

SPARSE AND VARIATIONAL MODELS FOR PAN-SHARPENING OF MULTISPECTRAL IMAGES

Thesis

Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by
RAJESH GOGINENI



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING,
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,
SURATHKAL, MANGALORE -575025

JULY, 2020

DECLARATION

I hereby *declare* that the Research Thesis entitled **SPARSE AND VARIATIONAL MODELS FOR PAN-SHARPENING OF MULTISPECTRAL IMAGES** which is being submitted to the *National Institute of Technology Karnataka, Surathkal* in partial fulfillment of the requirement for the award of the Degree of *Doctor of Philosophy* in **Department of Electronics and Communication Engineering** is a *bonafide report of the research work carried out by me*. The material contained in this research thesis has not been submitted to any University or Institution for the award of any degree.



RAJESH GOGINENI,
Reg. No.: 155012/EC15F06
Department of Electronics and
Communication Engineering.

Place: NITK-Surathkal.

Date:

CERTIFICATE

This is to certify that the Research Thesis entitled **Sparse and Variational Models for Pan-sharpening of Multispectral Images** submitted by **RAJESH GOGINENI** (Register Number: 155012EC15F06) as the record of the research work carried out by him, is accepted as the *Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy**.

Dr. Ashvini Chaturvedi.

Research Guide

Professor

Dept. of E & C Engg.

NITK Surathkal - 575025

Chairman - DRPC

(Signature with Date and Seal)

Acknowledgements

I want to express my sincere gratitude to my research supervisor Dr. Ashvini Chaturvedi for the continuous support of my Ph.D. study and related research for his patience, motivation, and immense knowledge. His advice on both research as well as on my career has been priceless.

I express my gratitude to Dr.Laxminidhi T, Head, Department of E and C Engineering, for his support, help, and encouragement.

I am grateful to my RPAC members, Prof. B.R. Shankar, Dept. of Mathematical and Computational Sciences and Prof. U. Shripathi Acharya, Dept. of E and C Engineering for giving vital comments and suggestions throughout the research, which helped in improving the quality of research. I want to thank Prof. M. S.Bhat and all the other faculty members and staff of E and C department, NITK Surathkal for their assistance. I want to thank B.S Daya Sagar, Professor and Formar Head, SSI Unit, ISI, Bangalore centre, for his insightful comments, suggestions, and support.

I want to express my gratitude to all friends and colleagues at NITK for encouraging me in the good and the bad times making a memorable stay in NITK. Thanks to Raghavendra MANS, Vasudeva reddy, Jayaram reddy, Shilpa suresh, Prasad naik R, BalaNarasimha G, Bharath M, Venkatadri Tekuri, Raghavendra P. Reddy and Sai krishna for the support during my research work.

I express my gratitude to the Ministry of Human Resource Development (MHRD), Government of India for providing financial assistance during the research. A very special gratitude goes out to Technical Education Quality Improvement Programme (TEQIP), to help and provide the funding for the research work.

I am grateful to my maternal uncle M.V. Seshaiah, my grandmother M.Tulasamma, and my wife G.Ramadevi, who have provided me with moral and emotional support in my life. I am also grateful to my other family members and friends who have supported me along the way.

And also to everyone in the Wireless Sensor Networks Lab, NITK. . . . it was excellent sharing laboratory with all of you during the last five years. Thanks for all your encouragement!

Finally, I would like to thank god for giving me good health, strength and bliss during my research work.

This Thesis is Dedicated to
My Grand Mother **Smt.M.Tulasamma,**
My Maternal Uncle **Sri M.V.Seshaiah,**
My Wife **Ramadevi** and My Son **Rishi Teja.**

Abstract

Remote sensing is defined as a mechanism facilitating the measurement of object's features on the earth's surface through the data obtained from platforms such as aircraft and satellites. Remote sensing provides the observation, mapping, analysis, and management of various resources present on the earth. In the past few decades, the tremendous progress in remote sensing technologies has enriched the techniques of acquisition, processing, and analysis of acquired data. The imaging data collected by the satellite sensors can be characterized using features like spatial resolution, spectral resolution, radiometric resolution, and temporal resolution. This thesis investigates the optical images whose spectral range spans visible and near-infrared (NIR) regions of the electromagnetic (EM) spectrum.

The images with high-spatial and high-spectral resolution are of immense interest for various remote sensing applications like land mapping, change detection, and object recognition. The current generation satellite sensors, namely QuickBird, IKONOS, WorldView, GeoEye, etc., incur constraints such as the trade-off between spatial and spectral resolutions, limited on-board storage of satellite platform, moderate signal to noise ratio of received signal energy. The afore-mentioned commercial satellites usually produce two kinds of images; Panchromatic (PAN) image with high spatial and low spectral resolution and multispectral (MS) image with high spectral and low spatial resolution characteristics. The physical and technological limitations of sensors prohibit the objective of achieving an image with the finest spatial and spectral resolution.

Pan-sharpening (PS) is a remote sensing image fusion method that produces a high-resolution multispectral (HRMS) image by synthesizing the low-resolution MS image with a corresponding high-resolution PAN image. The requirement of pan-sharpened data is steadily increasing, driven by the consistent diffusion of commercial products using high-resolution images like Google Earth and Bing Maps. To date, different classes of pan-sharpening methods such as component substitution, multi resolution analysis, and model based methods have been developed. Most of the conventional PS methods induce spectral distortion and spatial artifacts in the fused image. Further, there is a demand for an efficient fusion technique that yields a pan-sharpened image with balanced spatial and spectral qualities.

This research concentrates on developing pan-sharpening techniques using a sparse representation mechanism. In PS problems, the fused image is obtained by imparting the missing spatial features extracted from the PAN image into the MS image bands. The sparse representation (SR) based PS methods exploit the sparse nature of spatial

details using an appropriate basis, usually termed as a dictionary. The construction of a pertinent dictionary that promotes the sparsity of PAN and MS images is the fundamental task in SR based pan-sharpening problems. Motivated by the existing SR based techniques, PS methods based on two different dictionaries, namely dual dictionary and multi-scale dictionary are proposed in this thesis. To cope with the computational complexity realized by the large-sized satellite images, the SR based methods adapt patch-based strategies. The PS mechanism is implemented on overlapped patches extracted from the source images. The overlapping mechanism results in redundant and inconsistent image features in the fused image. An alternative mechanism termed as convolutional sparse representation (CSR) is deployed to deal with the drawbacks in patch-based pan-sharpening techniques. The CSR based methods are robust to misregistration between source images and produce the pansharpened image with enhanced spatial and spectral features. In addition to these methods, a variational pan-sharpening scheme is developed in this thesis to preserve the spatial details and to reduce the spectral distortion. The pan-sharpening process is formulated as a constrained optimization function using the appropriate and reliable prior terms. The developed optimization problem is solved using a suitable minimization algorithm yields a pan-sharpened image.

The proposed methods are evaluated using the datasets obtained over different geographical terrains. The experimentation is performed at full-scale and reduced-scale resolutions as specified by the Wald's and QNR protocols. The pan-sharpening techniques developed in this thesis are validated using visual and quantitative evaluation.

Keywords: Pan-sharpening; High-resolution multispectral image; Sparse representation; Dictionary learning; Convolutional sparse coding; Variational method.

TABLE OF CONTENTS

Acknowledgement	i
Abstract	iv
List of Figures	xi
List of Tables	xvi
Nomenclature	xviii
Abbreviations	xviii
1 INTRODUCTION	1
1.1 Pan-sharpening	4
1.2 Sparse Representation Mechanism	5
1.3 Dictionary Learning Mechanism	7
1.3.1 Multi-Scale Dictionary Learning Using Wavelets	9
1.4 Overview of Pan-sharpening Paradigms	10
1.5 Research Objectives	12
1.6 Structure of Thesis	13
2 QUALITY ASSESSMENT OF PAN-SHARPENED IMAGES AND SPECIFICATIONS OF DATASETS	15
2.1 Quality Metrics	16
2.2 Protocols for Pan-sharpening	18
2.2.1 Wald’s Protocol	19
2.2.2 QNR Protocol	19
2.3 Datasets	22

3	SPARSE REPRESENTATION DRIVEN TWO-STAGE PAN-SHARPENING ALGORITHM	24
3.1	A Review of Sparsity Based Pan-sharpening Methods	24
3.2	Two-Stage Pan-sharpening Algorithm	27
3.2.1	The Proposed Pan-sharpening Framework	28
3.2.2	Spectral Component (\widehat{X}_k^L) Estimation	28
3.2.3	Spatial Component (\widehat{X}_k^H) Estimation	29
3.2.4	Dual Dictionary Learning Mechanism	32
3.2.5	K-SVD Versus PAU-DL Training Algorithm Analysis	32
3.3	Results and Analysis	33
3.3.1	Evaluation Using Synthesis Property	35
3.3.2	Evaluation Using Consistency Property	40
3.3.3	Evaluation Using QNR Protocol	42
3.4	Conclusion	45
4	SPARSITY INSPIRED PAN-SHARPENING TECHNIQUE USING MULTI-SCALE LEARNED DICTIONARY	46
4.1	Multi-Scale Learned Dictionary	47
4.2	Pan-sharpening Using Multi-Scale Learned Dictionary (PS-MSLD)	48
4.2.1	HR Dictionary Construction	49
4.2.2	LR Dictionary Construction	50
4.2.3	Sparse Coefficients Estimation	51
4.2.4	Pan-sharpened Image Reconstruction	52
4.2.5	Pseudo code for PS-MSLD algorithm	53
4.3	Experimental Results	55
4.3.1	Experimental Results at Reduced-Scale	56
4.3.1.1	Experimental Results Using QuickBird Dataset	58

4.3.1.2	Experimental Results With IKONOS Dataset	59
4.3.1.3	Experimental Results Using Pléiades Dataset	61
4.3.2	Experimental Results at Full-Scale	63
4.4	Conclusion	66
5	PAN-SHARPENING ALGORITHMS BASED ON CONVOLUTIONAL SPARSE REPRESENTATION	67
5.1	Preliminaries	69
5.1.1	Convolutional Sparse Coding Mechanism	69
5.1.2	Sparse Representation Versus Convolutional Sparse Representation	70
5.1.3	Cartoon and Texture Decomposition	71
5.2	A Hybrid Pan-Sharpening Algorithm	75
5.2.1	Fusion of Texture Components	76
5.2.2	Fusion of Cartoon Components	77
5.2.3	Pan-sharpened Image Reconstruction	78
5.3	Results and Discussions	80
5.3.1	Evaluation Using Synthesis Property	81
5.3.2	Evaluation Using Consistency Property	84
5.3.3	Evaluation Using QNR Protocol	86
5.3.4	Analyzing Impact of Number of Filters	88
5.3.5	Proposed Algorithm's Execution Time Analysis	88
5.4	Pan-Sharpening Framework With Combined DWT and CSR	90
5.4.1	Decomposition Using Discrete Wavelet Transform	91
5.4.2	Fusion of Approximate Coefficient Bands	92
5.4.3	Fusion of Detail Coefficient Bands	94
5.5	Results and Analysis	95

5.5.1	Experimental Results With QuickBird Dataset	95
5.5.2	Experimental Results With IKONOS Dataset	98
5.5.3	Experimental Results With Tolouse (Pléiades) Dataset	100
5.6	Conclusion	103
6	A VARIATIONAL PAN-SHARPENING ALGORITHM TO ENHANCE THE SPECTRAL AND SPATIAL DETAILS	105
6.1	Preliminaries	107
6.1.1	Total Generalized Variation (TGV)	107
6.1.2	Spectral Angle Mapper (SAM)	108
6.2	Proposed Variational Model	109
6.2.1	Data Synthesizing Fidelity Term	109
6.2.2	Spatial Details Preserving Term	110
6.2.3	Inter-Band Correlation Preserving Term	111
6.2.4	Optimization Method	112
6.3	Results and Analysis	114
6.3.1	Parameters Selection	115
6.3.2	Reduced-Resolution Assessment	116
6.3.3	Full-Resolution Assessment	119
6.3.4	Comparative Analysis of Computational Cost	121
6.4	Conclusion	123
7	CONCLUSIONS AND FUTURE WORK	124
7.1	Conclusions	124
7.2	Future Work	126
	Bibliography	127

LIST OF FIGURES

1.1	Visible and InfraRed spectral regions(expanded) of an electro-magnetic spectrum	1
1.2	PAN, MS and Pan-sharpened images (WorldView-2)	4
1.3	Sparse representation schematic.	6
2.1	Assessment process of Wald’s synthesis property	20
3.1	Pan-sharpening results evaluated using synthesis property for Quick-Bird data (a) PAN image (200x200, 2.8-m) (b) Up-sampled MS image (200x200, 11.2-m) (c) Reference MS Image (d) FIHS (e) AWLP (f) SR-Li (g) SR-TD (h) SR-WT (i) SR-D (j) Proposed method	36
3.2	Difference between pan-sharpened images for each method and reference MS image in Fig.3.1. (Red colored pixels indicates large differences and black pixels indicate less difference) (a) FIHS (b) AWLP (c) SR-Li (d) SR-TD (e) SR-WT (f) SR-D (g) Proposed method	37
3.3	Pan-sharpening results evaluated using synthesis property for WorldView-2 data (a) PAN image (200×200, 2-m) (b) Up-sampled MS image (200×200, 8-m) (c) Reference MS Image (d) FIHS (e) AWLP (f) SR-Li (g) SR-TD (h) SR-WT (i) SR-D (j) Proposed method	38
3.4	Difference between pan-sharpened images for each method and reference MS image in Fig.3.3. (Bright pixels means large difference and black mens less difference)(a) FIHS (b) AWLP (c) SR-Li (d) SR-TD (e) SR-WT (f) SR-D (g) Proposed method	39
3.5	Pan-sharpening results evaluated using consistency property for IKONOS data (a) PAN image (256×256, 1-m) (b) Up-sampled MS image (256×256, 4-m) (c) FIHS (d) AWLP (e) SR-Li (f) SR-TD (g) SR-WT (h) SR-D (i) Proposed method.	41

3.6	Pan-sharpening results evaluated using QNR protocol for QuickBird data (a) PAN image (256×256 , 0.7m) (b) Up-sampled MS image (256×256 , 2.8-m) (c) FIHS (d) AWLP (e) SR-Li (f) SR-TD (g) SR-WT (h) SR-D (i) Proposed method	42
3.7	Pan-sharpening results evaluated using QNR protocol for Pléiades data (a) PAN image (256×256 , simulated) (b) Up-sampled MS image (256×256 , 60-cm) (c) FIHS (d) AWLP (e) SR-Li (f) SR-TD (g) SR-WT (h) SR-D (i) Proposed method	44
4.1	The block-diagram of PS-MSLD	52
4.2	Schematic diagram for dictionary learning	53
4.3	Pan-sharpening results with QuickBird data at reduced- scale (a) Degraded PAN image (512×512 , 2.8-m) (b) original MS image (2.8-m) (c) EXP (upsampled MS image) (d) GIHS (e) SW (f) GSA (g) MGC (h) AWLP (i) SR-TD (j) SR-WT (k) SR-D (l) PS-MSLD	58
4.4	Pan-sharpening results with IKONOS data at reduced-scale (a) PAN image (256×256 , 4-m) (b) Reference MS image (256×256 , 4-m) (c) EXP (upsampled MS image) (d) GIHS (e) SW (f) GSA (g) MGC (h) AWLP (i) SR-TD (j) SR-WT (k) SR-D (l) PS-MSLD	60
4.5	Pan-sharpening results with pléiades data at reduced-scale (a) PAN image (b) MS image (c) EXP (d) GIHS (e) SW (f) GSA (g) MGC (h) AWLP (i) SR-TD (j) SR-WT (k) SR-D (l) PS-MSLD	62
4.6	Pan-sharpening results with WorldView-2 data at full-scale (true color images (5-3-2)) a) PAN image b) Up-sampled MS image (EXP) (c) GIHS (d) SW (e) GSA (f) MGC (g) AWLP (h) SR-TD (i) SR-WT (j) SR-D (k) PS-MSLD	64
4.7	Pan-sharpening results with QuickBird data at full-scale (false color images(4-2-3)) a) PAN image (512×512 , 0.7-m) b) up-sampled MS image (EXP) (512×512 , 2.8-m) (c) GIHS (d) SW (e) GSA (f) MGC (g) AWLP (h) SR-TD (i) SR-WT (j) SR-D (k) PS-MSLD	65
5.1	Sparse representation schematic.	70
5.2	Convolutional sparse representation schematic.	71

