Some New Methods for Multi-resolution Image Fusion

by

Upla Kishorkumar Parsottambhai

(200821005)

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree

of

Doctor of Philosophy

in

Information and Communication Technology

to

Dhirubhai Ambani Institute of Information and Communication Technology

January 2016

Declaration

This is to certify that

1. the thesis comprises my original work towards the degree of Doctor of Philosophy in Information and Communication Technology at DA-IICT and has not been submitted elsewhere for a degree,

2. due acknowledgment has been made in the text to all other material used.

Signature of Student

Certificate

This is to certify that the thesis work entitled "Some New Methods for Multi-resolution Image Fusion" has been carried out by Upla Kishorkumar Parsottambhai (200821005) for the degree of Doctor of Philosophy in Information and Communication Technology at this Institute under my supervision.

Thesis Supervisor Prof. Manjunath Joshi

Acknowledgments

First and above all, I praise God, the almighty for providing me this opportunity and granting me wisdom, health and strength to undertake this research work and complete it successfully. I could never have accomplished this without the faith I have in the almighty. Although the work described here was performed independently, it would not have been possible without the help and support of many wonderful peoples. I would therefore like to offer my sincere thanks to them.

I am extremely grateful to my supervisor Prof. Manjunath V. Joshi for his valuable guidance, scholarly inputs and consistent encouragement throughout the research work. Completion of this doctoral work was possible only because of his unconditional support. He always made himself available to clarify my doubts despite of his busy schedules. His guidance, fruitful discussions, keen interest and dedication in the research motivated me to shape my research direction. I greatly appreciate his efforts for making my research work focused to achieve goal. His achievements, his work ethics, and his keen eye for every important detail have been an inspiration throughout all the years I have worked with him. Thank you sir for the patience you had to supervise and direct my work, I consider it as a great opportunity to do my doctoral programme under his guidance and to learn from his research expertise.

I express my sincere gratitude to Director, Prof. S.C. Sahasrabudhe, Registrar, Dean-AP, Convener-Ph. D. and all other faculties who directly or indirectly helped me throughout my time at DA-IICT. I would like to thank administrative and technical staff members of DA-IICT who have been kind enough to advise and help me in their respective roles. My gratitude also extend to staff of DA-IICT resource-center.

On a broader note, I wish to acknowledge all the professors of DA-IICT who have inspired me directly or indirectly. I would like to thank faculty members Prof. V. P. Sinha, Prof. Suman Mitra, Prof. Asim Banerjee and Prof. Aditya Tatu who have shaped my thinking about this research direction during the initial phases of thesis work. It is always nice remembering the excellent times of the lessons by Prof. V. P. Sinha and how this sweet experience moved me toward the initiating research work. Perhaps most importantly, he has taught us how to think critically and independently for carrying out research. I express my gratitude to Prof. Mehul Raval (Ahmedabad University,

Ahmedabad, India) for his technical support, valuable inputs and remarks at several occasions during this doctoral work.

I am grateful to Prof. N. M. Santoki, former Head of Electronics and Communication Department at Government Polytechnic, Palanpur (GP, Palanpur) for being a constant source of inspiration to me while I was serving at Government Polytechnic, Palanpur. Without his warm encouragement and support, I would not have gone back to school to pursue my PhD. Being an Assistant professor at Electronics department, SVNIT, Surat I express my thanks to Prof. B. R. Taunk, former Head of EC department for giving healthy research environment. I am also thankful to present Head, Dr. (Mrs) U. D. Dalal for granting leave for completion of research work. My colleagues at DA-IICT and SVNIT have all extended their support in a very special way, and I gained a lot from them, through their personal and scholarly interactions, their suggestions at various points of my research programme. I believe that I would not be able to name every colleague and friend separately but I thank all of them for everything that they did for me.

One person who has always been ready to help me was Dr. Prakash Gajjar, Government Polytechnic for Girls, Surat. He always stand along with me to help me as and when required. Also, his research skills and dedication towards work mold my career as an active acedemician. Thank you Dr. Prakash for all your support! I also thank Prof. S. Ravishanker (Amrita School of Engineering, Banglore) for generously sharing his time and knowledge that influenced my research work. I would like to thank anonymous reviewers of our publications for their constructive suggestions that have greatly improved the publications.

During doctoral programme I was able to attend many national and international conferences held in different countries. I am thankful to Director of SVNIT, Surat for giving academic and financial support for attending all those conferences. I thank Sheida Rahmani (University of California, Los Angeles, USA) for availing their code for comparing our results on fusion. Also, I am thankful to A. G. Mahyari (Shiraz University, Iran) for providing their code of fusion method.

I am deeply thankful to my entire family for their love and support. Words can not express how grateful I am to my mother, father, mother-in law, and my father-in-law for all of the sacrifices that you have made on my behalf. Your prayer for me was what sustained me thus far. I acknowledge my entire family for providing me healthy educated

atmosphere in our family. This last word of acknowledgment I have saved for my dear wife Dipika, who has been with me all these years and has made them the best years of my life. I always fall short of words and felt impossible to describe her support in words. Without your unconditional supports and encouragements, I could not have finished this work, it was you who kept the fundamental of our family, and I understand it was difficult for you. I see myself unable to even express my feelings about the love and patience that I observed from you. I can just say thanks for everything and may God give you all the best in return. Last, I would like to thank my son, Taksh who made these years unforgettable and enjoyable.

DA-IICT, Gandhinagar — Kishor P. Upla

September, 2014 — 200821005

Abstract

It is always an interest of a mankind to explore the various resources of the earth such as minerals, agriculture, forestry, geology, ocean etc., Before the invention of the remote sensing, in order to analyze the various resources it was required to visit the field to take the different forms of data samples and later on those were processed further. The revolution in terms of photography using satellite made it possible to view the earth's surface without being in touch with the area of interest. With the help of satellite technology it is also possible to view the locations on the earth which are not accessible by the mankind. Remote sensing has effectively enabled the mapping, studying, monitoring and management of various resources present on the earth. It has also enabled monitoring of environment and thereby helping in conservation. In the last four decades, the advancement of the remote sensing technologies have improved the methods of collection, processing and analysis of the data.

Remote sensing involves the acquiring of the pictorial data of the earth's surface without any type of physical contact. It provides the information which not only helps in managing and protecting the natural resources but also helpful in the development of land usage in terms of urban planning. One of the major advantages of the remote sensing satellites is the ability to provide the repetitive observations of the same area. This capability is very useful to monitor dynamic phenomena such as cloud evolution, vegetation cover, snow cover, forest fires, etc. A farmer may use thematic maps to monitor the health of his crops without going out to the field. A geologist may use the images to study the types of minerals or rock structure found in a certain area. A biologist may want to study the variety of plants in a certain location.

The image acquisition process of remote sensing system consists of sensing the reflected electromagnetic energy from the surface of the earth. The amount of energy reflected from the earth's surface depends on the composition of the material. The variations in the reflected energy are captured by the remote sensing sensors placed in the satellite or aircraft which are then quantized and digitized into the pictorial form i.e., images. The smallest element of an image i.e., pixel corresponds to an area of a few squared meters in the actual scene which is referred to the spatial resolution of the given sensor. The spatial resolution is limited by the instantaneous field of view (IFOV) of the remote sensing

system. Smaller the IFOV, lesser is the area covered by sensor and hence the amount of collected light energy is reduced. By keeping the small IFOV, one can increase the amount of light falling on the sensor i.e. photo detector element by increasing the spectral width of the sensor. This results in wider spectral width with high spatial resolution. Alternatively, one can use the sensor with wide IFOV that covers large surface area. This makes the sensor to collect more light energy but the image formed has lower spatial resolution. However, in this case the spectral width of the sensor can be made narrower in order to sense the data in that spectral width which results in an image with high spectral resolution having fine spectral details. The data with narrower spectral width always helps in better classification since the materials present in the scene reflect the light energy of different wavelengths based on their composition. If one can capture the reflected energy at different bands of wavelengths then it provides separate information about the same scene content. However, this set of images obtained at different spectral bands is possible with the compromise of poor spatial resolution. This trade off in high spatial and spectral resolutions imposes the limitations on the hardware construction in the remote sensing sensors.

It is always of interest to visualize the content of the scene with high spatial and spectral resolutions. However, constraints such as the tradeoff between high spatial and spectral resolutions of the sensor, channel bandwidth, on board storage capability of a satellite system place the limitations on capturing the images with high spectral and spatial resolutions. Due to this, many commercial remote sensing satellites such as Quickbird, Ikonos and Worldview-2 capture the earth's information with two types of images: a single panchromatic (Pan) and a number of multispectral (MS) images. Pan has high spatial resolution with lower spectral resolution while MS image has higher spectral resolving capability with low spatial resolution. An image with high spatial and spectral resolutions i.e., fused image of MS and Pan data can lead to better land classification, map updating, soil analysis, feature extraction etc. Also, since fused image increases the spatial resolution of the MS image it results in sharpening the image content which makes it easy to obtain greater details of the classified maps. The *pan-sharpening or multi-resolution image fusion* is an algorithmic approach to increase the spatial resolution of the MS image with the preservation of spectral contents by making use of the high spatial resolution Pan image. In this thesis we address some new multi-resolution image

fusion techniques.

In multi-resolution image fusion problem, the given MS and Pan images have high spectral and high spatial resolutions, respectively. One can think of obtaining the fused image using these two by injecting the missing high frequency details from the Pan image into the MS image. The quality of the final fused image will then depend on the method used for high frequency details extraction and also on the technique for injecting these details into the MS image. In the literature various approaches have been proposed based on this idea. Motivated from this, we first address the fusion problem by using different edge preserving filters in order to extract the high frequency details from the Pan image. Specifically, we have chosen the guided filter and difference of Gaussians (DoGs) for detail extraction since these are more versatile in applications involving feature extraction, denoising, etc. Using these edge preserving filters we extract the high frequency details from the Pan image and inject them into the upsampled MS image.

One of the drawbacks of the fusion methods using edge preserving filters is the upsampling operation required to perform on the MS image before the injection of high frequency details into the same. Since this operation do not consider the effect of aliasing it results in distortions in the final fused image. Solving the problem of fusion by model based approach is accurate since aliasing present due to undersampled MS observation can be taken care of while modeling. Many researchers have used the model based approaches for fusion with the emphasis on improving the fused image quality and reducing the color distortion. In a model based method, the low resolution (LR) MS image is modeled as the blurred and noisy version of its ideal high resolution (HR) fused image. Since this problem is ill-posed, it requires regularization to obtain the final solution. In the proposed model based approach a learning based method that uses Pan data is used to obtain the required degradation matrix that accounts for aliasing. We use sub-sampled as well as non sub-sampled contourlet transform based learning to obtain close approximation to fused image (initial estimate). Then using the proposed model, the final solution is obtained by solving the inverse problem where a Markov random field (MRF) smoothness prior is used for regularizing the solution.

We next address the fusion problem based on the concept of self similarity and compressive sensing (CS) theory. In the earlier proposed approach, the degradation matrix entries were estimated by modeling the relationship between the Pan derived initial esti-

mate of the fused MS image and LR MS image which may be inaccurate as the estimate depends on the low spectral resolution Pan data. If the initial fused estimate is derived using the available LR MS image only, then the transformation between the estimated high resolution initial estimate and the observed MS image would be more accurate. This makes the estimated degradation matrix to better represent the aliasing. In this case we obtain the initial estimate using the available LR MS image only. Here, we use the property of natural images that the probability of availability of same or similar information in the current resolution and its coarser resolution is high. We exploit this self similarity concept and combine it with CS theory in order to obtain the initial estimate of fused image which is then used in obtaining the degradation. Finally, in order to better preserve the spatial details and to improve the estimate of fused image, we solve the multi-resolution fusion problem in a regularization framework by making use of a new prior called Gabor prior. Use of Gabor prior ensures features at different spatial frequencies of fused image image to match those of the available HR Pan image. Along with Gabor prior we also include a MRF prior which maintains the spatial correlatedness among the HR pixels.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the major achievements of mankind is to record the data of what we observe in the form of photography which is dated to 1826. Man has always tried to reach greater heights (treetops, mountains, platform and so on) to observe the phenomenon of interest, to decide habitable places, farming and such other activities. This motivates man to take photographs of earth from elevated platforms. To take the elevated photographs initially balloons, pegions and kites were used to capture the scene. After the invention of the aircraft in 1903, the first aerial photograph with stable platform was made possible in 1909. The primary platform that was used to carry remotely sensed instruments shifted from aircrafts to satellites in 1960s and 1970s. Satellites can cover much more land space than planes and can monitor areas on a regular basis. During the same time the word '*remote sensing*' replaced the frequently used word '*aerial photograph*'.

The new era in remote sensing began when the United State launched the first earth observation satellite called earth resources technology satellite (ERTS-1) dedicated primarily for land observation. This was followed by many other satellites like Landsat 1-5, systeme pour l' observation de la terre (SPOT), Indian remote sensing (IRS), Quickbird, Ikonos, etc. Change in image format from analog to digital was another major step towards the processing and interpretation of remotely sensed data [1]. The digital format made it possible to display and analyze imagery using computers, a technology that was also undergoing rapid change during this period. Due to the advancement of technology and development of the new sensors capture of the earth's surface in several different portions of the electro-magnetic spectrum are possible these days. One could now view the same area by acquiring the data as several different images in portions of the spectrum

beyond what the human eye could view. The remote sensing technology made it possible to see things occurring on the earth's surface which may not be detected by human eye.

The appropriate definition of '*remote sensing*' can be given as follows [1]. "*It means sensing of the earth's surface from space by making use of the properties of electromagnetic wave emitted, reflected or diffracted by the sensed objects, for the purpose of improving natural resource management, land use and the protection of the environment.*" Remote sensing has enabled mapping, studying, monitoring and management of various resources like agriculture, forestry, geology, water, ocean etc. It has further enabled monitoring of environment thereby helping in conservation. One of the major advantage of the satellite is its ability to provide repetitive observations of the same area with intervals of few minutes to a few weeks depending on the sensor and the orbit. This capability is very useful to monitor dynamic phenomena such as cloud evolution, vegetation cover, snow cover, forest fires, etc. A farmer may use thematic maps to monitor the health of his crops without going out to the field. A geologist may use the images to study the types of minerals or rock structure found in a certain area. A biologist may want to study the variety of plants in a certain location. In the last four decades satellite imaging has grown as a major tool for collecting information on almost every aspect on the earth.

The imaging sensors on board can acquire information in different spectral bands on the basis of the exploited frequency or at different resolutions. Therefore, a wide spectrum of data can be available for the same observed site. For many applications the information provided by individual sensors are incomplete, inconsistent, or imprecise and additional sources may provide complementary data. Fusion of different information results in better understanding of the observed site thus decreasing the uncertainty related to the single sources. In the interpretation of a scene, contextual information is important. For example in an image labeling problem, a pixel considered in isolation may provide incomplete information about the desired characteristics. Context can be defined in the frequency, space and time domains. These bands may be acquired by using either a single multispectral sensor or by using number of sensors operating at different frequencies. The spectral context improves the separation between various ground cover classes compared to a single-band image.

1.1 Characteristics of Remote Sensing Imagery

The resolution is an important characteristic feature of aerial images. In general sense the term '*resolution*' is defined as the smallest physical quantity that is discernable by an instrument. In other words, resolution is the power of the instrument to record finer details. High resolution of an instrument enables one to measure the quantity with more precision. In image processing, the resolution refers to the ability of the imaging sensor to record smallest measurable detail in a visual presentation. High resolution of an image is important in image processing as it helps to derive precise and accurate information in various applications.

Remote sensing images are characterized by four types of resolution: spatial resolution, spectral resolution, radiometric resolution, and temporal resolution.

- *Spatial resolution:* In digital image sensors, the analog images produced by the optical system are spatially sampled by the detector. Spatial resolution is a measure of the optical sensor's ability to record closely spaced objects such that they are distinguished as separate objects. If the imaging scenes are oversampled with a spatial frequency higher than the Nyquist frequency, it results in high resolution image. However, in practice, most digital image sensors undersample the analog scene. As a consequence the resulting resolution is determined by the spatial sampling frequency. In remote sensing, it refers to area of land space represented by one pixel in an image. It can be thought as the projection of the photo detecting element on to the ground. Thus the resolution is directly related to the area on the ground that represents a pixel in the detector. A sensor with $1m \times 1m$ spatial resolution can give finer details of the scene compared to a sensor with $10m \times 10m$ spatial resolution. Thus high spatial resolution allows for sharp details and fine intensity transitions across all directions. For representing an object, high spatial resolution image has more pixels compared to low resolution image. In other words, as the spatial resolution increases, the associated file size increases. Capturing a high spatial resolution camera needs a high density image sensor with closely spaced photo detectors. Different applications require different spatial resolutions. For applications such as large area change detection, it is economical to use medium-resolution imagery with large swath widths, to observe into areas where changes of interest

Table 1.1: Spatial resolution of some satellites

Satellite	Multispectral image	Panchromatic image
Landsat	$30m \times 30m$	$15m \times 15m$
SPOT 2, 4	$20m \times 20m$	$10m \times 10m$
Ikonos	$4m \times 4m$	$1m \times 1m$
OrbView3	$4m \times 4m$	$1m \times 1m$
Quickbird	$2.4m \times 2.4m$	$0.6m \times 0.6m$

have occurred. Similarly, for planimetric applications, it is recommended that imagery with the highest possible resolution be used to extract various features such as pavements, roads, etc. Different satellites capture images at various resolutions. For example in Table 1.1 we list the spatial resolution of the various satellites for capturing multispectral (MS) and panchromatic (Pan) images.

- *Spectral resolution:* Spectral resolution refers to the frequency or spectral resolving power of a sensor and is defined as the smallest resolvable wavelength difference by the sensor. Spectral resolution represents the width of the band within the electromagnetic spectrum that can be sensed by a sensor. As the bandwidth becomes narrower, the spectral resolution becomes higher. The spectral resolution plays important role in satellite imaging. High spectral resolution images captured by remote sensing camera provide more detailed information about mineral resources and geographical structures of the earth or any other planet under observation. High spectral resolution images can be acquired by capturing images of narrow spectral range. These images consists of pixels that represent spectral response within the band. For example in the case of vegetation, maximum reflectance occurs at the near infrared (NIR) region. Hence images captured in the band of NIR give more details of vegetation compared to the images captured in red or green spectral bands. A set of images captured at different spectral bands can be used to monitor land and other natural resources, including vegetated areas, wetlands, and forests. In Table 1.2 we display the spectral resolutions of different MS bands and Pan image provided by two satellites namely, Landsat enhanced thematic mapper plus (ETM+) and Quickbird.

- *Radiometric resolution:* Pixels carry information of the image intensity in form of

Table 1.2: Comparison of the spectral resolutions of the Landsat ETM+ and Quickbird sensors' bandwidth (μm)

Spectral band	Landsat ETM+	Quickbird
Panchromatic	0.52-0.90	0.45-0.90
Blue (band-1)	0.45-0.51	0.45-0.52
Green (band-2)	0.52-0.60	0.52-0.60
Red (band-3)	0.63-0.69	0.63-0.69
Near-infrared (band-4)	0.75-0.90	0.76-0.90

Table 1.3: Radiometric resolution of some satellites

Satellite	Radiometric resolution (bits)
Landsat	8
IRS	7
SPOT	8
Quickbird	11
Ikonos	11
OrbView3	11

binary digits called 'bits'. The intensity at any location in a real world scene may vary from zero to infinity. However in digital image it is not possible to represent this entire range. In practice this range is divided into a finite levels and the real world intensity is quantized and assigned the nearest finite level. The radiometric or brightness resolution refers to the smallest change in brightness that can be represented in an image. Each radiometric level is assigned a binary code. The increase in the brightness resolution requires more number brightness levels and hence more number of bits for each level. A binary image has two levels; black and white, hence requires only one bit for each pixel. A gray scale image is usually quantized using 256 grey levels with each level represented using 8 bits. Similarly, if each color plane of an RGB image requires 8 bits then at least 24 bits are needed for representing each pixel. For the illustration purpose we display the radiometric resolution of different satellites in Table 1.3.

- *Temporal resolution:* The term temporal resolution is related to video signals. A video of an event is a sequence of images (frames) captured at regular and short time interval between them. Temporal resolution, also known as frame rate, is the measure of the capability of displaying smallest movement/ motion of the moving

Table 1.4: Temporal resolution of various satellites

Satellite/Sensor	Revisit period(days)
Landsat	8
IRS	7
SPOT	8
Quickbird	11
Ikonos	11
OrbView3	11

objects in the video. Thus it refers to the number of frames captured per second. A video captured with low temporal resolution exhibits flicker or transitions of the moving objects in the scene/event. With high temporal resolution, the movement of the moving objects appears smooth and continuous. For a given duration of time, a high temporal resolution video requires more memory for storage and large bandwidth for transmission. In remote sensing temporal resolution refers to the frequency at which a given geographical area is imaged. Higher temporal resolution enables monitoring occurrence of rapid changes such as forests, floods, etc. This also improves the probability of obtaining cloud-free imagery over areas that experience frequent cloud cover. The revisit period of different satellites are listed in the Table 1.4.

There exists a trade-off while selecting a sensor. For example, if we want a high spatial resolution then the requirement is to keep low IFOV which reduces the energy of the reflected light acquired by the sensor causing the reduction in signal to noise ratio. Thus the captured image is distorted. One can improve the spatial resolution by capturing the image using higher spatial width for the sensor. However, this is possible at the cost of poor spectral resolution. Thus in order to have sensor with optimum performance we are required to make the suitable choice as per the requirement. The high spatial resolution images have better details which help in accurate measurement of the feature extension in the image. On the other hand the images with high spectral resolution give the better classification of different regions which are benefitted in accurate identification of the object. In this work we address the problem of reconstructing remotely sensed images that posses both the high spatial and high spectral resolutions.

1.1.1 Multispectral Images

Objects appear different through red lenses, or through blue or green lenses. Hence certain satellite sensors can record reflected energy in the red, green, blue, or infrared bands of the spectrum for the purpose of better analysis of data. This process of acquiring few different band images is call multispectral (MS) imaging. The improved ability of multispectral sensors provides a basic remote sensing data resource for quantitative thematic information, such as the type of land cover. Resource managers use information from multispectral data to monitor fragile lands, vegetated areas, wetlands, forests, etc. These data provide unique identification characteristics leading to a quantitative assessment of the earth's features.

In the area of remote sensing we are interested in recognizing an object or a feature from the images which are captured by using sensing devices. These feature includes vegetation, soil, rocks, minerals, water/ocean, snow and man-made features. The recognition of such objects requires high spectral resolution of the sensor. Remote sensing satellites are fitted with a camera that has a multi-channel detector with a few spectral bands. Each detector is sensitive to radiation within a narrow wavelength band. The resulting MS image contains both the brightness and spectral (color) information of the targets being observed. The MS sensors can record reflected energy in the red, green, blue, or infrared bands of the spectrum. The improved ability of MS sensors provides a basic remote sensing data resource for various kinds of applications. Examples of multi-spectral satellite systems include: Landsat TM, MS scanner (MSS), SPOT high resolution visible multispectral (HRV-XS), Ikonos MS, QuickBird MS.

In order to capture MS images, the light reflected from the scene is passed through filters with different spectral characteristics. These filters decompose the light into different spectral components which are then collected by multi-channel detectors and converted into digital image. Since the optical power is divided into several components, the available power to each detector is reduced. This leads to poor signal to noise ratio resulting in low spatial resolution. Thus multispectral images are characterized by high spectral resolution i.e. narrow bandwidth and low spatial resolution. As an example in Fig. 1.1 we show images of MS bands captured by QuickBird satellite. The spectral range of these bands are: blue ($0.45 - 0.52\ \mu m$), green ($0.52 - 0.60\ \mu m$), red ($0.63 - 0.69\ \mu m$), near

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

Figure 1.1: Multispectral (MS) images with spatial resolution of $2.4m \times 2.4m$ corresponding to the area of Sundarban, India captured using Quickbird satellite: (a) blue (band-1, $0.45 - 0.52\ \mu m$), (b) green (band-2, $0.52 - 0.60\ \mu m$), (c) red (band-3, $0.63 - 0.69\ \mu m$) and (d) near-IR (band-4, $0.76 - 0.90\ \mu m$).

(a) (b)

Figure 1.2: Color composition of MS image: (a) Natural color composition (NCC) and (b) false color composition (FCC)

infrared ($0.76 - 0.90\ \mu m$). Each spectral band has the utility in different kinds of analysis. Band 1 (blue) images are useful for representing water bodies, land, soil, vegetation etc. Band 2 (green) images enable us to inspect health of vegetation. Band 3 (red) images help in discrimination of vegetation, delineation of soil and geologic boundaries. Band 4 (NIR) images identify crops, emphasize land-water contrasts, etc. In order to visualize the image in RGB color format it is required to combine the red, green and blue bands. The resulted image is said to have natural color composition (NCC). However, in the case of vegetation where there is a maximum reflectance occurring at the NIR region we are required to observe the effects in color images. This can be accomplished by combining near NIR, red and green bands in RGB color image format which is referred to false color composition (FCC) since the represented color is not the true color perceived by us. Examples of NCC and FCC images are displayed in Fig. 1.2.

Figure 1.3: Panchromatic (Pan) image with spatial resolution of $0.6m \times 0.6m$ corresponding to the same geographical area as shown in Fig. 1.1 acquired using Quickbird satellite. The spectral range of Pan sensor is $0.45 - 0.90\ \mu m$.

1.1.2 Panchromatic Image

Panchromatic (Pan) sensor is a single channel detector sensitive to radiation within a broad wavelength range. Since the wavelength range coincides with the visible range, the resulting image is called Pan (all inclusive) image. The physical quantity being measured is the apparent brightness of the targets. Since the amount of light falling on the Pan sensor is higher when compared to that of MS sensors, the signal to noise ratio is higher in Pan image if both Pan and MS captured at the same spatial resolution. This makes it possible to capture the Pan image with a high spatial resolution without compromising on the SNR. Reason for low spatial resolution in MS image is: SNR will be too low if the spatial resolution increases for MS sensor. Examples of Pan imaging systems are: Ikonos Pan, Quickbird Pan, SPOT Pan, Landsat ETM+ Pan. An example of Pan image captured by Quickbird satellite is shown in Fig. 1.3. One can see that the details are clearly visible in the Pan image. Such high spatial resolution image when used with low spatial resolution MS image (i.e. fused MS image) help us in improving the accuracy in classification and interpretation. The spectral range of the Pan image is $0.45 - 0.90\ \mu m$.

1.2 Low Spatial Resolution Imaging

In order to check the quality of the remotely sensed imagery, both spatial and spectral resolutions are very important. Images with high spatial and spectral resolutions provide the required information for many of the remote sensing tasks. Following is one of the important factors which limit the spatial resolution of the sensors used in remote sensing camera.

- *Instantaneous field of view (IFOV):* It is the angular cone of collected energy by the remote sensing camera. The narrower the IFOV the area of the land mass covered by sensor reduces and hence the amount of light energy collected by the sensor is also decreased resulting in noisy image. If we use a sensor with wide IFOV it covers large area on the earth and hence it results in increase of light energy. Keeping the IFOV small, one can still increase the amount of light falling on the sensor by increasing the spectral width of the sensor. This can result in high spatial resolution which is the case in Pan image. In the case of MS image, it is necessary to keep the spectral width narrow. This causes decrease in signal to noise ratio (SNR) and hence we must increase the IFOV of the sensor so to obtain acceptable SNR. A sensor with a wide IFOV acquires an image at lower spatial resolution. Therefore the sensor hardware in the satellite is constructed to capture low spatial resolution MS images. To overcome this limitation one can use algorithmic approach i.e. multi-resolution image fusion or pan-sharpening to combine the MS and Pan image into single image which has both high spatial and spectral resolutions.

In the remote sensing, earth observing satellites provide MS and Pan data having different spatial, spectral, temporal, and radiometric resolutions as illustrated in Table 1.1-1.4. The need for a single image having the complementary information from both the MS and Pan images has increased. MS image with high spatial and spectral resolutions provide feature enhancement by increasing the accuracy of classification and change detection. The designing of a sensor to provide both high spatial and spectral resolutions is limited by the tradeoff between spectral and spatial resolutions. Hence, there exist number of image processing techniques to combine the available high spectral resolution MS image and high spatial resolution Pan image to obtain an image that has both high spatial and spectral resolutions.

1.3 Fusion in Remotely Sensed Images

Due to the tradeoff between spatial and spectral resolutions of the sensors and other constraints such as bandwidth and on-board storage capabilities of satellite most of the commercial remote sensing satellites such as Quickbird, Ikonos and Worldview-2 capture a single panchromatic image and a set of multispectral images. Pan has high spatial resolution with low spectral resolution while MS image has higher spectral but lower spatial resolution. For example, the Ikonos satellite provides Pan image with $1m \times 1m$ spatial resolution and an MS image with $4m \times 4m$ spatial resolution. These two images are required for the accurate description of captured scene. Since the Pan image has high spatial resolution, it describes subtle details in the scene such as roads, cars, etc. Hence it gives us detailed information of objects and features on the earth's surface. The MS sensors provide multi-band images with color information but with low spatial resolution. They are better suited for the discrimination and/or identification of land type. Also, MS images provide the necessary spectral information for the applications such as classification and hence different objects can be easily identified. These two types of images allow identifying different regions on the ground using the spectral signature on one hand and using the geometrical information on the other hand. In many remote sensing applications, the spatial information is as important as the spectral information. In other words, it is necessary to have images that have spectral resolution of multi-spectral images and the spatial resolution of a panchromatic image. A sensor with high resolution for both is hardly feasible [2]. It is always an interest among remote sensing community to merge these images.

Given the low spatial resolution MS image and high spatial resolution Pan image the *pan-sharpening* or *multi-resolution image fusion* uses an algorithmic approach to enhance the spatial resolution of MS images to make it same as the Pan image. Ideally, the fused image should have spatial resolution of original Pan image and spectral resolution of given low resolution (LR) MS image. Such a fused image can lead to better land classification, map updating, soil analysis, feature extraction etc. The goal of multi-resolution image fusion is to integrate complementary and non redundant information to provide a composite image which could be used for better understanding of the entire scene.

1.4 Multi-resolution Image Fusion: An Ill-posed Inverse Problem

One of the limitations of the low spatial resolution imaging is the mechanism used in the image acquisition process. This mechanism includes the lens subsystem along with the optical sensors which may result in degradation due to out-of-focus and diffraction limit. Distortions may also rise due to the optical aberration or the atmospheric turbulence. In addition to this the speed of shutter and relative motion between camera and object also affect the quality of the captured image. Thus the observed images are degraded that also includes aliasing due to down sampling. In order to solve the image reconstruction problem, one can formulate a mathematical model that represents the image acquisition process. This model, known as observation or forward model, relates the original image to the observed image(s). The accurate formulation of the observation model plays an important role in the success of any image reconstruction approach. The most commonly used forward models incorporate translation, blur, aliasing and noise in the formulation.

The image fusion algorithms attempt to reconstruct the high spatial resolution MS image from the given low resolution MS image and a high resolution Pan image. Note that each MS image is sampled at a rate less than Nyquist rate thereby causing aliasing effect. This is an inverse problem wherein the original information is retrieved from the observed data. A schematic representation of the inverse problem is shown in Figure 1.4. Solving the inverse problem requires inverting the forward transformation. It is difficult to invert the forward model without amplifying the noise present in the observed data and multiple solutions are possible because of more number of unknowns than the knowns. Such problems are called ill-posed inverse problems. While solving the multi-resolution fusion problem the forward model of high resolution (HR) to LR transformation can be reduced to matrix manipulations. Hence it is logical to formulate the fusion problem in a restoration framework as a matrix inversion. Knowing the forward model alone is not sufficient to obtain satisfactory solution. Some form of constraints on the space of solutions must be included. Procedure adopted to stabilize the inversion of ill-posed problems is called regularization. The regularization based approaches solve the ill-posed inverse problems by making them better-posed using the prior information about the solution. It is a systematic method for adding more information to the recon-

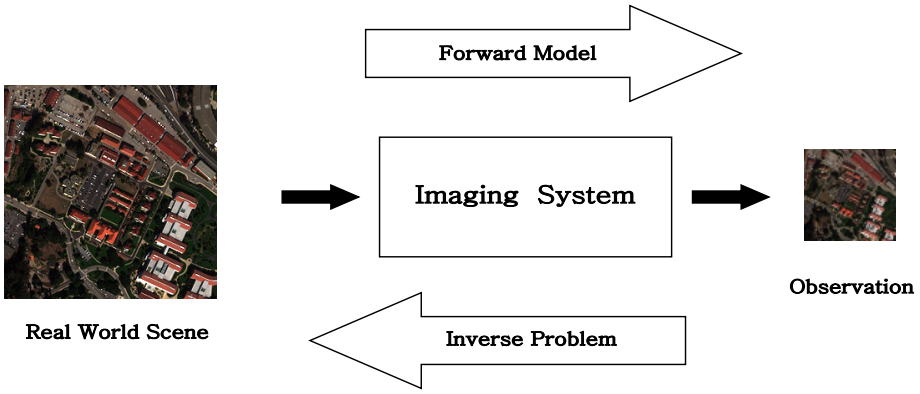


Figure 1.4: Schematic representation of inverse problem. The forward model is a mathematical description of the image degradation process. The inverse problem addresses the issue of reconstructing the original digital image corresponding to the real world scene.

struction system. Bayesian reconstruction approach is commonly employed for solving these problems. This method is used when a posterior probability density function of the original image given the observation can be established. Bayesian estimation distinguishes between the possible solutions by using an priori model for fused image. The major advantages of the Bayesian approach are its robustness and flexibility in modeling noise characteristics and a priori knowledge about the solution and solving using the optimization techniques. In case of convex optimization efficient gradient based methods can be used to obtain the solution which otherwise would require computationally expensive methods such as simulated annealing.

1.5 Applications of Image Fusion

The image fusion is the specific category of data fusion which was started in 1950s. Due to the rapid growth in the advancement of technology and inventions of new sensors a huge collection of data is possible where the provided information are of complementary in nature. Instead of processing this individual sensor output it is always desirable to merge these data in order to increase the throughput of the system. The data fusion process consists the combination and utilization of data originating from the different sources with an aim to obtain information with "greater quality". The meaning of "greater quality" depends upon the application [3,4]. When the given data is in the form of image resultant fusion process is called image fusion. The objectives of fusion differ with the applications. For example in the medical community the feature enhancement is often

required in order to carry out the diagnosis process. The diagnosis could be improved by fusing the different images such as computed tomography (CT), magnetic resonance imaging (MRI), and Positron emission tomography (PET). Similarly the *RGB* camera mounted with thermal sensor provide the images which are very useful in detecting the security threats into public places or military areas. Fusion of these complementary images enhance the capability of the surveillance systems. A single imaging sensor is often unable to provide a complete information of the scene. The process of fusion aims at integrating the complementary information provided by the different sensors for a better representation of the situation than which would have been possible by using any of the sensors individually.

In remote sensing, by using image interpretation an area can be studied without being physically present there. The processing and interpretation of remote sensing images also have specific use in various fields. In geology, for instance, remote sensing can be applied to study and analyze large geographical areas. Remote sensing interpretation makes it easy for geologists to identify the types of rocks and changes in an area occurred due to natural events such as a flood or landslide.

Remote sensing is also helpful in studying vegetation types. Interpretation of remote sensing images allows physical and biogeographers, ecologists, those studying agriculture, and foresters to easily detect which kind of vegetation is present in certain areas and its growth potential. Additionally, those studying urban and other areas of land use are also concerned with the remote sensing because it allows them to easily pick out the used land in an area. This can then be used as data in city planning applications and the study of species habitat. Because of its varied applications and ability to allow users to collect, interpret, and manipulate data over large often not easily accessible and sometimes dangerous areas, remote sensing has become a useful tool for all geographers.

Remote sensing techniques have proven to be powerful tools for the monitoring of the earth's surface and atmosphere on a global, regional, and even local scale, by providing important coverage, mapping and classification of land cover features such as vegetation, soil, water and forests. The volume of remote sensing images continues to grow at an enormous rate due to advances in sensor technology for both high spatial and temporal resolution systems. Consequently, an increasing quantity of image data from satellite sensors have been available, including multi-resolution images, multi-temporal images,

and multi-spectral bands images. The goal of multiple sensor data fusion is to integrate complementary and non redundant information to provide a composite image which could be used to better understanding of the entire scene. It has been widely used in many fields of remote sensing, such as object identification, classification, and change detection. Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times. Change detection is an important process in monitoring and managing natural resources and urban development because it provides quantitative analysis of the spatial distribution of the population of interest. Image fusion for change detection takes advantage of the different configurations of the platforms carrying the sensors. The combination of these temporal images in same place enhances information on changes that might have occurred in the area observed. Sensor image data with low temporal resolution and high spatial resolution can be fused with high temporal resolution data to enhance the changing information of certain ground objects.

MS images with high spatial resolution are desired in many remote sensing applications. High resolution MS images lead to better analysis, classification and interpretation and fusion technique can be considered to improve the spatial resolution of the land area. The fused images of land fields can lead to accurate estimate of types of crops. The fused images of geographical land area help in better segmentation of regions containing forests, rivers, roads and other geographical structures.

The remote sensing satellite captures the same geographical area at the regular interval depending on the temporal resolution of that satellite. The availability of multi-temporal data sets over the same scene makes it possible to extract valuable temporal characteristics of surface cover types that may be of interest to applications requiring the monitoring of spectral or spatial characteristic changes over time. It also helps in crop monitoring, climate change and during the period of natural disaster.

One of the shortcomings of the MS image is the limited number of bands in the electromagnetic spectrum with wide spectral width of the individual band. This do not provide the contiguous and dense spectral bands which is often required in order to accurate discrimination of the materials present in the scene. This can be accomplished by acquiring large number of images with relatively less spectral width. In remote sensing this is referred as hyperspectral imaging. The hyperspectral image provides a densely sampled and almost continuous spectral information over the given wavelengths. Thus

essentially it capture even minor variations in the scene reflectance. Although the hyper-spectral data results with better classification of the regions the processing and analysis of the same requires large computational since it includes many numbers of images. In this thesis we are addressing the problem of resolution enhancement of MS images only.

1.6 Contributions of the Thesis

In this thesis we solve the problem of multi-resolution image fusion. In this problem, the given high spectral resolution MS image and high spatial resolution Pan image are combined to give a single fused image which has both high spatial and high spectral resolutions. Since the MS sensor is sensitive to particular spectral range only, it is required to increase the size of detector to get adequate amount of light to have acceptable SNR. Because of this the spatial resolution of the MS sensor is restricted when compared to that of Pan sensor. On the other hand, Pan has wider spectral range which makes it with poor spectral resolution and due to this the size of detector is decreased and the spatial resolution becomes high. In other words, the multi-resolution image fusion requires to infer the missing high frequency details of MS image from the high spatial resolution Pan image. In this thesis we address two fusion approaches in which the extracted high frequency details from the Pan image are injected into the MS image using edge preserving filters. In addition to this, we also propose the model based solutions to the multi-resolution image fusion.

- The Pan image has high spatial resolution. One could think of extracting the high frequency details from the Pan image and inject the same into the MS image to obtain the fused image. In this case the quality of the fused image depends on two important points. First is the details extraction process and second is the injection model using which the extracted details are injected into the MS image. Hence we start our work of fusion using edge extraction methods. We use the edge preserving filters such as guided filter and difference of Gaussians in order to extract the details from the Pan image. Motivation behind choosing these filters is the versatile use of the same in the various applications of feature extraction in the computer vision community. The extracted details are injected to the upsampled MS image after weighting them with scaling factor calculated based on the MS pixel intensity values.

We conducted the experiments on different satellite images and also compared the results with the state of the art methods. Along with the qualitative evaluation we also perform the quantitative analysis by computing the different measures.

- The main limitation of the fusion techniques based on edge preserving filters is the upsampling operation of the MS image. In order to overcome this limitation we propose the model based approach. In this approach we model the given low resolution MS image as blurred and noisy version of the unknown fused MS image. In this model the degradation due to the downsampling can be estimated which takes care of aliasing. In order to estimate the degradation ideally we require the true fused MS image, however it is unknown in our case. In this situation we use the approximation of the true MS image. This initial HR approximation is obtained using the directional transforms such as contourlet transform (CT) and non-subsampled CT (NSCT). Motivation for choosing these transforms is due to their properties such as multi-scale decomposition and higher directionality. Using the initial HR approximation and the given MS image, we obtain the degradation. Since the problem of model based fusion is ill-posed inverse regularization is required to obtain the final solution, A maximum a *posteriori* Markov random field (MAP-MRF) framework is used to obtain the final cost function in which smoothness prior is used resulting in convex cost function and same is minimized using gradient descent optimization technique. In order to show the efficacy of the proposed fusion method, experiments are conducted on the different datasets captured using satellites such as Ikono-2, Quickbird and Worldview-2. Along with perceptual comparison of the proposed approach with the state of the art methods using degraded and un-degraded datasets we have also conducted the quantitative evaluation by calculating various spatial and spectral measures.

- In the same regularization framework we also cast the fusion problem with the patchwise estimation of degradation. To do this we first obtain pairs of LR and HR patches from the given MS observation using the concepts of self-similarity and compressive sensing. In the regularization framework we use a new Gabor prior to extract the bandpass spatial details from the Pan image. The potential of the proposed fusion approach is verified by conducting the different experiments on the

images of various satellites. Similar to the earlier approach, here also the efficacy of the proposed method is evaluated by conducting the experiments on degraded as well as on un-degraded datasets of three different satellites i.e., Ikonos-2, Quickbird and Worldview-2. The results are compared on the basis of traditional measures as well as recently proposed quality with no reference (QNR) measure which does not require the reference image.

1.7 Objective of the thesis

Images with high spectral and spatial resolution provide accurate details of the earth. This information is required in many of the remote sensing tasks such as classification, change detection, etc. However, due to the hardware limitation the acquired MS image low spatial resolution though it has high spectral resolution. Pan image has high spatial resolution with poor spectral details. Thus there exist need for a single image having complementary information from both the MS and Pan image. This has motivated us to propose algorithmic approach to combine MS and Pan image pixels which can better represent the information of both the images. Following are the objectives of the thesis:

- To develop an algorithm which capture the spatial details from Pan data to provide a high spatial and spectral resolution MS image which can be used for better understanding of the scene.
- Formulate the fusion process as an inverse problem to estimate degradation arising due to aliasing.
- Use regularization framework to obtain the better solution to the ill-posed problem.

1.8 Organizations of the Thesis

In this thesis we propose new approaches for multi-resolution image fusion. We first consider this problem by using detail extraction from Pan image. Then we address the fusion problem using model based approach. The effect of aliasing is considered in the model by estimating the degradation matrix. We then use the regularization framework

to obtain the fused image. Along with the smoothness prior we also propose a new prior called Gabor prior in order to extract the high frequency details from the Pan image.

The literature survey for different fusion methods are described in chapter 2. Different approaches for extraction of details along with the various injection models for carrying out multi-resolution fusion are the topics of discussion in this chapter.

In chapter 3, two new approaches for multi-resolution image fusion by using different edge preserving filters are proposed. We have chosen the guided filter and difference of Gaussians (DoGs) for detail extraction. In a guided filter based technique, the Pan and MS images are used to extract the missing high frequency details. In the second technique, the difference of Gaussians is used as a band-pass filter to discard all but a handful of spatial frequencies that are present in the Pan image. The comparison of the results obtained using these proposed fusion approaches is shown with various state of the art methods.

The undersampling of MS images introduces aliasing in the images. The filter based methods presented earlier do not take care of aliasing. In chapter 4, we present a model based approach that takes care of aliasing due to undersampling of MS observation. The the LR MS image is modeled as the blurred and noisy version of its ideal HR fused image. The degradation between LR and HR MS images is estimated by first estimating an initial approximation to fused image. Results obtained using the proposed model based approach are compared and discussed with the other existing approaches.

In order to increase the accuracy, one needs to derive the initial estimate using the available LR MS image only. With this motivation, we next tackle the problem of image fusion using the concepts of self-similarity and compressive sensing. We obtain the degradation estimation and propose a new prior based on Gabor filter in order to extract the details from the Pan image. The details of this are given in chapter 5.

We summarize our works in the form of conclusions and possible future directions in chapter 6.

Chapter 2

Literature Review

In many remote sensing applications, the spatial information of a fused image is as important as the spectral information. In other words, it is necessary to have images that have spectral resolution of multi-spectral images and the spatial resolution of a panchromatic image. A sensor with high resolution for both, spatial and spectral, at the same time is hardly feasible [2]. The coarse spatial resolution of MS images is the result of a trade-off due to physical and technical constraints. The quantity of energy which arrives onto the detector is proportional to the width of its spectral range and hence is smaller in the MS sensor than in the Pan sensor. It is therefore necessary to increase the energy that impinges on the MS detector to obtain satisfactory signal-to-noise ratio. However, this is not possible due to technological limitations. Further, if MS images were having high resolution, the amount of data to transmit would be larger. The on-board storage and data transmission to the ground also restrict the spatial resolution of MS images. This makes the remote sensing satellite sensors to acquire MS images with low spatial resolution and Pan image with high spatial resolution. Thus the MS images have high spectral but low spatial resolution and the Pan has high spatial but low spectral resolution.

The goal of multi-resolution fusion or pan-sharpening is to combine high spatial resolution of the panchromatic image with the high spectral information of the MS image. The fused MS image should have high spatial resolution in order to aid in the detection and classification tasks. It should also contain the same spectral (color) information as the original multi-spectral data for better identification of targets. In other words the fused image should possess both high spatial quality and spectral quality.

In the remote sensing community, multi-resolution image fusion or pan-sharpening

is a challenging problem. During the last two decades, a great amount of research has been carried out in this field. SPOT 1, the satellite launched in 1986 had provided the Pan and MS images with spatial resolution of $10m \times 10m$ and $20m \times 20m$, respectively. Since then the problem of multi-resolution image fusion has drawn significant attention of remote sensing community in developing algorithms for better fusion. In 1999, the first review article on the different fusion methodologies was presented by Pohl *et. al.* [5]. Here, authors have described different fusion techniques which are divided into different categories based on arithmetic operations, wavelet transforms and PCA transform. This article also includes the different applications of image fusion such as land usage, flood monitoring and applications based on geology.

The multi-resolution fusion techniques can be classified in number of ways. In [6] authors have classified these methods based on spatial and transform domains. In the methods based on spatial domain, the processing is performed directly on the pixel intensity values. However, in the transform domain methods, the test images are first converted to domains such as Fourier or discrete cosine transform, etc., and the processing takes place in the transform domain itself. Here, inverse transform is necessary to obtain the final fused image in the spatial domain. The fusion methods can also be grouped based on the approaches used in obtaining the final fused image. One such classification is reported in [7] where authors classify the pan-sharpening approaches into three different groups: projection substitution, relative spectral contribution and the fusion methods based on mlioration de la rsolution spatiale par injection de structures (ARSIS) concept. The ARSIS is a French word which means to improve the spatial resolution by structure injection. Authors in [8] discuss the framework of general image fusion (GIF) which consist of the classification as well as comparison and comparative analysis of the existing fusion methods.

To understand the concepts behind different multi-resolution image fusion techniques in a better way here we classify them into three different categories as projection-substitution methods, multi-resolution analysis (MRA) based methods and model based approaches. In the following sections we describe the different fusion methods under these categories. Note that in this thesis the terms “pan-sharpening” and “multi-resolution image fusion” are used interchangeably.

2.1 Projection Substitution based Techniques

A multispectral data captured by a remote sensing satellite has the collection of several monochromatic images. Each of this image is referred to as band which is captured using a sensor with a particular spectral range. For example Landsat-7 satellite captures the MS images having seven bands in the spectral range from 0.45 to 1.25 μm. One can visualize the MS image by consisting any three bands as red, green and blue images in the RGB color space format. However, same can also be projected onto other color space such as intensity hue saturation (IHS) where I describes the luminance component of the scene, H refers to the contribution of colors and saturation component represents the purity of colors. The human visual system also works on the principle of IHS color space to identify/describe the objects being imagined by us. The IHS transformation converts the given RGB image into IHS color space where intensity component corresponds to the spatial information of the image and hue and saturation components together represent the spectral details of the scene [9]. Since the IHS transformation separates the spatial and spectral details one can use the same to fuse the MS and Pan images i.e., the spatial component resulted in IHS color space can be manipulated using a mathematical operation without disturbing the color details. This category of the fusion classification not only includes the fusion methods based on IHS [10–17] but it also includes those based on principal component analysis (PCA) [18–21] which are proposed in early 1990s.

In these approaches one has to use suitable interpolation technique to make the size of MS image same as the Pan image before the suitable fusion method is applied. Also, since the MS and Pan images are of complement characteristics it is required to perform the radiometric corrections on the Pan image using the histogram matching between the intensity component of the upsmapled MS image and the Pan image. The histogram matched Pan image is then substituted into the intensity component in the IHS color space to get the pan-sharpened image in the IHS color space. The final fused image is obtained after taking the inverse IHS transform in the last step. An alternative to the IHS transform is the PCA which transforms the intercorelated bands into the set of uncorelated bands. After this, first principal component i.e. PC^1 is replaced with the Pan image since it represents an image with highest variance. Histogram matching of Pan to PC^1 is mandatory before substitution because the mean and variance of PC^1 are

different than those of Pan image. In addition to fusion methods based on IHS and PCA under this category of classification, we also include the pan-sharpening methods based on the concept of relative spectral contribution [7]. In this concept, it is assumed that the spatially degraded Pan image is a linear combination of acquired MS images. This assumption is based on the observation that the spectral response functions of Pan image overlaps with the various bands of MS image. Some of the common methods that fall under this category are Brovey [22, 23] and P+XS [24] fusion methods.

One of the drawbacks in the IHS based fusion methods is that it can be applied on three bands only. Due to this the Pan image do not correspond to better representation of luminance component in the entire set which affects the quality of fused image severely. Literature shows that the performance of the fusion methods based on IHS, PCA transform and relative spectral contribution is highly depended on the correlation between the intensity component of the MS and the Pan image [25]. Although these fusion methods result in a fused image with better preservation of geometrical structures [26–28] which are well suited for the applications such as cartography, visual analysis and target recognition [14, 28] the major drawback in these fusion methods is the spectral or color distortion which may be localized to certain land area [14, 29]. The reason behind this spectral distortion is due to the modification or alteration of the low frequency details present in the MS image during the injection of high frequency details from Pan image [7, 30]. In addition to this if the spectral responses of the MS bands are not perfectly overlapped with that of Pan image the performance of the fusion techniques based on PCA or IHS and relative spectral contribution results in poor fused data [31].

In order to overcome the three band limitation of IHS fusion method Tu *et. al.* [32, 33] have proposed a framework to use more than three bands for fusion using IHS transform. This work is considered to be a pioneer in extending the IHS transform for further use in multi-resolution image fusion. This fusion method is referred to as generalized IHS (GIHS) or fast IHS (FIHS) in the literature. Here weights are fixed to compute the intensity component. In order to improve the quality of the fused image researchers have modified the framework of FIHS method in [33].

Unlike the FIHS method [33] in which fixed weights are used to compute the intensity component, authors in [34] estimate the weights and the same are computed using the genetic algorithm. They also use the same algorithm to compute the appropriate

gain constants that are multiplied with the extracted details of the Pan image before injecting them into the upsampled MS image. Although this method obtains better pan-sharpened image it suffers from the computational issues since it requires the population based optimization techniques such as genetic algorithm to estimate the weights. The computational complexity of FIHS fusion method [33] is lesser than that of transform based fusion method. In [31] Gonzalez *et. al.* utilize the spectral response functions of MS and Pan images in order to compute the intensity component of the IHS transform which further reduces the computational burden of FIHS method. They verified their fusion method by conducting the experiments on the images captured using Ikonos satellite. The disadvantage of the this method is that one has to know the spectral responses of the given images in order to obtain better fusion performance. Similar to the method in [31] the approach proposed in [35] also has lesser computational complexity. In this case the radiometric properties of MS and Pan data are used in selecting the high frequency details from the Pan image and are injected into the upsampled MS image. In addition to this the weights on the extracted Pan details are also computed before injecting it to MS image. As done in [31] here also the authors verify the potential of their method using the images acquired by Ikonos satellite. Although the fusion methods proposed in [31, 35] have the advantage in terms of reduced computational complexity they require the knowledge of spectral response functions of the data and this restricts their usefulness only to limited sensors.

