Development of Novel Techniques for Image Fusion with Improved Depth of Field and Dynamic Range

A thesis

Submitted in fulfillment of the requirement for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Electronics and Communication Engineering

By

Vishal Chaudhary
Reg. No: 951406002

Under the supervision of

Dr. Vinay Kumar
Associate Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Department of Electronics and Communication Engineering

Thapar Institute of Engineering and Technology

(Deemed to be University),

Patiala-147004, India

December 2020

Certificate

I, Vishal Chaudhary, hereby certify that the work, which is being presented in the thesis, entitled "Development of Novel Techniques for Image Fusion with Improved Depth of Field and Dynamic Range" submitted in Thapar Institute of Engineering and Technology, Patiala in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Electronics and Communication Engineering, is an authentic record of my own work carried out during the period Aug 2014 to Dec 2020 under the supervision of Dr. Vinay Kumar.

I have also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken. The matter presented in this thesis has not been submitted elsewhere for the award of any other degree or diploma from any institution.

Date: 10.12.2020

Vishal Chaudhary
Candidate

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Dr. Vinay Kumar
Associate Professor

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Vinay Kumar for the continuous support of PhD research work, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank the rest of my thesis committee: Dr. Kulbir Singh, Dr. Amit Kumar Kohli and Dr. Neeraj Kumar for their insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives.

I would like to thank Dr. Alpana Agarwal (Head of Department, ECED) for the support by the department and Dr. Rafat Siddique (Dean of Research and Sponsored Projects) for all the support for doing research in the University.

Last but not the least, I would like to thanks my parents for continuous encouragement and keeping my morale high, my wife for her patience and support, and my son and daughter who missed me throughout these years.

Special thanks to almighty God for giving me chance to pursue PhD and letting me feel his presence all the time.

Vishal Chaudhary

Abstract

Image fusion technique is a popular technique to extract scene information by combining complementary information from pre-registered captured images. Image fusion has many applications like remote sensing, medical imaging, surveillance and photography. Present work has proposed fusion techniques in the field of photography. In photography, captured images suffered from low depth of field (DoF) and lack of dynamic range. We have tried to address both the issues.

The fusion can be performed in spatial domain or transform domain. In spatial domain direct operation is performed on images in order to extract complementary information whereas in transform domain, images are first converted to transformed images for processing and then converted back to synthesize fused image. Both techniques have their positives and negatives.

Evaluation of fusion technique is also become important in order to evaluate the quality and to compare with stat-of-the-art techniques. Two type of evaluation techniques are available; subjective and objective evaluation. In subjective evaluation, group of trained peoples observe the image and rate them on given scale. Whereas in objective evaluation, mathematics based technique are used to test the quality. In subjective evaluation, different people have different views which sometime affect the rating of quality image. Therefore objective evaluation gives more concrete results.

We have proposed three spatial domain fusion techniques by using filters. Three different spaces, block, feature and regions, are explored to increase depth of field and dynamic range. In block based technique, all pre-registered and captured images are splitted into local and global layers by using Neighbor distance filter. Local layer is processed with block based fusion whereas global layer is combined as weighted sum. Finally combined local and global layer are combined to synthesize fused image. It addressed both low depth of field as well as lack of dynamic range.

In feature based technique, combination of recursive and Gaussian filter separates three features, constant, low varying and high varying features, from pre-registered and captured images. These features are processed separately for features with high activity level and further combined together to synthesize fused image. In region based technique, pre-registered and captured images are splitted into frequency and base layer by using

recursive filter. The scene is divided into three regions based on intensity of captured images. The frequency layer is processed with region based fusion by using defined three regions, whereas base layer is combined as weighted sum. Finally fused image is synthesize by using processed frequency and base layer. Both feature and region based techniques addressed lack of dynamic range.

The proposed techniques are tested on randomly selected standard data set. Objective evaluation based on probability, edges transfer are used to evaluate the quality of proposed techniques and compared with state-of-the-art techniques.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRNOYMS

CCD	Charge Coupled Device
CRF	Camera Response Function
DoF	Depth of Field
EF	Exposure Fusion
HDR	High Dynamic Range
LDR	Low Dynamic Range
MI	Mutual Information
MSE	Mean Square Error
MRI	Magnetic Resonance Imaging
MGA	Multi-scale Geometric
MF	Median Filter
NDFM	Normalize Difference in Focus Measure
NMI	Normalized Mutual Information
ND	Neighbor Distance
PSNR	Peak Signal to Noise Ratio
RMSE	Root Mean Square Error
RF	Recursive Filter
SLR	Single Lens Reflex
SSIM	Structure Similarity based Metric
SIFT	Scale Invariant Feature Transform
VIFF	Visual Information Fidelity for Fusion

Chapter – 1

Introduction

Digital image processing is a process of performing various operations on digital images. The processes, like noise reduction, enhancement, segmentation etc. allow to extract attributes from a digital image up to, and including, the recognition of individual objects [1].

An image may be defined as two dimensional function $I(p, q)$, where (p, q) represents the co-ordinates. The value of $I(.)$, at any pair of co-ordinates, gives the intensity at that location. The function $I(.)$ is called as digital image, and each element is called as pixel or pel [1-2].

An image is captured by optical device called camera. Traditional camera uses photographic film to capture the reflected light of scene. Whereas digital cameras use electronic sensor, like charge coupled device (CCD) or CMOS sensor to capture reflected light. Broadly digital camera are of two types: compact and single lens reflex (SLR) camera.

The advancement in VLSI technologies has made possible to design very powerful central processing unit (CPU) to allow the processing of digital image with an ease. The smaller size and yet powerful CPU leads to development of affordable smart phones with integrated low dynamic range camera (LDR). Smartphones, now a days in the hand of almost every person, makes the images and its processing an inevitable part of their life.

1.1 Camera

A camera is an optical device to capture the photons reflected from the subject and fall on photographic film or electronic sensor. The photons pass through lens system which focuses it on to sensor. The information recorded on sensors is known as image. Apart from the basic function of recording an image some additional features are also provided in the camera:

- View-finder
- Control of exposure time
- Focus or un-focus of desired section

Fig. (1.1), shows the image of typical camera and its related functions [92]. The outer part of camera, which holds the capturing mechanism, is known as camera body. The capturing mechanism includes viewfinder, shutter release, data display, which helps the photographer to capture more suitable image. The front shows the lens system including aperture to identify focused or unfocused section. The central part shows the electronic sensor (cross-section) which records the image.

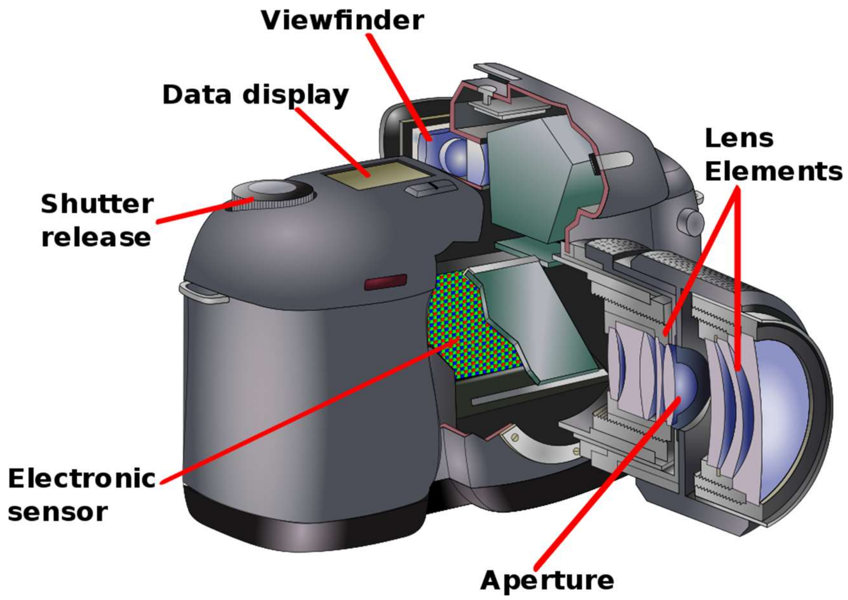


Fig. 1.1 Camera with labels

1.1.1 Classification

Camera can be classify into two categories based on sensor used to record an image; traditional and digital camera. Mostly digital camera is in use now a days.

Digital camera is further classified into two categories: compact camera and single lens reflex (SLR) camera. Compact camera is small, lightweight and inexpensive, refer Fig. (1.2) [93]. They are usually come with standard and automatic settings. It will select appropriate exposure automatically. The cameras like zoom compact camera and advanced compact camera from the family of compact camera has feature of zooming, manual selection of focus and exposure. These feature also increases the cost proportionally.

Fig. 1.2 Compact Camera

SLR cameras are professional cameras, refer Fig. (1.3) [94]. They are larger, heavier and costlier than compact camera. They offer interchangeable lens, zooming feature and variety of manual settings. The cost is a concern with SLR cameras as compared to compact camera.

Fig. 1.3 DSLR Camera

1.1.2Features of Digital Camera

A typical digital camera has many features, which influence the image quality. These features are discussed briefly in the following section

(i) Shutter speed: The shutter opening and closing of the camera allows the exposure of sensor to the incoming light. Speed of shutter decides the time for which the sensor is exposed to incoming light. The time is defined as exposure time and inversely proportional to shutter speed.

Shutter speed correlates amount of light and exposure time in 1:1 ratio. For example, amount of light entering doubles with double in exposure time.

(ii) Aperture: The aperture controls the amount of light enters into the camera through the lens by controlling its area. Larger aperture allows more amount of light and vice-versa.

It is specified as f-stop value, e.g. f/2.8. The increase in value of f-stop decreases the area of aperture. For same exposure, with the decrease in f-stop value to half (f/2.8 to f/1.4), results in increase in amount of light by four times.

(iii) ISO: ISO determines the sensitivity of camera to the incoming light. High ISO increases the sensitivity whereas low ISO decreases it. However low ISO is recommendable for normal use since high ISO besides increasing the sensitivity also increases the image noise.

(iv) Camera Lens: A camera lens is an optical lens which collects the light reflected from the subject and pass on to camera sensor. The light while passing through the lenses get converge in between the lens system followed by absorbing by sensor. The converging point of light is known as nodal point. The distance between the nodal point and sensor is known as focal length.

The camera lenses are classified into two major categories: Prime and Zoom lenses. Prime lenses have fixed focal length whereas Zoom lenses have varying focal length. Another classification based on focal length is wide angle lenses and telephoto lenses. Wide angle lenses have short focal length whereas telephoto lenses have long focal length.

The camera lenses of compact camera are mostly prime lenses, however some cameras have limited zoom lens capability. On the other hand SLR camera have very rich features in terms of shutter speed, ISO speed, aperture and zoom lenses. These rich features in SLR camera also increases the cost. SLR with full of features comes with higher cost and out of average person’s range.

1.1.3Fusion

The compact camera, also defined as LDR camera, is capable of capturing low dynamic range (LDR) images. These images are not able to capture all-inclusive details of scene due to limited focal capabilities of camera. The effects on captured images are many, like lack of focus or lack of exposure. One way to capture all details of scene into an image is by using HDR camera (SLR camera), but cost of camera limits its use. The other solution is fusion of images. Fusion is a process of combining data collected from different sources into more informative data. Example, Infrared and visual data, RADAR and SONAR [4].

With respect to image fusion, it is defined as a process of combining the complementary information from multiple images into single image. Multiple images are captured by using different sensors or by using different settings of camera.

Image fusion is classified, based on types of sources and image capturing conditions, as [32]

a) Multi-modal/Multi-sensor fusion: It fuses the images captured with different sensors, e.g. fusion of infrared and visible image of same scene.

b) Multi-view fusion: It fuses the captured images with different view point. It is useful for partial opaque scenes to improve image quality.

c) Multi-exposure fusion: It fuses differently exposed images, e.g. under and over exposed images are fused together. Shutter speed is utilized to vary exposure time.

d) Multi-focus fusion: It fuses the captured images with different focused sections, e.g. near and far focused images are fused together. Camera lens is responsible feature for capturing images with different focused sections. Aperture size and object distance also affects focused sections.

The proposed work has focused on multi-exposure and multi-focus image fusion. The following sections explains the basics of these two fusion methods.

1.2 Multi-focus Image Fusion

A camera captures image of scene by passing reflected light from scene (subject) to film plane through lens. The quality of subject image, created on film, is largely dependent on lens. If the distance between lens and film plane is appropriate then the quality of object image would be sharper otherwise it become blurry and underlying information is lost. So, on one hand lens helps to capture image, whereas on the other hand it become responsible to blur the image. The mechanism of sharpening one section and blurring other section of image is defined as Depth of Field.

The depth of field (DoF) of digital camera depends on geometrical optical model of the camera. In camera, for certain distance between subject and lens, subject image created on plane is sharply focused and other distances are blurred. Fig. (1.4) shows the geometric optical model of image formation [95]. The three dots white ('B'), yellow ('A') and brown ('C') represent three subjects and white ('b'), yellow ('a') and brown ('c') are the corresponding subject images on the image plane. The distance for subject 'A' is selected such that image "a" is sharply focused and image formed exactly on the image plane. The subject 'B' and 'C' are at larger and smaller distance and results in image formation in front and behind the image plane. These images are less focused as compare to "a". The position of images "b" and "c" on image plane is a circle and called as circle of confusion. The distance of dots 'B', 'C' increases from dot 'A', increases the circle of confusion and thus blurriness. The distance from "a" on the image plane up to which the sharpness is acceptable is known as depth of field (DoF). Fig (1.5) shows the range of object distance over which sharpness ($\Delta d_{i,acceptable}$) is acceptable on sensor (film plane) [96]. The tip of the arrow represents the dot in Fig (1.5). One would like to capture an image with large depth of field.

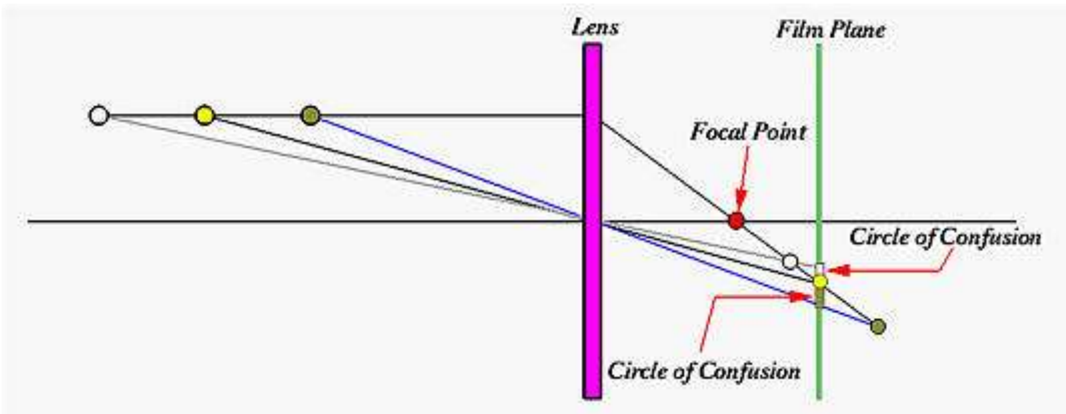


Fig. 1.4 Geometric Optic Model, circle of confusion

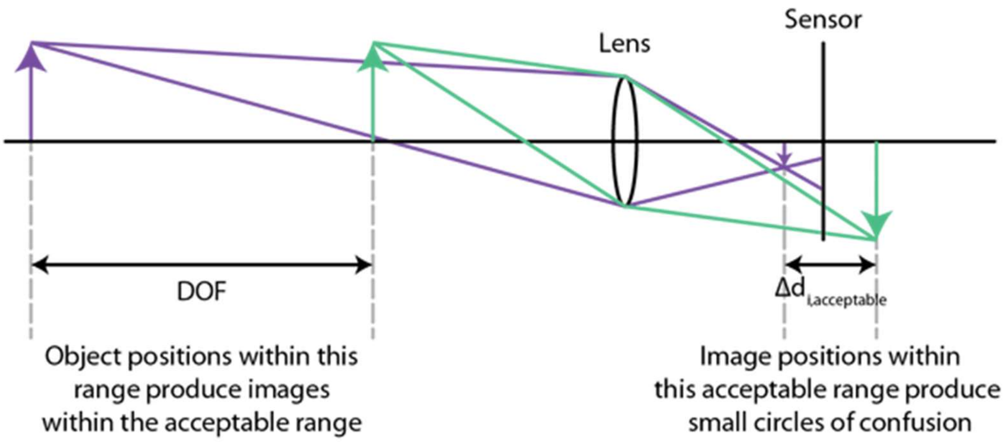


Fig. 1.5 Geometric Optic Model, depth of field

An image captured by LDR camera usually suffers from low DoF. It results in loss of information (due to blurriness) in various parts of image. One can use HDR camera to capture image with large DoF. The high cost of HDR camera limits its use to most of the people. The other solution is multi-focus image fusion.

All captured images, collects underlying information of scene with different and limited DoF. The limited DoF results in sharper as well as blurred sections. The sharper section is termed as focused and blurred section is termed as unfocused section. The fusion process collects focused sections from respective captured images and combined together to construct fused image. The constructed fused image has larger depth of field as compare to captured images. The process of fusion thus helps to increase the DoF of image without using HDR camera.

(a) **(b)** **(c)**

Fig. 1.6 Source images 'Pepsi' **a.** left focused, **b.** right focused, **c.** fused image

Fig. (1.6) [102] represents source images with Fig. (1.6(a)) has left focused sections, Fig. (1.6(b)) has right focused sections and Fig. (1.6(c)) is fusion of Fig. (1.6(a)-(b)). Clearly fused image has better focused area or high depth of field.

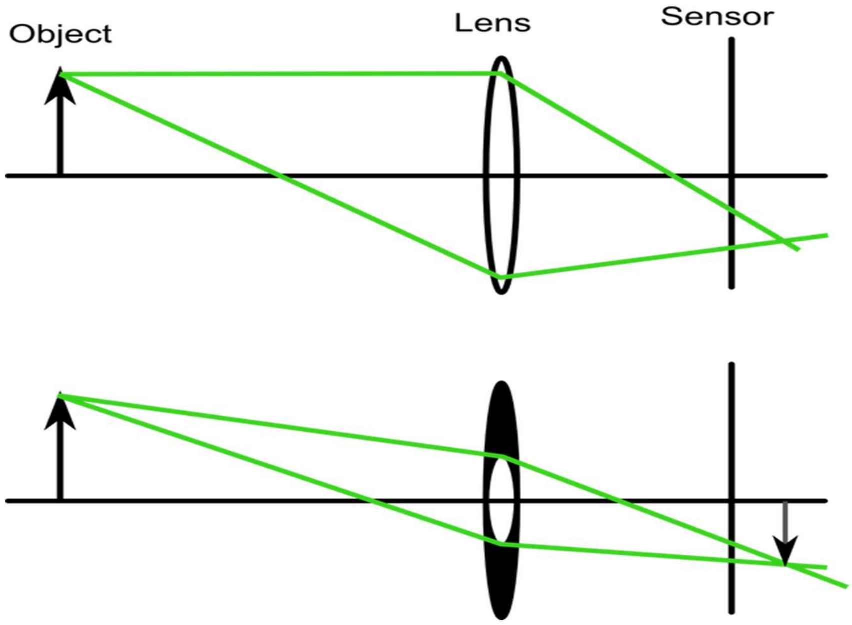


Fig. 1.7 Top: Full aperture, Bottom: Narrow aperture

Aperture and object distance are other two parameters which also affect the depth of field. Aperture controls the amount of light passing through the lens by changing its shape. The shrink in size of aperture narrows down the light entering into the camera and results in increase in DoF. The reverse is true for increasing the aperture size. Fig. (1.7) represent two positions of aperture; top figure represent the full aperture, bottom figure represent the narrow aperture [97]. Similarly, it is observed that increase in object distance from lens increases the depth of field.

1.3 Multi-exposure Image Fusion

Human can perceived scene luminance from 10^{-3} to 10^{5} cd/m^2, whereas the dynamic range of LDR camera is 30 to 10^2 cd/m^2; i.e. human will be able to perceive loss of information in captured images.

An image is captured by collecting reflected light of scene by camera sensor. The light passes through shutter of camera before striking on to sensor. The time for which shutter remain open, plays important role in transferring information of scene to image. If the shutter speed is high, it gives less time (exposure time) to capture the details whereas lower speed gives more time to capture the details. Short exposure time allows camera to collect details present in brighter section and large exposure time allows in darker section. In both cases, there is a possibility that complete details may not be captured. Fig. (1.9(a-c)) represent captured images by LDR camera in increasing exposure time.

The immediate solution is to use high dynamic range (HDR) camera to capture the detail of scene, but find limited use due to expense involved. Our challenge is to develop methods or algorithms to combine the details in one picture.

One way to combine details of multiple images is with two stage HDR-tone mapping process. At first stage LDR images are combined to create HDR image using camera response function (CRF). Followed by suitable tone mapping process to compress down to LDR image in second stage, refer Fig. (1.8). An image created by this process is called HDR like image. This process of constructing HDR image needs information of source images like exposure time and exposure value to construct CRF.

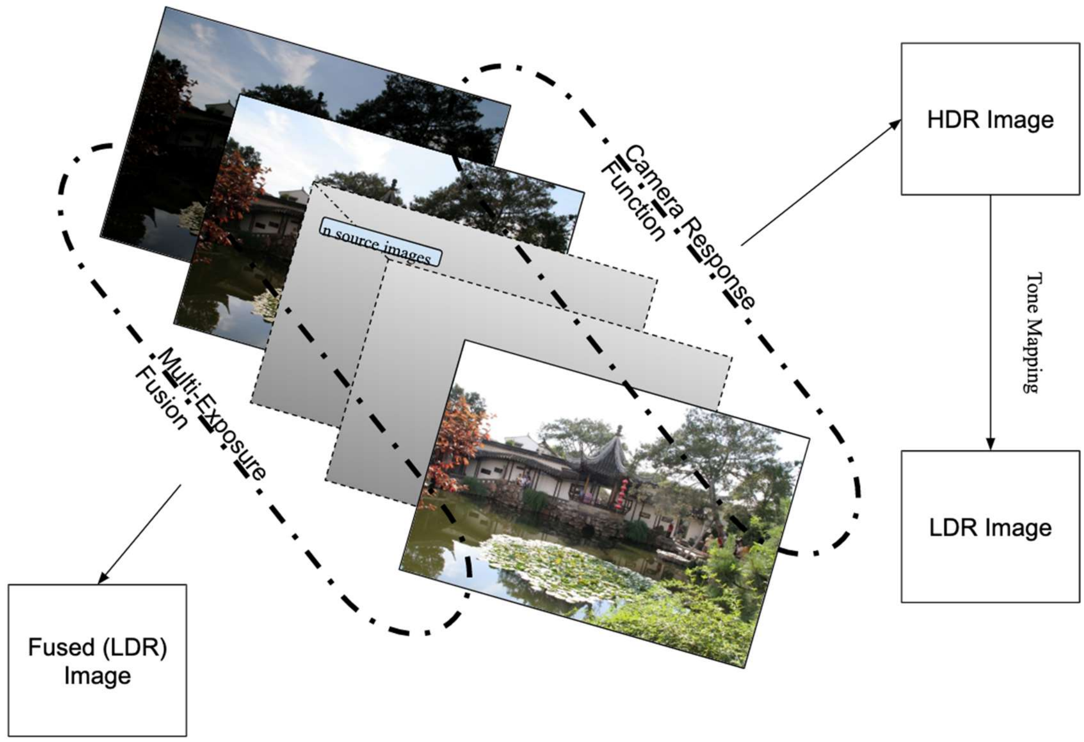
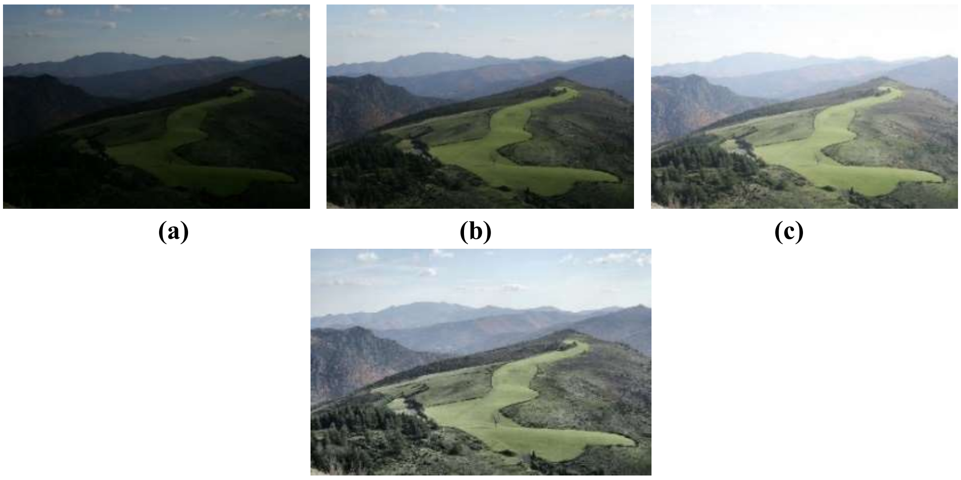


Fig. 1.8 Multi-exposure and tone-mapping

HDR-tone mapping is a complex and time consuming process due to two stage process. Requirement of image information to construct CRF poses another drawback. The other solution is multi-exposure image fusion. In multi-exposure image fusion, sections with maximum activity level are extracted from source images and combined to construct fused image. Unlike HDR-tone mapping process, this process requires no information of source images. It also reduces processing time as well as complexity due to reduction from two-step process to one-step process, refer Fig (1.8).



(a) (b) (c)

(d)

Fig. 1.9 Source images 'Landscape' **a.** low exposure time, **b.** medium exposure time, **c.** high exposure time, **d.** fused image

Figure (1.9), represents an example of images captured at varying shutter speed (or exposure time) and fused image [107]. Fig. 1.9(a) is a under exposed image which is captured at high shutter speed (or low exposure time). In under-exposed image, sections of scene with pixel intensity towards darker side become darkest (or black) like bottom left section whereas pixels with intensity on higher side usually captured at correct intensity. Fig. 1.9(c) is an over-exposed image which is captured at lower shutter speed (or high exposure time). This image collect darker sections correctly, whereas brighter sections become brightest like sky part of an image. Fig 1.9(b) is a normally exposed image captured at medium shutter speed (or medium exposure time). The image captures pixels with intensity in between under and over exposed image. Fig. 1.9(d) is a fused image of three captured image, clearly it has maximum information of scene.

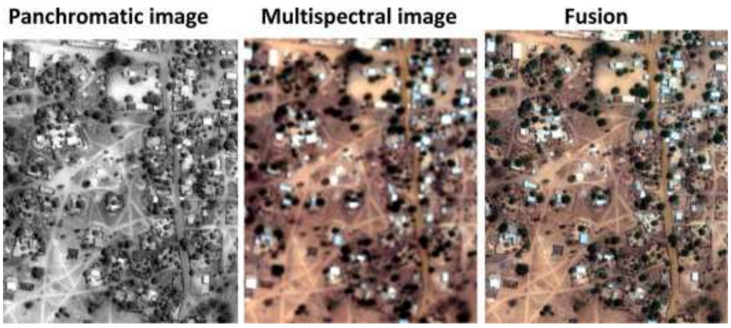
1.4 Applications

Image fusion has become a very important technique for image analysis and computer vision. The fused image obtained from multiple images gives a better view of image for further processing by human or machine and finds lot of applications like remote sensing, medical imaging, surveillance and navigation, and photographic applications [3].

1.4.1 Remote Sensing

Remote sensing is a technique to collect object information without making physical contact to it. Image processing plays important role in remote sensing applications such as crop monitoring, weather prediction.

Image fusion also helps to better understand the remotely collected data; e.g. fusion of panchromatic and multi-spectral images. Panchromatic images collect high spatial resolution information, whereas multi-spectral images collect the spectral information of object. Fusion of both images provide spatial as well as spectral information in one image, refer Fig. (1.10) [98].



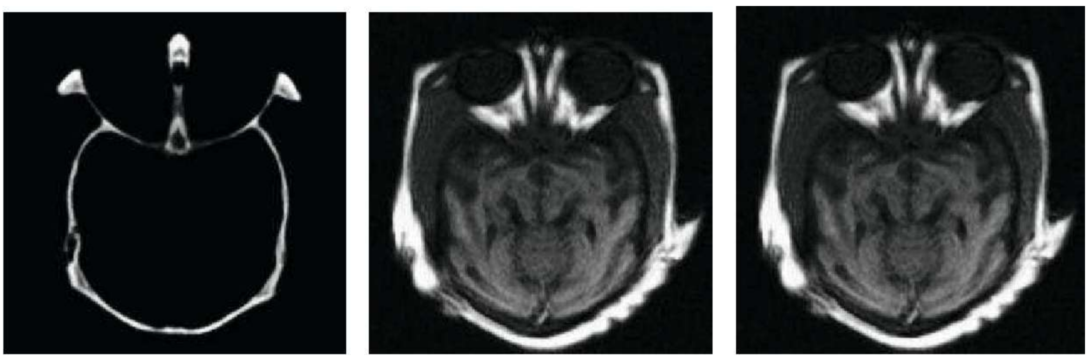
(a) (b) (c)

Fig. 1.10 Remote sensing **a.** Panchromatic image **b.** Multispectral image **c.** Fused image

1.4.2 Medical Imaging

Medical imaging is a process of representing internal parts/behaviour of body as visual representation for clinical analysis and medical intervention. Radiography, Magnetic Resonance Imaging (MRI), Ultrasound and Computerized Tomography (CT) are few examples of medical imaging.

Image fusion in medical imaging, combines information collected by different sensors into one image. For example fused image of MRI and CT gives the soft tissue structure and bone structure together. The combined information helps medical practitioners to analyse and take better decision, refer Fig. (1.11) [99].



(a) (b) (c)

Fig. 1.11 Medical Imaging **a.** CT image **b.** MRI image **c.** Fused image

1.4.3 Surveillance

Surveillance is monitoring of behavior of people or activities in area to direct or protect people. Surveillance is usually done by CCTV or thermal cameras in the form of images, therefore image processing is required to analyze it.

Image fusion techniques can contribute by combining visible image (day image) and infrared image (thermal image). Infrared image captures objects based on temperature but fails to collect spatial information, whereas visible images collects spatial resolution information, refer Fig. (1.12) [100]. The same fused image also be used as eye for day as well as night vision.

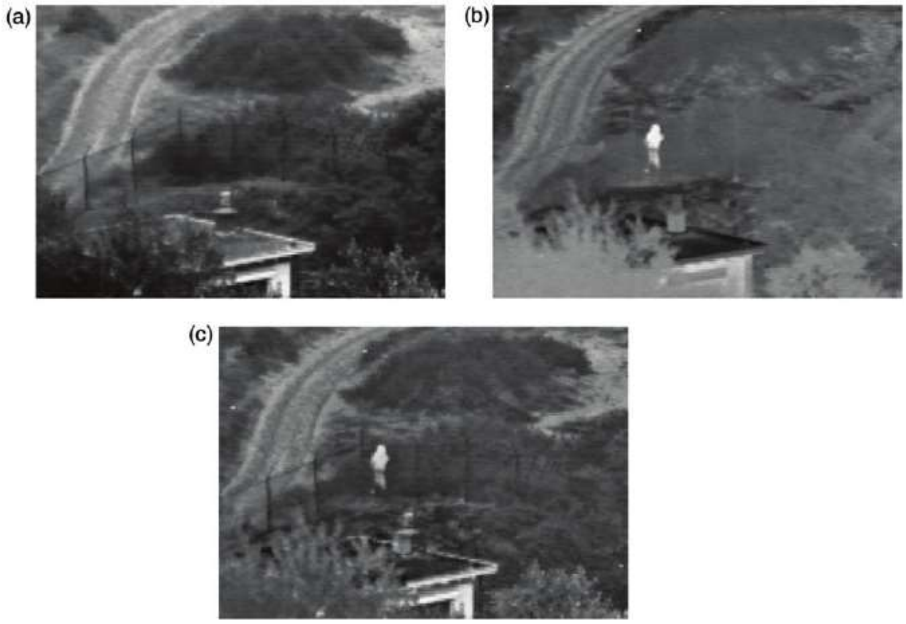


Fig. 1.12 Surveillance **a.** Visible image **b.** Infrared image **c.** Fused image

1.4.4 Photography

Photography, an image captured by camera, suffers from low depth of field (DoF) and lack of exposure level while capturing images. Both degrades the quality of image and thus become less useful. The quality can be significantly improved by using multi-focus/multi-exposure fusion algorithms.

Multi-focus image fusion combine captured images with different limited DoF to construct enhanced DoF fused image, refer Fig (1.6). Similarly multi-exposure image fusion combines captured images with different exposure level to construct well exposed image, refer Fig. (1.9).

Chapter – 2

Literature survey

Image fusion technique collects more information of scene by combining multiple captured images as compare to single image. Many researchers have attempted to develop number of fusion techniques for multi-focus and multi-exposure domain. In this chapter we will present major works available in literature. There are two major methods for image fusion: multi-focus and multi-exposure. We discuss them in rest of the chapter.

2.1 Multi-Focus Image Fusion

Multi-focus image fusion combines co-registered images captured with different focused sections, refer (Chapter 1). Fusion techniques are broadly divided into two categories: spatial domain and transform domain. In spatial domain direct pixels are processed to extract informative sections and synthesize fused image. Whereas in transform domain, processing takes place by transforming images into space-frequency domain. The resulting coefficients are processed to synthesize fused image by corresponding inverse transform. Spatial domain techniques are influenced by two factors; first, image segmentation into regions or features, second, selection of informative region or feature. Transform domain techniques are influenced by selection of transform method and identification of informative coefficients.

2.1.1 Spatial domain

In spatial domain, simplest way for fusion is by averaging pixel intensities of all captured images. Weighted average of pixels of captured images is another simple way of fusing images. The simplicity of mathematics of these two methods for image fusion also accompanied with loss of contrast in fused image.

In order to improve the contrast of fused image, research community has developed region based fusion techniques. Regions may be of predefined size, like blocks or arbitrary shape. Region based technique segments the captured images into predefined blocks or arbitrary shape regions. Most informative sections are selected from corresponding captured images to synthesize fused image.