5.3	The proposed pan-sharpening algorithm schematic.	79
5.4	Visual results of pan-sharpening methods for Pléiades data at reduced-scale (a) Degraded PAN image (b) Up-sampled MS image (EXP) (c) Reference MS image (d) IHS (e) AWLP (f) MTF-GLP (g) SR-Li (h) SR-TD (i) SR-D (j) SR-LD (k) SR-CD (l) Proposed Method	82
5.5	Visual results of pan-sharpening methods for IKONOS data at reduced-scale (a) Degraded PAN image (b) Up-sampled MS image (EXP) (c) Reference MS image (d) IHS (e) AWLP (f) MTF-GLP (g) SR-Li (h) SR-TD (i) SR-D (j)SR-LD (k) SR-CD (l) Proposed Method	83
5.6	Visual results of pan-sharpening methods for QuickBird data at full-scale (a) PAN image (b) Up-sampled MS image (EXP) (c) IHS (d) AWLP (e) MTF-GLP (f) SR-Li (g) SR-TD (h) SR-D (i)SR-LD (j) SR-CD (k) Proposed Method	85
5.7	Visual results of pan-sharpening methods for WorldView-2 data at full-scale (a) Degraded PAN image (b) Up-sampled MS image (EXP) (c) IHS (d) AWLP (e) MTF-GLP (f) SR-Li (g) SR-TD (h)SR-D (i) SR-LD (j) SR-CD (k) Proposed Method	87
5.8	The combined DWT and CSR based pan-sharpening schematic	91
5.9	Visual results of QuickBird data (a) PAN image (b) Reference MS image (c) Up-sampled MS image (EXP) (d) FIHS (e) AWLP (f) MTF-GLP (g) SR-LD (h) SR-WT (i) SR-D (j) SR-RS (k) SR-CD (l) CSR-D (m) CSR-R (n) Proposed method	96
5.10	Visual results of IKONOS data (a) PAN image (b) Reference MS image (c) Up-sampled MS image (EXP) (d) FIHS (e) AWLP (f) MTF-GLP (g) SR-LD (h) SR-WT (i) SR-D (j) SR-RS (k) SR-CD (l) CSR-D (m) CSR-R (n) Proposed method	98
5.11	Difference between the fused outcomes and the reference MS image for IKONOS data (a) FIHS (b) AWLP (c) MTF-GLP (d) SR-LD (e) SR-WT (f) SR-D (g) SR-RS (h) SR-CD (i) CSR-D (j) CSR-R (k) Proposed method	99
5.12	Visual results of Pléiades data (a) PAN image (b) Up-sampled MS image (EXP) (c) FIHS (d) AWLP (e) MTF-GLP (f) SR-LD (g) SR-WT (h) SR-D (i) SR-RS (j) SR-CD (k) CSR-D (l) CSR-R (m) Proposed method	101

6.1	Performance characteristics of SAM and ERGAS with regularization parameter λ variation (a) SAM Vs. λ (b) ERGAS Vs. λ	115
6.2	Performance characteristics of Q4, full-scale metrics with regularization parameter λ variation (a) Q4 Vs. λ (b) D_λ, D_s Vs. λ	115
6.3	IKONOS dataset and experimental results by different methods (a) PAN image (b) Reference MS image (c) EXP (up-sampled MS image) (d) IHS (e) AWLP (f) MTF-GLP (g) AVWP (h) SR-LD (i) DTV (j) V-L1 (k) SR-CD (l) Proposed method	117
6.4	Pléiades dataset and experimental results by different methods (a) PAN image (256×256) (b) Reference MS image (c) EXP (up-sampled MS image) (d) IHS (e) AWLP (f) MTF-GLP (g) AVWP (h) SR-LD (i) DTV (j) V-L1 (k) SR-CD (l) Proposed method	119
6.5	QuickBird dataset and experimental results by different methods (a) PAN image (400×400) (b) up-sampled MS image (c) IHS (d) AWLP (e) MTF-GLP (f) AVWP (g) SR-LD (h) DTV (i) V-L1 (j) SR-CD (k) Proposed method	120
6.6	Convergence speed of the proposed method for the datasets utilized in experimentation	122

LIST OF TABLES

1.1	Resolution characteristics of sensors	3
3.1	Gain at Nyquist frequency for different sensors	35
3.2	Quality metrics for QuickBird data evaluated using synthesis property	37
3.3	Quality metrics for WorldView-2 data evaluated using synthesis property	40
3.4	Quality metrics for IKONOS data evaluated using consistency property	41
3.5	Quality metrics for QuickBird data evaluated using QNR protocol	43
3.6	Quality metrics for Pléiades data evaluated using QNR protocol	44
4.1	Quality indices for QuickBird (India-Surderban) dataset at reduced-scale	59
4.2	Quality indices for IKONOS (China-Sichuan) dataset at reduced-scale	61
4.3	Quality indices for Pléiades (Toulouse, France) dataset at reduced-scale	63
4.4	Quality indices for WorldView-2 dataset at full-scale	64
4.5	Quality indices for QuickBird dataset at full-scale	66
5.1	Quantitative results for Pléiades data using synthesis property	82
5.2	Quantitative results for IKONOS data using synthesis property	84
5.3	Quantitative results for QuickBird data using consistency property	86
5.4	Quantitative results for WorldView-2 data using QNR protocol	87
5.5	Computational time of different methods for IKONOS dataset	89
5.6	Quantitative measures for QuickBird dataset	97
5.7	Quantitative measures for IKONOS dataset	100

5.8	Quantitative measures for Pléiades data	102
6.1	Reduced-resolution quality Metrics for IKONOS data set	118
6.2	Reduced-resolution quality Metrics for Pléiades data set	118
6.3	Full-resolution quality Metrics for QuickBird dataset	121
6.4	Average execution time comparison of different methods (in seconds)	122

Abbreviations

RS	Remote sensing
SAR	Synthetic aperture radar
LiDAR	Light detection and ranging
IFOV	Instantaneous field of view
SNR	Signal-to-noise ratio
NIR	Near Infrared
PS	Pan-sharpening
PAN	Panchromatic
MS	Multispectral
LRMS	Low-resolution multispectral
HRMS	High-resolution multispectral
SR	Sparse representation
MSLD	Multi-scale learned dictionary
OMP	Orthogonal matching pursuit
k-SVD	k-Means singular value decomposition
CS	Component substitution
MRA	Multi resolution analysis
IHS	Intensity-hue-saturation
PCA	Principal component analysis
ATWT	atrous wavelet transform
AWLP	Additive wavelet luminance proportional
MTF	Modulation transfer function
CC	Correlation coefficient
RMSE	Root mean square error
UIQI	Universal image quality index
SAM	Spectral angle mapper
ERGAS	Erreur relative globale adimensionnelle de synthese
RASE	Relative average spectral error
SSIM	Structural similarity index measure
PSNR	Peak signal to noise ratio
QNR	Quality with no reference
PAU-DL	Profile atom update-dictionary learning
APrU-DL	Atom-Profile updating dictionary learning
DWT	Discrete wavelet transform
BDS	Band dependent spatial details
CSC	Convolutional sparse coding
CPT	Cartoon plus texture decomposition
BPDN	Basis pursuit denoising
ADMM	Alternating direction method of multipliers
TGV	Total generalized variation
CSR	Convolutional sparse representation
GLP	Generalized Laplacian pyramid

Notations

P	Panchromatic image
X	Low-resolution multispectral image
\tilde{X}	Up-sampled multispectral image
\hat{X}	High-resolution multispectral image
R	Resolution ratio between PAN and MS images
B	Number of spectral bands
W_k^d	Wiener deblurring filter for k^{th} band
m_k	MTF matched filter for k^{th} band
D_λ	Spectral distortion index
D_s	Spatial distortion index
S, α	Sparse coefficients matrices
T_0	Admitted sparsity level
\mathbb{D}	Dictionary
\bar{E}	Error matrix in dictionary training
DI	Detail image
g_k	Insertion coefficients for k^{th} band
μ	Mean value
E	Expectation (spatial average)
σ	Standard deviation
v	Additive Gaussian noise
\mathbf{d}_m, \mathbf{M}	Dictionary filters in convolutional sparse representation (CSR)
\mathbf{x}_m	Sparse coefficient maps in CSR model
λ	Regularization parameter in CSR model
γ	Regularization parameter in CPT decomposition
$\check{E}(\cdot)$	Wavelet energy
$\ \cdot\ _p$	l_p norm of the quantity ‘ ’
T	Texture component
C	Cartoon component
g_x, g_y	Horizontal, vertical gradient operators

CHAPTER 1

INTRODUCTION

Remote sensing is the process of acquiring features of an object or an activity through sensors positioned at a distant platform, not being in direct contact with the object of interest. Aircrafts and satellites are the common platforms that facilitate data acquisition for remote sensing observation. Remote sensing sensors, especially those mounted on satellites, furnish the valuable imagery data of the earth's surface. The accessibility and analysis of such data have promoted the applications in fields like climatic changes, precision agriculture, land-cover, and land-use aspects, etc. The primary objective of remote sensing (RS) sensors is to capture the energy reflected or emitted by the earth's surface in a digital format. The recorded digital data is processed by the computing machines to produce imaging products useful for a wide range of applications. Depending upon the specifications of spectrum usage, the following three types of images are produced by the RS sensors: optical images, synthetic aperture radar (SAR) images, and light detection and ranging (LiDAR) images (Pohl and Van Genderen, 2016). This thesis primarily involves optical images that correspond to the visible and infrared spectral regions of the electro-magnetic (EM) spectrum.

The EM spectrum comprising a broad range of wavelengths and the corresponding frequencies is shown in Fig.1.1.

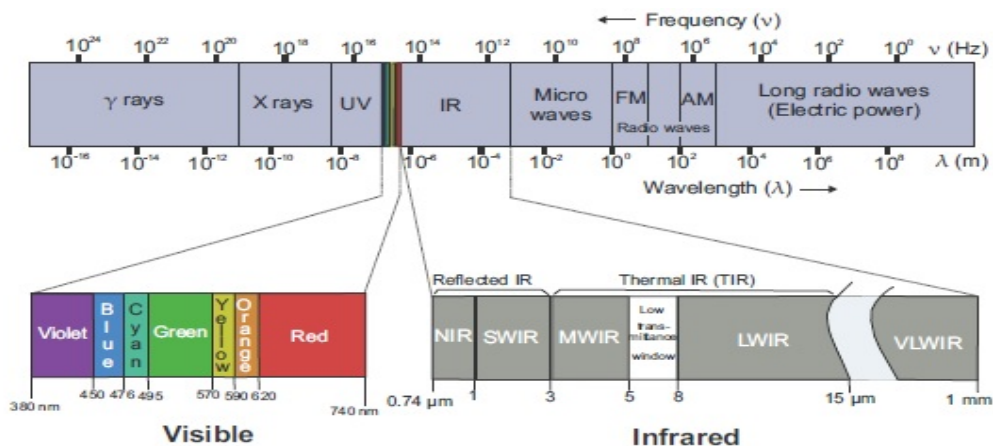


Figure 1.1: Visible and InfraRed spectral regions(expanded) of an electro-magnetic spectrum

The remote sensing optical images are characterized by the following sensor attributes: spatial resolution, spectral resolution, radiometric resolution, and temporal resolution. Spatial resolution refers to the geographical area on the ground represented by a pixel or a measure of the smallest object that can be resolved by the sensor. The amount of scene details that can be secured by the sensor increases with the decrease in pixel size. The spatial resolution of the sensor is associated with its instantaneous field of view (IFOV). The subtended angle through which a detector element or pixel captures the radiation energy is defined as IFOV. For a given number of pixels, the finer the IFOV (corresponding to small area coverage) is, higher the spatial resolution. The spatial resolution of a sensor is the minimum distance by which two separate objects are observed as distinct and is expressed in meters (m). The spectral resolution is the sensor's capability to resolve the wavelength intervals/bands of electro-magnetic spectrum. The narrower the spectral bandwidth is, the higher the spectral resolution. Depending on the supported spectral resolution, sensors can be classified as panchromatic, multispectral, hyperspectral and ultraspectral. Multispectral (MS) images are usually composed of 4-7 spectral bands. Whereas, the sensor that can measure number of spectral bands in the order of hundreds or thousands is referred to as hyperspectral (HS) data. In addition to the MS or HS images, satellite sensors produce a panchromatic (PAN) image. The PAN sensor integrates the radiation from visible and near-infrared range into a single broadband.

The radiometric resolution indicates the dynamic range or different intensities of radiation that the sensor can able to distinguish. The sensor with high dynamic range can simultaneously observe the low and high contrast objects in the scene. The dynamic range of QuickBird sensor is 11 bits, thus 2048 intensity levels are present in each spectral band. Temporal resolution is the time elapsed between two successive measurements of the same scene and is also named as revisit time. Temporal resolution characteristic is useful for land-cover change detection aspect. The resolution characteristics of some of the high-resolution sensors are presented in Table 1.1 (Pohl and Van Genderen, 2016).

The trade-off between sensor's specifications such as IFOV, signal-to-noise ratio (SNR), and on-board storage of satellite sensors impose constraints on the desired spatial and spectral resolutions. For a specific value of IFOV, to contend the image SNR, MS sensors (usually with reduced spectral bandwidth compared to PAN sensor) must support low spatial resolution. The PAN imaging sensor is sensitive to the radiation within a broader wavelength range. Conversely, each MS band covers a narrow spectral range. The higher amount of energy received by the PAN imaging sensor per pixel can

Table 1.1: Resolution characteristics of sensors

Features	QuickBird		Pléiades		IKONOS	
Dynamic range (bits/pixel)	11		12 or 16		11	
Spatial resolution (m)	PAN	MS	PAN	MS	PAN	MS
	0.7	2.8	0.5	2	1	4
Spectral range (nm)	455–745	B: 455–525 G: 530–590 R: 625–895 NIR1: 760–890	455–745	B: 455–525 G: 530–590 R: 625–895 NIR1: 760–890	450–900	B: 450–520 G: 510–600 R: 630–700 NIR1: 760–850

detect intensity variations with smaller pixel sizes compared with MS sensors. Hence, on the same satellite platform, the resolution of PAN sensor is higher than that of MS sensor. Owing to these limitations, the higher spatial resolution can be acquired with a compromise on spectral diversity. The current generation satellite sensors such as IKONOS, QuickBird, WorldView, Pléiades and GeoEye produce images with complementary spatial and spectral characteristics: A panchromatic (PAN) image with high-spatial resolution and a multispectral (MS) image with high-spectral resolution. The most effective solution for providing high-spatial resolution and high-spectral resolution remote sensing images is to develop efficient image fusion techniques.