In [36] the authors obtain better solution for fusion when compared to FIHS technique of [33] by applying adaptive component substitution. Here, the final fused image is obtained after estimating the initial low and high resolution component images by computing the statistical parameters such as correlation coefficient (CC) and standard deviation between MS and Pan images. An adaptive IHS method proposed in [37] estimates the different weights used in the component substitution. In addition, in this case the final fused image is obtained by inserting the edge details from the Pan image estimated using the exponential edge detector.

The IHS fusion model is also generalized to include any projection substitution techniques in [38]. Here, a generalized intensity component is modeled as the weighted average of the MS bands which is similar to the concept of fusion methods based on relative spectral contribution. The weights are obtained as regression coefficients between the MS

observations and the spatially degraded Pan image with an aim of capturing the spectral responses of the sensors. These weights are then used in obtaining the intensity component from the MS bands which is then injected into the upsampled version of the MS image to be fused in order to obtain the final fused image.

Authors in [39] introduce a tradeoff parameter in order to control the amount of injected details from Pan to MS image. They formulate the fusion problem as minimization problem in which the difference between upsampled and fused MS images is minimized. At the same time the fused image should be as close as possible to the high spatial resolution Pan image. The tradeoff parameter is involved between these two constraints which controls the spectral and spectral resolutions in the final fused image. The range of tradeoff parameter is from one to infinity which varies the fused image from MS image (no spatial details from Pan image) to the fused image obtained using FIHS method [33], respectively. They conclude that the spatial and spectral resolutions can not be obtained simultaneously. An improvement to this model is proposed by the authors in [40] where the tradeoff parameter is automatically adjusted. The idea of tradeoff parameter is further extended by researchers in [41] where the tradeoff parameter is obtained by using wavelet transform.

A method based on the Gram-Schmidt (GS) orthogonalization procedure is proposed in [42] in order to obtain the fused image. This approach is patented by Eastman Kodak and implemented in the environment for visualizing images (ENVI) software package. In this method the fused image can be obtained by selecting the one of the different versions of the spatially degraded Pan image. Later the authors in [38] modified this preprocessing step of Pan image by using the multiple regression weights estimation on the MS bands which has been proven to be an effective fusion method using GS orthogonalization. Another fusion method based on the concept of the matting model is proposed in [43]. This model decomposes the given MS image into three components, i.e., alpha channel, spectral foreground and background. Similar to the concept of IHS method, here also the alpha channel is substituted by Pan image and inverse operation is then applied to obtain the fused image. A fusion approach that uses the smoothing filter-based intensity modulation (SFIM) is proposed in [44]. Their method modulates the intensity values of low spatial resolution MS image based on the ratio of Pan and its degraded version. The proposed method was shown to perform superior when compared to fusion techniques

based on Brovey transform [22] and IHS transform [33]. In order to remove the distortion in fusion caused due to the bicubic interpolation in [44], a new fusion method is proposed in [45] which uses the nonlinear upscaling operation instead of bicubic interpolation that gives a sharper and better correlated upscaled image. The frequency details from the Pan image is extracted using a pair of upscaling and downscaling filters which are then added to the upscaled image. An adaptive pan-sharpening method based on smoothing filter intensity modulation (SFIM) is also proposed in [46].

Though the above fusion algorithms have certainly advantages they result in fused image with some degradations in terms of spectral as well as spatial contents. The IHS transform based approaches enhance the texture features however they make use of interpolated version of MS images. Due to this interpolation prior to the replacement of the high frequency details from the Pan image it causes the spatial degradation in the fused images. The projection substitution based fusion techniques also result in spectral distortion due to the dissimilarity between intensity component and the the Pan data [25]. The IHS based methods require much less computational burden when compared to fusion methods based on transform domain and are easy to implement. However, they still suffer from spatial and spectral distortions.

2.2 Multi-resolution based Techniques

In the last decade, pan-sharpening techniques based on multi-resolution analysis (MRA) have become significant due to their ability to capture the information present at different scales. These methods work on the following principle: extract the high frequency details not available in MS image from the Pan image and inject those details to the MS image. This concept of detail extraction and injection was pioneered by Chavez in [47] which was based on high pass filtering (HPF). Here, author extracted the high frequency details from the Pan image by taking the difference of Pan and its low pass filtered version usually blurred with the box type filter. These extracted details are then injected into the upsampled MS image to obtain the fused image. The comparison of fusion methods based on IHS, PCA and HPF was done in [18] which shows the better performance of HPF method when compared to IHS and PCA based fusion techniques. Later the improvement in terms of detail extraction and injection was effectively handled by the development of

MRA i.e., wavelet, contourlet, etc., to carry out data fusion task.

MRA calculates the approximation of signals at various scales, which is done by using wavelet basis functions. Theory of wavelet was introduced in the beginning of the last century and is applied to perform signal processing tasks since the year 1980. Due to its desirable properties such as multi-scale decomposition and time-frequency localization it is a very promising tool in the various signal and image processing applications. While performing fusion using MRA, the details extracted from an image using the wavelet transform can be injected into that of another image using number of methods. For example the method can be based on substitution, addition, or a selection on either frequency or spatial context. The first effective MRA in this domain is a decimated discrete wavelet transform (DWT) which is orthogonal, non-redundant and non-symmetric directional transform. It captures the detail information present at horizontal, vertical and diagonal directions at each stage of decompositions. Various approaches have been proposed for pan-sharpening that are based on the use of DWT [26, 48–52]. In these methods, first the histogram matching is performed between Pan and each band of MS images. The histogram matched Pan image and each band of interpolated MS image are then decomposed separately using DWT. Finally, the approximation and detail layers obtained at different levels of decompositions of MS and Pan images are used in obtaining the fused image. Here, the extracted details from Pan may be added to or they may be substituted for MS image pixels to obtain the fused image. The inverse DWT is performed on the modified MS image to obtain the final fused image in the spatial domain. It has been found that the fusion results obtained by using DWT with additive or substitution fusion rule perform similar in terms of quality of final pan-sharpened image [53]. Although MRA using DWT provides effective tool to carry out data-fusion tasks the process of detail injection may result in spatial distortions such as blurring, ringing as well as aliasing effects in the fused images which are mainly due to the shift variant property of DWT [54–56]. However, multi-resolution representation of image using DWT makes the fusion results better when compared to that of HPF or IHS based methods which do not use such hierarchical description to describe the image.

In order to avoid the problem of shift variance of DWT, researchers in the fusion community use either “a trous” wavelet transform (AWT) or Laplacian pyramid (LP) to decompose the image. Unlike the DWT which is critically subsampled, the AWT

and LP are oversampled which allow an image to be decomposed into nearly disjointed bandpass channels in the frequency domain without losing the spatial connectivity of its high frequency contents such as edges regions. Use of AWT or LP leads to a stack of image layers obtained from different band-pass filters with same dimensions and with the reduction in resolution by factor of 2 for each level. The AWT is a nonorthogonal, redundant, undecimated and symmetric directional transform. It was first proposed by authors in [57] for music synthesis. The term "a trous" ("with holes") was introduced by Dutilleux [58] and its theoretical analysis is described in [59]. Various fusion schemes using AWT are presented in [56, 60, 61]. In [56] AWT with additive fusion rule is used to obtain the fused image by making the use of three MS three bands. This approach was later generalized for more than three bands and is often referred as additive wavelet luminance proportional (AWLP) [60]. Authors in [60] also propose a model for multi-resolution fusion using the spectral response function of the MS and Pan images provided by the manufacturer. In [61] Ranchin *et. al.* propose a framework for fusion using AWT with different detail injection methods. The comparison of fusion results obtained using DWT and AWT is reported in [53] where authors conclude that AWT outperforms DWT.

In addition to AWT, the MRA based on LP also represents an undecimated transform [62]. The theory of LP was introduced prior to MRA by Burt and Adelson [63] for compact image representation. Similar to AWT, LP is a bandpass image decomposition derived using the Gaussian pyramid (GP) which is a multi-resolution image representation. This can be regarded as an AWT in which the image is recursively low pass filtered and downsampled to generate a low pass approximation image, which is re-expanded and subtracted pixel by pixel from the original image to yield the detail layer. The LP is used in [61, 64] to obtain the multi-resolution fusion of MS and Pan images while it is used by authors in [65] to obtain the fusion of hyperspectral data. One of the benefits from the AWT and LP based pan-sharpening methods is that by generalization of these methods it is easy to fuse MS and Pan images of having non-integer or even fractional resolution difference. One such implementation is presented in [54] where the authors propose the fusion of the MS and Pan images having non-octave or fractional resolution ratio. The generalization of these works are proposed in [66–68] in which the idea of multi-rate signal processing [69] is used. The common characteristic of the AWT and LP is that they are redundant multi-resolution transforms which are well suited for image

fusion as demonstrated in [53, 54]. It is interesting to note that the use of undecimated DWT (UDWT) as MRA to decompose the image may also be useful to obtain the fused image since it is also an oversampled transform. However, the computational cost along with complexity issues are high when it is intended to fuse images having non-octave or fractional resolution ratio.

A number of fusion methods based on MRA framework that use various methods for details extraction and injection are modeled in a generic way by using ARSIS concept [51,61]. Authors in [7,25] classify the fusion methods based on MRA using ARSIS concept which comprises of three different models. First one is the multi-scale model (MSM) that performs hierarchical description of the information content related to the spatial contents in an image. Examples of MSM include DWT, LP, AWT and UDWT which have been already discussed earlier. In addition to these it also includes Gaussian pyramids (GP) [70], HPF [18] and iterative filter banks [71]. The second one in the ARSIS concept is the inter band structure model (IBSM) which deals with the relationship between the approximation and/or details of MS and images. It uses a radiometric transformation such as gain or offset of spatial structures when injecting details from Pan to MS image. The final model is the high resolution IBSM (HRIBSM) which is concerned with the procedure to inject the extracted details from Pan to MS image. For example the fusion method proposed in [55] considers the modulation transfer function (MTF) of MS images during the injection process. Note that the fusion implementations with the choice of MSM, IBSM and HRIBSM make the different pan-sharpening methods affect the quality of the fusion method.

In ARSIS concept authors also elaborate the examples of successful implementation of image fusion using AWT and LP as MSMs and the results are compared using different IBSMs. IBSM aims to achieve the goal of making the fused image most similar to the original MS image. In [51] Ranchin and Wald demonstrate the use of two local models for image fusion by making use of details instead of approximations. The need for such IBSM in the fusion process is discussed in [72]. Here, they show the effect of both global and local IBSMs on the results of pan-sharpening using redundant transforms i.e., AWT and LP and conclude that the local IBSM outperforms when compared to global one. They use the local IBSM as proposed in [54] which consist of space-varying gain. Here, Pan detail coefficients are multiplied with gain constant to achieve local equalization of

the MS and Pan sensors.

Few researchers have worked on fusing Pan and MS images where the resolution difference is not a power of 2 e.g., fusion of systeme pour l' observation de la terre (SPOT) Pan and Landsat thematic mapper (TM) MS bands. Blanc *et. al.* [71] use the ARSIS concept with iterated rational filter banks to pan-sharpen the MS image and they compare their results with the method based on DWT which is designed for non-integer resolution factor and show that DWT performs poor for non-integer case. Use of the multi-band DWT to merge SPOT Pan and Landsat TM MS images is discussed in [73]. Using the framework of ARSIS concept Aiazzi *et. al.* [54] also propose the generalized LP using context-based decision (GLP-CBD) to merge the data of non-integer or fractional resolution difference between the Pan and MS images.

In order to exploit the benefits of projection substitution based fusion techniques several pan-sharpening approaches have been proposed by combining MRA and projection substitution. Nunez *et. al.* [56] combine the IHS transform and the wavelet based MRA to obtain a fused image. A fusion method that uses PCA and wavelet transform is discussed in [74]. The authors in [75] and [76] combine the IHS with bi-orthogonal and redundant wavelet transforms, respectively in order to obtain the pan-sharpened image. Hong and Zhang in [77] use integrated IHS and wavelet to fuse the MS and Pan images and they demonstrate their results on images acquired from Quickbird and Ikonos satellites. Gonzalez *et. al.* [12] discuss the framework to merge the MS and Pan images by integrating wavelet transform with the IHS and PCA methods, respectively. Recently, a detail injection model based on MTF [55] is used in the fusion approach of [56] to obtain the fusion [78]. The main advantage of combining MRA and projection substitution methods is to obtain the fused image with better perceptual and quantitative measures.

Though wavelet transform preserves spectral information efficiently, it lacks in preservation of spatial fidelity due to limited directionality. Moreover, the isotropic wavelets are scant of shift-invariance and multi-directionality and fail to provide an optimal solution for highly anisotropic edges and contours that are encountered in images. Also, due to the limited directionality the edge extraction and injection may not correspond to those in true MS images. In order to solve these limitations, the MRA based fusion methods using transforms such as curvelet [79] and contourlet [80, 81] have also been proposed. Curvelet has the property of anisotropy and has higher directional property which makes

it better suited for image fusion than the wavelet transform. In [82] Choi *et. al.* fuse the MS and Pan images using ARSIS concept that uses curvelet transform. They compare their fusion results with that of IHS and wavelet based methods. Similarly, the authors in [83] propose a fusion approach based on curvelet transform for resolution factor of 2 and 4 for Quickbird and Ikonos dataset. They use the curvelet transform to extract the high frequency details from the Pan image and inject them into the corresponding details of MS bands. They evaluate perceptual and quantitative test by comparing the fusion results with traditional methods.

The construction of curvelet transform requires a rotation operation and also it corresponds to a partition of the 2D frequency plane based on polar coordinates. This property makes the idea of curvelet transform simple in the continuous domain but causes problems in the implementation for discrete images. Similar to the curvelet transform, the contourlet transform (CT) proposed in [80] has the properties of wavelet transform. In addition it also possesses the characteristics such as multi-directionality and anisotropy which captures the details present in different directions. It employs LP [84] and directional filter bank (DFB) [85] to capture and link the discontinuities into geometric structures. The disadvantage of CT is that it is shift variant whenever subsampling is used. The shift invariance version of CT called non-subsampled CT (NSCT) [81] performs better in fusion due to the absence of decimation but it suffers from redundancy. The fusion methods based on CT and NSCT can be found in [86–88]. In [86] Shah and Younan obtain the fusion of MS and Pan images using the PCA and NSCT. Here, after applying the PCA on upsampled MS images they choose a principal component (PC) image that has high correlatedness with Pan image. The selected component and histogram matched Pan image are then decomposed using the NSCT. The detail coefficients of the Pan image are then replaced the corresponding selected component in the CT domain in order to obtain the fused image. A similar approach is presented in [87], where the transformed image corresponding to first PC and Pan images are decomposed using NSCT and details are injected based on local variance of Pan and first principal component images. Authors in [88] consider both spectral and spatial similarity while fusing. They decompose the Pan and MS images using NSCT and the detail coefficients of Pan image are injected to MS image based on spatial similarity measure. Similar to [86] they inject detail coefficients at all levels from Pan to MS image.

In addition to the standard DWT, the compl wavelet transform (CWT) is a complex-valued is a two-dimensional wavelet transform which provides multi-resolution, sparse representation, and useful characterization of the structure of an image [89–91]. Steerable filterbanks [92] proposed by Simoncelli was one of the first nonadaptive designs that provided an elegant way to obtain flexible orientation specificity with redundant representation. Attempts have been made to achieve near shift-invariance and improved directionality using the idea of near-analytic transform designs such as CWT with small redundancy. Salient features of these complex transform designs are shiftability i.e., near shift-invariance, increased directionality and the availability of phase information. In [93], Nguyen and Oraintara proposed a shiftable complex directional transform by combining Laplacian pyramid and complex directional filter bank. Here in order to obtain the analyticity, dual-tree structure of real DFBs is constructed where the fan filters used in two trees are constrained to satisfy the Hilbert pair criteria and certain conditions on phase responses. In [94] the same authors addressed the implementation issues such as border artifacts and the constraints on the designed finite impulse response (FIR) filters. However these FIR filters are truncated versions of infinite impulse response (IIR) filters and the transform is approximately shift-invariant. In [95], authors proposed complex-valued steerable filter banks for texture synthesis application and utilize the same for finding features based on local phase and energy. Use of subsampled MDFBs and CWTs in many image processing applications result in suboptimal performance due to the use of downsamplers. Down-sampling stages used in these transforms lead to large reconstruction error (caused by aliasing), limited directional flexibility, difficulty in filter design etc. Subsampled complex-valued transforms are used in [96–98] for the same. Approach in [96] uses DT-CWT [89], while the approaches in [97, 98] use shiftable complex directional transform [93]. Note that all these MRA-based pan-sharpening approaches are nonadaptive and to the authors best knowledge use of spatially adaptive transform-based pan-sharpening approach has not been proposed yet.

In the literature number of methods are discussed that use different approaches for fusion. In order to compare these methods the data fusion committee of the IEEE geoscience and remote sensing society (GRS-S) sponsors a yearly “Data Fusion Contest” focusing on one specific application each year. The Data Fusion Contest-2006 was held with an aim to identify the best pan-sharpening algorithm. Fusion techniques that were

based projection substitution as well as MRA were included in this contest. Outcomes of this contest are presented in [4]. An important observation of the evaluation results was that two best performing algorithms, GLP-CBD [54] and AWLP [60] based on MRA using AWT and LP, respectively share a common philosophy of taking into account the imaging sensor related physical models in the algorithm. These techniques are widely used as state of the art methods. However, AWLP and GLP-CBD methods have the limitation that the quality of the fused image depends on size of the target present in an image. Depending on the size of target either AWLP [60] or GLP-CBD [54] performs better. In [99] the authors have proposed an approach to use either AWLP [60] or GLP-CBD [54] based on the target size. They have shown that AWLP perform better when the target size is small and similarly GLP-CBD performs better when the target is larger.

2.3 Model based Fusion Approaches

While a majority of pan-sharpening methods adopt a projection-substitution, MRA or a combination of these two, techniques independent of these frameworks have also been proposed in the literature. The most relevant techniques to our discussion here are the methods that based on *model based approach*, where the multi-resolution image fusion is posed as minimizing a cost function based on a model of the imaging sensor. All those approaches addressed earlier do not use an explicit relation between the observed and fused images. The model based fusion methods are partly motivated by image restoration and super-resolution research which use the image formation model and obtain the solution by formulating it as an inverse optimization process.

The image formation involves the degradation of the true high resolution MS image. Because much information is lost in the this process, estimating a true MS image from its degraded version is an ill-posed inverse problem where solution is not unique. This means that according to the degradation process, many different high resolution MS images can produce the same low resolution MS image. Regularization theory is an effective framework to obtain the better solution for the ill-posed problem. Similarly, Bayesian framework has also been proven to be an important tool to solve such inverse problems. Within this framework, the fused image is extracted from Bayesian posterior distribution in which the prior knowledge and artificial constraints on the fusion results are incorpo-

rated. These model based fusion methods utilize the image formation models and regard the fusion process as an inverse optimization problem. During recent years, several model based fusion techniques have been proposed which are based on a regularization theory and/or on a Bayesian framework. Generally, there are two types of image formation models used in these methods, first is based on the LR MS formation which relates the high resolution MS image to its LR version and the second is the formation of Pan image which relates the given Pan image to the true MS image.

The main advantage of a Bayesian formulation is that the problem gets converted into the probabilistic framework. Using this idea authors in [100] have proposed the fusion of MS and Pan images captured using Ikonos satellite. They use the LR MS and Pan image formation models and assume that the resulting error components follow the Gaussian distribution. In a Bayesian framework it is also required to define the prior knowledge of the fused image. Here, authors used a non informative model for the fused image since the fused data itself is unavailable i.e. probability density function of the fused data is uniform. The weighting coefficient between spectral and spatial details is also introduced in the same Bayesian formulation in this method. Within Bayesian framework, maximizing a posterior distribution in the form of maximum a *posteriori* (MAP) has resulted in improved solution in many areas of image processing. An MAP estimation for enhancing the resolution of hyperspectral data was proposed by Hardie *et. al.* [101]. In their work they obtained two terms in the final cost function, first is due to the result of the LR MS image formation and second term is due to the correlation between the different data to be merged.

In [102, 103] authors use LR MS image formation model where they model an LR MS image as a blurred and decimated version of the true HR MS image. Using MAP estimation final cost function is obtained where appropriate priors are used in order to regularize the final solution. The Pan image formation model is not used in these fusion techniques instead same (i.e., Pan image) is used to estimate the different parameters of the priors with the assumption that all the MS bands have similar spatial structures to the Pan image. In [102] Joshi *et. al.* have used the autoregressive prior for fusion and different parameters of this prior were estimated using the Pan image. It is well known that Markov random field (MRF) based modeling for capturing the spatial correlation among pixels is the most general model and is often used as a prior during regularization while solving

the ill-posed problems. In the area of image processing and computer vision, many researchers use the MRFs as a convenient way of modeling the contextual entities such as image pixels, depth field, and other correlated features. In [103] authors have modeled the final fused image as an MRF and have used an edge preserving inhomogeneous Gaussian MRF (IGMRF) prior in the regularization framework. They estimate IGMRF prior parameters using the available Pan image. Since the final cost function in the methods proposed in [102, 103] is convex simple gradient based iterative optimization method is used to obtain the final solution. Using MRF based prior Aanaes *et. al.* [104] have also solved the multi-resolution fusion problem. Here, the spatial neighborhood weights for the MRF are computed from the Pan image to transfer edge information from Pan to the fused image.

The LR MS image formation model without decimation is used in [105] where authors have assumed that the interpolated LR MS image is the degraded versions of its fused MS image and hence they do not consider downsampling operation in their model. Along with the LR MS image formation they also use the Pan image formation where the original Pan image is modeled as a linear combination of all the bands of the fused MS image. The final cost function is derived using discrete sine transform and the solution is obtained by using constraint least squares method. Similar model is also employed in [106] where gradient field of modified Pan image is used for spatial enhancement. The modified Pan is obtained using an intensity modulation based fusion method i.e, SFIM in [44]. Here, authors have used the local correlation coefficients between the MS and Pan image to adjust the amount of spatial and spectral details in the fused image. Finally, they obtain the pan-sharpened or fused image by optimizing the cost function by using gradient descent optimization method.

In [107] Zhang *et. al.* have addressed the fusion problem using adjustable model based approach. They model the LR MS image as decimated, blurred and noisy version of its true MS image. Along with this they also include the regularization term related to the modeling of the Pan image which accounts for the spatial detail preservation. A Huber edge preserving prior is used in MAP framework to regularize the final solution. The regularization parameters are also estimated that take care of the amount of spectral and spatial details in the pan-sharpened image. Although this method results in better fusion quality it uses a prior which requires fine tuning to get better fused image.

Aly and Sharma [108] have proposed a model based approach for pan-sharpening using the image formation models of LR MS and Pan images. Here, the Pan image formation term in the final cost function includes a high pass filter in order to preserve the high frequency details from the Pan image. In addition to this they also add the regularization prior term which takes care of adding the spatial details of Pan image to the fused image based on the correlation between Pan and fused images. Since the final cost function is convex they use the gradient based optimization method to obtain the final fused image.

There are several model based methods that exploit the variational framework in order to obtain the fusion of MS and Pan images. The approach of total variation (TV) was initially introduced for the regularization of inverse problems in [109] for solving denoising problem. Due to its ability to preserve the sharp discontinuities such as edges it has obtained greater success in the applications such as image restoration, denoising, inpainting, etc. The pan-sharpening method based on variational framework was first attempted by authors in [110] which has used the geometric information of the Pan image by aligning the same with each band of MS image. In order to get the spectral information for the fused image, it makes the assumption that the Pan image is an approximated linear combination of the high resolution MS images. The other fusion methods based on similar concept are reported in [111–113]. In [111], Moeller *et. al.* combine the idea of wavelet transform for a higher spectral quality and use the concept of variational approach in order to obtain the spatial information. Similarly in [112] authors utilize the assumption that a linear combination of the fused images gives the Pan image. In addition to this they also assume that the decimation of the pan-sharpened image gives the observed MS image. Using these assumptions they cast the fusion problem in regularization framework where TV prior on the fused image is used which encourages images that are piecewise smooth between edges. Similar assumption was used by He *et. al.* in [113]. They incorporate the gradient of Pan image in addition with that of the high resolution MS image in the form of TV prior to obtain the final fused image. An extension of this work is presented in [114] by including the total variation sparsity priors based on the characteristics of the MS and Pan images.

One of the advantages of the variational framework is that the different constraints can be included in the form of different terms in the objective function which is to be minimized. In [115], authors utilize few constraints such as the gradient of the Pan image

is a linear combination of the gradient of the true MS image, the upsampled MS image is the degraded version of the true MS image and the gradient of fused MS band is approximated by that of the uspsampled MS image to obtain the cost function. They proved the effectiveness of the proposed fusion method by conducting subjective as well as quantitative evaluation on two sets of images acquired using Quickbird and Ikonos satellites. The TV prior is also extended to the non local TV (NLTV) [116, 117] method. A pan-sharpening method based on the NLTV was proposed by Buades *et. al.* in [118]. Here, the final cost function consists of two terms: the first term forces the pan-sharpened image to be consistent with Pan and MS data, and the second corresponds to a non local regularization term which acts as a neighborhood filter on Pan image. The Pan image is used to derive nonlocal relationships among patches describing the geometry of the desired fused image.

Although TV and NLTV regularization may give better fusion results they suffer from certain implementation issues. A TV prior has L_1 norm in the regularization term. Although this results in edge preservation, the computational complexity is increased due to the non-differentiability of term with L_1 norm. The difficulty of non-differentiability can be avoided by the small perturbation in the prior [109]. However, it results in the modification of the original cost function and hence it causes the deviation in the required solution [119]. In addition to this, the final output also depends on the value of regularization parameter. Various methods are proposed in the literature to estimate this regularization parameter [120–124] which are either computationally expensive or yield approximate solution. It is interesting to note that though there are various approaches proposed in the literature to minimize the cost function using the TV prior, close form solution does not exist for the cost function with TV prior [125]. Unlike the TV prior which utilizes same weights to its neighbors, NLTV prior incorporates the nonlocal interaction among the neighboring pixels and computes the weights accordingly. In order to obtain the better solution using NLTV regularization, the accuracy of this weight function is very important. Researchers often use a reference image with features similar to the original image to estimate the weight function since original image is unavailable [126]. The NLTV prior also suffers with the drawbacks such as the selection of the size of patch for computing the weight function and window size for computing the image gradients which are set empirically [126]. Besides this the NLTV regularization algorithms designed for

a particular penalty are often not applicable to other functionals [127]. Thus, although TV and NLTV regularization may give better fusion results, they suffer from the high computational complexity [128] and implementations issues.

Recently, the use of compressive sensing (CS) theory has become very popular due to its ability to recover the unknown sparse signal from a small set of linear projections. The key point of CS theory is the sparsity regularization which refers to the characteristic of many natural signals which makes it suitable to solve the inverse problem. In [129] Li and Yang introduced the CS theory to obtain the fusion for remotely sensed images. In this method, a dictionary is constructed using sample images having high spatial resolution. They obtain fused image as the sparse linear combination of HR patches available in dictionary. Authors in [130, 131] construct an overcomplete joint dictionary from the available MS and Pan images and extract the most relevant spectral and spatial information using the ℓ_1 minimization. On a similar line Zhu and Bamler [132] use a dictionary constructed using the Pan image and its downsampled LR version and obtain the fusion by exploring the sparse representation of HR/LR multi-spectral image patches. In [133], authors have used the CS framework with dictionary patches learned from the initial high resolution MS image obtained using AWLP method [60]. They obtain the final fused image using a trained dictionary obtained from K-singular value decomposition (K-SVD). Authors in [134] create the over-complete dictionary with basis functions of different transforms such as discrete cosine transform (DCT), wavelets, curvelets and ridgelets and the best bases for MS and Pan images are obtained with convex optimization. The final fused MS image is obtained by merging best bases coefficients of Pan and MS bands as per values of the local information parameter. In addition to this Harikumar *et. al.* [135] use the CS theory in order to obtain initial HR approximation of the final fused image. The final solution is obtained by forming a cost function where truncated quadratic smoothness prior is used to regularize the solution. A graph cut optimization method is used to obtain the pan-sharpened image.

Regardless of the fusion classification, most of these fusion techniques required the set of MS images to be perfectly aligned with the Pan image i.e., the MS and Pan images have to be registered. The discussion on the different effects of image misregistration on different pan-sharpening methods is extensively reviewed in [136]. The image interpolation used in many fusion methods may cause the misregistration between the expanded

MS and Pan images. The distortion due to the image interpolation is studied by authors in [137] in case of multi-resolution image fusion. It has been found that the visual and quantitative performances of fusion methods largely depend on interpolation accuracy. Here, authors have proposed the new bicubic interpolation technique which can perfectly align the expanded MS and Pan images. We mention here that in this thesis we present few new approaches for fusion in which the registered MS and Pan data are used. However, we do not inject the high frequency details into the interpolated version of the MS image in the proposed model based approaches as done in most of the fusion approaches.

The multi-resolution image fusion is one of the most emerging research area among the fusion community and is now quite matured problem. Several survey papers have been published that consisting of various approaches for pan-sharpening or multi-resolution image fusion [2, 25, 28, 138–144]. Similarly many books are also published on the fusion techniques and its applications [145–148].

Chapter 3

Image Fusion using Different Edge Preserving Filters

In this chapter we discuss the approaches of multi-resolution image fusion using guided filter and difference of Gaussians (DoGs). In multi-resolution image fusion problem, the given MS and Pan images have high spectral and high spatial resolutions, respectively. One can obtain the fused image using these two images by injecting the missing high frequency details from the Pan image into the MS image. The quality of the final fused image will then depend on the method used for high frequency details extraction and also on the technique for injecting these details into the MS image. In the literature various approaches have been proposed based on this that also includes the state of the art methods such as additive additive wavelet luminance proportional (AWLP) [60] and generalized Laplacian pyramid-context based decision (GLP-CBD) [54]. Motivated by these works, we first address the fusion problem by using different edge preserving filters in order to extract the high frequency details from the Pan image. Specifically, we have chosen the guided filter and difference of Gaussians (DoGs) for detail extraction since these are more versatile in applications involving feature extraction, denoising, etc.

3.1 Related Work

A large number of techniques have been proposed for the fusion of Pan and MS images which are based extracting the high frequency details from Pan image and injecting them into MS image. They are discussed in detail in the chapter on literature survey.

These methods have been broadly cover the different categories as projection-substitution methods i.e., based on principal component analysis (PCA) and intensity hue saturation (IHS) [20, 33] and multi-resolution approaches based on obtaining a scale-by-scale description of the information content of both MS and Pan images [56, 149]. Among these methods, the multi-resolution based methods have proven to be successful [25]. Most of the multi-resolution techniques have been based on wavelet decomposition [56, 149] in which the MS and Pan images are decomposed into approximation and detail sub-bands, and the detail sub-band coefficients of the Pan image are injected into the corresponding sub-band of MS image by a predefined rule in which the MS image is first interpolated to make it to the size of Pan image. This is followed by inverse wavelet transform to obtain the fused image. This concept was taken forward in "a trous" Wavelet Transform (AWT) based fusion [56] in which the image is first convolved with cubic spline filter and decomposed into wavelet planes. The GLP-CBD fusion method [54] is an improved and successful version of AWT. It proposes to add the wavelet coefficients not directly but after weighting them with a constant which is computed based on local correlation of MS and Pan images. Another technique known as AWLP [60] which is based on AWT method that intends to preserve the spectral signature between the bands of the MS image by injecting high frequency values proportional to their original values. A new multi-resolution based technique, which takes into account the characteristics of multiple information sources simultaneously is introduced in [150]. This method is based on bilateral filter [151]. Though this approach takes into account both images while extracting detail bands, no clear relationship is established between the Pan and the MS images.

In this chapter we discuss two separate fusion approaches based on guided filter and difference of Gaussians, respectively. Since the Pan and MS images correspond to the same scene captured with the different sensors, there exists a definite relation between their detail bands. We make use of this relationship in our proposed approach using multistage guided filter similar to the methods on MRA framework. Assuming that this relationship is linear, we derive a multistage guided filter in which the Pan or MS image is used as guidance image while extracting the details from the other image. Since the details extraction process consist not only Pan image but also MS image the spatial distortion of MS image is reduced. The proposed method of fusion is accomplished by using a two stage guided filter. The extracted high frequency details are added to the

corresponding MS image to obtain the final fused image.

Similar to guided filter the other fusion approach is proposed based on the difference of Gaussians (DoGs) for extracting the edge details from the Pan image. The difference of Gaussians can be used as a band-pass filter to discard all but a handful of spatial frequencies that are present in the Pan image. In the first level, the Pan image is filtered using Gaussian kernel and the resulting blurred image is subtracted from the original image to get the first level of high frequency details. We then extract second level of high frequency details by applying the same procedure on the blurred Pan image. The high frequency details obtained in the first and second levels are merged into the upsampled MS image to get the final result. Advantage of the proposed approach using DoGs is the reduction in spatial degradation of the final fused image due to the use of non subsampled Pan image while fusing.

3.2 Fusion Approach using Multistage Guided Filter (MGF)

We propose a multi-resolution image fusion approach based on multistage guided filter (MGF) in order to extract the details from MS and Pan images. The detail extraction process exploits the relationship between the Pan and MS images by utilizing one of them as a guidance image while extracting details from the other. This way the spatial distortion of MS image is reduced by consistently combining the details obtained using both types of images. The final fused image is obtained by adding the extracted high frequency details to corresponding MS image. The results of the proposed algorithm are compared with the traditional and the state of the art methods using the images captured with different satellites such as Quickbird, Ikonos-2 and Worldview-2. The quantitative assessment is evaluated using the conventional measures as well as using a relatively new index i.e., quality with no reference (QNR) which does not require a reference image. The results and measures clearly show that there is significant improvement in the quality of the fused image using the proposed approach. The multi-resolution image fusion problem using MGF can be posed as follows. The given high spectral resolution MS image and high spatial resolution Pan image we obtain the single fused image with high spatial and

spectral resolutions using multistage guided filter.

3.2.1 Multistage Guided Filter (MGF)

He *et. al.* [152] have introduced the guided filter which is an edge preserving smoothing filter. They have experimented successfully to use the guided filter for a variety of applications including detail enhancement, compression, flash/no-flash de-noising etc. Later on Li *et. al.* [153] extended the guided filter for fusion of multifocus, multitemporal and multiexposure images. Motivated from this here we use the guided filter explicitly for fusion of remotely sensed images. In order to extract more meaningful details from these images we extend the guided filter to multistage form. A brief discussion of guided filter is given below:

A guided filter is a filter which uses guidance image as one of the inputs to the filter. Given the guidance image $\mathcal{I}$, the guided filter output Λ is assumed to be a linear transformation of this image. This transformation is given for a local window ν_k centered at a pixel k as [152],

$$\Lambda_i = \mathtt{a}_k \mathcal{I}_i + \mathtt{b}_k, \forall i \in \nu_k, \tag{3.1}$$

where Λ_i and $\mathcal{I}_i$ represent the i^{th} pixel intensity of filter output and guidance images, respectively. The size of window ν_k is $r \times r$ and r is an integer. The coefficients $\mathtt{a}_k$ and $\mathtt{b}_k$ are constants in window ν_k. These constants are estimated by minimizing the squared difference between output image Λ and input image $\wp$ as,

$$E(\mathtt{a}_k, \mathtt{b}_k) = \sum_{i \in \nu_k} ((\mathtt{a}_k \mathcal{I}_i + \mathtt{b}_k - \wp_i)^2 + \varepsilon \mathtt{a}_k^2), \tag{3.2}$$

where ε is a regularization parameter set by the user. The constants $\mathtt{a}_k$ and $\mathtt{b}_k$ are given by [152] as,

$$\mathtt{a}_k = \frac{\frac{1}{|\nu|}\Sigma_{i \in \nu_k} \mathcal{I}_i \wp_i - \mu_k \bar{\wp}_k}{\sigma_k^2 + \varepsilon}, \quad and \quad \mathtt{b}_k = \bar{\wp}_k - \mathtt{a}_k \mu_k, \tag{3.3}$$

where σ_k^2 and μ_k are the variance and mean of $\mathcal{I}$ in ν_k, $|\nu|$ is the number of pixels in ν_k, and $\bar{\wp}_k = \frac{1}{|\nu|}\Sigma_{i \in \wp_i} \wp_i$ is the mean of $\wp$ in ν_k. In equation (3.1), the values of $\mathtt{a}_k$ and $\mathtt{b}_k$ are same for all the pixels in the window ν_k and they are computed for all the overlapping windows in an image. Due to this the filter output for i^{th} pixel (Λ_i) has different values for all the overlapping windows over a pixel i. The obtained values of $\mathtt{a}_k$ and $\mathtt{b}_k$ for the

overlapping windows are averaged first and then used in equation (3.1) to compute the filter output. Then the final guided filter as follows

$$\Lambda_i = \bar{\mathtt{a}}_i \mathcal{I}_i + \bar{\mathtt{b}}_i, \tag{3.4}$$

where $\bar{\mathtt{a}}_i = \frac{1}{|\nu|} \sum_{k \in \nu_i} \mathtt{a}_k$ and $\bar{\mathtt{b}}_i = \frac{1}{|\nu|} \sum_{k \in \nu_i} \mathtt{b}_k$.

For the purpose of fusing Pan and MS images, guided filter is extended to a multistage form. In order to get adequate amount of extracted details, either of the Pan or MS image is used as a guidance image for the other as explained below. Since the Pan and MS images have different sizes, resampling is done on the MS image to obtain the same size for both. Also, it is worth to mention here that since the spatial resolution of the MS images are poor the meaningful high frequency details from them are extracted from the intensity component (INT) which is obtained as the weighted average of all the resampled MS bands. Due to the average operation, the computed intensity component has the spectral response similar to the spectral response of the Pan image which helps in extracting the possible high frequency details from the MS image when it is used as the guidance image. To describe the multistage form of guided filter here we use $\digamma(\wp, \mathcal{I})$ notation as the guided filter function with $\wp$ and $\mathcal{I}$ as input and guidance images, respectively. At the first stage ($j = 1$) to extract the details from the Pan image, the intensity component (INT) is used as guidance image ($\mathcal{I}$) and Pan image is used as an input image ($\wp$). This can be represented as,

$$Pan_A^j = \digamma(Pan, INT), \tag{3.5}$$

where Pan_A^j represents the guided filtered output or approximation layer with Pan image as an input image. Similarly, the Pan image is used as a guidance image ($\mathcal{I}$) for INT component image as an input image ($\wp$) which can be given as,

$$MS_A^j = \digamma(INT, Pan), \tag{3.6}$$

where MS_A^j denotes the approximation layer with intensity component as an input image.

The details layers at this stage can written as,

$$Pan_D^j = Pan - Pan_A^j \text{ and} \tag{3.7}$$

$$MS_D^j = INT - MS_A^j, \tag{3.8}$$

For the j^{th} stage ($j > 1$), the guided filter equations can be depicted as follows:

$$Pan_A^j = F(Pan_A^{j-1}, MS_D^{j-1}) \text{ and} \tag{3.9}$$

$$MS_A^j = F(MS_A^{j-1}, Pan_A^{j-1}), \tag{3.10}$$

where Pan_A^{j-1} and MS_A^{j-1} are the approximation layers (guided filtered outputs) of Pan and INT component as input images, respectively for $(j-1)^{th}$ level. The detail layers for the j^{th} decomposition stage (Pan_D^j and MS_D^j) for the Pan and the MS images can be expressed as,

$$Pan_D^j = Pan_A^{j-1} - Pan_A^j \text{ and} \tag{3.11}$$

$$MS_D^j = MS_A^{j-1} - MS_A^j, \tag{3.12}$$

3.2.2 The Proposed Approach using Guided Filter

The proposed fusion technique is outlined in Fig. 3.1. An approach of two stage guided filter is used to combine the details of the Pan and MS images in a consistent manner. Note that unlike the conventional methods where the information is extracted from the Pan and injected into the MS image, the details in the proposed method are extracted from both MS and Pan images in which each one is functioning as a guidance image. The detail is then adjusted using a gaining constant to achieve a fused image with higher spatial as well as higher spectral content. Considering an MS image with l different bands and the Pan image (Pan), the proposed fusion technique obtains the fused image (Z) for each band. The steps of the proposed fusion method can be detailed as follows:

1. The low resolution (LR) MS image is resampled to the size of Pan image by using a suitable interpolation technique (bicubic interpolation is normally preferred).

2. Considering number of MS bands as 4, an intensity image INT is formed using

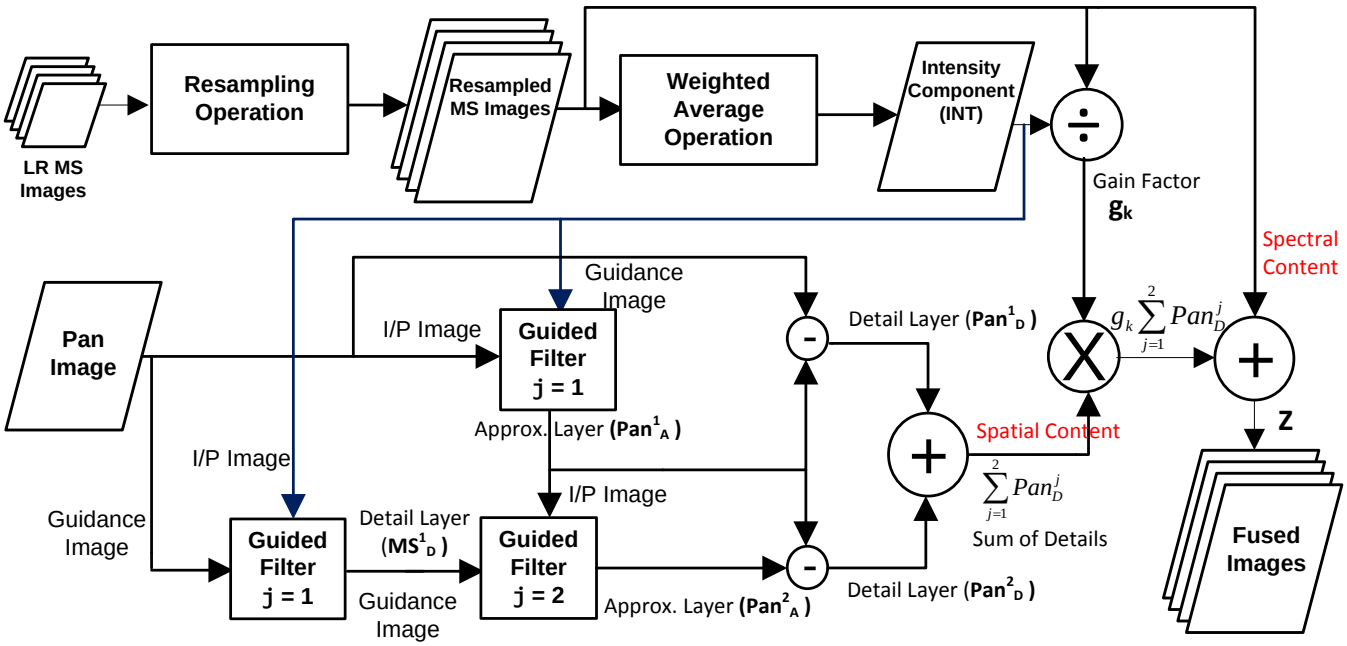


Figure 3.1: Block diagram of the proposed approach

them in order to make same spectral widths for both MS and Pan images. This is obtained using a weighted average of the four MS bands [33] i.e.,

$$INT = \frac{1}{l}\sum_{k=1}^{l} c_k \cdot MS_k, \tag{3.13}$$

where MS_k, $k = 1, \ldots, l$ denotes the resampled band of the MS image. The $c_k, k = 1, \ldots, l$ are the different constant estimated from the spectral response of MS bands.

3. The intensity component of MS image (INT) and the Pan image are used for multistage guided filter decomposition given in equations (3.5) to (3.12) and the detail and approximation layers are obtained. Here, a two stage guided filter decomposition is used.

4. In order to combine the extracted details in a consistent manner, they are merged band wise using a weighted gain. The gain factor g_k, calculated separately for each pixel in a band, is given by

$$g_k(i) = \frac{MS_k(i)}{INT(i)} \quad k = 1, \ldots, l, \tag{3.14}$$

where i denotes the pixel location in an image.

5. Finally, the extracted details are merged band wise with the MS image after weighting with g_k as,

$$Z_k = MS_k + g_k \sum_{j=1}^{N} Pan_D^j, \quad k = 1, \ldots, l, \tag{3.15}$$

where N represents the number of stages for guided filter. Here, the first term (MS_k) i.e. the resampled MS image contains spectral information while spatial information is carried by the second term i.e. the weighted sum of extracted details.

3.3 Fusion Approach using Difference of Gaussians (DoGs)

Use of the multistage form of guided filter to extract the details from the Pan image provides improvement in terms of detail extraction. However, due to multistage extension and patchwise computations, it results in increased computation in obtaining the final fused image. In this section, we describe a fast method for multi-resolution image fusion based on the difference of Gaussians (DoGs). The Pan and MS images are used to obtain a fused image having both high spectral and spatial resolutions. The method is based on two stage form of DoG on the Pan image. First, the Pan image is convolved with Gaussian kernel to obtain a blurred version and the high frequency details are extracted as the first level DoGs by subtracting the blurred image from the original. In order to get the second level DoG, same steps are repeated on the blurred Pan image. The extracted details at both DoGs are added to MS image to obtain the final fused image. Experiments have been conducted with different values of standard deviations of Gaussian blur with images captured from different satellite sensors such as Ikonos-2, Quickbird and Worlview-2. A QNR index along with the other measures are evaluated to check the efficacy of the proposed algorithm. The subjective and the quantitative assessment show that the proposed technique performs better, fast and less complex when compared to recently proposed state of the art techniques.

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

Figure 3.2: Example of DoGs. (a) Input test image, (b) Gaussian blurred image with standard deviation of $\sigma_1 = 4$, (c) Gaussian blurred image with standard deviation of $\sigma_2 = 2$ and (d) DoGs of (b) and (c), shown as inverted pixel intensity values.

3.3.1 Difference of Gaussians (DoGs)

Difference of Gaussians (DoGs) involves the subtraction of Gaussian blurred image from the image which is usually less blurred. The Gaussian blurred image can be obtained by convolving the Gaussian kernel with the input image. The DoG image $\tilde{D}^j(x, y)$ at j^{th} level can be given as,

$$\tilde{D}^j(x, y) = Ł(x, y, \sigma_1^j) - Ł(x, y, \sigma_2^j), \tag{3.16}$$

where $Ł(x, y, \sigma)$ is a blurred image obtained by convolving the input image $I(x, y)$ with the Gaussian kernel $£(x, y, \sigma_i)$ as,

$$Ł(x, y, \sigma_i^j) = £(x, y, \sigma_i^j) * I(x, y), \tag{3.17}$$

where σ_i^j, $i = 1, 2$ are the different values of standard deviation for Gaussian kernel at j^{th} level. Gaussian blurring suppresses high-frequency spatial information of the input image. Subtracting one image from the other preserves spatial information that lies between the range of frequencies that are preserved in the two blurred images. Thus, the DoGs is a band-pass filter that discards all but a handful of spatial frequencies that are present in the original image. The spreading parameter (σ) of the Gaussian can be chosen to select the edge details within the bandpass region of the Pan image. Fig. 3.2 displays the example of DoGs. In Fig. 3.2(a) we display the input test image. Two image blurred with $\sigma_1 = 4$ and $\sigma_2 = 2$ are displayed in Fig. 3.2(b) and (c), respectively. The DoG image is displayed in Fig. 3.2(d).

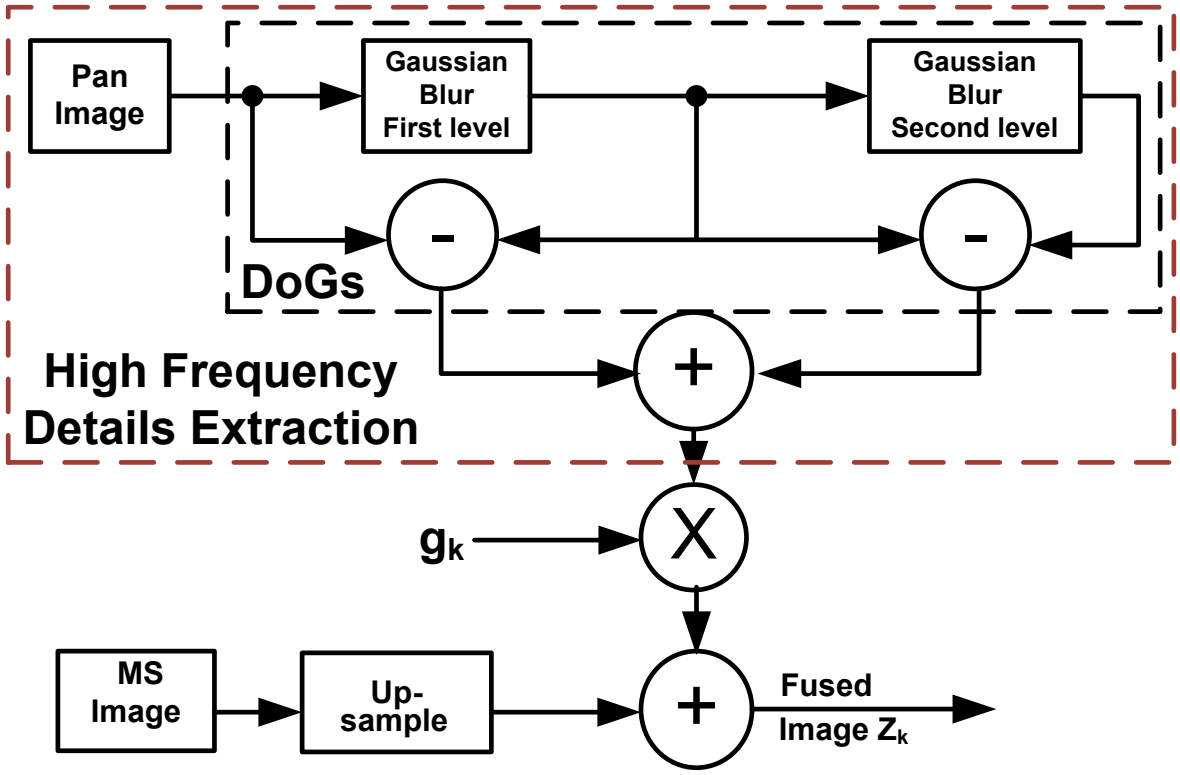


Figure 3.3: Block schematic of the proposed approach for k^{th} MS image. Here, g_k is a gain constant.

3.3.2 Proposed Approach using DoGs

Difference of Gaussians is widely used for feature extraction and edge detection. In fusion approaches the high frequency details present in the Pan image are extracted for injecting these details in the MS images. In the proposed method we use the DoGs for extracting these details at two different levels. Since we are interested in extracting the high frequency details from the Pan image, first level of DoG is performed by taking the difference between the original Pan image and its blurred version and the second level DoG is performed by taking the difference of second and first level. The extracted details of the first and second levels are merged with the MS image with appropriate weight g_k to get the final result. We conducted the experiments with more than two levels of DoGs but the improvement was not significant with the increased computation cost. Hence we used two stage approach. The block schematic of the proposed method is displayed in Fig. 3.3. The high frequency details extraction from the Pan image using DoGs is shown with dotted box in Fig. 3.3.

