2007, Special Issue on Image Fusion: Advances in the State of the Art.

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

[43] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Information Fusion*, vol. 12, no. 2, pp. 74 – 84, 2011.

[44] S. Zheng, W. Shi, J. Liu, G. Zhu, and J. Tian, "Multisource image fusion method using support value transform," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1831–1839, July 2007.

[45] K. P. Upla, M. V. Joshi, and P. P. Gajjar, "An edge preserving multiresolution fusion: Use of contourlet transform and MRF prior," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3210–3220, June 2015.

[46] J. Xu, L. Yang, and D. Wu, "Ripplet: A new transform for image processing," *Journal of Visual Communication and Image Representation*, vol. 21, no. 7, pp. 627 – 639, 2010.

[47] S. Das and M. K. Kundu, "Ripplet based multimodality medical image fusion using pulse-coupled neural network and modified spatial frequency," in *Proceedings of International Conference on Recent Trends in Information Systems*, Dec 2011, pp. 229–234.

[48] G. R. Easley, D. Labate, and F. Colonna, "Shearlet-based total variation diffusion for denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 260–268, Feb 2009.

[49] D. D. Po and M. N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1610–1620, June 2006.

[50] X. Liu, Y. Zhou, and J. Wang, "Image fusion based on shearlet transform and regional features," *AEU - International Journal of Electronics and Communications*, vol. 68, no. 6, pp. 471 – 477, 2014.

[51] X. Liu, W. Mei, and H. Du, "Structure tensor and nonsubsampled shearlet transform based algorithm for CT and MRI image fusion," *Neurocomputing*, vol. 235, pp. 131 – 139, 2017.

[52] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[53] T.-A. Teo and C.-C. Lau, "Pyramid-based image empirical mode decomposition for the fusion of multispectral and panchromatic images," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 4, Jan 2012.

[54] S. M. U. Abdullah, N. U. Rehman, M. M. Khan, and D. P. Mandic, "A multivariate empirical mode decomposition based approach to pansharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3974–3984, July 2015.

[55] A. Dogra, B. Goyal, and S. Agrawal, "From multi-scale decomposition to non-multi-scale decomposition methods: A comprehensive survey of image fusion techniques and its applications," *IEEE Access*, vol. 5, pp. 16 040–16 067, 2017.

[56] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2226–2238, Aug 2006.

[57] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000.

[58] B. K. Shreyamsha Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1193–1204, Jul 2015.

[59] K. N. Chaudhury and K. Rithwik, "Image denoising using optimally weighted bilateral filters: A sure and fast approach," *Computing Research Repository*, vol. abs/1505.00074, 2015.

[60] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1141–1151, Oct 2002.

[61] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Proceedings of Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 815–830.

[62] J. Hu and S. Li, "The multiscale directional bilateral filter and its application to multisensor image fusion," *Information Fusion*, vol. 13, no. 3, pp. 196 – 206, 2012.

[63] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864–2875, July 2013.

[64] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Information Fusion*, vol. 30, pp. 15 – 26, 2016.

[65] E. Lallier and M. Farooq, "A real time pixel-level based image fusion via adaptive weight averaging," in *Proceedings of the Third International Conference on Information Fusion*, vol. 2, July 2000, pp. WEC3/3–WEC313.

[66] C. W. Therrien, J. W. Scrofani, and W. K. Kreb, "An adaptive technique for the enhanced fusion of low-light visible with uncooled thermal infrared imagery," in *Proceedings of International Conference on Image Processing*, vol. 1, Oct 1997, pp. 405–408.

[67] A. Dogra, B. Goyal, S. Agrawal, and C. K. Ahuja, "Efficient fusion of osseous and vascular details in wavelet domain," *Pattern Recognition Letters*, vol. 94, pp. 189 – 193, 2017.

[68] A. Dogra, S. Agrawal, B. Goyal, N. Khandelwal, and C. K. Ahuja, "Color and grey scale fusion of osseous and vascular information," *Journal of Computational Science*, vol. 17, pp. 103 – 114, 2016.