Wald.L (Wald, 1999) defined image fusion as ” a formal framework which are expressed means and tools for the alliance of data of the same scene originating from different sources. It aims at obtaining information of greater quality; the exact definition of quality depend upon the application.” Fused image give better interpretation means and results as data with different characteristics are merged. The objective of image fusion technique is to integrate the useful information from multiple images of the same scene to generate a result which can provide more useful information than that can be obtained from any source image individually.

The remote sensing applications like land-use and land-cover classification, environmental monitoring, objection detection, and Google maps desire the images with high spatial as well as high spectral resolution. Hence, the objective is to integrate the geometric features constituted by the PAN image (but not present in MS image) and the spectral bands of the MS image (contrary to an unique band in PAN image) into a single image. The advanced data fusion techniques have been developed to exploit

increasing amount of the sophisticated multi-sensor data. Specific methodologies with acceptable apprehension of input data are desirable for image fusion. Remote sensing fusion mechanisms furnish the imagery with enhanced interpretation capabilities on combining essential spatial and spectral features.

1.1 Pan-sharpening

Pan-sharpening (PS) is a remote sensing image fusion technique that combines the complementary spatial and spectral characteristics from the individual images and synthesizes a high-resolution multispectral (HRMS) image. Thus, the pan-sharpening process produces images with high spatial resolution while preserving the essential spectral information (Alparone et al., 2015). The PAN, MS and HRMS images with the corresponding resolution properties are shown in Fig.1.2.

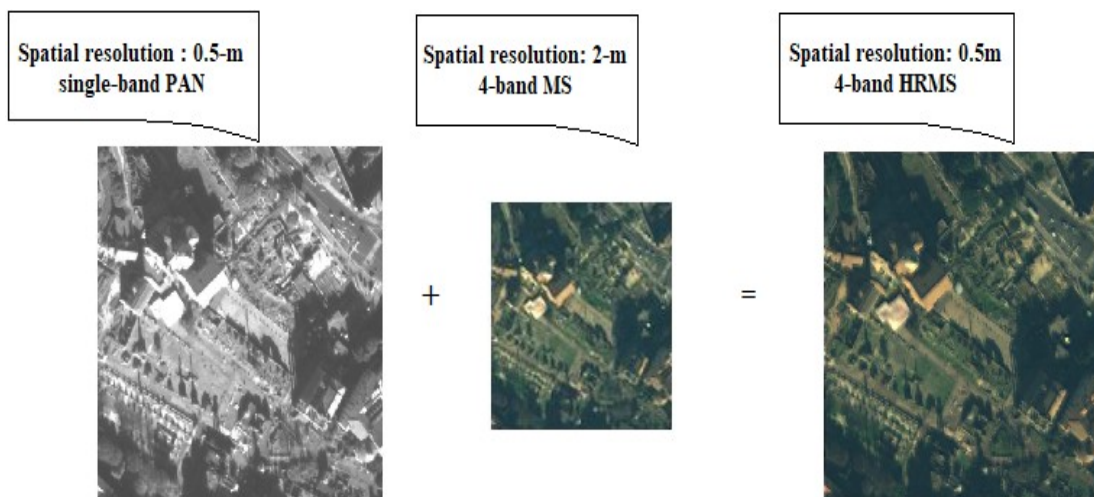


Figure 1.2: PAN, MS and Pan-sharpened images (WorldView-2)

For an illustration, if the multispectral image comprises of $256 \times 256 \times 4$ pixels, the corresponding panchromatic image is 1024×1024 pixels. These two images are treated as the observed data. The obtained pan-sharpened image consists of $1024 \times 1024 \times 4$ pixels. The number of pixels that are to be estimated is 3.2 times the number of pixels that synthesize the observed data.

The research related to sparse representation and reconstruction of images has expanded significantly in the recent years. Sparsity deals with the representation of images using an overcomplete dictionary that consists of image patches as columns (called as atoms). Images are described as linear combinations of few of these atoms. Designing

an appropriate dictionary that sparsely represents the given image is a challenging task. The sparse recovery of signals over an appropriate dictionary is one of the most actively developed theories for remote sensing applications. The initial attempts to apply the concept of sparsity to pan-sharpening of multispectral images are reported in the recent years.

The following two sections describe the concepts of sparse representation of signals/images, the dictionary learning mechanism and the concept of multi-scale dictionary learning process in wavelet domain.

1.2 Sparse Representation Mechanism

During the past two decades, sparsity has emerged as a prominent concept in a wide-range of image processing applications (image denoising, compression, restoration and super resolution; to name only a few applications) (Rish and Grabarnik, 2014). Much of the recent evolution of research made in image processing domain can be attributed to the sparse modelling of image content and a wise implementation of these models in a variety of applications.

A signal/image, \mathbf{y} is considered as an $n \times 1$ column vector in a finite-dimensional subspace of \mathbb{R}^n , is strictly or exactly sparse if most of the vector's elements are equal to zero, i.e., if its support $\Lambda_{\mathbf{y}} = \{1 \leq i \leq n \mid \mathbf{y}[i] \neq 0\}$ is of cardinality $k \ll n$. A k -sparse signal is a signal for which exactly k of its samples have a non-zero value. If a signal is not sparse in its inherent domain, it may be represented sparsely in an appropriate transform domain. For instance, if \mathbf{y} represents a sinusoidal signal it is clearly not sparse in time-domain, however, its Fourier transform is extremely sparse (actually 1-sparse, comprising of only one arbitrary frequency).

The natural images tend to be sparse in a redundant image domain which is usually known as dictionary where, every column of the dictionary is called an atom. The image can be modelled as a linear combination of K atoms of the dictionary such that,

$$\mathbf{y} = \mathbb{D}\mathbf{s} \tag{1.1}$$

where, \mathbf{s} is called sparse representation vector of an image \mathbf{y} with the associated dictionary \mathbb{D} . If the dictionary is a fixed basis, every signal can be exclusively represented as

a linear combination of the dictionary atoms. For an orthogonal dictionary, the sparse representation coefficients (elements of vector \mathbf{s}) can be measured as inner products of the signal \mathbf{y} and the dictionary atoms. Whereas, if the dictionary is non-orthogonal, the sparse coefficients (elements of \mathbf{s}) are the inner products of the signal, \mathbf{y} , and the inverse of dictionary \mathbb{D} , also referred to as the bi-orthogonal dictionary. An overcomplete dictionary ($\mathbb{D}^{n \times K}, n \ll K$) has more elements/ atoms than the dimension of the image. Hence, the dynamics in Eq.1.1 amounts to an under-determined system of linear equations. This leads to a possibility of existing infinitely many vectors, \mathbf{s} to satisfy the relation $\mathbf{y} = \mathbb{D}\mathbf{s}$. The schematic representation of a signal over a redundant dictionary is shown in Fig.1.3.

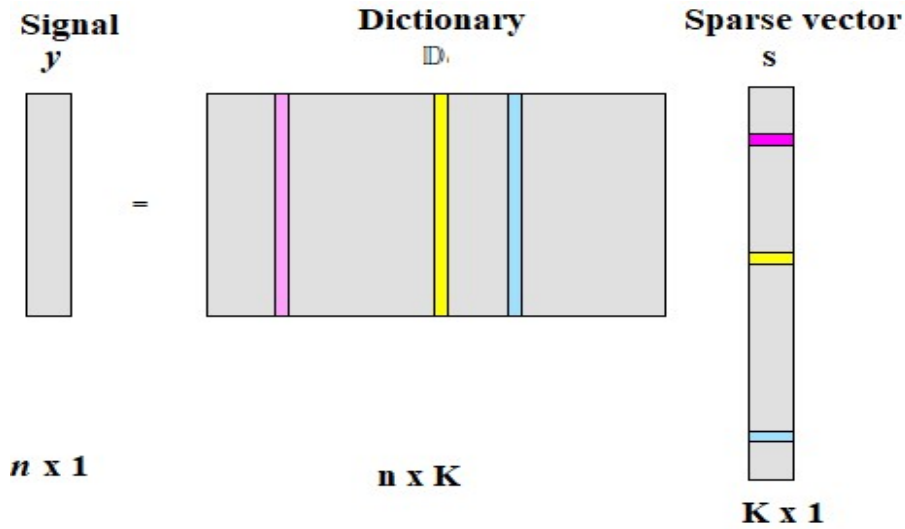


Figure 1.3: Sparse representation schematic.

The aim of sparse representation (also known as sparse coding) is to find a vector \mathbf{s} with fewest possible non-zero elements. Representing the image \mathbf{y} using an appropriate dictionary \mathbb{D} by enforcing the sparsity on \mathbf{s} , leads to solving the following optimization problem.

$$\hat{\mathbf{s}} = \operatorname{argmin} \|\mathbf{s}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbb{D}\mathbf{s}\|_2^2 = 0 \quad (1.2)$$

where, $\|\mathbf{s}\|_0$ is the ℓ_0 pseudo-norm that counts the number of non-zero elements in vector \mathbf{s} . It is well-known that solving the Eq.1.2 is a non-deterministic polynomial-time hard (NP-hard) problem. Therefore, the following two tractable alternative approaches have been practiced: greedy algorithms and convex relaxation.

The ℓ_1 minimization problem,

$$\hat{\mathbf{s}} = \operatorname{argmin} \|\mathbf{s}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbb{D}\mathbf{s}\|_2^2 \leq \epsilon \quad (1.3)$$

is regarded as a convex relaxation of Eq.1.2 (Donoho, 2006). Here, ε is the error tolerance. Usually, the Eq.1.3 is also termed as Basis Pursuit (BP). The ℓ_1 minimization problem can be solved with efficient convex optimization methods. The greedy algorithm called orthogonal matching pursuit (OMP) (Tropp, 2004) has proven to be impressive in solving sparse approximation problem (Eq.(1.3)).

OMP algorithm is an iterative algorithm to solve the sparse approximation problem. The fundamental principle of this algorithm is gradually finding the non-zero locations of sparse coefficients vector ‘s’ one at a time. The training signal \mathbf{y} is assumed as initial residual. In each step, the algorithm search for the column from the dictionary \mathbb{D} that best correlated with the current residual. Subsequently, by considering the new atom and its coefficient the residual is updated. In each iteration of algorithm, all the non-zero coefficients are updated by the Least-Squares method. The choice of the dictionary that sparsifies the signals is crucial for the success of the sparse coding problem.

1.3 Dictionary Learning Mechanism

The performance of the sparse coding algorithms in terms of attained approximation quality and possible choices for the optimal sparse coefficient vector \mathbf{s} depends not only on the signal \mathbf{y} but also on the over complete dictionary, \mathbb{D} . For a given class of images, there exist dictionaries that are more likely to lead to sparsest solutions than that can be obtained on using the other dictionaries. The explicit aim of dictionary learning methods is to find an optimized dictionary that constitute a set of specific atoms describing the most vital attributes of the target images.

The methods for designing a dictionary can be divided into the following two categories: the analytic method and the learning based method. In analytic method, the dictionaries are designed using a set of pre-determined functions/bases. The dictionaries of this category includes definite matrices formed exclusively by Fourier transform, Wavelets, Curvelets etc. These dictionaries are extremely structured and can be implemented in a speedy manner.

In the learning based class, the dictionary is learned from a set of training samples which is more adaptive to the data. The learned dictionaries yield better performance in specific applications, although at the cost of unstructured features. A typical set of such applications includes the images, which contain complicated and non-stationary information. K-Means singular value decomposition (K-SVD) algorithm (Aharon et al.,

2006) is a prominent dictionary learning/training algorithm to extract the most primitive features of images. The dictionary learning problem can be formulated as:

$$\underset{\mathbb{D}, S}{\operatorname{argmin}} \|\mathbf{Y} - \mathbb{D}\mathbf{S}\|_F^2 \quad \text{subject to } \|s_i\|_0 \leq T_0 \quad \forall i \quad (1.4)$$

where, $\mathbf{Y} \in \mathbb{R}^{n \times N}$ is a matrix of columns being training images. $\mathbf{S} \in \mathbb{R}^{K \times N}$, with s_i as columns is a matrix of sparse representations for the data elements in \mathbf{Y} . Here, T_0 is the admitted sparsity level for each column of \mathbf{S} and $\|\cdot\|_F$ represents the Frobenius norm.

The Frobenius norm of a 'p × q' matrix, \mathbb{D} is defined as the squareroot of sum of squares of its elements.

$$\|\mathbb{D}\|_F = \sqrt{\sum_{i=1}^p \sum_{j=1}^q |d_{ij}|^2}$$

K-SVD solves the Eq.1.4 iteratively by the execution of the following two stages: sparse coding and dictionary-update. The Dictionary \mathbb{D} is initialized arbitrarily and sparse coefficients α are computed in the first stage. The penalty term is written as

$$\|\mathbf{Y} - \mathbb{D}\mathbf{S}\|_F^2 = \sum_{i=1}^N \|y_i - \mathbb{D}s_i\|_2^2, \quad (1.5)$$

The problem formulated in Eq.1.4 is seperated into N sub problems.

$$\text{for } i = 1, 2, \dots, N. \quad \min_{s_i} \|y_i - \mathbb{D}s_i\|_2^2 \quad \text{subject to } \|s_i\|_0 \leq T_0$$

In second stage, the dictionary atoms are modified with the corresponding current sparse vectors. Presuming the dictionary \mathbb{D} and the sparse coefficient matrix \mathbf{S} are fixed, the column d_k from the dictionary and its corresponding coefficients in s , that is i^{th} row s_T^i to be considered for updation. In the formulated objective function (Eq.1.4), the term $\mathbb{D}\mathbf{S}$ is decomposed as

$$\|\mathbf{Y} - \mathbb{D}\mathbf{S}\|_F^2 = \left\| \mathbf{Y} - \sum_{i=1}^K d_i s_T^i \right\|_F^2 \quad (1.6)$$

$$= \left\| \left(\mathbf{Y} - \sum_{i \neq k} d_i s_T^i \right) - d_k s_T^k \right\|_F^2 \quad (1.7)$$

$$= \|\bar{\mathbf{E}}_k - d_k s_T^k\|_F^2 \quad (1.8)$$

\bar{E}_k is the error matrix for all the N sub-problems, except for the k^{th} atom in the dictionary. The following modifications are executed on the error term, \bar{E}_k , defined in Eq.1.8, in order to enforce the sparsity constraint. The error matrix, \bar{E}_k is modified such that it is restricted to include only the sub-problems (out of total N sub-problems) that are currently using the d_k atom, labeled as \bar{E}_k^R . The row vector, s_T^k is shrunk by discarding the zero entries and termed as s_R^k . Under these circumstances, the problem can be treated as the minimization of the function

$$\|\bar{E}_k^R - d_k s_R^k\|_F^2 \quad (1.9)$$

The singular value decomposition (SVD) is used to decompose the restricted error matrix as $\bar{E}_k^R = U\Delta V^T$. The first column of U is the new solution \tilde{d}_k , and the first column of V^T multiplied by $\Delta(i,j), i = j = 1$, is the new coefficient vector s_R^k . The values obtained from SVD, effectively minimize the error defined in Eq.1.8.

The prominent results with K-SVD can be achieved by implementing the following variations.

- The dictionary element which is being not frequently used (relatively with other atoms) during sparse coding stage can be replaced with least represented training vector.
- If any two most similar columns (if the absolute inner product of two columns exceeds a threshold value) are exists in the dictionary, one of the column can be replaced with least represented vector.