Let the MS images of the size $M \times M$. The high spatial resolution Pan image has

spatial resolution of $qM \times qM$, where q is the spatial resolution difference between the Pan and MS images. Our problem can now be posed as, given the LR MS image and a high spatial resolution Pan image (Pan), obtain the fused MS image (Z_k) with both high spatial and spectral resolutions. The proposed method is described in following steps:

1. The MS images are resampled to the resolution of Pan image using suitable interpolation technique. Here, we use bicubic interpolation due to its simplicity and fast output. This resampled MS band is denoted as MS_k, $k = 1, \ldots, l$. Here, l represents the number of bands.

2. The Pan image ($Pan(x, y)$) is convolved with Gaussian kernel $\pounds(x, y, \sigma_1)$. In order to obtain the blurred version $Ł_P(x, y, \sigma_1^1)$ which is given as,

$$Ł_P(x, y, \sigma_1^1) = \pounds(x, y, \sigma_1^1) * Pan(x, y). \tag{3.18}$$

3. The first level DoG is obtained by subtracting the blurred Pan image from the original Pan image without blur to obtain the high frequency details as

$$Pan_D^1 = Pan(x, y) - Ł_P(x, y, \sigma_1^1). \tag{3.19}$$

4. Gaussian kernel is again convolved with blurred Pan image $Ł_P(x, y, \sigma_1^1)$ that gives us the blurred Pan at second level i.e,

$$Ł_P(x, y, \sigma_2^2) = \pounds(x, y, \sigma_2^2) * Ł_P(x, y, \sigma_1^1). \tag{3.20}$$

5. The other possible details at the second level, Pan_D^2, are obtained by performing DoG between $Ł_P(x, y, \sigma_1^1)$ and $Ł_P(x, y, \sigma_2^2)$ as,

$$Pan_D^2 = Ł_P(x, y, \sigma_1^1) - Ł_P(x, y, \sigma_2^2). \tag{3.21}$$

6. The extracted details ($Pan_D^j, j = 1, 2$) are combined with MS image (MS_k) to get

Table 3.1: Details of datasets used in experimentation.

Satellite	Ikonos-2	Quickbird		Worldview-2
Downloaded from	[154]	[155]	[156]	[157]
Acquired Date	22^{nd} February, 2003	4^{th} July, 2005	28^{th} January, 2005	9^{th} October, 2011
Area covered	Mount Wellington area near Hobart Tasmania	Area around Boulder city USA	Sundarban India	San Francisco USA
Spatial Resolution: MS	$4m \times 4m$	$2.4m \times 2.4m$		$2m \times 2m$
Pan	$1m \times 1m$	$0.6m \times 0.6m$		$0.5m \times 0.5m$
Data format	11-bits	8-bits	11-bits	11-bits
No. of Pan and MS images	1 Pan, 4 bands	1 Pan, 4 bands	1 Pan, 4 bands	1 Pan, 8 bands
Original size: MS	3031×3287	828×815	16384×16384	16384×16384
Pan	12124×13148	3312×3260	4096×4096	4096×4096

the final fused image (Z_k) as,

$$Z_k = MS_k + g_k \sum_{j=1}^{2} Pan_D^j, \quad k = 0, 1, \ldots, l. \tag{3.22}$$

Here, g_k is a gain constant given in equation (3.14).

3.4 Experimental Results

The performance of the proposed fusion methods is evaluated by conducting the experiments on the images of different satellite sensors. We use the the images captured using Ikonos-2, Quickbird and Worldview-2 satellites. Various details of data captured using these satellites are listed in Table 3.1. In the experimentations, first MS and Pan images are co-registered. We prepare the datasets by cropping the original acquired MS and Pan images with the size of 256×256 and 1024×1024, respectively. Based on Wald's protocol [29], the MS and Pan images are spatially degraded and the experiments are carried out on them to quantitatively test the output of the fused results with the true MS images. In order to do so we downsample the MS and Pan images by a factor of 4. In all the experiments of degraded dataset, the size of MS and Pan images are 64×64 and 256×256, respectively. In addition to degraded dataset experiments, we also conducted the experiments on the un-degraded dataset. In the un-degraded dataset the size of MS and Pan images correspond to the size of original MS and Pan images. The results of the proposed fusion methods are compared with the other approaches on the basis of quality of images in terms of perceptual as well as quantitative measures. The following quanti-

tative measures are used for quantitative evaluation for those experiments conducted on degraded images.

- The correlation coefficient (CC) is calculated as

$$CC = \frac{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}[Z(i,j) - \bar{Z}][I_{ref}(i,j) - \bar{I}_{ref}]}{\sqrt{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}[Z(i,j) - \bar{Z}]^2}\sqrt{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}[I_{ref}(i,j) - \bar{I}_{ref}]^2}}, \tag{3.23}$$

  where I_{ref} and Z denote the reference and fused images with each of size $N_1 \times N_2$, respectively. The $\bar{I}_{ref}$ and $\bar{Z}$ indicate the mean of reference and fused images, respectively. The correlation coefficient indicates the degree of correlation between I_{ref} and Z. When Z and I_{ref} are same, the correlation coefficient approaches to one.

- The root mean squared error (RMSE) is the intensity difference between two images [60]. It is defined as

$$RMSE = \frac{1}{N_1 N_2}\sqrt{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}(I_{ref}(i,j) - Z(i,j))^2}. \tag{3.24}$$

  The ideal value of RMSE is 0 which indicates fused image is equal to the reference image.

- ERGAS is an acronym in French for "Erreur relative globale adimensionnelle de synthese" which translates to "relative dimensionless global error in synthesis". It gives a global spectral quality measure of the fused image (Z) [60], which is defined as

$$ERGAS = 100\frac{h}{\iota}\sqrt{\frac{1}{l}\sum_{i=1}^{l}(\frac{RMSE^2(i)}{MEAN^2(i)})} \tag{3.25}$$

  where h is spatial resolution of the Pan image; ι is spatial resolution of the MS images; l is the number of bands of the fused image Z ; $MEAN(i)$ is the mean value of the i^{th} band of the original MS image and $RMSE(i)$ is the RMSE between the i^{th} band of the original and fused MS images. Smaller the value of ERGAS indicates better fusion result. The ideal value of ERGAS is zero.

- The average quality (Q_{AVG}) index is defined as the average value of the universal image quality (UIQ) index of each band [60]. UIQ index models the distortion between two images as a combination of distortion due to loss of correlation, luminance, and contrast. This index is given as [60],

$$UIQ = \frac{4\sigma_{I_{ref}Z} \cdot \bar{I}_{ref} \cdot \bar{Z}}{(\sigma^2_{I_{ref}} + \sigma^2_Z)(\bar{I}^2_{ref} + \bar{Z}^2)}, \tag{3.26}$$

  where $\sigma_{I_{ref}Z}$ denotes the covariance between I_{ref} and Z, $\bar{I}_{ref}$ and $\bar{Z}$ are the mean of I_{ref} and Z, respectively. $\sigma^2_{I_{ref}}$ and σ^2_Z are the variances of I_{ref} and Z, respectively. Higher value of UIQ indicates better quality of the fused image. The ideal value of this measure is 1.

In addition to the above measures we also conducted the quantitative analysis of the fused image by computing a new measure called quality with no reference (QNR) which does not require the reference image.

- The QNR [158] is defined as,

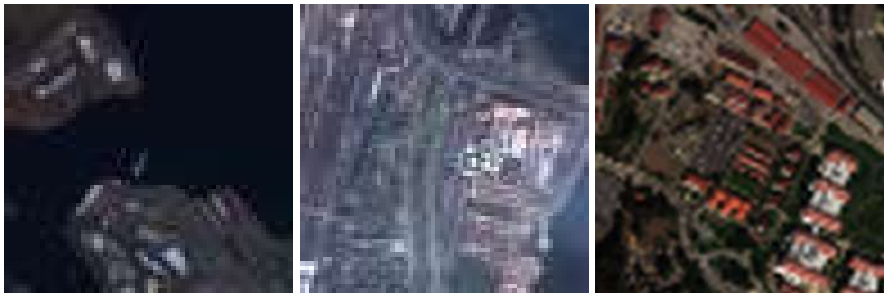
$$QNR = (1 - D_\lambda) \cdot (1 - D_s), \tag{3.27}$$

  where D_s and D_λ are the spatial and spectral distortions. The ideal value of QNR is 1. Spatial distortion index (D_s) is calculated as [158]:

$$D_s = \sqrt{\frac{1}{l}\sum_{i=1}^{l} |Q(MS_i, Pan_{LR}) - Q(Z_i, Pan)|}. \tag{3.28}$$

  Here, l is number of MS bands. MS_i and Z_i denote the i^{th} band of original and fused MS image, respectively. Pan is the original Pan image and Pan_{LR} represents the spatially degraded version of the Pan image obtained by passing it through lowpass filter having normalized cutoff frequency at the resolution ratio between MS and Pan, followed by decimation. The $Q(\cdot, \cdot)$ is quality index defined in equation (3.26). The index D_s attains its minimum velue of zero when quality measures are same for two images. Spectral distortion index (D_λ) is calculated as [158]:

$$D_\lambda = \sqrt{\frac{1}{l(l-1)}\sum_{i=1,}^{l}\sum_{r=1, r\neq i}^{l} |Q(MS_i, MS_r) - Q(Z_i, Z_r)|}. \tag{3.29}$$



(a) (b) (c)

Figure 3.4: Downsampled LR MS images. The LR test MS images of size 64×64 obtained using (a) Ikonos-2 satellite shown with the color composition of 4, 3 and 2 bands, (b) Quickbird satellite shown with the color composition of 3, 2 and 1 bands and (c) Worldview-2 satellite shown with the color composition of 5, 3 and 2 bands.

Here, D_λ is derived from the difference of inter-band quality index (Q) values calculated using the fused and the original MS bands. The ideal value of D_λ is 0.

The different parameters setting related to these proposed fusion approaches are as follows. In the proposed MGF based fusion method, the window of different size r is used for all three datasets while the regularization parameter ε is chosen as 10^{-6} for all the experiments. The values of r are $2, 7$ and 9 for the experiments on Ikonos-2, Quickbird and Worldview-2 satellite images, respectively. These parameters are selected by trial and error method. In the fusion approach based on DoGs, we use the standard deviation for Gaussian kernel as 2 and 1 at first and second stages, respectively i.e.$\sigma_1^1 = 2$ and $\sigma_2^2 = 1$. These values are tuned empirically. We extended the experimentation with more than two levels of DoG. However, the extracted high frequency details after second DoG level were less effective in improving the quality of final fused image. The effectiveness of the proposed fusion methods based on MGF and DoGs are demonstrated by comparing its fusion results with the results of other popular methods such as fast IHS (FIHS) [33], adaptive IHS [37] and AWLP [60].

In the following subsections we present the experimental results obtained using different satellite images. The subsections 3.4.1-3.4.3 discuss the fusion results obtained using the datasets of Ikonos-2, Quickbird and Worldview-2 satellites, respectively. The computational complexity of different fusion methods is the topic of discussion in subsection 3.4.4.

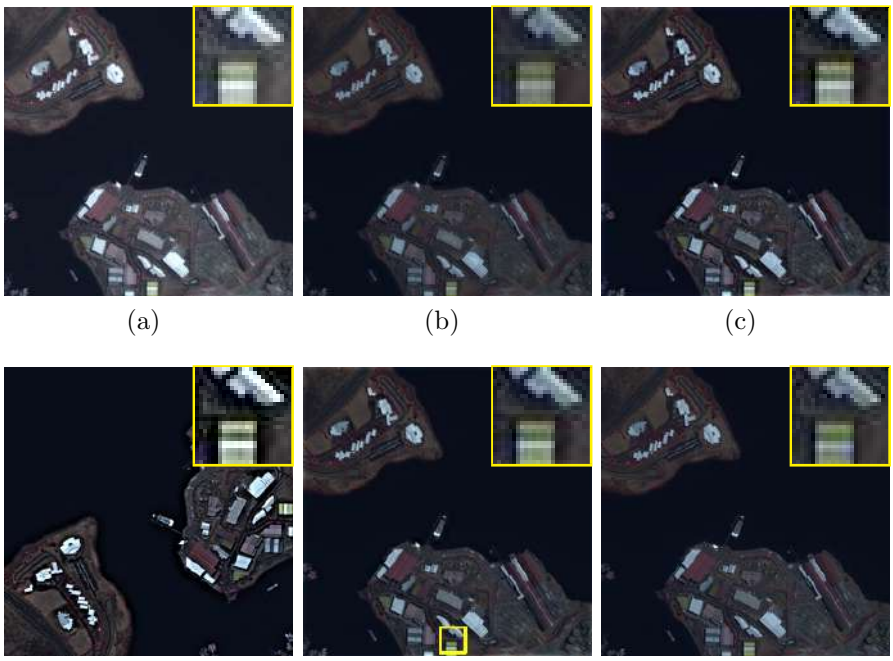


Figure 3.5: Results of multi-resolution image fusion on the degraded dataset of Ikonos-2 satellite shown with the color composition of 4, 3 and 2 bands. Fusion results of size 256 × 256 obtained using (a) FIHS method [33], (b) AIHS method [37] (c) AWLP approach [60], (d) proposed approach using MGF and (e) proposed method using DoGs. (f) Original MS image of size 256 × 256. The magnified region of a small squared area shown in (e) is displayed at top right corner.

3.4.1 Experimental results on the degraded and un-degraded datasets of Ikonos-2 satellite

We evaluate the performance of the proposed methods by conducting experiments on both degraded and un-degraded datasets. We first consider the images captured using Ikonos-2 satellite. The details of data captured using this satellite have been listed in Table 3.1 which correspond to the area of Mount Wellington, Hobart, Tasmania. The fusion results obtained using different methods are displayed in Fig. 3.5 and Fig. 3.6 for degraded and un-degraded datasets, respectively. Here, the results are shown with the color composition of 4, 3 and 2 bands. The downsampled test MS image for the degraded experiment is displayed in Fig. 3.4(a). In Fig. 3.5(a) and Fig. 3.6(a) we display the fusion results obtained using FIHS method [33] for degraded and un-degraded case, respectively. The results of AIHS method [37] are depicted in Fig. 3.5(b) and Fig. 3.6(b). In addition to this we have also compared our results with the fusion technique based on wavelet

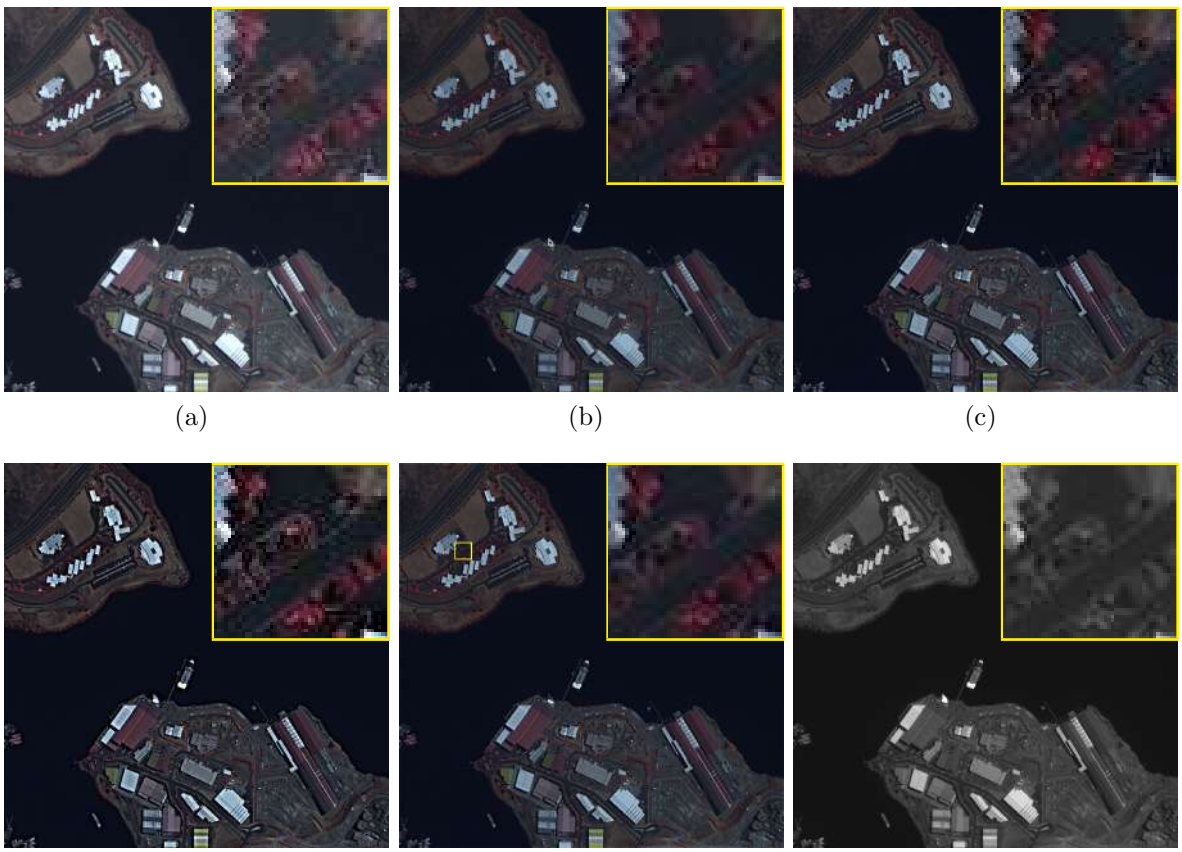


Figure 3.6: Results of multi-resolution image fusion on the un-degraded dataset of Ikonos-2 satellite shown with the color composition of 4, 3 and 2 bands. Fusion results of size 1024×1024 obtained using (a) FIHS method [33], (b) AIHS method [37] (c) AWLP approach [60], (d) proposed approach using MGF and (e) proposed method using DoGs. (f) Original Pan image of size 1024×1024. The magnified region of a small squared area shown in (e) is displayed at top right corner.

transform [60] and the results for the same are shown in Fig. 3.5(c) and Fig. 3.6(c), respectively. Finally, the results of the proposed fusion approaches using MGF and DoGs are displayed in Fig. 3.5(d, e) and Fig. 3.6(d, e) for degraded and original datasets, respectively. In the experiment of degraded dataset the original MS image is available and the same is displayed in Fig. 3.5(f). However, for the un-degraded case we display the original Pan image in Fig. 3.6(f) since we do not have the original MS image. In results displayed in Fig. 3.5 & 3.6, we show the magnified images corresponding to a small square portion shown with a yellow border in Fig. 3.5(e) and Fig. 3.6(e) and the same are displayed at top right corner in all the results for better comparison. The quantitative assessment for these experiments are depicted in Table 3.2.

Following observations are made from the fusion results displayed in Fig. 3.5 and

Fig. 3.6. In the results of the FIHS method [33] which are displayed in Fig. 3.5(a) and Fig. 3.6(a), one can see the preservation of color details along with the spatial contents. Since this approach is based on the spectral response of Ikonos-2 satellite, the results of this method show the improvement in the final fused image. By looking at the magnified images one can compare the results of FIHS [33] method with the results of AIHS [37] and AWLP [60] methods which are displayed in Fig. 3.5(a-c) and Fig. 3.6(a-c). Here, we may observe that the fusion results obtained using AIHS [37] and AWLP [60] methods lack in preserving the spectral and spatial contents when compared to the FIHS approach [33]. However, we can see the improvement in the results of the proposed MGF based approach which are displayed in Fig. 3.5(d) and Fig. 3.6(d) for degraded and un-degraded datasets, respectively over the FIHS method [33]. It is important to note that since the proposed MGF fusion method utilizes the high frequency details which are extracted not only from Pan image but also from MS image it shows better preservation of color information along with edge details. This observation is clearly visible in the magnified images of the proposed MGF fusion method (see Fig. 3.5(d) and Fig. 3.6(d)) which indicates that the results of proposed MGF fusion method have the spatial details such as boundary of houses and streets similar to the Pan image and spectral information close to original MS image. The results of the proposed fusion method based on DoGs are displayed in Fig. 3.5(e) and Fig. 3.6(e), respectively and its performance is also similar to these methods (AIHS and AWLP). Also, it is worth to mention that the performance of the proposed fusion method based on DoGs is close to that of the AWLP method [37] which is one of the state of the art fusion methods.

The quantitative measures for the experiments on degraded and un-degraded dataset are listed in Table 3.2. In this table, we also display the ideal values of these measures within brackets as a reference. Values shown as boldface indicate that they are closer to the ideal value. In the experiment of the degraded dataset, the quantitative performance of the proposed fusion method using MGF is better in terms of Q_{AVG} and RMSE. However, for the other measures such as CC and ERGAS the proposed DoGs fusion method performs better when compared to other fusion techniques. For the experiment of the un-degraded dataset, the quantitative evaluation in terms of QNR, D_s and D_λ are better for the proposed fusion approach based on MGF when compared to the other fusion methods. Looking at this quantitative assessment we can mention the

Table 3.2: Quantitative measures for the experiments of Ikonos-2 dataset shown in Fig. 3.5 and Fig. 3.6

Dataset	Measure	FIHS [33]	AIHS [37]	AWLP [60]	Proposed using MGF	Proposed using DoGs
Degraded (Fig. 3.5)	CC(1)	0.980	0.975	0.964	0.973	**0.988**
	ERGAS(0)	10.929	8.966	5.511	4.160	**4.210**
	$Q_{AVG}(1)$	0.860	0.879	0.896	**0.906**	0.944
	RMSE(0)	21.816	10.544	7.636	**5.645**	8.110
Un-degraded (Fig. 3.6)	QNR(1)	0.429	0.583	0.576	**0.640**	0.595
	$D_s(0)$	0.429	0.287	0.299	**0.228**	0.288
	$D_\lambda(0)$	0.326	0.181	0.177	**0.171**	0.164

following points. Although the visual assessment of FIHS method [33] for degraded and un-degraded datasets is better it lacks in showing better quantitative performance. Also, one may notice that though the qualitative performance of the proposed fusion method based on DoGs is close to AIHS [33] and AWLP [60] methods its quantitative measures have shown improvement in almost all measures except for RMSE. As we can see from the Table 3.2, the values of different measures for degraded and un-degraded datasets for the proposed methods are closer to the ideal values when compared to other fusion techniques.

3.4.2 Experimental results on the degraded and un-degraded datasets of Quickbird satellite

The second experiment is conducted on the degraded and un-degraded datasets of the Quickbird satellite which is downloaded from [156]. The details of the data captured using this satellite is presented in Table 3.1 and the captured area corresponds to an area of Sundarban, India. In Fig. 3.7 and Fig. 3.8 we display the fusion results with the color composition of 3, 2 and 1 bands. The test MS image for the experiment on the degraded dataset is displayed in Fig. 3.4(b). Similar to the earlier experiments on Ikonos-2 dataset here also we compare our fusion results with the recently proposed methods and the same are displayed in Fig. 3.7(a-c) and Fig. 3.8(a-c) for degraded and un-degraded datasets, respectively. The results of the proposed fusion methods based on MGF and DoGs are displayed in Fig. 3.7(d, e) and Fig. 3.8(d, e) for degraded and un-degraded datasets, respectively. For better visualization a small region shown in Fig. 3.7(e) and

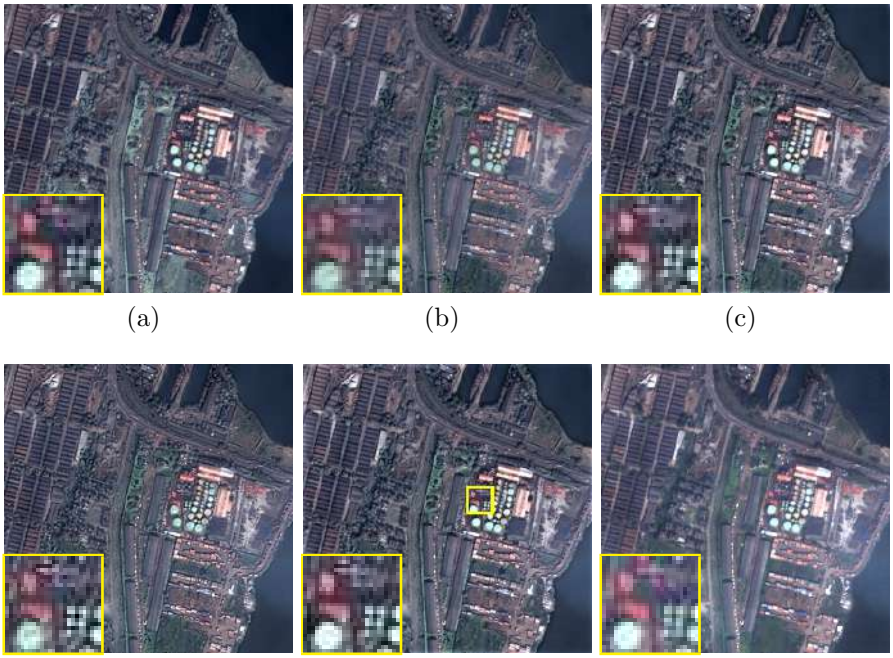


Figure 3.7: Results of multi-resolution image fusion on the degraded dataset of Quickbird satellite shown with the color composition of 3, 2 and 1 bands. Fusion results of size 256 × 256 obtained using (a) FIHS method [33], (b) AIHS method [37] (c) AWLP approach [60], (d) proposed approach using MGF and (e) proposed method using DoGs. (f) Original MS image of size 256 × 256. The magnified region of a small squared area shown in (e) is displayed at bottom left corner.

Fig. 3.8(e) with yellow border are magnified and shown in the top right corner of all the results. Looking at the results following observations can be made. The results of FIHS method [33] displayed in Fig. 3.7(a) and Fig. 3.8(a) show poor spectral and spatial contents due to the poor injection of details from the Pan to MS image in this method. Since the detail injection approach uses the the model based on spectral response of Ikonos-2 satellite it results in fused images with distortions when one tries to use it on the data captured using other satellites. These distortions can be seen in the fused images displayed in both Fig. 3.7(a) and Fig. 3.8(a). The other two fusion methods i.e., AIHS [37] and AWLP [60] perform better when compared to FIHS method [33]. Visual observation of the results obtained using the proposed method based on MGF which are displayed in Fig. 3.7(d) and Fig. 3.8(d) indicates that the edges and spectral changes are better for the proposed fusion method based on MGF when compared to the other methods. Similarly results obtained using proposed approach based on DoGs (Fig. 3.7(e) and Fig. 3.8(e)) have better preservation of homogenous as well as edge regions. This observation can be

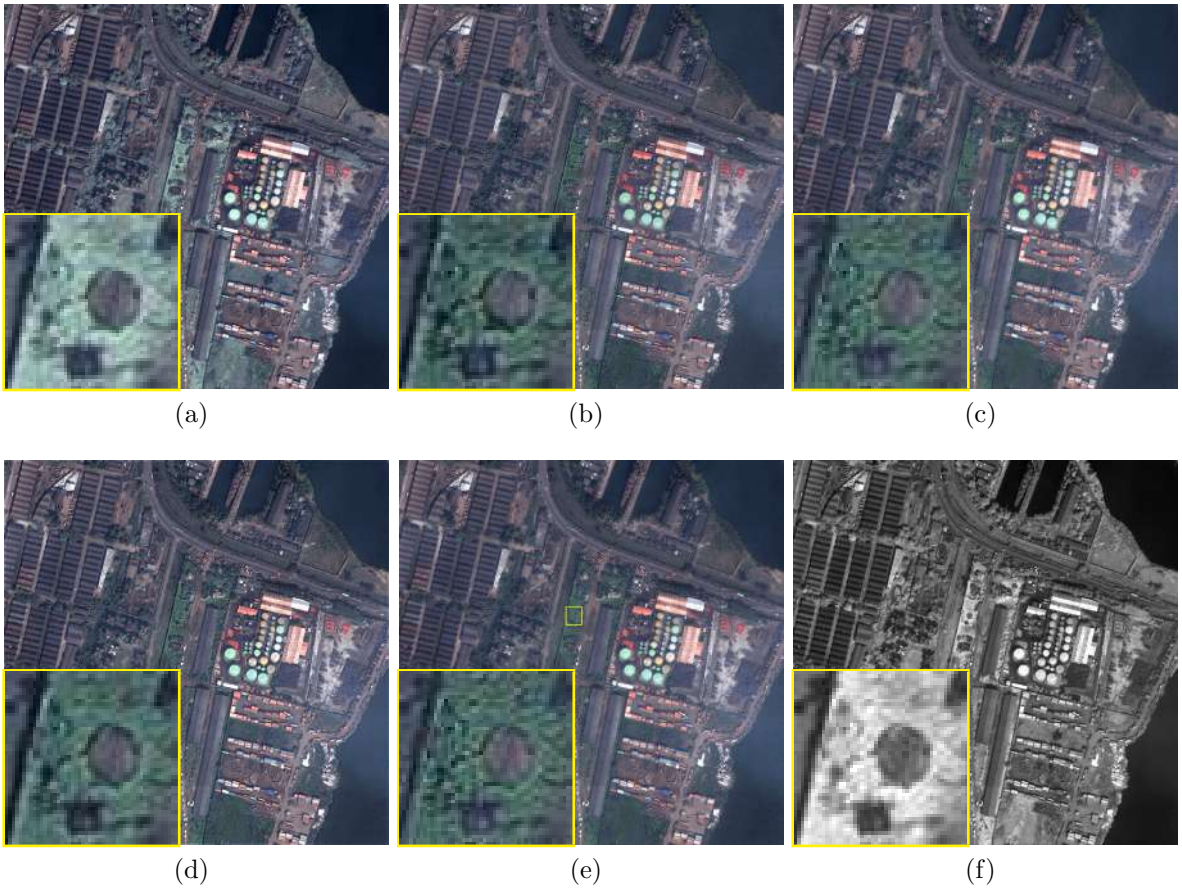


Figure 3.8: Results of multi-resolution image fusion on the un-degraded dataset of Quickbird satellite shown with the color composition of 3, 2 and 1 bands. Fusion results of size 1024×1024 obtained using (a) FIHS method [33], (b) AIHS method [37] (c) AWLP approach [60], (d) proposed approach using MGF and (e) proposed method using DoGs. (f) Original Pan image of size 1024×1024. The magnified region of a small squared area shown in (e) is displayed at bottom left corner.

easily verified by comparing the magnified images of these results displayed in Fig. 3.7(b, c, e) and Fig. 3.8(b, c, e) for degraded and un-degraded datasets, respectively. Also, it is of interest to note that the results of proposed fusion method based on MGF perform better when compared to that with the other proposed fusion approach based on DoGs. While comparing the magnified images of all the fusion results displayed in Fig. 3.7 and Fig. 3.8 one can say that the results of proposed fusion approach based MGF and DoGs perform better when compared to that with other methods.

The quantitative analysis for this experiment is depicted in Table 3.3 . The values of CC, ERGAS and RMSE measures are better for the proposed method based on MGF. However, Q_{AVG} is highest for the proposed fusion method using DoGs. Similarly, the quantitative evaluation of the results obtained using un-degraded dataset for the proposed

Table 3.3: Quantitative measures for the experiments of Quickbird dataset shown in Fig. 3.7 and Fig. 3.8

Dataset	Measure	FIHS [33]	AIHS [37]	AWLP [60]	Proposed using MGF	Proposed using DoGs
Degraded (Fig. 3.7)	CC(1)	0.867	0.917	0.920	**0.942**	0.929
	ERGAS(0)	6.228	3.996	4.476	**3.874**	4.661
	$Q_{AVG}(1)$	0.892	0.802	0.832	0.845	**0.875**
	RMSE(0)	20.645	13.413	16.037	**12.609**	13.959
Un-degraded (Fig. 3.8)	QNR(1)	0.281	0.408	0.410	0.559	**0.577**
	$D_s(0)$	0.605	0.453	0.447	**0.325**	0.373
	$D_\lambda(0)$	0.288	0.254	0.258	0.171	**0.079**

method based on DoGs is better in terms of the QNR and spectral distortion index (D_λ). Also, the spatial distortion measure (D_s) is lowest for the proposed fusion technique based on MGF. From the visual inspection and quantitative analysis one can conclude that our method on guided filters and difference of Gaussians have noticeable improvement over different regions when we compare the same with the results of other popular fusion methods.

3.4.3 Experimental results on the degraded and un-degraded datasets of Worldview-2 satellite

Finally, the last experiment is conducted on the degraded and un-degraded datasets of data captured using Worldview-2 satellite. The original acquired MS and Pan images of this satellite correspond to the area of San Francisco, USA. The spatial resolutions of images provided by this satellite are highest among the other satellites used in the previous two experiments. The fusion results obtained using different techniques are displayed in Fig. 3.9 and Fig. 3.10, respectively with the color composition of 5, 3 and 2 bands. Similar to the previous two experiments here also we display the input MS test image in Fig. 3.4(c) for the experiment on the degraded dataset. In Fig. 3.9(a-c) and Fig. 3.10(a-c), we display the fusion results obtained using different methods such as FIHS [33], AIHS [37] and AWLP [60] for degraded as well as un-degraded case. Similar to the earlier experiments on Ikonos-2 and Quickbird satellites here also we display the zoomed-in images for a small square area shown in Fig. 3.9(e) and Fig. 3.10(e) with yellow border. By looking at the fusion results obtained using FIHS method [33] which are displayed

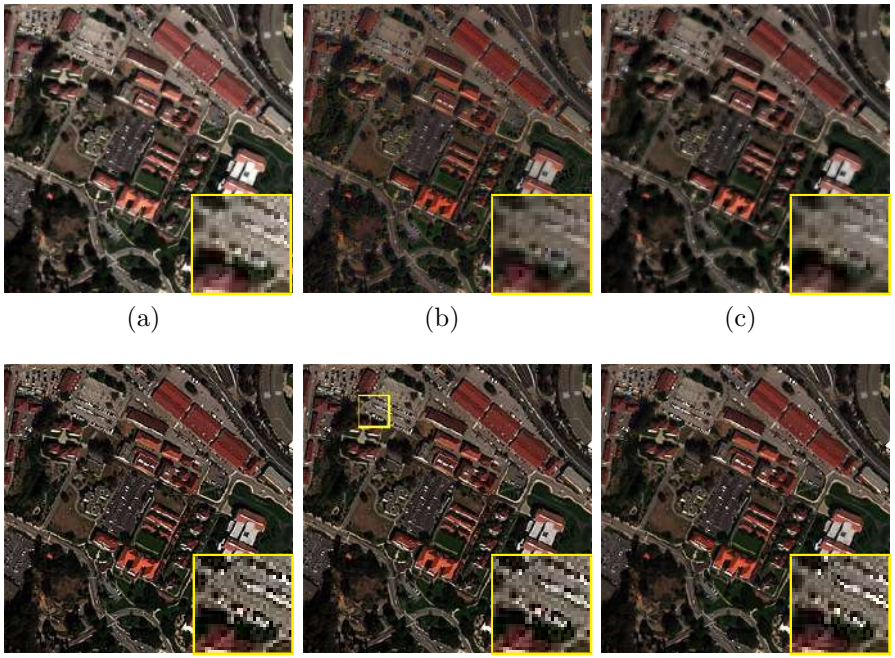


Figure 3.9: Results of multi-resolution image fusion on the degraded dataset of Worldveiw-2 satellite shown with the color composition of 5, 3 and 2 bands. Fusion results of size 256×256 obtained using (a) FIHS method [33], (b) AIHS method [37] (c) AWLP approach [60], (d) proposed approach using MGF and (e) proposed method using DoGs. (f) Original MS image of size 256×256. The magnified region of a small squared area shown in (e) is displayed at bottom right corner.

in Fig. 3.9(a) and Fig. 3.10(a) we can see that the performance of this method is better for this dataset when compared to the results of AIHS [37] and AWLP [60] which are displayed in Fig. 3.9(b, c) and Fig. 3.10(b, c), respectively. Results of these methods look blurred and also lack of preservation of color details. However, the results of FIHS fusion method (see Fig. 3.10(a) and Fig. 3.9(a)) appear closer to the original MS image which is displayed in Fig. 3.9(f). Similar observation also holds good for the results obtained using the un-degraded dataset. When we compare the results of our approaches based on MGF and DoGs displayed in Fig. 3.9(d, e) for degraded datasets with the results of all other fusion techniques displayed in Fig. Fig. 3.9(a-c) we can observe that their performance is improved in the preservation of spatial details with color information. Similarly, we notice the improvement in the results of our methods for the un-degraded case too. One can compare the magnified images of all these results of degraded and un-degraded cases and conclusion may be drawn that results of proposed fusion techniques using MGF and DoGs have better edge details with spectral information. For example, the zoomed-in

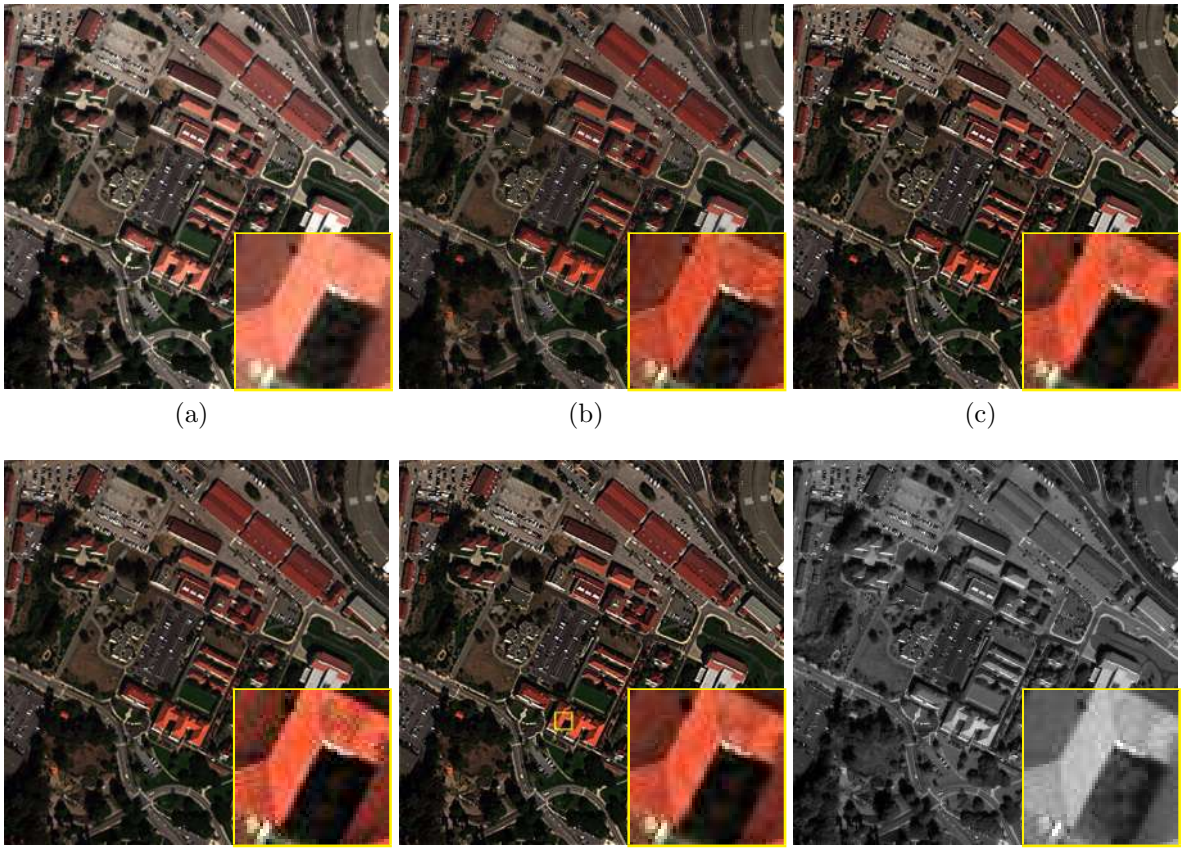


Figure 3.10: Results of multi-resolution image fusion on un-degraded dataset of Worldview-2 satellite shown with the color composition of 5, 3 and 2 bands. Fusion results of size 1024 × 1024 obtained using (a) FIHS method [33], (b) AIHS method [37] (c) AWLP approach [60], (d) proposed approach using MGF and (e) proposed method using DoGs. (f) Original Pan image of size 1024 × 1024. The magnified region of a small squared area shown in (e) is displayed at bottom right corner.

images displayed in Fig. 3.10(d, e) show that the fused images in the proposed methods appear sharper and are in consistent to Pan image which is displayed in Fig. 3.10(f).

The quantitative analysis for these experiments is given in Table 3.4. For the experiment on the degraded dataset, the measures such as CC and Q_{AVG} are closer to their ideal values for the proposed fusion method based on MGF. Similarly, the ERGAS and RMSE are better for the proposed method using DoGs. For the un-degraded case all the three measures i.e., QNR, D_s and D_λ are better for proposed fusion method based on DoGs.

It is of interest to note that the proposed fusion method using MGF extracts the details from both MS and Pan images and due to this it results in better fused image. However, in the fusion method using DoGs, it uses Pan image only to extract the meaningful

Table 3.4: Quantitative measures for the experiments of Worldview-2 dataset shown in Fig. 3.9 and Fig. 3.10

Dataset	Measure	FIHS [33]	AIHS [37]	AWLP [60]	Proposed using MGF	Proposed using DoGs
Degraded (Fig. 3.9)	CC(1)	0.908	0.919	0.916	**0.925**	0.911
	ERGAS(0)	9.031	8.360	7.612	7.334	**7.103**
	$Q_{AVG}(1)$	0.665	0.679	0.675	**0.695**	0.681
	RMSE(0)	30.448	22.320	20.915	19.807	**18.072**
Un-degraded (Fig. 3.10)	QNR(1)	0.4760	0.1587	0.6118	0.6150	**0.6658**
	$D_s(0)$	0.3600	0.5830	0.2390	0.1950	**0.1758**
	$D_\lambda(0)$	0.2550	0.6192	0.1960	0.2360	**0.1921**

Table 3.5: Computational complexity of different fusion methods.

Method	Average computation time (in seconds)	
	Degraded dataset	Un-degraded dataset
FIHS [33]	0.229	0.570
AIHS [37]	0.789	2.290
AWLP [60]	0.333	0.634
Proposed approach using MGF	1.236	4.896
Proposed approach using DoGs	0.208	0.480

details and performance of this method is slightly poor when compared to same with MGF fusion method. Although it is worth to mention that the performance of these two fusion methods are considerably improved when compared to the results of the state of the art methods.

3.4.4 Computational Complexity

The proposed algorithms as well as those used for comparison are executed on Matlab 7.6 installed on desk top computer with Intel dual core processor having 4 GB RAM. The proposed fusion methods are based on simply extracting the details from the Pan image using different filters, and hence are computationally efficient as compared to different state of the art techniques. However, since the fusion approach based on MGF is based on multistage it has slightly higher computational complexity when compared to the other approaches. Table 3.5 lists the average time taken to execute the main body of different methods. In this table one can clearly observe that the average run time for the proposed method based on DoGs is less when compared to other methods.

3.5 Conclusion

We have proposed two fusion methods based on guided filter and difference of Gaussians. The multistage form of guided filter is introduced to extract the details from Pan and MS images. Similarly, DoGs have been used to extract the high frequency details from the Pan image. The extracted details are transferred to MS image via a weighted average gain factor. The results have been compared with the state of the art methods by conducting the experiments on images of Quickbird, Ikonos-2 and Worldview-2 satellites. Comparison based on perceptual and quantitative measures clearly shows improvement in the quality of fused image obtained using the proposed fusion methods based on the MGF and DoGs. From the displayed results, quantitative measures and the computational complexity analysis one may say that the performance of the proposed fusion methods based on MGF and DoGs is better, fast and less complex when compared to the state of the art methods.

Chapter 4

Model based Approach for Multi-resolution Image Fusion

Recently, many researchers have attempted to solve the problem of multi-resolution image fusion by using the model based approaches with the emphasis on improving the fused image quality and reducing the color distortion [103, 107]. They model the low resolution (LR) MS image as the blurred and noisy version of it's ideal high resolution (HR) fused image. Solving the problem of fusion by model based approach is desirable since aliasing present due to undersampling of MS image can be taken care of while modeling. The fusion using the interpolation of MS images and edge preserving filters as given in Chapter 3 do not consider the effect of aliasing. In this chapter we propose a model based approach in which a learning based method is used to obtain the required degradation matrix that accounts for aliasing. Using the proposed model, the final solution is obtained by solving it as inverse problem. The proposed approach uses sub-sampled as well as non sub-sampled contourlet transform based learning and a Markov random field (MRF) prior for regularizing the solution.

4.1 Previous Work

Many researchers have used model based approach for fusion with the emphasis on improving fusion quality and reducing the color distortion [100, 101, 103–105, 107, 129, 130, 132]. In [104] the authors have proposed a spectrally consistent method for pixel-level fusion based on the model of the imaging sensor and using the pixel neighborhood regular-

ization. The fused image is obtained by optimizing an energy function consisting of a data term and a prior term. The image fusion based on restoration framework is suggested by authors in [105] which models the LR MS image as the blurred and noisy version of it's ideal. Also they model the Pan image as a linear combination of true MS images. The final fused image is obtained by using a constrained least squares (CLS) framework. The same model with maximum a *posteriori* (MAP) framework is used in [101, 107]. In [101], authors use the model based approach to enhance the hyperspectral images using the Pan image. Their framework takes care of enhancement of any number of spectral bands. In [107], authors use the same model as used in [101] for fusion and regularize the solution with Huber-Markov prior. They also estimate the regularization parameters adaptively with the use of spectral response of the sensor.

The authors in [103] use the image formation process using a linear modal and solve the problem of fusion by applying a discontinuity preserving prior constraint under the regularization framework. An Inhomogeneous Gaussian Markov Random Field (IGMRF) is used as a prior and its parameters are estimated using the Pan image that has high spatial resolution. However, since the learning of the spatial relationship is entirely based on the Pan data, it adds to spectral distortion in the fused image. The fusion performance of this method is also affected due to the approximate parameters estimation using maximum likelihood. It is also computationally taxing since IGMRF parameters are estimated at every location in the image. Our proposed approach in this chapter uses a transform domain method for using the high frequency details. Hence, we briefly explain few fusion techniques that are based on transform domain.

The favorable time-frequency localization to express the signal locally makes the wavelet a candidate for multi-sensor image fusion. However, wavelet bases are isotropic and represent limited directions. Therefore they fail to represent anisotropic edges and the contours in images in a better way. The MRA based fusion methods using the other transforms such as curvelet [82] and contourlet [86] have also been proposed. The contourlet transform (CT) has the characteristics of localization, multi-direction, and anisotropy. The CT gives the asymptotic optimal representation of contours and has been efficiently applied in image fusion techniques. However, the CT lacks shift-invariance which causes energy content at different levels of decomposition to differ. Hence it's performance while fusing MS and Pan images may be slightly inferior when compared to NSCT. In the non-

subsampled contourlet transform (NSCT) the input to filter banks is without subsampling and this makes it shift-invariance and hence it performs better. Fusion methods based on these transforms are reported in [86–88].

The authors in [86] obtain enhancement in spectral and spatial information by using principal component analysis (PCA) and non-subsampled CT. Here, all the detail coefficients at different scales of histogram matched Pan image are injected into MS image. A similar approach is presented in [87], where the first principal component i.e., PC^1 is also decomposed using NSCT and details are injected based on local variance of Pan details and PC^1. Authors in [88] obtain the fused image by using both spectral and spatial similarity. They decompose Pan and MS images using NSCT and the detail coefficients of Pan image are injected to MS image based on spatial similarity measure. Here, authors assume that the spectral frequency bands of MS images are same as the NSCT filter bank outputs. All these approaches do not consider aliasing and blurring present in the low resolution MS image and hence do not solve the problem of ill posedness inherently present in multi-resolution fusion. Although they use CT, no attempt is made to minimize the aliasing effect by exploiting the relationship between LR and HR MS images.

In this chapter, we propose a model based approach that uses subsampling as well as non subsampling CT for edge preservation. In our work a close approximation (initial estimate) to the final fused image is first obtained using the available Pan, MS image and the CT/NSCT which is then used in deriving the relation between LR and HR MS images and also the edges in the final fused image. We assume that the derived initial estimate has the desirable global characteristics of the fused image. The final fused MS image is obtained by using regularization for which we use a homogenous MRF prior that requires simple gradient based optimization in order to obtain the final solution. We take care of edge preservation in the final fused image by regularizing the non edge regions only. This is because our method uses CT/NSCT in order to obtain an initial estimates of fused image that has strong edges. Our approach in this paper differs from other CT based approaches and has the following advantages.

- As we are considering not only the Pan image but also the MS images while obtaining the initial estimate, it captures the smooth contours (i.e., spatial dependencies) as well as sharp edges effectively while still preserving the spectral content.

- Use of homogeneous MRF prior reduces the computational complexity. Though the prior is homogeneous MRF, the method still takes care of preserving edge pixels.
- Use of Canny edge detector which is an optimum edge extractor to locate the edge pixels in the initial estimate and retaining them as edges in the final fused image gives better edge details in the final solution.
- Optimization is carried out only on those pixels which do not belong to edges. This avoids oversmoothing of edge regions.
- Our method uses a degradation matrix estimated using the available MS image and the initial estimate, thus it avoids the need for assuming the degradation matrix in MS image formation process. The advantage here is that the degradation matrix is estimated from the given data itself and this leads to accurate estimation of the fused image.
- The proposed approach preserves the edge details without using the computationally taxing optimization methods [159, 160].

4.2 Description of the Proposed Approach using Block Schematic

The block schematic of our proposed multi-resolution fusion is shown in Fig. 4.1 in which an m^{th} low resolution MS image and the Pan image are fused giving Z_m as the fused image. The initial approximation of fused image (initial estimate) obtained using the CT/NSCT is used to recover the high frequency details for the fused MS image. Using this initial estimate and the given LR MS image, the degradation matrix entries are estimated by posing it as a least squares problem. This degradation matrix gives us the relationship between the LR and HR MS images. The discontinuities in the final fused image correspond to the edge pixels in the initial estimate. A Canny edge detector is used to extract these edge details from the initial estimate. The final solution is obtained by using the maximum a *posteriori*-Markov random field (MAP-MRF) formulation in which the MRF prior parameters are estimated using the initial estimate. With this MAP framework the final cost function consists of data fitting term and the MRF prior

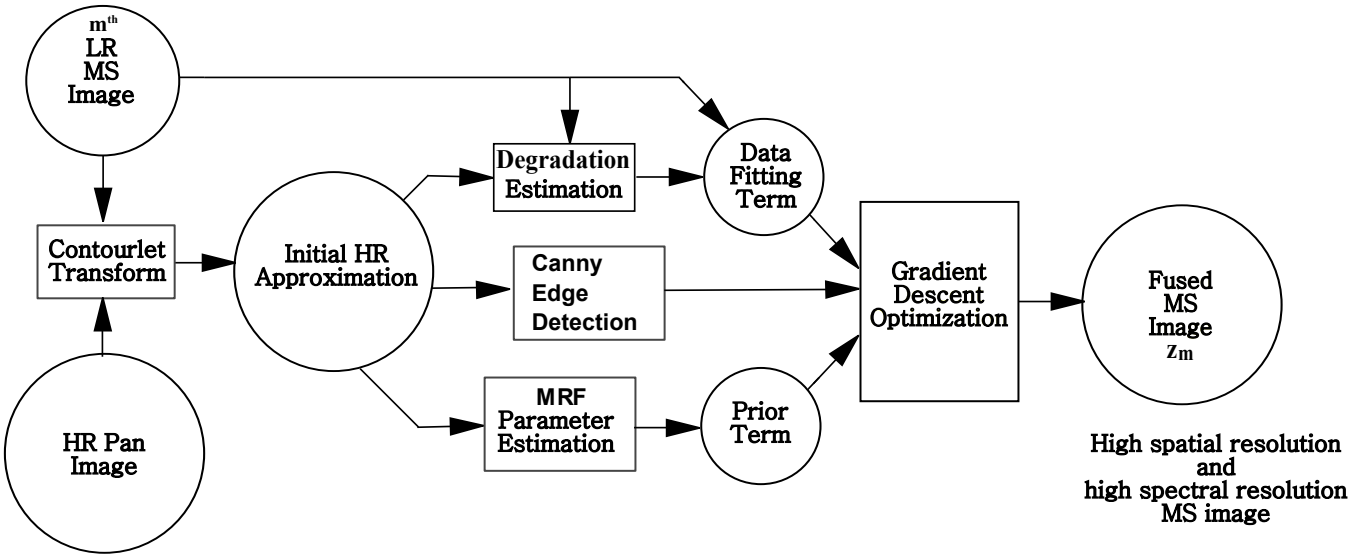


Figure 4.1: Block schematic of the multi-resolution fusion process for fusing an m^{th} MS and the Pan image. Here LR and HR correspond to low resolution and high resolution, respectively. The process is repeated for each of the MS image to obtain fused image separately for each of the LR observations.

term. Here the data fitting term forms the likelihood function. This cost function is optimized using gradient based optimization technique in order to smooth the non edge regions only. This takes care of preserving the homogeneity as well as edges in the final fused image without affecting the spectral content in the MS image.