[69] J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems*. Norwell, MA, USA: Kluwer Academic Publishers, 1990.

[70] R. Shen, I. Cheng, and A. Basu, "QoE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2469–2478, June 2013.

[71] Wright, "Fast image fusion with a Markov random field," in *Image Processing And Its Applications, Proceedings of Seventh International Conference on (Conf. Publ. No. 465)*, vol. 2, July 1999, pp. 557–561.

[72] R. Azencott, B. Chalmond, and F. Coldefy, "Markov fusion of a pair of noisy images to detect intensity valleys," *International Journal of Computer Vision*, vol. 16, no. 2, pp. 135–145, Oct 1995.

[73] D. Agrawal and J. Singhai, "Multifocus image fusion using modified pulse coupled neural network for improved image quality," *IET Image Processing*, vol. 4, no. 6, pp. 443–451, December 2010.

[74] Z. Wang, Y. Ma, F. Cheng, and L. Yang, "Review of pulse-coupled neural networks," *Image and Vision Computing*, vol. 28, no. 1, pp. 5 – 13, 2010.

[75] S. Mitra and S. K. Pal, "Fuzzy multi-layer perceptron, inferencing and rule generation," *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 51–63, Jan 1995.

[76] Y. Zhang, L. Chen, Z. Zhao, J. Jia, and J. Liu, "Multi-focus image fusion based on robust principal component analysis and pulse-coupled neural network," *Optik - International Journal for Light and Electron Optics*, vol. 125, no. 17, pp. 5002 – 5006, 2014.

[77] Z. Wang and Y. Ma, "Medical image fusion using M-PCNN," *Information Fusion*, vol. 9, no. 2, pp. 176 – 185, 2008.

[78] R. Wasserman, R. Acharya, C. Sibata, and K. H. Shin, "A data fusion approach to tumor delineation," in *Proceedings of International Conference on Image Processing*, vol. 2, Oct 1995, pp. 476–479.

[79] G. N. Brock, W. D. Beavis, and L. S. Kubatko, "Fuzzy logic and related methods as a screening tool for detecting gene regulatory networks," *Information Fusion*, vol. 10, no. 3, pp. 250 – 259, 2009, Special Issue on Natural Computing Methods in Bioinformatics.

[80] I. Bloch, O. Colliot, O. Camara, and T. Géraud, "Fusion of spatial relationships for guiding recognition, example of brain structure recognition in 3d MRI," *Pattern Recognition Letters*, vol. 26, no. 4, pp. 449 – 457, 2005, iCAPR 2003.

[81] A. Das and M. Bhattacharya, "Evolutionary algorithm based automated medical image fusion technique: Comparative study with fuzzy fusion approach," in *2009 World Congress on Nature Biologically Inspired Computing (NaBIC)*, Dec 2009, pp. 269–274.

[82] J. K. Avor and T. Sarkodie-Gyan, "An approach to sensor fusion in medical robots," in *Proceedings of IEEE International Conference on Rehabilitation Robotics*, June 2009, pp. 818–822.

[83] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, April 2010.

[84] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44.

[85] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Information Fusion*, vol. 13, no. 1, pp. 10 – 19, 2012.

[86] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4779–4789, Sept 2013.

[87] M. Yin, P. Duan, W. Liu, and X. Liang, "A novel infrared and visible image fusion algorithm based on shift-invariant dual-tree complex shearlet transform and sparse representation," *Neurocomputing*, vol. 226, pp. 182–191, 2017.

[88] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, Dec 2016.

[89] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147–164, 2015.

[90] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion: A review," *Computing Research Repository*, vol. abs/1702.03515, 2017.

[91] W. Wang, L. Jiao, and S. Yang, "Fusion of multispectral and panchromatic images via sparse representation and local autoregressive model," *Information Fusion*, vol. 20, pp. 73 – 87, 2014.