1.3.1 Multi-Scale Dictionary Learning Using Wavelets

The inherent features of a certain class of images can be acquired effectively by combining the features of multi-scale representations (such as Wavelets) with the capability of learned dictionaries. The traditional analytic and learning based approaches can be merged to construct exactly a multi-scale learned dictionary inheriting the advantages of both of these approaches. The notion of multi-scale dictionary learning is effective for sparse and redundant representation of signals which are naturally appear at different scales. The multi-scale dictionary learning using wavelets can be expressed by modifying Eq.(1.4), as :

$$\underset{D,S}{\operatorname{argmin}} \|\mathbb{W}_A Y - \mathbb{D}S\|_F^2 \quad \text{subject to } \|s_i\|_0 \leq T_0 \quad \forall i \quad (1.10)$$

where, W_A is the wavelet analysis operator. The Wavelet analysis domain is a collection of coefficient images or bands. Different bands consists of data at different scales and orientations (horizontal, vertical and diagonal).

According to this formulation, the data samples can be represented as sparse combination of atoms from a multi-scale core dictionary like wavelets. Since the wavelet domain is a combination of bands at different scales, the multi-scale dictionary learning process can be summarized as learning separate dictionaries for different bands which in turn are oriented in different directions, namely horizontal, vertical and diagonal. Thereby, the dictionary learning in the transform domain is expressed as:

$$\forall b \operatorname{argmin}_{\mathbb{D}, \mathbb{S}} \|(W_A Y)_b - \mathbb{D}_k \mathbb{S}_b\|_F^2 \quad \text{subject to } \|\mathbb{S}_{i,b}\|_0 \leq T_0 \quad \forall i \quad (1.11)$$

Here, b denotes index for wavelet sub-bands. The formulation in Eq.1.11 implies that the dictionary is trained in wavelet domain, further it leads to reasonable computational cost in learning as well as in sparse coding.

In contrast to the dictionary learning in image domain, a large area of image is affected with the increase of level of decomposition in wavelet domain learning approach. This feature creates a global as well as local outlook for the wavelet domain dictionary learning mechanism.

1.4 Overview of Pan-sharpening Paradigms

Pan-sharpening is a pixel level fusion technique that deals with a process of changing the low-spatial resolution multispectral bands to high-spatial resolution color images on fusing with a co-georegistered high spatial resolution panchromatic (black and white) image of the same area. Pan-sharpening has become important in many applications of remote sensing like change detection, monitoring hazards and other applications. Owing to great importance of pan-sharpened images, many pan-sharpening algorithms have been developed by the researchers. The comprehensive review articles present the development in remote sensing image fusion field in the last three decades ([Vivone et al., 2015](#); [Pohl and van Genderen, 2015](#); [Ghassemian, 2016](#); [Duran et al., 2017](#)). To date, many pan-sharpening methods have been developed which can be classified into different categories based on the mathematical framework used to implement the fusion process. The two well-known and established classes of algorithms are based on the concept of component substitution (CS) and multi-resolution analysis (MRA). The

other categories include fusion based on sparse representation of signals and variational schemes. Furthermore, several hybrid methods have been proposed which combines the modalities of different categories.

The CS and MRA methods can be summarized in a rational framework, so as to apparently express the differences between the two classes. Let $X_k, k = 1, 2, \dots, B$. be the MS image with B number of bands and P be the PAN image. The MS image up-sampled to the size of PAN image is denoted as \tilde{X}_k and \hat{X}_k represents the estimated HRMS image.

The pan-sharpened image is obtained by adding a detail image, DI, to the up-sampled MS image, \tilde{X}_k , as

$$\hat{X}_k = \tilde{X}_k + g_k \cdot DI, \quad k = 1, 2, \dots, B \quad (1.12)$$

The detail image, DI, is usually composed of spatial details missing in MS image bands. The size of detail image DI, is the same as that of up-sampled MS image and pansharpened images. The CS and MRA methods mainly differ in terms how the details of an image DI are estimated. The quality of the pansharpened image is predominantly affected by the detail image. In Eq.1.12, $g_k = [g_1, g_2, \dots, g_B]$ is a vector of insertion coefficients, which are band-specific. For CS based methods, the detail image is estimated as,

$$DI = P - \sum_{k=1}^B \omega_k \cdot \tilde{X}_k \quad (1.13)$$

The choice of parameters like insertion coefficients (g_k) and the band dependent weights (ω_k) are determined by the corresponding CS technique. The weight vector specifies the proportion in which each band has to be preferred for the weighted sum of MS bands. The spectral distortion in the pansharpened image is resulted by the difference between PAN image and the weighted sum of MS bands. The familiar CS based methods are intensity-hue-saturation (IHS) (Tu et al., 2001, 2004), principal component analysis (PCA) (Chavez et al., 1991), Gram-Schmidt transform (GS) (Laben and Brower, 2000) and adaptive GS (GSA) methods (Aiazzi et al., 2007).

The detail image for MRA techniques is estimated as:

$$DI = P - P_L \quad (1.14)$$

Where P_L represents the low-pass version of PAN image, P. The particular MRA algorithm determines the type of the filter used to determine P_L and the insertion co-

efficient vector g_k . The well-known MRA methods are based on high-pass filter (HPF) (Chavez et al., 1991), wavelets (Amolins et al., 2007b), trous wavelet transform (ATWT) (Nunez et al., 1999), additive wavelet luminance proportional (AWLP) (Otazu et al., 2005), curvelet transform (Nencini et al., 2007) and contourlet transform (Yang and Jiao, 2008).

The CS methods preserve the requisite spatial details, however, they induce spectral distortion in the fused image. The disparity between PAN image and the replaced spatial component, results in lack of retaining the consistent spectral information. Whereas, MRA methods are capable of enhancing the spectral information, though inferior to CS methods in spatial detail enhancement.

The concept of sparsity has become prevalent for the pan-sharpening of multispectral images in the recent years. The reliable and robust solutions to the pan-sharpening problem are designed based on the sparse representation framework in this thesis. The sparse coding is promoted by constructing the relevant dictionaries based on the features extracted from the source images. In addition, a variational based optimization paradigm is designed for the pan-sharpening mechanism.

1.5 Research Objectives

To acquire the legitimate balance between transferring spatial details and preserving spectral information of a high-resolution multispectral (HRMS) image, sparsity and variational based pan-sharpening methods have enticed huge attention. The general objective of this thesis is to solve the pan-sharpening problem by developing a variety of dictionaries to promote the sparsity of remote sensing images. Further, the pan-sharpening is formulated as a variational model driven by consistency priors in a unified optimization framework. The optimization problem is then solved to obtain the pan-sharpened image. The sub-objectives are:

1. To design a robust sparsity model by using the statistical measures during the learning phase of the dictionary to reduce spectral distortion in the fused image and memory requirements of the pan-sharpening algorithm. The scale-invariance notion is deployed in learning a dual dictionary to reconstruct the spatial details that are to be imparted into the fused image.

2. By Developing a sparsity based fusion method using multi-scale learned dictionary that inherits the merits of both adaptive and learning based dictionaries. The

principle of multi resolution analysis (MRA) is combined with the concept of sparse recovery to obtain the pansharpened MS image.

3. The shortcomings in patch based fusion are, limited ability in detail preservation and high sensitivity to misregistration. Formulating a fusion model for pan-sharpening based on convolutional sparse representation (CSR) to overcome the aforementioned drawbacks.

4. To develop an efficient variational model for pan-sharpening, by formulating an energy functional consists of pertinent regularization terms and an optimization algorithm to solve the energy function.

1.6 Structure of Thesis

The structure of the thesis chapter-wise is as follows.

Chapter-2: In this chapter, the protocols developed for evaluation of pansharpened images and different quality metrics used in the assessment of experimental results are discussed. The chapter also presents various datasets used for experimentation.

Chapter-3: A two-stage pan-sharpening algorithm for enhancing the spatial details is presented in this chapter. A dual dictionary is constructed with the patches having high-frequency details extracted from the PAN image and its low resolution variant. A de-blurring filter is designed and the up-sampled MS images are processed with the designed filter to reduce the spectral distortion in the fused image. An efficient training algorithm, namely, parallel atom-updating dictionary learning (PAU-DL) is used for dictionary learning.

Chapter-4: The pan-sharpening method that combines the MRA framework with the sparse representation over a multi-scale learned dictionary is illustrated in this chapter. The multi-scale dictionary is learned in wavelet domain. The multi scale learned dictionary enhances the representation of inherent features of the images with an efficient sparse coding.

Chapter-5: The Convolutional Sparse Representation (CSR) is a relatively recent development in many image processing applications, as an alternative to the generalized sparse representation. Two panshrpening methods based on Convolutional sparse representation (CSR) are presented in this chapter. The conventional sparse representation based pan-sharpening methods employ patch partition based strategies. An alternative

mechanism called convolutional sparse representation is applied for pan-sharpening. Cartoon plus texture (CPT) based image decomposition is used to enhance the spatial and spectral qualities of pan-sharpened image. Further, another pan-sharpening scheme is implemented in which wavelet decomposition is used and CSR based fusion is performed in wavelet domain instead of exercising it in spatial domain.

Chapter-6: A variational model for pan-sharpening is proposed in this chapter. The panshrpening process is formulated as a constrained optimization problem with the appropriate prior terms. The inter-band correlation is exploited to reduce the spectral distortion in the fused image. An efficient algorithm called alternating direction method of multipliers (ADMM) is developed to minimize the optimization function. The operator splitting framework assisted by the ADMM algorithm yields the pansharpened image as a solution of the proposed optimization problem.

Further, in chapters 3,4,5 and 6, the comprehensive fusion quality assessment relative to state-of-the-art methods is detailed.

Chapter-7: The conclusions are drawn and perspectives on future work are given in this chapter.

CHAPTER 2

QUALITY ASSESSMENT OF PAN-SHARPENED IMAGES AND SPECIFICATIONS OF DATASETS

The quality assessment of pan-sharpened images is very important in various remote sensing applications. The suitability of the fused product for a particular application depends on the spatial and spectral quality of the pansharpened images. The spectral information is useful for the applications like soil and vegetation analysis and lithography. The geometric feature detection and extraction based applications depend heavily on the spatial information. For the given input datasets, pan-sharpening process aims to produce a new dataset that acquires some of the properties of its components. The objective of quality metrics is to evaluate the consistency of preserving the spatial and spectral resolution features of fused image that are inherited from the given source images.

The critical limitations for fusion quality evaluation are: the reference image for comparison is unavailable and if the quality is evaluated at the highest resolution (at the resolution of PAN image), the measured spatial and spectral qualities may follow opposite trends. To cope with the problems of lack of reference image and spectral-spatial distortion trade-off, distinct protocols have been proposed for the quality evaluation. These protocols have evolved because a single performance measure is not suitable to validate the quality of pan-sharpened images. The lack of an ideal reference image for comparison limits the objective evaluation of fusion results, independent of the chosen assessment index ([Vivone et al., 2015](#)). Therefore, assessment of image fusion mechanism makes usage of two different approaches. The first, a qualitative approach, is as the visible inspection of fused images, comparing the outcome to the original input data. The second, a quantitative approach, uses statistics and other assessment methods to provide comparable quality measures also known as quality indices.

This chapter focuses about the overview on quality assessment of fusion of multispectral (MS) images with high-resolution panchromatic (PAN) observations. A detailed discussion about the protocols developed for the evaluation of pan-sharpened products and the datasets used for experimentation are presented. The notions of spatial and spectral quality and the quality metrics used for their measurements are discussed.

2.1 Quality Metrics

Quality metrics are defined to evaluate the similarity for both the scalar and vector valued images, as required by the established protocols. The list of metrics to assess the quality of pan-sharpened images is long. The most commonly used quality metrics including definitions are explained in this section.

- Correlation coefficient (CC) : Correlation coefficient indicates the degree of similarity between the original MS image, X and the pan-sharpened image, \hat{X} each having the size $p \times q$. The CC can be determined as:

$$CC = \frac{\sum_{i=1}^p \sum_{j=1}^q [\hat{X}_{i,j} - \bar{\hat{X}}][X_{i,j} - \bar{X}]}{\sqrt{\sum_{i=1}^p \sum_{j=1}^q [\hat{X}_{i,j} - \bar{\hat{X}}]^2 \cdot \sum_{i=1}^p \sum_{j=1}^q [X_{i,j} - \bar{X}]^2}} \quad (2.1)$$

where \bar{X} and $\bar{\hat{X}}$ represents the mean values of original MS image and pan-sharpened image, respectively. The optimal value of CC is one.

- Root mean square error (RMSE) : RMSE gives the standard measure of difference in pixel values between two images, the original MS image X and the pan-sharpened image \hat{X} .

$$RMSE = \sqrt{\frac{1}{pq} \sum_{i=1}^p \sum_{j=1}^q (X_{i,j} - \hat{X}_{i,j})^2} \quad (2.2)$$

For smaller values of RMSE the pan-sharpened image is in close match with the original MS image.

- Universal image quality index (UIQI) : UIQI measures the similarity between two images X and \hat{X} .

$$Q = \frac{4 \cdot \sigma_{X,\hat{X}} \cdot \bar{X} \cdot \bar{\hat{X}}}{(\sigma_X^2 + \sigma_{\hat{X}}^2)[\bar{X}^2 + \bar{\hat{X}}^2]} \quad (2.3)$$

in which $\sigma_{X,\hat{X}}$ denotes the covariance between X and \hat{X} , \bar{X} and $\bar{\hat{X}}$ are means, and σ_X^2 and $\sigma_{\hat{X}}^2$ are the variances of original MS image, X and the pan-sharpened image, \hat{X} , respectively.

This index can be re-written as product of three factors:

$$Q = \frac{\sigma_{X,\hat{X}}}{\sigma_X \sigma_{\hat{X}}} \cdot \frac{2 \cdot \bar{X} \cdot \bar{\hat{X}}}{[\bar{X}^2 + \bar{\hat{X}}^2]} \cdot \frac{2 \cdot \sigma_X \sigma_{\hat{X}}}{(\sigma_X^2 + \sigma_{\hat{X}}^2)}$$

First factor is the correlation coefficient between X and \hat{X} . Second factor indicates spectral distortion as bias in the mean of X with respect to \hat{X} . Third term reflects changes in contrast between X and \hat{X} .

- Spectral angle mapper (SAM): SAM calculates the spectral similarity between fused image and the original MS image.

$$\text{SAM}(V_X, V_{\hat{X}}) = \arccos \left(\frac{\langle V_X, V_{\hat{X}} \rangle}{\|V_X\|_2 \cdot \|V_{\hat{X}}\|_2} \right) \quad (2.4)$$

Where $\langle \cdot \rangle$ and $\|\cdot\|_2$ operators denote dot product and l_2 norm, respectively. V_X and $V_{\hat{X}}$ are the spectral vectors constructed from each pixel of the original MS image, X and fused MS image, \hat{X} . To provide an overall assessment of spectral distortion, SAM is averaged over the entire image. An ideal value (ZERO) of SAM indicates the absence of spectral distortion.

- ERGAS is a French acronym for “*Erreur relative globale adimensionnelle de synthese*“. It depicts global quality of the fused image.

$$\text{ERGAS} = 100 \frac{h}{l} \sqrt{\frac{1}{B} \sum_{i=1}^B \left(\frac{\text{RMSE}(i)}{\text{MEAN}(i)} \right)^2} \quad (2.5)$$

where $\left(\frac{h}{l}\right)$ is the resolution ratio between PAN and LRMS images. B is the number of bands. $\text{MEAN}(i)$ is mean (average) of i th band. A low ERGAS value indicates an optimal state of spectral information preservation.

- Q4 : Q4 is a multispectral extension of UIQI suitable for images having four spectral bands.