4.3 Contourlet Transform based Edge Learning

The approaches based on multi-scale model (MSM) are popular among the fusion community. The first MSM which is still widely used by the researchers corresponds to the wavelet transform (WT) based model. It has the properties of multi-scale analysis and time-frequency localization. However, use of WT in fusion process preserves the edge details present in the horizontal, vertical and diagonal directions only. One can overcome this limitation using more recently proposed multi-resolution and directional transforms called *contourlet transform (CT)* and *non-subsampled CT(NSCT)* presented in [80, 81], respectively. The NSCT is the shift invariant version of CT that avoids using subsampling. These contourlet transforms not only have the advantages of multi-scale and time-frequency-localization properties of wavelets, but it also provides a high degree of directionality and anisotropy. The CT and NSCT employ directional filter bank (DFB)

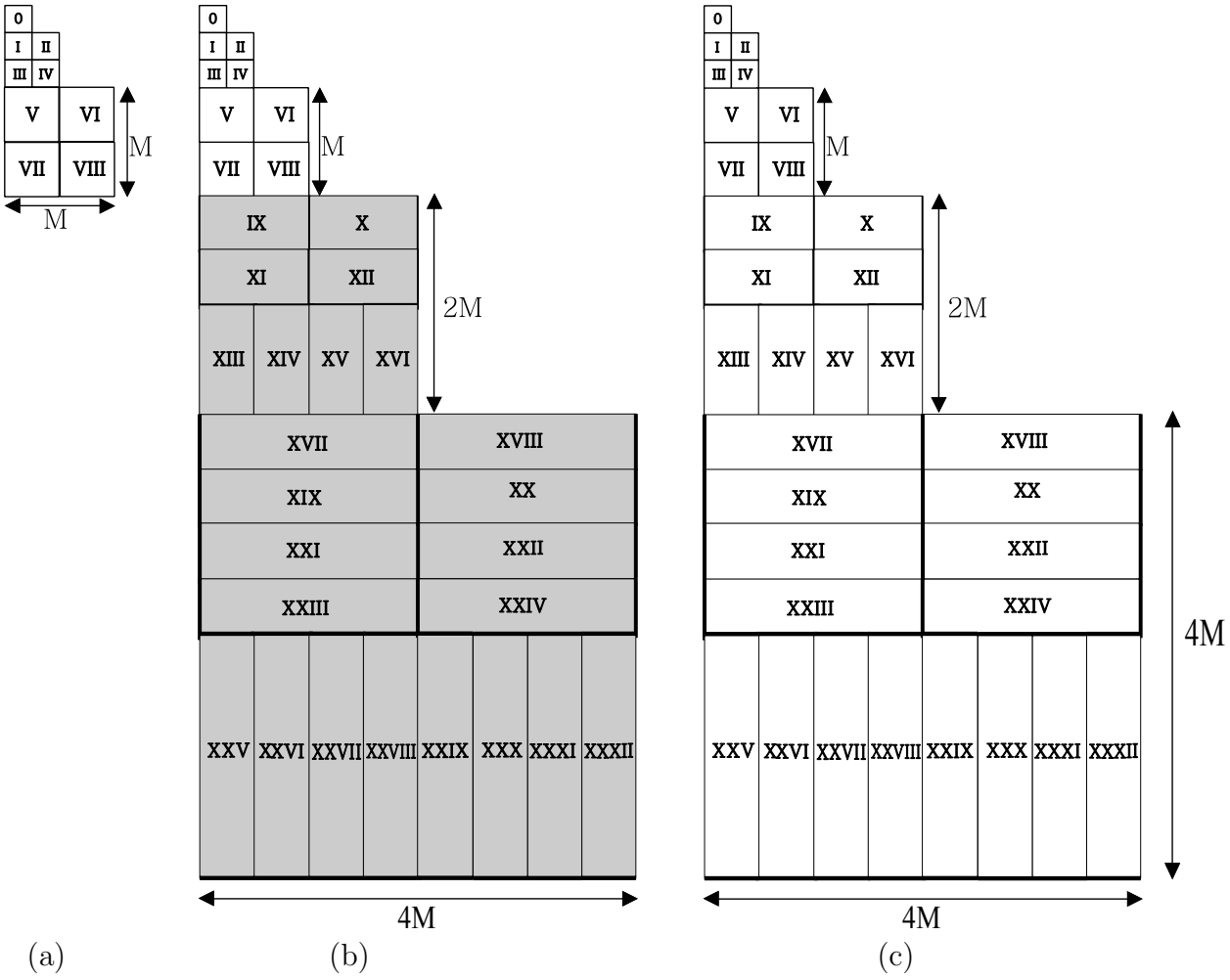


Figure 4.2: Learning the initial approximation to final fused image (initial estimate) using contourlet transform. (a) Two level contourlet decomposition of an MS image, (b) four level contourlet decomposition of initial estimate. Here shaded area sub-bands coefficients are to be learned from contourlet decomposition of the Pan image and (c) four level contourlet decomposition of the Pan image.

and non-subsampled DFB (NSDFB) that capture the edges in the different directions by linking the point discontinuities into linear structures. For details on contourlet transform the reader may refer to [63, 81, 85, 161].

We now explain the contourlet transform based method to obtain the initial estimate of the fused image for the given LR MS image. It may be mentioned here that though this procedure is explained using CT it can be easily extended to non-subsampled case by using the CT without using downsampling at every level. The contourlet decomposition is done on the LR MS and the Pan images. We take two level CT for the given low resolution MS image, which is shown in Fig. 4.2(a). Here, sub-band 0 corresponds to coarsest sub-band. The sub-bands $I - IV$ represent the directional sub-bands for first

level contourlet decomposition and the sub-bands $V-VIII$ correspond to those at second level. Considering a resolution difference of 4 between the Pan and MS, the Pan image is subjected to four levels of contourlet decomposition (see Fig. 4.2(c)). Considering both the computation complexity and the need for capturing the high frequency details, we use eight directional filters at the third level and sixteen at the fourth level while decomposing the Pan image. One may also compute the dominant directions of the initial estimate and use adaptive contourlet based approach in order to learn the edge details [162]. However, due to adaptive property computational complexity is increased much. We restrain from doing it in our approach as finding the dominant directions itself can be considered as a significant work and we wish to consider it in our future work. Similar to the MS image, the sub-band 0 represents the coarsest sub-band in Fig. 4.2(c), while the sub-bands $I-IV$ and $V-VIII$ correspond to first and second level decompositions, respectively. The sub-bands $IX-XVI$ and $XVII-XXXII$ correspond to third and fourth level decompositions of Pan image. In Fig. 4.2(b), we show the four level contourlet decomposition of the initial estimate to be obtained using Pan and MS image decompositions. In the LR MS image, lower frequencies are intact while high frequencies are missing. These missing frequencies can be obtained from the Pan data by using our learning approach. In order to do this, the low frequency details are duplicated in the initial estimate from the sub-bands $0-VIII$ of the MS image contourlet decomposition. The missing high frequency details i.e., CT coefficients of level 3 and 4 which is shown as the shaded area in Fig. 4.2(b), are obtained by copying the contourlet transform coefficients of the Pan image that correspond to the third and fourth levels (sub-bands $IX-XXXII$) of the initial estimate. If we denote the sub-band 0 of MS image contourlet transform decomposition as $\underset{MS}{\Psi}_0$, then one can describe the learning process as follows. The sub-bands $0-VIII$ of the initial estimate are copied from the sub-bands $0-VIII$ of the contourlet decomposition of the MS image. i.e.,

$$\underset{INT.}{\Psi}_i := \underset{MS}{\Psi}_i, \quad i = 0, I, II, \ldots, VIII. \tag{4.1}$$

In above equation (4.1), $\underset{INT.}{\Psi}_i$ denotes the i^{th} sub-band of initial estimate for the given MS band image. The sub-bands $IX-XXXII$ of Pan image contourlet decomposition

are copied into the corresponding sub-bands of the initial estimate. i.e.,

$$\underset{INT.}{\Psi}{}^{i} := \underset{PAN}{\Psi}{}^{i}, \quad i = IX, X, XI, \ldots, XXXII. \tag{4.2}$$

Here, $\underset{PAN}{\Psi}{}^{i}$ is the i^{th} sub-band of Pan image contourlet decomposition. Once the contourlet coefficients of the initial estimate are obtained, the inverse contourlet transform is taken to obtain initial estimate in spatial domain. This process is repeated for all the MS image bands. It is worth to mention here that it is not possible to obtain the true edge details of final fused image. Hence one has to look for those edge details that better approximate the edges in the final solution. Since the CT/NSCT gives better directional details it is reasonable to assume that the edges in the initial estimate correspond to true edges. A Canny edge detector is then used to obtain the edges from this initial estimate. Here, One may argue on the selection of the Canny edge filter for edge preservation. However, it is well known that the Canny edge detector represents an optimum edge detector and it performs better under noisy conditions. Also, it minimizes number of false edges due to number of reasons [163, 164]. In order to set a proper threshold value while extracting the edges from the initial estimate using the Canny edge detector, we first extract the edges of the given Pan image in which a proper threshold is selected to get the edges. The same threshold is then used on the initial estimate. Note that performing edge operation on the Pan is an offline operation.

4.4 Forward Model and Degradation Estimation

Since we cast our problem in a restoration framework, solving such a problem needs a forward model that represents the image formation process. Let l be the number of low-resolution MS images Y_m $(m = 1, 2, \ldots l)$, each captured with a different spectral band, of size $M \times M$ and Z_m be the corresponding fused HR MS image of size $qM \times qM$, where q is the decimation factor representing the spatial resolution difference between the LR and HR fused images. The forward model for the image formation can be written as,

$$\mathbf{y}_m = A_m \mathbf{z}_m + \mathbf{n}_m, \quad m = 1, 2, \ldots l. \tag{4.3}$$

In above equation (4.3), $\mathbf{y}_m$ and $\mathbf{z}_m$ represent the lexicographically ordered vectors of size $M^2 \times 1$ and $q^2M^2 \times 1$, respectively. A_m is the degradation matrix of size $M^2 \times q^2M^2$, which accounts for aliasing and blur. In equation (4.3), $\mathbf{n}_m$ is the independent and identically distributed (i.i.d.) noise vector with zero mean and variance $\sigma^2_{n_m}$ and has same size as $\mathbf{y}_m$. The multivariate noise probability density of $\mathbf{n}_m$ is given by

$$P(\mathbf{n}_m) = \frac{1}{(2\pi\sigma^2_{n_m})^{\frac{M^2}{2}}} e^{-\frac{1}{2\sigma^2_{n_m}}\mathbf{n}_m^T\mathbf{n}_m}. \tag{4.4}$$

Now the problem can be stated as follows. Estimate fused image $\mathbf{z}_m$ given $\mathbf{y}_m$. This is clearly an ill-posed inverse problem and needs additional constraint in the form of regularization prior to solve it. In order to obtain the solution by regularization one has to know A_m. However, this is unknown since we do not have the fused MS image. Since, we already know the approximation (initial estimate) to fused image, we use it for estimating A_m. One may write A_m as,

$$A_m = D_m H_m, \quad m = 1, 2, \ldots l. \tag{4.5}$$

where D_m is the decimation matrix of size $M^2 \times q^2M^2$ and H_m is the blur matrix of size $q^2M^2 \times q^2M^2$ which is assumed to be space invariant. While solving the fusion problem H_m is usually considered as an identity matrix [103]. However, in this work a non-identity blur matrix is assumed. When one assumes aliasing as the averaging of appropriate HR pixels then for a decimation factor of q the corresponding decimation/aliasing matrix can be written as [165]

$$D_m = \frac{1}{q^2}\begin{pmatrix} 1\,1\ldots 1 & & \mathbf{0} \\ & 1\,1\ldots 1 & \\ . & . & . \\ \mathbf{0} & & 1\,1\ldots 1 \end{pmatrix}. \tag{4.6}$$

In our work we consider decimation which has a different form for D_m as given in equation (4.7) below. By considering the LR pixel as linear combination of q^2 HR pixels with

appropriate weights, we choose D_m as,

$$D_m = \begin{pmatrix} a_1^m\ a_2^m \dots\ a_{q^2}^m & & \mathbf{0} \\ & a_1^m\ a_2^m \dots\ a_{q^2}^m & \\ . & . & . \\ \mathbf{0} & & a_1^m\ a_2^m \dots\ a_{q^2}^m \end{pmatrix}, \tag{4.7}$$

where $|a_i| \leq 1, i = 1, 2, \dots q^2$. Note that equation (4.7) replaces averaging effect (equal weights) as given in equation (4.6) by unequal weights. This matrix using unequal weights models the distortion caused due to aliasing that happens because of undersampling of MS image. It is of interest to compare the fusion results obtained using equation (4.7) and equation (4.6) and it is discussed in the experimental section for $q = 4$. By assuming a space invariant blur, the blur matrix H_m has the form

$$H_m = \begin{pmatrix} H_0^m & H_{qM^2-1}^m & H_{qM^2-2}^m \dots & H_1^m \\ H_1^m & H_0^m & H_{qM^2-1}^m \dots & H_2^m \\ . & . & . \dots & . \\ H_{qM^2-1}^m & H_{qM^2-2}^m & H_{qM^2-3}^m \dots & H_0^m \end{pmatrix}, \tag{4.8}$$

where m=1,2, …, l and each H_j can be written as,

$$H_j^m = \begin{pmatrix} h_{j,0}^m & h_{j,q-1}^m & h_{j,q-2}^m \dots & h_{j,1}^m \\ h_{j,1}^m & h_{j,0}^m & h_{j,q-1}^m \dots & h_{j,2}^m \\ . & . & . \dots & . \\ h_{j,q-1}^m & h_{j,q-2}^m & h_{j,q-3}^m \dots & h_{j,0}^m \end{pmatrix}. \tag{4.9}$$

Here $h_{.,.}$ are the values of the point spread function (PSF) for blur. Since we consider a space invariant blur, H_m is block circulant. The multiplication of D_m and H_m results in the A_m matrix which is given by

$$A_m = \begin{pmatrix} A_1^m & A_2^m & \dots & A_{q^2M^2-1}^m & A_{q^2M^2}^m \\ A_{q^2M^2-q+1}^m & \dots & A_1^m & A_2^m \dots & A_{q^2M^2-q}^m \\ . & . & . & . & . \\ \dots & \dots & \dots A_1^m & \dots & \dots \end{pmatrix}. \tag{4.10}$$

For $M = 2$ and $q = 2$ the A_m matrix has size of 4×16 and can be written as,

$$A_m = \begin{pmatrix} A_1^m & A_2^m & A_3^m & A_4^m & A_5^m & A_6^m & A_7^m & A_8^m & A_9^m & A_{10}^m & A_{11}^m & A_{12}^m & A_{13}^m & A_{14}^m & A_{15}^m & A_{16}^m \\ A_{15}^m & A_{16}^m & A_1^m & A_2^m & A_3^m & A_4^m & A_5^m & A_6^m & A_7^m & A_8^m & A_9^m & A_{10}^m & A_{11}^m & A_{12}^m & A_{13}^m & A_{14}^m \\ A_9^m & A_{10}^m & A_{11}^m & A_{12}^m & A_{13}^m & A_{14}^m & A_{15}^m & A_{16}^m & A_1^m & A_2^m & A_3^m & A_4^m & A_5^m & A_6^m & A_7^m & A_8^m \\ A_7^m & A_8^m & A_9^m & A_{10}^m & A_{11}^m & A_{12}^m & A_{13}^m & A_{14}^m & A_{15}^m & A_{16}^m & A_1^m & A_2^m & A_3^m & A_4^m & A_5^m & A_6^m \end{pmatrix}. \tag{4.11}$$

We estimate the entries in equation (4.10) using the available LR MS image and initial HR estimate image. A minimum norm approach has been used for the estimation of A_m. In literature various algorithms are available to estimate the same however we use the technique given in [166] which is computationally very efficient compared to other methods. Inclusion of non-identity blur matrix leads to under determinant set of equations. It is worth to mention here that we are not assuming the known entries for D_m and H_m matrices, instead estimate them from the given data i.e., using LR observation and the initial estimate (Z_m^0). Many researchers consider an identity matrix for blur which is not true in practice. Hence the estimated A_m matrix as done in our approach is closer to the true degradation.

4.5 MRF Prior Model

In order to obtain a regularized estimate of the high-resolution fused image, we define an appropriate prior term using MRF modeling. MRF provides a convenient and logical approach to model context dependent entities such as pixel intensities, depth of the object, and other spatially correlated features [167]. Recently, conditional random fields (CRFs) are being used as priors in solving many of the problems in the image processing applications. Though, it is a good idea to use CRF on fused MS image that depends on the LR resolution MS image, it is difficult to obtain true dependency between them. Also when we model an entity using CRF the parameter estimation requires the use of computationally taxing partition function [168]. Hence, in this work we prefer to use an MRF model which do not require the LR observations for parameter estimation. Also the computational complexity is reduced additionally by the use of homogeneous MRF where a single MRF parameter is estimated. An MRF prior for the unknown fused HR image can be described by using a energy function U expressed as Gibbsian density given

by

$$P(\mathbf{z}_m) = \frac{1}{Z_{m\theta}} e^{-U(\mathbf{z}_m)}, \tag{4.12}$$

where $\mathbf{z}_m$ is the fused HR image to be estimated and $Z_{m\theta}$ is partition function. One can choose U as a quadratic form with a single global parameter, assuming that the images are globally smooth. A method of specifying MRF prior involves considering the pair wise cliques c on a neighborhood and imposing a quadratic cost which is function of finite difference approximations of the first order derivative at each pixel location. This constitutes a homogeneous and non edge preserving smoothness prior. By using first order neighborhood, the energy function corresponding to the MRF prior can be written as,

$$U(\mathbf{z}_m) = \sum_{c \in \mathcal{C}} V_c(\mathbf{z}_m) \quad = \quad \gamma_m \sum_{k=1}^{qM} \sum_{l=1}^{qM} [(Z_{m(k,l)} - Z_{m(k,l-1)})^2 + (Z_{m(k,l)} - Z_{m(k-1,l)})^2], \tag{4.13}$$

where γ_m represents the MRF parameter that indicates the penalty for departure from smoothness in $\mathbf{z}_m$. $\mathcal{C}$ is the set of all cliques. The MRF parameter γ_m is known if the fused image is known. In our work, since the initial estimate is already available, we make use of the same to estimate γ_m. We use maximum pseudo likelihood for estimating it [167].

It is worth to mention few points on the various fusion techniques proposed using the discontinuity preserving priors. In order to obtain the better fusion, researchers have used different types of discontinuity preserving priors. In these techniques the resulting cost function may be convex or non-convex. When the cost function is non-convex the computationally taxing optimization techniques are required to obtain the final solution [159, 160, 169]. A Huber-MRF is an edge preserving prior and it is convex. So it could be a good choice for edge preservation which allows us to minimize the cost function by using gradient descent technique [107, 170]. However, the performance of this prior is highly dependent on selection of the threshold used in defining the prior. Also, the use of single threshold cannot take care of preserving the edges having different gradients present in the image. The estimation of this threshold is a very challenging task and it is image dependant. In our approach we avoid using such threshold based edge preservation.

Similar to the Huber prior one may argue to use total variation (TV) [109] or non local total variation (NLTV) [116, 117] as the discontinuity preserving prior for the edge preservation. The TV prior was first proposed by Rudin *et. al.* [109] that has L_1 norm in the regularization term. Although this results in edge preservation, the computational complexity is increased due to the non-differentiability of term with L_1 norm. The difficulty of non-differentiability can be avoided by the small perturbation in the prior [109]. However, it results in the modification of the original cost function and hence it causes the deviation in the required solution [119]. In addition to this, the final output also depends on the value of regularization parameter. Various methods are proposed in the literature to estimate this regularization parameter [120–124] which are either computationally expensive or yield approximate solution. It is interesting to note that though there are various approaches proposed in the literature to minimize the cost function using the TV prior, close form solution does not exist for the cost function with TV prior [125]. Unlike the TV prior which utilizes same weights to its neighbors, NLTV prior incorporates the nonlocal interaction among the neighboring pixels and computes the weights accordingly. In order to obtain the better solution using NLTV regularization, the accuracy of this weight function is very important. Researchers often use a reference image with features similar to the original image to estimate the weight function since original image is unavailable [126]. The NLTV prior also suffers with the drawbacks such as the selection of the size of patch for computing the weight function and window size for computing the image gradients which are set empirically [126]. Besides this the NLTV regularization algorithms designed for a particular penalty are often not applicable to other functionals [127]. Thus, although TV and NLTV regularization may give better fusion results, they suffer from the high computational complexity [128] and implementations issues. Due to these reasons we have resorted to a very simple approach of edge preserving fusion approach. The edges in the final fused image in the proposed method are obtained by extracting them from an initial estimate which itself is dependent on the given high resolution data i.e., Pan image. The extracted edges are preserved in the final fused image by retaining the edges corresponding to initial estimate and by applying MRF regularization on homogenous regions only. This gives the advantage in terms of using gradient based techniques for optimization and preserving the edges in the final solution.

4.6 MAP Estimation and Optimization Process

4.6.1 MAP Estimation

The MRF model on the fused image serves as the prior for the MAP estimation in which the prior parameter is already known. The data fitting term contains the degradation matrix estimated using the initial estimate. In order to use maximum *a posteriori* estimation to HR fused image, we need to obtain the estimate as

$$\hat{\mathbf{z}}_m = \underset{z_m}{^{argmax}} P(\mathbf{z}_m/\mathbf{y}_m). \tag{4.14}$$

Using the Bayes' rule we can write,

$$\hat{\mathbf{z}}_m = \underset{z_m}{^{argmax}} \frac{P(\mathbf{y}_m/\mathbf{z}_m)P(\mathbf{z}_m)}{P(\mathbf{y}_m)}. \tag{4.15}$$

While maximizing over all $\mathbf{z}_m$, denominator in equation (4.15) becomes a constant. Hence we can write

$$\hat{\mathbf{z}}_m = \underset{z_m}{^{argmax}} P(\mathbf{y}_m/\mathbf{z}_m)P(\mathbf{z}_m). \tag{4.16}$$

Now taking the *log* we get,

$$\hat{\mathbf{z}}_m = \underset{z_m}{^{argmax}} [logP(\mathbf{y}_m/\mathbf{z}_m) + logP(\mathbf{z}_m)]. \tag{4.17}$$

Finally, using equation (4.3) and equation (4.13), we can write

$$\hat{\mathbf{z}}_m = \underset{z_m}{^{argmin}} [E(\mathbf{z}_m)], \tag{4.18}$$

where $E(\mathbf{z}_m)$ is the final cost function to be minimized and can be expressed as

$$E(\mathbf{z}_m) = \frac{\| \mathbf{y}_m - A_m\mathbf{z}_m \|^2}{2\sigma^2_{n_m}} + \sum_{c\in\mathcal{C}} V_c(\mathbf{z}_m). \tag{4.19}$$

In equation (4.19), the first term ensures the fidelity of the final solution to the observed data through the image formation process and it is called the data term. The second term is the smoothness prior.

4.6.2 Optimization Process

Since the cost function given in equation (4.19) is convex, it can be minimized using a simple gradient based technique, which quickly leads to the minima. The gradient of this cost is given by differentiating it with respect to $\mathbf{z}_m$ as

$$\nabla E(\mathbf{z}_m) = -2A_m^T(\mathbf{y}_m - A_m\mathbf{z}_m)/2\sigma_{n_m}^2 + 2\gamma_m\hat{\mathbf{g}}_m, \tag{4.20}$$

where the first term corresponds to differentiation of the data term, while $\hat{\mathbf{g}}_m$ vector constructed after differentiating $\sum_{c\in\mathcal{C}} V_c(\mathbf{z}_m)$ at every location. The differentiation with respect to $Z_m(i,j)$ of the second term at a (i,j) location is given by

$$\hat{G}_m(i,j) = 2[4Z_{m(i,j)} - Z_{m(i,j-1)} - Z_{m(i,j+1)} - Z_{m(i-1,j)} - Z_{m(i+1,j)}]. \tag{4.21}$$

Note that $\nabla E(\mathbf{z}_m)$ is a vector of size $q^2M^2 \times 1$. The complete gradient descent optimization process is depicted in algorithm 4.1. The value of α_m decides the rate of convergence. Note that a smaller value of the step size α_m will lead to slower convergence and for the larger value the algorithm may not converge. This makes us to use a varying step size α_m^n and the same is estimated using the method given in [170] as

$$\alpha_m^n = \frac{[\nabla E(\mathbf{z}_m^n)]^T\nabla E(\mathbf{z}_m^n)}{[\nabla E(\mathbf{z}_m^n)]^T(A_m^TA_m/2\sigma_{n_m}^2 + \gamma_m\nabla^2\hat{\mathbf{g}}_m)\nabla E(\mathbf{z}_m^n)}. \tag{4.22}$$

Since the optimization process is iterative the choice of initial guess fed to the optimization process determines the speed of convergence. Use of the available initial estimate as an initial solution speed-up the convergence. Note that we obtain initial estimate separately for each of the MS observations and the optimization is carried out independently for every LR MS observation. We mention here again that in order to preserve the edges in the final solution we detect the edges from the initial estimate with the help of Canny edge detector and we do not perform the optimization on those edge pixels. In other words the edges in the final fused image correspond to the already learned edges in the initial estimate.

Algorithm 4.1: Gradient descent technique

Data: LR MS image ($\mathbf{y}_m$), initial estimate obtained using NSCT/CT ($\mathbf{z}_m^0$), A_m and γ_m
Result: Fused MS image ($\hat{\mathbf{z}}_m$)
foreach *LR MS image* $\boldsymbol{y}_m, m = 1, 2, \ldots, l$ **do**
  set $n = 0$;
  repeat
    find gradient with respect to $\mathbf{z}_m^n$ of cost function given in equation (4.19) using equation (4.20);
    estimate the step size α_m^n using equation (4.22);
    perform $\mathbf{z}_m^{n+1} = \mathbf{z}_m^n - \alpha_m^n(\nabla E(\mathbf{z}_m^n))$, where α_m^n is the step size of n^{th} iteration;
    $n = n + 1$;
  until $\frac{||z_m^n - z_m^{n-1}||^2}{||z_m^{n-1}||^2} \leq 10^{-06}$;
  set $\hat{\mathbf{z}}_m = \mathbf{z}_m^{n-1}$;
end

Table 4.1: Details of experimentation setup

	Degraded dataset	Un-degraded datset
Image size: MS Pan	64×64 256×256	256×256 1024×1024
Fused image size	256×256	1024×1024
Data format	Same as original data format	
Downsampling operation ($q = 4$)	Every 4×4 pixels of original images are averaged to get corresponding single pixel in downsampled image	Not required
Image format for display of results	24 bits RGB format	

4.7 Experimental Results

In this section, we present results of the proposed method for fusion. The fusion results using CT as well as NSCT are discussed and the comparison is shown with various state of the art approaches. In addition to fusion experiments conducted on both the degraded and un-degraded (original) images, we have also conducted experiments to show the effect of estimated decimation matrix entries (equation (4.7)) over the average one (equation (4.6)) and the effect of estimated MRF parameter on the fusion results. In order to show effectiveness of the proposed approach, we have conducted experiments on three sets of real images captured using Ikonos-2, Quickbird and Worldview-2 satellites. The detailed information of these satellites data are given in Table 3.1. The spatial resolution difference between the Pan and MS images is 4 for the images captured using these satellites. We performed experiments for a resolution difference of 2 as well as 4. Our approach for spatial resolution 2 outperforms when compared to other methods. Due to space constraints, we are demonstrating results for spatial resolution of 4 only. In Table 4.1 we show the details of setup used for degraded and un-degraded datasets. In all experiments, we use the "$9-7$" Laplacian pyramid (LP) filter bank based CT/NSCT as proposed in [171]. Along with this LP filter bank we use "*pkva*" and "$dmaxflat7$" directional filter banks for CT and NSCT decomposition, respectively [81, 172]. In order to compare the performance of proposed approach using CT/NSCT transform with the wavelet transform based approach we conducted the experiments where wavelet transform is used instead of CT/NSCT to obtain initial estimate. The experiments are conducted using MRF prior with first as well as second order neighborhood systems. The performance of the proposed method is compared with other approaches on the basis of quality of images in terms of perceptual as well as quantitative measures. These measures are computed using the fused MS and the original MS images. In order to make the availability of original for the quantitative comparison, the experiments are conducted on degraded images by using down sampling operation ($q = 4$) on both Pan and MS images. The following quantitative measures are used for performance evaluation for those experiments conducted on degraded images.

1. The CC is given in equation (3.23) and it is widely used by the fusion community. It finds the correlation between original MS band and fused MS band. Higher values

of CC indicates high spectral and spatial quality of the fused images.

2. Recently, Zhang *et al.* [173] have proposed a new image quality assessment called feature-similarity (FSIM). The FSIM is based on the value of phase congruency (PC) and gradient magnitude (GM) of a local structure. The high value of PC indicates highly informative features. To find the FSIM between two images f and $\hat{f}$, first the PC and the GM values for all pixels are computed as given in [173]. Using the phase congruency and the gradient magnitude, the similarity measures between f and $\hat{f}$ are calculated as

$$S_{PC}(i,j) = \frac{2PC_f(i,j) \cdot PC_{\hat{f}}(i,j) + T_1}{PC_f^2(i,j) + PC_{\hat{f}}^2(i,j) + T_1}, \tag{4.23}$$

and

$$S_G(i,j) = \frac{2G_f(i,j) \cdot G_{\hat{f}}(i,j) + T_2}{G_f^2(i,j) + G_{\hat{f}}^2(i,j) + T_2}, \tag{4.24}$$

where T_1 and T_2 represent small positive constants. PC_f, $PC_{\hat{f}}$ and G_f, $G_{\hat{f}}$ represent the phase congruency and the gradient magnitude of f and $\hat{f}$, respectively. Then, the $S_{PC}(i,j)$ and the $S_G(i,j)$ are multiplied to get the similarity $S_L(i,j)$ i.e.,

$$S_L(i,j) = S_{PC}(i,j) \cdot S_G(i,j). \tag{4.25}$$

Finally, the FSIM is computed as [173]:

$$FSIM = \frac{\sum_{i,j\in\Omega} S_L(i,j) \cdot PC_m(i,j)}{\sum_{i,j\in\Omega} PC_m(i,j)}. \tag{4.26}$$

Here Ω represents the entire image region in spatial domain and $PC_m(i,j) = max(PC_f(i,j),\ PC_{\hat{f}}(i,j))$. The range of FSIM is 0 to 1 and a higher value represents better performance.

3. The definition of RMSE is given in equation (3.24). It is the average MSE computed over all the fused MS bands. Low value of RMSE indicates minimum difference between original and fused image.

4. The *erreur relative globale adimensionnelle de synthse* (ERGAS) defined in equation (3.25) gives the global spatial quality in fused image. Ideal value of ERGAS is

zero.

5. Spectral information divergence (SID) [174] measures the information divergence between the probability distributions generated by the fused MS and original MS image. It gives the spectral variability of a single mixed pixel from a probabilistic point of view. Smaller SID value indicates better fused image.

6. A universal image quality index (Q_{avg}) is defined in equation (3.26). It models distortion as a combination of three different factors: loss of correlation, luminance distortion and contrast distortion. The highest value of (Q_{avg}) is 1 and it indicates that the fused image is same as the original image.

Among the above measures, CC, RMSE, ERGAS, SID and Q_{avg} correspond to the spectral distortion [37, 60, 86] while FSIM gives the spatial fidelity of the fused image.

In addition to experiments of degraded images, the performance of the proposed method is also verified by conducting the experiments on the un-degraded (original) datasets. In this experiment the reference fused image is not available for quantitative comparison and hence we use following measures in which reference image is not required.

1. The quality with no reference (QNR) is given in equation (3.27). The ideal value of this measure is 1 means that the fused image is same as original MS image. This measure is calculated using spatial distortion (D_s) and spectral distortion (D_λ).

2. The spatial distortion (D_s) defined in equation (3.28) gives the spatial detail by comparing the MS and fused images with Pan image. It's ideal value is 0.

3. The spectral distortion (D_λ) tells the spectral distortion present in the fused image in comparison with original MS image and it's ideal value is 0. Equation (3.29) defines the D_λ.

In following subsections 4.7.1 and 4.7.2, we show the effect of decimation matrix coefficients and MRF parameter on fusion, respectively. The fusion results of degraded datasets obtained for different satellites are depicted in subsection 4.7.3 to 4.7.5. In subsection 4.7.6, we describe the fusion results obtained for un-degraded datsets. Finally, the computation time involved for different fusion methods is presented in 5.6.5.

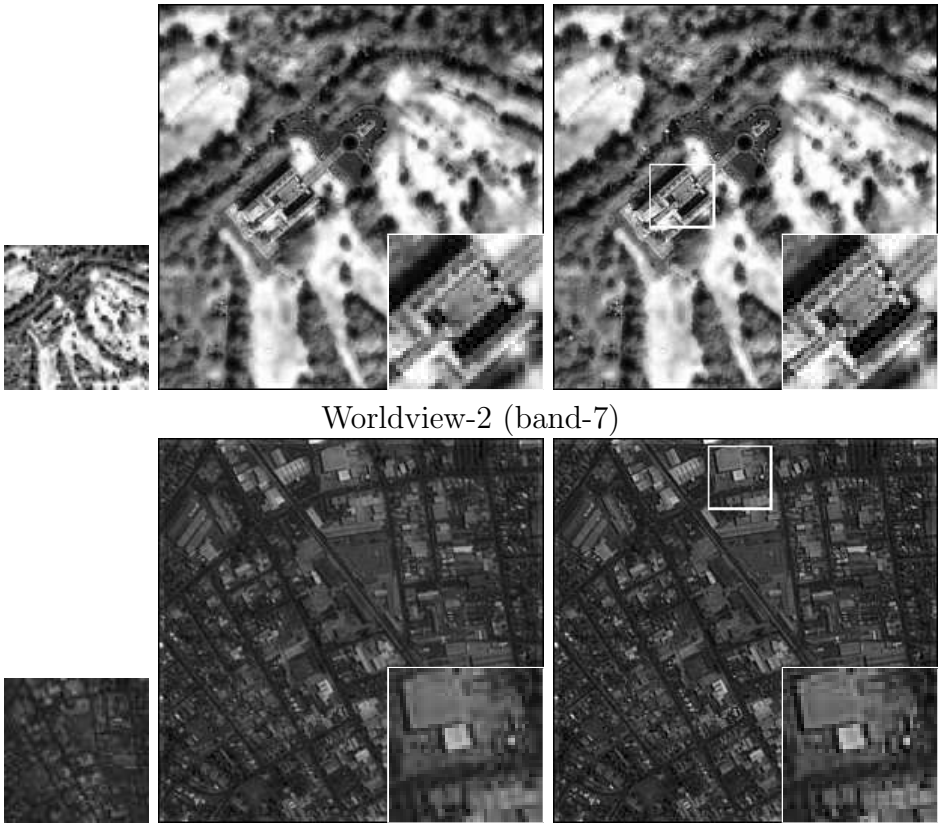


Figure 4.3: Experimental results to show the effect of decimation matrix with equal and unequal weights ($q = 4$). The first and second row consist fusion results for images captured using Worldview-2 (band-7) and Ikonos-2 (band-3) satellite sensors, respectively. (a) The downsampled LR MS image. The size of LR MS image is 64×64. Fused MS images of size 256×256 with (b) equal weights and (c) unequal weights. The zoomed-in version of small area shown with white color border in (c) are displayed at bottom right corner in (b, c).

4.7.1 Result on Effect of Decimation Matrix Coefficients

Before proceeding to discussing the fusion results using our proposed method based on CT and NSCT we first show the effect of decimation matrix on MS image fusion with unequal weights (equation (4.7)) and equal weights (equation (4.6)) by conducting the experiments on the degraded version images captured by two different satellites. Due to space constraint we display only results of band-7 and band-3 of Worldview-2 and Ikonos-2 satellites, respectively. The matrix with unequal weights were obtained by formulating the problem as minimization of $||\mathbf{y}_m - D_m\mathbf{z}_m^0||^2$ with respect to sparse entries in D_m, where $\mathbf{z}_m^0$ in the initial estimate derived from NSCT based learning. The locations of these entries correspond to those used for averaging (i.e., those locations where D_m entries were

Table 4.2: Quantitative measures for fused MS images shown in Fig. 4.3. Here values in boldface indicate better performance.

Image	Measure	Decimation matrix with equal weights	Decimation matrix with unequal weights
Worldview-2 (band-7)	MSE	0.0478	**0.0394**
	PSNR (in dB)	27.2941	**30.4381**
Ikonos-2 (band-3)	MSE	0.0491	**0.0411**
	PSNR (in dB)	25.9345	**28.8621**

having equal values). The unequal weights of the D_m are then obtained using least squares method. The final cost function given in equation (4.19) is then minimized by replacing A_m with D_m in order to obtain the fused image. The fusion results using equal and unequal weights are depicted in Fig. 4.3. In Fig. 4.3(a), we display the low resolution MS images. The fusion results with equal and unequal weights are displayed in Fig. 4.3(b) and Fig. 4.3(c), respectively. For better visualization we also display the zoomed-in version of a small area at bottom right corner in the fused images. Looking at these fused images we can observe that the results with unequal weights have reduced the blurring effect when compared to those of equal weights. One can also see the improvement in quantitative measures as given in Table 4.2. This Table shows the mean squared error (MSE) and the peak signal to noise ratio (PSNR) calculated between original and fused MS images. The results shown in Fig. 4.3 confirms that the use of decimation matrix with unequal weights (equation (4.7)) results in better fusion performance when compared to that obtained using equal weights (equation (4.6)). Note that in this experiment, the blur matrix is considered as an identity. However, in our work we take care of blurring as well as the aliasing by including the non-identity blur matrix and use the estimated degradation matrix (A_m) entries instead of D_m.

4.7.2 Effect of MRF parameter γ_m on Fusion

In the proposed approach, first the initial estimate image is obtained using the CT/NSCT transform, MS and Pan images. The final fused image is obtained by casting the problem in a regularization framework and solving the same by using MAP-MRF approach. In our regularization we use an MRF smoothness prior given in equation (4.13). Here, γ_m represents the MRF parameter. It serves as a weight that appropriately penalizes the departure from smoothness in the fused image and its value depends on the entity modeled as MRF. Since the fused image is modeled as MRF and same is unavailable,

true value of the MRF parameter is not known. However, since we have the initial estimate which represents close approximation to the final fused image, it can be used to estimate the MRF parameter. If this parameter is not estimated one may set it empirically however doing this increases the computational cost. Since the MRF parameter acts as regularizing parameter, it is interesting to see the effect of the same on the solution i.e. the fused image. In order to investigate the effect of γ_m on the solution, we conducted an experiment on band-1 of degraded Quickbird dataset [156] using the proposed approach in which different values of γ_1 are chosen manually. The results of this experiment are displayed in Fig. 4.4 for different choices of γ_1. Fig. 4.4(a-c) show the fused images with three different values of γ_1. Comparing the images displayed in Fig. 4.4(a-c), we observe that the fused image shown in Fig. 4.4(a) is smoothed out due to higher value of $\gamma_1 = 0.1$. The fused image in Fig. 4.4(b) looks a bit better when compared to Fig. 4.4(a) but edges are not preserved well. Although the fused image of Fig. 4.4(c) has minimum smoothness and has prominent edge details, it looks noisy. This is because of the reduction in weightage ($\gamma_1 = 0.001$) for the prior term in equation (4.19). In this case the final solution relies heavily on the first term (data term) in equation (4.19) forcing it to become noisy since it corresponds to least squares estimation with much less emphasis on prior. In Fig. 4.4(d) we display the fused image with estimated MRF parameter (γ_1) but without using the edge details extracted from the Canny edge detector. What we observe here is the loss of edge details. The entire image appears smooth since the edge regions are also included while minimizing the cost function. In Fig. 4.4(e) we display the fused image with estimated γ_1 but with edge pixels corresponding to those of the initial estimate which are estimated using the Canny edge detector. Thus we see that the Fig. 4.4(e) has better preservation of homogenous and heterogenous regions when compared to other results displayed in Fig. 4.4(a-d).

4.7.3 Fusion Results for Degraded Dataset: Ikonos-2

The Ikonos-2 satellite has one Pan and four MS images (band-1 to band-4) with the spatial resolution of $1m \times 1m$ and $4m \times 4m$, respectively. These images were downloaded from the Internet [154], which has images of urban as well as non-urban areas. All these images correspond to Mount Wellington area near Hobart Tasmania. They were captured

(a) $\gamma_1 = 0.1$ (b) $\gamma_1 = 0.01$ (c) $\gamma_1 = 0.001$ (d) $\gamma_1 = 0.0063$ (e) $\gamma_1 = 0.0063$

Figure 4.4: Effect of γ_1 on fusion result obtained for degraded Quickbird (band-1) satellite image. (a-c) Fused images with manually selected value of γ_1, fused MS images with estimated γ_1 (d) without using Canny edge detector and (e) with Canny edge detector.

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

Figure 4.5: LR test MS images obtained by downsampling the original MS images captured using different satellite sensors. The size of each LR test MS image is 64×64. The downsampled MS images for (a, b) Ikonos-2 satellite with color composition of bands-3, 2, 1 and bands-4, 3, 2 respectively, (c, d, e) Quickbird satellite with color composition of bands-3, 2, 1 and bands-4, 3, 2 corresponding to area around Boulder city, USA and bands-3, 2, 1 corresponding to Sundarban, India respectively and (d, e) Worlodview-2 satellite with color composition of bands-5, 3, 2 and bands-7, 5, 3, respectively.

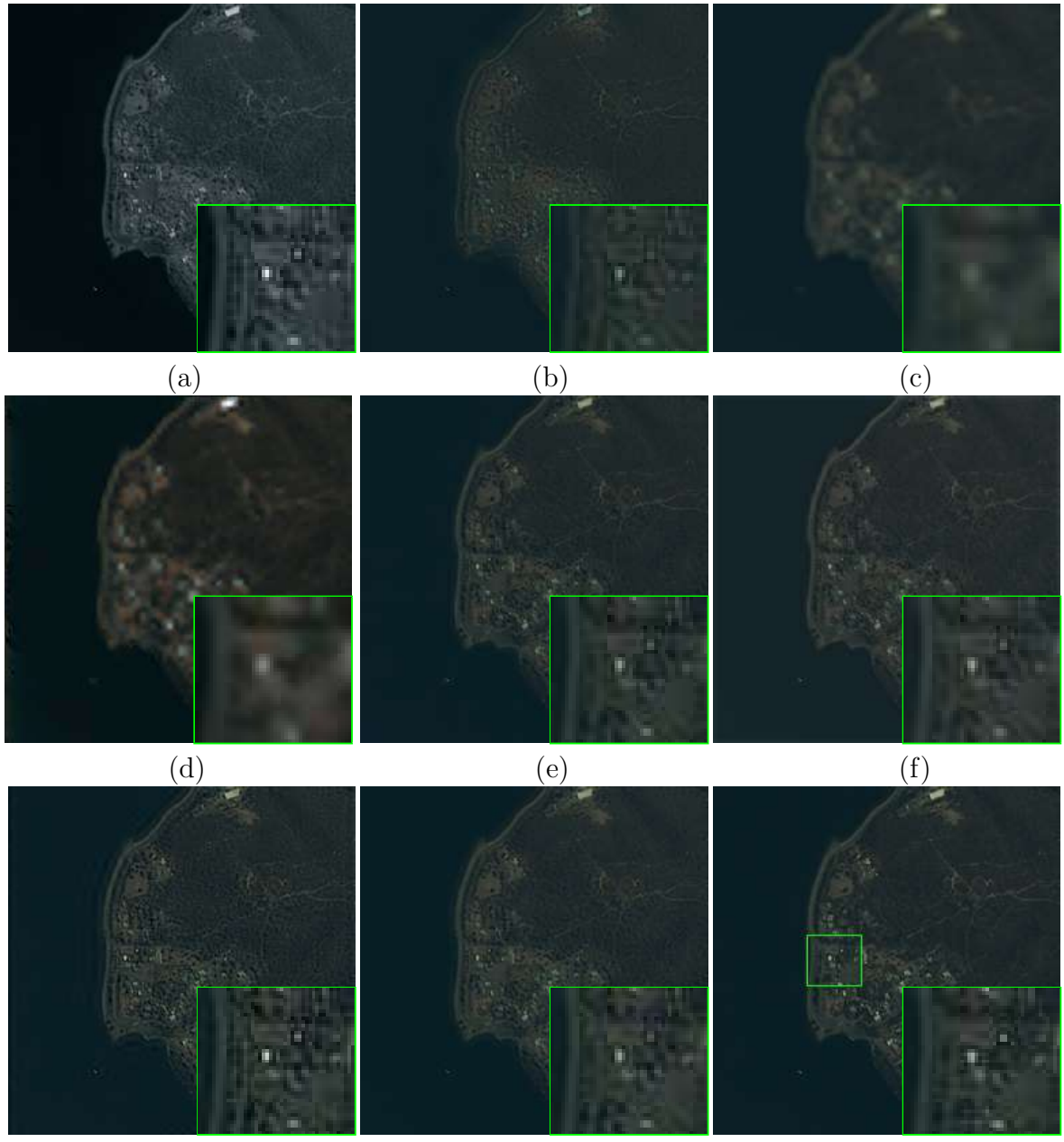


Figure 4.6: MS fusion results for downsampled Ikonos-2 satellite images consisting of non-urban area shown as color composite of bands-3, 2 and 1 ($q = 4$). Fused images obtained using (a) temporal Fourier transform (TFT) based approach [175], (b) approach in [87], (c) approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom right corner of all the images.

on 22^{nd} February 2003 and have radiometric resolution of 11-bits. The original Pan and MS images are of size 12124×13148 and 3031×3287, respectively. For experimentation, we prepared two datasets from these images by using cropping operation. These cropped

images are then coregistered. Our dataset used for experimentation consists of a Pan image and MS images of size 1024×1024 and 256×256, respectively. We downsample them by a factor of 4 and conduct the experiments using the degraded Pan and MS images of size 256×256 and 64×64, respectively. The radiometric resolution is kept same as the original for these images. Experiments are conducted on the available data without any preprocessing such as converting it into 8-bits or linear stretching etc. MRF parameters are estimated separately for each band using the corresponding initial estimate. The estimated value of MRF parameter γ_1 for band-1 is 0.00257. As mentioned earlier the experiments were conducted on first and second neighborhood MRF priors. From the results, we observed that using both these neighborhoods resulted in almost the same MSE and also visually similar results. However, with the use of second order neighborhood, the convergence speed was faster when compared to first order neighborhood. Similar observation was noticed in [176]. Here, the results are shown with second order neighborhood only. With the varying step size value in the gradient descent the convergence is obtained in less than 20 iterations which is fast compared to the fixed step size value.

The results are shown separately for images containing non-urban and urban area. In order to accommodate experiments on all the three sets of images, we show the fused results for non-urban area using bands 3, 2 and 1 while the results on urban area are shown using the bands 4, 3 and 2. The downsampled LR MS images for different satellites are displayed separately in Fig. 4.5 in which Fig. 4.5(a) and Fig. 4.5(b) correspond to LR MS images for Ikonos-2 satellite. The fusion results for the two cases are shown as color composites in Fig. 4.6 and Fig. 4.7, respectively and the comparison is shown with various state of the art fusion approaches. In both the figures, the magnified portion is shown at the bottom-right corner on each of the fused images. Fig. 4.6(a) and Fig. 4.7(a) show the fused images obtained using temporal Fourier transform (TFT) [175] method. The results obtained using the approach based on NSCT [87] are displayed in Fig. 4.6(b) and Fig. 4.7(b), respectively. In Fig. 4.6(c) and Fig. 4.7(c), we show the fused images for the approach in [86]. We also display the fusion results obtained using [103] in Fig. 4.6(d) and Fig. 4.7(d). Fig. 4.6(e) and Fig. 4.7(e) show the fusion obtained using the adaptive IHS method [37]. The MS fusion images using AWLP method [60] are shown in Fig. 4.6(f) and Fig. 4.7(f). It may be mentioned here that we choose these methods for comparison since

Table 4.3: Quantitative measures for fused MS images shown in Fig. 4.6. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 1(B)	1	0.8042	0.8672	0.8403	0.8404	0.9093	0.8469	0.9232	**0.9250**
	Band 2(G)	1	0.8814	0.8868	0.8630	0.8597	0.9211	0.9324	**0.9557**	0.9489
	Band 3(R)	1	0.9385	0.8994	0.8929	0.8837	**0.9655**	0.9596	0.9628	0.9633
	Band 4(NIR)	1	0.9519	0.9565	0.9312	0.9259	0.9646	0.9673	**0.9782**	0.9719
FSIM	Band 1(B)	1	0.8304	0.9269	0.8977	0.8907	0.9438	0.8809	0.9550	**0.9574**
	Band 2(G)	1	0.9104	0.9171	0.8577	0.8735	0.9536	0.9322	0.9621	**0.9673**
	Band 3(R)	1	0.9349	0.9160	0.8397	0.8476	0.9513	0.9500	0.9525	**0.9530**
	Band 4(NIR)	1	0.8881	0.8977	0.7629	0.7924	0.8749	0.9171	0.9173	**0.9230**
RMSE	Band 1-4	0	9.08	4.53	5.08	9.55	3.48	3.81	3.07	**3.01**
ERGAS	Band 1-4	0	7.57	4.16	4.59	6.52	3.12	3.28	2.88	**2.83**
SID	Band 1-4	0	0.0530	0.0075	0.0386	0.0840	0.0394	0.0386	0.0061	**0.0057**
Q_{avg}	Band 1-4	1	0.4989	0.9008	0.8960	0.5303	0.9003	0.8962	0.9104	**0.9146**

they perform the fusion using the edge details in Pan image which is in similar line with the proposed method. Finally, the results of the proposed approach using CT and NSCT are depicted in Fig. 4.6(g, h) and Fig. 4.7(g, h), respectively. For visual comparison we display the original MS images in Fig. 4.6(i) and Fig. 4.7(i), respectively.