[92] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Information Fusion*, vol. 8, no. 2, pp. 131 – 142, 2007, Special Issue on Image Fusion: Advances in the State of the Art.

[93] H. R. Shahdoosti and H. Ghassemian, "Combining the spectral PCA and spatial PCA fusion methods by an optimal filter," *Information Fusion*, vol. 27, pp. 150–160, 2016.

[94] Y. Jiang and M. Wang, "Image fusion with morphological component analysis," *Information Fusion*, vol. 18, pp. 107 – 118, 2014.

[95] S. Li, J. T. Kwok, and Y. Wang, "Combination of images with diverse focuses using the spatial frequency," *Information Fusion*, vol. 2, no. 3, pp. 169 – 176, 2001.

[96] B. Miles, I. B. Ayed, M. W. K. Law, G. Garvin, A. Fenster, and S. Li, "Spine image fusion via graph cuts," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1841–1850, July 2013.

[97] F. Chen, F. Qin, G. Peng, and S. Chen, "Fusion of remote sensing images using improved ICA mergers based on wavelet decomposition," *Procedia Engineering*, vol. 29, pp. 2938 – 2943, 2012, 2012 International Workshop on Information and Electronics Engineering.

[98] M. Ghahremani and H. Ghassemian, "A compressed-sensing-based pan-sharpening method for spectral distortion reduction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2194–2206, April 2016.

[99] C. He, Q. Liu, H. Li, and H. Wang, "Multimodal medical image fusion based on IHS and PCA," *Procedia Engineering*, vol. 7, pp. 280 – 285, 2010, 2010 Symposium on Security Detection and Information Processing.

[100] S. Li and B. Yang, "Hybrid multiresolution method for multisensor multimodal image fusion," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1519–1526, 2010.

[101] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3634–3646, Dec 2011.

[102] K.-L. Hua, H.-C. Wang, A. H. Rusdi, and S.-Y. Jiang, "A novel multi-focus image fusion algorithm based on random walks," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 951 – 962, 2014.

[103] H. Li, L. Li, and J. Zhang, "Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering," *Optics Communications*, vol. 342, pp. 1 – 11, 2015.

[104] D. Wu, A. Yang, L. Zhu, and C. Zhang, "Survey of multi-sensor image fusion," in *Life System Modeling and Simulation*, S. Ma, L. Jia, X. Li, L. Wang, H. Zhou, and X. Sun, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 358–367.

[105] S. S. Malik, S. P. P. Kumar, and G. B. Maruthi, "DT-CWT: Feature level image fusion based on dual-tree complex wavelet transform," in *Proceedings of International Conference on Information Communication and Embedded Systems (ICICES2014)*, Feb 2014, pp. 1–7.

[106] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75 – 89, 2016.

[107] S. Prabhakar and A. K. Jain, "Decision-level fusion in fingerprint verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861 – 874, 2002.

[108] A. J. Rashidi and M. H. Ghassemian, "A new approach for multi-system/sensor decision fusion based on joint measures," *International Journal of Information Acquisition*, vol. 01, no. 02, pp. 109–120, 2004.

[109] F. T. Mahmoudi, F. Samadzadegan, and P. Reinartz, "Object recognition based on the context aware decision-level fusion in multiviews imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 1, pp. 12–22, Jan 2015.

[110] J. Tian, L. Chen, L. Ma, and W. Yu, "Multi-focus image fusion using a bilateral gradient-based sharpness criterion," *Optics Communications*, vol. 284, no. 1, pp. 80 – 87, 2011.

[111] S. Li, X. Kang, J. Hu, and B. Yang, "Image matting for fusion of multi-focus images in dynamic scenes," *Information Fusion*, vol. 14, no. 2, pp. 147 – 162, 2013.

[112] S. Koduri, "Multisensor data fusion with singular value decomposition," in *Proceedings of UKSim 14th International Conference on Computer Modelling and Simulation*, March 2012, pp. 422–426.