S. Li et al. [8] were first among the few who used block based fusion technique to combine captured images. In their proposed technique, all captured images are divided into blocks of predefined size. Each block is then tested for underlying information using spatial frequency. S. Li et al. [9] proposed another block based technique by using artificial neural network. Three features, spatial frequency, visibility and edge feature, extracted from few blocks of images are used to train the neural network. These trained networks then identify informative blocks of source images to synthesize fused image. Similarly, W. Huang et al. [16] proposed block based technique, where Energy of image Laplacian (EOL) is used to train Pulse Coupled Neural Network (PCNN). PCNN is further used to identify blocks with better underlying information for image fusion. Apart from pre-defined block size based fusion few techniques [17, 26] are proposed where, prior to block based fusion, block sizes are first optimized by using Genetic or Differential Evolution algorithm.

For arbitrary shape region based fusion, S Li et al. [18] proposed technique where captured images are divided into regions using normalized cut method. Each region is tested by spatial frequency to test its activity level, most informative regions are then combined to synthesize fused image. Y. Zhang et al. [20] proposed technique in which images are initially divided into predefined size of blocks and converted into regions based on blurring measure. Finally region based fusion is performed to synthesize fused image. Similarly, I. De et al. [31], proposed region based fusion where images are initially divided into quad tree structure blocks. These blocks are converted into regions based on normalize difference in focus measure (NDFM). Another region based technique is proposed by X. Bai et al. [39], where pre-defined blocks are converted into regions. Regions are then tested by newly proposed activity level testing measure, sum of weighted modified Laplacian, and selected informative regions are combined to synthesize fused image. Y. Zhang et al. [34] perform region based fusion by using non-negative matrix factorization and difference images. D. Guo et al. [37] used self-similarity and depth information to perform fusion of images.

The block and arbitrary region based fusion technique, unlike pixel averaging techniques, is able to maintain the contrast of scene in fused image. But they also suffer from following limitations

(a) The block size plays important role, like, selection of smaller block size may results in loss of contrast, whereas larger block size may results in selection of both focused as well as unfocused section. In both cases possibility of selection of wrong pixels affects the quality of fused image.
(b) Intensity variations along the edges of blocks creates artefacts on the fused image.
(c) Heavily dependence on region segmentation technique.
(d) Identification of boundaries of regions.

Apart from region based fusion techniques, many authors have proposed feature based fusion techniques. Ishita De et al. [14] used multi-scale Top-Hat transform to identify dark or bright feature on all pixel locations. These feature values are used to create binary image for all captured images. The binary images and captured images are combined together to construct fused image. Similarly X. Bai et al. [33] proposed multi-scale Top-Hat transform technique to extract informative features and perform fusion. B. Yang et al. [25] proposed sparse representation based fusion technique. A. Saha et al. [32] used invariance property of natural image statistics for fusion. The average log amplitude of images which has a natural trend of varying is used to identify pixels with more information and to construct fused image. Y. Jiang et al. [35] performed fusion by using morphological component analysis on captured images. Y. Liu et al. [38] used dense SIFT to calculate descriptors on each pixel. These descriptors are further used to perform fusion. Paul et al. [42] proposed fusion technique in gradient domain.

Region and feature based techniques are effective techniques to perform fusion. However these techniques are not exploring information present on different scales, which is an inherent property of image. To explore information present at multi-scale positions, H. Zhao et al. [29] proposed a parallel pipe framework on captured images to split into multiple neighbour distance images and low frequency image. These images are further processed to synthesize fused image.

2.1.2 Transform domain

In transform domain, fusion techniques are classified into single scale and multi-scale transform. In single scale transform, images are transformed to frequency domain coefficients. The coefficients with more information are separated and combined from respective image transforms. Fused image is constructed by corresponding inverse transform. DCT [11], ICA [13] and PCA [22] are single scale transform techniques useful for fusion of images. These techniques fail to extract information present at different scales.

In multi-scale transform, images are decomposed into multi-scale transform coefficients. Each scale coefficients are processed and informative coefficients are selected. The selected coefficients are converted back to spatial domain by using corresponding inverse transform. Laplacian pyramid [5], gradient pyramid [12], ratio of low pass pyramid [6] and wavelet transform [10] are commonly used multi-scale transform. In these pyramid transform, images are splitted into pyramidal structure by down sampling process. The informative pixels are selected at each layer of pyramid and fused image is constructed by combining layers from pyramid by up-sampling process. The process of up-sampling and down-sampling makes pyramid transform more prone to shift variance. Complex wavelet transform [7] is proposed to address the problem of shift variance of wavelet transform. The multi-scale transform techniques are able to extract more information as compare to single scale transform but still lag in extracting multi-directional information present in image.

Multi-scale geometric (MGA) transform like curvelet [24], shearlet [28] and contourlet [23] transform are proposed to collect information present in multi-direction. However these techniques also suffer from shift-variance. In order to overcome shift variance, few techniques are proposed like non-subsampled shearlet transform (NSST) [30], non-subsampled contourlet transform (NSCT) [40]. These techniques are able to collect information present at different scales as well as in different directions. However the advantage is achieved at the cost of complexity.

Following are the observations for transform techniques

(i) Single scale transform, like DCT, fails to extract information present at different scales.
(ii) Multi-scale transform, like pyramid transform, is able to extract information present at different scales but suffers from shift-variance and lack of directional information.
(iii) The shift-invariant version of multi-scale transform like Complex wavelet transform solves the limitation of shift-variance but still suffers from lack of directional information.

(iv) Multi-scale geometric (MGA) transform like NSCT solves the problem of shift-variance and able to extract directional information. However these techniques have high complexity.

(v) Transform technique perform fusion process globally in frequency domain and any change in coefficient affects all pixels in spatial domain.

(vi) The images transformed to frequency domain for processing and converted back to spatial domain results in loss of data.

Some of the major works in spatial and transform based techniques are explained briefly.

S. Li et al. [8] decomposed pre-registered source images into blocks of pre-defined size. Each block is tested for underlying information using spatial frequency (SF). SF calculates activity level by first calculating pixel intensity difference row and column wise and then square root of sum of their squared values computes frequency information in terms of spatial frequency. Blocks with highest SF transferred from corresponding source images into fused image. If the corresponding block has similar SF, then average of two is transferred to fused image. This fused is an intermediate image which is further passed through majority filter to get the final fused image.

S. Li et al. [9] proposed fusion technique based on artificial neural network. All pre-registered source images are divided into pre-defined size blocks. Three features, spatial frequency, visibility and edge feature, are extracted from each image block. First, few image blocks are used to train the neural network based on extracted features. Later, all image blocks are processed through trained neural network. The clearer block among the co-located image blocks are transferred and combined together to synthesize fused image. The fused image is further processed by majority filter.

Ishita De et al. [14] explains the area based registration process to register the captured source images. For image fusion, morphological mathematics based multi-scale Top-Hat transform is utilized to detect the dark and bright feature on all pixel locations. The scalable structuring element extract features present at different scales. Binary images are created based on feature value of co-located pixels. The binary images are further processed with opening and closing filter to create continuous structure. The source images and their corresponding binary images are combined together on pixel to pixel basis to construct fused image.

W. Huang et al. [16] divide captured pre-registered source images into blocks of pre-defined size. Feature maps are calculated by using energy of image Laplacian (EOL) for each block. Normalized feature maps are fed into Pulse Coupled Neural Network (PCNN) as an external stimulus and corresponding output is calculated. The image block with maximum output value is transferred to fused image. The constructed fused image is further processed with majority filter. The quality of fused image is tested by using objective evaluation method.

J Zhang et al. [17] proposed a block based method for image fusion in which block size is first optimized using Genetic Algorithm (GA). Later, all captured source images are partitioned into blocks using optimized block size. Spatial quality assessment is done for each image block and block with maximum quality assessment value among the co-located blocks is selected. The selected blocks are combined to synthesize fused image. Quality analysis is performed using objective evaluation techniques.

S Li et al. [18] presents the region based fusion method. All pre-registered captured images are averaged together to construct intermediate fused image. The intermediate fused image is segmented into regions using normalized cut method. The activity level of all regions is calculated using spatial frequency and region with maximum spatial frequency among co-located regions are selected for fusion. The selected regions are merged together to construct final fused image. Objective analysis is performed to compare the quality with other methods.

Y. Zhang et al. [20] proposed region based method for multi-focus image fusion. Captured source images are divided into blocks of size 4×4 or 8×8. The co-located image blocks are compared by using blurring measure technique and binary images are constructed. The binary images are converted into fusion maps by removing or filling isolated blocks and results in formation of regions. Source images are combined with fusion maps to construct fused image. Subjective as well as objective evaluation is performed to check the quality of this method.

R. Redondo et al. [21] proposed multi-resolution based multi-focus image fusion. Log-Gabor Wavelet transform is used to decompose captured images into layers at defined scale and orientations. Decision map is defined by using area based activity level across orientations. All layers except low frequency layer are combined by using decision map, whereas low frequency layers are combined as average sum. Finally inverse transform is applied to construct fused image. Log-Gabor multi-resolution is compared with other multi-resolution methods in terms of percentage of error (E).

B. Yang et al. [25] proposed a fusion technique using sparse representation. All captured source images are divided into patches of size $n \times n$ using sliding window technique. The patches are transformed into vectors using lexicographic ordering and establish as single vector. The activity level of all coefficients is calculated and coefficients with maximum activity level among source images are selected and combined to construct final vector. Fused image is constructed by converting final vector into image.

V. Aslantas et al. [26] proposed a block based image fusion method in which block size is optimized using Differential Evolution (DE) algorithm. All captured images are divided into blocks of size $m \times n$. The activity level of each block is calculated and blocks with maximum activity level are combined together to construct intermediate fused image. The global activity level of intermediate fused image is used as fitness function for differential evaluation algorithm to further optimise the block size. The cycle again repeats until optimized solution is reached. The intermediate fused with optimized block size is the final fused image. Objective evaluation is performed to check the quality of fused image.

V. P. S. Naidu [27] proposed multi-resolution scheme to fuse the images. The scheme decomposed the captured pre-registered source images into L level using multi-resolution singular value decomposition. Each decomposed level coefficients with larger absolute value are selected for fusion whereas coarsest level coefficients are combined as an average. These fused L levels of multi-resolution singular value decomposition is converted back to fused image. The quality of scheme is evaluated using objective evaluation indices.

H. Zhao et al. [29] proposed a parallel piped framework for image fusion. The oriented distances of a pixel to a point is utilized to construct the neighbour distance filter. Neighbour distance filter is then utilized in a parallel piped framework on source images to split into multiple neighbour distance images and low frequency image. These images, from all source images, are tested on pixel basis using maximum activity level. The selected pixels at each level are combined to construct fused image. Consistency verification is performed by using small majority filter. Objective evaluation is performed to evaluate the performance.

G. Guorong at al. [30] proposed non-subsampled shearlet transform based image fusion method. Captured and pre-registered source images are splitted into low-pass sub-band and band-pass sub-band coefficient images. The low-pass and band-pass coefficient images are combined on the basis of ‘average’ and ‘absolute maximum selection’ respectively, and corresponding initial fused image is constructed by using inverse shearlet transform. Focus regions are identified on comparing each source image with initial fused image by using square error. These regions are used to perform region based fusion on low-pass and band-pass coefficient images. Objective evaluation is performed to assess the quality of fused image.

I. De et al. [31] proposed quad tree structure based image fusion method. Initially source images (level 0) are divided into four quadrants (level 1) and corresponding focus measure is calculated. Normalize difference in focus measure (NDFM) decides further division of each quadrant into four quadrants. If NDFM is less then threshold, corresponding blocks are further divided into four blocks otherwise block with largest focus measure is selected and become part of final image. The process of division into quadrants continues till the size of quadrants either become too small or NDFM does not permit it. Objective evaluation is performed to evaluate the performance.

A. Saha et al. [32] proposed a feature based image fusion method. The method utilized scale invariance property of natural image statistics, which says that the average log amplitude of ensemble of images shows a natural trend of varying. The log amplitude spectrum of each image is calculated and then convolves with average filter to generate the average log amplitude spectrum. The feature, saliency map, of each image is calculated based on mutual spectral residual. Based on maximum value of saliency at a particular pixel, final fusion will be performed. Objective evaluation is performed to assess the quality of fused image.

X. Bai et al. [33] used the multi-scale top-hat transform to develop weighted image fusion algorithm. Multi-scale structuring elements of same shape and increase in size are used to extract useful bright and dim features at each pixel location. The extracted bright and dim features of all scales are combined together. The weighted bright and dim features are combined with base image to construct fused image. The weights are calculated by using mean, deviation and entropy around each pixel. These three strategies along with other state-of-the-art techniques are compared using visual as well as qualitative analysis.

Y. Zhang et al. [34] proposed fusion of multi-focused images by using non-negative matrix factorization and difference images. All captured and pre-registered images are fused together to construct temporary fused image by using non-negative matrix factorization. The difference images based on difference between captured images and temporary fused image are constructed. The focused regions are detected by salient features of difference images. The detected regions and source images are combined to construct fused image. Qualitative analysis is performed to evaluate the performance of proposed method.

Y. Jiang et al. [35] proposed fusion technique by using morphological component analysis. Captured source images are separated into cartoon and texture components as sparse structure. The texture and cartoon components from respective source image are integrated to construct combined sparse texture and cartoon components. These components are transformed back to construct fused texture and cartoon images, which later combined to construct fused image. Objective evaluation is performed to assess the quality of fused image.

Y Liu et al. [36] proposed image fusion framework based on multi-scale transform and sparse representation. All captured and pre-registered source images are splitted into low pass and high pass coefficients using multi-scale transform. Sparse representation technique is used to combine low-pass coefficients, whereas high pass coefficients are combined based on maximum absolute value. These combined coefficients are converted back to fused image using inverse multi-scale transform. Six available multi-scale transform are used to test on different image sets. Subjective as well as objective evaluation is performed to test the quality.

D. Guo et al. [37] proposed fusion method using self-similarity and depth information. Shared similarity regions are identified for all pixel locations from pre-registered source images. These regions are tested for clarity using gradient and depth information. All similarity regions are compared for source images using calculated clarity measure. Votes are granted to a pixel based on clarity measure for all shared regions. The source images are combined together on pixel-to-pixel basis to construct fused image. Qualitative analysis is performed to evaluate the performance of proposed method.

Y. Liu et al. [38] used dense SIFT to fuse multi-focused images. All pre-registered source images are processed with dense SIFT to calculate corresponding dense SIFT image.

SIFT descriptors are combined at each pixel to construct activity level map for all SIFT images. Initial decision map of source images is constructed using SIFT descriptors to define definite focused, definite un-focused and indefinite regions. The pixels in indefinite regions are addressed further to create final decision map. Fused image is constructed by combining decision map and source images. Qualitative analysis is performed to evaluate the performance of proposed method.

X. Bai et al. [39] proposed fusion of pre-registered source images using a quad-tree structure. Source images are decomposed into blocks of optimum size in a quad-tree structure using MDFM and SMDG, and identifying focused regions. A new focused measure, sum of weighted modified Laplacian, is defined and used in decomposing images. Detected focused regions are selected form respective image and combined to construct fused image. Qualitative analysis is performed to compare the results with previous work.

Y. Yang et al. [40] proposed a novel image fusion frame work for multi-focus images, which relies on NSCT domain and focused area detection. The source image are divided into low and high frequency coefficients using Non-subsampled contourlet transform. An initial fused image is obtained using visual contrast rule and log-gabor energy rule for low and high frequency coefficients. The focused areas in source images are selected by RMSE between initial fused image and source images, which is further utilized to perform final fusion. Qualitative analysis is performed to evaluate the performance of proposed method.

L Cao. et al. [41] proposed algorithm based on spatial frequency in Discrete Cosine Transform (DCT) domain. The method proposed to use spatial frequency in DCT domain. The source images are divided into blocks and transformed into DCT domain. Spatial frequency is calculated for each block. The block of DCT coefficients with maximum spatial frequency is selected and combined together. The selected DCT coefficient are converted back to construct fused image. The selected DCT coefficients can further be processed for JPEG.

Paul et al. [42] proposed a method to process luminance of the source images in gradient domain. All captured and pre-registered source images are converted into gradient domain. The pixels with maximum gradient are transferred from respective image to construct gradient fused image. The fused gradient image is processed by Haar wavelet to reconstruct luminance fused image. Chrominance is processed separately and combined with luminance to construct the final image. Qualitative analysis is performed to evaluate the performance of proposed method.

H. Tang et al. [43] proposed learning based fusion technique. Pixel-wise convolutional neural network (p-CNN) recognize focused and un-focused pixel by using neighborhood information. Focused pixels are combined to construct fused image. The network is trained by comprehensive public image database. Results are compared with state-of-the-art techniques by using subjective as well as objective evaluation techniques.

G. Qi et al. [44] fused multi-focus images by using geometric information based dictionary and sparse representation. The source images are analyzed for geometric information and used to build sub-dictionary by principle component analysis. These sub-dictionaries are further merged to build combined dictionary. The comprehensive sampling matched pursuit algorithm extract sparse coefficients from respective source images to construct fused image.

M. S. Farid et al. [45] proposed region based fusion technique by using Content Blurring Algorithm (CAB). CAB induces blurring to pixel based on underlying information. The focused pixel received significant blur whereas unfocused pixel little or no blur. The difference in original and CAB blurred image segment the images into regions which are further used for fusion. Qualitative and Quantitative evaluation is performed to evaluate the quality.

B. Xiao et al. [46] proposed fusion technique based on Hessian matrix based decomposition. Multi-scale Hessian matrix decompose each source image into feature and background regions. The two regions based on two focus measures are used to construct initial decision maps respectively. Final decision map based on initial decision maps are used to combine source images into fused image. Results are verified by subjective as well as objective evaluation.

S. Aymaz et al. [47] proposed fusion technique based on super-resolution and stationary wavelet transform (SWT). All source images are processed with super-resolution method to enhance information. These processed images are splitted into sub-bands using SWT. Principle Component Analysis (PCA) is used to select most informative sub-band among co-located sub-bands. The selected sub-bands are combined to construct fused image by Inverse SWT. Quality testing is performed to evaluate the proposed method.

2.2 Multi-Exposure Image Fusion

In multi-exposure image fusion, co-registered images captured at different shutter speed are combined to construct fused image. Like multi-focus image fusion, multi-exposure fusion techniques are also proposed in spatial as well as transform domain.

2.2.1 Spatial domain

In spatial domain efforts were made to develop fusion techniques by using regions or features of captured images. Region of an image may be pre-defined block or arbitrary shape region. A. A. Goshtasby et al. [48] were among the first to proposed block based fusion technique by dividing each image into predefined blocks, and best blocks based on the entropy are selected for fusion. Similarly, A. R. Vaekonyi-Koczy et al. [51] divided three channels of colour image into non-overlapping blocks. The blocks with maximum gradient are selected and combined together for fusion. K. H. Jo et al. [55] proposed region based fusion in which images are divided into arbitrary shape regions using intensity levels. These regions are combined with source images as weighted average to construct fused image. X. Li et al. [63] also proposed region based fusion by using K-mean clustering. These techniques suffered from various limitations as with multi-focus image fusion techniques (refer section 2.1.1).

The limitation of region based techniques directed research community towards feature based fusion. Many fusion techniques have been proposed by using extracted features from captured images. T. Mertens [50] proposed classical exposure fusion technique by extracting three features; contrast, saturation and well-exposedness. These features are used to construct weights and combined with source images to synthesize fused image. Raman et al. [52] proposed fusion technique by using matting function. Matting function is calculated by difference between source image and its corresponding bilateral filter processed image. K. Kotwal et al. [54] proposed another technique based on matte function. S. Li et al. [57] extracted three features local contrast, brightness and colour dissimilarity by using median and recursive filter. These features are used to construct fused image as weighted sum. T. Jinno et al. [58] proposed irradiance of image to perform fusion. W. Zhang et al. [59] used two features visibility and consistency to perform exposure fusion. Y. Liu et al. [66] used three features; contrast, exposure quality and spatial consistency using dense SIFT to perform exposure fusion. K. Ma et al. [68] used three components: signal strength, signal structure and mean intensity to synthesize fused image. Apart from feature based technique some probabilistic technique are also proposed; R. M. Song et al. [60] proposed probabilistic fusion technique. Two features, visible contrast and scene gradient are extracted from images and is utilized to get the synthesised image as maximum a post priori framework. R. Shen et al. [56] proposed fusion technique based on Generalized Random Walks. Shen et al. [61] proposed Gaussian conditional random field model based exposure fusion.

2.2.2 Transform domain

In transform domain, I. Zafar et al. [49] proposed DCT transform based fusion. V. Radhika et al. [69] proposed block based fusion using various transform techniques. The multi-scale techniques are also proposed for multi-exposure fusion in order to extract more information. Wang et al. [53] used dual tree directional filter bank to use multi-scale positions in order to construct fused image. J. Shen et al. [62] used boosting Laplacian pyramid to construct fused image. A. Gladran [71] and Q. Wang et al. [73] used multi-scale transform to perform multi-scale exposure fusion.

Some of the major works in spatial and transform based techniques are explained briefly.

A. A. Goshtasby et al. [48] used the optimum block size and blending function to perform multi exposure fusion. The source images are divided into blocks of defined size and entropy is calculated for each block. The co-located blocks are tested for maximum entropy and combined together using monotonically decreasing blending function to synthesize fused image.

I. Zafar et al. [49] present the first ever method to address the problem of multi exposure fusion in transform domain. The proposed algorithm divide the source images into non-overlapping blocks of suitable size. Each corresponding block is transformed into DCT domain. Based on maximum L2 norm of AC coefficients, corresponding block is selected. Similar process is applied to all the blocks and fused image is synthesize by inverse DCT.

T. Mertens [50] proposed a fusion technique using multi-resolution pyramidal structure of pre-registered source images. Three quality measure, contrast, saturation and well-exposedness, are used to extract information at each pixel level. These are further combined to construct weight map for each image. All source images are decompose into Laplacian pyramidal structure and weight map into Gaussian pyramidal structure. These pyramidal structure are multiplied at corresponding scale. The fused image is constructed by reverse pyramidal structure.

A. R. Vaekonyi-Koczy et al. [51] develop block based method to fuse multi-exposure images. The colour source images are divided into blocks of fix size. Gradient of each block is calculated for all three channels separately. The block with maximum sum of gradient of all the three channels is selected and become a part of fused image. Blocking effects are visible by combining blocks directly and degrade the quality of fused image. To reduce this effect authors have used 2-D Gaussian function to blend the selected block from source images.

Raman et al. [52] proposed pixel level fusion process using matting function. Matting function is calculated on each pixel by using difference between source image and its corresponding bilateral filter processed image. Matting function gives more weightage to pixel with more information and less to other co-located pixels. Finally fused image is calculated by combining matting function and source images as weighted sum.

Wang et al. [53] proposed pyramidal dual tree directional filter bank based multi-scale fusion method. All captured and pre-registered source images are decomposed into high pass and low frequency coefficients. High pass coefficients with high saliency measure are selected for fused image. Saliency measure is calculated as local energy around a pixel. Low frequency coefficients are combined as average sum. These selected high pass and low frequency coefficients are combined and converted back to construct fused image. Proposed method is evaluated for quality using objective evaluation indices.

K. Kotwal et al. [54] proposed optimization based approach to perform fusion. Three objectives based on entropy, contrast and spatial correlation are defined to calculate matte function for each pixel and optimized using Euler-Lagrange technique. Qualitative analysis is performed to evaluate the performance of proposed method.

K. H. Jo et al. [55] presented the clustered based method for multi-exposure images. Initially source images are decomposed into regions based on intensity levels. Utility function is calculated for each region based on entropy and level of details. Binary masks are constructed for each source image by testing utility function on pixel to pixel basis. These binary masks are processed using bilateral filter, in order to smooth the transition along the boundaries, to construct corresponding weighted masks. The fused image is constructed as weighted sum of source images and weighted masks.

S. Li et al. [57] addresses the multi exposure fusion using median and recursive filter. The three features local contrast, brightness and colour dissimilarity are calculated and combined to construct weight for all source images. These weights are further refined using recursive filter to smooth the transitions. Fused image is calculated as weighted sum of source images. Qualitative analysis is performed to evaluate the performance of proposed method.

T. Jinno et al. [58] proposed fusion technique using irradiance of image. The irradiance is calculated for each pixel of captured image and corresponding weights are defined. The captured images are combined together as weighted sum to construct fused image. The method also addressed the motion blur in captured images. Qualitative analysis is performed to evaluate the performance of proposed method.

W. Zhang et al. [59] proposed gradient directed exposure fusion. Two quality measures; visibility and consistency are extracted from captured images. Weight map based on quality measures are defined and combined with captured images as seamless composition to construct fused image.

M. Song et al. [60] proposed a new scheme based on probabilistic model. The scheme is divided into scene detail extraction and SDR image synthesis. Visible contrast and scene gradient are extracted from the series of images with different exposure levels. Maximum a post priori framework is utilized to get the synthesised image based on visible contrast and scene gradient. For colour image RGB is converted into YUV and all the operations are performed on Y i.e. luminance and at the end converted back to RGB.

J. Shen et al. [62] used new exposure fusion approach based on boosting Laplacian pyramid. The source images are decomposed into base layer and detailed layers which are boosted separately. Weight maps for each pixel is calculated by combining local exposure weight, global exposure weight and saliency weight. The boosting process is guided by local exposure weight and saliency weight. Fusion is performed by combining weight maps and boosted layers. Qualitative and quantitative evaluation is performed to test the proposed method.

X. Li et al. [63] proposed layered based algorithm with global and detailed layers are constructed from source images. Global layer is constructed using bilateral filter of average of source images. Global layer is divided into regions using K-mean clustering. The entropy and gradients are used to select the best region among all the exposures. These regions along with detailed layer is used to perform the fusion at gradient level and fused image is reconstructed. Qualitative analysis is performed to evaluate the performance of proposed method.

A. V. Vanmali et al. [65] addresses the problem of ghost effects along with exposure fusion. The process is done in four steps viz. weight map calculation, new-object detection, modified weight map and exposure fusion. The weight map is calculated on the basis of contrast, well-exposedness and saturation for each pixel. For object detection, gamma correction, bilateral filtering and absolute difference between reference image and all images are combined to create binary images. These binary images are used to modify the weights. The fusion is performed using multi-scale structure.

Y. Liu et al. [66] proposed three feature based fusion technique. Three features; contrast, exposure quality and spatial consistency are extracted from captured and pre-registered source images. Contrast is calculated based on dense SIFT and used two strategies; weighted average and winners take all in weight contribution. Weights are calculated by combining these three features. Finally fused image is constructed by combining captured images as weighted sum. Qualitative analysis is performed to evaluate the performance of proposed method.

K. Ma et al. [68] proposed structural patch decomposition approach for multi-exposure image fusion. Each image patch is decomposed into three components: signal strength, signal structure and mean intensity. Each component is fused separately and recombine to construct patch and further fused image. The proposed method is further compared with state-of-the-art techniques.

V. Radhika et al. [69] proposed block based technique for exposure fusion. The captured and pre-registered source images are divided into pre-defined blocks. All corresponding blocks are processed with transform technique and coefficients are calculated. The block coefficient with high variance are selected and converted back to fused image by inverse transform. Existing transform techniques are used for fusion. Objective evaluation is performed to test the quality.

A. Gladran [71] proposed multi-exposure image fusion for image dehazing. An image is captured under haze condition and artificially transformed into varying exposure images by using gamma corrections. Weights for all images are calculated by using absolute contrast of pixel. The source images along with weights are converted into Laplacian pyramid structure. Fusion is done by combining each level of source and weights and fused image is constructed by inverse Laplacian transform. Subjective and objective evaluation is performed to evaluate the quality of fused image.

Q. Wang et al. [73] proposed multi-scale pyramid based multi-exposure image fusion. The fusion is performed as multi-scale weighted sum. The weights are calculated for each image by using three parameters; contrast, saturation and degree of exposure. All captured source images are splitted into Laplacian pyramid and weights as Gaussian pyramid. Fusion is performed at each level and fused image is constructed by inverse pyramid structure. Qualitative assessment is performed to evaluate the quality.

2.3 Challenges in Image Fusion

Broadly, there are two ways to perform image fusion; spatial domain and transform domain. The challenges for each category is as under

- Spatial domain fusion suffers from blocking effects which deteriorate fused image quality.
- Transform domain fusion suffers from data loss during transformation of image.

In present work, we proposed three spatial domain techniques and to resolve the problem of blocking effects, all three techniques works on layer based fusion. Each image is splitted into multiple layers and later processed layers are combined to synthesize fused image.

Chapter – 3

Quality Indices (Image Fidelity)

An image fusion is a process of synthesis of extracted information from captured source images into fused image. The fusion technique should retain underlying information of source images and does not degrade the visible quality of an image. The evaluation becomes necessary to evaluate fusion technique to test the quality. The image fidelity criteria are useful in accessing and evaluating the quality of fused image.

Image fidelity criteria are of two types to evaluate the performance: subjective and objective. The subjective criteria are based on visual inspection by group of people familiar with images, whereas in objective criteria well defined mathematical expressions are used.

3.1 Subjective evaluation

The subjective evaluation is performed by using rating scales; goodness scale and impairment scale [2]. Goodness scale is divided into two category; overall goodness scale and group goodness scale. Overall goodness scale is the global scale over which an image quality is rated ranging from ‘Excellent’ to ‘Unsatisfactory’, refer Table 3.1. Group goodness scale is second category of global scale which compares a set of images and rate them ranging from ‘Best’ to ‘Worst’, refer Table 3.1. The impairment scale test the quality decline of image in comparison with reference image. The rating is given ranging from ‘Not-noticeable’ to ‘Extremely objectionable’, refer Table 3.2. In subjective evaluation, number of trained observer are invited to rate the image based on goodness scale and mean value [2] is given as

$$R = \frac{\sum_{k=1}^{n} s_k n_k}{\sum_{k=1}^{n} n_k} \qquad (1)$$

where s_k is the score associated with the k^{th} rating, n_k is the number of observers with this rating and n is the number of grades in the scale.

3.2 Objective evaluation

Objective evaluation techniques are based on well-defined mathematical expressions. It is broadly divided into two categories, reference image and non-reference image based evaluation. In reference image based evaluation, an image (under evaluation) is compared with reference image to test the quality. Root Mean Square Error (RMSE) is a good example of reference image based evaluation. The non-reference based evaluation; an image

Table 3.1
Image Goodness Scales

Overall goodness scale		Group goodness scale	
Excellent	(5)	Best	(7)
Good	(4)	Well above average	(6)
Fair	(3)	Slightly above average	(5)
Poor	(2)	Average	(4)
Unsatisfactory	(1)	Slightly below average	(3)
		Well below average	(2)
		Worst	(1)

Courtesy: [2]

Table 3.2
Impairment Scales

Not noticeable	(1)
Just noticeable	(2)
Definitely noticeable	(3)
Impairment not objectionable	(4)
Somewhat objectionable	(5)
Definitely objectionable	(6)
Extremely objectionable	(7)

Courtesy: [2]

(under evaluation) is evaluated independently, without referencing reference image. Mutual Information (MI), Edge based Similarity Measure ($Q^{\frac{AB}{F}}$) are the good examples.

Different quality indices for reference and non-reference based evaluation are discussed below:

3.2.1 Reference based evaluation indices

Quality indices based on reference image are as follows

3.2.1.1 Mean Square (MSE) and Root Mean Square Error (RMSE)

Mean square error is simple and commonly used index. The squared difference between reference image and image under evaluation (fused image) on pixel-to-pixel basis is calculated and termed as error. The calculated error is averaged by total number of pixels. The expression for MSE is as follows

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_R(i,j) - I_F(i,j))^2 \tag{2}$$

where I_R, I_F are reference and fused image respectively, $M \times N$ is image size and (i,j) represents pixel location.

The more is the similarity between reference and fused image, lesser is the error and MSE value. Thus smaller MSE value represents better quality.

Root mean square error is a square root of mean square error. The expression for RMSE is as follows

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_R(i,j) - I_F(i,j))^2} \tag{3}$$

3.2.1.2 Peak Signal to Noise Ratio (PSNR)

It is a ratio of peak power of an image to power of noise. Power of an image is calculated as square of maximum intensity level present in an image whereas noise power is calculated as mean square error (MSE), refer Eq. (2). For 8-bit image, PSNR between reference image (I_R) and image under evaluation (I_F) is calculated as

$$PSNR = 10log\left[\frac{255^2}{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_R(i,j) - I_F(i,j))^2}\right] = 10log\left[\frac{255^2}{MSE}\right] \tag{4}$$

where I_R, I_F are reference and fused image respectively, 255 represents maximum value present for 8-bit image, $M \times N$ is image size and (i,j) represent pixel location.

An increase in similarity between I_R and I_F results in decrease in MSE and consequently increases PSNR. Therefore larger the value of PSNR, better is the quality of fused image $I_{F.}$

3.2.1.3 Standard Deviation

It measures the dispersion in intensity of pixels around the mean intensity of an image. Higher value of dispersion represents more information in image and vice-versa. The expression for image under evaluation I_F is as follows

$$\sigma_F = \sqrt{\frac{1}{M.N}\sum_{i=1}^{M}\sum_{j=1}^{N}(I_F(i,j) - \mu_F)^2} \tag{5}$$

Similarly σ_R is calculated for reference image I_R.

where I_R, I_F are reference and fused image respectively, μ_F is mean intensity of an image under evaluation.

The similarity between I_F and I_R is depend on the closeness between their deviations. More is closeness better is the similarity.

3.2.2 Non - reference based evaluation indices

Quality indices without referring reference image are as follows

3.2.2.1 Mutual Information (MI)

Mutual Information [74] is based on information theory to measure the statistical dependence between two random variables. It measures the information shared by one variable about other variable.