Looking at the results displayed in Fig. 4.6 for non-urban area the following points can be noticed. One can see that fused image using TFT method (see Fig. 4.6(a)) although shows improvement in spatial information it has spectral distortions. This can be seen in the magnified region in which the artifacts are clearly visible. The approaches based on NSCT [86, 87] (Fig. 4.6(b) and Fig. 4.6(c)) fail in both spectral and spatial information preservation. Similarly, we can observe the poor enhancement of details in Fig. 4.6(d) [103]. Further, the AWLP approach [60] (Fig. 4.6(f)) preserves the spectral details but the improvement of spatial details is lacking. This may be because of the non usage of prior information. Although the adaptive IHS method (see Fig. 4.6(e)) gives better fused image perceptually the results of the proposed approach shown in Fig. 4.6(g, h) are better both in terms of perception and quantitative measures. We can see from the magnified details that in the proposed method the area nearer to the sea have better preservation of spatial details when compared to the same in Fig. 4.6(a-f). Here, we can also observe that the proposed approach with NSCT (Fig. 4.6(h)) has less spatial and color distortion and also appears closer to original MS image (Fig. 4.6(i)) when compared to the result using CT (Fig. 4.6(g)). The quantitative comparison using all 4 bands for non-urban area (Fig. 4.6) is given in Table 4.3 that also has the ideal value for each measure as the reference. The values marked in boldface indicate that their value is

closer to the ideal. It is clearly observed that the proposed method has better CC when compared to all other methods except for band 3. For this band though the CC value for the proposed method is less, it is still closer to the adaptive IHS technique [37] which has the highest value. The FSIM measure is based on phase congruency (PC), and is highest when there is significant pixel intensity difference between neighboring pixels. Hence a higher value of this indicates that the edges are well preserved. From the Table 4.3, one can observe that the proposed method with the NSCT has the high value of FSIM when compared to other approaches. The other spatial measures such as RMSE and ERGAS are also indicated in Table 4.3. As expected for these measures the proposed approach shows better performance when compared to all the other approaches. The SID measure which indicates the spectral distortion present in the image is better when compared to other methods. The Q_{avg} which quantifies spatial as well as the spectral details is also highest for the proposed method. From the Table 4.3, we can say that the proposed method using CT and NSCT performs better in terms of quantitative measures when compared to other methods except for correlation coefficient which is marginally less than adaptive IHS approach for band 3. However, we can see that fusion using adaptive IHS is visually poor when compared to the proposed method.

Similar observations can be made for Fig. 4.7 which has urban area. Here also we display the magnified portion separately as seen at the bottom right corner. We see that all other approaches except the AWLP method [60] show inferiority in terms of visual perception when compared with our approach. But looking at the magnified portion which consists of the areas around large building one can say that preservation of spatial details in the proposed approach are better when compared to AWLP approach [60]. A careful observation shows that distinguishing a region such as the one shown in the magnified area from the other regions is more easier in the proposed methods, indicating better preservation of spectral features. Interestingly, one can also see that the artifacts are less dominant in Fig. 4.7(h) indicating that NSCT performs better due to non sub-sampling property. The quantitative comparison for the results on urban area is shown in Table 4.4. As indicated the proposed method performs better in terms of quantitative measures when compared to other approaches. We see from the Table 4.4 that the CC measures for both CT and NSCT are better among all other approaches. Similarly, we can see the improvement in the FSIM for the proposed method when compared to

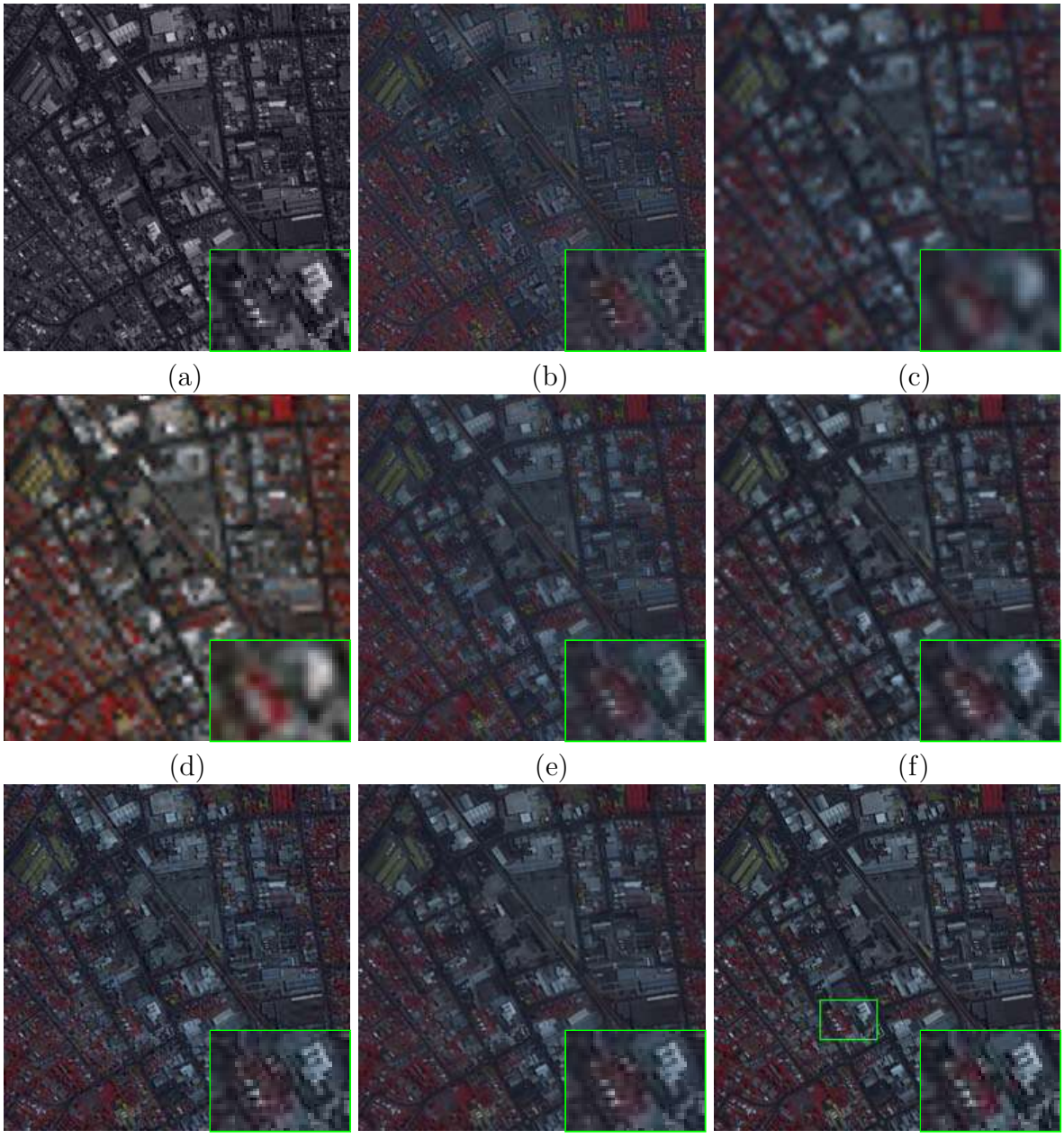


Figure 4.7: MS fusion results for degraded dataset of urban area images captured using Ikonos-2 satellite shown as color composite of bands-4, 3 and 2 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [87], (c) approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom right corner of all the images.

other methods and the values of RMSE and ERGAS are closer to the ideal value. The measures that show the spectral details preservation i.e., SID and Q_{avg} are also better when compared to other methods. We mention here that the use of NSCT improves the

Table 4.4: Quantitative measures for fused MS images result shown in Fig. 4.7. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 1(B)	1	0.8892	0.8359	0.8107	0.8121	0.9405	0.9323	0.9406	**0.9441**
	Band 2(G)	1	0.9215	0.8553	0.7988	0.8009	0.9487	0.9380	**0.9543**	**0.9543**
	Band 3(R)	1	0.9415	0.8702	0.7860	0.7890	0.9493	0.9322	**0.9498**	0.9497
	Band 4(NIR)	1	0.8787	0.8589	0.7121	0.7155	0.9016	0.8925	0.9032	**0.9200**
FSIM	Band 1(B)	1	0.8303	0.8568	0.7611	0.7893	0.9043	0.9115	**0.9155**	0.9083
	Band 2(G)	1	0.9118	0.8496	0.7203	0.7728	0.8860	0.9083	0.9255	**0.9294**
	Band 3(R)	1	0.9324	0.8607	0.7131	0.7736	0.8760	0.8972	**0.9394**	0.9320
	Band 4(NIR)	1	0.9049	0.8035	0.6618	0.7572	0.8228	0.8719	0.9091	**0.9164**
RMSE	Band 1-4	0	11.65	12.99	14.02	15.19	9.56	9.49	9.04	**7.87**
ERGAS	Band 1-4	0	5.44	6.07	6.52	6.27	4.46	4.42	4.25	**3.69**
SID	Band 1-4	0	0.0257	0.0144	0.0182	0.0312	0.0145	0.0139	0.0125	**0.0116**
Q_{avg}	Band 1-4	1	0.4945	0.7135	0.7390	0.7276	0.7558	0.7454	0.7582	**0.7695**

RMSE and ERGAS when compared to CT method, however this improvement is not significant for other measures. From the results displayed in Fig. 4.6 and Fig. 4.7 and the quantitative measures shown in Table 4.3 and Table 4.4, we can see that the proposed method performs well in terms of both perceptual as well as quantitative assessment for the results on images of Ikonos-2 satellite.

4.7.4 Fusion Results for Degraded Dataset: Quickbird

We now consider images captured using Quickbird satellite, which provides Pan and MS images with spatial resolutions of $0.6m \times 0.6m$ and $2.4m \times 2.4m$, respectively. These images were downloaded from the website [155]. This satellite data covers an area around Boulder city, USA. These images were captured on 4^{th} July, 2005 and their radiometric resolution is of 8-bit. The Pan and MS images were of size 3312×3260 and 828×815, respectively. After cropping operation, the Pan image has the size of 1024×1024 and the MS images have size of 256×256. Their LR versions i.e., the downsampled images are of size 256×256 and 64×64, respectively. The results for this experiment are shown in Fig. 4.8 and Fig. 4.9, respectively. In Fig. 4.8, we show the fusion results with color composite of bands-3, 2 and 1 covering the forest area, while Fig. 4.9 shows the fused images with color composite of bands-4, 3 and 2 and this consists of semi-urban area. The LR MS images are shown in Fig. 4.5(c) and Fig. 4.5(d). The fused images obtained for all other approaches except the proposed method are shown in Fig. 4.8(a-f) and Fig. 4.9(a-f). In Fig. 4.8(g, h) and Fig. 4.9(g, h) we show the results obtained using the proposed

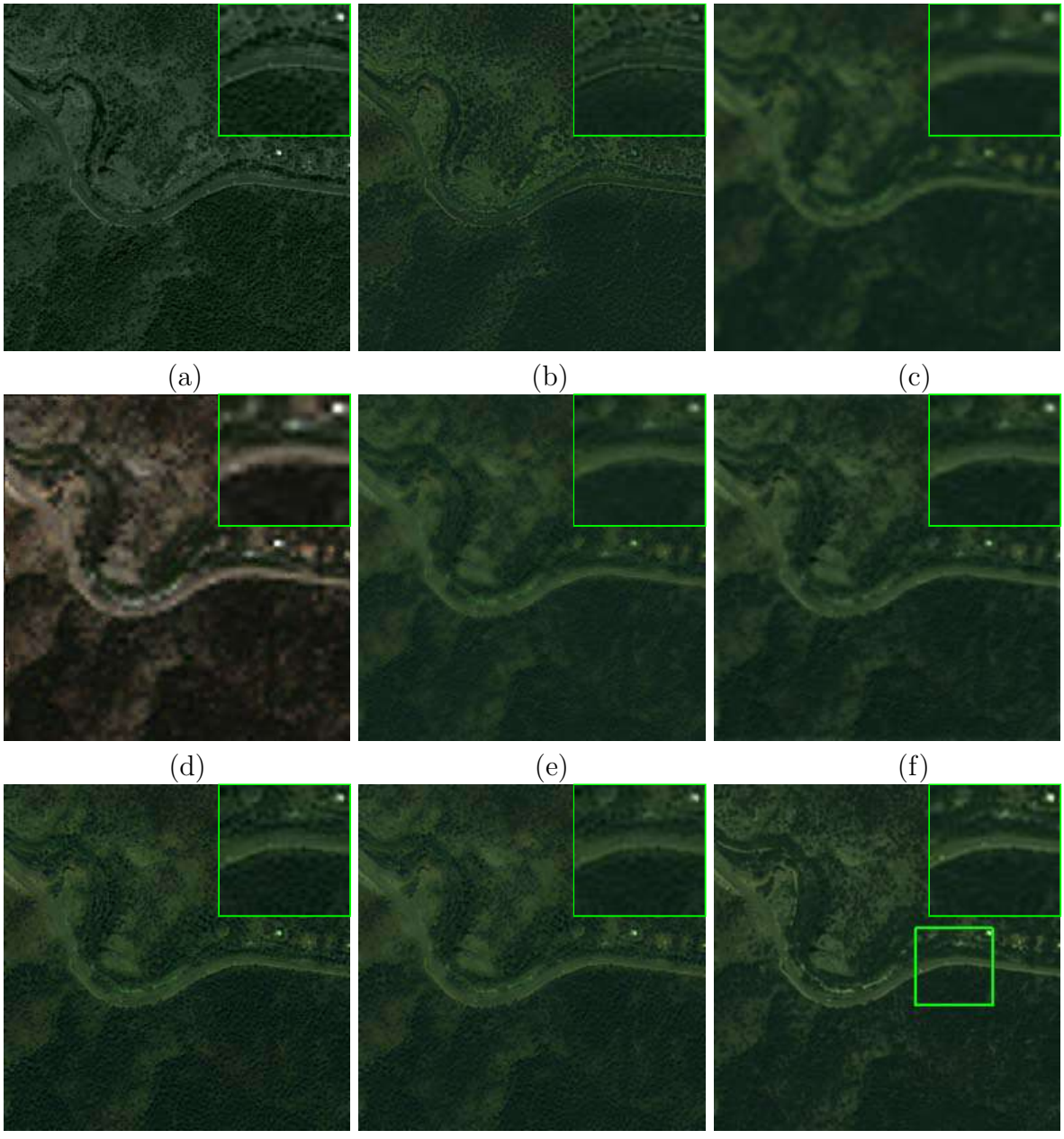


Figure 4.8: MS fusion results for degraded dataset of forest area captured using Quickbird satellite shown as color composite of bands-3, 2 and 1 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [87], (c) approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the top right corner of all the images.

method using CT and NSCT. Finally, we display the original MS images in Fig. 4.8(i) and Fig. 4.9(i). For better visual clarity, we also show the magnified regions in all the fused images.

Similar to the experiments on Ikonos-2 images, here also we observe better preservation of spatial details in Fig. 4.8(a) but with spectral distortion. The results shown in Fig. 4.8(b-d), do not show comparable performance with the result of proposed method. The MS fused images shown in Fig. 4.8(e) and Fig. 4.8(f) using the adaptive IHS and AWLP show the improvement in spectral information but they lack in preserving the spatial content. From the magnified region consisting of road and tree areas one can see that the proposed approach using NSCT (see Fig. 4.8(h)) has better spatial and spectral contents when compared to AWLP [60] as well as adaptive IHS method [37]. A closer look indicates that the proposed approach using NSCT has better edge details and also texture is close to original MS image when compared to AWLP and adaptive IHS methods. Similar observations can be made from the images displayed in Fig. 4.9, where the improvement in the proposed method is evident in the magnified region. In Fig. 4.9(g, h), we can compare the performance of the proposed method using CT and NSCT. It is clearly visible that the proposed approach using NSCT performs better in terms of preservation of spatial details as well as in terms of spectral enhancement compared to the method using CT. In conjunction with the results shown in Fig. 4.8 and Fig. 4.9, the quantitative comparisons are given in Table 4.5 and Table 4.6, respectively. Following comments can be made by observing the values given in the two Tables. Looking at Table 4.5, the values of CC for the proposed CT and NSCT methods are better when compared to all the other approaches. Also, all the other measures except FSIM for band 4 and ERGAS computed using bands $1-4$ are closer to their ideal for the proposed approach when compared to other approaches. These two measures are better for AWLP method [60]. Similar observations can be made from Table 4.6, in which except for SID measure our proposed fusion approach performs better when compare to other approaches. This indicates that barring few bands, the performance of the proposed approach is better when compared to other methods. As expected edge enhancement is poor for the AWLP approach since it is based on WT which has limited directionality. From the fusion results displayed in Fig. 4.8 and 4.9 and the quantitative measures shown in Tables we may conclude that the proposed algorithm performs better even on Quickbird images.

In continuation with above experiments on Quickbird images, we also experimented on the images downloaded from the link [156]. These images cover an area of Sundarban, India and radiometric resolution of these image is of 11-bits. The Pan and MS images

Table 4.5: Quantitative measures for fused MS images result shown in Fig. 4.8. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 1(B)	1	0.7152	0.7223	0.8829	0.8794	0.9274	0.8835	0.9375	**0.9381**
	Band 2(G)	1	0.7854	0.7552	0.8819	0.8786	0.9408	0.9312	0.9458	**0.9465**
	Band 3(R)	1	0.7417	0.7387	0.8871	0.8817	0.9313	0.9266	0.9318	**0.9401**
	Band 4(NIR)	1	0.7901	0.7308	0.8235	0.8282	0.9101	0.9116	0.9150	**0.9223**
FSIM	Band 1(B)	1	0.7581	0.8831	0.8695	0.8235	0.9332	0.9217	**0.9390**	0.9136
	Band 2(G)	1	0.8463	0.8526	0.7895	0.8398	0.8845	0.8870	0.8945	**0.9017**
	Band 3(R)	1	0.8270	0.8467	0.8041	0.8253	0.8698	0.8798	0.8874	**0.8934**
	Band 4(NIR)	1	0.8263	0.7515	0.6792	0.7864	0.8098	**0.9025**	0.8735	0.8765
RMSE	Band 1-4	0	10.49	14.85	7.40	13.21	5.83	5.54	**5.48**	5.56
ERGAS	Band 1-4	0	6.58	6.76	3.85	5.26	3.08	**3.01**	3.04	3.21
SID	Band 1-4	0	0.1051	0.0574	0.0308	0.0492	0.0259	0.0252	0.0236	**0.0233**
Q_{avg}	Band 1-4	1	0.8786	0.8053	0.9483	0.7536	0.9463	0.9470	0.9492	**0.9575**

Table 4.6: Quantitative measures for fused MS images result shown in Fig. 4.9. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 1(B)	1	0.6621	0.6814	0.8676	0.8565	0.9162	0.9169	0.9151	**0.9278**
	Band 2(G)	1	0.7022	0.7061	0.8613	0.8518	0.9270	0.9210	**0.9324**	0.9149
	Band 3(R)	1	0.6507	0.6852	0.8631	0.8559	0.9189	0.9133	0.9303	**0.9313**
	Band 4(NIR)	1	0.6195	0.5609	0.6966	0.7952	0.8746	0.8751	0.8766	**0.8820**
FSIM	Band 1(B)	1	0.7463	0.8242	0.8323	0.8514	0.8856	0.8892	0.8849	**0.8901**
	Band 2(G)	1	0.7933	0.7976	0.7690	0.7862	0.8519	0.8533	0.8559	**0.8590**
	Band 3(R)	1	0.7860	0.8011	0.7848	0.8119	0.8488	0.8437	0.8505	**0.8511**
	Band 4(NIR)	1	0.7589	0.6586	0.6966	0.7617	0.7709	0.8378	0.8414	**0.8451**
RMSE	Band 1-4	0	17.87	17.44	12.43	18.93	10.32	9.61	6.82	**6.64**
ERGAS	Band 1-4	0	8.91	8.20	5.62	8.72	4.35	4.39	2.88	**2.81**
SID	Band 1-4	0	0.2038	0.1088	0.0507	0.1904	0.0686	**0.0318**	0.0490	0.0514
Q_{avg}	Band 1-4	1	0.7678	0.7924	0.8981	0.7130	0.9042	0.9011	0.9067	**0.9090**

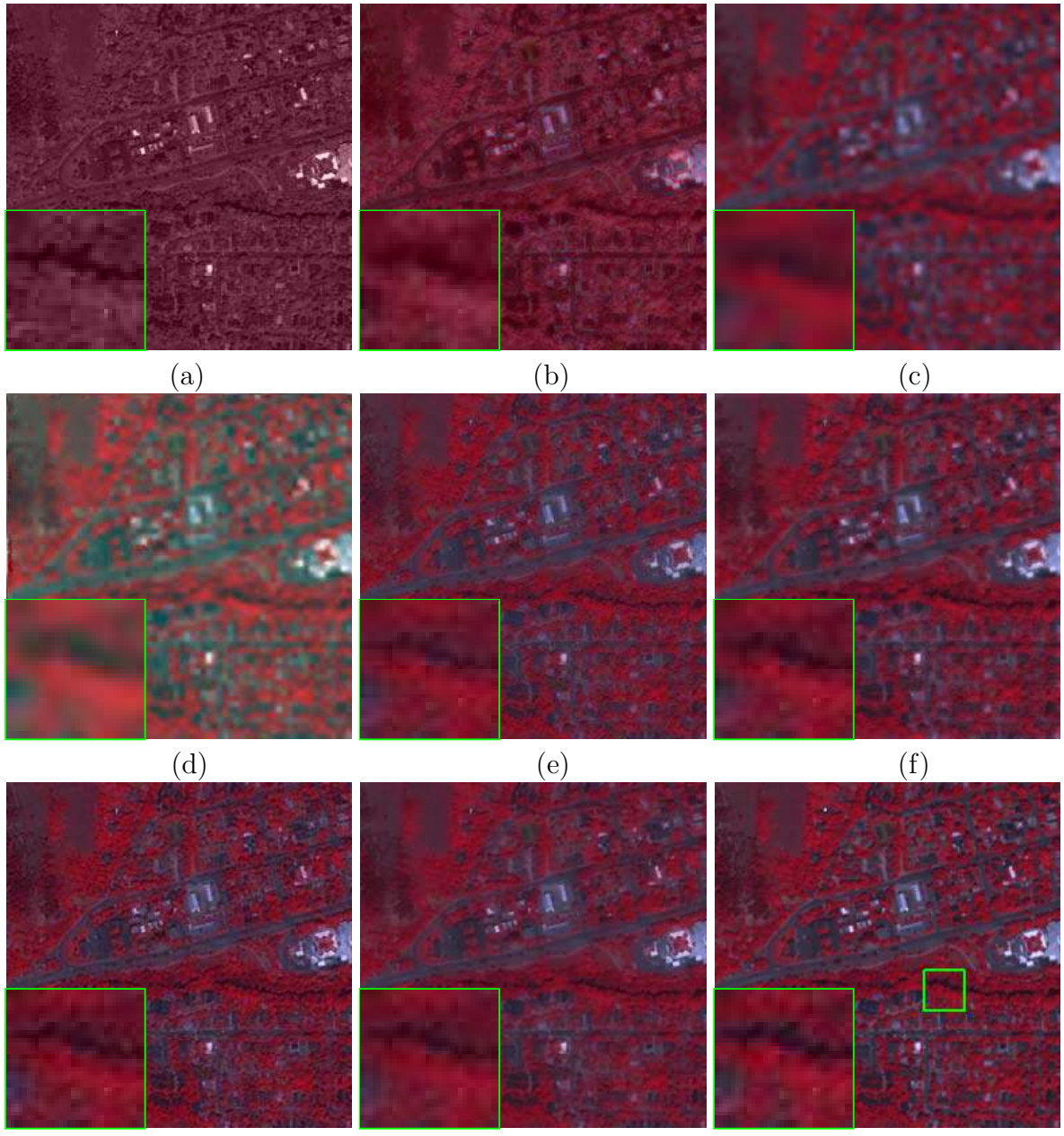


Figure 4.9: MS fusion results for downsampled Quickbird satellite images consisting of semi-urban area shown as color composite of bands-4, 3 and 2 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [87], (c) approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

are of size 16384×16384 and 4096×4096, respectively. The test images were prepared by cropping and co-registering these images and the experiments were conducted on degraded as well as on un-degraded images. The results of experiments conducted on

Table 4.7: Quantitative measures for fused MS images result shown in Fig. 4.10. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 1(B)	1	0.5466	0.5756	0.9250	0.9099	0.9137	0.9002	0.9244	**0.9256**
	Band 2(G)	1	0.5505	0.5309	0.9208	0.9071	0.9304	0.9316	9220	**0.9363**
	Band 3(R)	1	0.4581	0.5159	0.9225	0.9146	0.9251	0.9380	9165	**0.9389**
	Band 4(NIR)	1	0.9182	0.5091	0.8967	0.8816	0.9759	0.9614	0.9578	**0.9742**
FSIM	Band 1(B)	1	0.6792	0.7236	0.7730	0.7858	0.8052	0.8076	0.8129	**0.8417**
	Band 2(G)	1	0.7446	0.7559	0.7568	0.7492	0.8485	0.8457	0.8390	**0.8509**
	Band 3(R)	1	0.6972	0.7241	0.7541	0.7392	0.8160	0.8333	**0.8577**	0.8501
	Band 4(NIR)	1	0.8108	0.7156	0.7023	0.6074	0.8976	0.9079	0.8763	**0.9183**
RMSE	Band 1-4	0	19.90	20.01	20.90	21.63	17.47	16.76	14.67	**14.02**
ERGAS	Band 1-4	0	9.89	8.05	5.78	7.12	5.25	5.01	5.76	**4.95**
SID	Band 1-4	0	0.4381	0.4370	0.0320	0.1119	0.0686	0.0381	0.0431	**0.0378**
Q_{avg}	Band 1-4	1	0.6936	0.5381	0.7930	0.5377	0.8034	0.8827	0.8571	**0.8901**

degraded images are displayed in Fig. 4.10 and the quantitative comparison for the same is depicted in Table 4.7. One can see that the results of temporary Fourier transform (TFT) [175] and NSCT based [87] methods are very poor. Also, results of the approaches proposed in both [86] and [103] (see Fig. 4.10(c, d)) appear blurred indicating spatial and spectral distortions. We display the results of adaptive IHS and AWLP methods in Fig. 4.10(e) and Fig. 4.10(f), respectively and they appear perceptually similar to fused results of the proposed method using CT and NSCT shown in Fig. 4.10(g, h). We may see that the spectral improvement of the proposed method is similar to adaptive IHS and AWLP methods. However, we observe that both these methods lack in preserving the spatial information when compared to the proposed method which is visible in magnified region that show clear edge details for the proposed approach. We include the original MS image in Fig. 4.10(i). The quantitative comparison is presented in Table 4.7 where the proposed method performs better for all the measures. Before we proceed for next experiment, we would like to mention the following points related to this experiment. In the fusion results shown in Fig. 4.10 for the degraded dataset of color composite of bands-3, 2 and 1, the original band-1 MS image acquired from the satellite lacks sharpness. The down sampling operation further degrades it. Hence we notice spectral distortions in the results of this experiment. However, we see that in the proposed method though there exists spectral distortion, it still performs better in terms of the preservation of edge regions and spectral content as seen from the quantitative measures.

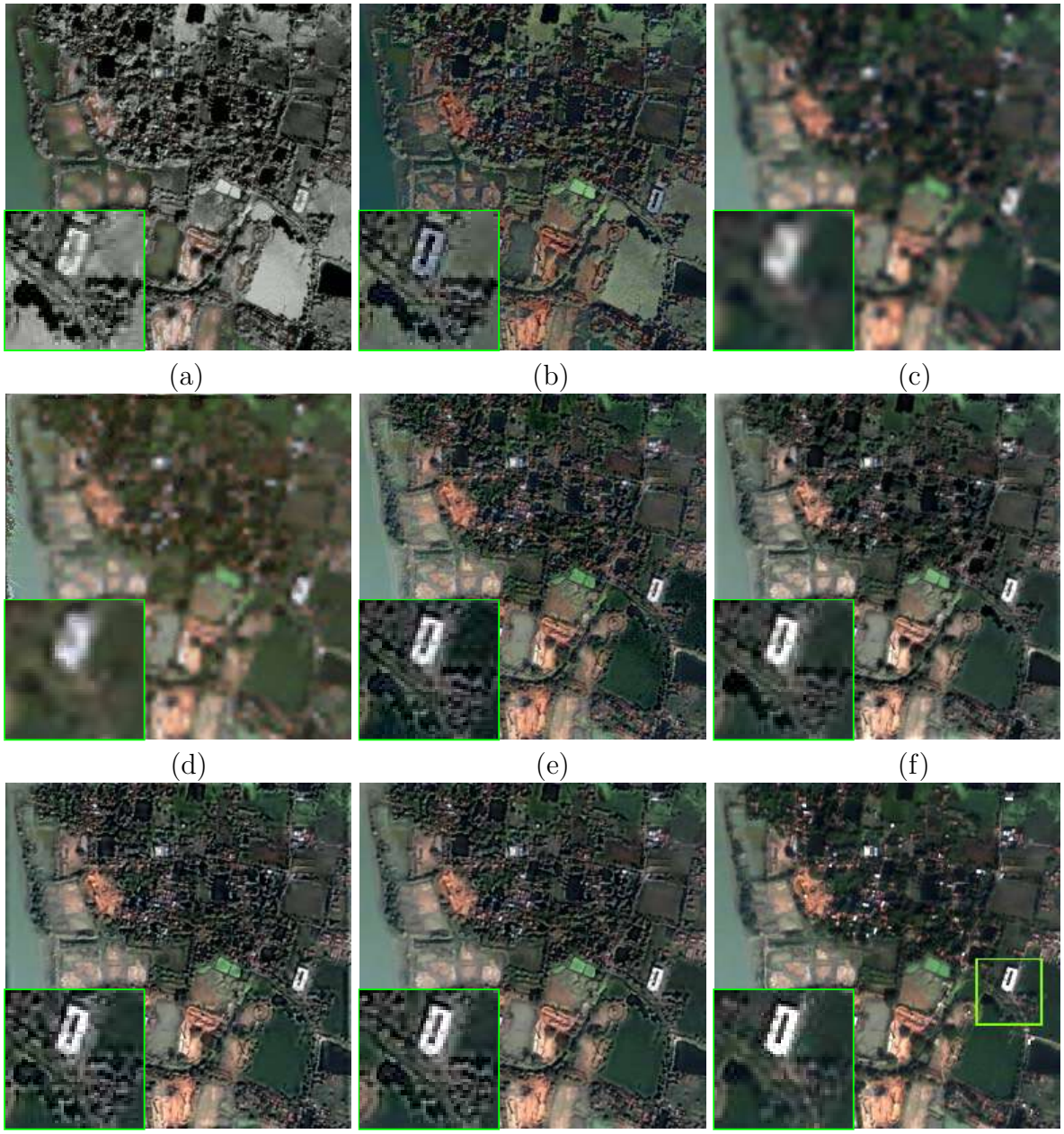


Figure 4.10: MS fusion results for degraded dataset of semi-urban area of Sundarban, India captured using Quickbird satellite shown as color composite of bands-3, 2 and 1 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [87], (c) approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

4.7.5 Fusion Results for Degraded Dataset: Worldview-2

Finally, we conduct the experiments on the images captured using Worldview-2 satellite sensor which has more than four MS images and also provides images with higher spatial

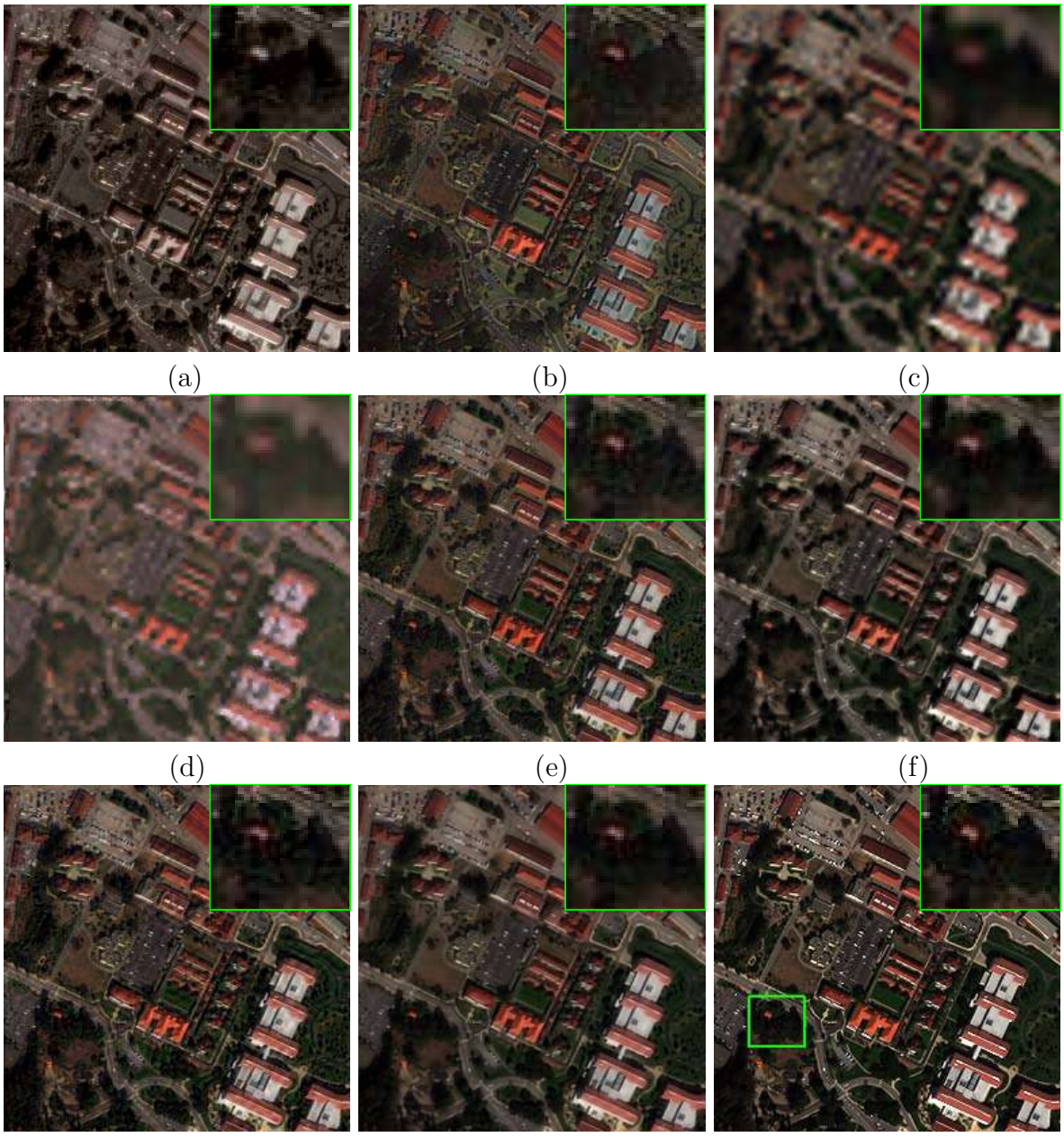


Figure 4.11: MS fusion results for degraded dataset of urban area image captured using Worldview-2 satellite shown as color composite of bands-5, 3 and 2 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [87], (c) approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the top right corner of all the images.

resolution compared to other satellites. This satellite captures a Pan image and eight MS images with spatial resolution of $0.5m \times 0.5m$ and $2m \times 2m$, respectively. The images used in this experiment are downloaded from [157] and they cover urban and semi urban areas

of San Francisco, USA. These images were captured on 9^{th} October, 2011 and have 11-bits of radiometric resolution. Here, also we use the images with same radiometric resolution without using any correction such as contrast stretching. The dataset consisting of Pan image of size 1024×1024 pixels and the MS images of size 256×256 pixels were used for experiments. The results are shown in Fig. 4.11 and Fig. 4.12, respectively. In Fig. 4.11 we display the fused images with color composite of bands-5, 3 and 2 and this has the urban area. The results in Fig. 4.12, showing semi-urban area are displayed by choosing color composite of bands-7, 5 and 3. The degraded LR MS images of urban and semi urban areas are shown in Fig. 4.5(e) and in Fig. 4.5(f), respectively. In Figs. 4.11 and 4.12 we display the fusion results obtained using various approaches along with their original MS images. Similar to previous two experiments here also we display the magnified portions of a region. The results of the proposed approach are shown in Figs. (4.11, 4.12)(g, h). Once again, looking at the images in Fig. 4.11 we see that the proposed method preserves high frequency edge details better when compared to other approaches. In the image showing urban area we select a region consisting of a house surrounded by trees for viewing it as magnified. From the magnified regions displayed in Fig. 4.11, we can see that spatial details in the tree and the color of the house are better preserved for the proposed method using NSCT when compared to all other methods. These details appear blurred for the AWLP method [60]. For adaptive IHS method [37] the same region has poor preservation of edges. Similarly, in Fig. 4.12 too one can observe the enhancement of spatial details for the proposed method when compared to the other approaches. Once again looking at the magnified portion we can see that texture preservation is better for the proposed approach based on NSCT. Here, one may notice blur in the fusion results obtained using methods in [87], [86], [103] (Fig. 4.12(b-d)). In Fig. 4.12(b) we display the result of fusion method [87] which has significant blur. This is due to the lesser correlatedness between MS and Pan images which yields the saturated fused image when using the method of [87]. Since the method proposed in [86] (Fig. 4.12(c)) is based on similar concept it also results in blur. The result of fusion approach proposed in [103] is displayed in Fig. 4.12(d) which uses the IGMRF parameters. Since it requires accurate registration of Pan and MS images in order to obtain the better fused image a small error in registration leads to degradation of edge/non edge regions in the fused image which is clearly visible as a distortion in Fig. 4.12(d).

Table 4.8: Quantitative measures for fused MS images result shown in Fig. 4.11. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 2(B)	1	0.8340	0.8395	0.7781	0.7811	0.9194	0.8955	0.9254	**0.9260**
	Band 3(G)	1	0.8800	0.8574	0.7769	0.7786	0.9367	0.9105	0.9386	**0.9412**
	Band 5(R)	1	0.8778	0.8632	0.8176	0.8164	0.9461	0.9314	**0.9501**	0.9489
	Band 7(NIR)	1	0.8113	0.5223	0.7861	0.7816	0.8747	0.8829	0.8828	**0.8866**
FSIM	Band 2(B)	1	0.8262	0.7916	0.6975	0.7054	0.8527	0.8440	0.8540	**0.8650**
	Band 3(G)	1	0.8659	0.8151	0.6824	0.6700	0.8749	0.8621	0.8909	**0.8949**
	Band 5(R)	1	0.7908	0.8124	0.6825	0.6839	0.8514	0.8584	0.8785	**0.8786**
	Band 7(NIR)	1	0.7169	0.6123	0.6490	0.6186	0.7790	0.7805	**0.7873**	0.7826
RMSE	Bands 2,3,5,7	0	27.26	35.11	31.13	39.73	22.32	21.72	**21.10**	21.30
ERGAS	Bands 2,3,5,7	0	10.23	12.56	12.08	11.35	8.36	8.18	7.88	**7.58**
SID	Bands 2,3,5,7	0	0.7670	0.2816	0.2709	0.2614	0.2541	0.2554	**0.2511**	0.3047
Q_{avg}	Bands 2,3,5,7	1	0.5141	0.5625	0.6644	0.5588	0.6795	0.6788	0.6819	**0.6837**

Table 4.9: Quantitative measures for fused MS images result shown in Fig. 4.12. Here boldface indicate values closer to the ideal.

Measure	Image	Ideal value	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
CC	Band 2(B)	1	0.7482	0.8286	0.8824	0.8852	0.9345	0.9329	0.9336	**0.9378**
	Band 3(G)	1	0.9013	0.8322	0.8944	0.8932	0.9677	0.9553	0.9746	**0.9754**
	Band 5(R)	1	0.7687	0.8265	0.8777	0.8789	**0.9427**	0.9363	0.9291	0.9387
	Band 7(NIR)	1	0.8898	0.7228	0.8896	0.8845	0.9319	0.9317	0.9347	**0.9377**
FSIM	Band 2(B)	1	0.7861	0.7846	0.7946	0.7761	0.8336	0.8357	0.8402	**0.8460**
	Band 3(G)	1	0.8721	0.7979	0.7800	0.7577	0.9200	0.8937	**0.9227**	0.9186
	Band 5(R)	1	0.7694	0.7546	0.7573	0.7203	0.8401	0.8424	**0.8440**	0.8364
	Band 7(NIR)	1	0.6912	0.7407	0.6823	0.6843	0.7661	**0.8393**	0.7698	0.7716
RMSE	Bands 2,3,5,7	0	25.62	29.40	22.38	28.85	17.40	15.69	14.82	**14.52**
ERGAS	Bands 2,3,5,7	0	14.59	18.32	12.02	13.37	8.70	8.55	6.88	**6.69**
SID	Bands 2,3,5,7	0	1.5759	0.4488	0.4542	0.5746	0.4892	0.4388	**0.4100**	0.4110
Q_{avg}	Bands 2,3,5,7	1	0.7517	0.4780	0.6581	0.6104	0.8374	0.8480	**0.8505**	0.8485

The quantitative comparisons for the same experiments of urban and semi urban images are displayed in Table 4.8 and Table 4.9, respectively. As can be seen from Table 4.8 that all the quantitative measures are better for the proposed methods. However, in Table 4.9, the CC measure of band 5 for adaptive IHS [37] shows minor improvement and the FSIM value for band 7 is better for AWLP method [60] when compared to the proposed approach. Except CC and FSIM measures the proposed method using CT and NSCT perform better for other measures. Thus, one can say that the our approach works well for Worldview-2 satellite images too.

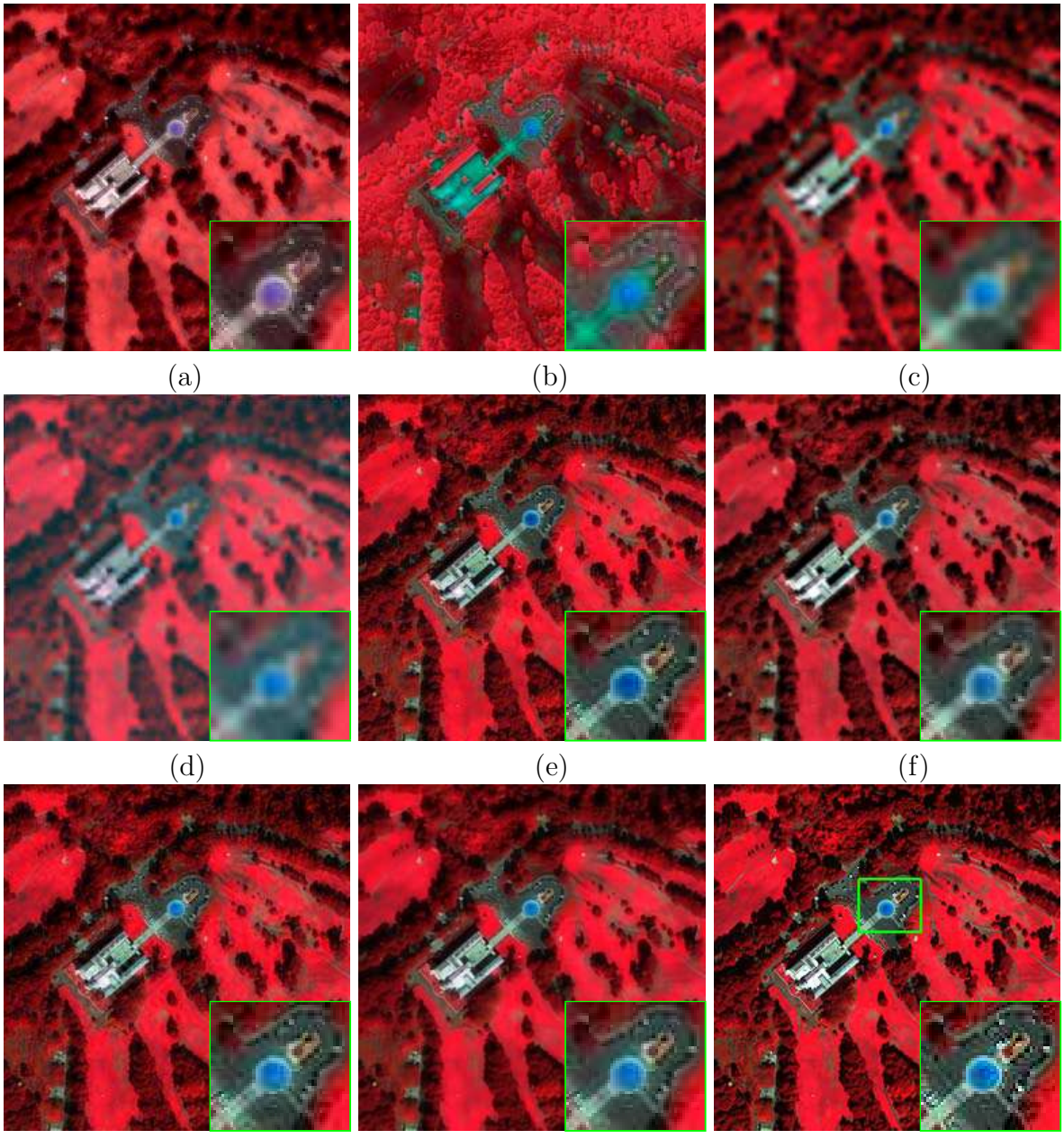


Figure 4.12: MS fusion results for degraded dataset of semi-urban area images captured using Worldview-2 satellite shown as color composite of bands-7, 5 and 3 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [87], (c)approach in [86], (d) approach in [103], (e) adaptive IHS approach [37], (f) AWLP [60], (g) proposed approach using CT and (h) proposed approach using NSCT. (i) Original MS image. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom right corner of all the images.

4.7.6 Fusion Results for Un-degraded (original) Datasets: Ikonos-2, Quickbird and Worldview-2

The potential of the proposed method is also checked by conducting the experiments on un-degraded (original) MS and Pan images. The same datasets consisting of images from Ikonos-2 [154], Quickbird [156] and Worldview-2 [157] satellites were used here. Unlike the previous experiments in this case, Pan and MS images are not downsampled which results in the size of them as 1024×1024 and 256×256, respectively. Similar to degraded dataset experiments, the radiometric resolution is not altered. The original MS images are displayed in Fig. 4.13. The fused MS images of various approaches along with proposed method for all the datasets are displayed in the Fig. 4.14 to Fig. 4.16. Fig. 4.14 shows the fusion results for images of Ikonos-2 satellite having semi-urban area while Fig. 4.15 and Fig. 4.16 correspond to results on urban area images captured using Quickbird and Worldview-2 satellites. The size of the fused MS image is 1024×1024. The performance of the proposed approach is compared with the methods used in the earlier experimentations on degraded images. Here, we also include the results of Brovey method [22] along with other approaches. The results of TFT based method [175], approach in [86], method in [103], adaptive IHS [37], AWLP [60] and Brovey method [22] are shown in Fig. 4.14(a-f) to Fig. 4.16(a-f). The results of the proposed approach based on CT and NSCT transforms are depicted in Fig. 4.14(h, i) to Fig. 4.16(h, i) for the three datasets. Due to space constraints we have not compared our results with NSCT based fusion method [87]. In order to compare the fusion performance of using CT/NSCT for initial estimate over the use of wavelet transform for the same, we also conducted the experiments to obtain the fused MS image using wavelet transform based initial estimate. Here we use "db3" wavelet in order to obtain the initial estimate. This initial estimate is used in the regularized framework to obtain the final fused image. The fused images obtained with this initial estimate are displayed in Fig. 4.14(g) to Fig. 4.16(g). In all these results we also include the zoomed-in versions of a small portion of the image for better visual comparison. The quantitative comparison is presented in Table 4.10 for all the un-degraded images where the quality index with no reference (QNR) [158] is used which does not require the reference image. The definition of QNR involves two terms: spectral distortion (D_λ) and spatial distortion (D_s). These measures

along with their ideal values are displayed in the Table 4.10. The boldface value in the table indicates that the value is best among the methods.

The following observations can be made by looking at the results displayed in Fig. 4.14 to Fig. 4.16. The fused images obtained using TFT method [175], approach in [86] and method in [103] show high spectral and spatial distortions. The other methods such as adaptive IHS [37], AWLP [60] and Brovey [22] show improvement in spectral and spatial details. In the results of Ikonos-2 dataset which is displayed in Fig. 4.14 one may see that the performance of the proposed method is comparable to those in [22, 37, 60] methods. However a close observation of these results indicate a small improvement in spatial details in the result of proposed method when compared to the other methods. The spectral content of the proposed method looks almost similar to other state of the art methods showing better preservation of colors. The quantitative performance for this experiment is listed in Table 4.10 in which the values of QNR and spectral distortion D_λ are better for the proposed NSCT based fusion approach and the spatial distortion factor D_s is better for Brovey method. But one can see that the difference between the values of D_s for the proposed and Brovey methods is insignificant. The performance of the proposed method is improved in the results of Quickbird and Worldview-2 satellites displayed in Fig. 4.15 and Fig. 4.16, respectively. In Fig. 4.15 we can see that the use of adaptive IHS method (Fig. 4.15(d)) causes color artifacts as well as spatial distortions. Similarly, black spots are visible in the result of adaptive IHS method for Worldview-2 satellite image (Fig. 4.16(d)). The results of AWLP method displayed in Fig. 4.15(e) and 4.16(e) show improvement in spectral and spatial details when compared to adaptive IHS method. However, the spectral distortion and the loss of edges details are less in the proposed method (see Fig. 4.15(i) and Fig. 4.16(i)) when compared to the results of AWLP method (Fig. 4.15(e) and Fig. 4.16(e)). The fusion results of Brovey method are shown in Fig. 4.15(f) and Fig. 4.16(f) for Quickbird and Worldview-2 datasets, respectively. In the result of Worldveiw-2 dataset (Fig. 4.16(f)) color artifacts are visible when compared to the result of the proposed method using NSCT displayed in Fig. 4.16(i). Thus, we observe that the performance of the proposed method in terms of perceptual quality is better when compared to the other approaches. Similar to the visual performance, the quantitative analysis (see Table 4.10) also shows better performance for the proposed method using NSCT when compared to all others. From these results on un-degraded

(a) (b) (c)

Figure 4.13: Un-degraded (original) MS images captured using different satellite sensors. The size of each MS image is 256×256. The MS images for (a) Ikonos-2 satellite with color composition of bands-3, 2, 1, (b) Quickbird satellite with color composition of bands-3, 2, 1 and (c) Worlodview-2 satellite with color composition of bands-7, 5, 3.

Table 4.10: Quantitative measures for fused MS image results shown in Fig. 4.15 to Fig. 4.16 for un-degraded datasets. Here boldface indicate values closer to the ideal.

Dataset	Measure	Ideal value	TFT method [175]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Brovey method [22]	Proposed using CT	Proposed using NSCT
Ikonos-2 (Fig. 4.14)	QNR	1	0.5772	0.4804	0.4305	0.6521	0.6662	0.6617	0.6664	**0.6772**
	D_λ	0	0.2639	0.2726	0.2658	0.1848	0.1653	0.1751	0.1683	**0.1543**
	D_s	0	0.2159	0.3396	0.4137	0.2001	0.2017	**0.1979**	0.1988	0.1992
Quickbird (Fig. 4.15)	QNR	1	0.5430	0.5581	0.5524	0.6176	0.6125	0.5955	0.6217	**0.6256**
	D_λ	0	0.2726	0.2190	0.2251	0.1825	0.1919	0.2006	0.1845	**0.1811**
	D_s	0	0.2535	0.2854	0.2871	0.2445	0.2420	0.2551	0.2376	**0.2360**
World-view-2 (Fig. 4.16)	QNR	1	0.5629	0.5200	0.5635	0.5852	0.5929	0.5859	0.6002	**0.6080**
	D_λ	0	0.2302	0.2298	0.1914	0.1769	0.1804	0.1996	0.1720	**0.1716**
	D_s	0	0.2688	0.3249	0.3031	0.2890	0.2766	0.2680	0.2751	**0.2660**

datasets and quantitative comparison one may conclude that the proposed method has better spectral and spatial details when compared to other state of the art methods. It is important to note that the fused image has significant spatial and color distortions if we choose the wavelet transform instead of CT/NSCT for obtaining the initial estimate. In Fig. 4.14(g) to Fig. 4.16(g), we display these fused images which have blockiness effect as well as color distortion. This may be due to the limited directionality property of wavelet transforms. We mention here that we conducted the experiments using degraded and un-degraded datasets having different resolutions with different types of regions and the obtained results using proposed method were better when compared to the state of the art approaches. Thus we conclude that our fusion method performs better on exhaustive dataset.