The Q4 index is defined as :

$$Q4 = \frac{4 |\sigma_{Z_X} \sigma_{Z_{\hat{X}}}| \cdot |\bar{Z}_X| \cdot |\bar{Z}_{\hat{X}}|}{(\sigma_{Z_X}^2 + \sigma_{Z_{\hat{X}}}^2) |\bar{Z}_X|^2 |\bar{Z}_{\hat{X}}|^2} \quad (2.6)$$

where,

$$\begin{aligned} Z_X &= a_x + ib_x + jc_x + kd_x \\ Z_{\hat{X}} &= a_{\hat{x}} + ib_{\hat{x}} + jc_{\hat{x}} + kd_{\hat{x}} \end{aligned} \quad (2.7)$$

are the quaternions and represent the reference MS image and fused image. Let a, d, c and d denote the radiance values of an image pixel in B, G, R and NIR bands, respectively. The optimal value of Q4 is 1, which indicates the original MS image and fused image are identical.

- Relative average spectral error (RASE): RASE is an error index that gives the global quality of fused image.

$$\text{RASE} = \frac{100}{\sum_{i=1}^B \mu(i)} \sqrt{B \sum_{i=1}^B \text{MSE}(i)} \quad (2.8)$$

where $\mu(i)$ is the mean of i^{th} band and B denotes the number of spectral bands. MSE is the mean square error between two spectral bands b_1 and b_2 , which is defined as $\text{MSE} = E[(b_1 - b_2)^2]$. The operator E (expectation) denotes the spatial average. The lower values of RASE indicates the similarity between multi spectral bands.

- Structural similarity index measure (SSIM): SSIM estimates the structural similarity between the pansharpened image and the reference image.

$$\text{SSIM} = \frac{(2 \cdot \mu_x \cdot \mu_{\hat{x}} + c_1)(2 \cdot \sigma_{x, \hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \quad (2.9)$$

The constants c_1 and c_2 are necessary to avoid a division by zero. They depend on the dynamic range of the pixel values. SSIM compares luminance, contrast, and structure using means and standard deviations of fused and reference image. The higher the value of the measure, the better is the expected quality of the fused image.

2.2 Protocols for Pan-sharpening

Quantitative image quality assessment is based on mathematical modeling and often referred to as objective analysis. The goal is to determine the closeness of the two data sets or in other words their similarity. For the quality assessment of fused images, several protocols have been established in the literature; Wald's protocol (Wald et al., 1997), Zhou's protocol (Zhou et al., 1998), Khan's protocol (Khan et al., 2009), quality with no reference (QNR) protocol (Alparone et al., 2008) etc. are the prominent one. The Wald's protocol is widely accepted for pan-sharpened image quality estimation by the research community and is reviewed further in (Ranchin et al., 2003; Thomas et al., 2008).

The approved protocols are reliable and authentic for quality evaluation. These protocols develop a combination of metrics that provide a comparable framework for quality. The basis admitted by researchers is, legitimate quality assessment depends on both visual inspection and a quantitative approach.

2.2.1 Wald's Protocol

The fused image is supposed to maintain the following three properties proposed by this protocol. The first property, **consistency** specifies that any fused image \hat{X} , if degraded to its basic resolution, should be identical to the original MS image X . The evaluation under this property is established as: The pan-sharpening process is performed on the original dataset to achieve the fused image \hat{X} . The obtained pan-sharpened image, \hat{X} is down-sampled to the resolution of source MS image, X . The comparison is made between the images \hat{X} and X . Thus, the consistency property quantifies the spectral quality of spatially enhanced image.

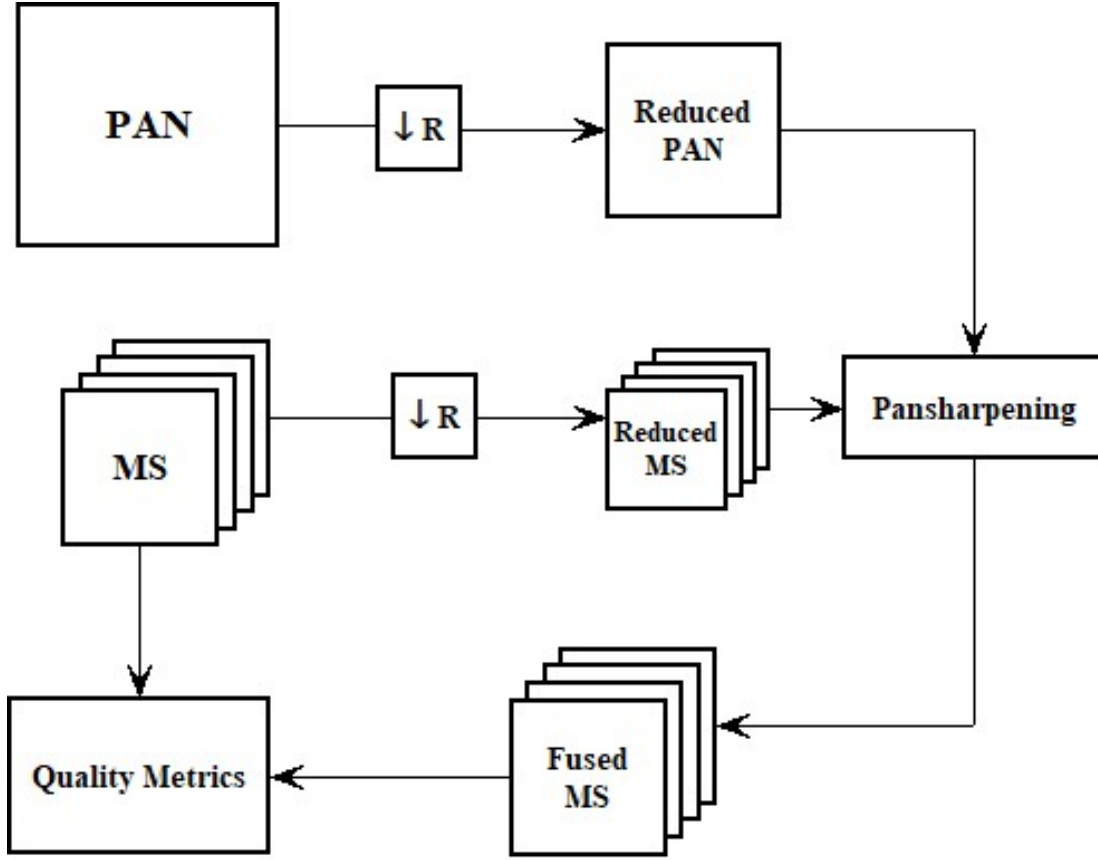
According to the second property, **synthesis**, the pan-sharpened image should be similar to ideal image that a corresponding sensor would observe at the spatial resolution of HRMS image. The images are viewed as scalar images, that is the property is checked for individual bands of multispectral image. For vector images, that is for all the bands of MS image, the synthesis property requires that the multispectral vector of fused (pan-sharpened) images should be as identical as possible to the multispectral vector of ideal images that the corresponding sensor, if it exists, would observe at the spatial resolution of the high-resolution image (same as the resolution of PAN image). The later part of synthesis property proposed for vector images is also known as the third property of Wald's protocol.

The second and third properties are difficult to implement since, the ideal image requisite for comparison does not exist. Hence, the synthesis property is implemented on the images at reduced-scale. The source images are degraded to a scale equal to the resolution ratio between PAN and MS images (usually the ratio is four). The pan-sharpening process is performed on down-sampled images, produces the fused image at the resolution of original MS image. The fused image is then compared with the original MS image. The MS image is treated as a reference image, hence the second and third properties can be tested.

The schematic of quality evaluation at reduced-scale is presented in Fig.2.1.

2.2.2 QNR Protocol

The lack of reference for quality evaluation has motivated the evolution of another protocol known as quality with no reference (QNR) protocol. The QNR protocol assess the quality of pan-sharpened images without requiring a reference HRMS image. The QNR



R - Resolution ratio between PAN and MS images

Figure 2.1: Assessment process of Wald's synthesis property

protocol operates at full-scale and involves two metrics to evaluate spatial and spectral quality of the fused image, namely spectral distortion index D_λ and spatial distortion index D_s .

The spectral distortion index D_λ is estimated between low-resolution MS image bands and the fused image bands. Two sets of UIQI values are estimated at low and high-resolution to determine the spectral distortion index. The differences between corresponding UIQI values indicates the spectral distortion produced by fusion process.

$$D_\lambda = \sqrt{\frac{1}{B(B-1)} \sum_{i=1}^B \sum_{j=1, j \neq i}^B |Q(X_i, X_j) - Q(\hat{X}_i, \hat{X}_j)|^{\theta_1}} \quad (2.10)$$

where X_i is the i^{th} band of low resolution MS image, \hat{X}_i is i^{th} band of pan-sharpened

MS image and B is the number of spectral bands. Q represents the the index UIQI and the exponent θ_1 is an integer selected to assert the large difference values and its default value is set as one.

The spatial distortion index (D_s) is estimated by calculating the UIQI values between the MS image bands and low resolution version of PAN and as well as between pan-sharpened MS image bands and the PAN image. The difference between the two values produce spatial distortion index:

$$D_s = \sqrt{\frac{1}{N} \sum_{i=1}^B |Q(X_i, P_L) - Q(\hat{X}_i, P)|^{\theta_2}} \quad (2.11)$$

with P_L is the PAN image down-sampled to the resolution of MS image. The default value for the exponent θ_2 is considered as unity.

The global quality measure QNR is evaluated as

$$QNR = (1 - D_\lambda)^{a_1} \cdot (1 - D_s)^{a_2} \quad (2.12)$$

QNR measures both the spatial and spectral qualities of the fused image. Further, a_1 and a_2 are the tuning parameters used in order to assign the priority to the spectral quality and spaial quality evaluation respectively.

If $a_1 = a_2 = 1$, the spatial and spectral aspects are considered with equal precedence in quality assessment of fused image.

The quality assessment paradigm using QNR protocol is developed based on the following hypothesis.

(i) The inter-band similarities between the low-resolution MS image bands measured using the index UIQI should not change with resolution.

(ii) The relation between a low-resolution version of PAN image and each low-resolution MS image band should be identical to the relation between high-resolution PAN image and each band of pan-sharpened MS image.

(iii) The estimated spectral (MS-MS) or spatial (MS-PAN) differences in similarity values at low and high-resolution are the quantitative measures of distortion present.

2.3 Datasets

This section discusses about the various datasets used in the domain of remote sensing research. The remote sensing datasets from four different sensors- IKONOS, QuickBird, WorldView-2, and a simulated dataset acquired from Pléiades are used in this research work.

1. IKONOS dataset ¹:

IKONOS was the first high resolution satellite launched in 1999, consists of optical very high resolution (VHR) sensor. The images generated by the IKONOS sensor are appropriate for small-area analysis due to its high resolution and pointing capabilities. The radiometric resolution of the IKONOS data is quantized to 11 bits.

The IKONOS dataset used in the current research work is composed of mountainous area with trees of sichuan from China. The spatial resolution of PAN and MS images are 1-m and 4-m, respectively. The resolution ratio between PAN and MS images is four. The MS image is composed of 4-bands: Red, Green, Blue and Near InfraRed (NIR).

2. QuickBird dataset ²:

QuickBird satellite was launched in 2001 for commercial purposes. The data acquired is quantized to 11 bits. A single sensor captures both the PAN and MS images. The QuickBird sensor offers PAN image with 0.7-m resolution and 4-band (R,G,B and NIR) MS image with 2.8-m resolution. The images used for this thesis work, include a forest area of the geographical region "Surderban" situated in the eastern part of India.

3. WorldView-2 dataset ³:

DigitalGlobe company launched the WorldView-2 (WV-2) satellite in 2009. The WV-2 dataset used for experimentation in this work represents an area from Sydney, Australia. The WorldView-2 sensor produces a single band PAN image and an eight band MS image. The MS image composed of four standard color bands (red, green, blue, and near- infrared 1) and four new bands (coastal, yellow, red edge, and near-infrared 2).

WorldView-2 sensor provides 0.5-m panchromatic image and 8-band MS image with each spectral band maintaining 2-m resolution. The resolution ratio is 4 and the radiometric resolution is 11 bits. The spectral range of PAN acquired by WorldView-2 is narrower than that of acquired from IKONOS and QuickBird sensors.

¹<http://www.glcf.umiacs.umi.edu/data/ikonos>

²<http://www.glcf.umiacs.umi.edu/data/quickbird>

³https://www.apollomapping.com/image_downloads/WV2/

4. Pléiades dataset ⁴:

For the Pléiades dataset, the resolution of MS bands is 60-cm. The high resolution PAN image is simulated from the available MS bands. The resolution ratio between PAN and MS images is selected as four and the radiometric resolution is 11 bits. This dataset represents the urban area of Toulouse in France.

⁴<http://www.openremotesensing.net/>

CHAPTER 3

SPARSE REPRESENTATION DRIVEN TWO-STAGE PAN-SHARPENING ALGORITHM

This chapter presents a pan-sharpening (PS) technique exploiting sparse representation of spatial details over a dual-dictionary learned from the high-frequency details of panchromatic image. The proposed method is implemented in two stages. The spectral components and spatial components are estimated separately and subsequently integrated to obtain the required high-resolution multispectral (HRMS) image. The section 3.1 presents the state-of-the-art sparsity based pan-sharpening methods. The mathematical framework and the implementation details of the proposed method are described in section 3.2. Section 3.3 presents the comprehensive comparative performance analysis of proposed method with the help of visual outcomes and quality metrics. Section 3.4 concludes the chapter.

3.1 A Review of Sparsity Based Pan-sharpening Methods

The sparse representation based pan-sharpening methods lead to the promising results compared with the conventional CS and MRA based methods. SR based fusion methods exploit the fact that the image patches are sparse in a redundant dictionary. The SR based remote sensing fusion method was first proposed in (Li and Yang, 2011) (authors used the term compressive sensing). The high-resolution MS (HRMS) image is reconstructed based on sparsity regularization. A random raw patches dictionary derived from HRMS images is used to promote the sparsity of fused image patches. However, the inherent limitation of this scheme is that the dictionary construction requires a large set of training images which may not be readily available.

To overcome this problem, pan-sharpening techniques with sparse coding over two different kinds of dictionaries have been proposed in the literature. The dictionary constructed from the available source images i.e., the PAN and MS images. (Jiang et al., 2012) proposed a trained dictionary synthesized using PAN and low-resolution MS

(LRMS) images for the sparse representation of fused image patches. The HRMS image patches are assumed to be sparse and reconstructed from a joint dictionary learned from both the source images. In (Li et al., 2013b; Cheng et al., 2014), authors consider the advantage of using the source images to produce adaptive dictionaries. The dictionary for unknown HRMS image is learned without using a set of training images. All these methods exploit the relationship between source images and fused image using an established remote sensing image formation model.

The remote sensing image formation model can be illustrated as following:

The observed LRMS image is modeled as the decimated and noisy version of the HRMS image.

$$\mathbf{X} = \mathbf{M}_1 \widehat{\mathbf{X}} + \mathbf{v}_1 \quad (3.1)$$

where, $\widehat{\mathbf{X}}$ and \mathbf{X} are the column vectors (lexicographically ordering its pixels) representing the HRMS image and MS image respectively. The matrix \mathbf{M}_1 is introduced to implement the blurring and spatial degradation operations and \mathbf{v}_1 represents the zero mean additive gaussian noise vector.