For two random variables P and Q, Mutual Information is calculated as

$$I_{XY}(x,y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p_{XY}(x,y)\log(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}) \tag{6}$$

where $p_{XY}(x,y)$ is the joint probability distribution and $p_X(x)$, $p_Y(y)$ are the marginal probability distributions.
For images, marginal and joint probability distributions are calculated by normalisation of corresponding marginal and joint histogram. For 2 input images X_1, X_2, $I_{X_1X_2}$ represents mutual information and given by

$$I_{X_1X_2}(X_1, X_2) = \sum_{X_1}\sum_{X_2} p_{X_1X_2}(X_1, X_2)\, log\frac{p_{X_1X_2}(X_1, X_2)}{p_{X_1}(X_1)p_{X_2}(X_2)} \tag{7}$$

For N input images $X_1, \dots, X_N$, Mutual Information between set of images X_i $(i\ \epsilon\ [1, N])$ and fused image (F) is calculated as

$$I_{FX_i}(X_1, \dots, X_N, F) = I_{FX_1}(X_1, F) + \dots + I_{FX_N}(X_N, F) \tag{8}$$

Larger value of MI represents better quality of image.

3.2.2.2 Normalised Mutual Information (NMI)

Normalised Mutual Information, normalise the Mutual Information calculated as per equation (8). In [75], it was point out that calculation of Mutual Information is biased towards image with highest entropy. The solution suggested is to normalise Mutual Information by corresponding entropy value. The modified expression for N input images $X_1, \dots, X_N$, is as follows

$$I_{NMI}(X_1, \dots, X_N, F) = 2\left[\frac{I_{FX_1}(X_1,F)}{H(X_1)+H(F)} + \dots + \frac{I_{FX_N}(X_N,F)}{H(X_N)+H(F)}\right] \tag{9}$$

where, I_{FX_i} represents mutual information (refer (6)), $H(X_i)$ and $H(F)$ are marginal entropies. Higher value of NMI represents better quality.

3.2.2.3 Edge-based Similarity Index ($Q^{AB/F}$)

It evaluates the amount of information in terms of edges transferred from source images to fused images [76]. Sobel edge operator is used to calculate relative edge strength and orientation of each source image with respect to fused image. The calculated strength, orientation are used to derive corresponding preservation value and combined together to represent the perceptual loss of information in fused image. For N input images $X_1, \dots, X_N$, the relative strength of each source image X_i ($i \in [1, N]$) in fused image (F) is

$$Q^{X_iF}(m,n) = Q_g^{X_iF}(m,n).\, Q_\alpha^{X_iF}(m,n) \tag{10}$$

where $Q_g^{X_iF}, Q_\alpha^{X_iF}$ are the edge strength and orientation preservation value of source image X_i.

These are combined together as normalised weighted sum, represented as follows

$$Q^{AB/F} = \frac{\sum_{m=1}^{P}\sum_{n=1}^{Q}[Q^{X_1F}(m,n)\, w^{X_1}(m,n) + \dots + Q^{X_NF}(m,n)\, w^{X_N}(m,n)]}{\sum_{m=1}^{P}\sum_{n=1}^{Q}[w^{X_1}(m,n) + \dots + w^{X_N}(m,n)]} \tag{11}$$

The range for $Q^{AB/F}$ is in between (0, 1), where '0' represents complete loss of edges and '1' represents complete transfer of edges.

3.2.2.4 Structure similarity-based metric (SSIM)

It reveals the degree of similarity between the two images, covering three aspects luminance, contrast and structure. Luminance is calculated as mean intensity of an image. For source image X_i and fused image F, μ_{X_i} and μ_F represents luminance. To extract contrast information from images, standard deviation is used. σ_{X_i} and σ_F represents the deviation for source image and fused image respectively. For structure information, normalised signals are used, e.g. for source image X_i, normalised signal is $\frac{X_i - \mu_{X_i}}{\sigma_{Xi}}$. These three factors are combined to represents the similarity measure [77]. The expression for N input images $X_1, \dots, X_N$ is as follows

$$SSIM(X_1, F) = \left(\frac{2\mu_{X1}\mu_F + c1}{\mu_{X1}^2 + \mu_F^2 + c1}\right)^{\alpha} \cdot \left(\frac{2\sigma_{X1}\sigma_F + c2}{\sigma_{X1}^2 + \sigma_F^2 + c2}\right)^{\beta} \cdot \left(\frac{\sigma_{X1F} + c3}{\sigma_{X1}\sigma_F + c3}\right)^{\gamma} \tag{12}$$

$$SSIM\ (X_1, \dots, X_N, F) = \frac{1}{N}(SSIM(X_1, F) + \cdots .. + SSIM(X_N, F)) \tag{13}$$

where μ_{X_i}, μ_F represents mean, σ_{X_i}, σ_F represents standard deviation, and σ_{X_iF} represents covariance of X_i and F. $c1, c2, c3$ are constants. α, β, γ control the contribution.
Higher value of SSIM represents better quality.

3.2.2.5 Entropy

It measure the information content of an image by using probability of grey levels. The histogram of image gives the probability values of all grey levels. The expression for entropy of 8-bit image is as follows

$$E = -\sum_{j=0}^{255} p(j) log_2(p(j)) \tag{15}$$

where $p(j)$ represents the probability of j^{th} grey level.
The low entropy means less variations in grey levels or represents flat image, similarly high entropy represents high variations in grey level and indicates high information.

3.2.2.6 Average Gradient

It measure the average gradient value of image. The gradient of pixel is the difference in intensity with neighbour pixels in horizontal and vertical direction. The higher gradient of pixel represents more information. The expression for image under evaluation I_F is as follows

$$AG = \frac{1}{M.N} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{(I_F(i,j) - I_F(i-1,j))^2 + ((I_F(i,j) - I_F(i,j-1))^2}{2}} \quad (16)$$

where $M.N$ represents the number of pixels and (i, j) represents the pixel location.

Higher value of average gradient represents better quality.

3.2.2.7 Visual Information Fidelity for Fusion (VIFF)

It is Visual Information Fidelity (VIF) based quality index [78]. The images are divided into sub-bands by using wavelet transform. Each sub-band is evaluated by VIF with and without distortion. VIFF of respective sub-band is calculated as ratio of VIF with and without distortion. Final VIFF is calculated as weighted sum VIFF of all sub-bands [78]. VIFF for k^{th} sub-band and final expression are as follows

$$VIFF_k(X_1, \ldots, X_N, F) = \frac{\sum FVID\ (X_1, \ldots, X_N, F)}{\sum FVIND\ (X_1, \ldots, X_N, F)} \quad (17)$$

$$VIFF(X_1, \ldots, X_N, F) = \sum_{p_k} p_k . VIFF_k(X_1, \ldots, X_N, F) \quad (18)$$

where p_k represents the weight for k^{th} sub-band.

3.2.2.8 Standard Deviation (Non-reference)

Standard deviation can also be used as non-reference based evaluation index. It measures the dispersion in intensity of a pixels around the mean intensity of an image, refer Eq. (5). Higher value of standard deviation represents more information of image and vice-versa.

Chapter – 4

Block based fusion

4.1 Introduction

The present chapter discusses a novel block based image fusion algorithm. The algorithm processes pre-registered source images to generate fused image primarily to enhance depth of field (DoF) and further extended to increase dynamic range. The source images are processed with block based fusion to synthesize fused image.

The block based techniques usually faces challenges like selection of block size, artifacts around edges of block. These factors also degrades the quality qualitatively as well as visually. To overcome from this limitation we have used layer based fusion in the proposed method. Each source image is splitted into two layers; local and global. Local layer holds the high frequency information whereas global layer holds the low frequency information. Local layer is processed with block based fusion and combined with fused global layer. The addition of block based processed local layer with global layer improves visual quality. The qualitative value depends on the frequency information separated from captured images, which is also addressed in the proposed technique.

4.2 Local and Global Layer

Every image holds two type of information: the high frequency components (like strong edges), and low frequency components (like weak edges or texture). The high frequency information is separated from each image and termed as local layer whereas remaining low frequency information is termed as global layer.

4.2.1 Neighbor Distance Filter

The neighbor distance (ND) filter [29] is based on oriented distance between a point and its neighbor, refer Fig (4.1) [101]. The oriented distance, as per differential geometry, between a point and its neighbor measures the bending of surface in a given direction. Fig. (4.1) shows the oriented distance between (x, y) and $(x + d, y + d)$. The curvature between 'foot of perpendicular' and $(x + d, y + d)$ represents the oriented distance. In image, the two points are replaced by two neighbor pixels and oriented distance then calculates difference in intensity between two neighbour pixels.

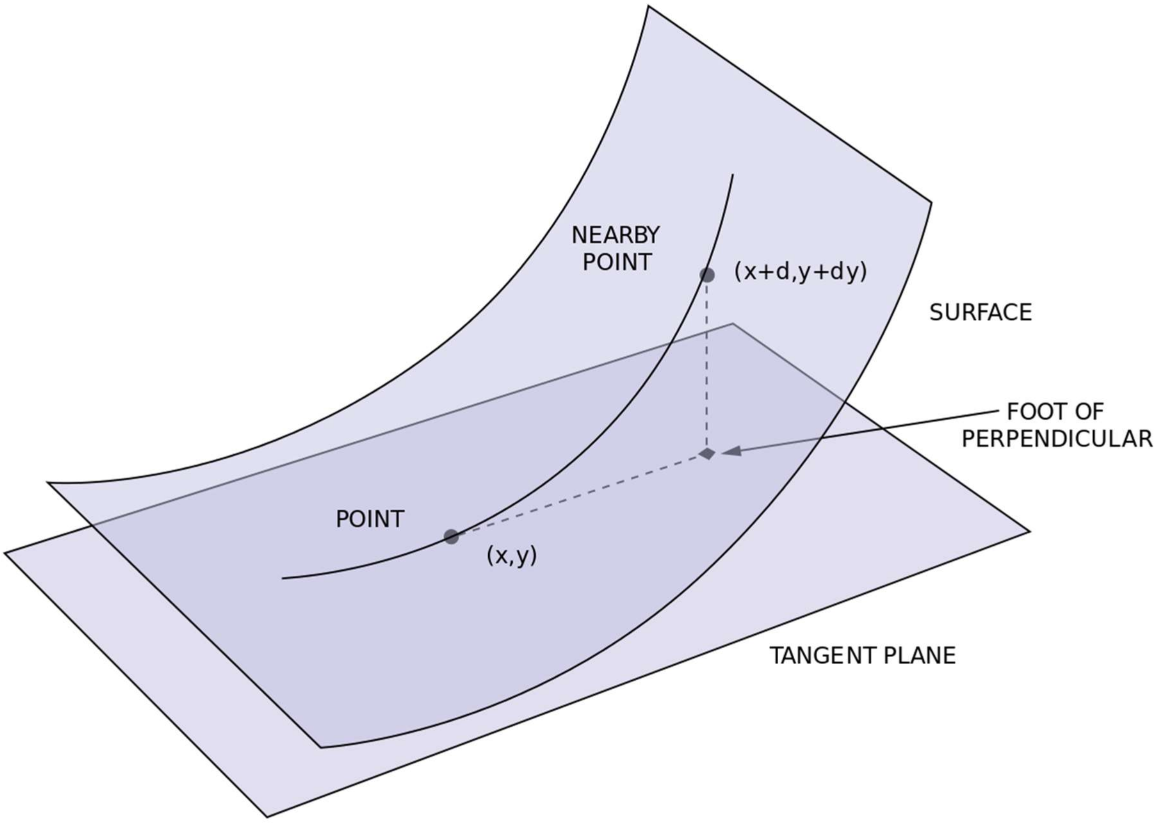


Fig. 4.1 Oriented distance between (x, y) and $(x + d, y + d)$.

A pixel of an image is surrounded by eight pixels. The oriented distance of target pixel with neighbor pixels gives the information in eight directions. The summation of eight oriented distances around target pixel is defined as Neighbor distance (ND), refer Eq. (19). The higher value of neighbor distance represents higher information around pixel.

$$ND(u, v) = \sum_{k=1}^{8} OD\big((u_0, v_0), (u_k, v_k)\big) \quad (19)$$

where 'OD' represents oriented distance and 'ND' represents Neighbour distance.

The calculated neighbor distance is further processed to construct neighbor distance (ND) filter by defining block size $R \times C$ and bandwidth h. The kernel [29] with block size 5×5 and bandwidth h = 0.4 is as follows

$$ND = \begin{bmatrix} -0.0004 & -0.0103 & -0.0274 & -0.0103 & -0.0004 \\ -0.0103 & -0.1256 & -0.0760 & -0.1256 & -0.0103 \\ -0.0274 & -0.0760 & 1.0000 & -0.0760 & -0.0274 \\ -0.0103 & -0.1256 & -0.0760 & -0.1256 & -0.0103 \\ -0.0004 & -0.0103 & -0.0274 & -0.0103 & -0.0004 \end{bmatrix} \quad (20)$$

We use the same kernels to extract high frequency components.

4.2.2 Frequency Components Extraction and Layers Construction

Let ϕ_i be the set of pre-registered N source images of size $P \times Q$, where $i \in 1 \dots . N$. Each source image is processes with ND filter (refer Eq. (20)) to split it into high frequency components (image) and remaining low frequency component (image). The separation is done on multi-scale basis by zero padding filter kernels. The zero layer padding is equivalent to single zero around the neighbor of previous zero padded kernels.

The multi-scale analysis is performed by varying the kernel size instead of varying image size. This helps to keep it less prone to shift-variant. The zero padding layers around the kernels is proportional to number of high frequency images, refer Eq. (21).

For any source image ϕ_i, multi-scale extraction of high frequency component images ($\boldsymbol{H_{ij}}$) and remaining low frequency component image (L_i) is as follows

$$\phi_i = L_i + [\boldsymbol{I}] \times [\boldsymbol{H_{ij}}]^T \, ; i \in [1,2, \dots , N] \;\&\; j \in [1,2, \dots . , M] \tag{21}$$

and $\boldsymbol{H_{ij}} = [H_{i1}\ H_{i2}\ H_{i3} \dots H_{ij} \dots H_{iM}]$

The pseudo code to generate high frequency images is as follows

$ND_1 = ND, Im = \phi_i$

$For\ j = 1{:}\,M$

$\quad H_{ij} = Im \circledast ND_j$

$\quad Im = Im - H_{ij}$

End (22)

where i, j represents i^{th} source image and j^{th} frequency image, ND_j is j^{th} zero padded kernel, '$\circledast$' is convolution operator and ND_1 is the original filter kernels.

The extracted M high frequency images from source image ϕ_i are combined together and represented as Local layer (I_{hfi}), refer Eq. (23). The remaining low frequency components are represented as Global layer (L_i), refer Eq. (24). Fig. (4.2), shows the splitting of source image into high frequency images and remaining low frequency image. Fig. (4.3), shows the combining of high frequency images to construct local layer.

$$I_{hfi} = \sum_{j=1}^{M} H_{ij} \, ; i \in [1,2 \dots . N] \tag{23}$$

$$L_i = \phi_i - I_{hfi} \tag{24}$$

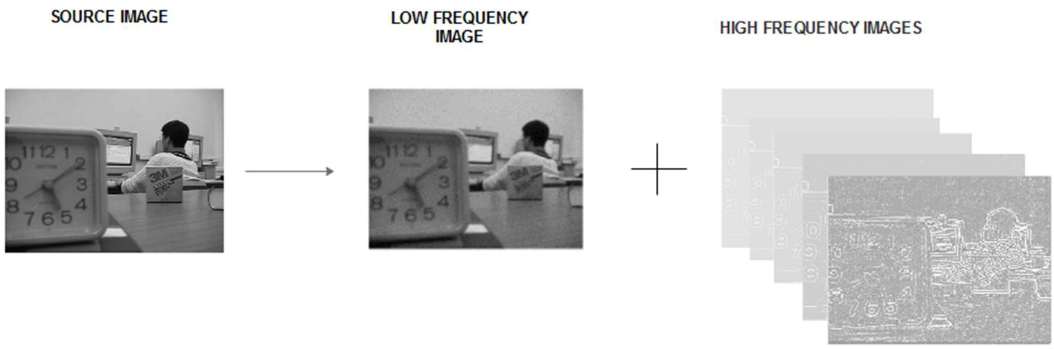


Fig. 4.2 Splitting of source image into low and high frequency images

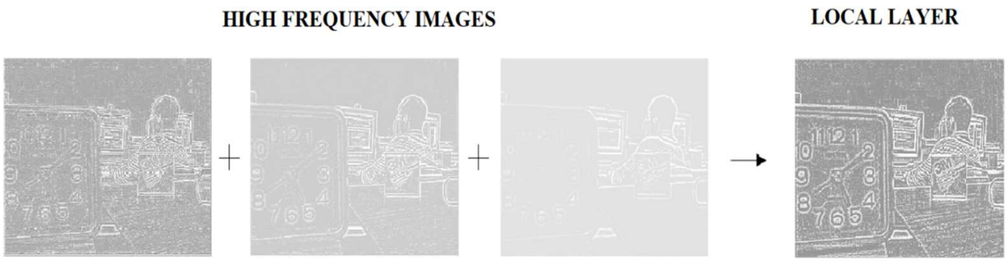


Fig. 4.3 High frequency images to local layer

4.3 Fusion of Local Layer

Local layers (I_{hfi}) holds the high frequency information of the scene through differently captured images. These images holds informative sections in different focused/exposed sections. The fusion of local layers combine these informative sections together into one image.

There are many spatial based techniques available to perform fusion like pixel-to-pixel basis, block based fusion, region based fusion, and similarly transform techniques like DCT, Wavelet etc. To maintain simplicity of spatial operations and to avoid mis-registration problem, we have adopted block based fusion.

4.3.1 Block based Fusion: Multi-Focus

All local layers are divided into blocks of predefined size $m \times n$ ($m \times n \subset (P \times Q)$). The blocks with most informative section are identified by any of activity level measurement techniques available like gradient, deviation etc. Spatial frequency (SF) [8], a gradient based technique, is adopted to test the activity level of co-located blocks. The block with highest value of SF are selected to transfer into fused image.

For simplicity, we are using two source images to simplify the explanation, but it is extendable up-to N source images. Let ϕ_1 and ϕ_2 are two source images and corresponding local layers (I_{hf1}, I_{hf2}) are divided into blocks ($I_{hf1,b}, I_{hf2,b}$) of size $m \times n$. Spatial Frequency is calculated for ($I_{hf1,b}, I_{hf2,b}$) to evaluate the underlying information. Binary weight of value '1' is assigned to entire block of local layer, which has maximum SF, and '0' to other co-located blocks.

Let I_{w1} and I_{w2} are the binary weights for blocks of two corresponding local layers ($I_{hf1,b}, I_{hf2,b}$) and allocation of their values are as follows

$$\begin{aligned} &if\ SF_1 > SF_2, \quad [I_{w1}] = \boldsymbol{J_{mn}}\ \&\ [I_{w2}] = \mathbf{0_{mn}} \\ &\qquad\quad \text{else,} \quad [I_{w1}] = \mathbf{0_{mn}}\ \&\ [I_{w2}] = \boldsymbol{J_{mn}} \end{aligned} \tag{25}$$

where $\boldsymbol{J_{mn}}\ and\ \mathbf{0_{mn}}$ are unit and null matrix of size $m \times n$, respectively.

Local layers (I_{hf1}, I_{hf2}) are combined along with binary weights $\boldsymbol{I_{w|w1}}, \boldsymbol{I_{w|w2}}$ as weighted sum. Equation (26) represents the weighted sum.

$$I_{hf} = I_{hf1} \,.\, \boldsymbol{I_{w|w1}} + I_{hf2} \,.\, \boldsymbol{I_{w|w2}} \tag{26}$$

where '.' represents bitwise logical AND operation, $\boldsymbol{I_{w|w1}}\ and\ \boldsymbol{I_{w|w2}}$ represents weights for entire local layer and are given by

$$\boldsymbol{I_{w|w1}} = \begin{bmatrix} I_{w1} & \cdots & I_{w1} \\ \vdots & \ddots & \vdots \\ I_{w1} & \cdots & I_{w1} \end{bmatrix}$$

$$\boldsymbol{I_{w|w2}} = \begin{bmatrix} I_{w2} & \cdots & I_{w2} \\ \vdots & \ddots & \vdots \\ I_{w2} & \cdots & I_{w2} \end{bmatrix}$$

4.3.2 Block based Fusion: Multi-Exposure

Similar to Multi-focus fusion of local layers, Multi-exposure fusion of local layers is done by using block based fusion. All layers are divided into blocks of pre-defined size $m \times n$ $(m \times n \subset (P \times Q))$. Two activity level measurement techniques, spatial frequency (SF) and exposedness, are used to identify the informative blocks. Spatial frequency measures the underlying information, whereas exposedness evaluate the intensity value of block. Spatial frequency (SF) and exposedness are explained later in this section.

Two source images (ϕ_1, ϕ_2) and corresponding local layers (I_{hf1}, I_{hf2}) are divided into blocks $(I_{hf1,b}, I_{hf2,b})$ of size $m \times n$. Spatial frequency and exposedness is calculated for each block $(I_{hf1,b}, I_{hf2,b})$ and binary weights are allocated. I_{w1}, I_{w2} (refer Eq. (25)) are the weights for corresponding layers by using SF and I_{w3}, I_{w4} are the weights by using exposedness, refer Eq. (27, 28).

$$\begin{aligned} &if\ T \leq EX_1 \leq 255 - T, && [I_{w3}] = \boldsymbol{J_{mn}} \\ &\qquad else, && [I_{w3}] = \boldsymbol{0_{mn}} \end{aligned} \tag{27}$$

$$\begin{aligned} &if\ T \leq EX_2 \leq 255 - T, && [I_{w4}] = \boldsymbol{J_{mn}} \\ &\qquad else, && [I_{w4}] = \boldsymbol{0_{mn}} \end{aligned} \tag{28}$$

where $\boldsymbol{J_{mn}}$ and $\boldsymbol{0_{mn}}$ are unit and null matrix of size $m \times n$ respectively, T represents threshold value varies between 10 and 30 [57], and EX_1, EX_2 represents the block intensity value.

Combined weight are calculated as a product of weights using SF and Exposedness, refer Eq. (29-30).

$$[I_{w5}] = [I_{w1}] \,.\, [I_{w3}] \tag{29}$$

$$[I_{w6}] = [I_{w2}] \,.\, [I_{w4}] \tag{30}$$

where '.' represents bitwise AND operation.

Local layers (I_{hf1}, I_{hf2}) are combined along with binary weights $\boldsymbol{I_{w|w5}}, \boldsymbol{I_{w|w6}}$ as weighted sum, refer Eq. (31).

$$I_{hf} = I_{hf1}.\boldsymbol{I_{w|w5}} + I_{hf2}\,.\boldsymbol{I_{w|w6}} \tag{31}$$

where '.' represents bitwise AND operation, $\boldsymbol{I_{w|w5}}$ and $\boldsymbol{I_{w|w6}}$ represent weights for entire layer and are given by

$$\boldsymbol{I_{w|w5}} = \begin{bmatrix} I_{w5} & \cdots & I_{w5} \\ \vdots & \ddots & \vdots \\ I_{w5} & \cdots & I_{w5} \end{bmatrix}$$

$$\boldsymbol{I_{w|w6}} = \begin{bmatrix} I_{w6} & \cdots & I_{w6} \\ \vdots & \ddots & \vdots \\ I_{w6} & \cdots & I_{w6} \end{bmatrix}$$

4.3.3 Spatial Frequency

Spatial frequency (SF) measures the variations in intensity of pixels in horizontal as well as vertical direction. The variations are calculated by taking difference in pixel intensity in both the directions. These differences are higher for edges or textures as compare to flat section. Mathematically SF is calculated as

$$SF = \sqrt{(RF)^2 + (CF)^2} \tag{32}$$

$$RF = \sqrt{\frac{1}{mn}\sum_{x=1}^{m}\sum_{y=2}^{n} \phi(\,(x,y)\; - I(x,y-1))^2} \tag{33}$$

$$CF = \sqrt{\frac{1}{mn}\sum_{y=1}^{n}\sum_{x=2}^{m} \,(\,\phi(x,y)\; - \; I(x-1,y))^2} \tag{34}$$

where $\phi(x,y)$ is the intensity of image ϕ at location(x,y), $m \; \times n$ is the size of section. Higher value of SF represents more information content.

4.3.4 Exposedness

Exposedness reflects exposure level of a section in a source image. Darker sections in under exposed image and brighter sections in over exposed image have improper exposure level and very little information about the underlying details. To exclude these undesired sections and to replace them with normal exposure sections (obtained from other source images), exposedness factor is adopted in the present technique.

The mean value of region is used as exposedness of region, whereas intensity of pixel is used as exposedness of pixel. For 8-bit image exposure level between 'T' to '255-T' is considered normally exposed. Where T represents threshold and its value can be anywhere in between 10-30 [57]. The exposure level for image I having section of size $m \times n$ is given by

$$EX = \frac{\sum_{y=1}^{m}\sum_{x=1}^{n} I(x,y)}{m \times n} \tag{35}$$

4.4 Fusion of Global Layer:

Global layer (L_i) holds the remaining low frequency components of source images after removal of high frequency components. The proper selection of these low frequency components are as important as the selection of high frequency components. Global layer or low frequency component image holds very less intensity variations, which makes it difficult to identify correct pixel.

In order to identify correct low frequency components, we have used normalized average intensities of source images. The weights are calculated from these normalized average intensities. For two source images ϕ_1,ϕ_2 weights w_1, w_2 are given as

$$w_1 = \frac{Int_1}{Int} \tag{36}$$

$$w_2 = \frac{Int_2}{Int} \tag{37}$$

$$Int = Int_1 + Int_2 \tag{38}$$

where $Int_1 = \frac{\sum_x \sum_y \phi_1(x,y)}{P \times Q}$, $Int_2 = \frac{\sum_x \sum_y \phi_2(x,y)}{P \times Q}$ and $P \times Q$ is the size of source image.

Global layer (L_i) are combined together as weighted sum and described in Eq. (39)

$$I_{lf} = [w_1 \ \ w_2] \times [L_1 \ L_2]^T \tag{39}$$

4.5 Fused Image

Fused Image (I_f) is finally constructed by combining fused local and global layer on pixel-to-pixel basis. For multi-focus image fusion, fused local layer (refer Eq. (26)) and fused global layer (refer Eq. (39)) are combined to construct fused image. For multi-exposure image fusion, fused local layer (refer Eq. (31)) and fused global layer (refer Eq. (39)) are combined to construct fused image.

$$I_f = I_{hf} + I_{lf} \tag{40}$$

where I_f represents fused image.

4.6 Grey and Color Source Images

The proposed algorithm can process both grey as well as color images. For grey source images, ϕ_i is processed directly whereas for colour images, their luminance and chrominance components are processed separately. Luminance holds the underlying information of the image whereas chrominance holds the color information. The RGB components of source images (ϕ_i) is converted into luminance-chrominance (YC_rC_b) and back to RGB by using following expressions [79]

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ C_b &= 0.564(B - Y) \\ C_r &= 0.713(R - Y) \end{aligned} \tag{41}$$

$$\begin{aligned} R &= Y + 1.402C_r \\ G &= Y - 0.344C_b - 0.714C_r \\ B &= Y + 1.772C_b \end{aligned} \tag{42}$$

All processes of proposed technique like layers construction, layers fusion and fused image are applied on luminance of source images. The chrominance is processed separately to preserve the color information of the scene. We have combined chrominance C_{r_i}, C_{b_i} of source images as weighted sum. The weights are calculated based on deviations of respective chrominance image. The image has more deviation has more weightage and vice-versa, refer Eqs. (24-25).

$$C_{rf} = \sum_i WCr_i \times C_{r_i}: i \in [1,2..N] \tag{43}$$

$$C_{bf} = \sum_i WCb_i \times C_{b_i}: i \in [1,2..N] \tag{44}$$

where i represents number of source images, WCr_i, WCb_i are weights and given by

$$WCr_i = \frac{dev\left(C_{r_i}\right)}{\sum_i dev\left(C_{r_i}\right)}$$

$$WCb_i = \frac{dev\left(C_{b_i}\right)}{\sum_i dev\left(C_{b_i}\right)}$$

Finally processed luminance and chrominance are combined together to move back into RGB format.

4.7 Algorithm

The flow of proposed multi-focus and multi-exposure fusion technique are presented in algorithm form below. Fig. (4.4) represents the flow chart of the algorithm.

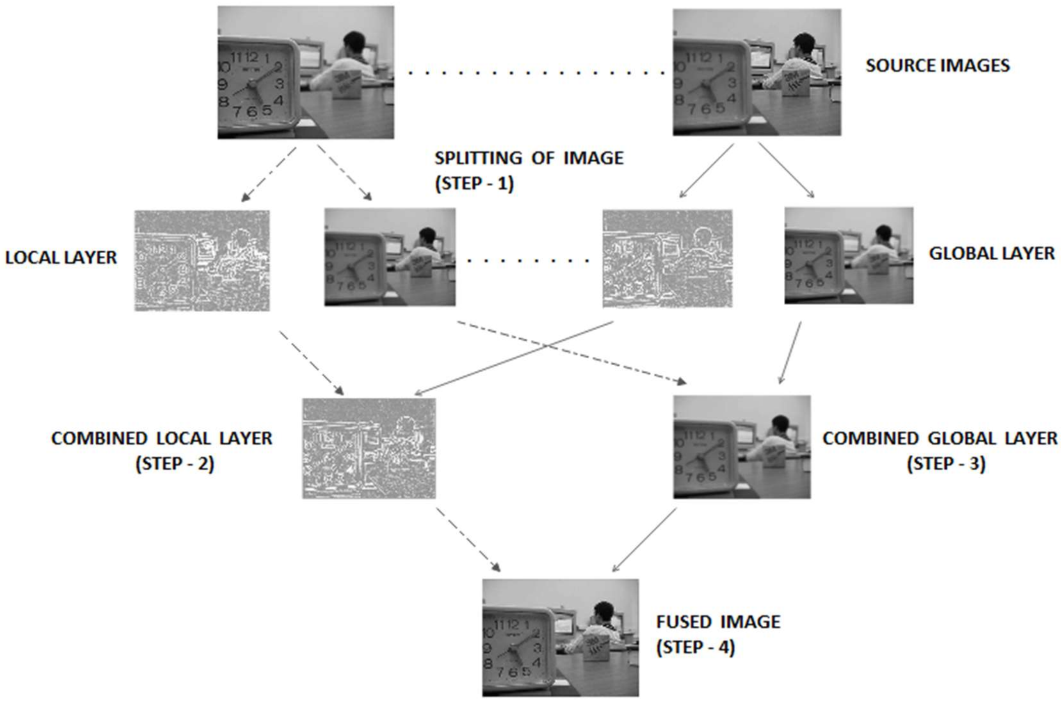


Fig. 4.4 Flow chart of proposed method

Algorithm: Block based multi-focus and multi-exposure fusion

Capture pre-registered images of scene at different exposure time and focused sections.

Input:

Pre-registered color source images ($\phi_{i,RGB}: i \in [1, \mathrm{N}]$);

Begin

1. Colour image ($\phi_{i,RGB}$) converted to (ϕ_{i,YC_rC_b})

   $Y_i = 0.299R_i + 0.587G_i + 0.114B_i$;

   $C_{bi} = 0.564(B_i - Y_i)$;

   $C_{ri} = 0.713(R_i - Y_i)$;

   Luminance (Y_i) and Chrominance(C_{bi}, C_{ri}) processed separately;

   Luminance

   a. Split source images into Local and Global layer

      Local layer: $I_{hfi} = \sum_{j=1}^{M} H_{ij}$;

      Global layer: $L_i = Y_i - I_{hfi}$;

b. Compute Local and Global fused image

i. Local fused image

Multi Focus: $I_{hf} = I_{hf1} . \boldsymbol{I_{w|w1}} + I_{hf2} . \boldsymbol{I_{w|w2}}$;

Multi Exposure: $I_{hf} = I_{hf1} . \boldsymbol{I_{w|w5}} + I_{hf2} . \boldsymbol{I_{w|w6}}$;

ii. Global Fused image

$I_{lf} = [w_1 \;\; w_2] \times [L_1 \, L_2]^T$;

c. Construct fused image

$$I_f = I_{hf} + I_{lf}$$

Chrominance

d. Process chrominance

$$C_{rf} = \sum_i WCr_i \times C_{r_i};$$
$$C_{bf} = \sum_i WCb_i \times C_{b_i};$$

2. Image (I_{i,YC_rC_b}) converted back to ($I_{i,RGB}$) using

$R = I_f + 1.402C_r$;

$G = I_f - 0.344C_b - 0.714C_r$;

$B = I_f + 1.772C_b$;

End

Output:

Color fused image ($I_{F,RGB}$);

4.8 Result and Analysis:

The proposed algorithm is tested on number of image sets in order to evaluate its performance. The performance is evaluated by using objective evaluation techniques, as in subjective evaluation view changes from person to person. The objective evaluation is based on mathematical formula and provide concrete results. Various objective indices are already discussed in chapter-3.

The algorithm is tested on randomly selected image sets in two ways. Firstly, effect of block size on the fused image is tested and compared with other block based techniques. This is done only for multi-focus image fusion. Secondly, quality of fused image is evaluated by comparing it with state-of-the-art techniques for multi-focus as well as multi-exposure image fusion.

4.8.1 Multi-focus Image Fusion

The local layers is combined together by using block based fusion technique. The effect of block based fusion is visible in two ways, Firstly the quality of fused image changes with change in block size. Secondly, the intensity value along the edges of block, affects the quality of fused image and it also deteriorate the visible appearance of an image. Combining all this together is known as blocking effect.

The proposed method has shown improvement in terms of removal of blocking effects while maintaining the quality of fused image. The addition of combined local layer over global (base) layer decreased the impact of intensity values along edges and improves the visible appearance. To test the effect of block size on the quality of fused image, we have varied the block size and evaluate the quality of fused image using standard quality indices. It is further compared with other block based techniques [8, 9]. Table 4.1 shows comparison chart in terms of $Q^{AB}/_F$.

The comparison is done in two ways, Firstly, the value of $Q^{AB}/_F$ is directly compared with other block based techniques and it is observed that for all block sizes proposed method (PM) has performed much better. Secondly, the deviation of values of $Q^{AB}/_F$ is calculated for all block sizes (refer Table 4.1, column 10) to assess the effect on the quality of fused image with change in block size. Deviation represents the deviation of $Q^{AB}/_F$ values of corresponding row. It is observed that deviation has a very small value as well as less compared to other block based techniques suggesting that block size variation has less (or negligible) effect on fused images.