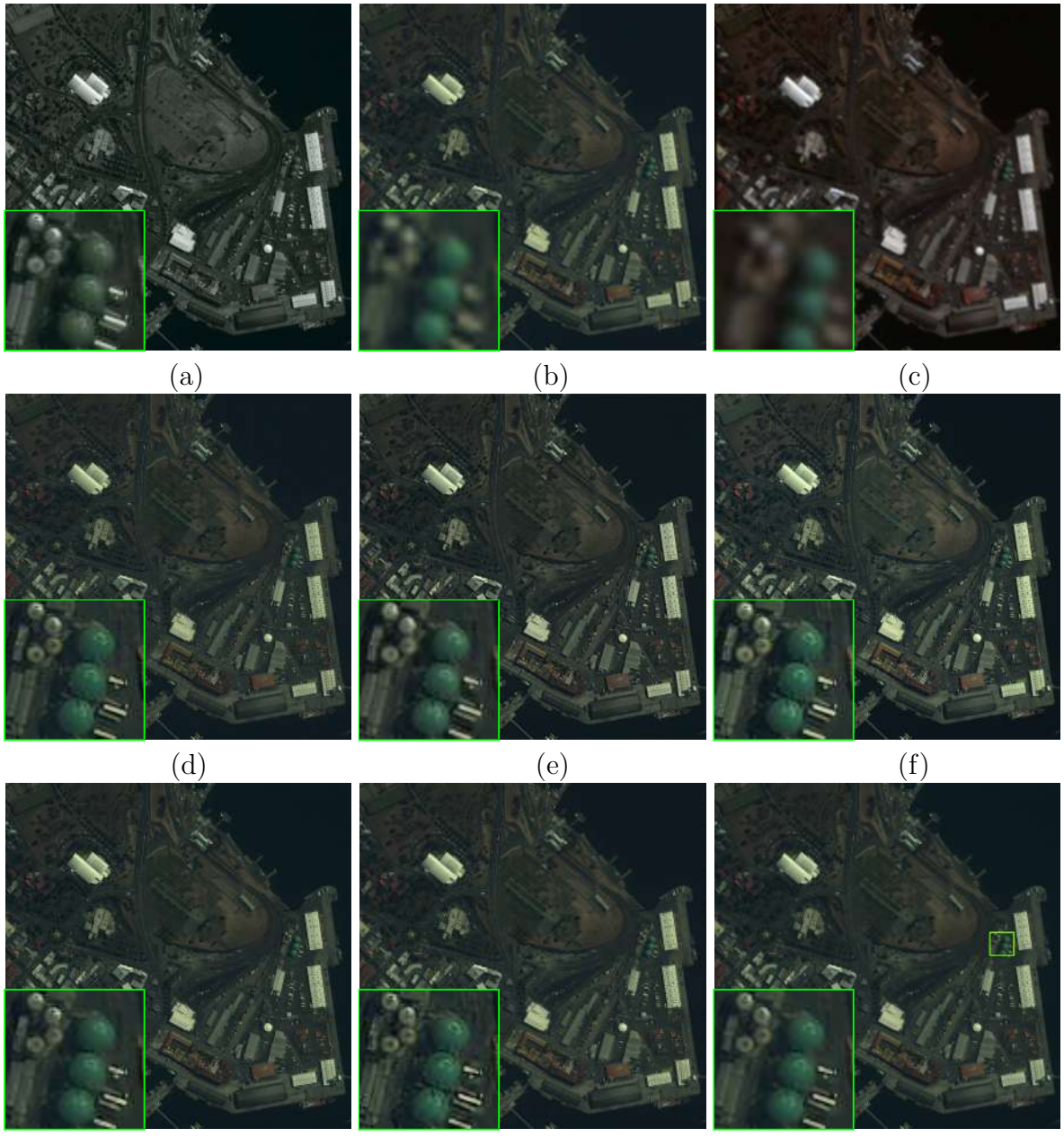


Figure 4.14: MS fusion results for un-degraded dataset consisting of semi-urban area captured using Ikonos-2 satellite shown as color composite of bands-3, 2 and 1 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [86], (c) approach in [103], (d) adaptive IHS approach [37], (e) AWLP [60], (f) Brovey method [22], (g) Wavelet based initial estimate, (h) proposed approach using CT and (i) proposed approach using NSCT. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

4.7.7 Spectral distortion at edge pixels

Since our approach is based on extracting the edges from the initial estimate and incorporating them in the final fused image it would be of interest to check spectral distortion

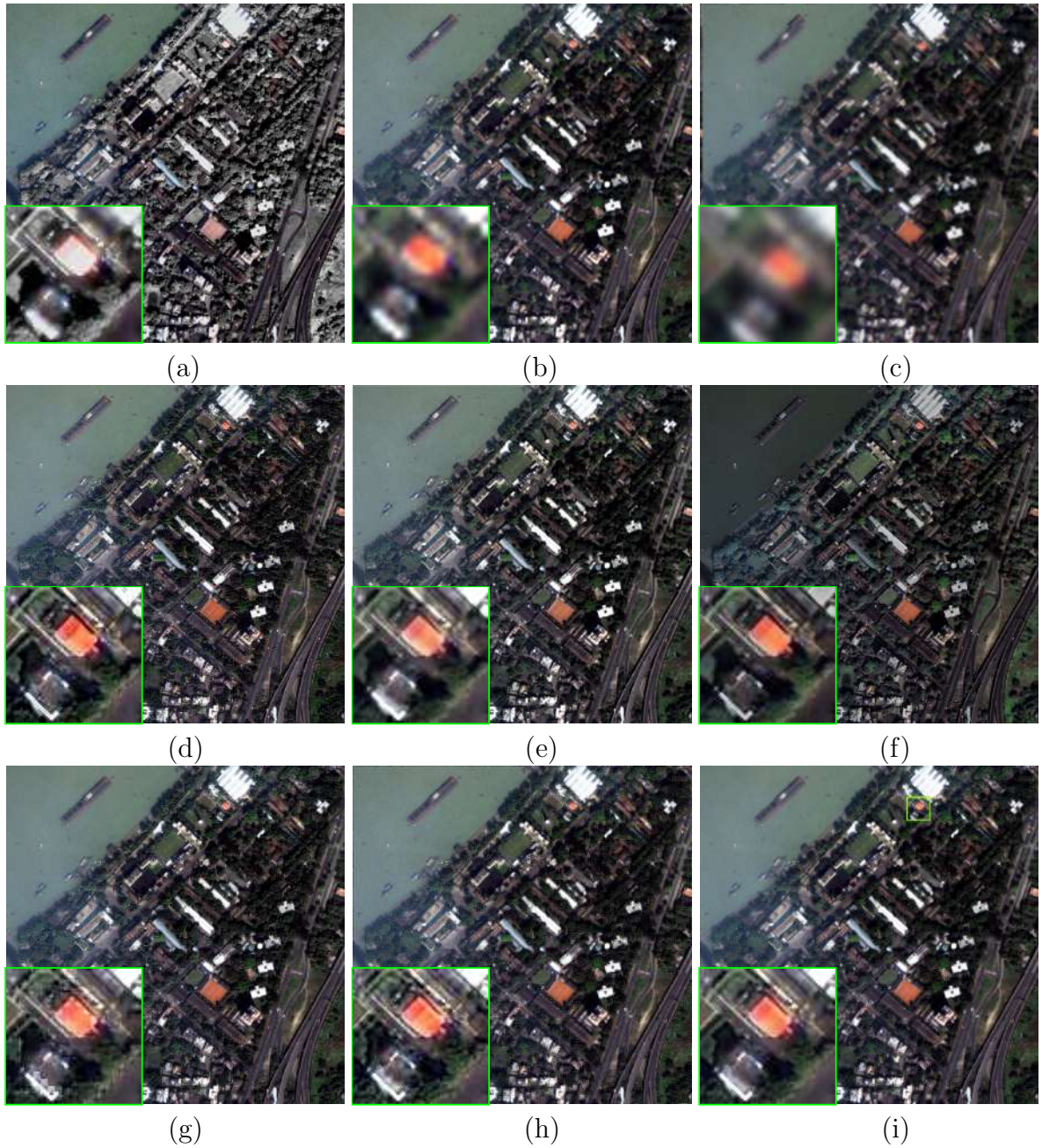


Figure 4.15: MS fusion results for un-degraded dataset consisting of urban area captured using Quickbird satellite shown as color composite of bands-3, 2 and 1 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [86], (c) approach in [103], (d) adaptive IHS approach [37], (e) AWLP [60], (f) Brovey method [22], (g) Wavelet based initial estimate, (h) proposed approach using CT and (i) proposed approach using NSCT. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

caused due to this. In order to check same we compute the spectral distortion measures on edge pixels. It is of interest to mention here that if a sudden change in color occurs over a region it also results in spatial intensity change in the fused image [177]. So the

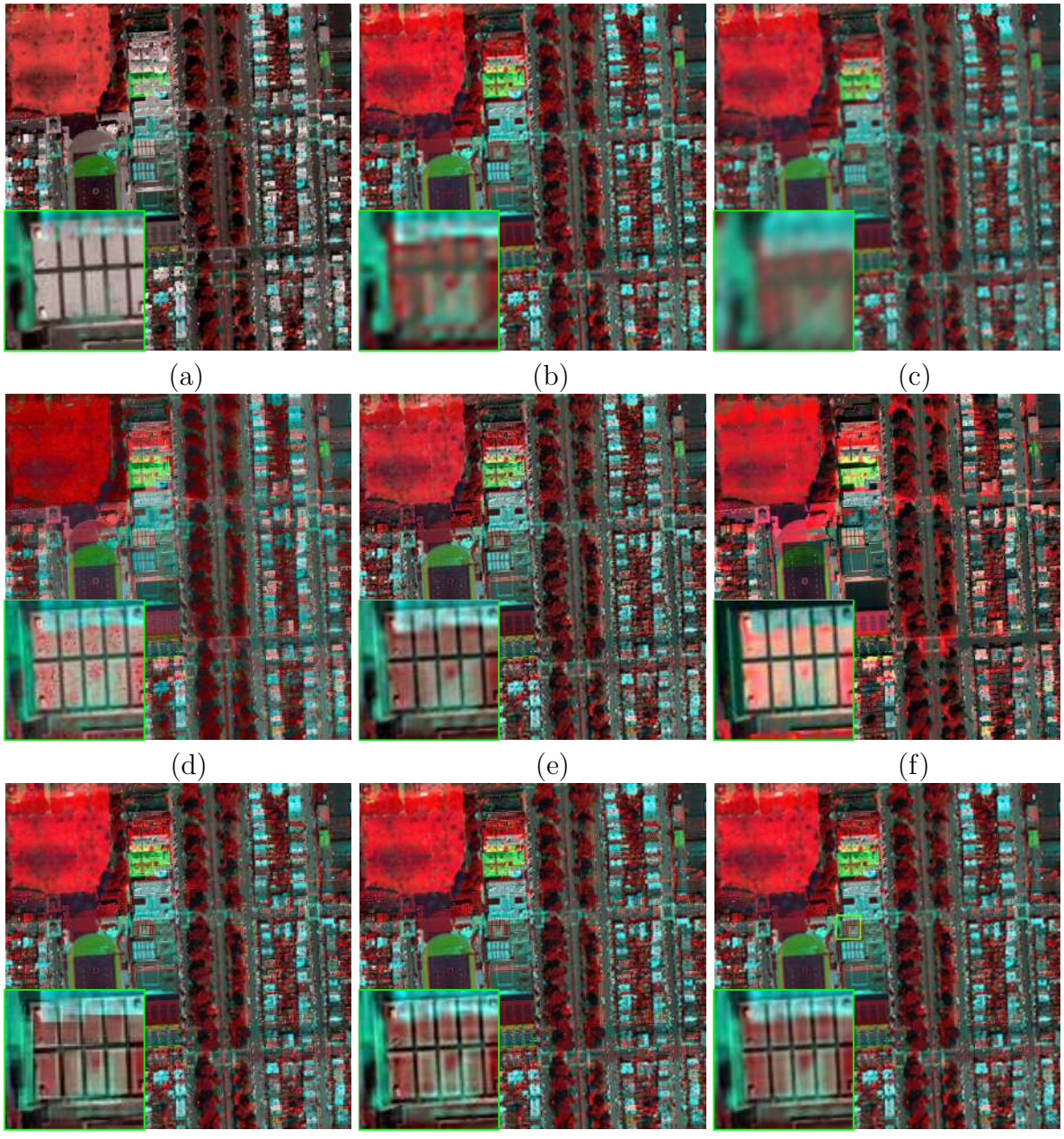


Figure 4.16: MS fusion results for un-degraded dataset consisting of urban area captured using Worldview-2 satellite shown as color composite of bands-7, 5 and 3 ($q = 4$). Fused images obtained using (a) TFT based approach [175], (b) approach in [86], (c) approach in [103], (d) adaptive IHS approach [37], (e) AWLP [60], (f) Brovey method [22], (g) Wavelet based initial estimate, (h) proposed approach using CT and (i) proposed approach using NSCT. The magnified image for a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

quantitative measures such as CC and D_λ computed on the fused image pixels are useful in evaluating the performance of the fusion method over the edge regions. Ideally the values of both CC and D_λ must be 1 and 0, respectively which indicates that the fused

edges are same as the original image.

Both CC and D_λ measures are computed over the edge pixels of the selected regions of Fig. 4.6-4.12 and they are displayed in Table 4.11 for degraded case. In Fig. 4.6 & 4.7, we observe that the zoomed-in portions of selected regions of non-urban and urban area acquired using Ikonos-2 satellite have the significant spectral details. One can observe that the proposed method has lower values of CC and D_λ when compared to other methods. The spectral distortion measures computed for the results shown in Fig. 4.8-4.10 correspond to the results of Quickbird satellite. In this also, the values of the spectral distortion measures are better for the proposed method. It is important to note that although Fig. 4.10 has the spectral distortion in all the methods, the proposed method still works better for this dataset. In the images of Worldview-2 satellite we have chosen a region which has significant spectral change (see Fig. 4.11 & 4.12). Similar to earlier experiments of Ikonos-2 and Quickbird, here also the proposed method has lowest spectral distortion which is clearly indicated by the values shown in Table 4.11. One can see from the results displayed in Fig. 4.12(b) that the fusion result of [87] appears highly distorted. This is also reflected in the values of CC and D_λ in Table 4.11 with a high value of D_λ and significantly low values for CC.

Similar to the degraded dataset experiments, we have also performed the quantitative analysis on edge pixels for the results of un-degraded datasets. In this case since the original MS image is not available we compute D_λ only since it does not require the original image. These values of D_λ are displayed in Table 4.12 for the results shown in Fig. 4.14-4.16. From the zoomed-in images displayed in these figures one can visually observe that the proposed fusion method has better preservation of the spectral details. This observation is also reflected by having lowest values of D_λ for the proposed method when compared to the other fusion techniques. This quantitative analysis performed on degraded and un-degraded dataset also clarifies that the fusion techniques based on the multi-resolution analysis (MRA) are proven to be effective against the spectral distortions when compared to other fusion methods which is also valid for the proposed method since it is based on CT/NSCT.

Table 4.11: Spectral distortion measures computed on edge pixels of selected regions shown in Fig. 4.6 to Fig. 4.12 for degraded datasets. Here, boldface indicate values closer to the ideal.

Dataset	Measure	Band	TFT method [175]	Approach in [87]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Proposed using CT	Proposed using NSCT
Ikonos-2 (Fig. 4.6)	CC	Band 1	0.8170	0.8464	0.8084	0.8173	0.9164	0.9253	0.9236	**0.9372**
		Band 2	0.8579	0.8200	0.8632	0.8516	0.9245	0.9372	0.9406	**0.9453**
		Band 3	0.8878	0.8756	0.8989	0.8425	0.9515	0.9626	0.9570	**0.9679**
	D_λ	Band 1-3	0.2549	0.2107	0.2204	0.2088	0.1484	0.1366	0.1120	**0.0983**
Ikonos-2 (Fig. 4.7)	CC	Band 2	0.9351	0.8445	0.7952	0.8109	0.9556	0.9559	0.9660	**0.9699**
		Band 3	0.9514	0.8846	0.7851	0.7906	0.9553	0.9610	0.9676	**0.9717**
		Band 4	0.8602	0.8601	0.7254	0.7299	0.9252	0.9269	0.9162	**0.9343**
	D_λ	Band 2-4	0.2813	0.2513	0.2389	0.2647	0.1976	0.1890	0.1941	**0.1856**
Quickbird (Fig. 4.8)	CC	Band 1	0.7032	0.7181	0.8966	0.8575	0.9112	0.9116	**0.9271**	0.9251
		Band 2	0.7773	0.7472	0.8760	0.8663	0.9368	0.9284	0.9339	**0.9416**
		Band 3	0.7303	0.7486	0.8861	0.8526	0.9345	0.9303	0.9410	**0.9471**
	D_λ	Band 1-3	0.2459	0.2366	0.2504	0.2524	0.1582	0.1486	0.1363	**0.1238**
Quickbird (Fig. 4.9)	CC	Band 2	0.6928	0.6865	0.8592	0.8471	0.9352	0.9280	0.9373	**0.9469**
		Band 3	0.6740	0.6608	0.8595	0.8419	0.9020	0.9111	0.9415	**0.9458**
		Band 4	0.6042	0.5516	0.7061	0.8076	0.8915	0.9011	0.9134	**0.9180**
	D_λ	Band 2-4	0.5248	0.3479	0.2655	0.1967	0.1261	0.1344	0.1243	**0.1121**
Quickbird (Fig. 4.10)	CC	Band 1	0.5322	0.5856	0.8900	0.8768	0.9203	0.9103	0.9165	**0.9308**
		Band 2	0.5670	0.5281	0.9077	0.8932	0.9288	0.9335	0.9356	**0.9413**
		Band 3	0.4647	0.4911	0.9126	0.8994	0.9370	0.9385	0.9255	**0.9397**
	D_λ	Band 1-3	0.2624	0.2154	0.2290	0.2116	0.1620	0.1670	0.1546	**0.1489**
Worldview-2 (Fig. 4.11)	CC	Band 2	0.8641	0.8318	0.8038	0.8048	0.9218	0.9114	0.9331	**0.9439**
		Band 3	0.9146	0.8857	0.7988	0.7838	0.9412	0.9333	0.9475	**0.9522**
		Band 5	0.8745	0.8919	0.7870	0.7894	0.9346	0.9270	0.9387	**0.9393**
	D_λ	Band 2,3,5	0.2198	0.2203	0.2045	0.2153	0.1533	0.1538	0.1499	**0.1415**
Worldview-2 (Fig. 4.12)	CC	Band 3	0.8914	0.8228	0.8629	0.8973	0.9511	0.9489	0.9553	**0.9607**
		Band 5	0.7820	0.7658	0.8593	0.8657	0.9433	0.9323	0.9497	**0.9546**
		Band 7	0.8527	0.6945	0.8810	0.8820	0.9286	0.9280	0.9325	**0.9360**
	D_λ	Band 3,5,7	0.2151	0.5996	0.2436	0.2525	0.1301	0.1346	0.1206	**0.1145**

Table 4.12: Spectral distortion measure computed on edge pixels of selected regions results shown in Fig. 4.14 to Fig. 4.16 for un-degraded datasets. Here, boldface indicate values closer to the ideal.

Dataset	Measure	Band	TFT method [175]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Brovey method [22]	Proposed using CT	Proposed using NSCT
Ikonos-2 (Fig. 4.14)	D_λ	Band 1-3	0.2384	0.2263	0.2408	0.1596	0.1588	0.1657	0.1516	**0.1415**
Quickbird (Fig. 4.15)	D_λ	Band 1-3	0.2237	0.1928	0.1943	0.1365	0.1358	0.1309	0.1246	**0.1224**
Worldview-2 (Fig. 4.16)	D_λ	Band 3,5,7	0.2451	0.2667	0.2637	0.1645	0.1595	0.1895	0.1419	**0.1394**

Table 4.13: Average computation time involved for different fusion approaches

Experiment on	Average computation time to obtain single fused band (in *seconds*)							
	TFT method [175]	Approach in [86]	Approach in [103]	Adaptive IHS [37]	AWLP method [60]	Brovey method [22]	Proposed using CT	Proposed using NSCT
Degraded dataset	0.5075	56.4426	1.9721	0.7892	0.3263	0.3464	1.3934	3.5403

4.7.8 Computational Time

Finally, here are the points that correspond to computational expenses of different fusion approaches. All these methods have been implemented in Matlab 7.6 installed on Intel core 2 Duo processor with 2.4 GHz, 4 GB memory. In Table 5.5 we display the average computation time to obtain single fused image for degraded dataset using different fusion methods. Here, TFT, Adaptive IHS, AWLP and Brovey methods have less computational time. However, the fusion method in [86] takes more time due the computation of PCA and cross correlation coefficient. The average time to run the proposed CT based algorithm is about 1.4 seconds while it is slightly more (3.5 seconds) for NSCT based approach. This computation time is bit more compared to the non-regularizing techniques such as TFT, AWLP and Brovey. However due to the regularization frame work of the proposed approach, the advantages in terms of preservation of edge and color details are possible at the cost of slight increment in the computational time. In the proposed method, much of time is needed in obtaining contourlet decomposition at different levels as it is implemented in Matlab. But it can be reduced drastically if algorithm is implemented using C language and executed on a machine with graphical processing unit. Note that our approach do not add to time complexity when compared to the method of fusion involving CT/NSCT [86], since the proposed approach does not involve training database. Even we avoid using computationally expensive methods such as simulated annealing for preserving the edges. Much of the computational burden is reduced due to the use of simple optimization method and the use of estimated MRF parameter while optimizing the cost function. Use of estimated MRF parameter avoids trial & error based selection of the parameter during optimization and hence reduces the computational burden.

4.8 Conclusion

We have presented a new technique for multi-resolution image fusion using contourlet based learning and MRF prior. In the proposed method we first obtain the initial high resolution MS image by the available Pan image and the test MS image. Since the initial estimate has high spatial and spectral resolutions it is used to obtain the degradation between fused MS and test MS image where the blur is assumed to be a non-identity matrix. We cast the fusion problem in restoration framework and obtain the final solution

by using regularization framework. The final cost function is obtained using the MAP-MRF approach where MRF smoothness prior is used to regularize the solution. The edge details in the final fused image are obtained by applying Canny edge detector on the initial estimate. This gives us the advantage of preservation of edges in the final solution without using discontinuity preserving prior. The MRF parameter is also estimated using the initial estimate image which is used during optimization. The potential of the proposed fusion method is demonstrated by conducting experiments on the datasets acquired using various satellites. These datasets have the pixels corresponding to different regions of ground scene. Along with the qualitative analysis, we also show the quantitative performance by computing various measures. Since the final solution is obtained using Canny edge detector and smoothness MRF prior, the proposed method not only recovers the finer details with minimum spectral distortion it also avoids the use of computationally expensive optimization technique. The perceptual and the quantitative analysis show that the proposed technique yields better solution when compared to the state of the art approaches. In this chapter the initial estimate was obtained using Pan data. In the next chapter, we derive the initial estimate using the LR MS image itself. We also propose the use of a new prior based on the Gabor filter.

Chapter 5

Regularized Pan-sharpening based on Self-similarity and Gabor Prior

We address the fusion problem based on the concept of self similarity and a Gabor prior. In chapter 4, degradation matrix entries were estimated by modeling the relationship between the Pan derived initial estimate of the fused MS image and the LR MS image which may be inaccurate as the estimate depends on the low spectral resolution Pan data. If the initial fused estimate is derived using the available LR MS image only, then the transformation between the estimated high resolution fused image (initial estimate) and the LR MS image would be more accurate. This makes the estimated degradation matrix to better represent the aliasing. In that case we are required to obtain the initial estimate using only the available LR MS image as the true fused image is not available. Here, we use the property of natural images that the probability of availability of redundant information in the image and it's downsampled versions is high [178]. We exploit this self similarity concept along with CS theory in order to obtain the initial estimate of fused image. Finally, in order to preserve the spatial details and obtain a better estimate of fused image, we solve the pan-sharpening or multi-resolution image fusion problem in a regularization framework by making use of a new prior called Gabor prior.

5.1 Related Work

Li and Yang [129] have applied the theory of compressed sensing (CS) to obtain fusion for remotely sensed images. In this method, a dictionary is prepared from the sample

images having high spatial resolution. They obtain fused image as the linear combination of HR patches available in dictionary. The performance of this method depends on the availability of the high resolution MS images that have spectral components similar to that of the test image. Considering this limitation, the authors in [130, 131] propose to construct an overcomplete joint dictionary from available MS and Pan images and extract the most relevant spectral and spatial information using the l_1 minimization. They consider the MS image as the decimation of the original high resolution MS image and take the Pan as the linear combination of all bands of the high resolution MS images. On a similar line the authors in [132] use a dictionary constructed using the Pan image and its downsampled LR version and obtain the fusion by exploring the sparse representation of HR/LR multi-spectral image patches. In [133], authors used the CS framework with dictionary patches learned from the initial high resolution MS image obtained using AWLP method [60]. They obtain the final fused image using a trained dictionary obtained from K-singular value decomposition (K-SVD). Authors in [134] create the over-complete dictionary with basis functions of different transforms such as discrete cosine transform (DCT), wavelets, curvelets and ridgelets and the best bases for MS and Pan images are obtained with convex optimization. The final fused MS image is obtained by merging best bases coefficients of Pan and corresponding MS band as per values of the local information parameter.

The draw back of all these CS based fusion approaches is that the dictionary is constructed either using a database or by the use of MS and Pan images. The use of database requires the availability of high resolution MS images having characteristics similar to the test image and the use of Pan data suffers from the disadvantage of using lower spectral resolution images for dictionary construction. In the present work, we propose a model based pan-sharpening method using a new concept called self-similarity and a new prior referred to as Gabor prior. The self-similarity is based on the concept of the probability of occurrence of same or similar patches within and across the scale of image [178]. Literature shows that on an average 23% best match patches (exactly same patches) are available across the image for a downsampling factor (q) of 2 [178]. After obtaining the similar patches across the scale of LR MS image and obtaining their corresponding HR patches, the HR patches for which matches are not found are obtained by using the CS theory that uses the dictionaries constructed from the already found LR-

HR patches. This constitutes the initial pan-sharpened image estimate in our approach. The same is used in estimating the degradation between LR and unknown pan-sharpened images. A regularization frame work is then used to obtain the final solution in which the Gabor prior is used to extract the bandpass details from the Pan image to embed it into the final pan-sharpened image. Our regularization terms include the Gabor prior as well as the MRF smoothness prior that maintains spatial smoothness in terms of patch continuity in the final pan-sharpened image. The MRF parameter is estimated using the initial estimate. The final cost function is convex which allows us to use gradient based optimization technique for minimization. It is worth to mention that in this chapter we use the concept of the self-similarity in order to find the best matched patches in the image itself which is different from the term"learning used in the pattern recognition community or machine learning community. Note that learning in the pattern recognition corresponds to estimating of the parameters by training the network for a given training data set. Such training and the parameter estimation are not used in our work. In our work the meaning of learning is to get the high resolution patch by using the LR image patch matching. We use the LR MS image patch matching to learn the HR patch. In this process the complexity is much less when compared learning as used in pattern recognition community [179].

The proposed approach has the following advantages which makes it different from the other CS and regularization based pan-sharpening approaches.

- Instead of using Pan and MS images, only the LR MS image is used to obtain the initial estimate which results in better degradation estimation between the LR and HR images.

- Spatial details in the final pan-sharpened image are better preserved due to the use of Gabor prior which not only extracts the high frequency edge details from Pan image, but the features at different bandpass frequency bands are also extracted from the Pan image.

- Use of homogeneous MRF prior keeps the patch continuity and hence spatial details in the final pan-sharpened image are better preserved.

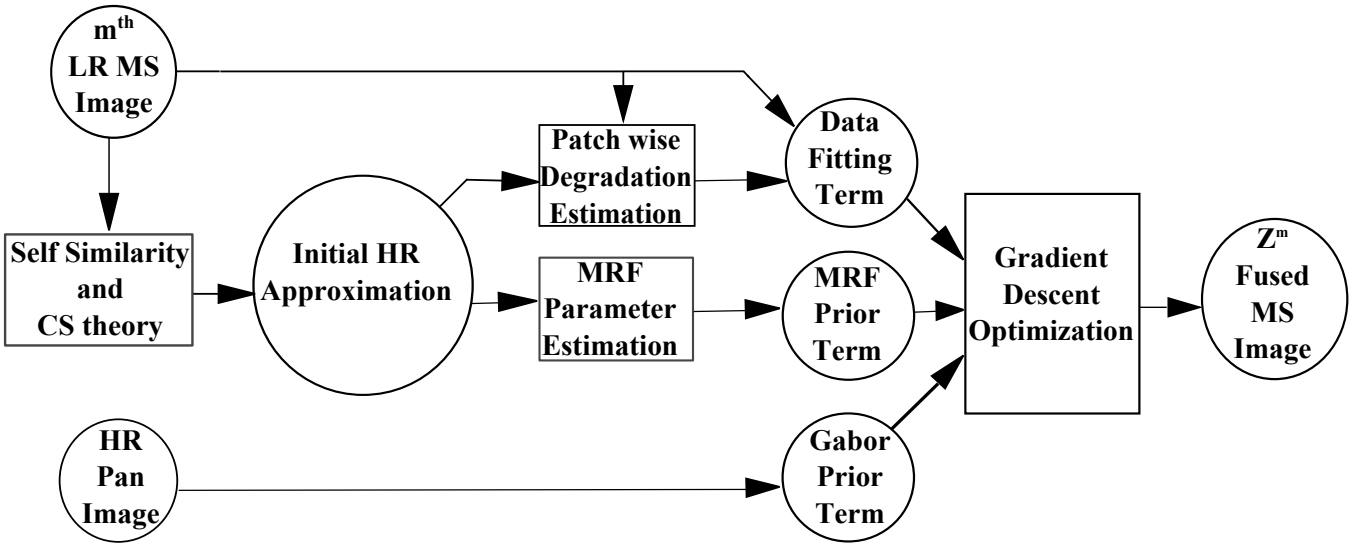


Figure 5.1: Block schematic of the proposed pan-sharpening method for fusing an m^{th} MS and the Pan images. Here LR and HR correspond to low resolution and high resolution, respectively.

5.2 Block Schematic of the Proposed Method

The block schematic to obtain the pan-sharpened MS image Z^m for the m^{th} low resolution MS is shown in Fig. 5.1. As shown in this figure, first the initial HR approximation to pan-sharpened image (initial estimate) is obtained using the LR MS image by making use of the concept of self-similarity and the theory of CS. Assuming that the initial estimate is close to the final pan-sharpened image we then obtain the patch wise degradation between the pan-sharpened image and LR MS image by using the initial estimate. This degradation matrix estimated for each patch are then used in the data fitting term of the final cost function. With an MAP framework, the final cost function consists of prior terms along with the data fitting term. Here, we use Gabor prior that extracts the bandpass details from the Pan image. We also use MRF prior to maintain continuity between patches. The MRF parameter is estimated using the initial estimate. Gradient descent optimization technique is then used to obtain the final pan-sharpened image.

5.3 Initial HR Approximation

In the proposed method we obtain an initial HR approximation (initial estimate) which is used in obtaining the degradation between LR and HR patches and to obtain the MRF parameter. The initial approximation is obtained using the LR MS image itself

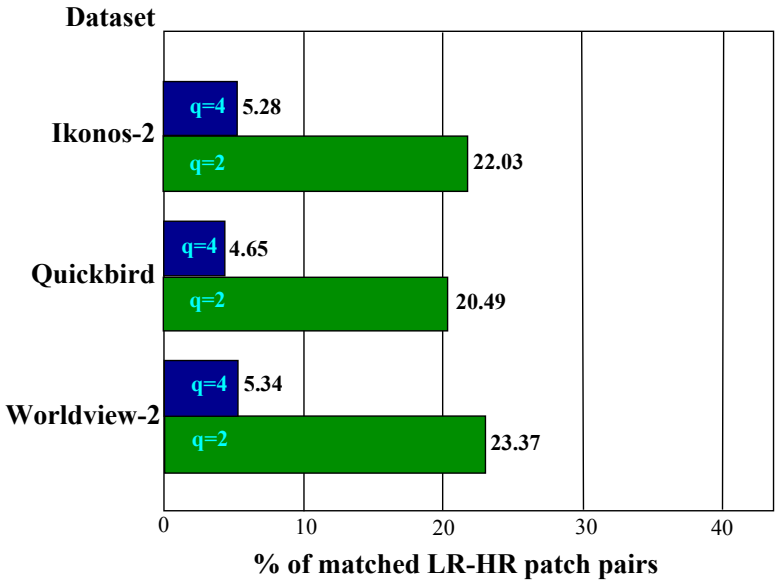


Figure 5.2: The statistics for number of matched LR-HR patch pairs using the images of different satellites for $q = 2$ and 4.

by using the concept of self-similarity and compressive sensing. This concept is based on the observation that the same or similar spatial information is repeated within or across the coarser resolution of the images [178]. The recurrence of patches in the same image or its coarser resolution is called *self-similarity*. This means that depending on the resolution difference, the HR patches are available in the given LR MS image itself. In [178] authors used this observation to obtain the super-resolution (SR) of LR image. Here, they first construct the LR cascade consisting of different coarser resolutions of the given image. They search for the similar patches available in the LR cascade and use them to estimate the SR of the given image. For the unmatched LR patch, interpolation is used to obtain the corresponding HR patch in the final SR image. Due to the interpolation of unmatched LR patches, this approach [178] suffers from the spatial distortions. In the proposed method we use the concept of self-similarity to find the same or similar patches in the coarser resolution of the given LR MS image. Unlike [178], matched LR-HR patches are found using only the coarser resolution of the given LR MS image, thus avoiding construction of the LR cascade. The unmatched patches are estimated using the compressive sensing (CS) framework which are interpolated in [178].

In order to investigate the possible number of matched LR-HR pairs in different types of satellite images we carried out the statistics for downsampling factor (q) of 2 and 4. To do this, we conducted an experiment on different images captured using various

satellites such as Ikonos-2, Quickbird and Worldview-2. Here, we use 50 MS images each of size 256×256 for every satellite. The coarser resolution images for $q = 2$ and $q = 4$ are obtained using Gaussian blurring and downsampling operation. In each band of MS image, we consider a LR patch of size $w \times w$ (here $w = 5$) and this patch is searched in its coarser resolution image. The *approximate nearest neighborhood* (ANN) [180] is used to find the same patch using the sum of squared difference (SSD) distance criteria resulting in LR-LR match. The threshold value for SSD is set to be 0 for all the datasets. Since the resolution difference is known, the corresponding $qw \times qw$ matched HR patch can be easily located in the LR MS band and this results in LR-HR matched pair. In order to explain this consider the resolution difference of 2. We then have two images of the MS band having the size of 256×256 and 128×128 pixels, respectively. If a $w \times w$ patch in MS image of size 256×256 is matched (similar) to a patch centered at a location (l, m) in its coarser resolution of size 128×128, then the patch of size $2w \times 2w$ centered at $(2l, 2m)$ in the MS image of size 256×256 corresponds to matched HR patch giving a matched LR-HR patch pair. We repeat this procedure for all the non overlapping patches and obtain the matched LR-HR patch pairs in the given band. We then calculate the percentage of LR-HR matched patch pairs by using the total number of available and matched LR-HR patch pairs for that band. This experiment is performed for all the 50 images of that particular satellite to obtain the different number of matched LR-HR patch pairs and we compute the average value in order to display the statistics. This statistics is presented in Fig. 5.2 for $q = 2$ as well as for $q = 4$. For example, for Ikonos-2 images, out of total patches of 63504 we obtained $22.03\%(\approx 13990)$ and $5.28\%(\approx 3354)$ matched LR-HR patch pairs for $q = 2$ and 4, respectively. It is worth to mention that since there are more patches marched for $q = 2$ than for $q = 4$, we use the coarser resolution image of $q = 2$ only while finding the matches. Hence in order to obtain the initial estimates for our MS bands where q is 4, we repeat the procedure for $q = 2$. We can see that on an average the number of matched LR-HR patch pairs for $q = 2$ and 4 are 22% and 5%, respectively.

In order to reconstruct the initial HR image, we need the matched HR patch for every LR patch. Since the unmatched LR-LR patches do not result in HR patches they have to be estimated. We use CS framework for this. The CS theory described in [181–183]

demonstrates the recovery of sparse vector. A vector $\mathbf{x}$ in $R^{\aleph}$ can be represented as

$$\mathbf{x} = \Psi\mathbf{v}, \tag{5.1}$$

where Ψ is an $\aleph \times \aleph$ basis matrix and $\mathbf{v}$ is $\aleph \times 1$ column vector of weighting coefficients. If the vector $\mathbf{v}$ has L nonzero elements i.e., $\mathbf{x}$ is L sparse in Ψ domain, then only L basis vectors are required to form to the vector. In CS theory a measurement vector $\mathbf{u}$ is generated from $\mathbf{x}$ as

$$\mathbf{u} = \Phi\mathbf{x}, \tag{5.2}$$

where Φ is a measurement matrix of size $\kappa \times \aleph$ with $\kappa < \aleph$. Substituting from equation (5.1) we get,

$$\mathbf{u} = \Phi\Psi\mathbf{v} = \Theta\mathbf{v}, \tag{5.3}$$

where Θ is an $\kappa \times \aleph$ matrix and allows reconstruction of $\mathbf{v}$ (and hence $\mathbf{x}$). Note that $\mathbf{v}$ has dimension of $\aleph \times 1$, but has only L nonzero coefficients. The recovery algorithm aims to find the sparse $\mathbf{v}$ subject to $\mathbf{u} = \Theta\mathbf{v}$. This can be solved using ℓ_1 minimization by posing the problem as [181],

$$min \parallel \mathbf{v} \parallel_1 \quad \text{subject to } \mathbf{u} = \Theta\mathbf{v}, \tag{5.4}$$

where $\| \cdot \|_1$ represents ℓ_1 minimization. Equation (5.4) recovers exactly L sparse components of the $\mathbf{v}$ and can be converted to linear programming (LP) of basis pursuit [181,182] with computational complexity of $O(\aleph^3)$.

In Fig. 5.3, we illustrate this concept of self-similarity. Fig. 5.3(a) displays the LR MS image (I_0) of Quickbird satellite shown with color composition of bands 3, 2 and 1. The coarser resolution (I_{-1}) for downsampling factor (q) 2 of the same image is displayed in Fig. 5.3(b). Consider a patch P_0 of size $w \times w$ which is marked with green border in I_0. We can observe in the Fig. 5.3(b) that the same or similar patch is also available in its coarser image I_{-1} as P_{-1} which is also marked with green border. Since these two patches form LR-LR match, parent HR patch (R_0) of P_{-1} (marked with blue color border in Fig. 5.3(a)) of size $qw \times qw$ is available in the LR MS image itself forming the corresponding LR-HR matched pair i.e., P_0 and R_0 constitute LR-HR matched pair. Note that R_0 represents the matched HR patch for P_0. As already discussed the ANN

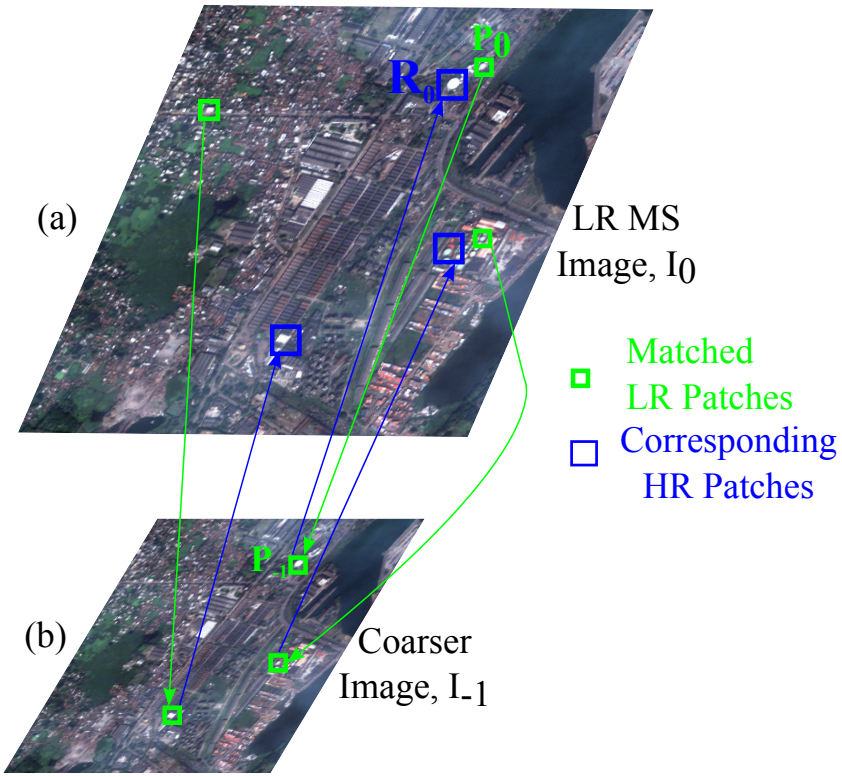


Figure 5.3: Patch recurrence for LR MS image into its coarser resolution image. (a) LR MS image, I_0 and (b) coarser resolution of (a), I_{-1}. Here, patches shown with green border are the matched LR patch pairs and corresponding to these matched pairs HR patches are shown with blue border in (a).

can be used to find the patch match in the LR and its coarser versions and hence its HR patch can be found from the LR image. To find the initial approximation, all the available LR-HR matched pairs are found by considering non overlapping patches in the LR MS image. We then construct two separate dictionaries using these LR and HR pairs. Let N_{best} be the number of matched patch pairs in the given LR MS image and corresponding to this D_{LR} and D_{HR} be the LR and HR dictionaries of size $w^2 \times N_{best}$ and $q^2w^2 \times N_{best}$, respectively. Here, $w^2 \times 1$ and $q^2w^2 \times 1$ correspond to size of lexicographically ordered LR and HR patch pixels, respectively. The unmatched HR patches are obtained using CS framework as depicted in Fig. 5.4. Given an LR patch for which no HR match is available and the LR dictionary D_{LR}, we can estimate the sparse coefficient vector $\mathbf{v}$ using l_1 minimization. This is represented in Fig. 5.4(a). Assuming that both the LR and HR patches have the same sparseness, the estimated sparse vector can be utilized to compute the unknown HR patch as displayed in Fig. 5.4(b). Repeating this for all those LR patches for which no HR matches are available, we obtain LR-HR matched

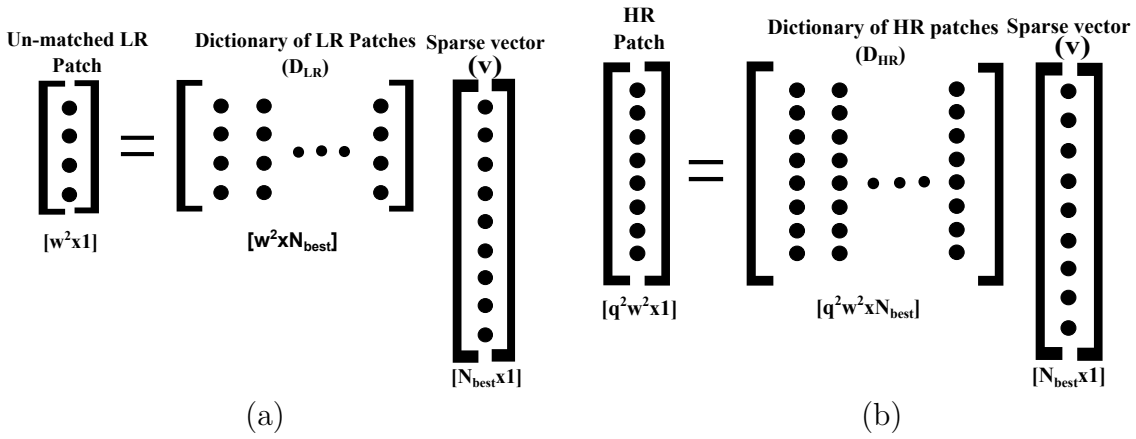


Figure 5.4: (a) CS framework to obtain the spare coefficient vector $\mathbf{v}$ for unmatched LR patch. (b) The corresponding HR patch estimated using dictionary of HR patches (D_{HR}) and the spare vector $\mathbf{v}$ in (a). Here the CS framework is depicted for $q = 2$.

pairs for the complete image. Placing these HR patches at their corresponding locations results in the initial approximation image. It is important to note that as the number of matched LR patches are significant, the quality of initial estimate would be better since we compute the unmatched LR-HR patches using the linear sparse combination of matched LR-HR patches. The procedure is repeated to obtain initial estimate separately for each LR MS image.

5.4 Image Formation Model and Degradation Estimation

In the proposed method, the LR MS image is modeled as blurred and noisy version of the true MS image. This linear image formation model for the i^{th} patch of m^{th} LR MS image of size $N_1 \times N_2$ can be written as,

$$\mathbf{y}_i^m = A_i^m \mathbf{z}_i^m + \mathbf{n}_i^m, \quad m = 1, 2, \ldots, l, \quad i = 1, 2, \ldots, N. \tag{5.5}$$

In equation (5.5), l represents the number of bands in LR MS image and N is the total number of patches each of size $w \times w$ in the LR MS image Y. The $\mathbf{y}_i^m$ and $\mathbf{z}_i^m$ are the lexicographically ordered vectors of i^{th} patch for m^{th} LR MS and unknown pan-sharpened MS images, respectively. The $\mathbf{n}_i^m$ is the corresponding noise vector which is assumed to be as *independent and identical distribution (iid)* Gaussian with zero mean and unit variance.

The degradation matrix for i^{th} patch of m^{th} LR MS image is given by A_i^m. However, this is unknown since we do not have the pan-sharpened MS image. Since, we already know the approximation to pan-sharpened image (initial estimate), we use it for estimating A_i^m. One may write A_i^m as,

$$A_i^m = D_i^m H_i^m, \quad m = 1, 2, \ldots l. \tag{5.6}$$

where D_i^m is the decimation matrix of size $w^2 \times q^2 w^2$ and H_i^m is the blur matrix of size $q^2 w^2 \times q^2 w^2$ which is assumed to be space invariant. By considering the LR pixel as linear combination of q^2 HR pixels with appropriate weights, we choose D_i^m as,

$$D_i^m = \begin{pmatrix} a_{i1}^m \; a_{i2}^m \ldots \; a_{iq^2}^m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & a_{i1}^m \; a_{i2}^m \ldots \; a_{iq^2}^m & \mathbf{0} \\ . & . & . \\ \mathbf{0} & \mathbf{0} & a_{i1}^m \; a_{i2}^m \ldots \; a_{iq^2}^m \end{pmatrix}. \tag{5.7}$$

Also, the space invariant blur matrix H_i^m has the form

$$H_i^m = \begin{pmatrix} H_{i0}^m & H_{iqM^2-1}^m & H_{iqM^2-2}^m \ldots & H_{i1}^m \\ H_{i1}^m & H_{i0}^m & H_{iqM^2-1}^m \ldots & H_{i2}^m \\ . & . & . \quad \ldots & . \\ H_{iqM^2-1}^m & H_{iqM^2-2}^m & H_{iqM^2-3}^m \ldots & H_{i0}^m \end{pmatrix}, \tag{5.8}$$

where $m = 1, 2, \ldots, l$ and each H_{ij} can be written as,

$$H_{ij}^m = \begin{pmatrix} h_{ij,0}^m & h_{ij,q-1}^m & h_{ij,q-2}^m \ldots & h_{ij,1}^m \\ h_{ij,1}^m & h_{ij,0}^m & h_{ij,q-1}^m \ldots & h_{ij,2}^m \\ . & . & . \quad \ldots & . \\ h_{ij,q-1}^m & h_{ij,q-2}^m & h_{ij,q-3}^m \ldots & h_{ij,0}^m \end{pmatrix}. \tag{5.9}$$

Here $h_{.,.}$ are the values of the point spread function (PSF) for blur. Since we consider a space invariant blur, H_i^m is block circulant. The multiplication of D_i^m and H_i^m results in

the A_i^m matrix which is given by

$$A_i^m = \begin{pmatrix} A_{i1}^m & A_{i2}^m & \dots & A_{iq^2M^2-1}^m & A_{iq^2M^2}^m \\ A_{iq^2M^2-q+1}^m & \dots & A_{i1}^m & A_{i2}^m \dots & A_{iq^2M^2-q}^m \\ . & . & . & . & . \\ \dots & \dots & \dots A_{i1}^m & \dots & \dots \end{pmatrix}. \tag{5.10}$$

For $w = 2$ and $q = 2$ the A_i^m matrix has size of 4×16 and can be written as,

$$A_i^m = \begin{pmatrix} A_{i1}^m & A_{i2}^m & A_{i3}^m & A_{i4}^m & A_{i5}^m & A_{i6}^m & A_{i7}^m & A_{i8}^m & A_{i9}^m & A_{i10}^m & A_{i11}^m & A_{i12}^m & A_{i13}^m & A_{i14}^m & A_{i15}^m & A_{i16}^m \\ A_{i15}^m & A_{i16}^m & A_{i1}^m & A_{i2}^m & A_{i3}^m & A_{i4}^m & A_{i5}^m & A_{i6}^m & A_{i7}^m & A_{i8}^m & A_{i9}^m & A_{i10}^m & A_{i11}^m & A_{i12}^m & A_{i13}^m & A_{i14}^m \\ A_{i9}^m & A_{i10}^m & A_{i11}^m & A_{i12}^m & A_{i13}^m & A_{i14}^m & A_{i15}^m & A_{i16}^m & A_{i1}^m & A_{i2}^m & A_{i3}^m & A_{i4}^m & A_{i5}^m & A_{i6}^m & A_{i7}^m & A_{i8}^m \\ A_{i7}^m & A_{i8}^m & A_{i9}^m & A_{i10}^m & A_{i11}^m & A_{i12}^m & A_{i13}^m & A_{i14}^m & A_{i15}^m & A_{i16}^m & A_{i1}^m & A_{i2}^m & A_{i3}^m & A_{i4}^m & A_{i5}^m & A_{i6}^m \end{pmatrix}. \tag{5.11}$$

We estimate the entries in equation (5.10) using the i^{th} matched LR-HR patch pair. Inclusion of non-identity blur matrix leads to under determinant set of equations and hence it cannot be solved by using pseudo inverse. We use a minimum norm approach for estimating A_i^m for all i and m. Here degradation matrix consists of decimation and blur matrices which means that our model takes care of both aliasing due to downsampling as well as blur. It is worth to say that since the initial estimate is obtained by using LR MS image only avoiding the use of available Pan image, our approach results in better degradation matrix entries.