PAN image can be approximated as a linear combination of HRMS image bands with some additive Gaussian noise, since both the images maintain same spatial resolution.

$$\mathbf{P} = \mathbf{M}_2 \widehat{\mathbf{X}} + \mathbf{v}_2. \quad (3.2)$$

where, $\mathbf{M}_2 = [w_1 \mathbf{I}, w_2 \mathbf{I}, w_3 \mathbf{I}, w_4 \mathbf{I}]$, \mathbf{I} is an identity matrix. w_1, w_2, w_3 and w_4 are the weights corresponding to four MS spectral bands and \mathbf{v}_2 is the additive Gaussian noise vector.

On combining equations (3.1) and (3.2)

$$\mathbf{G} = \mathbf{M}_c \widehat{\mathbf{X}} + \mathbf{v} \quad (3.3)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{X} \\ \mathbf{P} \end{bmatrix} \quad \mathbf{M}_c = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}$$

The objective of pansharpening is to recover the HRMS image $\widehat{\mathbf{X}}$, from the source image set \mathbf{G} . The reconstruction of pan-sharpened image from the source images (Eq.3.3) is formulated as an optimization problem with sparsity constraint.

$$\hat{\alpha} = \arg \min \|\alpha\|_0 \text{ s.t. } \|\mathbf{G} - \phi \alpha\|_2^2 \leq \epsilon \quad (3.4)$$

where, $\phi = M_c \mathbb{D}$, \mathbb{D} is a dictionary and the HRMS image can be represented as a linear combination of atoms (columns) of the dictionary and the sparse coefficients vector α as $\hat{X} = \mathbb{D}\alpha$. The equation 3.4 is solved for the sparsest α by using orthogonal matching pursuit (OMP) algorithm. Finally, the HRMS image is estimated as $\hat{X} = \mathbb{D}\hat{\alpha}$ where $\hat{\alpha}$ is an optimally sparse vector. Since it is very difficult to solve the problem in Eq.3.4 especially for large images, the OMP algorithm and dictionary learning algorithms are operated on image patches.

Another variant of SR based methods is based on designing dictionaries only from PAN image and its low resolution version. The underlying principle for these methods is the direct coherence between low resolution image patches and its high resolution counterparts. In (Zhu and Bamler, 2013) (authors named it as SparseFI) a PS technique is proposed based on the dictionary that explores sparse coding of MS images over the dictionary learned from PAN image. An extension of SparseFI, a two-step sparse coefficient estimation method (Joint SparseFI) is proposed in (Zhu et al., 2016). A two step sparse coding method with the exploitation of patch normalization (PN-TSSC) is presented in (Jiang et al., 2014) to reduce the spectral distortion. Vicinanza et al. (Vicinanza et al., 2015) proposed a fusion method that relies upon patch similarity paradigm integrated with spatial details injection scheme. The HF details of MS image patches are sparsely represented over a dictionary learned from degraded PAN image. The principle of CS and MRA methods is combined with the sparse representation of high-frequency details is investigated in few of the recent works (Yin, 2015; Cheng et al., 2015). The high-frequency details extracted from the PAN image are sparse coded and reconstructed over a dictionary and are injected into the MS bands. The pan-sharpening process is formulated as a restoration problem under sparsity constraint over a trained dictionary in (Wang et al., 2017). The dictionary is composed of several sub-dictionaries, learned using PAN image and corresponding MS image.

The critical issue in all these sparse representation based pan-sharpening methods is the dictionary construction. If the sampled patches are appropriate, then dictionary can represent all the details of remote sensing images. Large sized dictionaries lead to computational complexity. Since these methods use patch based processing, size of the image patch is an important parameter to be considered in yielding the fused image of required quality. Small patches contain little texture information while the large patches lacks generalization capability and involves high computational complexity. In addition, most of the current generation satellite sensors produce PAN and MS images whose spectral ranges are not exactly aligned.

A pan-sharpening method is designed in order to mitigate the effects of spectral range mismatch between PAN and MS image bands, and the computational burden in dictionary learning process. The main principle underlying this method is scale invariance between the sparse coefficients, i.e., the direct correspondance between the coefficients of PAN image and it's low-resolution variant. Since, the PAN and MS images observe the same scene, the MS image patches can be sparsely represented over the dictionary learned from PAN image. The spatial details extracted from PAN image are sparsely reconstructed using a dual dictionary. The HRMS image is obtained by injecting the spatial details into the interpolated and deblurred MS image bands. The detailed implementation of the pan-sharpening method is described in the following section.

3.2 Two-Stage Pan-sharpening Algorithm

An efficient two-stage pan-sharpening algorithm is proposed to moderate the spectral distortion and to preserve the essential spatial details in the fused image. The composition of high-resolution MS (HRMS) image, i.e., the spatial details and spectral information are estimated separately. The spectral mismatch between PAN and MS image bands degrades the quality of fused outcome. The spectral mismatch effect is mitigated by using the sensor's modulation transfer function (MTF) based pre-processing between PAN and MS images combined with the joint dictionary learning paradigm.

The implementation mechanism of the proposed method can be summarized as:

- An MTF-matched deblurring filter is designed and up-sampled MS image is processed with the designed filter in order to reduce the spectral distortion in the fused outcome. Since, the LRMS image can be treated as blurred and downsampled version of HRMS image.
- To extract the most relevant spatial details from each band, a preprocessing step is performed on PAN and MS bands with the filters that are matched with sensor's MTF. The dictionaries are trained with the high frequency details extracted from the PAN image.
- Two dictionaries are trained jointly to justify the coherence between sparse coefficients at different scales.
- In place of K-SVD, a fast and efficient training algorithm namely, profile atom update-dictionary learning (PAU-DL) is used to construct the dictionaries while reducing the computational complexity without compromising the output quality.

3.2.1 The Proposed Pan-sharpening Framework

Let X_k with $k = 1, 2, \dots, B$. denotes MS image with B bands and P be the PAN image. The objective of the pan-sharpening is to produce HRMS image \hat{X}_k , ($k = 1, 2, \dots, B$) having the spatial resolution of PAN image while preserving the spectral characteristics of MS image.

The HRMS image having high spatial and rich spectral resolution characteristics, can be realized on fusing the high-spatial resolution PAN image with the MS image possessing higher spectral resolution. The HRMS image can be represented using linear combination of its low and high-frequency components as,

$$\hat{X}_k = \hat{X}_k^L + \hat{X}_k^H, \quad k = 1, 2, \dots, B. \quad (3.5)$$

The reconstruction of desired high-resolution MS image, \hat{X} consists of the following two phases: The low-frequency components \hat{X}_k^L infers spectral details and can be derived from the given MS image. Whereas, the extricated spatail details from the PAN image are utilized to construct the high-frequency components.

3.2.2 Spectral Component (\hat{X}_k^L) Estimation

Observation of remote sensing image formation model describes that the given MS image can be treated as blurred and down-sampled version of the HRMS image. The spectral component of HRMS image, \hat{X}_k^L can be obtained by interpolating the MS image to the scale of PAN image followed by a deblurring operation. Deblurring is the process of removing blur by applying a filter on the image, considered as an important technique during image restoration process. Wiener filter, Regularized filter and the filter designed based on Lucy-Richardson algorithm are few of the recognized deblurring filters (Gonzalez, 2016). While considering the execution speed and quality of the output, Wiener filter is preferred for deblurring in most of the image processing applications.

The MS image processed with MTF deblurring filter is successfully used for pan-sharpening in CS and MRA techniques (Palsson et al., 2016). The process of generating interpolated and MTF deblurred MS image is described as:

The interpolated MS image (\tilde{X}_k) is obtained by employing bi-cubic interpolation method:

$$\tilde{X}_k = (X_k \uparrow R) \quad (3.6)$$

where $\uparrow R$ represents the up-sampling operation by a factor of R .

The blurring filters that are matched with MTF of the MS sensor are approximated using Gaussian low-pass filters. The gain at Nyquist frequency is usually provided by the sensor specifications. The Wiener deblurred filter, W_k^d , for each band of MS image can be estimated as:

$$W_k^d = \frac{1}{H_k} \cdot \frac{|H_k|^2}{|H_k|^2 + \frac{1}{S_\eta}} \quad (3.7)$$

where, $|H_k|^2 = H_k^* \cdot H_k$.

H_k is the Fourier transform of the blurring filter for k^{th} band.

H_k^* is the complex conjugate of H_k ,

$\frac{1}{S_\eta}$ is the noise to signal power ratio, usually approximated by a constant.

The Fourier transform of the required spectral (low-frequency) component can be determined as:

$$\bar{X}_k^L = W_k^d \cdot (\bar{X}_k) \quad (3.8)$$

\bar{X}_k is the Fourier transform of the \tilde{X}_k . Invoking the inverse Fourier transform of \bar{X}_k^L , the MTF deblurred version of MS image (\hat{X}_k^L) is obtained.

3.2.3 Spatial Component (\hat{X}_k^H) Estimation

The high-frequency details (\hat{X}_k^H) can be estimated by the spatial details derived from PAN image and must be relevant to the each band of MS image. The sparsity is more appropriate for the components having high intensity variations like edges and image textures. In this work, the sparse coefficients invariance over different resolutions is adapted to estimate the high-frequency components. The image patches at different resolutions share the similar sparse coefficients over a jointly learned dictionary.

The PAN image is histogram matched with each band of the up-sampled MS image to produce a modified PAN image P_k . The high-frequency components of the PAN

images can be obtained as:

$$P_k^h = P_k - (P_k) * m_k, \quad k = 1, 2, \dots, B. \quad (3.9)$$

The symbol $*$ denotes the convolution operator, and m_k is the filter matched to the sensor modulation transfer function (MTF) of k^{th} band of MS image. Here, m_k is approximated by a Gaussian filter offering specific sensor gain at the Nyquist frequency.

The PAN image is down-sampled to the resolution of original MS image to get a low-resolution PAN image, LP. The band specific high-frequency components of LP, LP_k^h , for the considered bands can be obtained as:

$$LP_k^h = LP_k - (LP_k) * m_k, \quad k = 1, 2, \dots, B. \quad (3.10)$$

Each band of MS image (X_k) is processed by a MTF matched filter to obtain corresponding low-frequency version. The high-frequency details of each MS band are obtained as:

$$X_{k,h} = X_k - X_k * m_k \quad (3.11)$$

The sparse representation for the patches of each $X_{k,h}$ component is evaluated with respect to a low-resolution dictionary \mathbb{D}_k^{LR} as

$$\min \|\alpha_k\| \quad \text{s.t.} \quad \|X_{k,h} - \mathbb{D}_k^{\text{LR}} \alpha_k\|_2^2 \leq \epsilon \quad (3.12)$$

Following the estimation of sparse coefficients vector, α_k the desired spatial components, $(\hat{Y}_k^{\text{H,ms}})$, can be reconstructed using a high-resolution dictionary \mathbb{D}_k^{HR} as

$$\hat{X}_k^{\text{H}} = \mathbb{D}_k^{\text{HR}} \cdot \alpha_k \quad (3.13)$$

The HRMS image, \hat{X}_k is obtained by merging the spectral component, \hat{X}_k^{L} and the spatial component, \hat{X}_k^{H} .

Pseudo code to implement the proposed method is given as algorithm-1.

Algorithm 1 Two-Stage Pan-sharpening Algorithm

INPUT: PAN image P and low-resolution MS image bands X_1, X_2, \dots, X_B .

For $k= 1$ to B (i.e., for each band).

I. Spectral component (\widehat{X}_k^L) estimation:

1. $(\widetilde{X})_k = X_k \uparrow R$. (\uparrow indicates up-sampling operation).
2. Construct a Wiener deblurring filter for each band of MS image using Eq.3.7.
3. By multiplying the fourier transform of Interpolated MS image with the corresponding Wiener filter yields the spectral component in frequency domain.
4. The inverse Fourier tranform of the outcomes obtained in step (3) results in the required spectral component (\widehat{X}_k^L).

II. Spatial component (\widehat{X}_k^H) estimation:

1. Perform histogram matching between PAN image, P, and each band of MS image X_k to produce image set P_k .
2. Down-sample the image P by a scale factor four to get a low resolution version LP. Perform the histogram matching between P and MS image bands to produce LP_k .
3. Attain the high-frequency components of image set P_k using Eq. 3.9.
4. Attain the high-frequency components of image set LP_k using Eq.3.10
5. $p_{k,h}^i$ and $p_{k,l}^i$ are the column vectors extracted from the patches of P_k and LP_k respectively.
6. Train the dictionaries D_k^{HR} and D_k^{LR} using PAU-DL algorithm as given in Eq.3.14.
7. Rearrange $X_{k,h}$ into column vectors and calculate sparse coefficients corresponding to spatial details to be injected in to k-th band ($\hat{\alpha}_k$) using Eq.3.12.
8. Estimate the spatial details (\widehat{X}_k^H) on multiplying sparse coefficients vector (α_k) with the corresponding dictionary D_k^{HR} as per Eq.3.13.

Reconstruct the HRMS image (\widehat{X}_k) using Eq. 3.5.

End For

OUTPUT: The pansharpened MS image bands $\widehat{X}_1, \widehat{X}_2, \dots, \widehat{X}_B$.

3.2.4 Dual Dictionary Learning Mechanism

To ensure that the high and low-resolution patches have the identical sparse coefficients, the corresponding dictionaries must learn jointly. The high-frequency components of PAN image and its low-resolution version are used to generate the dual dictionaries. The dual-dictionary learning process is formulated as an optimization problem.

For each band, $k = 1, 2, \dots, B$.

$$\langle \mathbb{D}_k^{\text{HR}}, \mathbb{D}_k^{\text{LR}}, \alpha_k \rangle = \underset{(\mathbb{D}_k^{\text{HR}}, \mathbb{D}_k^{\text{LR}}, \alpha_k)}{\text{argmin}}$$

$$\|p_{k,h}^i - \mathbb{D}_k^{\text{HR}} \alpha_k^i\|_2^2 + \|p_{k,l}^i - \mathbb{D}_k^{\text{LR}} \alpha_k^i\|_2^2 \text{ subject to } \forall i \|\alpha_k^i\|_0 \leq T_0 \quad (3.14)$$

with $i = 1, 2, \dots, n$ indicates the number of patches. Where, $p_{k,h}$ denotes the matrix containing columns as patches extracted from the high frequency version of original PAN image P_k^h . Similarly, $p_{k,l}$ represents the patches obtained from the high-frequency version of PAN image realized at low resolution, LP_k^h and $\alpha_k = [\alpha_k^1, \alpha_k^2, \dots, \alpha_k^n]$ is the sparse coefficient matrix for the signals $p_{k,h}$ and $p_{k,l}$ and T_0 is the upper bound of sparsity index.

To impose the sparse coefficients equality, the joint learning of dual dictionaries can be converted into a single sparse decomposition problem as :

$$\langle \mathbb{D}, \alpha_k \rangle = \underset{(\mathbb{D}, \alpha)}{\text{argmin}} \|p_k - \mathbb{D} \alpha_k\|_2^2 \text{ subject to } \forall i \|\alpha_k^i\|_0 \leq r_0 \quad (3.15)$$

where,

$$p_k = \begin{bmatrix} p_{k,h} \\ p_{k,l} \end{bmatrix} \text{ and } \mathbb{D}_k = \begin{bmatrix} \mathbb{D}_k^{\text{HR}} \\ \mathbb{D}_k^{\text{LR}} \end{bmatrix}$$

3.2.5 K-SVD Versus PAU-DL Training Algorithm Analysis

K-SVD is approved as most efficient among the dictionary learning algorithms. In each iteration one atom is updated in the second stage (update stage) of K-SVD algorithm. The objective function of dictionary training/learning using K-SVD algorithm is given as:

$$\min_{(\mathbb{D}, S)} \|Y - \mathbb{D}S\|_2^2 \text{ subject to } \forall i \|s_i\|_0 \leq T_0 \quad (3.16)$$

While updating the i^{th} atom, d_i of the dictionary \mathbb{D} , the non-zero entries in its associated row vector of S (referred as 'profile' of the atom) are also updated. The update process of the atom d_i along with non-zero entries of the i^{th} row of S , $s_{\mathbb{T}}^i$ leads to solving the following minimization problem.

$$\text{minimize } \|\bar{E}_i^{\mathbb{R}} - d_i s_{\mathbb{R}}^i\|_{\mathbb{F}}^2 \quad (3.17)$$

where, $\bar{E}_i^{\mathbb{R}}$ denotes the error matrix without considering the atom d_i and $s_{\mathbb{R}}^i$ denotes the row vector by discarding the zero entries from $s_{\mathbb{T}}^i$. The SVD gives the closest rank-1 approximation of $\bar{E}_i^{\mathbb{R}}$ which is the solution of Eq.3.17.