Proposed method is further compared with some of the recent algorithms to show its effectiveness. The state-of-the-art selected to be compared are MSVD [27], DCT [41] and MFEF [42]. MSVD [27] is based on singular value decomposition and have compared their results with wavelets. DCT [41] have used spatial frequency in DCT domain and have compared with previous DCT based techniques and NSCT. MFEF [42] is pixel based most recent technique. To test proposed method with these techniques, three sets of images are used. First set has three sets of grey source images without reference image; Boy, Pepsi and Disk. Second set contains two sets of grey source image with reference image; circuit, building. Last set has one color image; grass. Two well-known non-reference objective indices $Q^{AB}/_F$ and MI and two reference objective indices

PSNR and RMSE are adopted to verify the quality of fused images. Table-4.2 shows the values of various quality indices and Fig. (4.11) represents its corresponding bar graph with y-axis represents quality index scale and x-axis represents techniques under comparison.

Table 4.1

Comparison chart for varying block sizes

Image	Method	8 × 8	16 × 8	8 × 16	16 × 16	16 × 32	32 × 16	32 × 32	Deviation
Boy	BB [8]	0.5075	0.5133	0.5044	0.5074	0.5044	0.5021	0.4980	0.0048
	ANN[9]	0.3289	0.3367	0.3597	0.3559	0.3574	0.3574	0.3574	0.0123
	PM	0.6849	0.6825	0.6879	0.6839	0.6821	0.6832	0.6819	0.0021
Circuit	BB [8]	0.5697	0.5645	0.5641	0.5611	0.5575	0.5589	0.5579	0.0044
	ANN[9]	0.4484	0.4287	0.4374	0.4304	0.4244	0.4224	0.4187	0.0101
	PM	0.7416	0.7417	0.7435	0.7422	0.7338	0.7419	0.7335	0.0042

Number of high frequency images (M) – 60, PM – Proposed Method

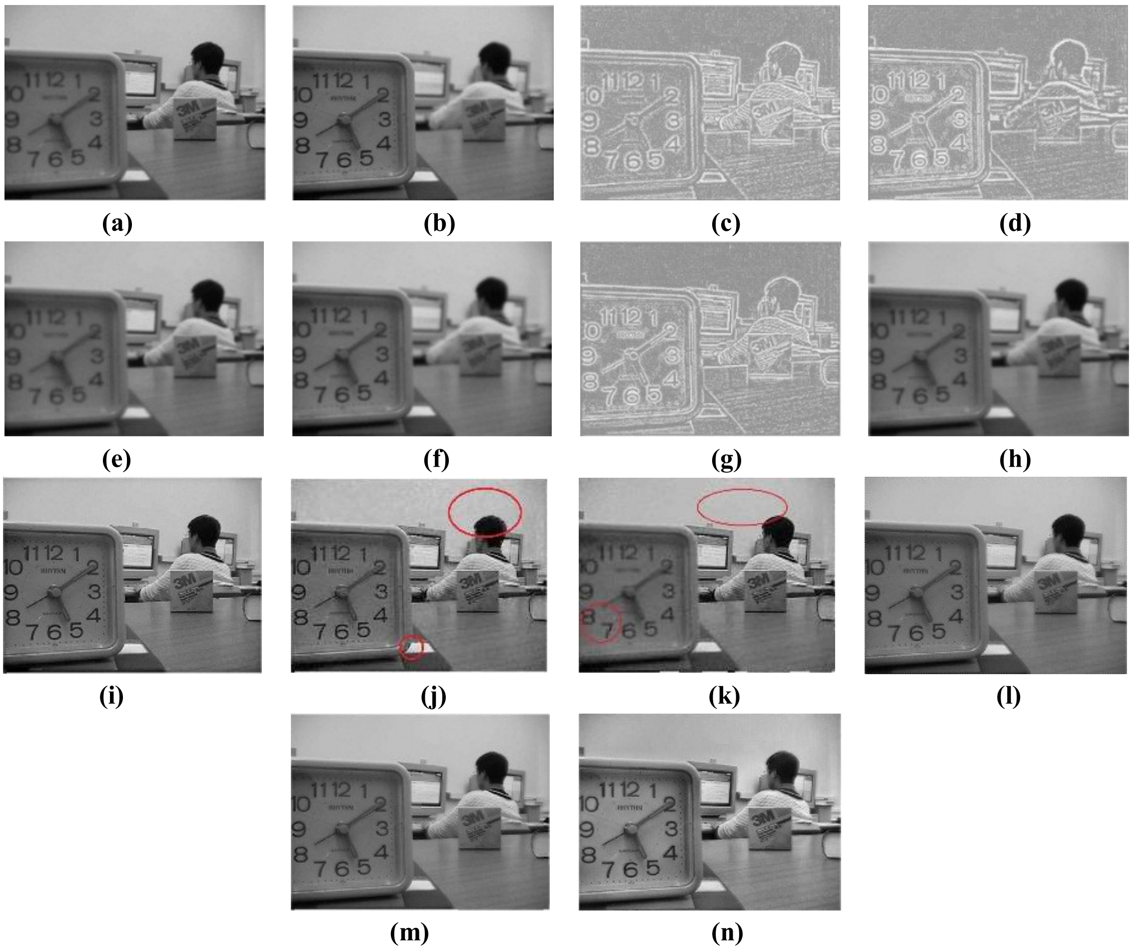


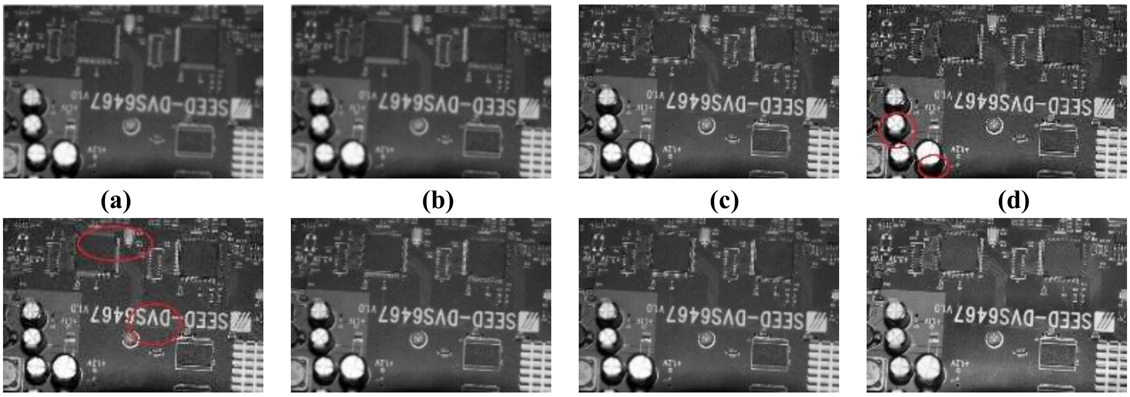
Fig. 4.5 “Image Boy,” source images and fusion results: **a – b** source images, **c – d** corresponding local layers, **e – f** corresponding global layers, **g** combined local layer, **h** combined global layer, **i – n** fused images for PM, BB, ANN, MSVD, DCT, MFEF

Table 4.2

Comparison chart on quality indices for multi-focus fusion

Image	Method	$Q^{AB/F}$	MI	PSNR	RMSE
Boy	MSVD [27]	0.4868	6.1849	Only applicable where reference image is available	
	DCT [41]	0.5259	6.6338		
	MFEF [42]	0.6219	3.7390		
	PM	0.6849	6.2641		
Pepsi	MSVD [27]	0.6666	6.5306	Only applicable where reference image is available	
	DCT [41]	0.6320	6.8613		
	MFEF [42]	0.5736	4.0979		
	PM	0.7620	7.0716		
Disk	MSVD [27]	0.5572	5.8261	Only applicable where reference image is available	
	DCT [41]	0.5448	6.2488		
	MFEF [42]	0.6734	3.7837		
	PM	0.7127	6.4572		
Grass	MSVD [27]	0.5547	4.3536	Only applicable where reference image is available	
	DCT [41]	0.5798	5.0336		
	MFEF [42]	0.6074	3.1665		
	PM	0.7291	5.4808		
Building	MSVD [27]	0.5744	5.2669	75.897	0.0411
	DCT [41]	0.6267	6.4964	82.004	0.0203
	MFEF [42]	0.7139	3.2683	67.256	0.1110
	PM	0.7312	8.0571	89.055	0.0090
Circuit	MSVD [27]	0.5108	5.0105	75.075	0.0451
	DCT [41]	0.6561	6.0973	79.974	0.0257
	MFEF [42]	0.6921	3.0311	70.244	0.0787
	PM	0.7416	7.5504	92.059	0.0064

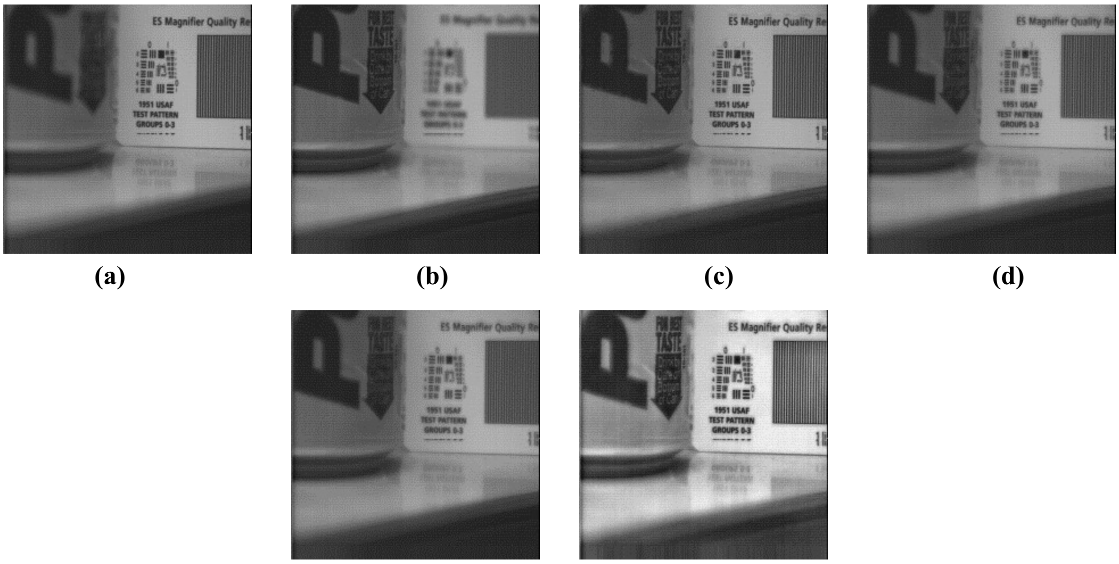
Number of high frequency images (M) – 60, Block size - 8×8, PM – Proposed Method



(a) (b) (c) (d)

(e) (f) (g) (h)

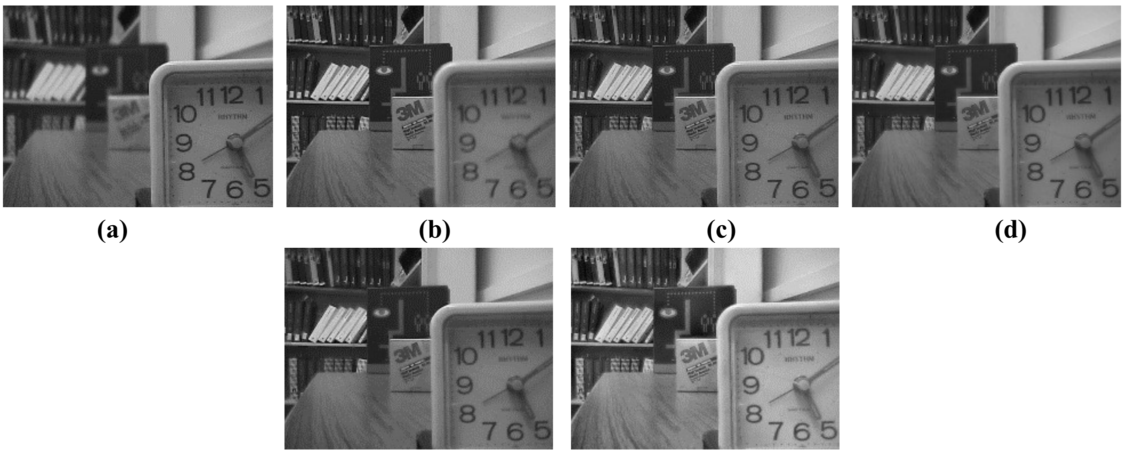
Fig. 4.6 "Image Circuit," source images and fused results: **a – b** source images, **c – h** fused images for PM, BB, ANN, MSVD, DCT, MFEF



(a) (b) (c) (d)

(e) (f)

Fig. 4.7 "Image Pepsi," source images and fused results: **a – b** source images, **c – f** fused images for PM, MSVD, DCT and MFEF



(a) (b) (c) (d)

(e) (f)

Fig. 4.8 "Image Disk," source images and fused results: **a – b** source images, **c – f** fused images for PM, MSVD, DCT and MFEF

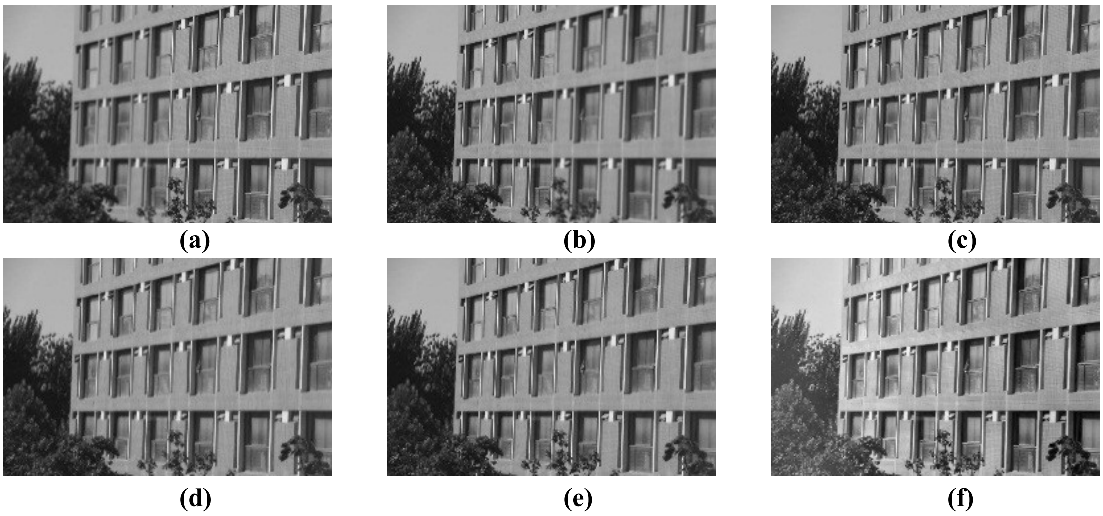


Fig. 4.9 “Image Building,” source images and fused results: **a – b** source images, **c – f** fused images for PM, MSVD, DCT and MFEF

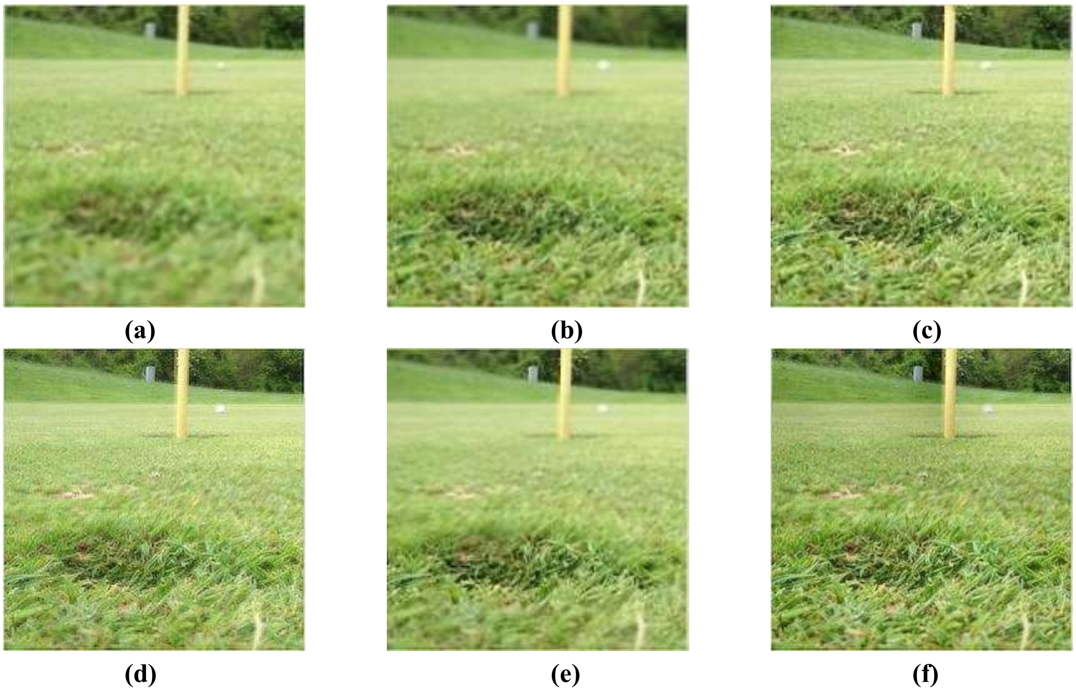


Fig. 4.10 “Image Grass,” source images and fused results: **a – b** source images, **c – f** fused images for PM, MSVD, DCT and MFEF

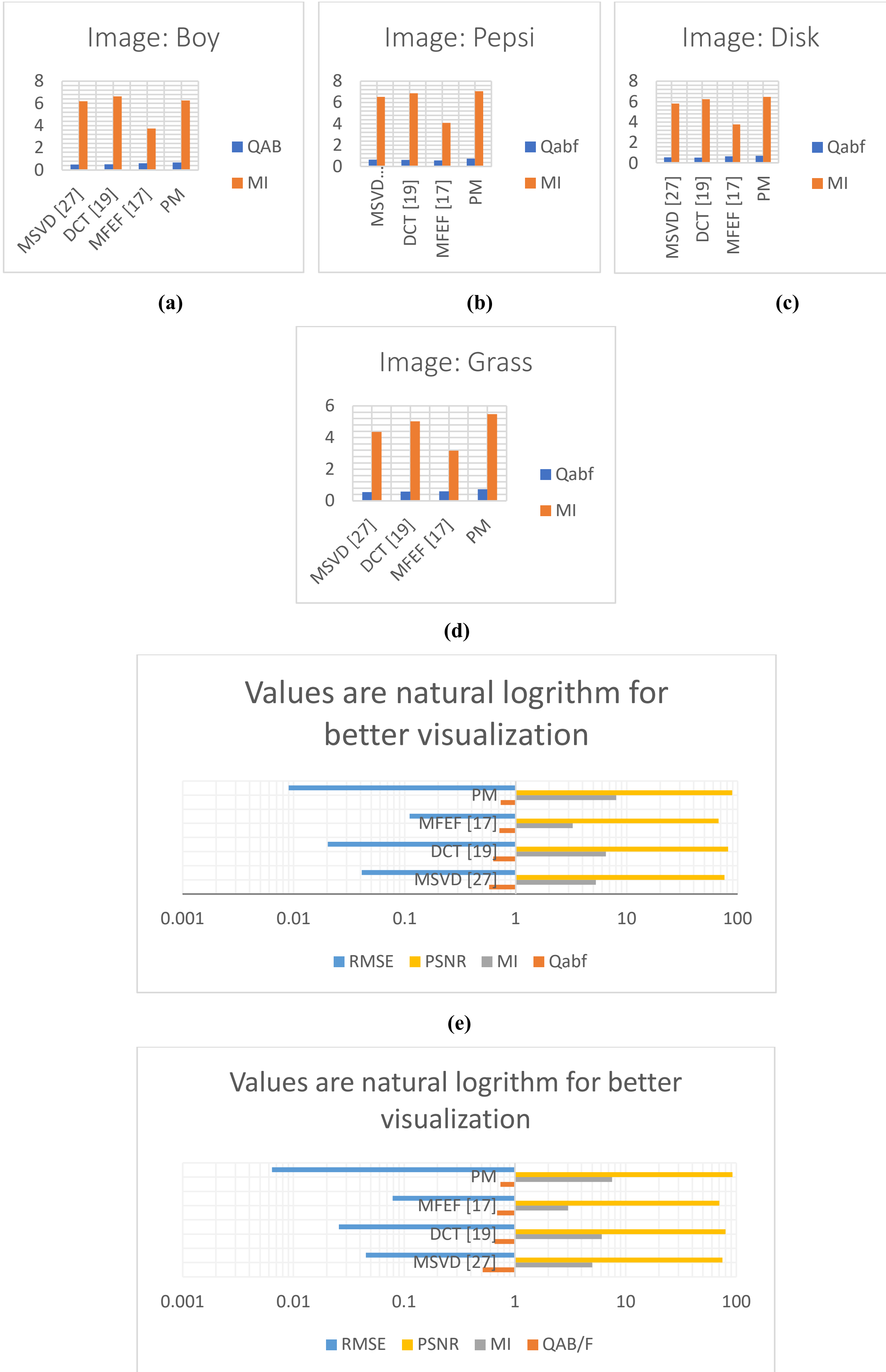


Fig. 4.11 Bar graphs corresponding to Table 4.2, **a** Image 'Boy', **b** Image 'Pepsi', **c** Image 'Disk', **d** Image 'Grass', **e** Image 'Building', **f** Image 'Circuit'

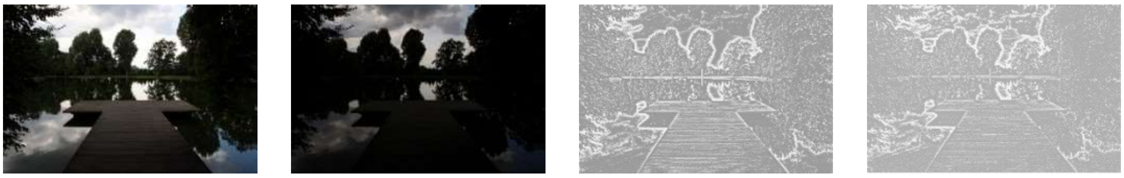
Fig. (4.5) gives the view of entire sequence of algorithm using source image 'Boy'. Fig. (4.5(a-b)) represents the two source images, Fig. (4.5(c-d)) represents the corresponding local layers, Fig. (4.5(e-f)) represents corresponding global layers, Fig (4.5(g)) represents block based fused local layer, Fig. (4.5(h)) represents weighted sum of global layers, Fig. (4.5(i)) represents the final fused image and remaining are fused images of state-of-the-art techniques, refer Fig. (4.5(j-n)). Fig. (4.6-4.10) shows the fused images of other five image sets.

From Table 4.2 it is clear that the proposed method has shown highest $Q^{AB}/_{F}$ as well as MI for all sets of images. For image sets circuit and building, where reference images are available, peak signal to noise ratio (PSNR) is highest for both images. Image circuit has shown more than 92% PSNR and root mean square error is lowest with large margin for both the images.

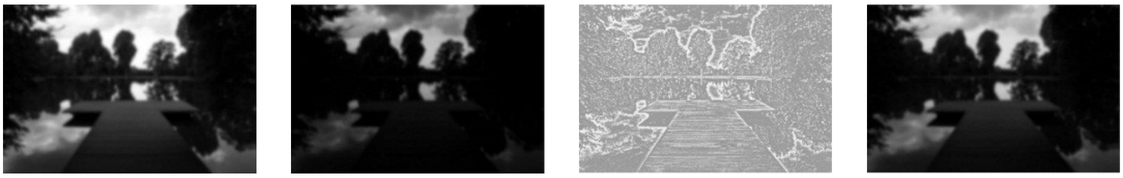
4.8.2 Multi-exposure Image Fusion

The proposed method for multi-focus image fusion is further extended for multi-exposure image fusion. To test multi exposure image fusion three sets of color images (Pond, Cathedral and Garden) are used. The state-of-the-art selected for comparison are Exposure Fusion (EF) [50], Median filter (MF) [57] and MFEF [42]. Exposure fusion [50] is a classical multi-exposure image fusion method. MF [57] is pixel based method and considered to check the performance of proposed method against tone mapping techniques. MFEF [42] is a gradient based fusion method and most recent technique. To evaluate the performance of proposed method, three quality indices $Q^{AB/F}$, VIFF and MI are adopted to compare proposed method with state of art techniques.

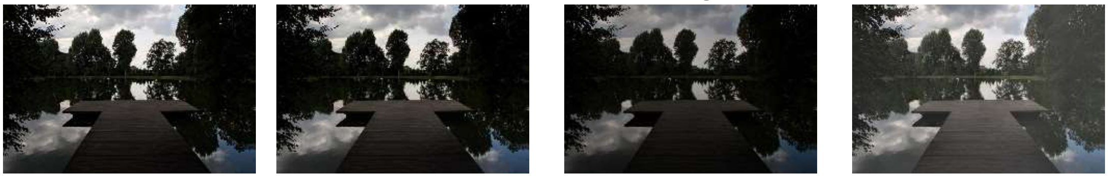
Fig. (4.12-4.14), shows the results of proposed method on three image sets. Table 4.3 shows the results of quality indices and Fig. (4.15), represents the corresponding bar graphs. For image pond, $Q^{AB/F}$ and VIFF have highest value for proposed method and MI is second highest (after MF). For image cathedral, proposed method has shown highest value for all indices, and for image garden, proposed method has performed better in terms of VIFF, MI whereas $Q^{AB/F}$ has similar results to EF.



(a) (b) (c) (d)



(e) (f) (g) (h)



(i) (j) (k) (l)

Fig. 4.12 "Image pond[5]," Source images and fusion results: **a – b** source images, **c – d** corresponding local layers, **e – f** corresponding global layers, **g** combined local layer, **h** combined global layer, **i – l** fused images for PM, EF, MF and MFEF

(a) (b) (c) (d)

(e) (f)

Fig. 4.13 "Image cathedral," source images and fusion results: **a – b** source images, **c – f** fused images for PM, EF, MF and MFEF

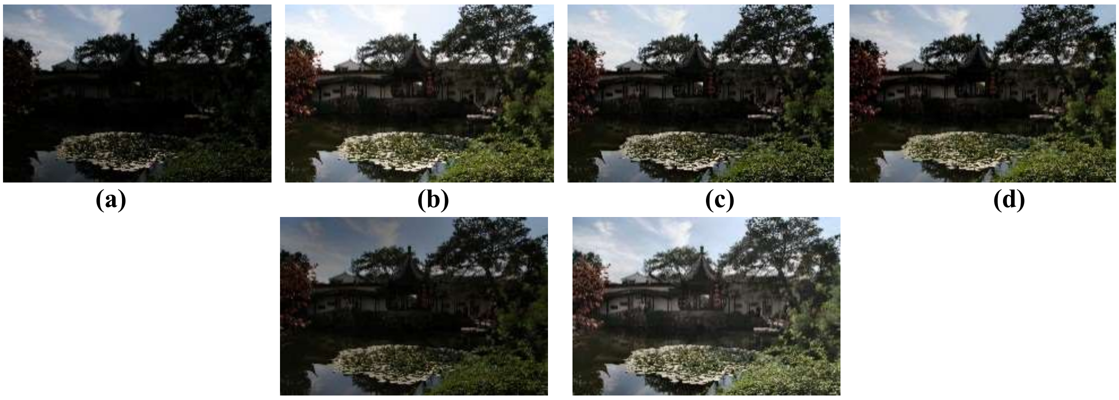


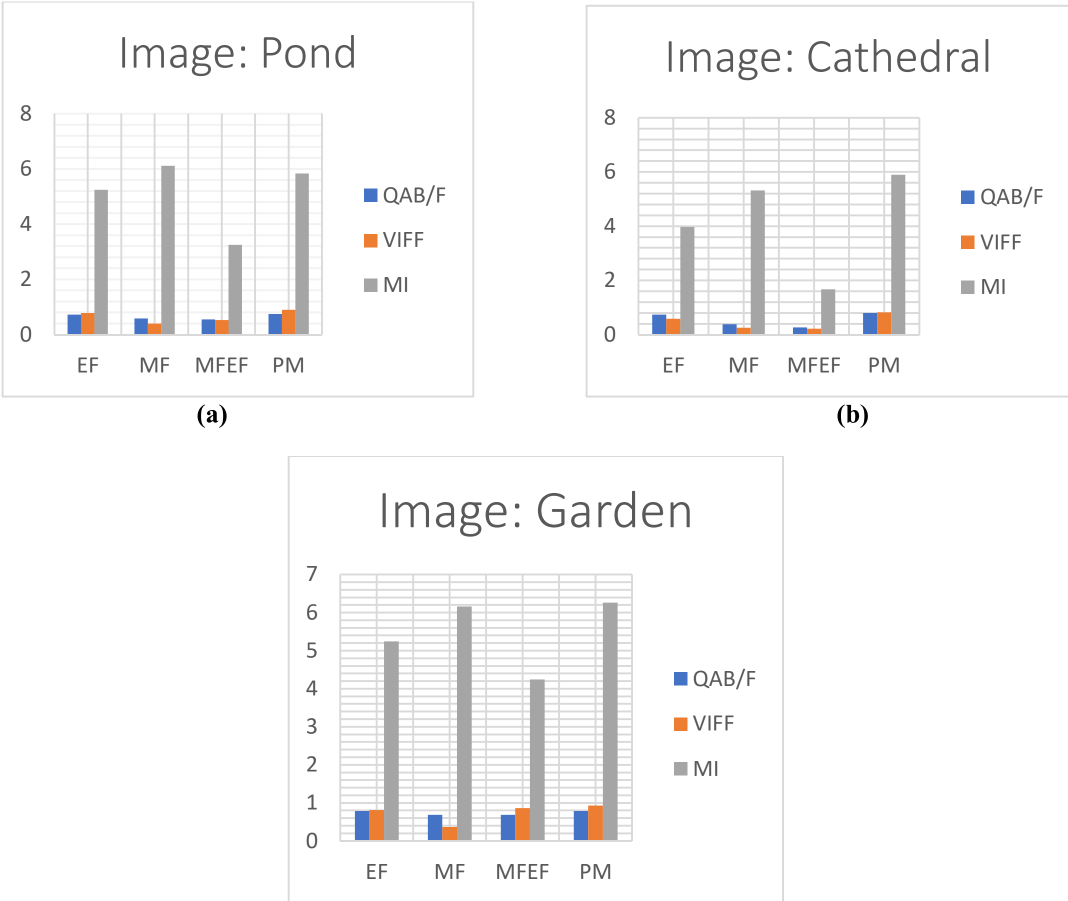
Fig. 4.14 “Image garden,” source images and fusion results: **a – b** source images, **c – f** fused images for PM, EF, MF and MFEF

Table 4.3

Comparison chart on quality indices for multi-exposure fusion

Image		Pond		
	Method	$Q^{AB/F}$	VIFF	MI
	EF [50]	0.7244	0.7825	5.2401
	MF [57]	0.5799	0.4013	6.1086
	MFEF [42]	0.5561	0.5250	3.2514
	PM	0.7466	0.9027	5.8337
Image		Cathedral		
	Method	$Q^{AB/F}$	VIFF	MI
	EF [50]	0.7351	0.5870	3.9738
	MF [57]	0.3827	0.2564	5.3230
	MFEF [42]	0.2702	0.2126	1.6684
	PM	0.7918	0.8236	5.8923
Image		Garden		
	Method	$Q^{AB/F}$	VIFF	MI
	EF [50]	0.7881	0.8098	5.2404
	MF [57]	0.6888	0.3696	6.1600
	MFEF [42]	0.6861	0.8619	4.2467
	PM	0.7876	0.9263	6.2649

Number of high frequency images (M) – 50, Block size - 8 × 8, PM – Proposed Method



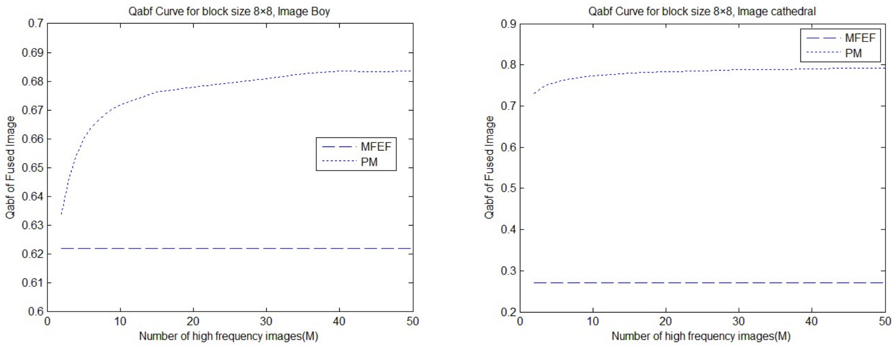
(a) **(b)** **(c)**

Fig. 4.15. Bar graphs corresponding to Table 4.3, **a** Image 'Pond', **b** Image 'Cathedral', **c** Image 'Garden'

4.8.3 Controlling parameter M

Number of high frequency images (M), refer (1-2), act as controlling parameter, it controls the information to be passed into fused image. Higher value of M transfers more information from source images into local layer (I_{hf}). The computational time also varies with change in value of M; image Grass (size 242×242) takes 0.15 sec and 0.9783 sec for value of M equal to 2 and 60, respectively.

The optimum value of M decided by user based on application. For real time operations M can be in between 2 to 5, whereas for high quality fused image, M can be set anywhere in between 60-80. Fig. 4.16 shows the variation in quality index $Q^{AB/F}$ with respect to M for proposed method and MFEF [42].