[113] J. Tian and L. Chen, "Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure," *Signal Processing*, vol. 92, no. 9, pp. 2137 – 2146, 2012.

[114] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864–2875, 2013.

[115] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Information Fusion*, vol. 23, pp. 139 – 155, 2015.

[116] Y. Liu and Z. Wang, "Dense SIFT for ghost-free multi-exposure fusion," *Journal of Visual Communication and Image Representation*, vol. 31, pp. 208 – 224, 2015.

[117] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, Jan 2016.

[118] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Information Fusion*, vol. 36, pp. 191–207, 2017.

[119] Y. Fei, G. Wei, and S. Zongxi, "Medical image fusion based on feature extraction and sparse representation," *International Journal of Biomedical Imaging*, vol. 2017, no. 3020461, 2017.

[120] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on information measure for performance of image fusion'," *Electronics Letters*, vol. 44, no. 18, pp. 1066–1067, August 2008.

[121] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[122] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[123] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 94–109, Jan 2012.

[124] V. Petrović, "Subjective tests for image fusion evaluation and objective metric validation," *Information Fusion*, vol. 8, no. 2, pp. 208 – 216, 2007, Special Issue on Image Fusion: Advances in the State of the Art.

[125] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Information Fusion*, vol. 9, no. 2, pp. 156–160, 2008.

[126] M. Choi, R. Y. Kim, and M.-G. Kim, "The curvelet transform for image fusion," *International Society for Photogrammetry and Remote Sensing, ISPRS 2004*, vol. 35, pp. 59–64, 2004.

[127] W. Kong and J. Liu, "Technique for image fusion based on nonsubsampled shearlet transform and improved pulse-coupled neural network," *Optical Engineering*, vol. 52, no. 1, pp. 017 001–017 001, 2013.

[128] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Information Fusion*, vol. 31, pp. 100 – 109, 2016.

[129] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol. 45, pp. 153 – 178, 2019.

[130] S. Li, J. T. Kwok, and Y. Wang, "Multifocus image fusion using artificial neural networks," *Pattern Recognition Letters*, vol. 23, no. 8, pp. 985 – 997, 2002.

[131] M. Yin, W. Liu, X. Zhao, Y. Yin, and Y. Guo, "A novel image fusion algorithm based on nonsubsampled shearlet transform," *Optik - International Journal for Light and Electron Optics*, vol. 125, no. 10, pp. 2274 – 2282, 2014.

[132] Q. guang Miao, C. Shi, P. fei Xu, M. Yang, and Y. bo Shi, "A novel algorithm of image fusion using shearlets," *Optics Communications*, vol. 284, no. 6, pp. 1540 – 1547, 2011.

[133] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proceedings of Fourth International Conference on Computer Vision*. IEEE, 1993, pp. 173–182.

[134] L. Cao, L. Jin, H. Tao, G. Li, Z. Zhuang, and Y. Zhang, "Multi-focus image fusion based on spatial frequency in discrete cosine transform domain," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 220–224, 2015.

[135] P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1485–1490, 2005.

[136] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[137] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[138] R. Huang, "Some inequalities for the Hadamard product and the fan product of matrices," *Linear Algebra and Its Applications*, vol. 428, no. 7, pp. 1551–1559, 2008.

[139] M. R. Mitashe, A. R. B. Habib, A. Razzaque, I. A. Tanima, and J. Uddin, "An adaptive digital image watermarking scheme with PSO, DWT and XFCM," in *Proceedings of IEEE International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, Feb 2017, pp. 1–5.

[140] D. Summers, "Harvard whole brain atlas: www. med. harvard. edu/aanlib/home. html," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 74, no. 3, p. 288, 2003.

[141] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," in *Proceedings of Information Processing in Medical Imaging*, vol. 3, no. 6, 1995, pp. 263–274.

[142] R. Wang, F. Bu, H. Jin, and L. Li, "A feature-level image fusion algorithm based on neural networks," in *Proceedings of First International Conference on Bioinformatics and Biomedical Engineering*, July 2007, pp. 821–824.