5.5 Regularization using Gabor and MRF Priors

As discussed in the previous section, the initial estimate is obtained by using LR MS image only without using the Pan image. Due to this the spectral details are intact in the initial estimate since MS image has high spectral resolution and it results in better degradation estimation. The factor that contributes for avoiding the spectral distortion in the initial estimate is due to the reconstruction of this initial HR estimate using the patches from the same MS observation. Pan-sharpening problem formulated in equation (5.5) is ill-posed which requires regularization in order to obtain a final solution. To regularize the solution we need suitable prior that preserves the spatial features at various frequencies. We also need to take care of the spatial dependencies in the final pan-sharpened image. In the literature many researcher have used various edge preserving

priors for pan-sharpening. However, these priors are suitable for preserving the high frequency content only. Also, they are computationally inefficient since they use costly optimization techniques in order to obtain the final solution. One may use Gabor filter which is a linear filter used to extract the band pass details present in the signal/image at various frequencies and orientations. Since the Pan image has high spatial resolution, these details can be extracted from the Pan image and can be used as prior in order to improve the final solution. Gabor filter has been used for texture representation and synthesis in the computer vision community [184,185]. The impulse response of $2D$ Gabor filter can be obtained by modulating a $2D$ sinusoidal using $2D$ Gaussian function as,

$$g(x, y, f, \theta, \sigma_x, \sigma_y) = e^{-\frac{1}{2}(\frac{x'^2}{\sigma_x^2}+\frac{y'^2}{\sigma_y^2})} \cos(2\pi f x'), \tag{5.12}$$

where (x, y) denotes the spatial coordinates and (x', y')=$(x\cos(\theta)+y\mathcal{SIN}(\theta), -x\mathcal{SIN}(\theta)+y\cos(\theta))$ represents the rotated (x, y) coordinates with angle θ. σ_x and σ_y are the variances in x and y directions for Gaussian kernel and f represents the frequency of the sinusoid. Gabor filter given in equation (5.12) describes the band pass filter with varying frequency f and orientation θ.

The use of Gabor prior in the proposed approach is illustrated in Fig. 5.5. It consists of the Gabor filter bank having different frequencies and orientations. The Pan and unknown pan-sharpened images (to be estimated) are passed through this filter bank. Since Gabor filter extracts the bandpass features, the outputs of the filter bank for Pan and unknown pan-sharpened correspond to the features extracted at different frequencies and orientations. In Fig. 5.5(a) and Fig. 5.5(b), we display the Gabor filter bank outputs when an i^{th} patch of the Pan and pan-sharpened images are applied as input to the filter bank. Here R represents the number of filters in the Gabor filter bank. Using this set up for prior, we seek for such a pan-sharpened image which when pass through Gabor filter bank has the band pass details similar to that the available in the Pan image (see Fig. 5.5(b)).

Along with the Gabor prior which takes care of preserving the spatial details in the final pan-sharpened image, we also use Markov random field (MRF) prior that preserves the patch continuity. Since our approach is based on patch based learning, the final output may result in blockiness which can be taken care by using MRF as smoothness

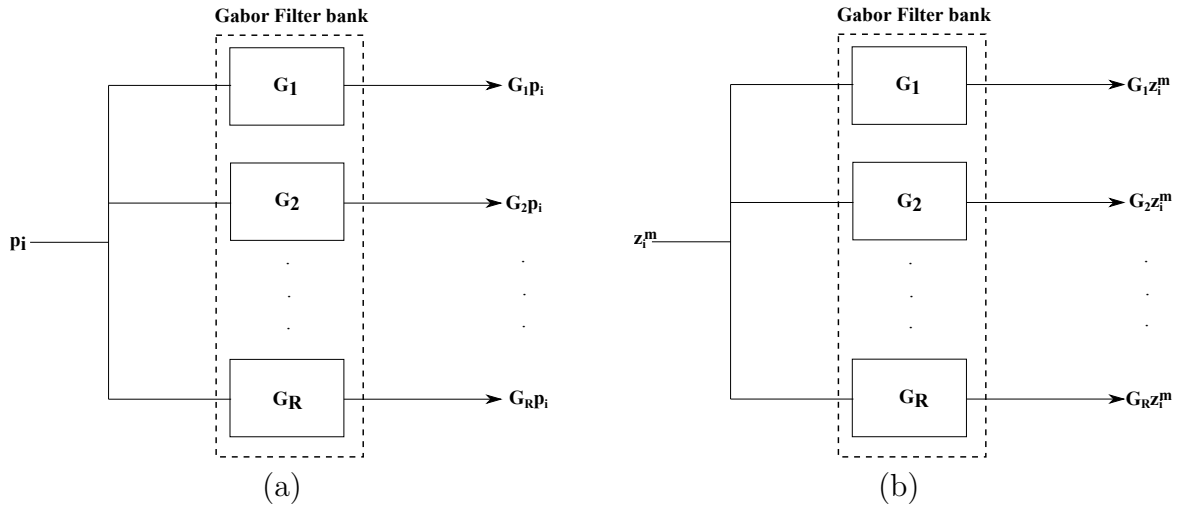


Figure 5.5: Illustration of Gabor prior. The outputs of Gabor filter bank when input is the i^{th} patch of (a) Pan image and (b) unknown pan-sharpened image.

prior. The MRF gives the spatial contextual dependency among the neighboring pixels and same can be used to preserve the continuity in the final pan-sharpened image. By using first order neighborhood, the MRF prior can be written as,

$$\sum_{c\in\mathcal{C}} V_c(\mathbf{z}^m) \quad = \quad \gamma^m \sum_{k=1}^{qN_1}\sum_{l=1}^{qN_2}[(Z^m_{k,l} - Z^m_{k,l-1})^2 + (Z^m_{k,l} - Z^m_{k-1,l})^2], \tag{5.13}$$

where γ^m represents the MRF parameter that indicates the penalty for departure from smoothness in $\mathbf{z}^m$. $\mathcal{C}$ is the set of all cliques. The parameter γ^m should be estimated using the pan-sharpened image. However, this is unknown and it has to be estimated. In our work, since the initial estimate is already available, we make use of the same to estimate γ^m. We use the method of maximum pseudo likelihood for estimating it [167]. Note that in equation (5.13), Z^m represents the entire image rather than a patch.

Using MAP-MRF regularization framework one can obtain the final cost function as

$$\hat{\mathbf{z}}^m = \underset{z^m}{argmin}\left[\sum_{i=1}^{N}\left(\parallel \mathbf{y}_i^m - A_i^m\mathbf{z}_i^m \parallel^2 + \sum_{j=1}^{R} \parallel G_j\mathbf{p}_i - G_j\mathbf{z}_i^m \parallel^2 \right) + \sum_{c\in\mathcal{C}} V_c(\mathbf{z}^m)\right], \tag{5.14}$$

where, A_i^m corresponds to the estimated degradation between the i^{th} low resolution MS image patch and the corresponding pan-sharpened patch. Here, N represents the total number of patches in the LR MS image which is also the number of patches in the pan-sharpened image. R denotes the total number of Gabor filters in the filter bank.

The second and third terms in equation (5.14) are Gabor and MRF smoothness priors, respectively. In the second term G_j represents a Gabor filter operator that corresponds to j^{th} filter impulse response. $\mathbf{p}_i$ denotes the i^{th} patch of the Pan image. This cost function is convex and hence can be minimized by using simple optimization technique such as gradient descent in order to obtain the final solution. It is worth to mention here that the minimization of first and second terms in equation (5.14) is performed patch wise, however the minimization of the last term in the same equation is carried out on the entire image.

5.6 Experimental Results

The effectiveness of the proposed pan-sharpening approach is verified by conducting the experiments on datasets of the different satellites such as Quickbird, Ikonos-2 and Worldview-2. The spatial resolution difference between Pan and MS images captured by using these satellites is 4. We conducted the experiments on degraded as well as on un-degraded (original) images. Details of the datasets used for experimentation are given in Table 3.1. The data for the experiments were prepared from the original MS and Pan images using cropping operation. The original MS and Pan images are cropped appropriately to obtain the images of size 256×256 and 1024×1024, respectively and these cropped images are used in all our experiments. These datasets are co-registered before conducting the experiment. It is worth to mention here that we use the datasets with the original radiometric resolution without doing any pre-processing such as contrast stretching. The performance of the proposed fusion method is compared with the other state of the art pan-sharpening techniques namely fast IHS (FIHS) [33] and adaptive IHS (AIHS) [37], AWLP [60], sparseFI [132] and Li *et. al* [129]. The quantitative analysis on the experiments of degraded dataset is performed by calculating the different measures such as correlation coefficient (CC) [29], erreur relative global adimensionnelle de synthse (ERGAS) [186], average quality index (Q_{AVG}) [187], root mean square error (RMSE) [29] and the spectral angle mapper (SAM) [188]. In the case of un-degraded experiments the quality with no reference (QNR) [158] is computed which does not require the reference image. The QNR [158] measure is a combination of the spectral and spatial distortions which are represented by D_λ and D_s, respectively.

Table 5.1: Gabor filter parameters.

Parameters	Ikonos-2	Quickbird	Worldview-2
Frequencies	0.025, 1.5	0.015, 1.15	0.035, 1.50
Orientations (*in degree*)	0, 45, 90, 145	0, 35, 95, 105	0, 25, 85, 145

5.6.1 Experimental Set-up

In order to perform the experiments on the degraded dataset we pass the original Pan and MS images through Gaussian filtering with different Nyquist cutoff frequencies and downsampling operation [189]. The downsampling factor (decimation) is chosen as $q = 4$ since the spatial resolution difference between MS and Pan images is 4. Thus the size of degraded MS and Pan images are 64×64 and 256×256, respectively. However, for un-degraded experiments they remain same as the original size of MS and Pan image. We use a patch size (w) of 5×5 while obtaining the initial estimate. In order to obtain the unmatched HR patches we use the ℓ_1 minimization using ℓ_1 *MAGIC* toolbox provided by Justin Romberg [190]. In the Gabor filter bank we use 8 filters having different frequencies and orientations. These parameters for the experimentations on various types of satellite images are listed in Table 5.1 which are chosen empirically based on the quality of the pan-sharpened image. We experimented with more number of filters in the bank, however the improvement in the final pan-sharpened images were not very significant. The step size for gradient descent optimization is kept as 0.01 for all the experiments.

5.6.2 Experimental Results on Degraded and Un-degraded Ikonos-2 Datasets

The first experiment is conducted on the degraded and un-degraded images of Ikonos-2 satellite. Details of the data captured using this satellite can be found in Table 3.1. The images in this dataset consists of the urban area of Mount Wellington nearer Hobart Tasmania and the pan-sharpening results of the same are displayed using the color composition of 4, 3 and 2 bands. The results obtained using different state of the art approaches in addition to the proposed approach are displayed in Fig. 5.6 and Fig. 5.7 for degraded and un-degraded images, respectively. The degraded MS and Pan images are displayed in Fig. 5.6(a) and Fig. 5.6(b), respectively and their originals i.e., the un-degraded images are displayed in Fig. 5.7(a) and Fig. 5.7(b), respectively. The pan-sharpened results of FIHS

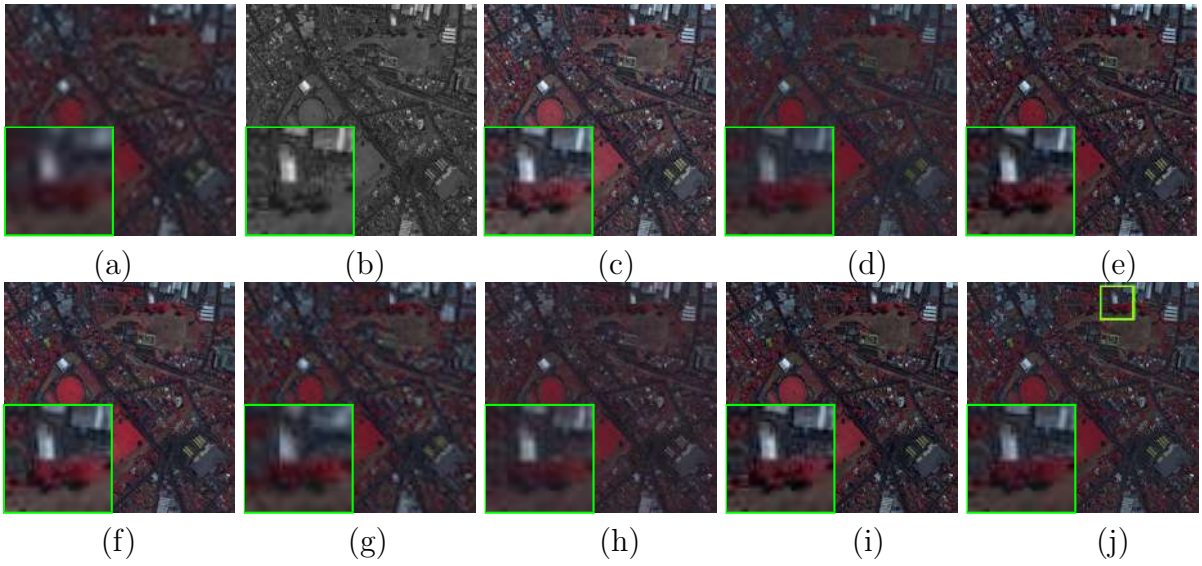


Figure 5.6: Results of pan-sharpening on degraded dataset of Ikonos-2 satellite consisting of urban area shown as color composite of bands-4, 3 and 2 ($q = 4$). (a) LR MS image of the size 64×64 upsampled to the size of Pan image. (b) Pan image of size 256×256. Pan-sharpened images obtained using (c) fast IHS (FIHS) approach [33], (d) AIHS [37], (e) AWLP [60], (f) Brovey method [22], (g) sparseFI [132], (h) Li *et. al* [129] and (i) proposed approach. (j) Original MS image. The magnified image of a small square region with a green border shown in (j) is displayed at the bottom left corner of all the images.

method [33] on degraded and original data are depicted in Fig. 5.6(c) and Fig. 5.7(c), respectively. In Fig. 5.6(d) and Fig. 5.7(d), we display the results of AIHS method [37]. The pan-sharpening results of AWLP [60] and Brovey [22] methods are displayed in Fig. 5.6(e) and Fig. 5.6(f) for the degraded dataset while the results of these methods for the original dataset are displayed in Fig. 5.7(e) and Fig. 5.7(f), respectively. Since our approach is based on CS, comparison is also shown with the other CS based techniques such as sparseFI [132] and Li *et. al* [129]. The pan-sharpened results obtained using these methods for degraded dataset are displayed in Fig. 5.6(g) and Fig. 5.6(h), respectively and the same for the original dataset are displayed in Fig. 5.7(g) and Fig. 5.7(h). In Fig. 5.6(i) and Fig. 5.7(i) we display the results of the proposed pan-sharpening approach using the degraded and un-degraded datasets, respectively. The original image is available for the experiment on degraded data and the same is displayed in the Fig. 5.6(j). In order to have better visual comparison among the results we also display the enlarged version of a small area marked with green border in Fig. 5.6(j) and Fig. 5.7(i). This is shown in each of the images at the bottom left corner. The quantitative measures of this experiment using the degraded dataset are summarized in Table 5.2. For the experiment

Table 5.2: Quantitative measures for Ikonos-2 imagery shown in Fig. 5.6 and Fig. 5.7. Here, boldface values indicate the value is better amongst the other methods.

Dataset	Measure	FIHS [33]	AIHS [37]	AWLP [60]	Brovey [22]	sparseFI [132]	Li *et. al* [129]	Proposed
Degraded (Fig. 5.6)	CC(1) Band 1	0.8729	0.9148	0.8693	0.8781	0.8059	0.8654	**0.9245**
	Band 2	0.9305	0.9346	0.9156	0.9283	0.7933	0.8997	**0.9370**
	Band 3	0.9394	0.9393	0.9309	0.9434	0.7782	0.8995	**0.9575**
	Band 4	0.9484	0.8908	0.8988	0.9153	0.7466	0.8590	**0.9676**
	ERGAS(0)	7.0320	4.4546	5.2946	5.0216	6.3824	4.9275	**4.0861**
	$Q_{AVG}(1)$	**0.8082**	0.7028	0.6897	0.6983	0.6909	0.5843	0.8042
	RMSE(0)	18.0100	8.9372	10.6475	11.5493	12.6856	9.7400	**7.4933**
	SAM(0)	5.9656	5.9377	5.7666	5.7753	5.9351	7.1801	**5.5847**
Un-degraded (Fig. 5.7)	$D_s(0)$	0.4163	0.4044	0.3824	0.2228	0.3674	0.1783	**0.1210**
	$D_\lambda(0)$	0.2353	0.1326	0.2769	0.0865	0.1284	0.0773	**0.0710**
	QNR(1)	0.4462	0.5166	0.4465	0.7098	0.5512	0.7580	**0.8165**

on un-degraded dataset the reference pan-sharpened image is not available and hence the quantitative evaluation is carried out with the help of QNR index which is included in the same Table 5.2. The boldface values in the table indicate that the value is better amongst the other methods. Also, in Table 5.2 we display the ideal value of different measures in the bracket.

Following points can be observed from the results on the degraded and un-degraded datasets of Ikonos-2 satellite. In the FIHS [33] and AIHS [37] methods, the extracted high frequency details from the Pan image are injected in the upsampled MS image. One can see from Fig. 5.6(c) and Fig. 5.7(c) that due to the appropriate weighting factors used in the FIHS approach [33], color information is better preserved in addition to the preservation of spatial details. This weighting factor is calculated using the spectral response functions of the Ikonos-2 satellite sensor. The results of AIHS method [37] (see Fig. 5.6(d) and Fig. 5.7(d)) lack in the preservation of color details when compared to other approaches. We observe that although the images displayed in Fig. 5.6(e) and Fig. 5.6(f) corresponding to AWLP [60] and Brovey [22] methods indicate enhancement in terms of spatial details, the spectral contents are not well preserved. One can compare the magnified regions of FIHS [33], AWLP [60] and the proposed methods displayed in Fig. 5.6(c, e, i) for the degraded datset and observe that the preservation of the spectral details such as the colors of trees along with the edge features of houses are better in the result of the proposed method when compared to that of FIHS [33] and AWLP [60]. Both the pan-sharpening results based on the CS theory displayed in Fig. 5.6(g, h) and Fig. 5.7(g, h) do not perform well as far as the preservation of the spatial details are concerned. They also show color saturation. This effect can be clearly seen in the magnified regions. Visual comparison of different approaches indicate that AIHS [37],

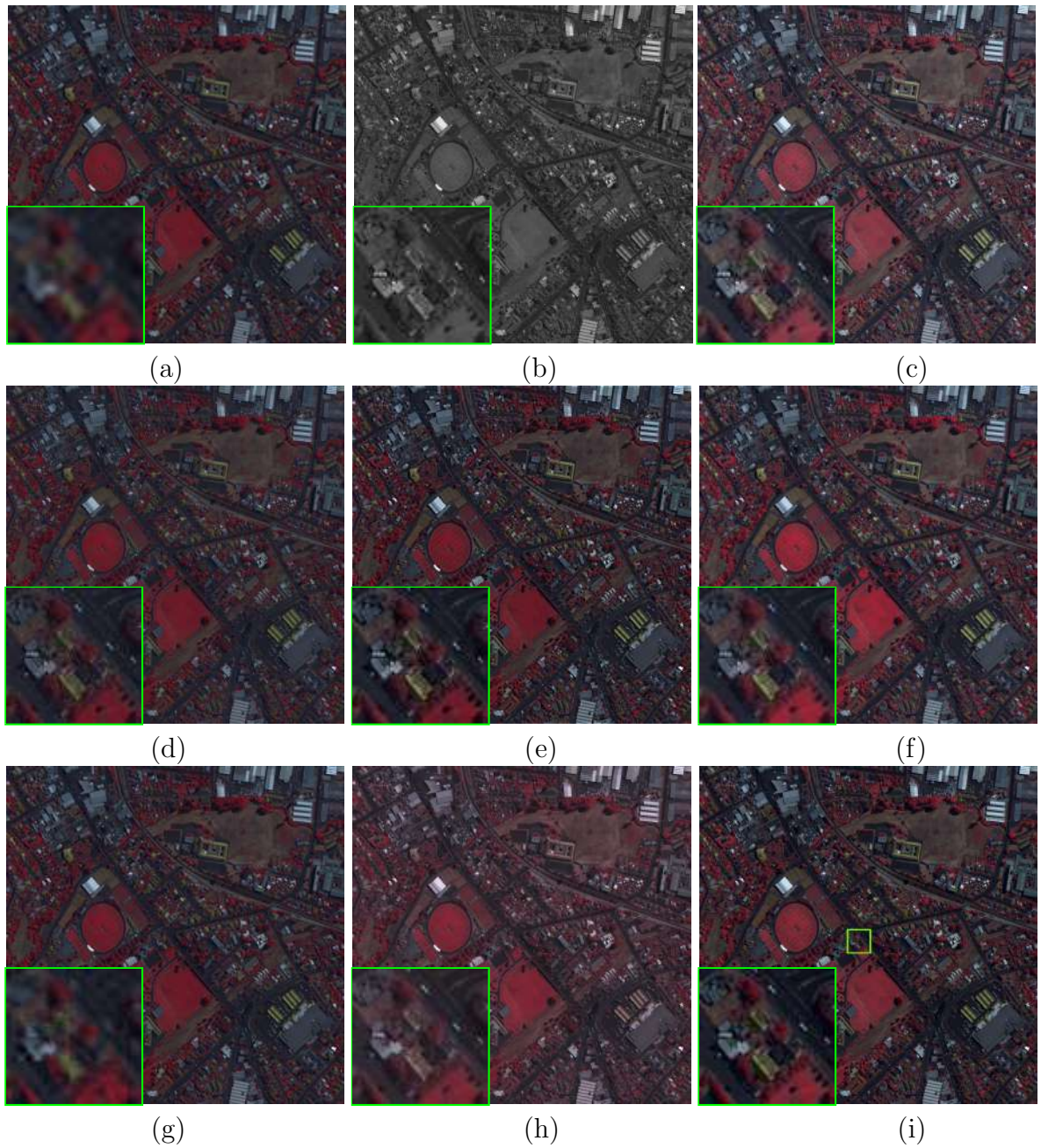


Figure 5.7: Results of pan-sharpening on un-degraded (original) dataset of Ikonos-2 satellite consisting of urban area shown as color composite of bands-4, 3 and 2 ($q = 4$). (a) Original MS image of size 256×256 upsampled to the size of Pan image. (b) Original Pan image of size 1024×1024. Pan-sharpened images obtained using (c) FIHS [33], (d) AIHS [37], (e) AWLP [60], (f) Brovey [22], (g) sparseFI [132], (h) Li *et. al* [129] and (i) proposed approach. The magnified image of a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

Brovey [22], sparseFI [132] and LI *et. al* [129] methods do not preserve the spatial details better when compared to the methods of FIHS [33] and AWLP [60]. The high frequency details in FIHS [33], AWLP [60] and the proposed approaches look similar to that of the Pan image (see Fig. 5.6(b) and Fig. 5.7(b)). We can also observe that in the magnified region of the proposed method in the result of un-degraded dataset (see Fig. 5.7(i)) the shape of the vehicles on the roads are similar to that of the Pan image shown in Fig. 5.7(b) indicating spatial resolution enhancement for the proposed approach.

Quantitative measures for this experiment are shown in Table 5.2. When we compare the quantitative measures of the proposed approach with the other methods, we can see that the proposed method shows better performance except for the Q_{AVG} index indicating that the spatial as well as spectral features are better preserved in the proposed method. Although the value of Q_{AVG} is higher for FIHS method [33] when compared to our approach, the difference between these two values is not significant. Since the original pan-sharpened image is unavailable in the experiment on un-degraded data, the performance in this case is tested in terms of QNR measure and the same is shown in Table 5.2 where the spectral distortion (D_s) and spatial distortion (D_λ) are lowest for the proposed method. One can conclude from the experiments on degraded and un-degraded images of Ikonos-2 satellite that the preservation of the spatial and spectral details in the proposed method is better when compared to other approaches.

5.6.3 Experimental Results on Degraded and Un-degraded Quickbird Datasets

Here, we describe the results obtained using the degraded and original datasets of Quickbird satellite. Table 3.1 lists the different details of the data captured using this satellite. Similar to the previous experiment on Ikonos-2 satellite, the results of degraded and un-degraded datasets for Quickbird satellite are displayed in Fig. 5.8 and Fig. 5.9, respectively. Here, experiments are conducted on the dataset consisting of semi-urban area of Sundarban, India and the results are displayed using the color composition of bands 3, 2 and 1. Use of different color composition than the Ikonos-2, helps us in testing the performance of our approach over various spectral bands. The degraded and original set of test images (MS and Pan) are displayed in Fig. 5.8(a, b) and Fig. 5.9(a,

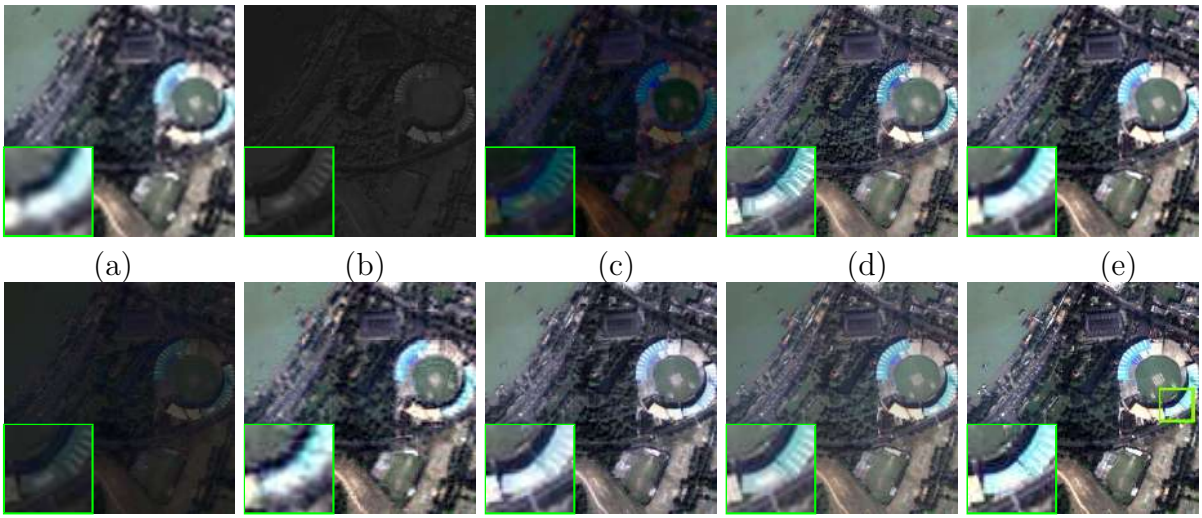


Figure 5.8: Results of pan-sharpening on degraded dataset of Quickbird satellite consisting of semi-urban area shown as color composite of bands-3, 2 and 1 ($q = 4$). (a) LR MS image of size 64×64 upsampled to the size of Pan image. (b) Pan image of size 256×256. Pan-sharpened images obtained using (c) FIHS approach [33], (d) AIHS [37], (e) AWLP [60], (f) Brovey method [22], (g) sparseFI [132], (h) Li *et. al* [129] and (i) proposed approach. (j) Original MS image. The magnified region of a small square region shown with a green border in (j) is displayed at the bottom left corner of all the images.

b), respectively. In Fig. 5.8(c-h) we display the pan-sharpening results using FIHS [33], AIHS [37], AWLP [60], Brovey [22], sparseFI [132] and Li *et. al* [129] approaches on the degraded dataset. The results of these pan-sharpening techniques for un-degraded dataset are shown in Fig. 5.9(c-h). Finally, in Fig. 5.8(i) and Fig. 5.9(i) we display the pan-sharpening results for the proposed method for degraded and un-degraded datasets, respectively. The quantitative evaluation corresponding to these experiments are listed in the Table 5.3. In each of these results displayed in Fig. 5.8 and Fig. 5.9, we also display the zoomed-in version of a small square area on the bottom left corner and are marked with green border. It corresponds to small square region shown in Fig. 5.8(j) and Fig. 5.9(i) consisting of the roof area of stadium.

Looking at the results of degraded dataset in Fig. 5.8, one can make the following observations. The FIHS method [33] fails in preserving both the spatial and spectral contents (see Fig. 5.8(c)). The result due to AIHS technique [37] improves the spatial details but the color distortion is clearly seen in the magnified region displayed in Fig. 5.8(d). The result of AWLP method [60] shown in Fig. 5.8(e) has lesser extent of color and spatial distortions. However, by comparing the zoomed-in versions of the results of AWLP [60]

and the proposed methods displayed in Fig. 5.8(e) and Fig. 5.8(i), respectively, we can clearly see that the proposed method has better preservation of spectral features in addition to the preservation of the edge details when compared to that of AWLP method [60]. Similar to FIHS method [33], the result of Brovey method [22] (Fig. 5.8(f)) also suffers from color distortion. The sparseFI pan-sharpening technique [132] shown in Fig. 5.8(g) has significant loss of spatial and spectral details. The other CS based pan-sharpening method i.e., Li *et. al.* approach [129] which is displayed in Fig. 5.8(h) has improved the spectral content. However, one can see that the preservation of high frequency details using this method is very poor when compared to the proposed method. By comparing the zoomed-in images of AWLP [60], Li *et. al.* [129] and the proposed methods (see Fig. 5.8(e, h, i)) we may conclude that the spectral attributes such as the blue and white color stripes appear distorted in the results of AWLP [60] and Li *et. al.* [129] methods when compared to the result of the proposed method. Also the proposed approach has better preservation of spatial features such as the different segments in the roof area of the stadium when compared to the other approaches.

Similar to degraded dataset the potential of the proposed method is also evaluated by comparing it with the results of un-degraded images and this is shown in Fig. 5.9. Looking at the magnified regions of the various approaches we can say that pan-sharpened images of FIHS method [33] and Brovey technique [22] displayed in Fig. 5.9(c) and Fig. 5.9(f), respectively have comparatively higher spectral distortions than other methods. The result of AIHS method [37] suffers in preserving both the spatial and spectral details. The blockiness effect can be clearly seen in the result of this method displayed in Fig. 5.9(d). The CS based methods, sparseFI [132] and Li *at. al.* [129] are displayed in Fig. 5.9(g, h) and both these approaches fail to output better pan-sharpened images when compared to that of proposed method (Fig. 5.9(i)). On comparing the magnified images of AWLP [60] and the proposed methods displayed in Fig. 5.9(e) and Fig. 5.9(i), respectively one can see that each region in the roof top are well separated in the result of the proposed method (see Fig. 5.9(i)). Similar to the degraded experiment here also we can conclude that the proposed method has better perceptual performance when compared to the other approaches. The pan-sharpened result of the proposed method has less color distortion with significant improvement in edge features.

The quantitative comparison shown in Table 5.3 indicates that the proposed approach

Table 5.3: Quantitative measures for Quickbird imagery shown in Fig. 5.8 and Fig. 5.9. Here, boldface values indicate the value is better amongst the other methods.

Dataset	Measure	FIHS [33]	AIHS [37]	AWLP [60]	Brovey [22]	sparseFI [132]	Li *et. al* [129]	Proposed
Degraded (Fig. 5.8)	CC(1) Band 1	0.7066	0.9384	0.9396	0.8538	0.9205	**0.9572**	0.9463
	Band 2	0.8461	0.9649	0.9498	0.9271	0.9301	**0.9813**	0.9716
	Band 3	0.8057	0.9139	0.9445	0.9246	0.9255	0.9654	**0.9722**
	Band 4	0.7764	0.9671	0.9542	0.8247	0.9321	0.9751	**0.9788**
	ERGAS(0)	57.3926	5.0335	4.8138	71.4647	5.6189	3.9499	**3.7549**
	$Q_{AVG}(1)$	0.6448	0.8564	0.8729	0.3433	0.8597	0.7795	**0.8871**
	RMSE(0)	72.2450	18.4657	17.8102	79.2845	20.6792	13.7457	**10.7645**
	SAM(0)	18.5125	5.6591	4.7019	4.1529	5.0594	6.2890	**4.1169**
Un-degraded (Fig. 5.9)	$D_s(0)$	0.3636	0.3145	0.2245	0.6477	0.1775	0.2298	**0.2148**
	$D_\lambda(0)$	0.5169	0.1650	0.1704	0.2337	0.1961	0.0620	**0.0480**
	QNR(1)	0.3074	0.5723	0.6432	0.2699	0.6611	0.7224	**0.7474**

performs better in terms of quantitative measures as well, showing better values for all the measures except for the values of CC for bands 1 and 2. The CC measure for these two bands are slightly better for Li *et. al.* [129]. Over all, the spectral and spatial fidelity indices are better in the results of the proposed method when compared to other pan-sharpening techniques. One may arrive at following conclusions from the displayed results and the quantitative assessment. Due to the Gabor filter which extracts the bandpas details present at different regions, the proposed method exhibits better preservation of spatial features. Estimation of initial HR approximation using the self-similarity and CS theory improves the performance of the proposed method in obtaining accurate degradation matrix between the LR and HR images and this in effect leads to better final output.

5.6.4 Experimental Results on Degraded and Un-degraded Worldview-2 Datasets

Finally, we explain the results on the degraded and un-degraded images of Worldview-2 satellite. This satellite provides the MS and Pan images with the highest spatial resolution among the other satellites as indicated in Table 3.1 which also shows the other details of the data captured using this satellite. Similar to the earlier experimentations here also we compare the results of various state of the art pan-sharpening techniques with that obtained using the proposed method for degraded and un-degraded datasets and these results are shown as color composite for bands 7, 5 and 3 in Fig. 5.10 and Fig. 5.11, respectively. We display the pan-sharpening results along with the display of MS and Pan images. A magnified region is also displayed in all these images as done in the

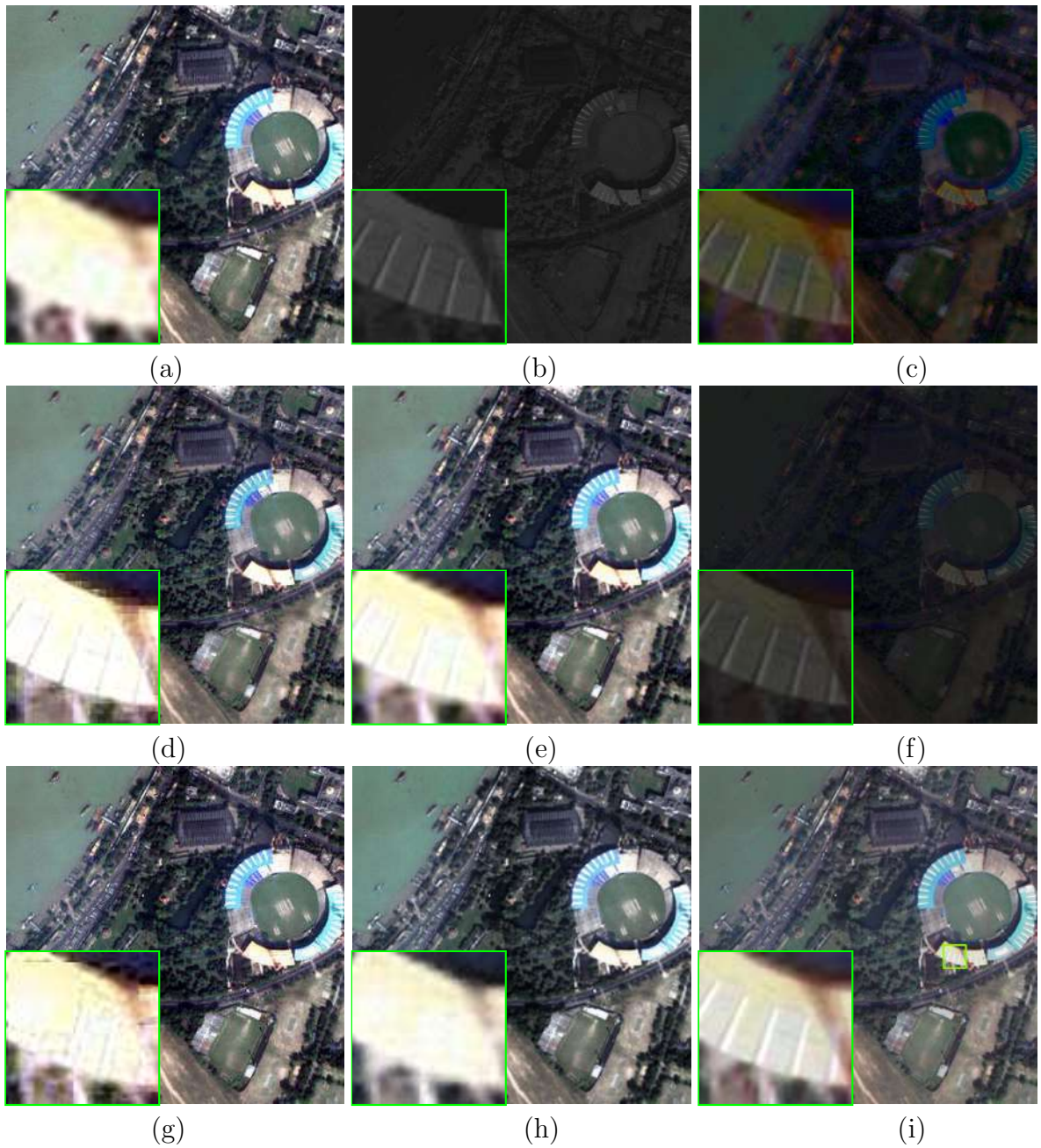


Figure 5.9: Results of pan-sharpening on un-degraded dataset of Quickbird satellite consisting of semi-urban area shown as color composite of bands-3, 2 and 1 ($q = 4$). (a) Original MS image of size 256×256 upsampled to the size of Pan image. (b) Original Pan image of size 1024×1024. Pan-sharpened images obtained using (c) FIHS [33], (d) AIHS [37], (e) AWLP [60], (f) Brovey [22], (g) sparseFI [132], (h) Li *et. al* [129] and (i) proposed approach. The magnified image of a small square region with a green border shown in (i) is displayed at the bottom left corner of all the images.

experiments on Ikonos-2 and Quickbird satellites. The quantitative measures of the these results have been listed in Table 5.4.

Looking at the results of FIHS approach [33] displayed in Fig. 5.10(c) and Fig. 5.11(c) for degraded and un-degraded datasets we observe the spectral distortions in these results. Similarly, the pan-sharpening results of Brovey method [22] (see Fig. 5.10(f) and Fig. 5.11(f)) also lack in preserving the color details. Comparing the results of AIHS [37] and AWLP [60] methods for degraded dataset displayed in Fig. 5.10(d) and Fig. 5.10(e), respectively we may say that AWLP [60] has better spatial features with lesser spectral distortions when compared to the result of AIHS method [37]. This observation is also true for un-degraded dataset which is evident by comparing their results displayed in Fig. 5.11(d, e). However, comparing the result of AWLP [60] shown in Fig.5.10(e) with that of the proposed method which is displayed in Fig. 5.10(i) we notice that in the proposed method, various regions in the urban area have textures comparable to the original MS image displayed in Fig. 5.10(j). The same conclusion holds for the results of un-degraded dataset displayed in Fig. 5.11(e, i). Similar to the earlier experiments, the results of sparseFI approach [132] suffers from blockiness effect (see Fig. 5.10(g) and Fig. 5.11(g)). The results of the other CS based approach for pan-sharpening i.e., the approach proposed by Li *et. al.* [129] displayed in Fig. 5.10(h) and Fig. 5.11(h) and show better preservation of the color details. However, one can see that the result of the proposed method (Fig. 5.10(i) and Fig. 5.11(i)) show significant improvement in spatial details with reduced color distortion when compared to that of Li *et. al.* approach [129].

In the results using the degraded dataset we highlight the few objects consisting of buildings, roads, trees, etc., by showing this region magnified. Comparing the zoomed-in regions of all the results we can conclude that the edge features of buildings and the color of trees are better preserved in the result of the proposed method (see Fig. 5.10(i)) when compared to the results of other state of the art methods such as AWLP [60] (Fig. 5.10(e)) and Li *et. al.* [129] (Fig. 5.10(h)). Similarly, from the results of the un-degraded dataset depicted in Fig. 5.11 we observe that the proposed method shows significant improvement when compared to the other pan-sharpening techniques. The quantitative assessment for these results listed in Table 5.4 shows that the performance of the proposed method for the degraded dataset is better in terms of every measure except that the CC for band 7 is better for the method proposed by Li *et. al.* [129]. However, we can see that the difference

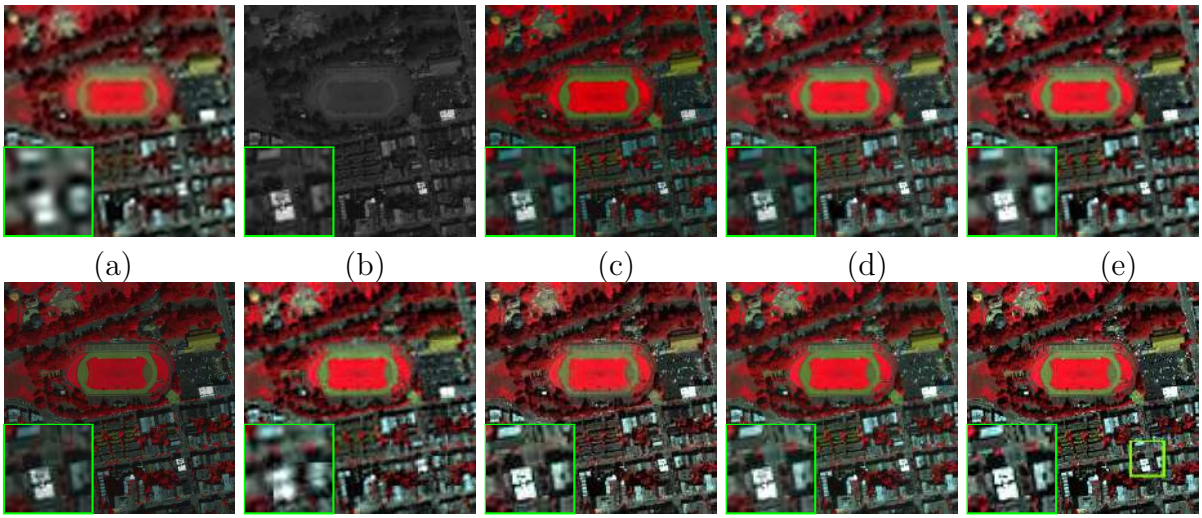


Figure 5.10: Results of pan-sharpening on degraded dataset of Worldview-2 satellite consisting of urban area shown as color composite of bands-7, 5 and 3 ($q = 4$). (a) LR MS image of size 64×64 upsampled to the size of Pan image. (b) Pan image of size 256×256. Pan-sharpened images obtained using (c) FIHS approach [33], (d) AIHS [37], (e) AWLP [60], (f) Brovey method [22], (g) sparseFI [132], (h) Li *et. al* [129] and (i) proposed approach. (j) Original MS image. The magnified region of a small square region shown with a green border in (j) is displayed at the bottom left corner of all the images.

between the values of CC index between Li *et. al.* [129] and the proposed method is not very significant. Visual comparison of Li *et. al.* [129] and the proposed methods show that the proposed method performs better in preserving both spatial as well as color details. Note that the pan-sharpening result of Li *et. al.* [129] approach depends on the dictionary which is constructed using the databases of original samples of MS and Pan images. For the case of un-degraded dataset, the spectral distortion (D_λ) for the proposed method is smaller when compared to all the other methods. The spatial distortion (D_s) and QNR index are slightly better for the approach of Li *et. al.* [129]. Looking at the displayed results and the quantitative measures obtained using the dataset of three satellites one may conclude that the pan-sharpened image obtained using the proposed method results in better preservation of edge features, object shapes and colors in the scene.

5.6.5 Computation Complexity

The proposed approach and all the other pan-sharpening methods have been implemented using Matlab 7.6 installed on Intel $i3$ processor with 2.4 GHz, 4 GB memory. In order to calculate the computation time for pan-sharpening of individual band we compute

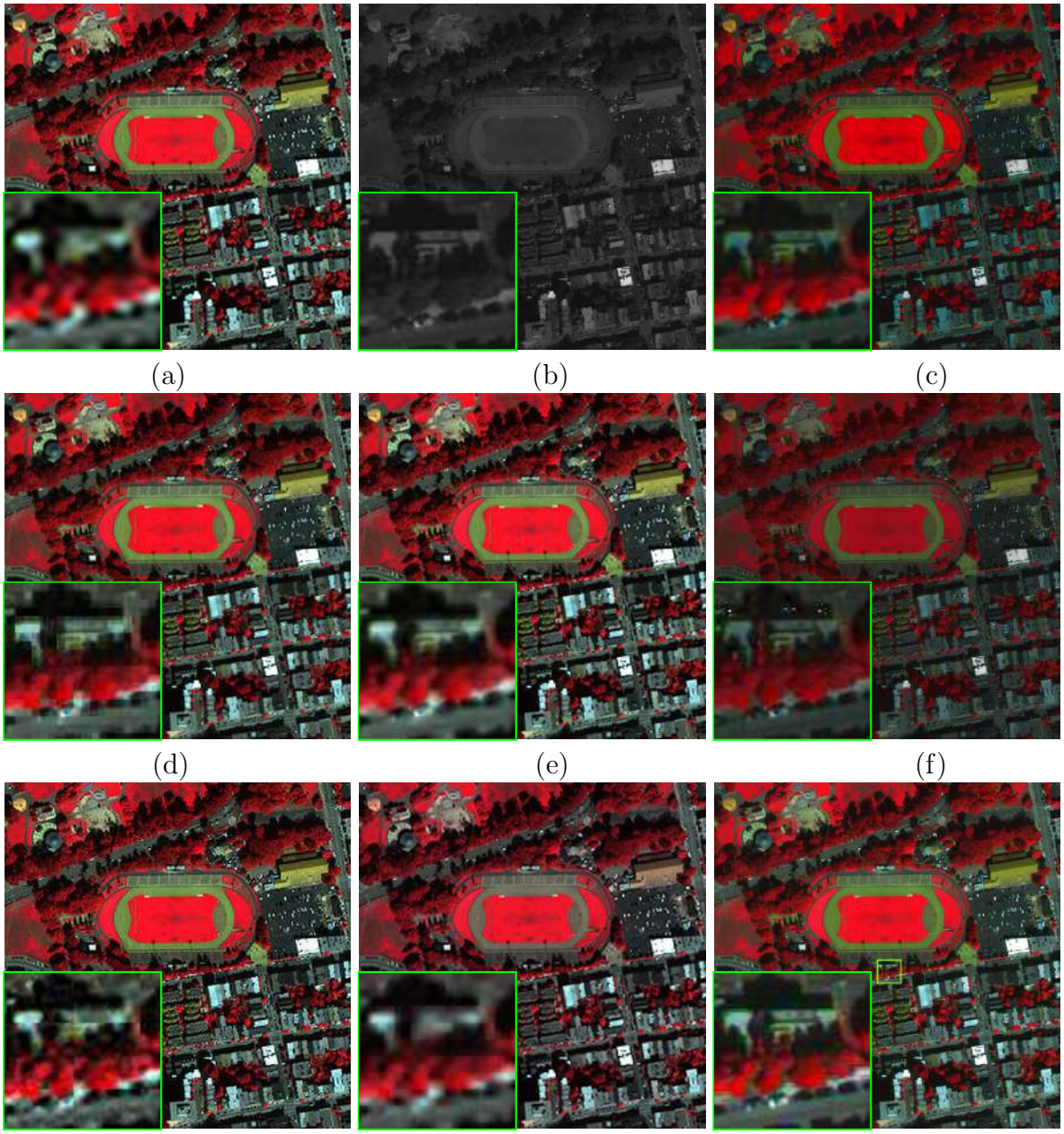


Figure 5.11: Results of pan-sharpening on un-degraded dataset of Worldview-2 satellite consisting of urban area shown as color composite of bands-7, 5 and 3 ($q = 4$). (a) Original MS image of size 256×256 upsampled to the size of Pan image. (b) Original Pan image of size 1024×1024. Pan-sharpened images obtained using (c) FIHS [33], (d) AIHS [37], (e) AWLP [60], (f) Brovey [22], (g) sparseFI [132], (h) Li *et. al* [129] and (i) proposed approach. The magnified image of a small square region with a green border shown in (i) is displayed at the bottom right corner of all the images.

Table 5.4: Quantitative measures for Worldview-2 imagery shown in Fig. 5.10 and Fig. 5.11. Here, boldface values indicate the value is better amongst the other methods.

Dataset	Measure	FIHS [33]	AIHS [37]	AWLP [60]	Brovey [22]	sparseFI [132]	Li *et. al* [129]	Proposed
Degraded (Fig. 5.10)	CC(1) Band 2	0.8829	0.9220	0.9061	0.9182	0.5526	0.9292	**0.9309**
	Band 3	0.9238	0.9277	0.9092	0.9352	0.5642	0.9485	**0.9552**
	Band 5	0.9201	0.8600	0.9145	0.9351	0.6120	0.9623	**0.9679**
	Band 7	0.8948	0.8968	0.9033	0.8875	0.7365	**0.9593**	0.9569
	ERGAS(0)	22.5893	9.9920	9.5293	22.0300	17.7815	7.2197	**6.1249**
	$Q_{AVG}(1)$	0.6468	0.6187	0.6016	0.5674	0.3538	0.7869	**0.8541**
	RMSE(0)	32.5583	23.2829	22.2374	35.5597	42.8931	17.4373	**16.0521**
	SAM(0)	16.2389	11.8724	11.1375	12.4531	13.5631	8.9023	**7.2100**
Un-degraded (Fig. 5.11)	$D_s(0)$	0.3068	0.4723	0.2748	0.3028	0.3487	**0.0256**	0.1552
	$D_\lambda(0)$	0.3339	0.1264	0.1537	0.1789	0.2298	0.1781	**0.0959**
	QNR(1)	0.4616	0.4609	0.6136	0.5723	0.4939	**0.8008**	0.7637

the total time required for all the bands and then use the average time as the time for pan-sharpening of single band. The average computation time for each method is listed in the Table 5.5. Following observations can be made by looking at the timing given in Table 5.5. Since the CS based methods i.e, sparseFI [132] and Li *et. al.* [129] depend on the construction of the dictionary and the use of optimization methods, they take more time when compared to the other pan-sharpening methods such as FIHS [33], AIHS [37], AWLP [60] and Brovey [22]. From the table it is clear that the proposed approach takes higher execution time when compared to the other approaches excluding the approach of Li *et. al.* [129] which has highest computational time. Increase in computation time for the proposed method is due to the estimation of initial HR approximation and also due to the time for performing regularization using iterative optimization technique. The methods proposed in [22, 33, 37, 60] are not iterative. Note that though they are computationally efficient their results suffer from the effect of aliasing. However, our approach considers the aliasing effect by using a suitable model for degradation. Since the pan-sharpening is an ill-posed problem, it is advantageous to use a degradation model in order to obtain better solution which is done in the proposed method by estimating the degradation between LR and HR MS images. Regularization using a new prior based on Gabor filtering also adds to improving the solution. Although the time complexity of the proposed approach is higher when compared to non iterative approaches, it has lesser distortions in both spatial and spectral contents. One can speed up the processing time of the proposed method by implementing the optimized code on the graphical processing unit (GPU). The fastest pan-sharpening approach is the FIHS method [33] which gives comparable results only for the Ikonos-2 dataset. Note that the execution time of Li *et.*

Table 5.5: Average computation time involved for different pan-sharpening approaches.

Experiment	Average computation time to obtain single pan-sharpening band (in *seconds*)						
	FIHS [33]	AIHS [37]	AWLP [60]	Brovey [22]	sparseFI [132]	Li *et. al.* [129]	Proposed
Degraded dataset	0.32	0.79	0.33	0.35	60.58	150.45	122.39

al. [129] is highest. This is due to the construction of the dictionary from the original MS and Pan images and also due to the use of computationally taxing optimization such as basis pursuit to obtain the pan-sharpened image. The time taken by our approach is not only less than the approach by Li *et. al.*, it performs better in terms of qualitative as well as quantitative comparison.