(a) **(b)**

Fig. 4.16.$Q^{AB/F}$ vs. number of high frequency images (M) for PM and MFEF: **a** MFIF, source image 'boy' **b** MEIF, source image 'cathedral'

Chapter – 5

Feature based fusion for dynamic range

5.1 Introduction

A scene varies largely by exposure level, therefore a single LDR (low dynamic range) camera can never collect all underlying information of the scene. The captured image may experience under or over-exposed areas due to high or low shutter speed of the camera, respectively. High shutter speed or low exposure time results in information loss in darker regions and vice-versa (refer Fig. 5.1). An image constructed by collecting complementary information from varying exposure time LDR images (source images) can preserve all the details of the scene. Two major approaches to perform the operation are: HDR (high dynamic range) tone mapping and multi-exposure image fusion. LDR source images are combined to construct HDR tone mapped image by extracting Camera Response Function (CRF). Extraction of CRF requires additional information of source images; like, exposure time and exposure value. Next, dynamic range of resulting tone mapped HDR image is compressed to LDR while preserving details [80-84].

In multi-exposure image fusion, LDR images are combined directly by extracting sections with high activity level to construct a fused image. The fused image is LDR or 'HDR like' image, which preserves all the details of source images without constructing an HDR image. Fig. 5.2 shows block diagram of both techniques.

The proposed algorithm is spatial domain multi-exposure image fusion technique. It uses the features of captured and pre-registered source images to synthesize fused image.

(a) (b) (c)

Fig. 5.1 Source images, 'Garden' **a** Low exposure time **b** Medium exposure time **c** Large exposure time

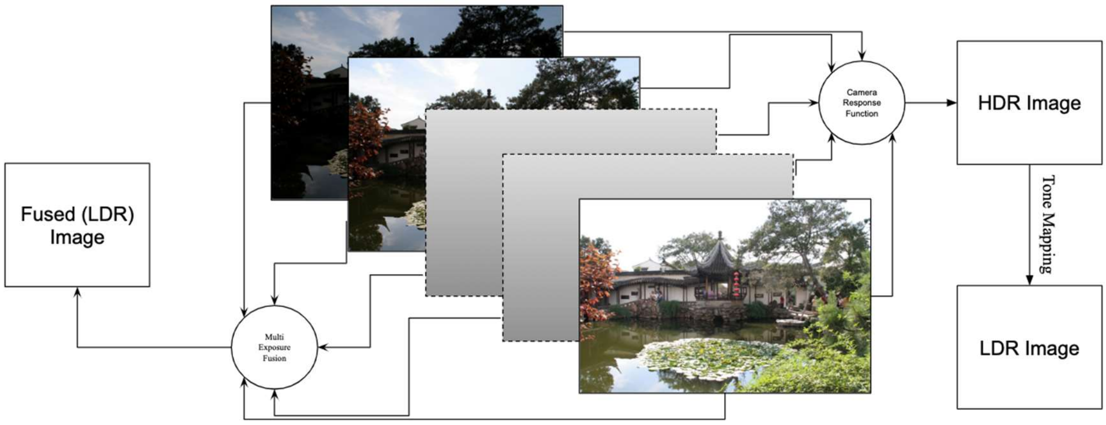


Fig. 5.2. Flow diagram for tone-mapping and multi-exposure image fusion

5.2 Disintegration of Image into three features

Variations in intensity of pixels are strong, weak or even negligible in different sections of an image. We have separated them into three different features. Strong, weak and negligible varying pixels are defined as high varying, low varying and constant components respectively.

The combination of Gaussian filter and Recursive filter (RF) [85] is used to separate these features. Gaussian filter is well known averaging filter, which smooth the variations to constant level. The smoothness controlling parameter is standard deviation (σ_g). Recursive filter (RF) is an edge sensitive averaging filter, which smooths the variations while preserving edges. This is performed by defining two variable parameters: standard deviation in range (σ_r) and space (σ_s). Standard deviation in space (σ_s) considers the distance, while standard deviation in range (σ_r) consider the intensity of neighbour region. In order to preserve the edges, value of σ_r needs to be kept small [85].

Let I_i be the set of pre-registered N source images of size $P \times Q$, where $i \in 1 \dots . N$. All source images I_i are filtered by RF $(rf(\cdot))$ with pre-defined values of σ_{r1}, σ_{s1}. RF preserves the edges or high varying features as per the value of σ_r and smooths low varying features. So RF processed images (E_i) holds high varying features and very low varying or constant features. RF filtered images (E_i) are further processed by a Gaussian filter $(gf(\cdot))$ with defined value of σ_g to smooth retained high varying features. The resultant images (B_i) holds constant features only. Three features are extracted by using I_i, E_i, and B_i, refer Eq. (45-48). Fig. 5.3, represents the flow of application of these filters.

$$E_i = rf(I_i) \tag{45}$$

$$L_i = I_i - E_i \tag{46}$$

$$B_i = gf(E_i) \tag{47}$$

$$H_i = E_i - B_i \tag{48}$$

where $gf(\cdot), rf(\cdot)$ represents Gaussian and recursive filter, respectively.

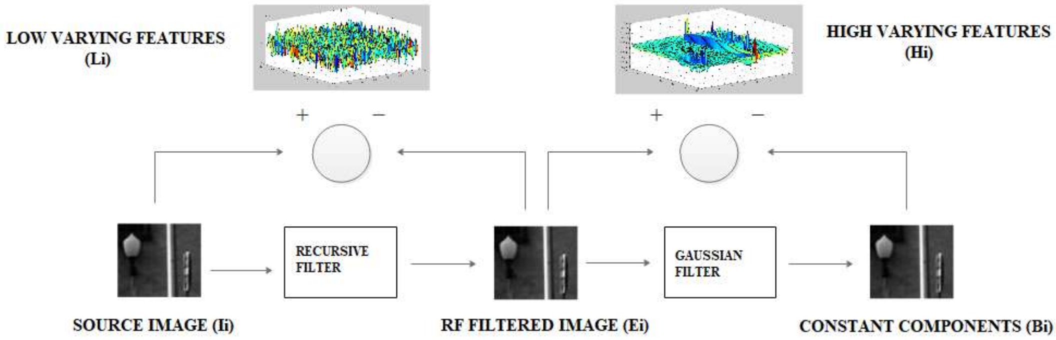


Fig. 5.3. Proposed flow diagram to extract high, low varying and constant components

Eq. (45) is the RF processed image and Eq. (47) represents the Gaussian filtering on E_i. The difference in source image (I_i) and RF processed image (E_i) separates the low varying feature images (L_i), refer Eq. (46). The difference in RF processed images E_i and Gaussian processed images G_i extracts the high varying feature images H_i, refer Eq. (48). B_i hold the constant features, refer Eq. (47).

5.3 Fusion of high and low varying feature images

High and low varying feature images (H_i, L_i) usually hold the edges and texture of source images respectively. Fusion of these feature images is performed by extracting features with maximum information and combining them. The extracted features are combined to construct combined high (H) and low (L) varying feature image.

The simplest process to test underlying information of any pixel in feature image is to use absolute maximum intensity value of pixel. The possibility of selecting noise instead of informative pixel by using absolute maximum is more probable and thus makes it an incorrect choice. Since neighboring pixels are interdependent, therefore testing pixels along their neighbors is preferred. Neighbor Distance (ND) filter [29], a high pass filter based on oriented distance, considers surrounding pixels while testing the activity level. Based on this, ND filter is adapted to test the activity level.

All feature images are processed with ND filter and corresponding high frequency values are calculated for each pixel. Let the corresponding high frequency value be represented as NDV. The NDV for high varying feature images (H_i) and low varying feature images (L_i) are denoted as $HNDV_i$ and $LNDV_i$ respectively. The pixel with highest value of NDV indicates that the corresponding pixel holds maximum underlying information.

All co-located pixels for high varying as well as low varying feature images are tested for highest underlying information based on NDV calculated using ND filter. The pixel with largest value of NDV among co-located pixels is selected and combined to construct fused high varying and low varying feature images. Eq. (49-50) represent ND filtering operation on H_i and L_i. Eq. (51-52) represents the construction of fused high (H) and low varying (L) images.

$$HNDV_i(p,q) = K_{ND} \circledast H_i(p,q)\,;\ i \in [1, \mathrm{N}-1] \tag{49}$$

$$LNDV_i(p,q) = K_{ND} \circledast L_i(p,q)\,;\ i \in [1, \mathrm{N}-1] \tag{50}$$

$$H(p,q) = H_i(p,q)\ for\ \max_{1\le i\le N}(|HNDV_i(p,q)|) \tag{51}$$

$$L(p,q) = L_i(p,q)\ for\ \max_{1\le i\le N}(|LNDV_i(p,q)|) \tag{52}$$

where N is number of source images, (p,q) represents pixel location, K_{ND} is ND filter kernel and '$\circledast$' represents convolution process.

5.4 Fusion of constant feature images

The extraction of high (H_i) and low (L_i) varying feature images from source images (ϕ_i) leaves very small or no variations in intensity in source images. The remaining flat images are the base images without frequency components and thus known as constant feature images (B_i). The selection of featured pixels for constant feature images are equally important as for high and low varying feature images. The wrong selection of pixels for constant feature images results in degradation of fused image.

To select pixels from B_i, ND filter is applied on source images (ϕ_i) and corresponding high frequency value (NDV) is computed and denoted as $BNDV_i$. $BNDV_i$ is further used to compute weights for all source images and normalized in the range of [0, 1]. These weights are combined with constant feature images B_i as weighted sum. Eq. (53) represents ND filtering on the source images, and calculation of weights are given by equations (54-55). Eq. (56) represents the weighted sum of constant feature images and calculated weights, refer fig. (5.4(a)).

$$BNDV_i(p,q) = K_{ND} \circledast I_i(p,q) \tag{53}$$

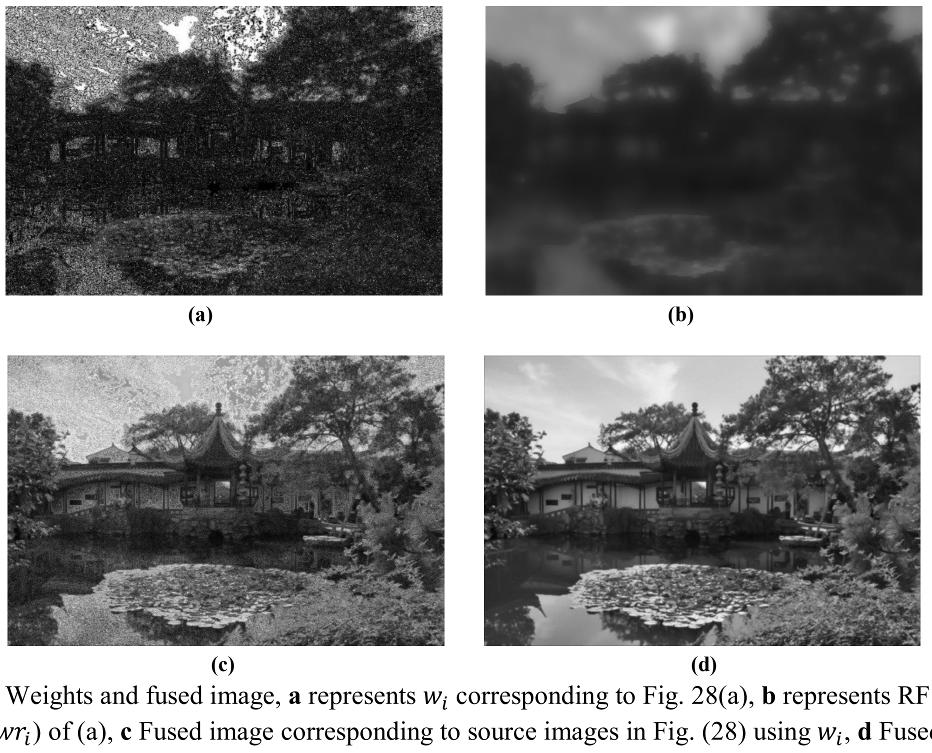
where K_{ND} is ND filter kernel.

$$w1_i(p,q) = |BNDV_i(p,q)| \tag{54}$$

$$w_i(p,q) = \frac{w1_i(p,q)}{\sum_i w1_i(p,q)};\ i \in [1, \mathrm{N}] \tag{55}$$

$$B(p,q) = \sum_{i=0}^{N}(w_i(p,q) \times B_i(p,q)) \tag{56}$$

where (p,q) represents pixel location.



(a) (b) (c) (d)

Fig. 5.4 Weights and fused image, **a** represents w_i corresponding to Fig. 28(a), **b** represents RF filtered image (wr_i) of (a), **c** Fused image corresponding to source images in Fig. (28) using w_i, **d** Fused image corresponding to source images in Fig. (28) using wr_i

Fig. 5.4(a) shows the weights calculated using Eq. (55) for source image of Fig. 5.1(a). It is clear that it has an abrupt variations in intensity. These calculated weights are combined with constant feature images, refer Eq. (56), and then to construct final fused image (to be discussed next). The abrupt variations clearly affects the quality of fused image, refer Fig. 5.4(c).

In order to smooth down these abrupt variations, calculated weights in Eq. (55) are processed by RF with parameter (σ_{r2} and σ_{s2}), refer Fig. 5.4(b). The smooth weights are combined with constant feature images to construct fused constant feature image. Eq. (57) represents the processed weights and Eq. (58) represents the construction of fused constant feature image by weighted sum of N constant feature images.

$$wr_x = rf(w_x);\ x \in [0, \mathrm{N}-1] \tag{57}$$

where $rf(\cdot)$ represents RF filter.

$$B_r(p,q) = \sum_{x=0}^{N-1}(wr_x(p,q) \times B_x(p,q)) \tag{58}$$

Fig. (5.4(b)) and (5.4(d)) represent the weights and fused image respectively.

5.5 Fused image:

Final fused image is constructed by combining constant (B_r), low varying (L) and high varying (H) fused images. Fused low varying image (refer Eq. (52)), fused high varying image (refer Eq. (51)) and fused constant feature image (refer Eq. (58)) are combined on pixel to pixel basis, refer Eq. (59).

$$I_f(p,q) = H(p,q) + L(p,q) + B_r(p,q) \tag{59}$$

where I_f represents fused image.

5.6 Grey and Color Source Images:

The proposed algorithm processes both grey as well as color images. For color images, RGB components are converted into luminance-chrominance (YC_rC_b), refer Eq. (41). Luminance holds the underlying information of the image whereas chrominance holds the color information. Both luminance and chrominance are processed separately and converted back to RGB components, refer Eq. (42) [79].

For grey images, ϕ_i represents the grey scale intensity images. Whereas for colour images, all operations are performed on luminance component. The chrominance components (C_{r_i}, C_{b_i}) of source images are processed by combining as weighted sum. The weights for chrominance are calculated as deviation of each chrominance pixel; refer (60-65).

$$C_r(p,q) = \sum_{i=0}^{N} W_{ri}(p,q) \times C_{ri}(p,q) \tag{60}$$

$$C_b(p,q) = \sum_{i=0}^{N} W_{bi}(p,q) \times C_{bi}(p,q) \tag{61}$$

$$W2_{ri}(p,q) = \sqrt{(C_{ri}(p,q) - \mu_{C_{ri}})^2} \tag{62}$$

$$W3_{bi}(p,q) = \sqrt{(C_{bi}(p,q) - \mu_{C_{bi}})^2} \tag{63}$$

$$W_{ri}(p,q) = \frac{W2_{ri}(p,q)}{\sum_i W2_{ri}(p,q)} \tag{64}$$

$$W_{bi}(p,q) = \frac{W3_{bi}(p,q)}{\sum_i W3_{bi}(p,q)} \tag{65}$$

where W_{ri} and W_{bi} are the normalized weights, $W2_{ri}$ and $W3_{bi}$ are un-normalized weights, $\mu_{C_{ri}}$ and $\mu_{C_{bi}}$ are means of corresponding chrominance images.

5.7 Algorithm:

This section presents the step by step operation of proposed multi-exposure fusion technique followed by the algorithm.

The process of multi-exposure fusion technique is as follows

Step 1: Let pre-registered source images be $(\phi_i; i \in [1, \mathrm{N}])$ of size $P \times Q$.

Step 2: RGB components are converted to YC_rC_b(Luminance - Chrominance) format using Eq. (41).

In the following steps, step 3-6 processes luminance (Y_i) and step 7-8 processes chrominance (C_{bi}, C_{ri}). For processing grey images, chrominance part is not required.

Step 3: Separate H_i, L_i and B_i from respective source images, refer section 5.2.

Step 4: Compute H, L using defined fusion rule, refer section 5.3.

Step 5: Compute B_r using defined fusion rule, refer section 5.4.

Step 6: H, L and B_r are combined together on pixel to pixel basis to construct final fused image, refer section 5.5.

Fig. (5.4), represents the fused image (luminance) for source images represented in Fig. (28). For grey source images, this is the final fused image.

Step 7: Chrominance C_{ri} and C_{bi} are processed to retain the colour information, refer section 5.6.

Step 8: Chrominance C_r, C_b along with fused luminance Y (refer (16)) are combined on pixel to pixel basis to get colour fused image, refer equation (42).

Algorithm: Feature based multi-exposure fusion

Capture pre-registered images of scene at different exposure time.

Input:

Pre-registered color source images ($\phi_{i,RGB}: i \in [1, \mathrm{N}]$);

Begin

Colour image ($\phi_{i,RGB}$) converted to (ϕ_{i,YC_rC_b})

$Y_i = 0.299R_i + 0.587G_i + 0.114B_i$;

$C_{bi} = 0.564(B_i - Y_i)$;

$C_{ri} = 0.713(R_i - Y_i)$;

Luminance (Y_i) and Chrominance(C_{bi}, C_{ri}) processed separately;

Luminance

a. Extract three feature images H_i, L_i, B_i from Y_i

$E_i = rf(Y_i)$;

$L_i = Y_i - E_i$;

$B_i = gf(E_i)$;

$H_i = E_i - B_i$;

b. Compute combined feature images

i. High varying feature image

$H(p,q) = H_i(p,q)\ for\ \max_{1 \le i \le N}(|HNDV_i(p,q)|)$;

ii. Low varying feature image

$$L(p,q) = L_i(p,q) \; for \; \max_{1\le i\le N}(|LNDV_i(p,q)|);$$

iii. Constant feature image

$$Br(p,q) = \sum_{i=1}^{N}(wr_i(p,q) \times B_i(p,q));$$

c. Construct fused image

$$Y(p,q) = H(p,q) + L(p,q) + B_r(p,q)\,;$$

Chrominance

d. Process chrominance

$$C_r(p,q) = \sum_{i=1}^{N} W_{ri}(p,q) \times C_{ri}(p,q);$$

$$C_b(p,q) = \sum_{i=1}^{N} W_{bi}(p,q) \times C_{bi}(p,q);$$

Image (I_{i,YC_rC_b}) converted back to ($I_{i,RGB}$) using

$R = Y + 1.402C_r;$

$G = Y - 0.344C_b - 0.714C_r;$

$B = Y + 1.772C_b;$

End

Output:

Color fused image ($I_{F,RGB}$);

5.8 Result and Analysis

The proposed method is tested on ten sets of random source images to evaluate its performance. The performance is evaluated with mathematical formula based objective evaluation techniques, unlikely to subjective evaluation which is person dependent. Various objective indices are already discussed in chapter 3. Three indices are selected to evaluate the performance: Mutual Information (MI) [74], Normalized Mutual Information (NMI) [75] and Edge-based Similarity Index ($Q^{\frac{AB}{F}}$) [76]. MI and NMI evaluate probability based information, and $Q^{\frac{AB}{F}}$ evaluates edges transferred to fused image.

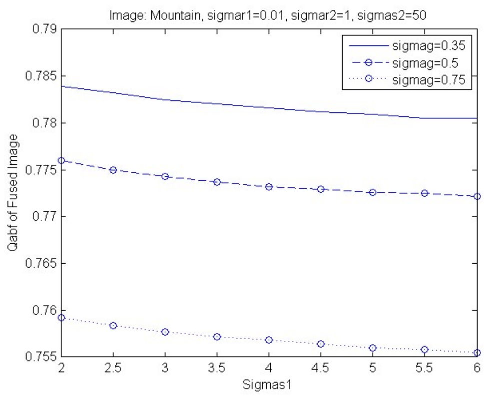
The performance of proposed method is compared with state-of-the-art techniques: SIFT [66], MFEF [42], and SPD [68]. SIFT [66] extracts local features which are used and used to construct fused image. MFEF [42] performs fusion in gradient domain, which is further processed by Haar wavelet to reconstruct fused image. SPD [68] decomposed source images into three components: signal strength, signal structure and mean intensity. These three components are combined separately to construct a fused image.

5.8.1 Parameter Selection

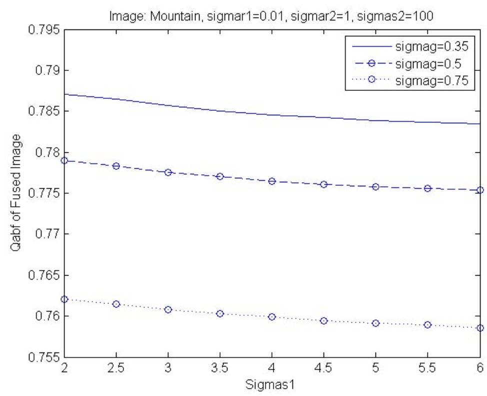
There are five free parameters in the proposed method σ_{r1}, σ_{s1}, σ_{r2}, σ_{s2} and σ_g. σ_{r1} and σ_{s1} decides the operation of recursive filter to extract the low varying features. σ_g controls the averaging operation and is used to extract high varying features. σ_{r2} and σ_{s2} are variables for recursive filtering to smooth the abrupt variations in weights. It is important to select optimum values for all the parameters through experimentation.

In section 5.2 and step – 3 of algorithm, recursive filtering is done to smooth the source images while preserving strong edges. To preserve strong edges, value of σ_{r1} should be on the lower side. Therefore we proposed (0.01 – 0.05) range and selected 0.01 for proposed method. In section 5.4 and step – 5 of algorithm, Recursive Filtering is used to strongly smooth the weights without preserving edges, To implement this, σ_{r2} should be on the higher side, therefore we proposed range of (0.9 – 1) and used 1 for proposed method. Remaining parameters σ_{s1}, σ_{s2} and σ_g are selected by experimentation.

Initial values of standard deviation (range) is set to σ_{r1} = 0.01, σ_{r2} = 1 as per the defined range. Firstly σ_{s1} is analysed for all three selected quality indices by varying its value in range (2 - 6) and keeping other two parameters (σ_{s2}, σ_g) at fixed. We will take two values (50 and 100) of σ_{s2}, and three values (0.35, 0.5 and 0.7) of σ_g to observe σ_{s1}. The results are similar for all images, therefore results of two images, Mountain and Kluki, are only reported here, refer Fig. (5.5(a-l)).



(a)



(b)

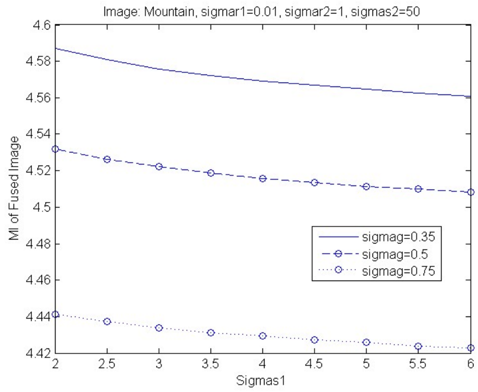


Image: Mountain, sigmar1=0.01, sigmar2=1, sigmas2=50
MI of Fused Image
Sigmas1
sigmag=0.35
sigmag=0.5
sigmag=0.75

(c)

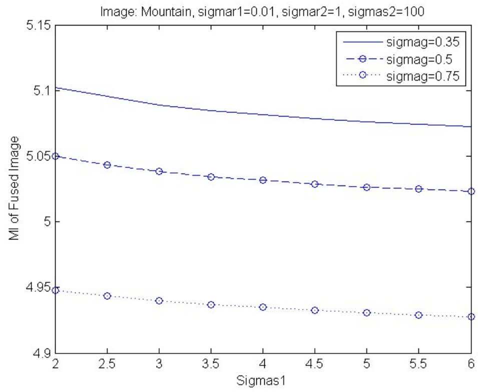


Image: Mountain, sigmar1=0.01, sigmar2=1, sigmas2=100
MI of Fused Image
Sigmas1
sigmag=0.35
sigmag=0.5
sigmag=0.75

(d)

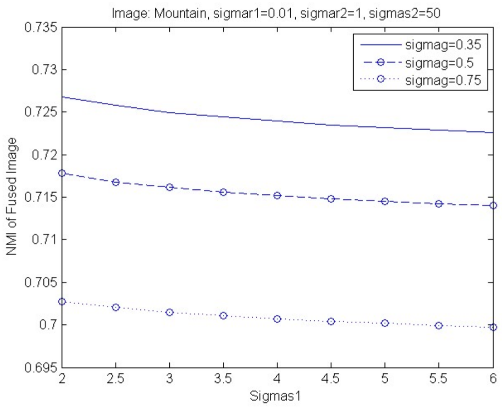


Image: Mountain, sigmar1=0.01, sigmar2=1, sigmas2=50
NMI of Fused Image
Sigmas1
sigmag=0.35
sigmag=0.5
sigmag=0.75

(e)

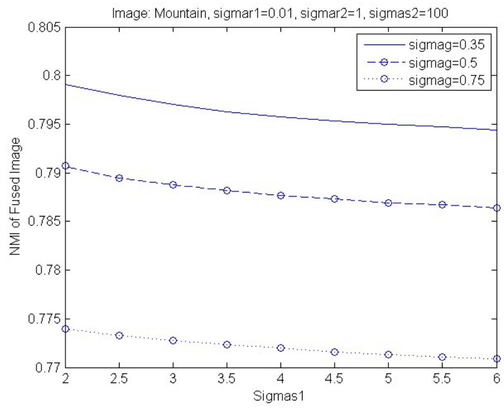


Image: Mountain, sigmar1=0.01, sigmar2=1, sigmas2=100
NMI of Fused Image
Sigmas1
sigmag=0.35
sigmag=0.5
sigmag=0.75

(f)

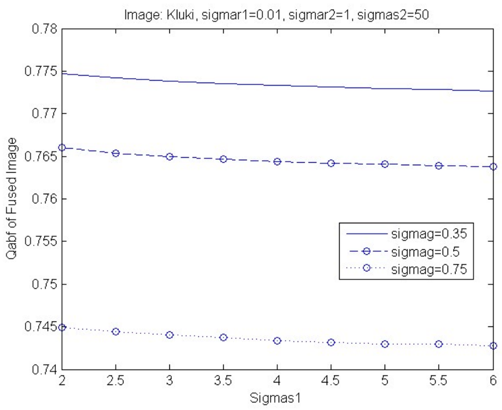


Image: Kluki, sigmar1=0.01, sigmar2=1, sigmas2=50
Qabf of Fused Image
Sigmas1
sigmag=0.35
sigmag=0.5
sigmag=0.75

(g)

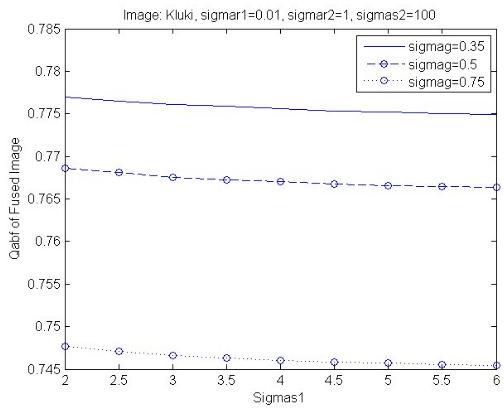
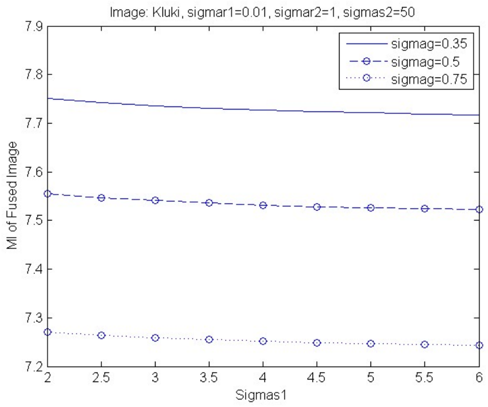
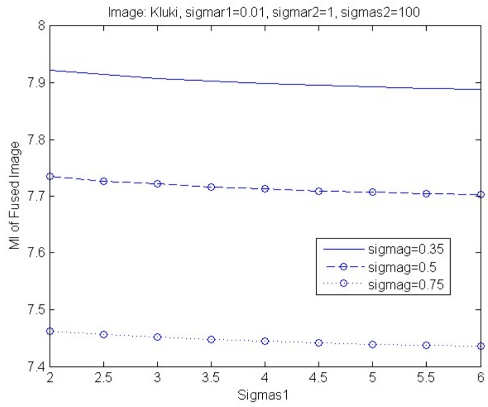


Image: Kluki, sigmar1=0.01, sigmar2=1, sigmas2=100
Qabf of Fused Image
Sigmas1
sigmag=0.35
sigmag=0.5
sigmag=0.75

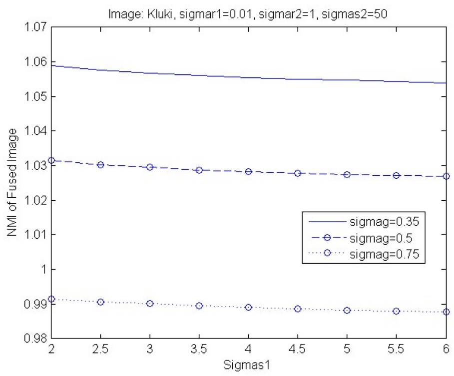
(h)



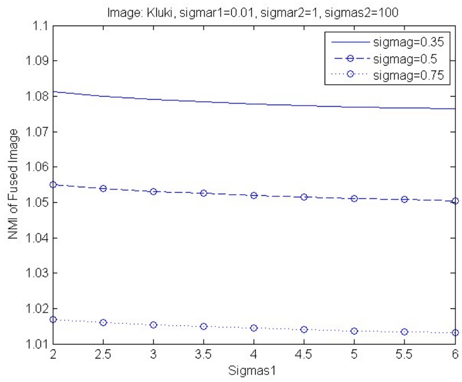
(i)



(j)



(k)

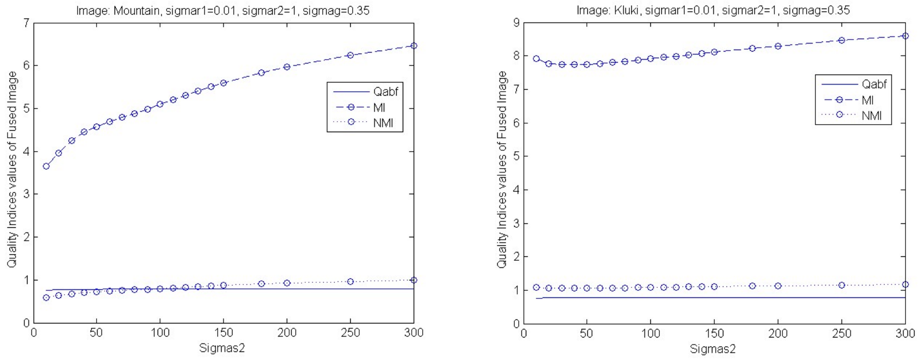


(l)

Fig. 5.5 Graphs **a-b** Image: Mountain, Qabf50, Qabf100, **c-d** Image: Mountain, MI50, MI100, **e-f** Image: Mountain, NMI50, NMI100, **g-h** Image: Kluki, Qabf50, Qabf100, **i-j** Image: Kluki, MI50, MI100, **k-l** Image: Kluki, NMI50, NMI100

For each image, two figures per quality indices are plotted based on the values of σ_{s2}; e.g Fig. 5.5(a-b), Image: Mountain, represents values of $Q^{\frac{AB}{F}}$ w.r.t σ_{s1} and at fixed value of $\sigma_{s2} = 50, 100$. Both figures represents three curves corresponding to three values (0.35, 0.5, and 0.75) of σ_g. Similarly Fig. 5.5(c-l) represents plots of other quality indices. Each figure has three curves for corresponding fixed values of σ_g. All three quality indices ($Q^{\frac{AB}{F}}, MI, NMI$) performed similar and close observation suggest that the value of σ_{s1} is appropriate in the range of (2, 2.5) and σ_g at 0.35. For this proposed method σ_{s1} is fixed at 2.5.

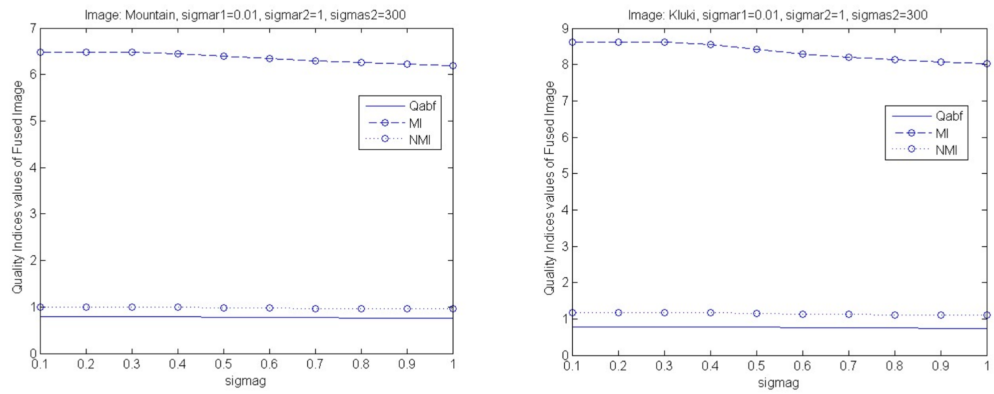
Same procedure is used to determine the range of remaining parameters. For fixed values of $\sigma_{r1} = 0.01$, $\sigma_{r2} = 1$ and $\sigma_{s1} = 2.5$, the remaining two parameters, σ_{s2} and σ_g, are varied and analysed. We keep σ_g at 0.35 and vary σ_{s2} from 10 to 300. Fig. 5.6(a-b) shows the plots for two source images, Mountain and Kluki. Each figure shows three curves corresponding to three quality indices (Q^{AB}_{F}, MI, NMI).