[143] S. Zhao, X. Chen, S. Wang, J. Li, and W. Yang, "A new method of remote sensing image decision-level fusion based on support vector machine," in *Proceedings of International Conference on Recent Advances in Space Technologies*, Nov 2003, pp. 91–96.

[144] W. Q. Lim, "The discrete shearlet transform: A new directional transform and compactly supported shearlet frames," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1166–1180, May 2010.

[145] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[146] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25 – 46, 2008.

[147] Y.-P. Lin and P. Vaidyanathan, "A Kaiser window approach for the design of prototype filters of cosine modulated filterbanks," *IEEE Signal Processing Letters*, vol. 5, no. 6, pp. 132–134, 1998.

[148] F. Cruz-Roldán, P. Martín-Martín, J. Sáez-Landete, M. Blanco-Velasco, and T. Saramaki, "A fast windowing-based technique exploiting spline functions for designing modulated filter banks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 1, pp. 168–178, 2009.

[149] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[150] R. Dony, "Karhunen-Loeve transform," *The Transform and Data Compression Handbook*, vol. 1, pp. 1–34, 2001.

[151] X. Luo, Z. Zhang, C. Zhang, and X. Wu, "Multi-focus image fusion using HOSVD and edge intensity," *Journal of Visual Communication and Image Representation*, vol. 45, no. Supplement C, pp. 46 – 61, 2017.

[152] B. Liao, L. Yan, W. Mo, J. Shen, and W. Zhang, "Coherence restricted stomp and its application in image fusion," *Journal of Visual Communication and Image Representation*, vol. 40, no. Part B, pp. 559 – 573, 2016.

[153] I. Riaz, T. Yu, Y. Rehman, and H. Shin, "Single image dehazing via reliability guided fusion," *Journal of Visual Communication and Image Representation*, vol. 40, no. Part A, pp. 85 – 97, 2016.

[154] J. Adu, S. Xie, and J. Gan, "Image fusion based on visual salient features and the cross-contrast," *Journal of Visual Communication and Image Representation*, vol. 40, no. Part A, pp. 218 – 224, 2016.

[155] M. Manchanda and R. Sharma, "A novel method of multimodal medical image fusion using fuzzy transform," *Journal of Visual Communication and Image Representation*, vol. 40, no. Part A, pp. 197 – 217, 2016.

[156] N. Kaplan, I. Erer, and O. Ersoy, "Fusion of multifocus images by lattice structures," *Journal of Visual Communication and Image Representation*, vol. 38, no. Supplement C, pp. 848 – 857, 2016.

[157] D. B. H. Tay and N. G. Kingsbury, "Flexible design of multidimensional perfect reconstruction FIR 2-band filters using transformations of variables," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 466–480, Oct 1993.

[158] W. F. Schreiber, "Wirephoto quality improvement by unsharp masking," *Pattern Recognition*, vol. 2, no. 2, pp. 117 – 121, 1970.

[159] X. Qu, J. Yan, and G. Yang, "Sum-modified-Laplacian-based multifocus image fusion method in cycle spinning sharp frequency localized contourlet transform domain," *Optics and Precision Engineering*, vol. 13, no. 2, 2005.

[160] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2898–2909, 2012.

[161] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.

[162] ——, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Information Fusion*, vol. 13, no. 1, pp. 10–19, 2012.

[163] Y. Liu and Z. Wang, "Multi-focus image fusion based on sparse representation with adaptive sparse domain selection," in *Proceedings of Seventh International Conference on Image and Graphics (ICIG)*. IEEE, 2013, pp. 591–596.

[164] H. Yin, S. Li, and L. Fang, "Simultaneous image fusion and super-resolution using sparse representation," *Information Fusion*, vol. 14, no. 3, pp. 229–240, 2013.

[165] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, April 1997.