5.7 Comparison of Different Proposed Fusion Approaches

Here, we compare the results of different fusion approaches proposed in this thesis. In this chapter we have proposed the model based fusion approach using the concept of self-similarity and Gabor prior. The fusion results of this approach are compared with the other model based approach proposed in chapter 4 as well as with the results obtained using edge preserving filters proposed in chapter 3. Although we have used different datasets for comparison of fusion results here we illustrate the comparison with the images acquired using Quickbird satellite where the spatial resolution difference between Pan and MS images is 4. In Fig. 5.12 and Fig. 5.13 we show the results with the color composition of 3, 2 and 1 bands obtained using different fusion methods for degraded and un-degraded datasets, respectively. The quantitative evaluation is performed by computing different measures and the same is displayed in Table 5.6.

Similar to the previous experiments here also we display the upsampled MS image for the experiments of degraded and un-degraded dataset in Fig. 5.12(a) and Fig. 5.13(a), respectively. The Pan images for the same are displayed in Fig. 5.12(b) and Fig. 5.13(b). In chapter 3, we have proposed two fusion approaches based on details extractions using multi-stage guided filter (MGF) and difference of Gaussians (DoGs). The results of these approaches are displayed in Fig. 5.12(c, d) and Fig. 5.13(c, d) for degraded and un-degraded datasets, respectively. In Fig. 5.12(e) and Fig. 5.13(e) we depict the pan-

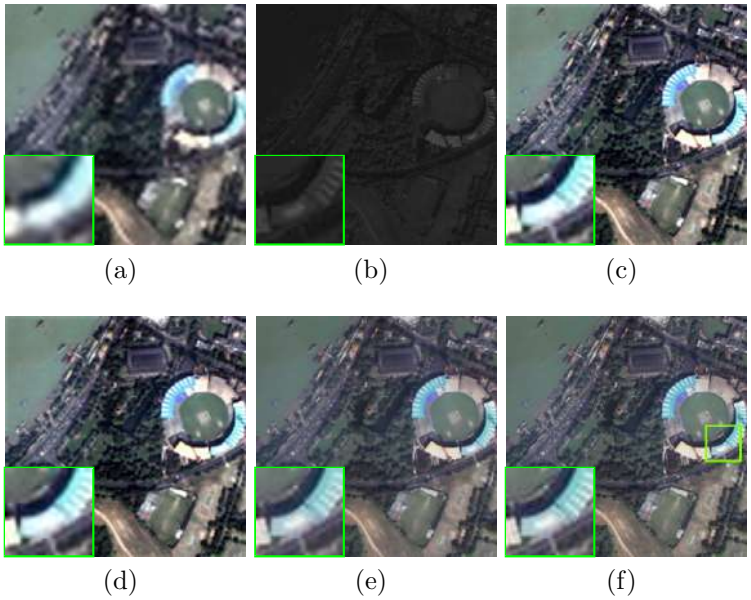


Figure 5.12: Results of pan-sharpening on degraded dataset of Quickbird satellite consisting of semi-urban area shown as color composite of bands-3, 2 and 1 ($q = 4$). (a) LR MS image of size 64×64 upsampled to the size of Pan image. (b) Pan image of size 256×256. Fusion results obtained using (c) MGF, (d) DoGs, (e) model based approach using NSCT and (f) model based approach using the concept of self-similarity and Gabor prior. The magnified region of a small square region shown with a green border in (f) is displayed at the bottom left corner of all the images.

sharpening results for degraded and un-degraded dataset obtained using the model based approach described in chapter 4 which uses NSCT to obtain the initial HR approximation. Finally, the fusion results of model based approach using the concept of self-similarity and Gabor prior are displayed in Fig. 5.12(f) and Fig. 5.13(f) for degraded and un-degraded cases, respectively. One may observe the following points while looking at the results displayed in Fig. 5.12 and Fig. 5.13.

One may see that the fused images displayed in Fig. 5.12(c, d) and Fig. 5.13(c, d) obtained using edge preserving filters preserve the spatial details of the Pan image although the spectral fidelity is lacking due to the upsampling of MS image. By looking at the zoomed-in images of degraded dataset (see Fig. 5.12(c, d)) we can see that spectral details of blue colors are blurred. Similarly, the performance of proposed fusion techniques based on edge preserving filters for the un-degraded case looks inferior when compared to same with other proposed fusion methods using model based approaches. In the model based approaches we avoid making use of the upsampling of MS image while fusing and we use the initial HR approximation before the final fused image is obtained. The use of initial

Table 5.6: Quantitative measures for Quickbird imagery shown in Fig. 5.12 and Fig. 5.13. Here, boldface values indicate the value is better amongst the other methods.

Dataset	Measure	Proposed fusion approaches using			
		MGF	DoGs	model based method using NSCT	model based method using self-similarity and Gabor prior
Degraded (Fig. 5.12)	CC(1) Band 1	0.9410	0.9390	**0.9501**	0.9463
	Band 2	0.9541	0.9631	0.9639	**0.9716**
	Band 3	0.9477	0.9444	0.9596	**0.9722**
	Band 4	0.9600	0.9530	0.9690	**0.9788**
	ERGAS(0)	4.8071	4.7810	3.9048	**3.7549**
	$Q_{AVG}(1)$	0.8858	0.8860	**0.8911**	0.8871
	RMSE(0)	12.9022	13.2090	11.8926	**10.7645**
	SAM(0)	4.6980	4.6821	4.5210	**4.1169**
Un-degraded (Fig. 5.13)	$D_s(0)$	0.2189	0.2290	0.2160	**0.2148**
	$D_\lambda(0)$	0.1629	0.1687	0.1020	**0.0480**
	QNR(1)	0.6538	0.6509	0.7040	**0.7474**

estimate in these proposed fusion methods is to estimate the degradation between LR MS and fused MS images. The MRF parameter is also estimated using this initial estimate. The first model based approach is described in chapter 4 which uses the directional transforms such as CT and NSCT to obtain the initial estimate. The pan-sharpening results obtained using this model based approach are displayed in Fig. 5.12(e) and Fig. 5.13(e) for degraded and un-degraded cases. We can see here that the improvement in terms of features such as colors of stadium roof and boundary of various targets appear sharp when compared to that of fusion results obtained using edge preserving filters. This is due to the estimation of the degradation caused by the downsampling of LR MS image which is possible with the help of initial estimate. However such estimation is missing in the case of proposed fusion approaches using edge preserving filters.

In the second model based proposed fusion approach we chiefly use the given LR MS observation to derive the initial HR approximation using the concepts of self-similarity and CS. Also, the patchwise estimation of coefficients of degradation matrix in this approach yield the better values of degradation matrix when compared to same with the other model based fusion approach. In addition to this, the proposed fusion method also uses the prior based on Gabor filter in order to extract the details from the Pan image. Fig. 5.12(f) and Fig. 5.13(f) show the performance of the this fusion technique for degraded and un-degraded cases, respectively. Here, one can compare these results with that of model based fusion method using NSCT and observe that fusion based on self-similarity and Gabor prior outperforms when compared to fusion using NSCT. The enhancement in the quality of the fused images is due to better patchwise estimation of degradation matrix and Gabor prior which extracts the bandpass details from the Pan

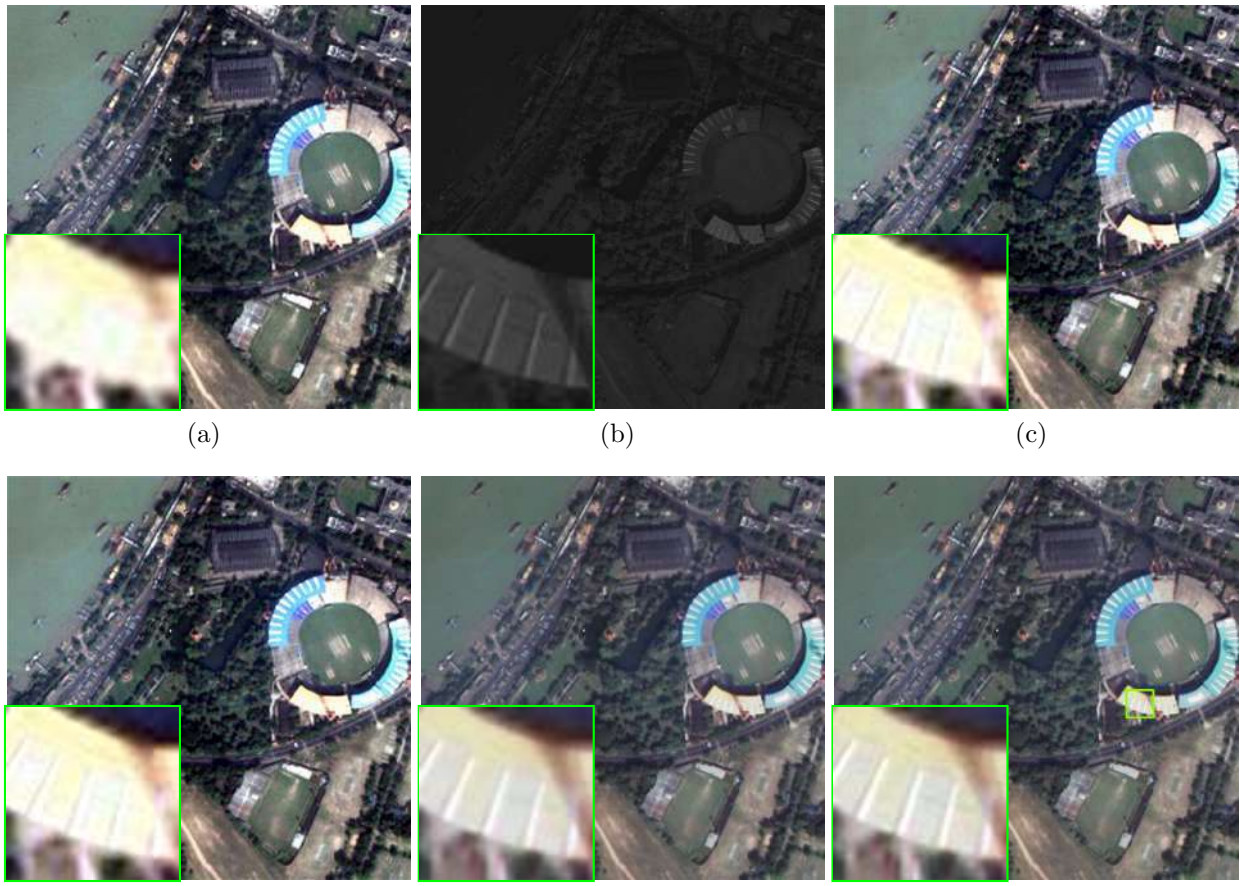


Figure 5.13: Results of pan-sharpening on un-degraded dataset of Quickbird satellite consisting of semi-urban area shown as color composite of bands-3, 2 and 1 ($q = 4$). (a) LR MS image of size 256×256 upsampled to the size of Pan image. (b) Pan image of size 1024×1024. Fusion results obtained using (c) MGF, (d) DoGs, (e) model based approach using NSCT and (f) model based approach using the concept of self-similarity and Gabor prior. The magnified region of a small square region shown with a green border in (f) is displayed at the bottom left corner of all the images.

image. Looking at the results displayed in Fig. 5.12 & Fig. 5.13 and the quantitative measures given in Table 5.6 we can conclude that the performance of the proposed fusion methods using model based approach is better when compared to same with the edge preserving filters.

It is of interest to note the computational efficiency of these fusion approaches. Table 5.7 lists the computation time to obtain the fused image using the fusion approaches proposed in this thesis. Here, one can see that the computation complexity for the fusion approach using edge preserving filters i.e., MGF and DoGs is less when compared to other proposed model based approaches. This is because fusion approaches based on edge preserving filters are non-iterative and hence they are computationally less intensive.

Table 5.7: Average computation time involved for different fusion approaches.

Experiment	Average computation time to obtain single pan-sharpening band (in *seconds*)			
	Proposed pan-sharpening algorithm using			
	MGF (Chapter 3)	DoGs (Chapter 3)	Model based approach using NSCT (Chapter 4)	Model based approach using self-similarity (Chapter 5)
Degraded dataset	1.23	0.20	3.54	122.39

However, these approaches i.e., using MGF and DoGs based methods do not consider the effect of aliasing while fusing to obtain the final fused image. In the model based approaches, we consider the aliasing effect by using a suitable model for degradation. Here, we use the initial HR approximation which is used to estimate the degradation between LR MS and final fused image. This initial HR approximation is also used to estimate the regularization parameter in the final cost function. Although the time complexity of the proposed fusion techniques using model based approach is higher when compared to that of the fusion approaches based on edge preserving filters, it has less distortions in spatial and spectral details. For example, the computational complexity of the pan-sharpening approach based on self-similarity and CS is highest. However, the fused images obtained using this approach are better when compared to the other proposed fusion approaches. One can note that the computation time complexity of the model based fusion approaches can be decreased by implementing the optimized code run on a GPU.

5.8 Conclusion

We have proposed a novel pan-sharpening method using the concept of self-similarity and Gabor prior. We cast the pan-sharpening problem in a restoration framework and obtain the final solution by regularizing the cost function. A new Gabor prior is proposed in order to extract the bandpass features from the Pan image. The final cost function is optimized using simple gradient based optimization technique. The potential of the proposed fusion method has been verified by conducting the experiments on both the degraded and un-degraded images captured by using the different satellite images such as Ikonos-2, Quickbird and Worldview-2. Quantitative evaluation of our results involves the traditional as well as new quality indices. From the pan-sharpened image and the measures one can conclude that the proposed method has better preservation of color and edge information when compared to the other techniques.

Chapter 6

Conclusions and Future Research Work

6.1 Conclusions

Remote sensing satellites capture the data in the form of images which are processed and utilized in various applications such as land area classification, map updating, whether forecast, urban planning, etc. However, due to the constraints on the hardware of the sensors and the available transmission bandwidth of the transponder, many commercial satellites provide the earth information by capturing images which have complementary characteristics. In this thesis, we have addressed the problem of multi-resolution image fusion. Here, the low spatial resolution MS image and high spatial resolution Pan image are combined to obtain a single fused image which has both high spectral and spatial resolutions. We seek for the fused image which has spectral resolution of MS image and spatial resolution of the Pan image. Although the MS and Pan images capture the same geographical area, the complement nature of these images in terms of the spatial and spectral resolutions give rise to variation in the two images. Because of this when we fuse the given images by using direct pixel intensity values, the resultant fused data suffers from the spatial as well as spectral distortions. Another important issue in the problem of multi-resolution image fusion is the registration of MS and Pan images. Accurate registration is a different task and in this thesis we use the registered data. Here, we present the conclusions which are drawn based on the different proposed methods for

pan-sharpening/image fusion.

We began our work by proposing two new fusion techniques based on the edge preserving filters. The Pan image has high frequency details that can be extracted with the help of edge preserving filter. These extracted details are injected into the upsampled MS image. In our work, we use two edge preserving filters namely the guided filter and difference of Gaussians (DoGs) in order to extract the required details present in the Pan image. The extension of the guided filter in the form of multistage is introduced which effectively extracts the details from Pan and MS images. Similarly, the concept of DoGs is used to extract the high frequency features from the Pan image. The potential of the proposed methods were evaluated by conducting the experiments on the original as well as degraded datasets captured using various satellites. The results were compared with the state of the art methods. In addition to the qualitative evaluation we have also checked the quantitative performance of the proposed approaches by calculating the traditional measures as well as a new measure called quality with no reference (QNR).

The main drawback of the edge preservation based fusion techniques is the use of upsampling of MS image and performing fusion without using the regularization. The use of upsampling on the test image introduces distortion in the fused image due to aliasing. Hence our next two approaches avoid using interpolated image directly and these methods are based on using a model for image formation. Here, the given low resolution MS image is modeled as aliased, blurred and noisy version of the unknown high resolution MS image i.e. it represents the degraded version of fused image. Since this is an ill-posed problem it requires suitable regularization in order to obtain the better solution.

In the first model based approach, we obtain an initial HR approximation (initial estimate) with the help of directional transforms such as sub-sampled and non sub-sampled contourlet transforms. Use of these transforms are due to their desirable properties such as anisotropy, directionality etc. Since the edge features in an image occur with the different directions and also along the smooth contours the limited directional transform such as wavelet fails to capture those features. Using CT and NSCT transforms, the high frequency details present in the Pan image are extracted in the transform domain to obtain the initial HR approximation separately. The estimated initial estimate was used in obtaining the transformation between LR and HR MS images by estimating the degra-

dation matrix. An MAP-MRF regularization framework was finally used to obtain the final solution where we model the unknown high resolution MS image as an MRF. Here, the initial estimate was also used in estimating the MRF parameter which avoids the use of empirically chosen parameter as done in many of the MRF based techniques. The final cost function being convex, a simple gradient based optimization was used to minimize the same. The edges in the final fused image were preserved by extracting the edges of the initial estimate using Canny edge operator and hence the optimization was restricted on the non edge pixels only. The proposed method preserves fine details present in different directions with minimum spectral distortion. The experimental results demonstrate that the proposed technique yields better solution as compared to those obtained using the recent state of the art approaches.

Our next model based fusion approach uses the concept of self-similarity and CS theory. Motivation behind using the concept of self-similarity is the use of redundant details present at the different resolutions of an image. Here, the initial estimate was obtained with the help of self-similarity and CS theory. It is important to note in this case, only the LR MS image was used in obtaining the initial approximation. This resulted in accurate estimate of degradation matrix since we avoid the use of low spectral resolution Pan image. The degradation matrix is estimated on the LR-HR patch pairs instead of the entire image. The pan-sharpening problem is cast as a restoration framework and the final solution is obtained by using regularization. A new Gabor prior was proposed in order to extract the high frequency details from the Pan image. The final cost function was minimized using the gradient descent optimization. To show the effectiveness of the proposed approach we conducted experiments on the datasets acquired from the different satellites and the results are compared with the other state of the art methods. From the results and the measures we conclude that the proposed method gives the pan-sharpened images with better preservation of color and edge regions when compared to the state of the art approach.

Before we end the conclusion it may be of interest to mention the following points regarding the use of initial estimate in our works. We assume that the initial estimate represents close approximation to the final fused. Note that though, it can be considered as closer to the fused image one cannot accept it as the final solution. This is because the problem of fusion is ill-posed and a solution space of initial estimate without regularization

is larger when compared to that obtained after regularization. However, one may still consider the initial estimate to represent the global characteristics such as edge details, spatial dependencies etc., of the final solution. In addition to this, we often require an accurate estimate of input-output relationship where the output is not known. In such situations it is reasonable to derive an initial estimate of the output from the given data and use the same for finding the transformation. Due to these reasons we derive an initial estimate and use it for obtaining an improved solution.

6.2 Future Research Work

In multi-resolution image fusion, the captured data in the form of MS and Pan images by the satellites are used to obtain single fused image with high spatial and spectral resolutions. Most of the fusion techniques suffer from the drawbacks such as poor preservation of spectral details, insufficient injection of spatial details, registration error and computational complexity. In this thesis we have addressed the fusion problem using the model based approach and also using the methods based on details injection into the MS image. In the model based approach, we model the given LR MS image as degraded versions of the unknown fused MS image. Since this is an ill-posed inverse problem, regularization is required to obtain the better fused image. We model the true MS image as MRF and use the smoothness prior in the final cost function. In addition to MRF prior, we also use the Gabor prior in order to obtain the high frequency details from the Pan image. However, there is still scope for improving the quality of the final fused image. In this section we describe the future directions for the work carried out in this thesis.

- One of the limitations of many fusion techniques including the proposed approach is the requirement of registration of given MS and Pan images. Since the MS and Pan images are captured by two different sensors there are always chances of registration error. Also, due to the motion of the satellite in its orbital path a slight time difference in capturing the MS and Pan data leads to acquiring images of different geographical area and it results in the registration error. In the proposed fusion methods we assumed that the given MS and Pan data were co-registered. However, it is always an interest of the fusion community to overcome this limitation. One can extend the fusion work proposed in this thesis in order to remove this limitation.

- In this thesis, we have proposed pan-sharpening techniques in which the extracted details from the Pan image using the different edge preserving filters are injected into the upsampled MS image with the appropriate scaling factor. This scale factor accounts for the proportionate detail injection of the Pan image which was calculated using the intensity values of the MS images. Literature on satellite imaging indicates that there is a considerable difference between the spectral response of the MS and Pan sensors. For example in the dataset of Ikonos-2 satellite, the spectral response of the Pan sensor covers the NIR spectral band completely. However, the blue spectral band is not complete under the spectral response of Pan sensor. Also, the spectral response function of the Pan sensor goes beyond the range of blue band. This difference in spectral characteristic of the two sensors i.e., MS and Pan affects the fused data in terms of spectral or spatial distortions. One can add this observation of spectral response of the two different sensors in order to calculate the scaling factor. The scaling factor based on the spectral response can inject accurate details into the MS image and it leads to better fused image. Apart from this the interpolation in the MS image leads to aliasing effect. It would be of interest to look into this as well while injecting the details into the MS image.

- In the proposed model based fusion approach we estimated the degradation that occurs due to the downsampling and blur in the MS image. In order to do this we require the original fused MS image. In such situations we use the close approximation to the final fused (initial HR approximation). In order to get the accurate values of degradation matrix coefficients we require the initial HR approximation as close as possible to the true high resolution MS image. Because of the different desirable properties of the directional transforms such as CT and NSCT we used them in our work to obtain the initial HR approximation of the final fused image. These transforms decompose the images using the bases that yield edges along the smooth contour as well. We decompose the MS and Pan images using CT and NSCT with limited number of levels. It would be of interest to increase the number of decomposition levels making them adaptive to the details present in the image without considerably increasing the time complexity. One can use adaptive directional transforms which fulfil this objective. Use of adaptive directional transform

may increase the complexity of the method. However it can be reduced by using the present day modern high computing processors such as graphical processing unit.

- While using the regularization framework, we have modeled the unknown fused MS image as an MRF and smoothness constraint was used as the prior term in the final cost function. We used Canny edge detector in order to preserve the edges into the final fused image. This gave us the advantages of reduced complexity with edge preservation without using discontinuity preserving priors in the final cost function. However, one can use the edge preserving prior such as MRF with line fields to preserve the edges in the final fused image. The resulting cost function may lead to the non-convex function which cannot be minimized by computationally efficient gradient based optimization techniques. But one may try using graph cuts or grab cuts in order to minimize such non-convex cost functions. These optimization technique are computationally less complex and they yield a solution closer to the optimum.

- The linear image formation model for LR MS image has been used in the proposed model based approach. This model gives the simplicity in the form of mathematical framework in order to find the solution. In the proposed method we have estimated the degradation between LR MS and fused images using a linear image formation model. The degradation consists of the decimation along with blur. In order to add the blur as the non-identity matrix we use the space invariant blur that results in the simpler structure of final degradation matrix. The better solution of the final fused image can be obtained by using the space variant blur instead of the space invariant. The patchwise use of space variant blur matrix may serve as more accurate model for the image formation and the same can be used to obtain the final fused image in order to get a better fusion.

- We use a model based approach to obtain a final fused image where initial estimate has been obtained with MS and Pan images both or only with MS image. In both the cases we use the linear image formation model to obtain the final fused image. However, it is interesting to model the image formation chain including relations between MS bands, Pan, and an ideal HR MS image, Bayesian inference and these could be applied to extract the optimal matrices while marginalizing over

the unknown HR MS image, yielding an iterative, but more rigorous approach.

Bibliography

[1] G. Joseph, *Fundamental of remote sensing.* 2nd Ed. University Press, 2005.

[2] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa, "Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images," *Inf. Fusion*, vol. 6, no. 3, pp. 193–212, 2005.

[3] L. Wald, "Some terms of reference in data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, 1999.

[4] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 10, no. 45, pp. 3012–3021, Oct. 2007.

[5] C. Pohl and J. L. V. Genderen, "Multi-sensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.

[6] R. A. Schowenderdt, "Remote sensing: Models and methods for image processing," *Second ed. Orlando, FL: Academic*, 1997.

[7] L. Wald, "Data fusion: Definitions and architectures. fusion of images of different spatial resolutions," *Paris, France: Les Presses de lcole des Mines*, p. 197, 2002.

[8] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 81–84, Jun. 2005.

[9] T. Lillesand and R. Kiefer, "Remote sensing and image interpretation," *Third ed., John Wiley and Sons*, 1994.

[10] W. J. Carper, T. M. Lillesand, and R. W. Kiefer, "The use of intensity-hue-saturation transform for merging SPOT panchromatic and multispectral image data." *Photogram. Eng. Remote Sen.*, vol. 56, no. 4, pp. 459–467, 1990.

[11] E. M. Schetselaar, "Fusion by the IHS transform should we use cylindrical of spherical coordinates?" *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 759–765, 1998.

[12] M. Gonzàlez, A. Audicana, J. L. Saleta, R. G. Catalàn, and R. Garcia, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004.

[13] R. Haydn, G. W. Dalke, J. Henkel, and J. E. Bare, "Application of the IHS color transform to the processing of multisensor data and image enhancement," *Proc. Int. Symp. Remote sens. arid semi-arid lands*, pp. 599–607, 1982.

[14] A. H. J. M. Pellemans, R. W. L. Jorddans, and R. Allewiijin, "Merging multispectral and panchromatic spot images with respect to the radimetric properties of the sensor," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 1, pp. 81–87, 1993.

[15] J. Vrabel, "Multispectral imagery band sharpening study," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1075–1083, 1996.

[16] H. Jing and L. Liu, "Color space conversion methods' applications to the image fusion," *Optical Technology*, no. 4, pp. 44–48, 1997.

[17] R. C. Gonzalez and R. E. Woods, "Digital image processing," *Addison-Wesley*, 1992.

[18] P. S. J. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[19] M. Ehlers, "Multisensor image fusion techniques in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 46, pp. 19–30, 1991.

[20] P. Chavez and A. Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, pp. 339–348, 1989.

[21] H. Yesou, Y. Besnus, and Y. Rolet, "Extraction of spectral information from landsat TM data and merger with SPOT panchromatic imagery– a contribution to the study of geological structures," *ISPRS J. Photogramm. Remote Sens.*, vol. 48, no. 5, pp. 23–36, 1993.

[22] A. Eshtehardi, H. Ebadi, M. J. Valadan, and A. Mohammmadzadeh, "Image fusion of Landsat ETM+ and SPOT setellite images using IHS, Brovey and PCA," *ISPRS J. Photogramm. Remote Sens.*, 2007.

[23] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color ehnacement of highly correlated images II. channel ratio and chromaticity transformation technique," *Photogramm. Eng. Remote Sens.*, vol. 22, no. 3, pp. 343–365, 1987.

[24] Guide des Utilisateurs des Donnes SPOT. Anonymous, 1986., CNES and SPOT Image (Eds.), Toulouse, France, 3 volumes, revised January 1991.

[25] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Sens.*, vol. 46, no. 5, pp. 1301–1312, 2008.

[26] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsate TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 3, pp. 743–757, 1998.

[27] V. K. Shettigara, "A generalised component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[28] V. Vijayaraj, C. G. O. Hara, and N. H. Younan, "Quality analysis of pansharpened images," *in Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, pp. 85–88, 2004.

[29] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[30] W. Shi, C. Zhu, Y.Tian, and J. Nichol, "Wavelet-based image fusion and assessment," *Int. J. Appl. Earth Observation Geoinformation*, vol. 6, no. 3-4, pp. 241–251, Mar. 2005.

[31] M. Gonzlez-Audcana, X. Otazu, O. Fors, and J. A. Alvarex-Mozos, "A low computational-cost method to fuse ikonos images using the spectral response function of its sensors," *IEEE Trans. Geosci. Remote. Sens.*, vol. 44, no. 6, pp. 1683–1691, 2006.

[32] T. M. Tu, S. C. Su, H. C. Shyu, and P. S. Huang, "A new look at ihs-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, 2001.

[33] T. M. Tu, P. S. Huang, C. L. Hung, and C. P. Chang, "A fast intensity-hue-saturation fusion techniques with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, 2004.

[34] A. Garzelli and F. Nencini, "Fusion of panchromatic and multispectral images by genetic algorithms," *in Proc. IEEE Int. Geosci. Remote Sens.*, pp. 3810–3813, 2006.

[35] W. Dou, Y. Chen, X. Li, and D. Z. Sui, "A general framework for component substitution image fusion: An implementation using the fast image fusion method," *Comput. Geosci.*, vol. 33, no. 2, pp. 219–228, 2007.

[36] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, 2011.

[37] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, 2010.

[38] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote. Sens.*, vol. 45, no. 10, pp. 3230–3239, 2007.

[39] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1672–1682, Jun. 2006.

[40] T. Tu, W. Cheng, C. Chang, P. S. Huang, and J.-C. Chang, "Best tradeoff for high-resolution image fusion to preserve spatial details and minimize color distortion," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 4, pp. 302–306, 2007.

[41] M. Lillo-Saavedra and C. Gonzalo, "Spectral or spatial quality for fused satellite imagery? A trade-off solution using the wavelet a trous algorithm," *Int. J. Remote Sens.*, vol. 27, no. 7, pp. 1453–1464, 2006.

[42] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," *U.S. Patent 6011875*, pp. 219–228, 2000.

[43] X. Kang, S. Li, and J. A. Benediktsson, "Pansharpening with matting model," *IEEE Trans. Geosci. Remote. Sens.*, vol. 52, no. 8, pp. 5088–5099, 2014.

[44] J. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.

[45] M. M. Khan, J. Chanussot, L. Condat, and A. Montanvert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 98–102, 2008.

[46] T. M. Tu, C. L. Hsu, P. Y. Tu, and C. H. Lee, "An adjustable pansharpening approach for ikonos/quickbird/geoeye-1/worldview-2 imagery," *J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 125–134, 2012.

[47] P. S. Chavez, "Digital merging of Landsat TM and digitized NHAP data 1: 24,000 scale image mapping," *Photogramm. Eng. Remote Sens.*, vol. 52, no. 10, pp. 1637–1647, 1986.

[48] T. Ranchin, L. Wald, and M. Mangolini, "Eflcient data fusion using wavelet transforms: the case of spot satellite images," *In Proc. Int. Symp.Optics, Imaging In-*

strum. Mathematical Imaging: Wavelet Applications in Signal and Image Processing, vol. 2034, pp. 171–178, 1994.

[49] D. A. Yocky, "Image merging and data fusion by means of the discrete two dimensional wavelet transform," *J. Opt. Soc. America: A*, vol. 12, no. 8, pp. 1834–1841, 1995.

[50] B. Garguet-Duport, J. Girel, J. M. Chasseny, and G. Pautou, "The use of multiresolution analysis and wavelet transform for merging spot panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1057–1066, 1996.

[51] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, 2000.

[52] D. A. Yocky, "Multiresolution wavelet decomposition image merger of landsat thematic mapper and spot panchromatic data," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1067–1074, 1996.

[53] M. Gonzalez-Audicana, X. Otazu, O. Fors, and A. Seco, "Comparison between Mallat's and the 'a trous' discrete wavelet transform based algorithms for the fusion of fusion multispectral and panchromatic images." *Int. J. Remote Sens.*, vol. 3, no. 3, pp. 595–614, 2005.

[54] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

[55] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, , and M. Selva, "MTF-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.

[56] J. Nunez, X. otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.

[57] R. Kronland-Martinet, J. Morlet, and A. Grossman, "Analysis of sound patterns through wavelet transforms," *Int. J. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 273–301, 1987.

[58] P. Dutilleux, "An implementation of the algorithme a trous to compute the wavelet transform," *In: Combes, J.M., Grossman, A., Tchamitchian, Ph. (Eds.), Wavelets: Time Frequency Methods and Phase Space. Springer, Berlin*, pp. 298–304, 1989.

[59] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *IEEE Trans. Signal Proces.*, vol. 40, no. 10, pp. 2464–2482, 1992.

[60] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, 2005.

[61] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion - the ARSIS concept and some successful implementation schemes," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1-2, pp. 4–18, Jun. 2003.

[62] A. P. Bradley, "Shift-invariance in the discrete wavelet transform," *in Proc. 8th Digital Image Comput.: Techniques and Appl.*, pp. 29–38, 2003.

[63] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, 1983.

[64] B. Aiazzi, L. Alparone, S. Baronti, and R. Carl, "A pyramid approach to fusion of landsat tm and spot-pan data to yield multispectral high-resolution images for environmental archaeology," *in Proc. SPIE Remote Sensing for Geography, Geology, Land Planning and Cultural Heritage, vol. 2960, EUROPTO Series*, pp. 153–162, 1996.

[65] T. A. Wilson, S. K. Rogers, and M. Kabrisky, "Perceptual based image fusion for hyperspectral data," *Photogramm. Eng. Remote Sens.*, vol. 35, pp. 1007–1017, 1997.

[66] B. Aiazzi, L. Alparone, S. Baronti, V. Cappellini, R. Carl, and L. Mortelli, "A laplacian pyramid with rational scale factor for multisensor image data fusion," *In*

Proc. Int. Conf. Sampling Theory and Applications-SampTA, vol. 97, pp. 55–60, 1997.

[67] B. Aiazzi, L. Alparone, S. Baronti, and I. Pippi, "Fusion of 18 m moms-2p and 30 m landsat tm multispectral data by the generalized laplacian pyramid," *In Proc. ISPRS Int. Arch. Photogramm. Remote Sens.*, vol. 32, no. 7-4-3W6, pp. 116–122, 1999.

[68] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi, "Multispectral fusion of multisensor image data by the generalized laplacian pyramid," *In Proc. Geo. Remote Sens. Symp.*, pp. 1183–1185, 1999.

[69] P. P. Vaidyanathan, "Multirate systems and filter banks," *Englewood Cliffs, NJ: Prentice-Hall*, 1992.

[70] V. T. Tom, "System for and method of enhancing images using a multiband information," *US Patent 4683496*, 1987.

[71] P. Blanc, T. Blu, T. Ranchin, L. Wald, , and R. Aloisi, "Using iterated rational filter banks within the ARSIS concept for producing 10m landsat multispectral images," *Int. J. Remote Sens.*, vol. 19, no. 2, pp. 2331–2343, 1998.

[72] A. Garzelli and F. Nencini, "Interband structure modeling for pan-sharpening of very high-resolution multispectral images," *Int. J. Inf. Fusion.*, vol. 6, pp. 213–224, 2005.

[73] W. Z. Shi, C. Q. Zhu, C. Zhu, and X. M.Yang, "Multi-bandwavelet for fusing spot panchromatic and multi-spectral images," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 5, pp. 513–520, 2003.

[74] D. C. Tseng, Y. L. Chen, and S. C. Liu, "Wavelet-based multispectral image fusion," *In Proc. Geo. Remote Sens. Symp.*, pp. 1956–1958, 2001.

[75] R. L. King and J. W. Wang, "A wavelet based algorithm for pan sharpening landsat 7 imagery," *In Proc. Geo. Remote Sens. Symp.*, pp. 849–851, 2001.

[76] Y. Chibani and A. Houacine, "The joint use of ihs transform and redundant wavelet decomposition for fusing multispectral and panchromatic images," *Int. J. Remote Sens.*, vol. 23, pp. 3821–3833, 2002.

[77] G. Hong and Y. Zhang, "High resolution image fusion based on wavelet and IHS transformations," *In Proceedings of the IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Berlin*, pp. 99–104, 2003.

[78] Y. Kim, C. Lee, D. Han, and Y. Kim, "Improved additive-wavelet image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 263–267, 2011.

[79] E. J. Candes and D. L. Donoho, "Curvelets-A surprisingly effective nonadaptive representation for objects with edges," *in Curve and Surface Fitting : Saint-Malo*, A. Cohen, C. Rabut, and L.L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press, 1999.

[80] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, 2005.

[81] A. L. da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, 2006.

[82] M. Choi, R. Y. Kim, M. R. Nam, and H. Kim, "Fusion of multispectral and panchromatic satellite images using the curvelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 2, no. 2, pp. 136–140, Apl. 2005.

[83] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Information Fusion*, vol. 8, no. 2, pp. 143–156, 2007.

[84] M. Do and M. Vetterli, "Framing pyramids," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2329–2342, 2003.

[85] H. R. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, 1992.

[86] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, 2008.

[87] H. Shi, B. Tian, and Y. Wang, "Fusion of multispectral and panchromatic satellite images using principal component analysis and nonsubsampled contourlet transform," *in Proc. Seventh Int. Conf. Fuzzy Syst. Knowledge Discovery (FSKD)*, pp. 2313–2315, 2010.

[88] A. G. Mahyari and M. Yazdi, "Panchromatic and multispectral image fusion based on maximization of both spectral and spatial similarities," *IEEE Trans. Geosci. Sens.*, vol. 49, no. 6, pp. 1976–1985, 2011.

[89] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, p. 123151, 2005.

[90] F. C. Fernandes, R. L. van Spaendonck, and C. S. Burrus, "A new framework for complex wavelet transforms," *IEEE Trans. Signal Process.*, vol. 51, no. 7, p. 18251837, 2003.

[91] ——, "Multidimensional, mapping-based complex wavelet transforms," *IEEE Trans. Image Process.*, vol. 14, no. 1, p. 110124, 2005.

[92] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," *in International Conference on Image Processing (ICIP)*, vol. 3, no. 3, p. 34443447, 1995.

[93] T. T. Nguyen and S. Oraintara, "The shiftable complex directional pyramid-part i: theoretical aspects," *IEEE Trans. Image Process.*, vol. 56, no. 10, p. 46514660, 2008.

[94] ——, "The shiftable complex directional pyramid-part ii: Implementation and applications," *IEEE Trans. Image Process.*, vol. 56, no. 10, p. 46614672, 2008.

[95] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, p. 4970, 2000.

[96] S. Ioannidou and V. Karathanassi, "Investigation of the dual-tree complex and shift-invariant discrete wavelet transforms on quickbird image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 1, p. 166170, 2007.

[97] L. W. Q. Zhang, H. Li, and Z. Ma, "Similarity-based multimodality image fusion with shiftable complex directional pyramid," *Pattern Recognit. Lett.*, vol. 32, no. 13, p. 15441553, 2011.

[98] J. Saeedi and K. Faez, "A new pan-sharpening method using multiobjective particle swarm optimization and the shiftable contourlet transform," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, p. 365381, 2011.

[99] B. Luo, M. M. Khan, T. Bienvenu, J. Chanussot, and L. Zhang, "Decision-based fusion for pansharpening of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 19–23, 2013.

[100] D. Fasbender, J. Radoux, and P. Bogaert, "Bayesian data fusion for adaptable image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1847–1857, Jun. 2008.

[101] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, 2004.

[102] M. V. Joshi, L. Bruzzonne, and S. Chaudhuri, "A model based approach to multiresolution fusion in remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2549–2562, Sep. 2006.

[103] M. V. Joshi and A. Jalobeanu, "MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1245–1255, Mar. 2010.

[104] H. Aanaes, J. R. Sveinsson, A. A. Nielsen, T. Bϕvith, and J. A. Benediktsson, "Model based satellite image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, May 2008.

[105] Z. Li and H. Leung, "Fusion of multispectral and panchromatic images using a restoration-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1482–1491, May 2009.

[106] Z. Zhou, S. Peng, B. Wang, Z. Hao, and S. Chen, "An optimized approach for pansharpening very high resolution multispectral images," *IEEE Trans. Geosci. Lett.*, vol. 9, no. 4, pp. 735–739, July 2012.

[107] L. Zhang, H. Shen, W. Gong, and H. Zhang, "Adjustable model-based fusion method for multispectral and panchromatic images," *IEEE Trans. Syst. Man Cybern. B*, vol. 42, no. 6, pp. 1693–1704, 2012.

[108] H. A. Aly and G. Sharma, "A regularized model-based optimization framework for pan-sharpening," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2596–2608, 2014.

[109] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259 – 268, 1992.

[110] C. Ballester, V. Casellesa, L. Igual, and J. Verdera, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," *Int. J. of Computer Vision*, vol. 69, no. 1, pp. 43–58, 2006.

[111] M. Moeller, T. Wittman, A. Bertozzi, and M. Burger, "A variational approach for sharpening high dimensional images," Ph.D. dissertation, Inst. Computational & Applied Math., Westfalische Wilhelms Universitat Munster, Mnster, Germany, 2009.

[112] F. Palsson, J. Sveinsson, and M. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, 2014.

[113] X. He, L. Condat, J. Chanussot, and J. Xia, "Pansharpening using total variation regularization," *In Proc. Geo. Remote Sens. Symp.*, pp. 166–169, 2012.

[114] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *In Proc. Geo. Remote Sens. Symp.*, vol. 23, no. 9, pp. 4160–4174, 2014.

[115] F. Fang, F. Li, C. Shen, and G. Zhang, "A variational approach for pan-sharpening," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2822–2834, 2013.

[116] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *J. Sci. Comput.*, vol. 42, pp. 185–197, 2010.

[117] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model Simul.*, vol. 7, pp. 1005–1028, 2008.

[118] J. Duran, A. Buades, B. Coll, and C. Sbert, "A nonlocal variational model for pansharpening image fusion," *SIAM J. Imaging Sciences*, vol. 7, no. 2, pp. 761–796, 2014.

[119] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 987–995, 1996.

[120] P. L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1213–1222, 2004.

[121] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Process.*, vol. 1, no. 3, pp. 322–336, 1992.

[122] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.

[123] V. A. Morozov, "Methods for solving incorrectly posed problems," *Springer-Verlag, New York*, 1984.

[124] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization-minimization approach," *Signal Process.*, vol. 89, no. 9, pp. 1683–1693, 2009.

[125] Y.-W. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1770–1781, 2012.

[126] D. Liang, H. Wang, Y. Chang, and L. Ying, "Sensitivity encoding econstruction with nonlocal total variation regularization," *Magnetic Resonance in Medicine*, vol. 65, pp. 1384–1392, 2011.

[127] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: a unified variational framework," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3192 – 3203, 2013.

[128] H. Chang, X. Zhang, X.-C. Tai, and D. Yang, "Domain decomposition methods for nonlocal total variation image restoration," *J Sci Comput*, vol. 60, pp. 79–100, 2014.

[129] S. Li and B. Yang, "A new pansharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, 2011.

[130] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing based pan sharpening method," *IEEE Trans. Geosci. Lett.*, vol. 9, no. 4, pp. 629–633, 2012.

[131] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Sens.*, vol. 51, no. 9, pp. 4779–4789, 2013.

[132] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, 2013.

[133] M. Cheng, C. Wang, and J. Li, "Sparse representation based pansharpening using trained dictionary," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 293–297, Jan. 2014.

[134] M. Iqbal, L. Chen, X.-Z. Wen, and C.-S. Li, "Remote sensing image fusion using best bases sparse representation," *in Proc. IEEE Int. Geosci. Remote Sens. Symp.*, pp. 5430–5433, 2012.

[135] V. Harikumar, P. P. Gajjar, M. V. Joshi, and M. S. Raval, "Multiresolution image fusion: Use of compressive sensing and graph cuts," *Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 5, pp. 1771–1780, 2014.

[136] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 446–453, 2011.

[137] B. Aiazzi, S. Baronti, M. Selva, and L. Alparone, "Bi-cubic interpolation for shift-free pan-sharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 65–76, 2013.

[138] Y. Zhang, "Understanding image fusion," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 6, pp. 657–661, 2004.

[139] Q. Du, N. Younan, R. King, and V. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geosci. and Remote Sens. Lett.*, vol. 4, no. 4, pp. 518 – 522, 2007.

[140] Q. Du, O. Gungor, and J. Shan, "Performance evaluation for pansharpening techniques," *in Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 6, pp. 4264–4266, 2005.

[141] G. Vivone, L. Alparone, J. Chanussot, M. D. Mura, A. Garzelli, G. Licciardi, R. Restaino, and L. Wald, "A critical comparison of pansharpening algotithms," *In Proc. Geo. Remote Sens. Symp.*, 2014.

[142] W. Ha, P. H. Gowda, and T. A. Howell, "A review of potential image fusion methods for remote sensing-based irrigation management: part II," *Irrigation Science*, vol. 31, no. 4, pp. 851–869, 2013.

[143] J. Zhang, "Multi-source remote sensing data fusion: Status and Trends," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010.

[144] G. Vivone, L. Alparone, J. Chanussot, M. D. Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[145] T. Stathaki, *Image Fusion: Algorithms and Applications.* Academic Press, 2008.

[146] H. B. Mitchell, *Image Fusion: Theories, Techniques and Applications.* Springer, 2010.

[147] R. S. Blum and Z. Liu, *Multi-Sensor Image Fusion and Its Applications.* CRC Press, 2005.

[148] S. Chaudhuri and K. Kotwal, *Hyperspectral Image Fusion.* Springer, 2013.

[149] S. Li, J. T. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge landsat tm and spot panchromatic images," *Inf. Fusion*, vol. 3, no. 1, pp. 17–23, 2002.

[150] J. Hu and S. Li, "Fusion of panchromatic and multispectral images using multiscale dual bilateral filter," in *IEEE Inf. Conf. Image Proc.*, 2011.

[151] J. T. S. Paris, P Kornprobst, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, 2009.

[152] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

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

[154] Sep. 2012. [Online]. Available: http://www.isprs.org/data/ikonos_hobart/default.aspx

[155] Sep. 2012. [Online]. Available: http://www.digitalglobe.com/product-samples

[156] Feb. 2014. [Online]. Available: http://www.glcf.umd.edu/data/quickbird/sundarbans.shtml

[157] May 2013. [Online]. Available: http://www.digitalglobe.com/downloads/microsites/dfc/optical/WV2_9Oct2011.zip

[158] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, 2008.

[159] D. Rajan and S. Chaudhuri, "Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1102–1117, 2003.

[160] A. N. Rajagopalan and S. Chaudhuri, "An MRF based approach to simultaneous recovery of depth and restoration from defocussed images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 7, pp. 577–589, Jul. 1999.

[161] M. Do and M. Vetterli, "Framing pyramids," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2329–2342, 2003.

[162] Y. Shi, X. Yang, and Y. Guo, "Translation invariant directional framelet transform combined with gabor filters for image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 44–55, Jan. 2014.

[163] E. Nadernejad, S. Sharifzadeh, and H. Hassanpour, "Edge detection techniques: Evaluations and comparisons," *Applied Mathematical Sciences*, vol. 2, no. 31, pp. 1507–1520, 2008.

[164] M. C. Shin, D. B. Godlgof, K. W. Bowyer, and S. Nikiforou, "Comparison of edge detection algorithms using a structure from motion task," *IEEE Trans. Syst. Man Cybern. B*, vol. 31, no. 4, pp. 589–601, Aug. 2001.

[165] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, 1994.

[166] V. N. Katsikis and D. Pappars, "Fast computing of the moore-penrose inverse matrix," *Electronic Journal of Linear Algebra: A publication of the International Linear Algebra Society*, vol. 17, pp. 637–650, Nov. 2008.

[167] S. Z. Li, *Markov Random Field Modeling in Computer Vision.* Tokyo: Springer-Verlag, 1995.

[168] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," *in Proc. Ninth IEEE Int.Conf. Comput. Vis.*, pp. 1150–1159, 2003.

[169] S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution and the bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, Nov. 1984.

[170] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, 1996.

[171] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2207–2232, 1993.

[172] S. M. Phoong, C. W. Kim, P. P. Vaidyanathan, and R. Ansari, "A new class of two-channel biorthogonal filter banks and wavelet bases," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 649–665, 1993.

[173] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.

[174] C. Chang, "Spectral information divergence for hyperspectral image analysis," *in Proc. Geosci. Remote Sens. Symp.*, vol. 1, pp. 509–511, 1999.

[175] A. G. Mahyari and M.Yazdi, "Fusion of panchromatic and multispectral images using temporal Fourier transform," *IET Image Process.*, vol. 4, no. 4, pp. 255–260, 2010.

[176] M. Xu, H. Chen, and P. K. Vershney, "An image fusion approach based on markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5116–5127, Dec. 2011.

[177] J. S. Bhatt, M. V. Joshi, and M. S. Raval, "A data-driven stochastic approach for unmixing hyperspectral imagery," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1936 – 1946, 2014.

[178] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," *IEEE Int. Conf. Computer Vision*, pp. 349–356, 2009.

[179] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision ECCV 2014*, ser. Lecture Notes

in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8692, pp. 184–199.

[180] S. Arya and D. M. Mount, "Approximate nearest neighbor queries in fixed dimensions," in *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1993, pp. 271–280.

[181] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[182] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[183] R. G. Baraniuk, "Lecture notes: Compressed sensing," *IEEE Signal Process. Mag.*, pp. 118–124, 2007.

[184] J.-K. Kamarainen, V. Kyrki, and H. Klviinen, "Invariance properties of gabor filter-based featuresoverview and applications," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1088–1099, 2006.

[185] T. P. Weldon, W. E. Higgins, and D. F. Dunn, "Efficient gabor filter design for texture segmentation," *Pattern Recognition*, vol. 29, no. 12, pp. 2005–2015, 1996.

[186] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion?" *Proc. Int. Conf. Fusion Earth Data*, pp. 46–61, 2000.

[187] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Trans. Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.

[188] R. H. Yuhas, A. F. H. Goetz, and J. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," *in Proc. Summar. 4th JPL Airborne Earth Sci. Workshop*, pp. 147–149, 1992.

[189] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Sens.*, vol. 47, no. 11, pp. 3880–3891, 2009.

[190] E. Candes and J. Romberg, "$l1$-magic: Recovery of sparse signal via convex programmingm." [Online]. Available: http://www.l1-magic.org.

List of Publications

- **International Journals**

1. K. P. Upla, P. P. Gajjar, and M. V. Joshi, "An Edge Preserving Multi-resolution Fusion: Use of Contourlet Transform and MRF Prior", **IEEE Trans. on Geoscience and Remote Sensing** Vol. 53 No. 6, pp. 3210–3220, May 2015
2. K. P. Upla, M. V. Joshi, Sharad Joshi and P. P. Gajjar,, "Multi-resolution Image Fusion using Edge Preserving Filters", **SPIE Journal of Remote Sensing** Vol. 9 No. 1 PP.096025-26, July 2015

- **International Conferences**

1. K. P. Upla, P. P. Gajjar, and M. V. Joshi, "Multiresolution fusion using contourlet transform based edge learning," IEEE Int. Geosci. Remote Ses. Symp. (IGARSS), pp. 523-526, 2011.
2. K. P. Upla, P. P. Gajjar, M. V. Joshi, A. Banerjee, and V. Singh, "A fast approach for edge preserving super-resolution," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), pp. 1-6, 2011.
3. R. C. Patel, P. P. Gajjar, M. V. Joshi, and K. P. Upla, "A model based approach to multiresolution image fusion: Use of CS theory," IEEE Int. Geosci. Remote Ses. Symp. (IGARSS), 2012.
4. K. P. Upla, P. P. Gajjar, and M. V. Joshi, "Pan-sharpening: Use of Difference of Gaussians", IEEE Int. Geosci. Remote Ses. Symp. (IGARSS), pp. 4922-4925, 2014.

- **National Conferences**

1. K. P. Upla, P. P. Gajjar, and M. V. Joshi, Pan-sharpening based on non-subsampled contourlet transform detail extraction, in Proc. of Fourth IEEE Natinal conf. on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), IIT Jodhpur, pp. 1-4, December 2013.
2. S. Joshi, K. P. Upla, and M. V. Joshi, Multi-resolution image fusion using multistage guided filter, in Proc. of Fourth IEEE Natinal conf. on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), IIT Jodhpur, pp. 1-4, December 2013.