In solving the dictionary learning problem using K-SVD is computationally complex especially in high dimensions. Parallel Atom-Updating Dictionary Learning (PAU-DL) algorithm, is computationally efficient and mainly differs with K-SVD in the dictionary update phase. The optimization problem in Eq.3.15 is solved by using the PAU-DL algorithm (Sadeghi et al., 2014). The prime difference between K-SVD and PAU-DL lies in the second stage of learning algorithm i.e., the dictionary column update mechanism. The parallel-atom update mechanism also accerates the convergence rate of the algorithm. The update mechanism using PAU-DL is described as following: The overall error matrix is written as

$$\bar{E} = Y - (A_1 + A_2 + \dots + A_J). \quad \forall i, A_i = d_i \cdot s_{\mathbb{R}}^i. \quad (3.18)$$

In K-SVD each A_i is updated by performing 'J' alternative updates of d_i and $s_{\mathbb{R}}$. To update A_i , the error \bar{E}_i is computed using the updated versions of $A_1 \dots A_i$, while $A_{i+1} \dots A_J$ have not yet updated. In similar way all A_i are partially updated. Instead of updating one atom at a time, PAU-DL implements the parallel updation of all the atoms, results in accelerating the convergence rate of the algorithm. The parallel atom update mechanism is described in Algorithm 2. The exhaustive implementation details of PAU-DL are presented in (Sadeghi et al., 2014).

3.3 Results and Analysis

The validity of the proposed method is appraised with four different sensor's datasets namely, QuickBird, IKONOS, Pléiades (Vivone et al., 2015) and WorldView-2. The analysis of obtained results is performed on considering six prominent pan-sharpening methods from different categories. FIHS (Tu et al., 2004) from CS class, AWLP (Otazu

Algorithm 2 Parallel atom update mechanism

Error matrix: $\bar{\mathbf{E}} = \mathbf{Y} - \mathbb{D}\mathbf{S}$

```
for j = 1.....J do
  for i = 1.....N do
     $\bar{\mathbf{E}}_i = \bar{\mathbf{E}} + \mathbf{d}_i \mathbf{s}_i^T$ 
    update  $\mathbf{s}_i^T$ 
    update  $\mathbf{d}_i$ 
     $\bar{\mathbf{E}} = \bar{\mathbf{E}}_i - \mathbf{d}_i \mathbf{s}_i^T$ 
  end for
end for
```

et al., 2005) from MRA, along with SR-Li (Li and Yang, 2011), SR-TD (Cheng et al., 2014), SR-WT (Cheng et al., 2015) and SR-D (Vicinanza et al., 2015) from sparse representation methods family are used for comparison.

For a generous comparison, the low resolution patch size is considered as 2×2 with 1 pixel overlap for all the SR based methods and OMP algorithm is used to estimate sparse coefficients. The proposed method is implemented by constructing two dictionaries: HR dictionary of size 64×512 , and LR dictionary of size 4×512 . The size of patches extracted to train the HR and LR dictionaries are 8×8 pixels and 2×2 pixels, respectively. During the training stage of dictionaries, the number of iterations in PAU-DL is set as 35. The dictionary is initialized with randomly selected training signals. Mutual coherence threshold is selected as 0.95. OMP algorithm with error threshold 1 is used for sparse coefficients estimation.

The performance evaluation of the proposed method is exercised using two distinguished protocols labeled as Wald's protocol (Wald et al., 1997) and QNR protocol (Alparone et al., 2008). The synthesis property of Wald's protocol is considered for the evaluation at reduced-scale. The original images are processed using MTF matched filters and down-sampled by a ratio of 4. The gain at Nyquist frequency to design Gaussian based MTF matched filters for different datasets are given in Table 3.1. The fusion is performed on the degraded data, further the result is compared with the original MS image which is treated as reference image.

The full-scale assessment is carried out with consistency property of Wald's protocol and QNR protocol. According to the consistency property, pan-sharpening is executed on source PAN and MS images. The fusion outcome is degraded to the scale of original

Table 3.1: Gain at Nyquist frequency for different sensors

Sensor	MS Image Bands				PAN
	Blue	Green	Red	NIR	
QuickBird	0.34	0.32	0.30	0.22	0.15
IKONOS	0.26	0.28	0.29	0.28	0.17
Pléiades	0.29	0.29	0.29	0.29	0.15
WorldView-2	0.35	0.35	0.35	0.35	0.11

MS image which is considered as reference image for comparison purpose. Six quality metrics are used for quantitative comparison of the proposed method on considering synthesis and consistency properties are defined as:

- CC: Correlation Coefficient is a similarity index with optimal value of one.
- ERGAS: The French acronym for Relative dimensionless global error in synthesis is a global error index. Low values of ERGAS imply similarity between fused and source multispectral data.
- SAM: Spectral Angle Mapper is usually expressed in degrees and equal to zero if both the images used for comparison are spectrally identical.
- SSIM: It express the structural similarity of two images, its higher value indicates more similarity.
- Q4: It is a multi-spectral extension of Universal Image Quality Index (UIQI) and is suitable for the images with four spectral bands. Its optimal value is one when two images are identical.
- PSNR: Peak Signal to Noise Ratio, higher value indicates less noise ingradient.

Since no reference image is available for comparison in practice, QNR protocol is more reliable for quality assessment at full-scale. QNR protocol comprises of 3 indices; spectral distortion index (D_λ), spatial distortion index (D_s) and Quality with no Reference (QNR).

QNR: Its optimal value is one and is obtained when indices D_λ and D_s are zero.

3.3.1 Evaluation Using Synthesis Property

The QuickBird dataset represents the forest area from Sunderbans-India region. The spatial resolution is 0.7-m for PAN image and 2.8-m for four band MS image (Blue, Green, Red and Near InfraRed (NIR)). The visual results at reduced-scale with Quick-Bird data are shown in Fig.3.1.

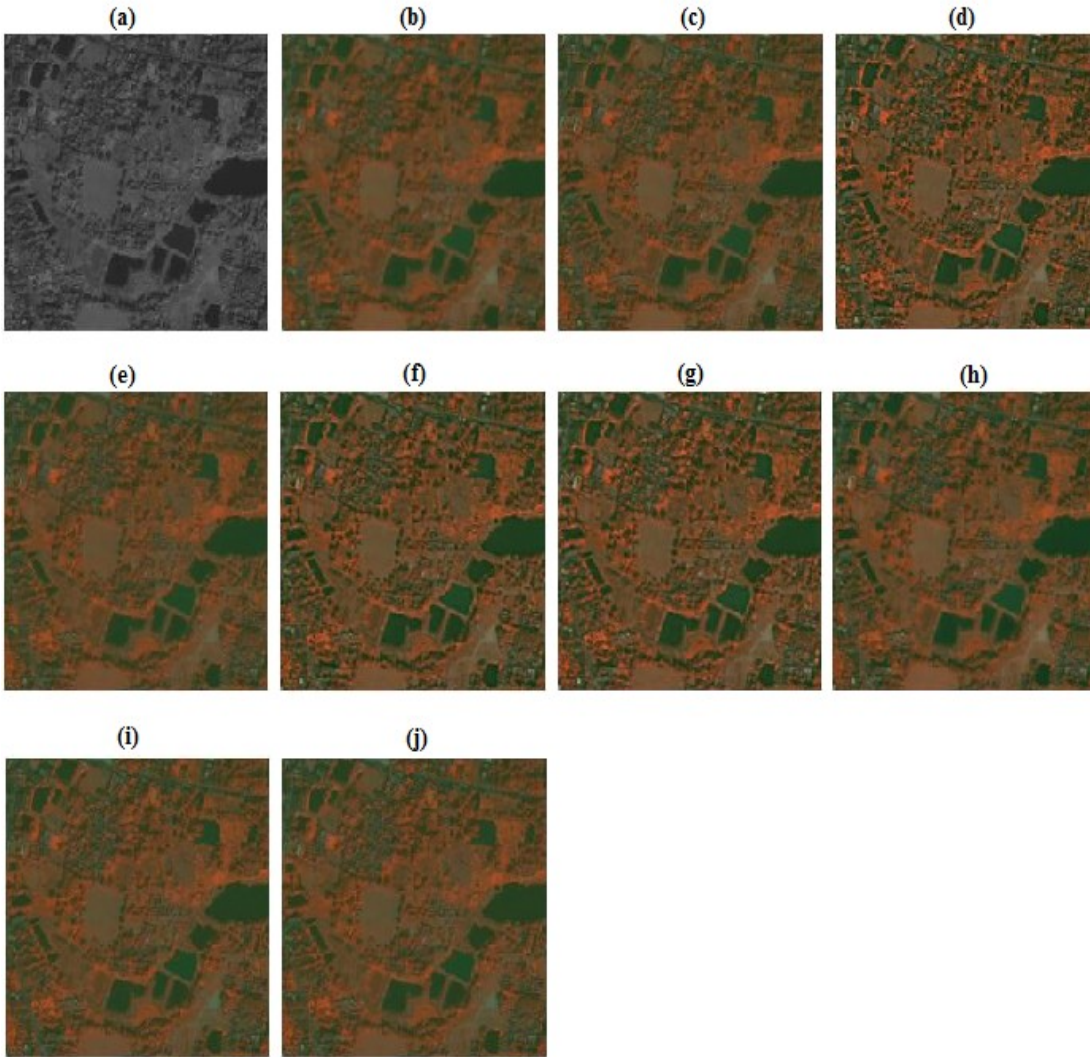


Figure 3.1: Pan-sharpening results evaluated using synthesis property for QuickBird data (a) PAN image (200x200, 2.8-m) (b) Up-sampled MS image (200x200, 11.2-m) (c) Reference MS Image (d) FIHS (e) AWLP (f) SR-Li (g) SR-TD (h) SR-WT (i) SR-D (j) Proposed method

The PAN image of size 200×200 pixels at 2.8-m resolution and up-sampled MS image with 11.2-m resolution are shown in Fig.3.1(a) and (b) respectively. The Reference MS image used for comparison of pan-sharpened outcomes is shown in Fig.3.1(c). The pan-sharpened outcomes of different methods are presented in Fig.3.1(d)-(j). Table 3.2 summarizes the quality metrics corresponding to the visual results shown in Fig.3.1. The optimal value for each index is highlighted in bold case. The outcome of FIHS method (Fig.3.1.(d)) shows color distortion and modest blurring of the spatial details. The highest value of performance measures like SAM and ERGAS obtained for FIHS method is in match with the visual result. The outcome of AWLP method (Fig.3.1(e)) appears to be natural and preserves most of the spatial details that are present in refer-

Table 3.2: Quality metrics for QuickBird data evaluated using synthesis property

	CC	ERGAS	SAM	SSIM	Q4	PSNR
FIHS	0.8735	2.2136	3.0137	0.7643	0.8837	27.0008
AWLP	0.9304	1.9853	2.8798	0.7841	0.8986	29.1164
SR-Li	0.8983	2.0171	2.9537	0.7823	0.8814	27.5725
SR-TD	0.9008	1.8725	2.3471	0.7978	0.8952	28.0172
SR-WT	0.9017	1.8862	2.3646	0.8031	0.8965	28.4213
SR-D	0.9189	1.6342	2.3161	0.8015	0.8978	29.2297
Proposed	0.9285	1.3339	2.1155	0.8163	0.9036	29.1568

ence image. The blocking artifacts are visible particularly at the edges in the result of SR-Li method. There is no noticeable difference between the results of SR-WT and SR-D methods. The proposed method outcome has more proximity with the reference image compared outcomes obtained on using other reported methods.

The absolute difference between pixels value of each fused image and the reference MS image is presented in Fig.3.2. The difference between the fused and reference

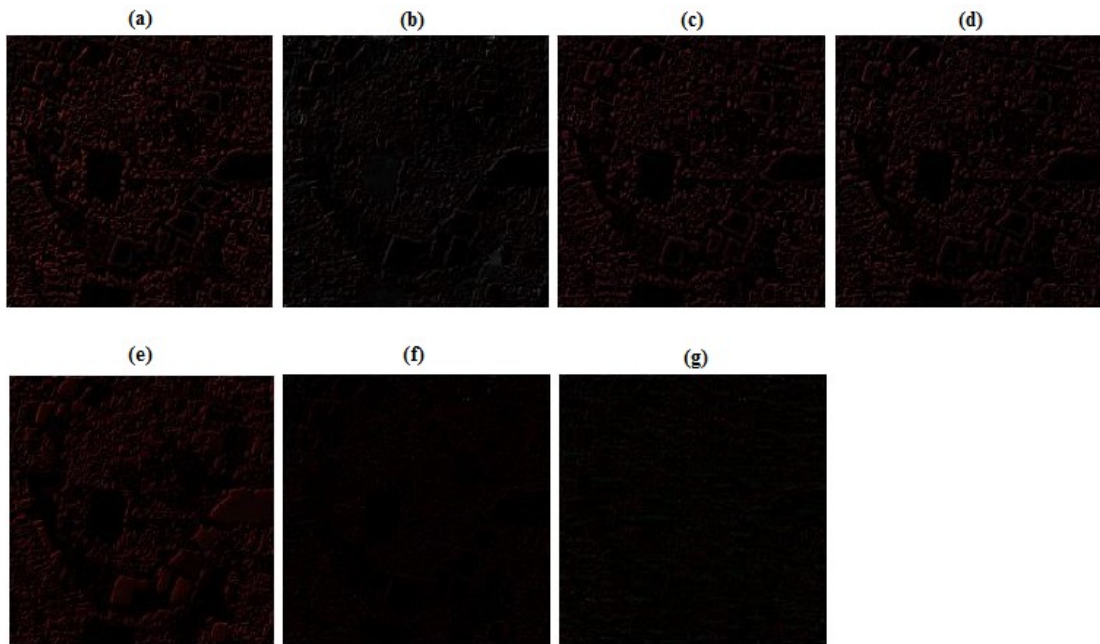


Figure 3.2: Difference between pan-sharpened images for each method and reference MS image in Fig.3.1. (Red colored pixels indicates large differences and black pixels indicate less difference) (a) FIHS (b) AWLP (c) SR-Li (d) SR-TD (e) SR-WT (f) SR-D (g) Proposed method

images are indicated by the extent of brightness in difference images. The areas which appear in bright color indicates the amount of difference, whereas, the smallest pixel differences are indicated by dark regions. On performing analysis of the difference

images presented in Fig.3.2, it can be interpreted that the result obtained using the proposed method is comparatively almost similar to the reference image with minimal accompanied distortion. The difference image obtained using the proposed method (Fig.3.2(g)) appears as black infers minute difference between the pan-sharpened image and the original MS image. The quality metrics are consistent with the visual outcomes.