[166] D.-C. Chang and W.-R. Wu, "Image contrast enhancement based on a histogram transformation of local standard deviation," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 518–531, Aug 1998.

[167] E. Elboher and M. Werman, "Cosine integral images for fast spatial and range filtering," in *Proceedings of 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 89–92.

[168] https://homepages.cae.wisc.edu/~ece533/images/ (access on 26/10/2018).

[169] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Information Fusion*, vol. 23, pp. 139–155, 2015.

[170] Y. Yang, S. Tong, S. Huang, and P. Lin, "Multifocus image fusion based on NSCT and focused area detection," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2824–2838, 2015.

[171] V. Papyan and M. Elad, "Multi-scale patch-based image restoration," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 249–261, Jan 2016.

[172] D. Ghosh and N. Kaabouch, "A survey on image mosaicing techniques," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 1 – 11, 2016.

[173] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, Jan. 2006.

[174] K. I. Okumura, S. Raut, Q. Gu, T. Aoyama, T. Takaki, and I. Ishii, "Real-time feature-based video mosaicing at 500 fps," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 2665–2670.

[175] G. Gao and K. Jia, "A new image mosaics algorithm based on feature points matching," in *Proceedings of Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, Sept 2007, pp. 471–471.

[176] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[177] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proceedings of Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 430–443.

[178] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.

[179] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proceedings of Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417.

[180] Y. Xiong and K. Turkowski, "Registration, calibration and blending in creating high quality panoramas," in *Proceedings of Fourth IEEE Workshop on Applications of Computer Vision, 1998. WACV '98.*, Oct 1998, pp. 69–74.

[181] A. Pandey and U. C. Pati, "A novel technique for non-overlapping image mosaicing based on pyramid method," in *Proceedings of Annual IEEE India Conference (INDICON)*, Dec 2013, pp. 1–6.

[182] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or "how do I organize my holiday snaps?"," in *Proceedings of Computer Vision — ECCV 2002*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 414–431.

[183] R. Song and J. Szymanski, "Auto-sorting scheme for image ordering applications in image mosaicing," *Electronics Letters*, vol. 44, no. 13, pp. 798–799, June 2008.

[184] A. Vishwakarma, M. K. Bhuyan, and Y. Iwahori, "Non-subsampled shearlet transform-based image fusion using modified weighted saliency and local difference," *Multimedia Tools and Applications*, vol. 77, no. 24, pp. 32 013–32 040, Dec 2018.

[185] S. Kirthica and R. Sridhar, "Horizontal scaling and aggregation across heterogeneous clouds for resource provisioning," *Computers and Electrical Engineering*, vol. 69, pp. 301 – 316, 2018.

[186] K. Sharma, "Feature-based efficient vehicle tracking for a traffic surveillance system," *Computers and Electrical Engineering*, vol. 70, pp. 690 – 701, 2018.

[187] C. Wang, X. Jing, and C. Zhao, "Local upsampling Fourier transform for accurate 2d/3d image registration," *Computers and Electrical Engineering*, vol. 38, no. 5, pp. 1346 – 1357, 2012, Special issue on Recent Advances in Security and Privacy in Distributed Communications and Image processing.

[188] E. Musoll, "Variable-size mosaics: A process-variation aware technique to increase the performance of tile-based, massive multi-core processors," *Computers and Electrical Engineering*, vol. 37, no. 6, pp. 1193 – 1211, 2011.

[189] Z. Hossein-Nejad and M. Nasri, "An adaptive image registration method based on SIFT features and RANSAC transform," *Computers and Electrical Engineering*, vol. 62, pp. 524 – 537, 2017.

[190] L. Cheng, M. Li, Y. Liu, W. Cai, Y. Chen, and K. Yang, "Remote sensing image matching by integrating affine invariant feature extraction and RANSAC," *Computers and Electrical Engineering*, vol. 38, no. 4, pp. 1023 – 1032, 2012.

[191] B. Shi, S. Bai, Z. Zhou, and X. Bai, "Deeppano: Deep panoramic representation for 3-D shape recognition," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2339–2343, Dec 2015.