(a) **(b)**

Fig. 5.6 Graph1 **a** Image : Mountain **b** Image : Kluki

The close observation shows that Q^{AB}_{F} remain constant for both images whereas MI and NMI increase substantially. Therefore for better performance, value of σ_{s2} is fixed at 300. Next, σ_g is varied from 0.1 to 1.0 at following parameter values: σ_{r1} = 0.01, σ_{r2} = 1, $\sigma_{s1} = 2.5$ and $\sigma_{s2} = 300$. Fig. 5.7(a-b), show the values of all three quality indices and the appropriate range for σ_g is (0.1 – 0.3). For our algorithm, we use $\sigma_g = 0.2$.



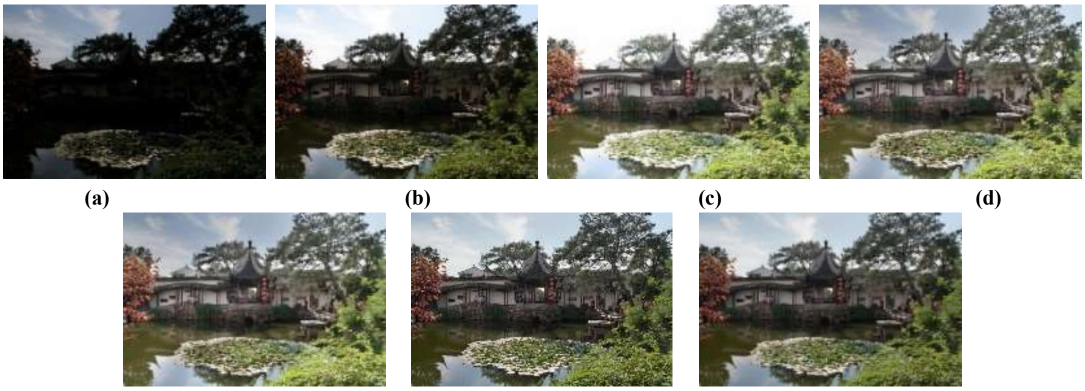
(a) **(b)**

Fig. 5.7 Graph2 **a** Image : Mountain **b** Image : Kluki

5.8.2 Experimental Results:

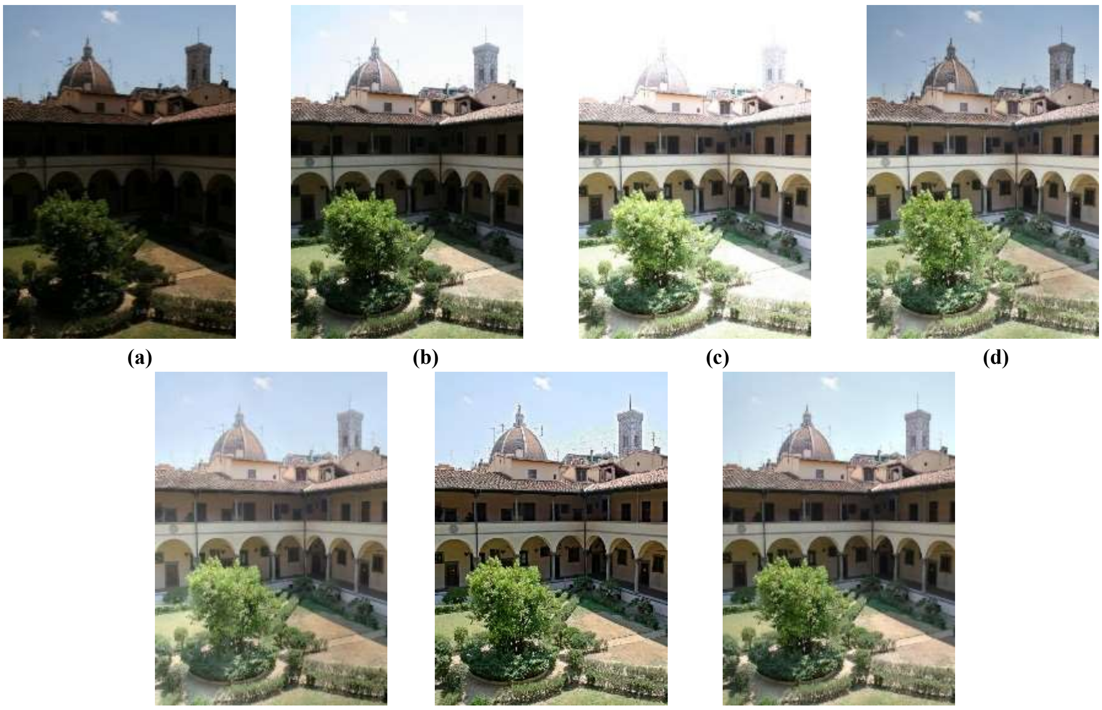
Proposed method is tested on ten random source images and compared with state-of-the-art techniques [42, 66, 68] using selected quality indices. The parameter values selected for implementing proposed algorithm are as follows: σ_{r1} = 0.01, σ_{s1} = 2.5, σ_{r2} = 1, σ_{s2} = 300, σ_g = 0.2. Source images along with fused images of proposed method and state-of-the-art techniques are shown in Figs (5.8-5.17). Quality indices Q^{AB}_{F}, MI and NMI are used to test the quality of fused images. Table 5.1 show the objective evaluation for all fused images.

For source images ‘Garden,’ ‘Building2,’ ‘Garage,’ ‘Landscape,’ ‘Cabin,’ ‘Estate’ and ‘Kluki,’, Proposed method (PM) performs much better as compared to state-of-art techniques. For source images ‘Pond’, ‘Venice’ value of Q^{AB}_{F} is highest for SIFT [66], whereas value of MI and NMI is highest for Proposed Method (PM). For images ‘Cave,’ Q^{AB}_{F} is similar for proposed method, whereas SIFT [66] is better in terms of MI and NMI.



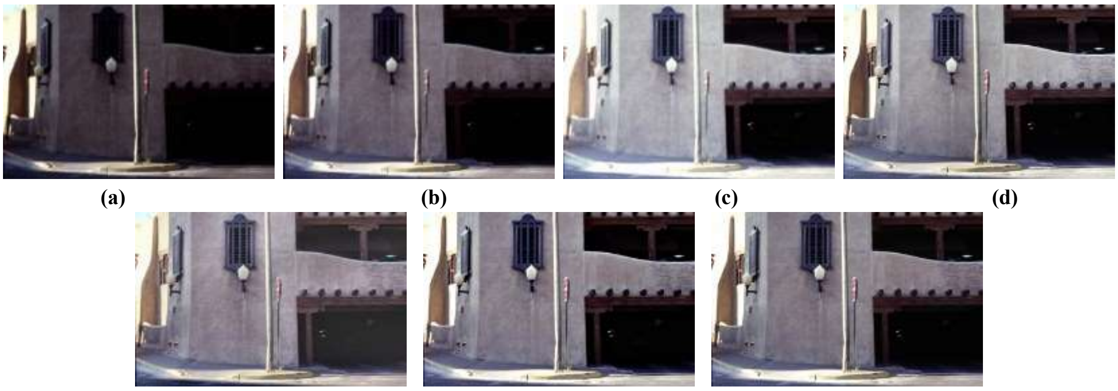
(a) (b) (c) (d) (e) (f) (g)

Fig. 5.8: Image ‘Garden’ and fused images **a-c** Garden, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM



(a) (b) (c) (d) (e) (f) (g)

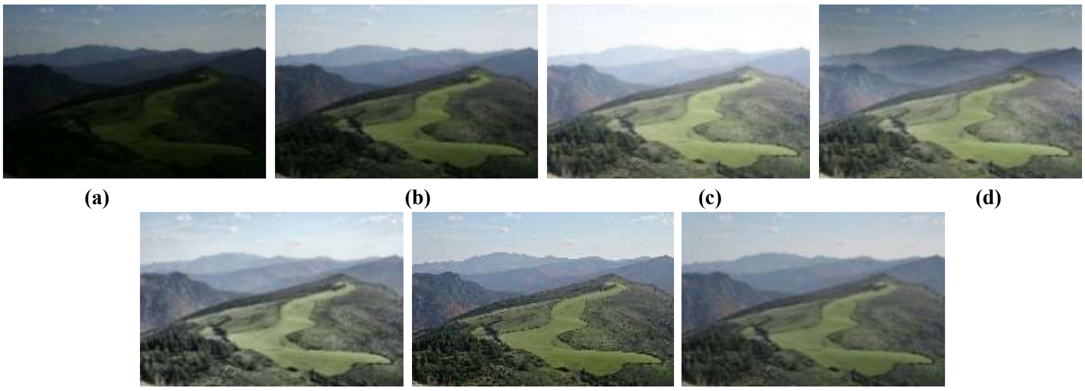
Fig. 5.9 Image ‘Building2’ and fused images **a-c** Building2, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM



(a) (b) (c) (d)

(e) (f) (g)

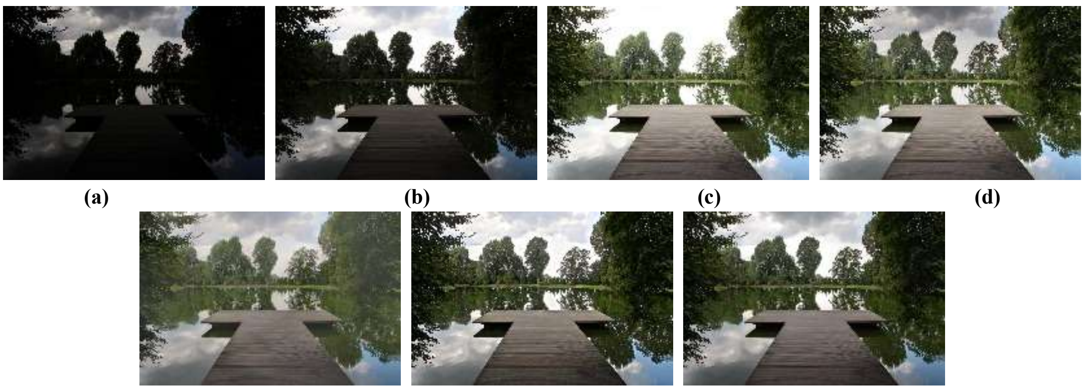
Fig. (5.10) Image 'Garage' and fused images **a-c** Garage, represents source images with increasing exposure time, **d-f** represents fused images for SIFT, MFEF, SPD and PM



(a) (b) (c) (d)

(e) (f) (g)

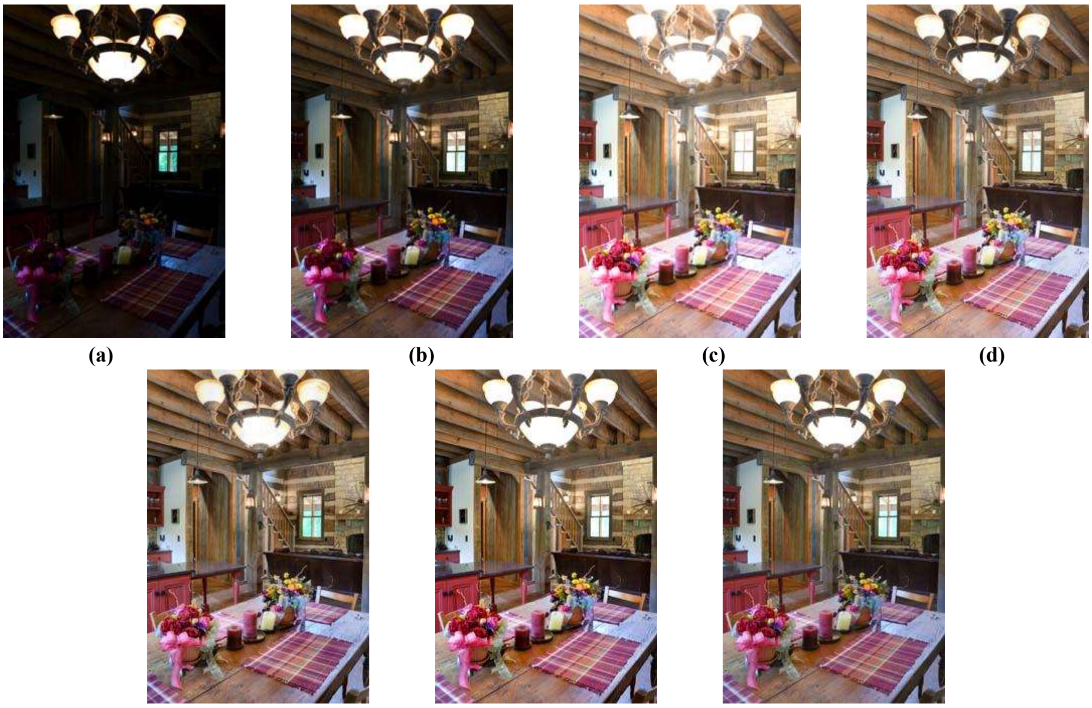
Fig. 5.11 Image 'Landscape' and fused images **a-c** Landscape, represents source images with increasing exposure time, **d-f** represents fused images for SIFT, MFEF, SPD and PM



(a) (b) (c) (d)

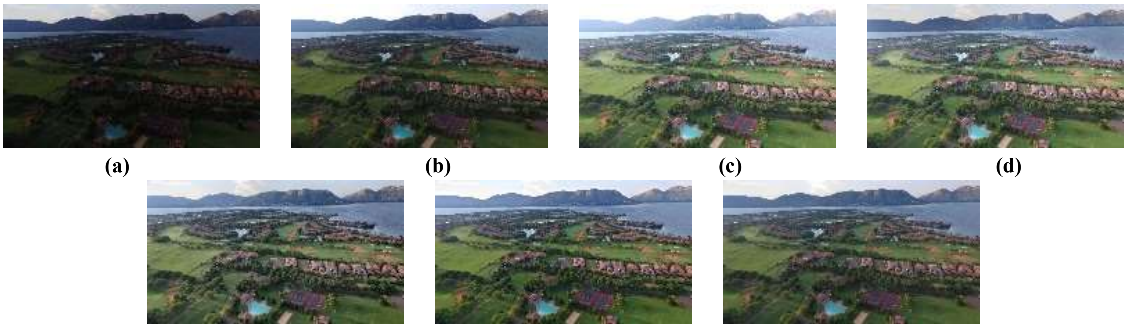
(e) (f) (g)

Fig. 5.12 Image 'Pond' and fused images **a-c** Pond, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM



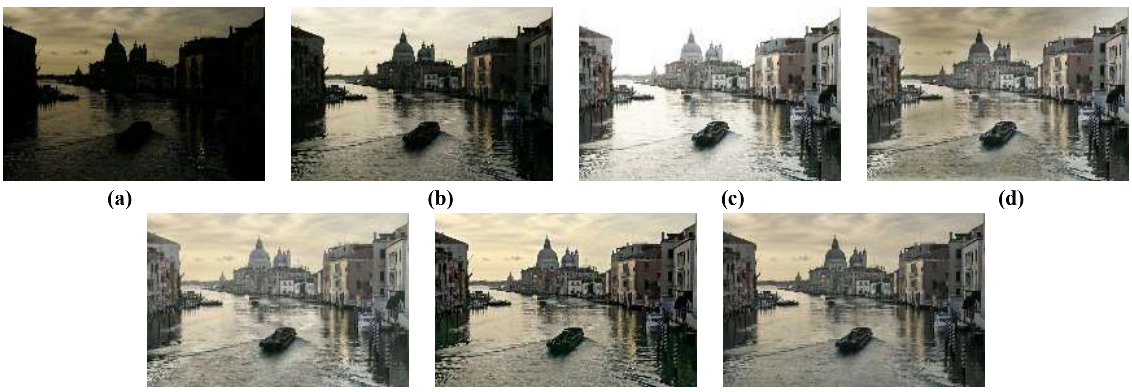
(a) (b) (c) (d) (e) (f) (g)

Fig. 5.13 Image 'Cabin' and fused images **a-c** Cabin, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM



(a) (b) (c) (d) (e) (f) (g)

Fig. 5.14 Image 'Estate' and fused images **a-c** Estate, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM



(a) (b) (c) (d) (e) (f) (g)

Fig. 5.15 Image 'Venice' and fused images **a-c** Venice, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM

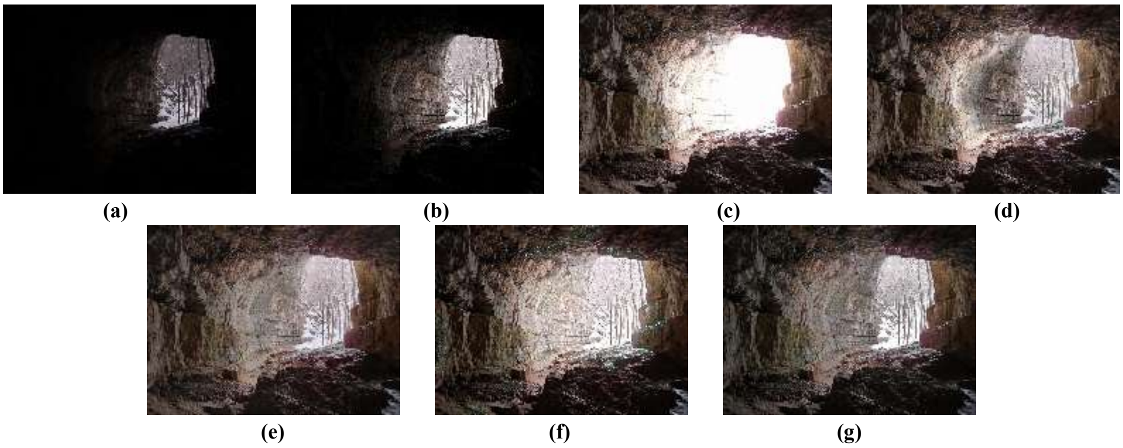


Fig. 5.16 Image 'Cave' and fused images **a-c** Cave, represents source images with increasing exposure time, **d-g** represents fused images for SIFT, MFEF, SPD and PM

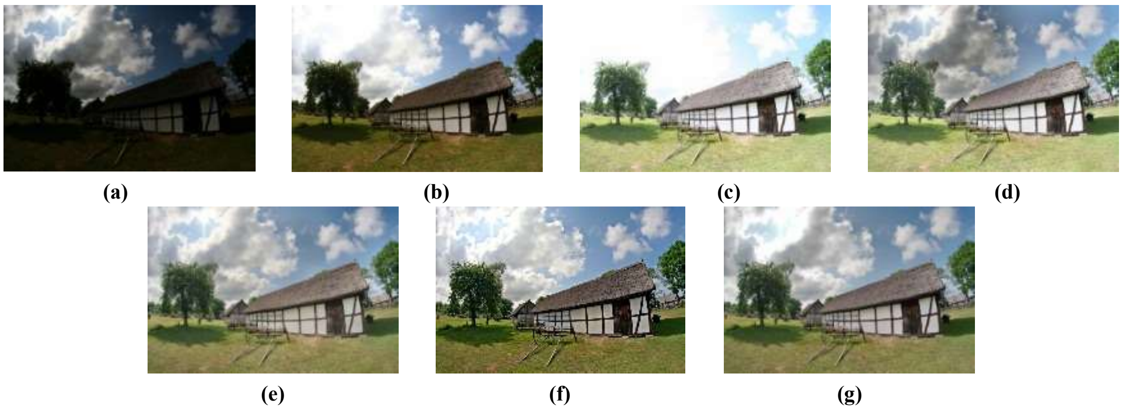


Fig. 5.17 Image 'Kluki' and fused images **a-c** Kluki, represents source images with increasing exposure time, **d-f** represents fused images for SIFT, MFEF, SPD and PM

Table 5.1

Comparison chart of quality indices for feature based fusion

Method	$Q^{AB}/_F$	MI	NMI	Method	$Q^{AB}/_F$	MI	NMI
	Image : Garden				Image : Building2		
SIFT[66]*	0.7272	6.7272	0.9252	SIFT[66]*	0.6675	6.6316	0.9133
MFEF[42]*	0.7067	4.9798	0.6918	MFEF[42]*	0.5901	5.4878	0.7550
SPD[68]*	0.6585	7.3417	1.0112	SPD[68]*	0.5876	8.0833	1.0949
PM*	0.7419	8.4378	1.1719	PM*	0.6669	9.5835	1.3021
	Image : Garage				Image : Landscape		
SIFT[66]*	0.7736	7.9364	1.1122	SIFT[66]*	0.7012	6.8049	0.9748
MFEF[42]*	0.7220	6.8623	0.9717	MFEF[42]*	0.6213	7.6463	1.0834
SPD[68]*	0.6866	8.2131	1.1842	SPD[68]*	0.5354	9.9376	1.3662
PM*	0.8013	9.9918	1.4434	PM*	0.7447	11.3580	1.5752
	Image : Pond				Image : Cabin		
SIFT[66]*	0.6444	7.4259	1.0184	SIFT[66]*	0.7390	6.5897	0.8961
MFEF[42]*	0.4245	5.1156	0.7395	MFEF[42]*	0.6895	5.9183	0.8101
SPD[68]*	0.4252	6.6880	0.9366	SPD[68]*	0.6993	7.3498	1.0043
PM*	0.6143	7.8534	1.1057	PM*	0.7670	8.0960	1.1279
	Image : Estate				Image : Venice		
SIFT[66]*	0.7092	9.8430	1.3858	SIFT[66]*	0.5938	4.4510	0.6144
MFEF[42]*	0.5635	6.2270	0.8695	MFEF[42]*	0.5359	5.1896	0.7139
SPD[68]*	0.5835	8.5968	1.2043	SPD[68]*	0.5129	6.6766	0.8967
PM*	0.7501	10.4996	1.4958	PM*	0.5317	7.1520	0.9697
	Image : Cave				Image : Kluki		
SIFT[66]*	0.7544	7.4686	1.0804	SIFT[66]*	0.7444	6.0114	0.8229
MFEF[42]*	0.7220	4.9983	0.7620	MFEF[42]*	0.7431	6.5565	0.8940
SPD[68]*	0.5982	5.4493	0.8190	SPD[68]*	0.6264	8.3529	1.1171
PM*	0.7569	6.7519	1.0250	PM*	0.7790	8.6159	1.1820

*PM – Proposed Method, MFEF - Multi Focus and Exposure Fusion, SPD – Structural Patch Decomposition, SIFT – Scale Invariant Feature Transform

Chapter – 6

Region based multi-exposure image fusion

6.1 Introduction

In this chapter we propose another fusion technique to enhance dynamic range. The proposed technique uses region based technique to synthesize fused image. Pre-registered source images with varying exposure are combined to enhance the dynamic range.

The lag in dynamic range of LDR camera images is due to under or over exposed conditions. These conditions depend on shutter speed of the LDR camera. Shutter speed is the time for which shutter remains open and allow the reflected waves to pass through it. This time is also known as exposure time and is inversely related to shutter speed. High shutter speed of camera means small exposure time, which results in underexposed conditions in some sections of image. Similarly low shutter speed results in overexposed conditions in some sections of image. These under or over exposed conditions on one hand deteriorate some part of the image and on the other hand capture quality information of remaining part. Fig. 5.1 shows three images with different exposure time. First image, Fig. 5.1(a), clearly captures quality information of the sky, while losing other. Similarly, Fig. 5.1(c) captures quality information of constructed portion but losses information of the sky. Fig. 5.1(b) captures information in between for both the regions. The aim is to transfer informative sections from different images to construct fused image.

(a) (b) (c)

Fig. 5.1 Source images 'garden' **a.** low exposure time, **b.** middle exposure time, **c.** high exposure time. (reproduced from chapter 5)

6.2 Division of scene into regions

The scene to be captured through multiple images is divide into three regions. The identification of regions is done by using pixel intensity of captured images. The captured image contains three types of sections: darker pixels, brighter pixels, and pixels with intensities between darker and brighter. The image captured at low exposure time darkens the darker section of scene. The intensity of such pixels is close to ‘0’ and described as darker pixels. Similarly high exposure time image brighten up the brighter section of scene and intensity is close to ‘255’. These pixels are described as brighter pixels. Fig.5.1 represents three images with exposure time in increasing order. Fig. 5.1(a) is captured at low exposure time and has large sections with pixel intensity close to 0. This area is known as under-exposed area. Similarly, Fig. 5.1(c) is captured at high exposure time and has large number of pixels having intensity close to 255 (for 8-bit image), and is known as over exposed area.

The under and over exposed areas keeps pixel intensity flat and results in loss of underlying information. In order to target these sections, we divide the scene (through captured images) into three regions. Two regions target under and over exposed sections, whereas third region covers remaining section. The pixel intensity of source images are utilized to define these regions. These regions are termed as black region (B_r), white region (W_r) and grey region (G_r). Black region (B_r) cover pixels from lowest exposer time image (refer Fig. 5.1(a)) with intensity value below ‘L’. White region (W_r) covers pixels from highest exposer time image (refer Fig. 5.1(c)) with intensity above ‘255-L’ (for 8-bit image). Grey region (G_r) covers remaining pixels of scene having intensity in-between ‘L’ and ‘255-L’. ‘L’ is a threshold value and varies from 10-30, refer [57].

The selection of ‘L’ is important in separating B_r and W_r from the source images. Many researcher [57] have used the threshold value in range (10-30) in order to remove the effect of under and over exposed pixels. We also use the same range for ‘L’. Eq. (66-68) defines the division of scene into three regions by using the pixel intensity of source images. Each region is represented by white pixels, refer Fig 6.1(a-c).

Let ϕ_i be the set of pre-registered N source images of size $P \times Q$, where $i \in 1 \dots . N$. The exposure time to capture these images are in increasing order; e.g. ϕ_1 and ϕ_N are captured at lowest and highest exposure time respectively.

$$B_r(p,q) = \begin{cases} 1, & if\ \phi_1(p,q) \leq L \\ 0, & else \end{cases} \quad (66)$$

$$W_r(p,q) = \begin{cases} 1, & if\ \phi_N(p,q) \geq 255 - L \\ 0, & else \end{cases} \quad (67)$$

$$G_r\ (p,q) = \begin{cases} 1, & if\ B_r(p,q) \neq 1\ and\ W_r(p,q) \neq 1 \\ 0, & else \end{cases} \quad (68)$$

where (p, q) represents pixel location.

Closer observation of Fig. (6.1(a-c)) suggest that the white region is not continuous and essential condition for good image fusion. We use morphological operation [1] to make the white region continuous or to create blob like structure. The structuring element (SE), a matrix of defined size and shape, is used to perform these operations.

The opening operation is performed as erosion followed by dilation on target image. It shrinks the foreground pixels (usually white pixels), refer Eq. (69). The closing operation is done as dilation followed by erosion on target image. It is best suited to fill the black spots inside the white section. It enlarges the boundaries of foreground pixels (usually white pixels), refer Eq. (70).

Opening and closing operation for any image ϕ_i is as follows

$$\phi_i \circ SE = (\phi_i \ominus SE) \oplus SE \quad (69)$$

$$\phi_i \bullet SE = (\phi_i \oplus SE) \ominus SE \quad (70)$$

$$SE = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (71)$$

where '∘' and '•' represents opening and closing morphological operation.

We have used a closing operation to make the region continuous with SE as square matrix of size 5×5, refer Eq. (71). The processed regions ((B_r), (W_r) and (G_r)) are shown in Figs. (6.1(d-f)).

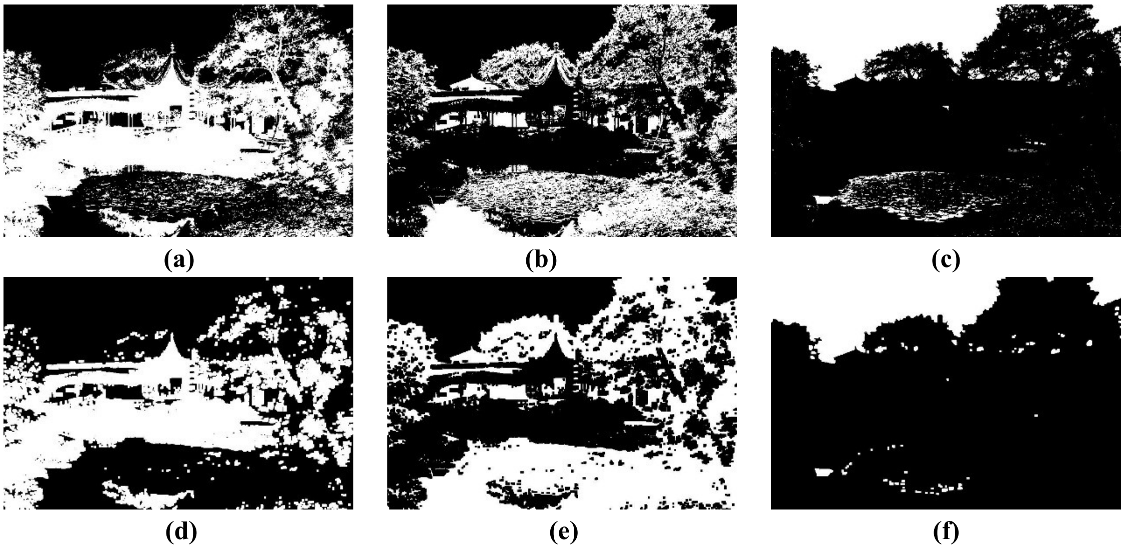


Fig. 6.1 Three regions represented by white pixels **a.** Black region B_r, **b.** Grey region G_r, **c.** White region W_r, **d.** Morphological Black region, **e.** Morphological Grey region, **f.** Morphological White region (SE = 5 by 5)

6.3 Frequency and base layer

The captured source images are splitted into two parts: frequency and constant component image. Frequency components are the intensity variations of pixels whereas constant component covers base information. There are many filters available in literature to extract frequency components like Laplacian filter. Edge aware filters, like bilateral filter [86] and anisotropic diffusion [87], are also useful in extracting frequency. Filters like Laplacian has fixed set of kernels, whereas in edge aware filters value of kernels can be varied. Thus edge aware filters are more useful than fixed kernel filters.

Bilateral filter and anisotropic diffusion are the two basic edge aware filters. Bilateral filter uses the distance and intensity information to extract frequency components. Whereas, partial differential equations are used to model anisotropic diffusion and implemented as an iterative process to separate frequency components. Recursive Filter (RF) [85] is another edge aware filter and works on the principle of bilateral filter.

We use Recursive Filter (RF) given its advantages like speed, direct operation on color images. It has two variable parameter: deviation in space (σ_s) and range (σ_r). The deviation in space (σ_s) uses the distance between the pixels, whereas deviation in range (σ_r) uses the pixel intensity to perform the blurring while preserving edges. Parameters σ_s and σ_r decide the level of blurring and preservation of edges. Specifically σ_r can be used to extract intensity variations (frequency) of an image by varying its value.

To extract intensity variations, value of σ_r is decremented from '1' (for image intensity normalized between [0, 1]) by predefined factor and information is extracted for each step; e.g., source image 'X' is processed for values of σ_r = 1 and 0.9 using RF resulting in images Y and Z respectively. The differences X-Y and Y-Z, allows to extract the intensity variations lying in that range. Information in terms of intensity variations are extracted by gradually decrementing the value of σ_r. The extracted information is combined together as frequency image. Further to exhaust the available intensity variations in last processed image, it is processed with well-known Gaussian filter. Deviation (σ_g) is the controlling parameter. The intensity variations are separated from Gaussian processed image and combined with frequency image, refer (72-74). The combined frequency image is defined as frequency layer and remaining image is defined as base layer.

Fig. 6.2, shows the frequency layer and base layer for source image 'garden', refer Fig. 5.1. Three parameters σ_s, σ_g and step size ($step$), are the free parameters whose values need to be fixed. Following pseudo code represents the construction of frequency (HF_i) and base layer (LF_i) by setting values of σ_s, σ_g and step size ($step$) as 0.5, 1 and 0.05 respectively. Let ϕ_i be the set of pre-registered N source images of size $P \times Q$, where $i \in 1 \dots . N$.

$\sigma_s = 0.5;\ \sigma_r = 1;\ step = 0.05;\ HF_i = 0;$
$Loop: r = 1:1:(1/step)$
$\quad D_i = RF(\phi_i, \sigma_s, \sigma_r);$
$\quad\quad if\ r = 1;\ HF_{if} = \phi_i - D_i;$
$\quad\quad else;\ HF_{if} = D1_i - D_i;$
$\quad\quad end$
$\quad D1_i = D_i;$
$\quad \sigma_r = \sigma_r - step;$
$\quad HF_i = HF_{if} + HF_i;$
End

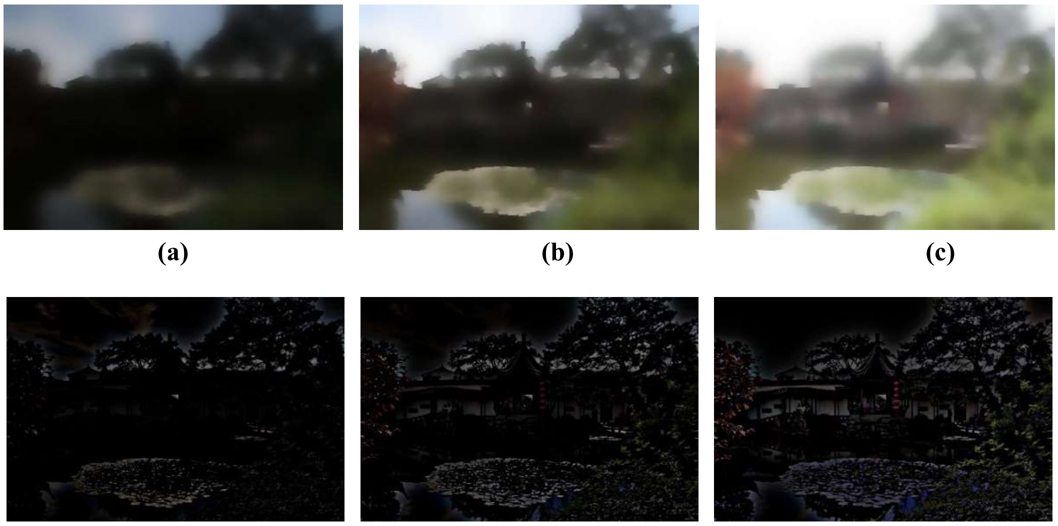
$G_i = gaussian(D_i; \sigma_g);$

$$HF_i = (D_i - G_i) + HF_i; \tag{72}$$

$$\phi_i = HF_i + LF_i;\quad i = 1,2 \dots, N \tag{73}$$

$$LF_i = \phi_i - HF_i \tag{74}$$

where HF_i, LF_i are frequency and base layer of i^{th} image, ϕ_i is i^{th} source image. $RF(.)$ and $gaussian(.)$ are recursive and Gaussian filtering respectively.



(a) (b) (c)

(d) (e) (f)

Fig. 6.2 Frequency and base layer **a-c.** Constant layer for source images, Fig. 5.1(a-c), **d-f.** Frequency layer for source images, Fig. 5.1(a-c).

6.4 Dense SIFT

Scale Invariant Feature Transform (SIFT) was first introduced by Lowe [88]. The method extracts large number of unique features efficiently by using cascade filtering approach. In first stage images are processed with difference of Gaussian (DoG) and stable features are extracted across all scales. In second stage weak features among the selected features are rejected. These features are described as key-point descriptors. Third stage extracts gradient and eight directions orientation information of neighbor area around each feature. For example, a region of 16×16 neighboring pixels divided into cell array of 4×4. Each cell is a group of 4×4 pixel. The orientations of each cell is quantized into 8 bins, resulting in $128 (= 4 \times 4 \times 8)$ dimensional feature vector per key point descriptor.

These key-point descriptors holds the gradient information of its neighborhood area and very useful in testing the activity level of pixel or region. However non-uniformity of key-point descriptors restricts its use in image fusion. Dense SIFT [89] is a SIFT based feature extraction method but unlike to SIFT consider all pixels as a key-point descriptors. This makes Dense SIFT useful in testing the activity level of pixel or region. The gradient information is calculated for all descriptors as per SIFT. Each pixel has gradient information in term of $128 (= 4 \times 4 \times 8)$ dimensional feature vector (f_r) per pixel, refer Eq. (75). All these features per pixel are added together to represent the activity level of each pixel (F_p), refer (76).

$$f_r(p,q) = [f_0 \dots f_{127}] \tag{75}$$

$$F_p(p,q) = \sum_{r=0}^{127} |f_r| \tag{76}$$

where (p, q) represents the pixel co-ordinate.

6.4.1 Activity Measurement of Regions

Dense SIFT calculates the gradient information of all key-point descriptors (pixels), refer Eq. (76). To use it for region, features per pixel (F_p) for all pixels are combined together to reflect the activity level of region. This information is defined as feature value (FV). The proposed method has divided the scene into three regions (B_r, W_r, G_r), refer section (6.2). The feature value (FV) of all three regions for i^{th} image is as under

$$FV_{Bi} = \sum_{Region\ B} F_p \ ; i \in [1, N]$$

$$FV_{Wi} = \sum_{Region\ W} F_p \ ; i \in [1, N]$$

$$FV_{Gi} = \sum_{Region\ G} F_p \ ; \ i \in [1, N] \tag{77}$$

where $FV_{Bi}, FV_{Wi}, FV_{Gi}$ are the activity levels of black, white and grey regions of i^{th} source image, respectively.

6.5 Fusion of Frequency Layer

All source images (ϕ_i) are splitted into frequency (HF_i) and base layer (LF_i), refer Eq. (72-74). Frequency layers are fused together using region based fusion. Three regions are defined as black region (B_r), white region (W_r) and grey region (G_r), refer Eq. (66-68). All frequency layers are tested for activity level over these three regions. For each defined region, the region with maximum feature value (FV) from respective frequency layer is selected for fusion, refer (78-80).

$$HL_B = (HF_i.* B_r) \max_{1 \le i \le N} FV_{Bi} \tag{78}$$

$$HL_W = (HF_i.* W_r) \max_{1 \le i \le N} FV_{Wi} \tag{79}$$

$$HL_G = (HF_i.* G_r) \max_{1 \le i \le N} FV_{Gi} \tag{80}$$

where '.*' represents pixel to pixel multiplication.

$HF_i.* B_r$ represents the separation of black region from respective frequency layers (HF_i) and similarly for white and grey regions.

The selected three regions from frequency layers are combined together on pixel to pixel basis, refer Eq. (81).

$$HL_{fused} = HL_B + HL_W + HL_G \tag{81}$$

6.6 Fusion of base layer

Base layer (LF_i) is the remaining information of source images after separating frequency layer (HF_i) in terms of intensity variations. The base layer usually holds very less or no variations in intensity. The best way to fuse base layers is to fuse them as average of all base layers. Weighted sum is another way to combine base layers in which weights are allotted to all base layer pixels. We have tested both strategy and defined in following equations

$$BL_{fused,1} = \sum_i W_i . BL_i \, ; i \in [1, N] \tag{82}$$

$$W_i = \frac{FV_i}{\sum_i FV_i} \tag{83}$$

where $FV_i = FV_{Bi} + FV_{Wi} + FV_{Gi}$ and W_i is the weight allotted to respective base layer.

$$BL_{fused,2} = (\frac{1}{N}) \sum_i BL_i \, ; i \in [1, N] \tag{84}$$

$BL_{fused,1}$ and $BL_{fused,2}$ are the fused base layer by using weighted sum and average sum respectively.

6.7 Fused Image

The fused frequency layer (FL_{fused}) and base layer (BL_{fused}) are combined together on pixel-to-pixel basis. Based on type of base layer used, proposed method construct two fused image, refer Eq. (85-86).

$$I_{fused,1} = FL_{fuse} \;\; + BL_{fused,1} \tag{85}$$

$$I_{fused,2} = FL_{fused} + BL_{fused,2} \tag{86}$$

6.8 Algorithm

This section presents the step by step operation of proposed multi-exposure fusion technique followed by the algorithm. Fig. 6.3 represents the flow of algorithm.

The process of proposed technique is as follows

Step 1: Let ($\phi_i; i \in [1, \mathrm{N}]$) be the pre-registered source images of size $P \times Q$.

Step 2: Divide scene into three regions (B_r, W_r and G_r) using pixel intensity, refer section 6.2.

Step 3: Split each source image into frequency (HF_i) and base (LF_i) layer, refer section 6.3.

Step 4: Calculate activity level for each region (FV) using Dense SIFT, refer section 6.4.

Step 5: Compute fused frequency layer, refer section 6.5.

Step 6: Compute fused base layer, refer section 6.6.

Step 7: Compute final fused image by combining fused frequency and base layer, refer section 6.7.

Algorithm: Region based multi-exposure fusion

Capture pre-registered images of scene at different exposure time.

Input:

Pre-registered color source images ($\phi_{i,RGB}: i \in [1, \mathrm{N}]$);

Begin

1. Construction of three regions (B_r, W_r and G_r) using pixel intensity of source images

$$B_r(p,q) = \begin{array}{l} 1,\ if\ \phi_1(p,q)\ \leq L \\ 0,\ else \end{array}$$

$$W_r(p,q) = \begin{array}{l} 1, if\ \phi_N(p,q) \geq 255 - L \\ 0,\ else \end{array}$$

$$G_r\ (p,q) = \begin{array}{l} 1,\ if\ B_r(p,q) \neq 1\ and\ W_r(p,q) \neq 1 \\ 0,\ else \end{array}$$

These three regions are processed with morphological closing operation to create blob like structure. Colour images are converted into grey images for region identification.

2. Split each image into frequency (HF_i) and base layer (LF_i) using Recursive filter

$$\phi_i = HF_i + LF_i\ ; \quad i = 1,2 \ldots, N$$

3. Calculate feature value (FV) of regions using features per pixel $\left(F_p\right)$. Dense SIFT is used to calculate F_p

$$FV_{Bi} = \textstyle\sum_{Region\ B} F_p\ \ ; i \in [1, N]$$

$$FV_{Wi} = \textstyle\sum_{Region\ W} F_p\ ; i \in [1, N]$$

$$FV_{Gi} = \textstyle\sum_{Region\ G} F_p\ ;\ i \in [1, N]$$

4. Compute region based frequency layer fusion

$$HL_B = (HF_i.*\ B_r) \max_{1 \leq i \leq N} FV_{Bi}$$

$$HL_W = (HF_i.*\ W_r) \max_{1 \leq i \leq N} FV_{Wi}$$

$$HL_G = (HF_i.*\ G_r) \max_{1 \leq i \leq N} FV_{Gi}$$

$$HL_{fused} = HL_B + HL_W + HL_G$$

5. Compute fused base layer using weighted and average sum

$$BL_{fused,1} = \sum_i W_i . BL_i \, ; i \in [1, N]$$

$$BL_{fused,2} = (\frac{1}{N}) \sum_i BL_i \, ; i \in [1, N]$$

6. Compute fused image using fused frequency and base layer

$$I_{fused,1} = FL_{fused} + BL_{fused,1}$$

$$I_{fused,2} = FL_{fused} + BL_{fused,2}$$

End

Output:

Color fused image ($I_{fused,RGB}$);

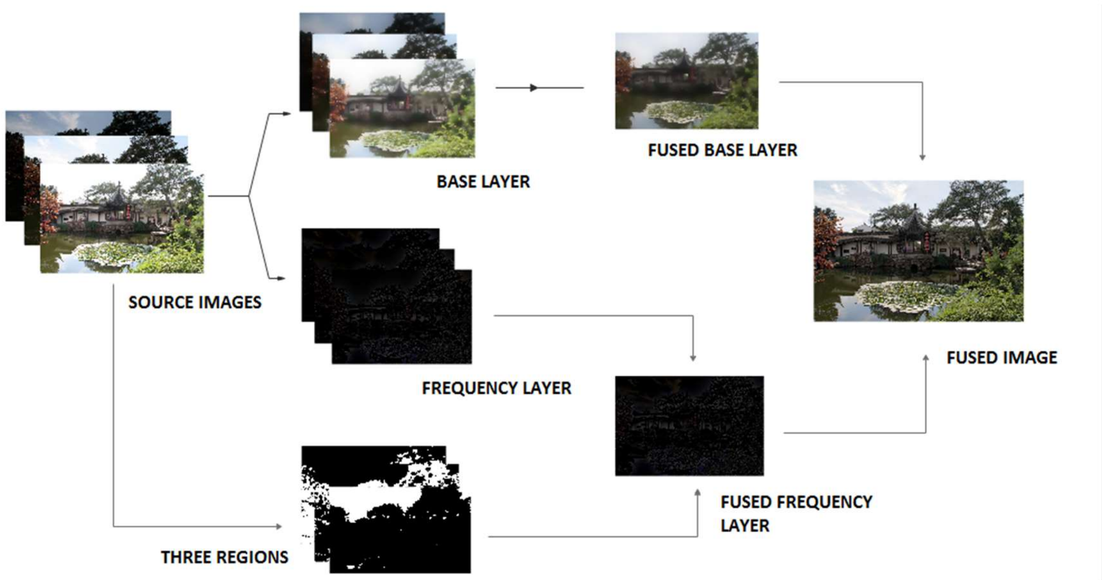


Fig. 6.3 Flow of Algorithm

6.9 Result and Analysis

Proposed method is experimented on 14 random set of images. Performance of proposed method is evaluated by using objective evaluation technique. Various objective indices are available and discussed in chapter 3. Four indices selected to perform the qualitative analysis are Mutual Information (MI) [74], Normalized Mutual Information (NMI) [75], Edge-based Similarity Index ($Q^{AB/F}$) [76] and Structure similarity-based metric (SSIM) [77]. MI and NMI evaluates probability based information. $Q^{AB/F}$ evaluates edges transferred from source images to fused image. SSIM tested the structural information of fused image.

The state-of-the-art selected for comparing proposed method are gradient domain based (MFEF) [42], patch based (SPD) [68], block based (BB1) [90] and recursive filter (RG) [91]. MFEF [42] performs fusion in gradient domain and processed by Haar wavelet to reconstruct fused image. SPD [68] fused images using three components: signal strength, signal structure and mean intensity. BB1 [90] is block based fusion using local and global layer. RG [91] is feature based fusion technique.

6.9.1 Parameter Selection:

There are five free parameters for this proposed method: L, structuring element (SE) size, step size ($step$), σ_s, and σ_g. 'L' plays a role in dividing scene into three regions using source images. The range for 'L' is '10-30' [57]. Since, it has been observed that the performance remain same in the specified range therefore we have selected '10' for our algorithm. Structuring element (SE) is a matrix of size 3×3 or 5×5 with value '1'. We have adopted 5×5 size for structuring element.

Step size ($step$) defines the value by which σ_r is decremented to separate frequency components from source images. We have selected '0.05' value, sufficient small value to give sufficient number of steps for separating frequency components. Remaining two variables σ_s, σ_g are responsible for blurring the image and to keep blurring at moderate level we have selected their values as 0.5, 0.35 respectively.

6.9.2 Experimental Results:

Experimentation is performed on randomly selected images using defined values of parameters. Objective evaluation is done by comparing their results with state-of-the-art techniques. Quality indices $Q^{AB}/_{F}$, MI, NMI and SSIM are used to test the quality of fused images. Fig. (6.4-6.17) shows the results of selected source images. Table 6.1 show the objective evaluation for all fused images.

The proposed method has two strategies based on base layer fusion and termed as PM-I and PM-II. PM-II represents the use of weighted sum for base layers whereas PM-I represents the use of average sum of base layers. The comparison is performed by first comparing PM-II with state-of-the-art and then PM-II is compared with PM-I.

Table 6.1 shows that performance of PM-II, in comparison with state-of-art, is better for following images: Garden, Estate, Cabin, Landscape, Mountains, Sunset, Building2, Garage, Pond and Kluki. For remaining images, performance is similar except for $Q^{AB/F}$. The comparison between PM-I and PM-II by referring Table 6.1 suggest that PM-II performed better than PM-I on the basis of MI, NMI, and SSIM. Whereas in terms of $Q^{AB/F}$ PM-I performed better.

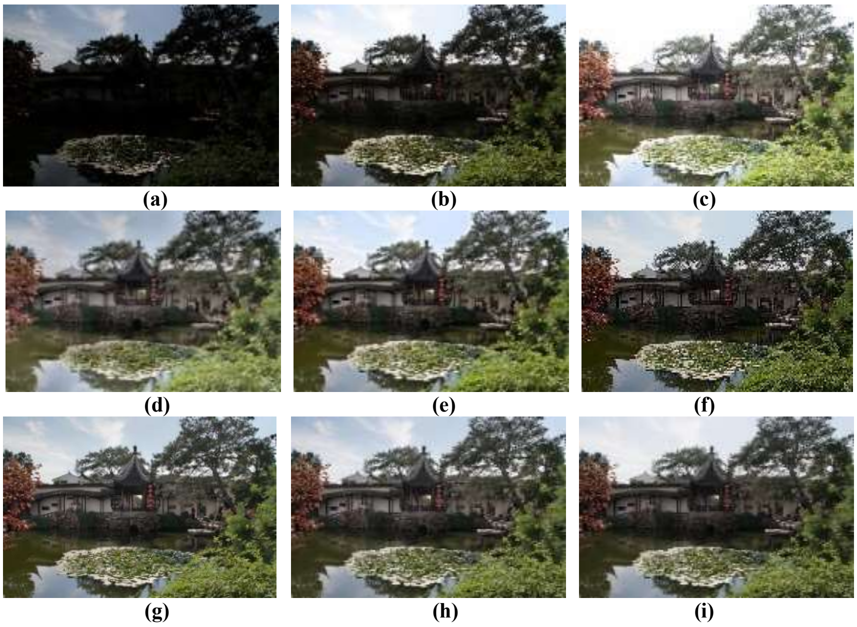


Fig. (6.4) Image 'Garden' and fused images **a-c** Garden, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

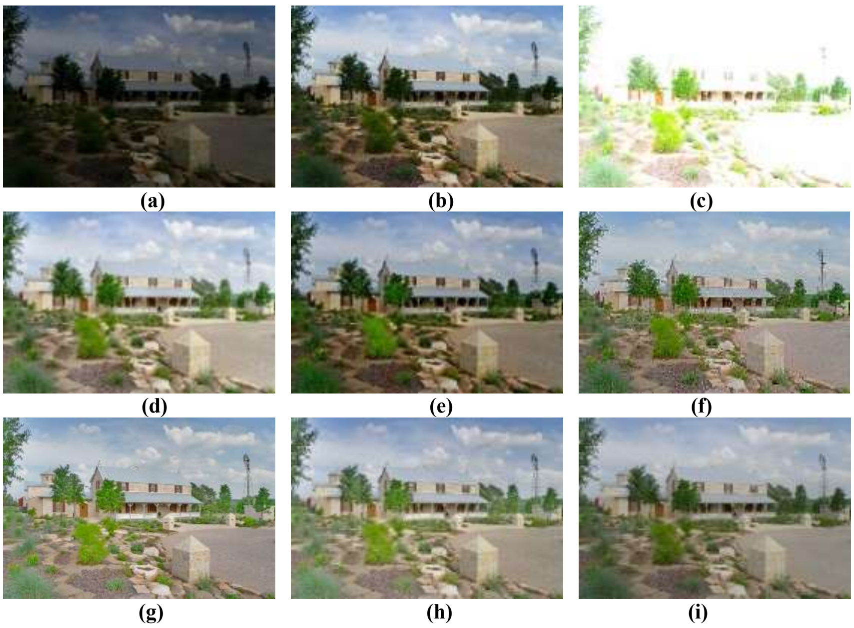


Fig. (6.5) Image 'Gasmire' and fused images **(a-c)** Gasmire, represents source images with increasing exposure time, **(d-i)** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

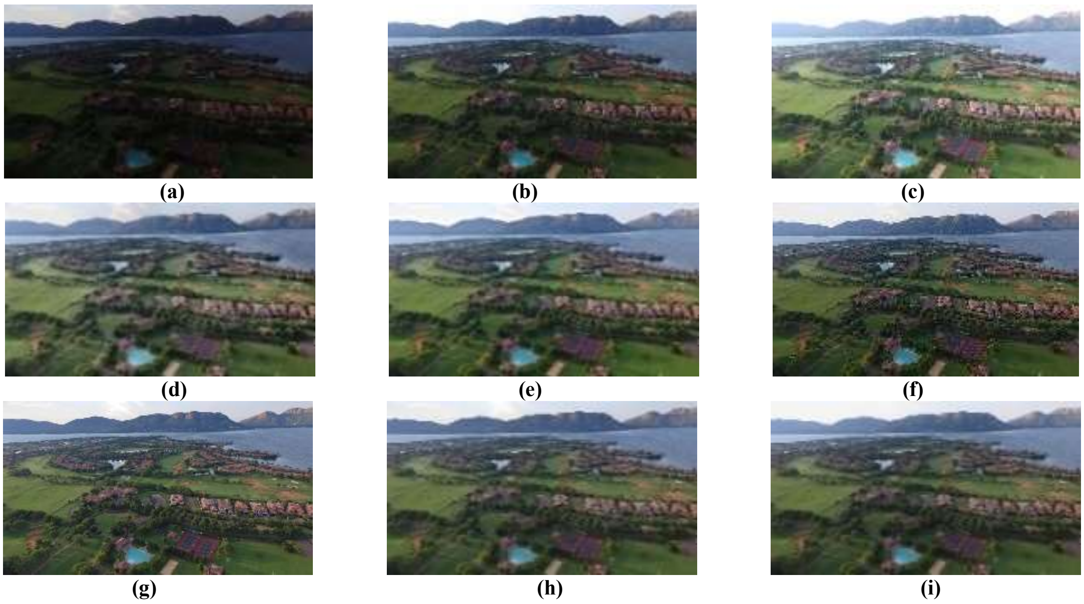


Fig. (6.6) Image 'Estate' and fused images **a-c** Estate, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

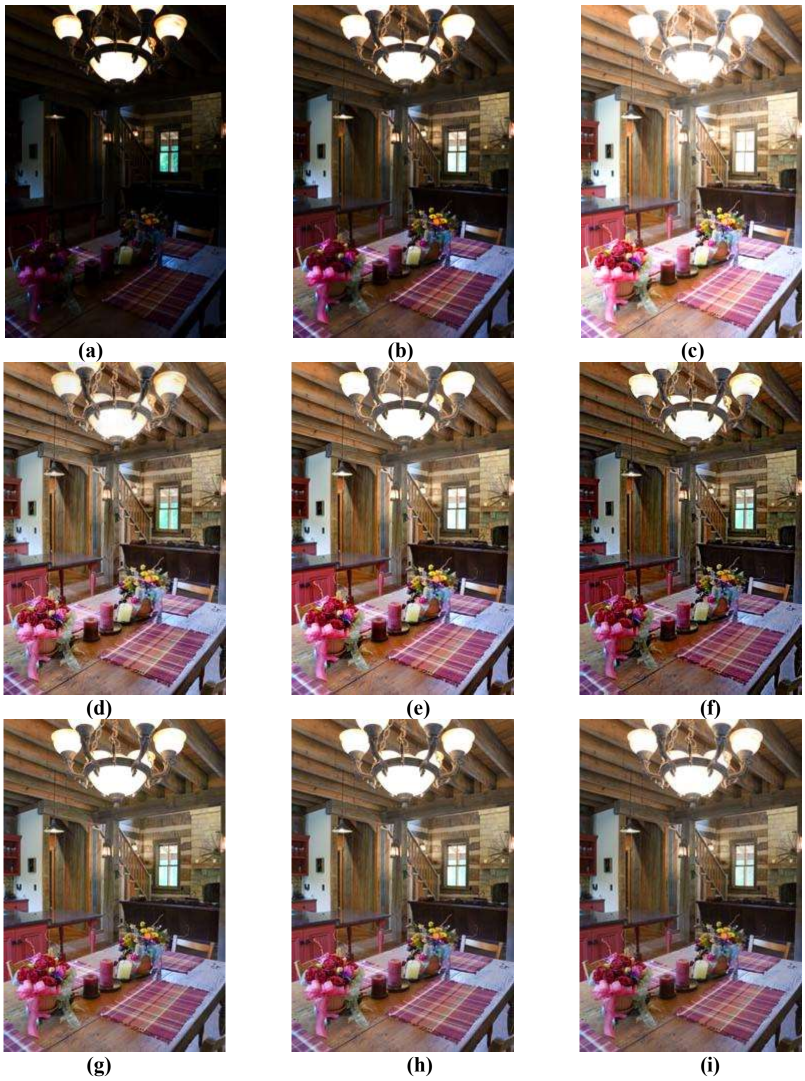


Fig. (6.7) Image 'Cabin' and fused images **(a-c)** Cabin, represents source images with increasing exposure time, **(d-i)** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

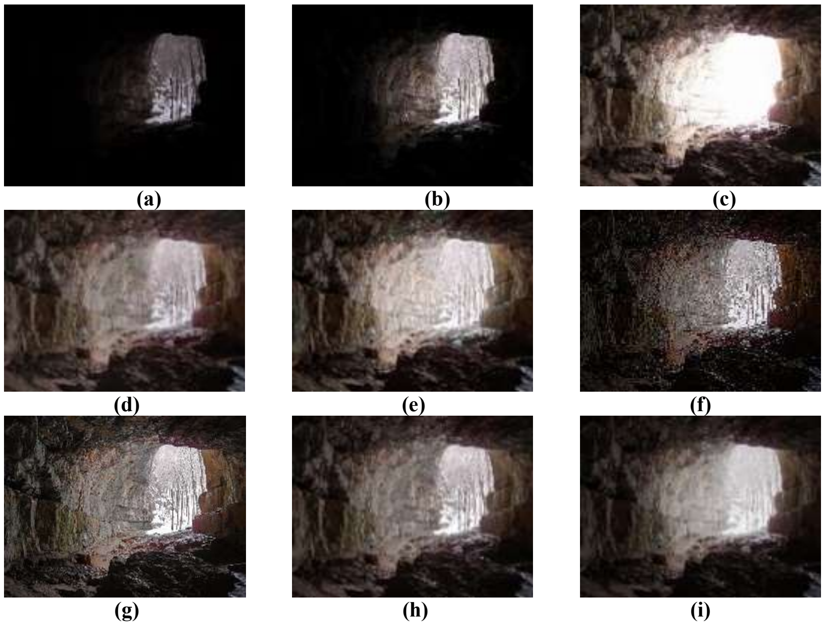


Fig. (6.8) Image 'Cave' and fused images **a-c** Cave, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

Fig. (6.9) Image 'Lighthouse' and fused images **a-c** Lighthouse, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, RG PM- I and PM-II.

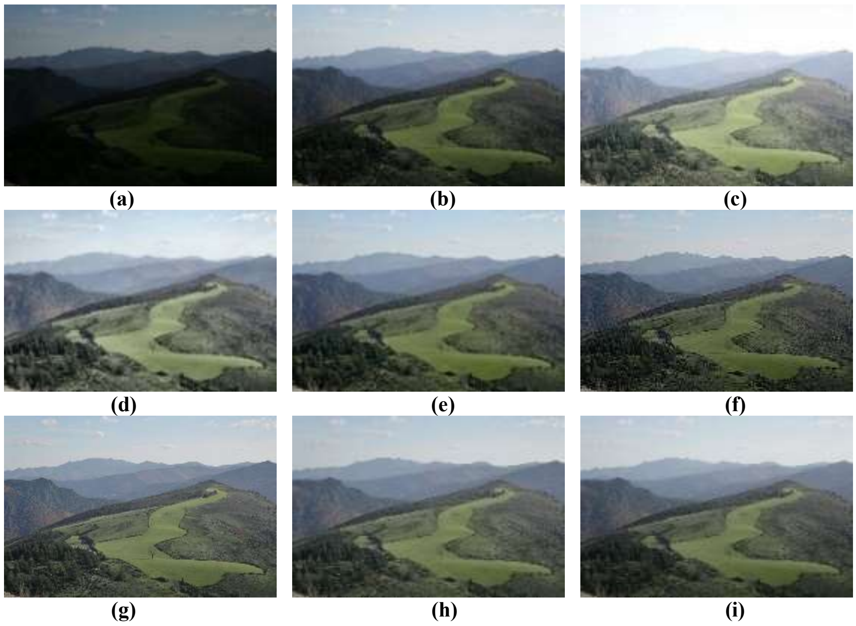


Fig. (6.10) Image ‘Landscape’ and fused images **a-c** Landscape, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

Fig. (6.11) Image ‘Sunset’ and fused images **a-c** Sunset, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, RG, PM- I and PM-II.

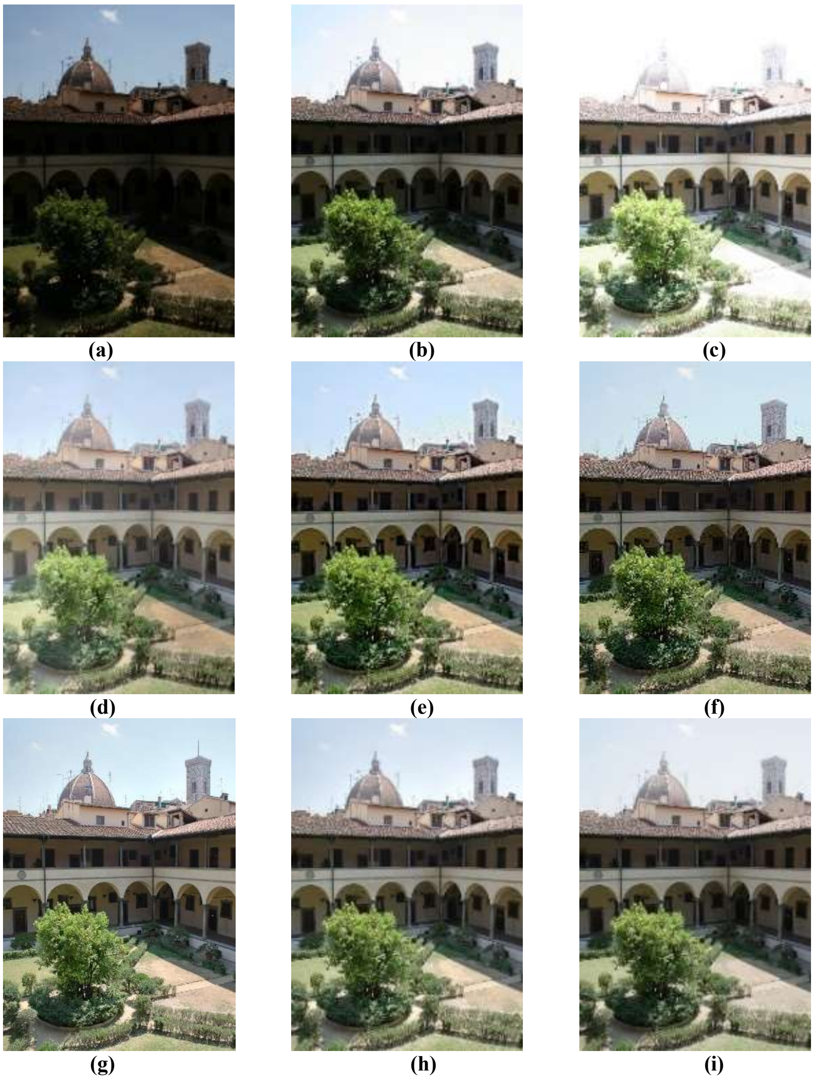


Fig. (6.12) Image 'Building2' and fused images **a-c** Building2, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

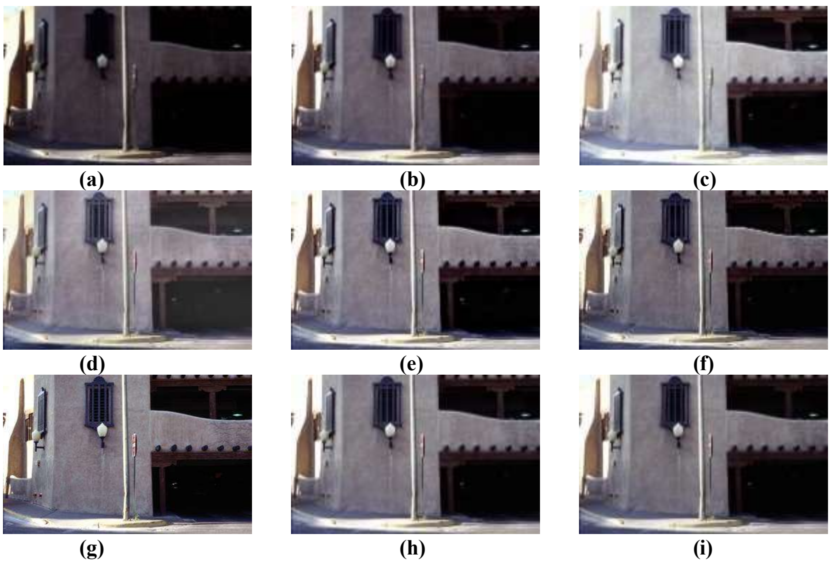


Fig. (6.13) Image 'Garage' and fused images **(a-c)** Garage, represents source images with increasing exposure time, **(d-i)** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

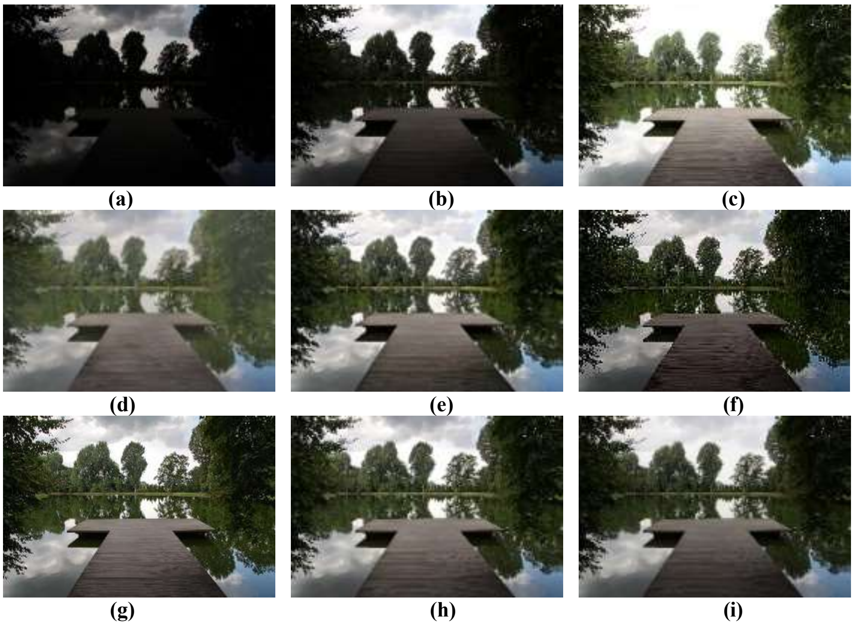


Fig. (6.14) Image 'Pond' and fused images **a-c** Pond, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

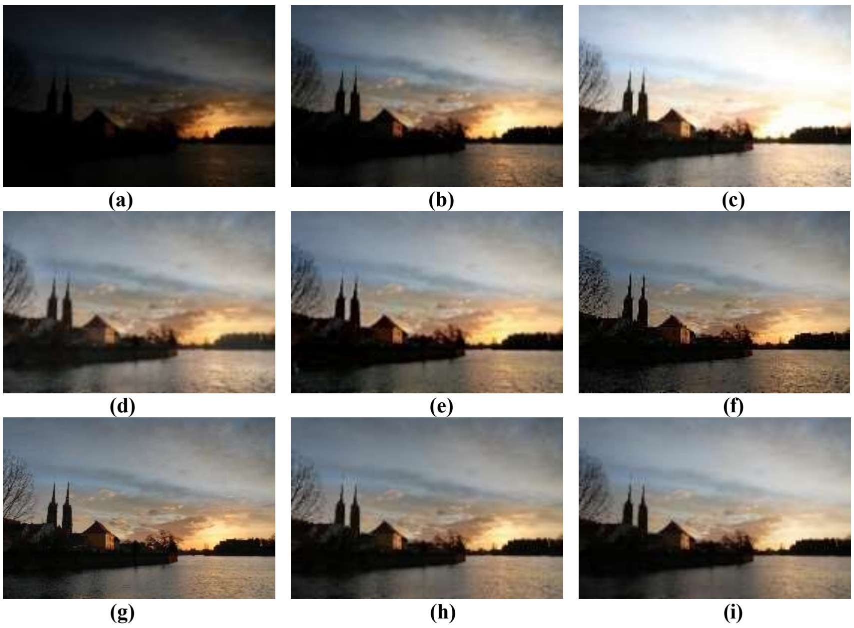


Fig. (6.15) Image 'Tumski' and fused images **(a-c)** Tumski, represents source images with increasing exposure time, **(d-i)** represents fused images for MFEF, SPD, BB, RG PM- I and PM-II.