[192] Y. Wu and J. Li, "Concentric mosaic compression with rebinning of slits (ROSS)," *IEEE Signal Processing Letters*, vol. 9, no. 9, pp. 269–271, Sept 2002.

[193] W. Yao and Z. Li, "Instant color matching for mobile panorama imaging," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 6–10, Jan 2015.

[194] M. Y. Ju, C. Ding, D. Y. Zhang, and Y. J. Guo, "Gamma-correction-based visibility restoration for single hazy images," *IEEE Signal Processing Letters*, pp. 1–1, 2018.

[195] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío, "Fusion-based variational image dehazing," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 151–155, Feb 2017.

[196] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 323–327, March 2018.

[197] http://www.mr-tip.com/serv1.php (access on 26/10/2018).

[198] https://sites.google.com/a/umich.edu/eecs442-winter2015/homework/image-stitching?tmpl=%2Fsystem%2Fapp%2Ftemplates%2Fprint%2F&showPrintDialog=1 (access on 26/10/2018).

[199] https://in.mathworks.com/matlabcentral/fileexchange/46148-image-mosaicing (access on 26/10/2018).

[200] N. Gracias, M. Mahoor, S. Negahdaripour, and A. Gleason, "Fast image blending using watersheds and graph cuts," *Image and Vision Computing*, vol. 27, no. 5, pp. 597–607, 4 2009.

[201] https://github.com/daeyun/Image-Stitching/tree/master/img (access on 26/10/2018).

[202] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proceedings of the 2011 International Conference on Computer Vision*, ser. ICCV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2548–2555.

List of Publications

International Journals

[1] A. Vishwakarma, M. K. Bhuyan, and Y. Iwahori, "Non-subsampled shearlet transform-based image fusion using modified weighted saliency and local difference," *Multimedia Tools and Applications, Springer*, vol. 77, no. 24, pp. 32013–32040, Dec 2018.

[2] A. Vishwakarma, M. K. Bhuyan, and Y. Iwahori, "An optimized non-subsampled shearlet transform-based image fusion using hessian features and unsharp masking," *Journal of Visual Communication and Image Representation, Elsevier*, vol. 57, pp. 48 – 60, 2018.

[3] A. Vishwakarma and M. K. Bhuyan, "Image fusion using adjustable non-subsampled shearlet transform," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–12, 2018 (available online).

Book Chapters

[1] A. Vishwakarma, M. K. Bhuyan, and Y. Iwahori, "An efficient algorithm for medical image fusion using nonsubsampled shearlet transform," in *Proceedings of 2nd International Conference on Computer Vision & Image Processing*, Singapore, LNCS, Springer, 2018, pp. 243–252.

[2] S. K. Kashyap, D. Jat, M. K. Bhuyan, A. Vishwakarma, and P. Gadde, "Zernike Moment and Mutual Information based Methods for Multi-modal Image Registration," in *Proceedings of 3rd International Conference on Computer Vision & Image Processing*, LNCS, Springer, 2019.

National and International Conferences

[1] A. Vishwakarma and M. K. Bhuyan, "Infrared and visible image fusion using NSST and phase stretch transform," in *Proceedings of IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, March 2017, pp. 472–476.

Manuscript Communicated

[1] A. Vishwakarma and M. K. Bhuyan, "Image Mosaicking using Improved Auto-sorting Algorithm and Local Difference-based Harris Features", communicated to *Computers and Electrical Engineering, Elsevier.*

Manuscript under Preparation

[1] A. Vishwakarma and M.K. Bhuyan, "Pixel-level image fusion: a research overview".

[2] A. Vishwakarma and M. K. Bhuyan, "Multi-focus image fusion using hybrid cosine and K-SVD learned dictionary".

[3] A. Vishwakarma and M.K. Bhuyan, "An application of fused images for the general classification of human brain diseases using deep CNN". (Apart from thesis)