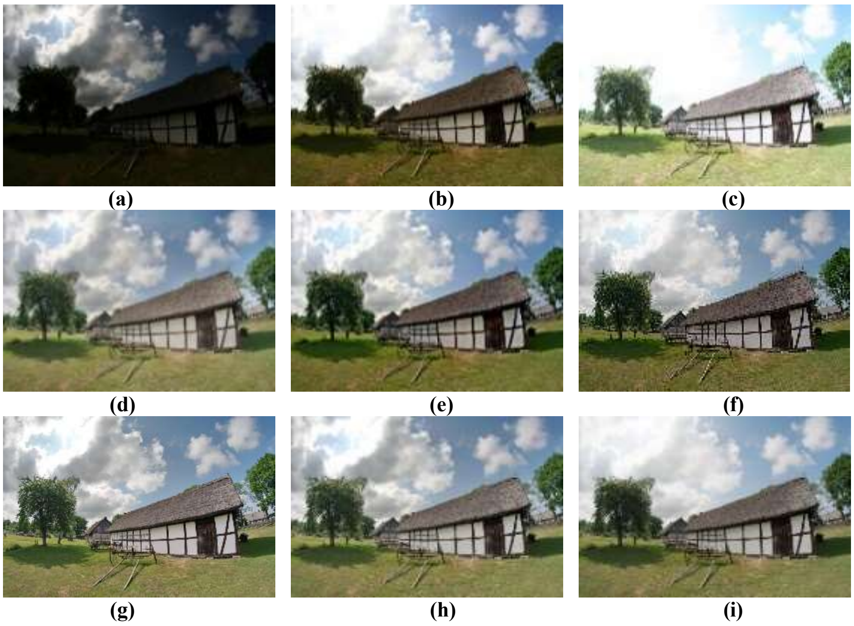


Fig. (6.16) Image 'Kluki' and fused images **a-c** Kluki, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

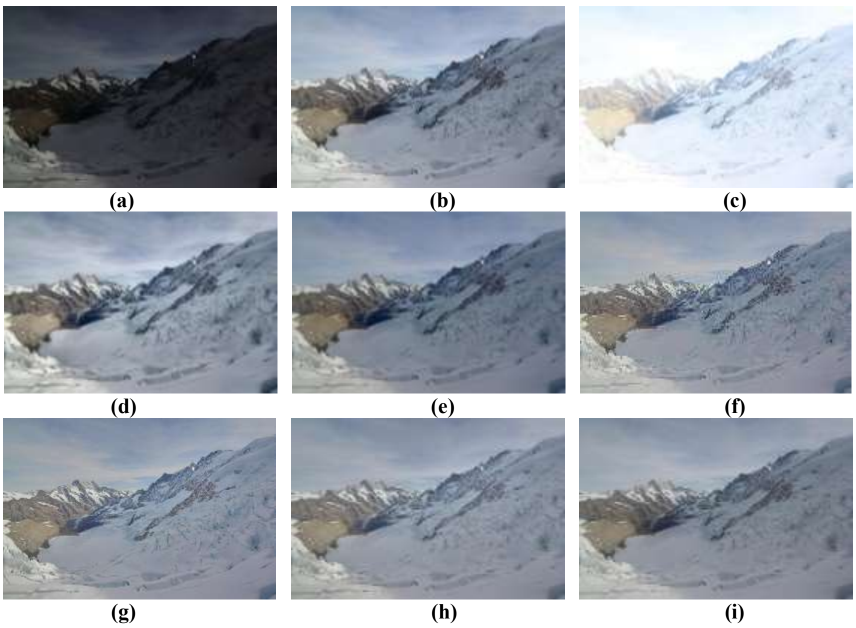


Fig. (6.17) Image 'Mountains' and fused images **a-c** Mountains, represents source images with increasing exposure time, **d-i** represents fused images for MFEF, SPD, BB, RG, PM- I and PM-II.

Table 6.1
Comparison chart of quality indices for region based fusion

Method	$Q^{AB}/_F$	MI	NMI	SSIM	$Q^{AB}/_F$	MI	NMI	SSIM
	Image : Garden				Image : Gasmire			
MFEF[42]*	0.7067	4.9798	0.6918	0.6949	0.6094	5.4507	0.7889	0.6304
SPD[68]*	0.6585	7.3417	1.0112	0.7762	0.5785	6.7549	0.9579	0.6907
BB1[90]*	0.6906	7.5474	1.0558	0.8409	0.6128	7.1086	1.0345	0.6548
RG[91]*	0.7419	8.4378	1.1719	0.7964	0.6515	6.2818	0.9112	0.6094
PM - I*	0.7445	8.1138	1.1246	0.7933	0.6521	4.1838	0.6197	0.5521
PM - II*	0.7249	10.7856	1.5050	0.8156	0.5980	9.7023	1.4026	0.6758
Method	Image : Estate				Image : Cabin			
MFEF[42]*	0.5635	6.2270	0.8695	0.8293	0.6895	5.9183	0.8101	0.6842
SPD[68]*	0.5835	8.5968	1.2043	0.8608	0.6993	7.3498	1.0043	0.6868
BB1[90]*	0.6845	9.6544	1.3666	0.9125	0.6890	7.2870	1.0162	0.8079
RG[91]*	0.7501	10.4996	1.4958	0.8797	0.7670	8.0960	1.1279	0.7295
PM - I*	0.7554	10.5319	1.4944	0.8790	0.7583	8.0142	1.1171	0.7302
PM - II*	0.7565	12.9356	1.8310	0.8885	0.7461	10.6894	1.4733	0.7639
Method	Image : Cave				Image : Lighthouse			
MFEF[23]*	0.7220	4.9983	0.7620	0.5808	0.5727	6.4011	0.9104	0.7252
SPD[24]*	0.5982	5.4493	0.8190	0.5073	0.4733	9.3407	1.2795	0.8332
BB[26]*	0.6831	4.7595	0.7722	0.8095	0.6115	10.0945	1.3927	0.8295
RG[30]*	0.7569	6.7519	1.0250	0.5984	0.6147	7.3308	1.0245	0.7742
PM - I*	0.7565	6.3947	0.9658	0.6259	0.6175	6.7296	0.9593	0.7295
PM - II*	0.5553	9.1381	1.3830	0.6522	0.5496	11.8506	1.6498	0.8130
Method	Image : Landscape				Image : Mountains			
MFEF[23]*	0.6213	7.6463	1.0834	0.7869	0.6425	5.6793	0.8398	0.6724
SPD[24]*	0.5354	9.9376	1.3662	0.8498	0.6059	9.4093	1.4095	0.7300
BB[26]*	0.6115	10.0945	1.3927	0.8295	0.7429	10.2705	1.5576	0.7104
RG[30]*	0.7447	11.3580	1.5752	0.8410	0.7892	6.4759	0.9993	0.6587
PM - I*	0.7498	10.6775	1.4889	0.8352	0.7914	5.0419	0.7893	0.6162

PM - II*	0.7351	14.1964	1.9622	0.8564	0.7883	12.3605	1.8663	0.7179
Method	Image : Sunset				Image : Building2			
MFEF[23]*	0.4774	6.1806	0.8600	0.7955	0.5901	5.4878	0.7550	0.7415
SPD[24]*	0.4839	10.1504	1.3917	0.8191	0.5876	8.0833	1.0949	0.8430
BB[26]	0.6051	11.0518	1.5255	0.8737	0.6299	9.1159	1.2484	0.8576
RG[30]*	0.6082	9.4944	1.3092	0.8316	0.6669	9.5835	1.3021	0.8291
PM - I*	0.6155	9.0790	1.2592	0.8284	0.6680	9.3218	1.2616	0.8290
PM - II*	0.5889	13.2280	1.8095	0.8357	0.6375	11.6358	1.6024	0.8449
Method	Image : Garage				Image : Pond			
MFEF[23]*	0.7220	6.8623	0.9717	0.8697	0.4245	5.1156	0.7395	0.6479
SPD[24]*	0.6866	8.2131	1.1842	0.9479	0.4252	6.6880	0.9366	0.7303
BB[26]	0.7286	9.6469	1.3993	0.9836	0.6001	6.4876	0.9302	0.8389
RG[30]*	0.8013	9.9918	1.4434	0.9622	0.6143	7.8534	1.1057	0.7591
PM - I*	0.7939	10.0872	1.4651	0.9640	0.6226	7.5660	1.0582	0.7609
PM - II*	0.7807	12.1687	1.7594	0.9811	0.5446	9.4500	1.3307	0.7981
Method	Image : Tumski				Image : Kluki			
MFEF[23]*	0.5864	6.6836	0.9326	0.8114	0.7431	6.5565	0.8940	0.6832
SPD[24]*	0.4666	10.05	1.3913	0.8549	0.6264	8.3529	1.1171	0.7815
BB[26]*	0.6012	11.3232	1.5771	0.8970	0.7041	10.3523	1.3820	0.7909
RG[30]*	0.6055	8.7795	1.2246	0.8574	0.7790	8.6159	1.1820	0.7436
PM - I*	0.6103	8.1940	1.1536	0.8569	0.7743	8.0747	1.1023	0.7228
PM - II*	0.5587	13.5664	1.8716	0.8562	0.7404	13.9196	1.8607	0.7657

*PM – I - Propose Method (normal average), PM – II - Proposed Method (weighted average), MFEF - Multi Focus and Exposure Fusion, SPD – Structural Patch Decomposition, BB1 – Block based Method

Chapter 7
Conclusion and Future Scope

7.1 Conclusion

The underlying information of scene captured by low dynamic range camera (LDR) suffered from lack of focus as well as low dynamic range. Image fusion is one of the inexpensive solution to increase the focused section and dynamic range.

In image fusion, multiple images are captured of same scene with different parameter settings like different focus points or shutter speed. The sections with most activity level are extracted and combined to synthesize the fused image. We have proposed three fusion techniques to combine multiple focus and exposure images to address lack of focus and dynamic range. These three fusion techniques have used predefined block, features and regions.

Block based technique, is intended to fuse differently focused images and further extended for exposure images. The captured images are splitted into local and global layers. These layers are processed separately to synthesize fused image. The outcomes of block based fusion are

- Proposed technique is based on spatial domain operations. Unlike to transform domain techniques, it doesn't suffered from data loss.
- Neighbour distance filter extracts frequency information on multi-scale basis by zero padding kernels. The image is free from up-sampling and down sampling process and thus less prone to shift-variance.
- Blocking effects are less visible.
- Controlling parameter 'M' gives liberty to choose between speed and quality.

Feature based technique is to address low dynamic range. The combination of Gaussian and recursive filter extract three features, high varying, low varying and constant features. These features are processed separately and combine together to synthesize fused image. The outcomes of feature based fusion are

- All operations are performed in spatial domain.

- Clean extraction of features by using combination of Gaussian and recursive filters.
- Three level activity levels testing of image based on three features gives best chance to select best part of captured image.
- Less affected by blocking effects.

Region based technique is to address low dynamic range. An image is splitted into frequency and base layer based on intensity variations by using recursive filter. Three regions of scene are defined by using intensity of captured images. Finally region based fusion is performed to synthesize fused image. The outcomes of region based fusion are

- All operations are performed in spatial domain.
- A unique three region theory is proposed.
- SIFT is used to identify the activity level of regions.
- Less dependence on region separation algorithm, unlike to region based fusion techniques.

All proposed techniques are tested on randomly selected differently focused and exposure images. Objective evaluation based on probability like Mutual information and edges like $Q^{\frac{AB}{F}}$ are used to evaluate the quality. The evaluation is also compared with state-of-the-art techniques.
As a conclusion of this research work, we are able to develop spatial domain fusion technique based on arbitrary shapes or features. The techniques have performed well as compare to state-of-the-art technique. We are able to reduce the effect of blocks, reduces the dependence on region algorithm. In region based fusion technique, we are able to process color images directly.

7.2 Future Scope

The image fusion technique depends on two factors; extraction of frequency information and activity level testing. Quality of fused technique will increase by improving any one or both factors. Therefore, we would like to see improvement in both the factors in future. We would also like to see more technique that would directly operate on color images.

References

[1] R. C. Gonzalez and R. E. Woods, "Digital Image Processing" 4th ed. Pearson (2018).

[2] A. K. Jain, "Fundamentals of Digital Image Processing," 1st ed. Pearson (2015).

[3] S. Li, X. Kang and L. Fang, "Pixel level image fusion: A survey of the state of the art," *Information Fusion*, vol. 33, pp. 100-112, Jan 2017.

[4] P. K. Varshney, "Multi-sensor data fusion," *Electronics and Communication Engineering Journal*, vol. 9, no. 6, pp. 245-253, Dec 1997.

[5] B. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transaction on Communications*, vol. 31, no. 4, pp. 532-540, Apr. 1983.

[6] A. Toet, "Image fusion by a ratio of low pass pyramid," *Pattern Recognition Letters*. vol. 9, no. 4, pp. 245-253, May 1989.

[7] N. Kingsbury, "A dual tree complex wavelet transform with improved orthogonality and symmetry properties," *Proceedings IEEE conference on Image Processing*, Vancouver 2000.

[8] 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, Sept. 2001.

[9] 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, June 2002.

[10] G. Piella, "A general framework for multi-resolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259-280, Dec. 2003. (Wavelet)

[11] J. Tang, "A contrast based image fusion technique in the DCT domain," *Digital Signal Processing*, vol. 14, no. 3, pp. 218-226, May 2004.

[12] V. S. Petrovic and C. S. Xydeas, "Gradient-based multi-resolution image fusion," *IEEE Transaction on Image Processing*, vol. 13, no. 2, 228-237, Feb. 2004.

[13] N. Mitianoudis and T. Stathaki, "Adaptive image fusion using ICA bases," *IEEE International Conference on Acoustics, Speech and Signal Processing*, May. 2006.

[14] I. De, B. Chanda and B. Chattopadhyay, "Enhancing effective depth of field by image fusion using mathematical morphology," *Image and Vision Computing*, vol. 24, no. 12, pp. 1278-1287, Dec. 2006.

[15] Q. Miao and B. Wang, "A novel adaptive multi-focus image fusion algorithm based on PCNN and sharpness," *Proceedings Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV SPIE*, 2005, pp. 704-712.

[16] W. Huang and Z. Jing, "Multi-focus image fusion using pulse coupled neural network," *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1123-1132, 2007.

[17] J. Zhang, X. Feng, B. Song, M. Li and Y. Lu, "Multi-focus image fusion using quality assessment of spatial domain and genetic algorithm," *Conference on Human System Interactions*, 2008.

[18] S. Li and B. Yang, "Multi-focus image fusion using region segmentation and spatial frequency," *Image and Vision Computing*. vol. 26, no. 7, pp. 971-979, July 2008.

[19] R. Nava, B. E. Ramirez and G. Cristobal, "A novel multi-focus image fusion algorithm based feature extraction and wavelets," *Proc. SPIE 7000, Optical and Digital Image Processing*, May 2008.

[20] Y. Zhang and L. Ge, "Efficient fusion scheme for multi-focus images by using blurring measure," *Digital Signal Processing*, vol. 19, no. 2, pp. 186-193, March 2009.

[21] R. Redondo, F. Sroubek, S. Fischer and G. Cristobal, "Multi-focus image fusion using the log-Gabor transform and Multi-size windows technique," *Information Fusion*, vol. 10, no 2, pp. 163-171, April 2009.

[22] M. R. Metwalli, A. H. Nasr, O. S. Farag Allah and S. A. Rabaie, "Image fusion based on Principle component analysis and high-pass filter," *IEEE International Conference on Computer Engineering and Systems*, Dec. 2009

[23] X. B. Qu, J. W. Yan and G. D. Yang, "Multi-focus image fusion method of sharp frequency localized Contourlet transform domain based on sum-modified Laplacian," *Optics Precision Engineering*, vol. 17, no. 5, pp. 1203-1211, May 2011.

[24] Q. Fu, F. Ren, L. Chen and Z. Xiao, "Multi-focus image fusion algorithms research based on Curvelet Transform," *Third IEEE International Conference on Genetic and Evolutionary Computing*, Feb 2010

[25] B. Yang and S. Li, "Multi-focus image fusion and Restoration with sparse representation," *IEEE Trans. on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884-892, April 2010.

[26] V. Aslantas and R. Kurban, “Fusion of multi-focus images using differential evolution algorithm,” *Expert Systems with Applications*, vol. 37, pp. 8861-8870, Dec. 2010.

[27] V. P. S. Naidu, “Image fusion technique using multi-resolution singular value decomposition,” *Defense science journal,* vol. 61, no. 5, pp. 479-484, Sept. 2011.

[28] Q. Miao, C. Shi, P. Xu, M. Yang and Y. Shi, “Multi-focus image fusion algorithm based on shearlets,” Chinese Optics Letters, vol. 9, no. 4, pp. 041001, 2011

[29] H. Zhao, Z. Shang, Y. Y. Tang and B. Fang, “Multi-focus image fusion based on the neighbor distance,” *Pattern Recognition*, vol. 46, no. 3, pp. 1002-1011, March 2013.

[30] G. Guorong, X. Luping and F. Dongzhu, “Multi-focus image fusion based on non-sub-sampled shearlet transform,” *IET Image Processing*, vol. 7, no. 6, pp. 633-639, August 2013.

[31] I. De and B. Chanda, “Multi-focus image fusion using morphology based focus measure in a quad-tree structure,” *Information Fusion*, vol. 14, no. 2, pp. 136-146, 2013.

[32] A. Saha, G. Bhatnagar and Q. M. J. Wu, “Mutual spectral residual approach for multifocus image fusion,” *Digital Signal Processing*, vol. 23, no. 4, pp. 1121-1135, July 2013.

[33] X. Bai, S. Gu, F. Zhou and B. Xue “Weighted image fusion based on multi-scale top-hat transform: Algorithms and comparison study,” *Optik*, vol. 124, pp. 1660-1668, 2013.

[34] Y. Zhang, L. Chen, J. Jia and Z. Zhao, “Multi-focus image fusion based on non-negative matrix factorization and difference images,” *Signal Processing*, vol. 105, pp. 84-97, 2014.

[35] Y. Jiang and M. Wang, “Image fusion with morphological component analysis,” *Information Fusion*, vol. 18, pp. 107-118, 2014.

[36] 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. 107-118, 2015.

[37] D. Guo, J. Yan and X. Qu, “High quality image fusion using self-similarity and depth information,” *Optics Communications*, vol. 338, pp. 138-144, 2015.

[38] Y. Liu, S. Liu and Z. Wang, “Multi-focus image fusion with dense SIFT,” *Information Fusion*, vol. 23, pp. 139-155, May-2015.

[39] X. Bai, Y. Zhang, F. Zhou and B. Xue, "Quadtree-based multi-focus image fusion using weighted focus measure," *Information Fusion*, vol. 22, pp. 105-118, 2015.

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

[41] 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, Feb. 2015.

[42] S. Paul, I. S. Sevcenco and P. Agathoklis, "Multi-exposure and Multi-focus image fusion in gradient domain, *Jn. of Circuits, Systems and Computers*, World Scientific. vol. 25, no. 10, DOI: 10.1142/S0218126616501231.

[43] H. Tang, B. Xiao, W. Li and G. Wang, "Pixel Convolutional Neural Network for multi- focus image fusion," *Information Sciences*, vol. 433-434, pp. 125-141, April. 2018.

[44] G. Qi, Q. Zhang, F. Zeng, J. Wang and Z. Zhu, "Multi-focus image fusion via morphological similarity-based dictionary construction and sparse representation," *CAAI Transactions on Intelligence Technology*, vol.3, no. 2, pp. 83-94, June. 2018.

[45] M. S. Farid, A. Mahmood and S. A. Al-Maadeed, "Multi-focus image fusion using Content Adaptive Blurring," *Information Fusion*, vol. 45, pp. 96-112, Jan. 2019.

[46] B. Xiao, G. Ou, H. Tang, X. Bi and W. Li "Multi-focus image fusion by Hessian matrix based decomposition," *IEEE Transaction on Multimedia*, DOI: 10.1109/TMM.2019.2928516

[47] S. Aymaz and C. Kose, "A novel image decomposition based hybrid technique with super-resolution method for multi-focus image fusion," *Information Fusion*, vol. 45, pp. 113-127, Jan. 2019.

[48] A. A. Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing*, vol. 23, no. 6, pp. 611-618, June 2005.

[49] I. Zafar, E. A. Edirisinghe and H. E. Bez, "Multi-exposure and multi-focus image fusion in transform domain," *IET International Conference on Visual Information Engineering (VIE 2006)*, pp. 606-611, 2006.

[50] T. Mertens, J. Kautz and F. V. Reeth, "Exposure fusion," *IEEE International Conference on Computer Graphics and Applications*, pp. 382-390, 2007.

[51] A. R. Varkonyi-Koczy, A. Rovid and T. Hashimoto, "Gradient – based synthesized multiple exposure time colour HDR image," *IEEE Transaction.*

on Instrumentation and Measurement, vol. 57, no. 8, pp. 1779-1785, Aug. 2008.

[52] S. Raman and S. Chaudhuri, "Bilateral Filter based Compositing for Variable Exposure Photography," *In proceedings of Euro graphics*, 2009.

[53] J. Wang, S. Feng and Q. Bao, "Pyramidal dual tree directional filter bank based exposure fusion for two complementary images," IEEE 10th International conference on Signal Processing Proceedings, 24-28 Oct 2010.

[54] K. Kotwal and S. Chauduri, "An optimization based approach to fusion of multi-exposure, low dynamic range images," IEEE 14th International conference on Information Fusion, 5-8 July 2011.

[55] K-H Jo and A. Vavilin, "HDR image generation based on intensity clustering and local feature analysis," *Computers in Human Behaviour*, vol. 27, no. 5, pp. 1507-1511, Sep. 2011.

[56] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized Random Walks for Fusion of Multi-Exposure Images," *IEEE Transaction on Image Processing*, vol. 20, no. 12, pp. 3634-3646, Dec. 2011.

[57] S. Li and X. Kang, "Fast multi exposure image fusion with median filter and recursive filter," *IEEE Trans. on Consumer Electronics*, vol. 58, no. 2, pp. 626-632, May 2012.

[58] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," IEEE Transaction on Image Processing, vol. 21, no. 1, pp. 358-365, Jan 2012

[59] W. Zhang, and W.K. Cham, "Gradient-Directed Multi-exposure Composition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2318-2323, April 2012.

[60] M. Song, D. Tao, C. Chen, J. Bu, J. Luo and C. Zhang, "Probabilistic Exposure Fusion," *IEEE Transaction on Image Processing*, vol. 21, no. 1, pp. 341- 357, Jan. 2012.

[61] R. Shen, I. Cheng and A. Basu, "QoE-based Multi-exposure image fusion in Hierarchical Multivariate Gaussian CRF," *IEEE Transaction on Image Processing*, vol. 22, no. 6, pp. 2469-2478, June 2013.

[62] J. Shen, Y. Zhao, S. Yan and X. Li, "Exposure fusion using boosting laplacian pyramid," *IEEE Trans. on Cybernatics*, vol. 44, no 9, pp. 1579-1590, Nov. 2013.

[63] X. Li, F. Li, L. Zhu and D. D. Feng, "Layered based exposure fusion algorithm," *IET Image Processing*, vol. 7, no. 7, pp. 701-711, 2013.

[64] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted Guided Image Filtering," *IEEE Transaction on Image Processing*, vol. 24, no. 1, pp. 120-129, Jan. 2015.

[65] A. V. Vanmali, S. G. Kelkar and V. M. Gadre, "Multi-exposure image fusion for dynamic scenes without ghost effects," *IEEE National conf. on Communications (NCC), Mumbai*, pp. 1-6, March 2015.

[66] 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, August 2015.

[67] M. Yu, E. Song, R. Jin, H. Liu, X. Xu, and G. Ma, "A novel method for fusion of differently exposed images based on spatial distribution of intensity for ubiquitous multimedia," *Multimedia Tools and Applications*, vol. 74, no. 8, pp. 2745-2761, April 2015.

[68] Keda Ma, H Li, H Yong, Z Wang, D Meng, and L Zhang, "Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach," *IEEE Transactions on Image Processing*, vol. 26, No 5, May 2017.

[69] V. Radhika, K. Veeraswamy and S. S. Kumar, "Digital Image Fusion using HVS in Block Based Transforms," *Journal of Signal Processing Systems*, vol. 90, no. 6, pp. 947-957, June 2018.

[70] Y. Kinoshita, S. Shiota, H. Kiya and T. Yoshida, "Multi-exposure image fusion based on exposure compensation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2018.

[71] A. Galdran, "Image Dehazing by artificial multiple exposure image fusion," *Signal Processing*, vol. 149, pp. 135-147, Aug. 2018.

[72] Y. Liu, D. Wang, J. Zhang and X. Chen, "A fast fusion method for multi-exposure image in YUV colour space," *IEEE 3rd Advanced Information Technology, Electronics and Automation Control Conference*, Oct. 2018.

[73] Q. Wang, W. Chen, X. Wu and Z. Li, "Detail Preserving Multi-scale exposure fusion," IEEE International Conference on Image Processing, Oct. 2018.

[74] G. Qu, D. Zhang, P. Yan, "Information measure for performance of image fusion," *Electronic Letters*. Vol. 38, no. 7, pp. 313-315, March 2002.

[75] M. Hoosny, S. Nahavandi, and D. Creighton, "Comments on Information Measure for Performance of Image Fusion," *Electronics Letters*, vol. 44, no. 18, Aug. 2008.

[76] C. S. Xydes and V. Pertovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308-309, Feb 2000.

[77] 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.

[78] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information Fusion*, vol. 14, no. 2, pp. 127-135, April 2013.

[79] I. E. Richardson, "The H.264 Advanced Video Compression Standard," 2nd Edition, Wiley, Aug. 2010.

[80] Y. Monobe, H. Yamashita, T. Kurosawa, and H. Kotera, "Dynamic range compression preserving local image contrast for digital video camera," *IEEE Transaction on Consumer Electronics*, vol. 51, no. 1, pp. 1–10, Mar. 2005.

[81] J. Kuang, G.M. Johnson, and M.D. Fairchild, "iCAM06: A refined image appearance model for HDR image rendering", *Journal of Visual Communication and Image Representation* vol. 18, no. 5, pp. 406-414, Oct. 2007.

[82] J. W. Lee, R. H. Park, and S. Chang, "Tone mapping using color correction function and image decomposition in high dynamic range imaging," *IEEE Transaction on Consumer Electronics*, vol. 56, no. 4, pp. 2772–2780, Nov. 2010.

[83] Q. Shan, J. Jia, and M.S. Brown, "Globally optimized Linear Windowed Tone Mapping," *IEEE Transaction on Visualization and Computer Graphics*, vol. 16, no. 4, pp. 663–675, July 2009.

[84] K. Kim, J. Bae, and J. Kim, "Natural HDR image tone mapping based on retinex," *IEEE Transaction Consumer Electronics*, vol. 57, no. 4, pp. 1807–1814, Nov. 2011.

[85] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge aware image and video processing," *ACM Transactions on Graphics*, vol. 30, no. 4, July 2011.

[86] C. Tomasi and R. Manduchi, "Bilateral Filtering for grey and color images," *IEEE International Conference on Computer Vision*, 1998.

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

[88] D. G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no.2, pp. 91-110, 2004.

[89] C. Liu, J. Yuen and A. Torralba, "SIFT Flow: Dense Correspondence across Scenes and its Applications," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978-994, May 2011.

[90] V. Chaudhary and V. Kumar, "Block based image fusion using multi-scale analysis to enhance depth of field and dynamic range," *Signal, Image and Video Processing, Springer*, vol. 12, no. 2, Feb. 2018.

[91] V. Chaudhary and V. Kumar, "Fusion of multi-exposure images using recursive and Gaussian filter," *Multi-dimensional Systems and Signal Processing, Springer* (2019) https://doi.org/10.1007/s11045-019-00655-6.

[92] Astrocog, *Reflex camera simple labels*, Nov. 2013. Accessed on Dec. 12, 2019. https://commons.wikimedia.org/wiki/File:Reflex_camera_simple_labels.svg

[93] D-Kuru, *Canon digital IXUS850*, Dec. 2009. Accessed on Dec. 12, 2019. https://commons.wikimedia.org/wiki/File:Canon_Digital_IXUS_850_IS-ar_5to4-fs_PNr%C2%B00268b.jpg

[94] Empernornie, *Nikon D810*, July 2014. Accessed on Dec. 12, 2019. https://commons.wikimedia.org/wiki/File:Nikon_D810-Crop-White_background.jpg

[95] C-K-Shene, *Depth of Field*. Accessed on Dec. 12, 2019. https://pages.mtu.edu/~shene/DigiCam/User-Guide/990/EXPOSURE/EV-depth-of- field.html

[96] Edward Lochocki, *How do object distance and focal length affect depth of field*. Accessed on Dec. 12, 2019. https://physicssoup.wordpress.com/2015/04/26/how-do- object-distance-and-focal-length-affect-depth-of-field/

[97] Edward Lochocki, *Why does a small aperture increase Depth of Field*. Accessed on Dec. 12, 2019. https://physicssoup.wordpress.com/2012/05/18/why-does-a-small- aperture-increase-depth-of-field/

[98] K. Ose et al., *Multispectral Satellite Image Processing*, 2016. Accessed on Dec. 12, 2019. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/spatial-resolution

[99] Kaushal Parmar et al., *Analysis of CT and MRI Image Fusion using Wavelet Transform*, 2012. Accessed on Dec. 12, 2019. https://www.semanticscholar.org/paper/Analysis-of-CT-and-MRI-Image-Fusion-Using-Wavelet-Parmar-Kher/a03ee42126b0134297ab40844ab34d439d429cd8

[100] Z. Liu et al., *The use of the contrast sensitivity function in the perceptual quality assessment of fused image*, March 2011. Accessed on Dec. 12, 2019. https://www.researchgate.net/publication/232887869_The_use_of_the_contrast_sensit ivity_function_in_the_perceptual_quality_assessment_of_fused_image

[101] *Neighbour distance*. Accessed on Dec. 12, 2019. https://www.cleanpng.com/png-second-fundamental-form-first-fundamental-form-dif-2732357/download-png.html

[102] *Multifocus Image Fusion*. Accessed on Nov 11, 2015. http://dsp.etfbl.net/mif/

[103] *Multifocus Image Dataset*. Accessed on Sep. 07, 2016. http://ucassdl.cn/resource.asp

[104] *Multifocus Image Dataset*. Accessed on Nov 22, 2015. http://camyx.comnews201304focustwist-app-re-focus-photos

[105] *HDR Photography gallery*. Accessed on Feb 02, 2019. https://www.easyhdr.com/examples/

[106] *Pangea Software*. Accessed on Feb 02, 2019. http://pangeasoft.net/pano/bracketeer/examples.html

[107] *Multifocus Image Dataset*. Accessed on Feb. 02, 2019. https://www.hdrsoft.com/index.html

[108] B. K. Shreyamsha Kumar, "Multifocus and Multispectral Image Fusion based on Pixel Significance using Discrete Cosine Harmonic Wavelet Transform," *Signal, Image and Video Processing, 2012, DOI: 10.1007/s11760-012-0361-x*.

[109] Harbinder Singh, Vinay Kumar and Sunil Bhooshan, "A Novel approach for detail-enhanced exposure fusion using guided filter," The Scientific World Journal, Hindwai Publishing Corporation, 2014, DOI:http://dx.doi.org/10.1155/2014/659217

List of Publications

1. Vishal Chaudhary and Vinay Kumar, “Block Based Image Fusion using Multi-Scale Analysis to Enhance Depth of Field and Dynamic Range” *Signal, Image and Video Processing (SIVP), Springer*, vol. 12, no. 2, pp. 271-279, Feb. 2018.

2. Vishal Chaudhary and Vinay Kumar, “Fusion of Multi-Exposure Images using Recursive and Gaussian Filter” *Multidimensional System and Signal Processing, Springer*, vol. 31, pp. 157-172, Jan 2020

3. Vishal Chaudhary and Vinay Kumar, “Target Based Image Fusion using Multi-Scale Feature Selection in Three Regions.” *Multimedia Tools and Application*, *Springer* (Accepted)

Annexure – I

Image Details

Image	Size	Format	Courtesy
Boy	$632 \times 474 \times 1$	JPEG	[102]
Building	$720 \times 480 \times 1$	JPEG	[103]
Building2	$800 \times 114 \times 3$	JPEG	[105]
Circuit	$720 \times 480 \times 1$	JPEG	[103]
Cathedral	$768 \times 1024 \times 3$	JPEG	[109]
Cabin	$221 \times 332 \times 3$	JPEG	[106]
Cave	$870 \times 653 \times 3$	JPEG	[105]
Disk	$640 \times 480 \times 1$	BMP	[108]
Estate	$870 \times 489 \times 3$	JPEG	[105]
Garden	$870 \times 578 \times 3$	JPEG	[105]
Grass	$242 \times 242 \times 3$	JPEG	[104]
Garage	$348 \times 222 \times 3$	JPEG	[66]
Gasmire	$700 \times 467 \times 3$	JPEG	[106]
Kluki	$870 \times 580 \times 3$	JPEG	[105]
Landscape	$790 \times 526 \times 3$	JPEG	[107]
Lighthouse	$796 \times 529 \times 3$	JPEG	[107]
Mountains	$870 \times 578 \times 3$	JPEG	[105]
Pepsi	$512 \times 512 \times 1$	TIF	[102]
Pond	$800 \times 533 \times 3$	JPEG	[105]
Sunset	$870 \times 577 \times 3$	JPEG	[105]
Tumski	$870 \times 580 \times 3$	JPEG	[105]
Venice	$790 \times 526 \times 3$	JPEG	[107]