The performance of the proposed method is further estimated at reduced-scale with another dataset from WorldView-2 sensor. The WorldView-2 sensor provides the PAN and MS images at the spatial resolution of 0.5-m and 2-m, respectively. The visual outcomes for degraded dataset obtained from WorldView-2 are presented in Fig.3.3. Degraded PAN image of size 200×200 pixels at 2-m resolution and MS image at 8-

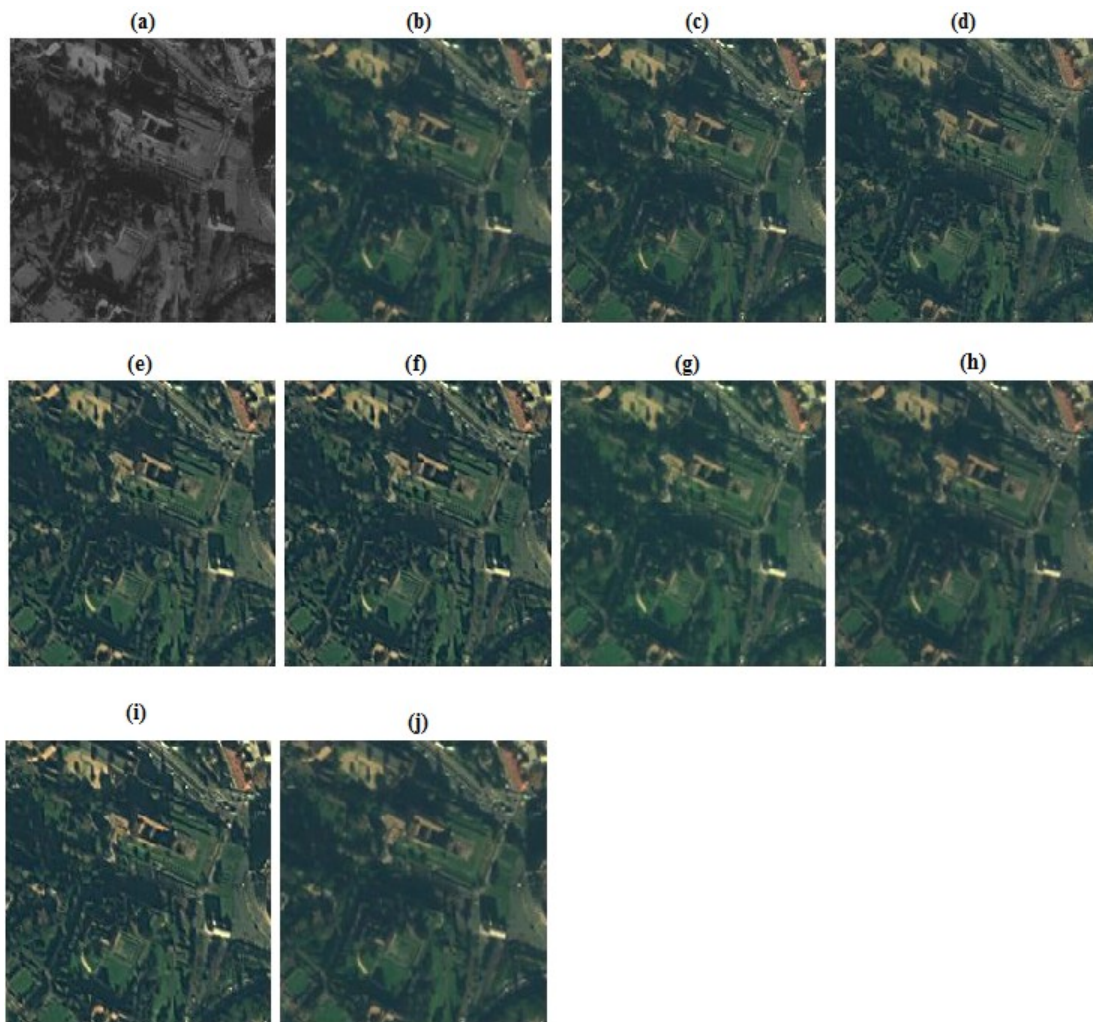


Figure 3.3: Pan-sharpening results evaluated using synthesis property for WorldView-2 data (a) PAN image (200×200 , 2-m) (b) Up-sampled MS image (200×200 , 8-m) (c) Reference MS Image (d) FIHS (e) AWLP (f) SR-Li (g) SR-TD (h) SR-WT (i) SR-D (j) Proposed method

m resolution are shown in Fig.3.3(a) and (b), respectively. The reference MS image is shown in Fig.3.3(c) and the outcomes of different methods are reported in Fig.3.3(d)-(j). The pixels difference between obtained pan-sharpened images on using different methods and reference MS image (Fig.3.3(c)) is presented in Fig.3.4.

The difference image observation as indicated in the Fig.3.4 demonstrates that the proposed method yields least possible pixel difference with the given reference MS image. The outcomes of FIHS and SR-Li methods experience a firm spectral distortion particularly for the regions composed of trees. The details of buildings roof and roads are not sharp enough in the outcome of AWLP method. The visual perception obtained from methods SR-TD, SR-WT and SR-D is nearly same as that of the reference image. Though, the apprehension of slight color changes is difficult to recognize from the visual analysis; the superiority of proposed method is evidenced by quality metrics reported in Table 3.3.

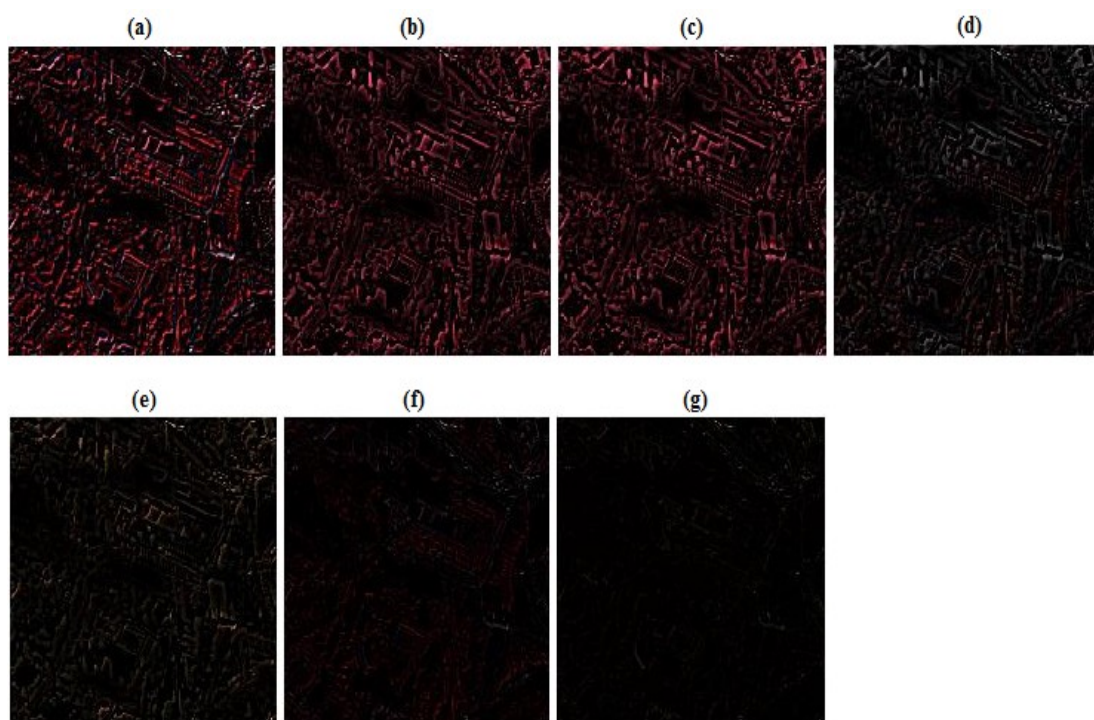


Figure 3.4: Difference between pan-sharpened images for each method and reference MS image in Fig.3.3. (Bright pixels means large difference and black mens less difference)(a) FIHS (b) AWLP (c) SR-Li (d) SR-TD (e) SR-WT (f) SR-D (g) Proposed method

The optimal values are achieved by the proposed method for all the metrics except SSIM and PSNR. Herein, the obtained values of quality metrics demonstrate the adequate performance of the proposed method in terms of retaining desired spatial details

Table 3.3: Quality metrics for WorldView-2 data evaluated using synthesis property

	CC	ERGAS	SAM	SSIM	Q4	PSNR
FIHS	0.9381	3.1275	4.0219	0.8354	0.9025	19.2142
AWLP	0.9533	2.8538	3.6435	0.8571	0.9277	20.5779
SR-Li	0.9458	2.9164	3.9127	0.8473	0.9148	20.3426
SR-TD	0.9482	2.8563	3.6349	0.8681	0.9285	20.4872
SR-WT	0.9615	2.8619	3.6426	0.8489	0.9321	21.2420
SR-D	0.9673	2.8673	3.7011	0.8563	0.9362	21.6537
Proposed	0.9720	2.5932	3.2762	0.8632	0.9401	21.4823

and preserving required spectral contents.

Analysis of visual results and quality metrics at reduced-scale reveal that the proposed method predominantly shows its eminence in enhancing the spatial details as well as the desired spectral information. The difference images presented in Fig.3.2(g) and Fig.3.4(g) manifest that the proposed method results into fused image that is almost similar to the reference MS image.

3.3.2 Evaluation Using Consistency Property

The data captured by IKONOS sensor are investigated to test and validate the proposed method at full-scale using consistency property of Wald's protocol. The images of IKONOS are acquired over China-sichuan region, which is composed of mountainous and vegetated area.

The PAN and 4-band MS image resolution is 1-m and 4-m, respectively. The data without degradation having specifications as: PAN image of size 256×256 at 1-m resolution and re-sampled MS image at 4-m resolution are shown in Fig.3.5(a) and Fig.3.5(b), respectively. The pan-sharpened images realized using different methods are degraded to the scale of original MS image for performance estimation. The visual outcomes of different methods are presented in Fig.3.5(c)-(i).

The quality metrics values presented in Table 3.4 indicate the highest degree of similarity between the pan-sharpened image outcome of the proposed method and the reference MS image. The SAM value attained on using the proposed method is also better than that of obtained from the other methods except for the SR-WT and SR-D. The Q4, SSIM and CC value of all the methods except for FIHS method differs by a narrow margin. It indicates the significant differences that can be observed from image analysis.

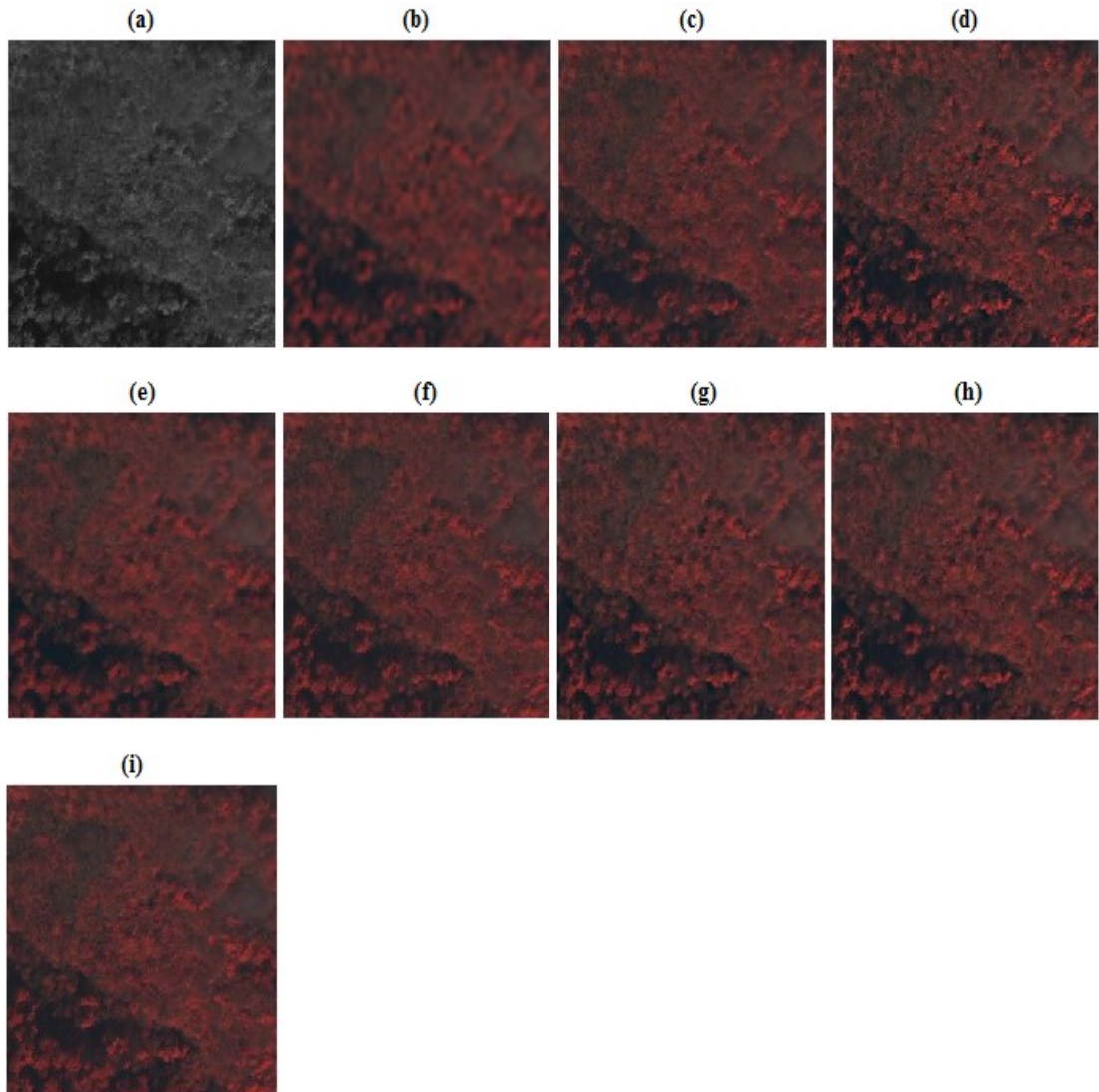


Figure 3.5: Pan-sharpening results evaluated using consistency property for IKONOS data (a) PAN image (256×256 , 1-m) (b) Up-sampled MS image (256×256 , 4-m) (c) FIHS (d) AWLP (e) SR-Li (f) SR-TD (g) SR-WT (h) SR-D (i) Proposed method.

Table 3.4: Quality metrics for IKONOS data evaluated using consistency property

	CC	ERGAS	SAM	SSIM	Q4	PSNR
FIHS	0.9411	2.0517	1.5748	0.8826	0.8960	23.8261
AWLP	0.9883	1.8765	1.2472	0.9609	0.9684	24.6439
SR-Li	0.9764	1.9274	1.3216	0.9535	0.9628	24.1192
SR-TD	0.9796	1.5876	1.3107	0.9574	0.9800	25.2472
SR-WT	0.9834	1.5770	1.1452	0.9624	0.9739	25.5769
SR-D	0.9873	1.3937	1.1483	0.9642	0.9752	26.3123
Proposed	0.9864	1.2962	1.1604	0.9783	0.9821	25.8954

3.3.3 Evaluation Using QNR Protocol

In this section, the performance of the proposed method is evaluated using the datasets without degradation. Fig.3.6(a) and (b) shows the QuickBird images; PAN image (256×256 pixels) at 0.7-m resolution and MS image at 2.8-m resolution, respectively. The visual outcomes for different methods at full-scale are presented in Fig.3.6(c)-(i).

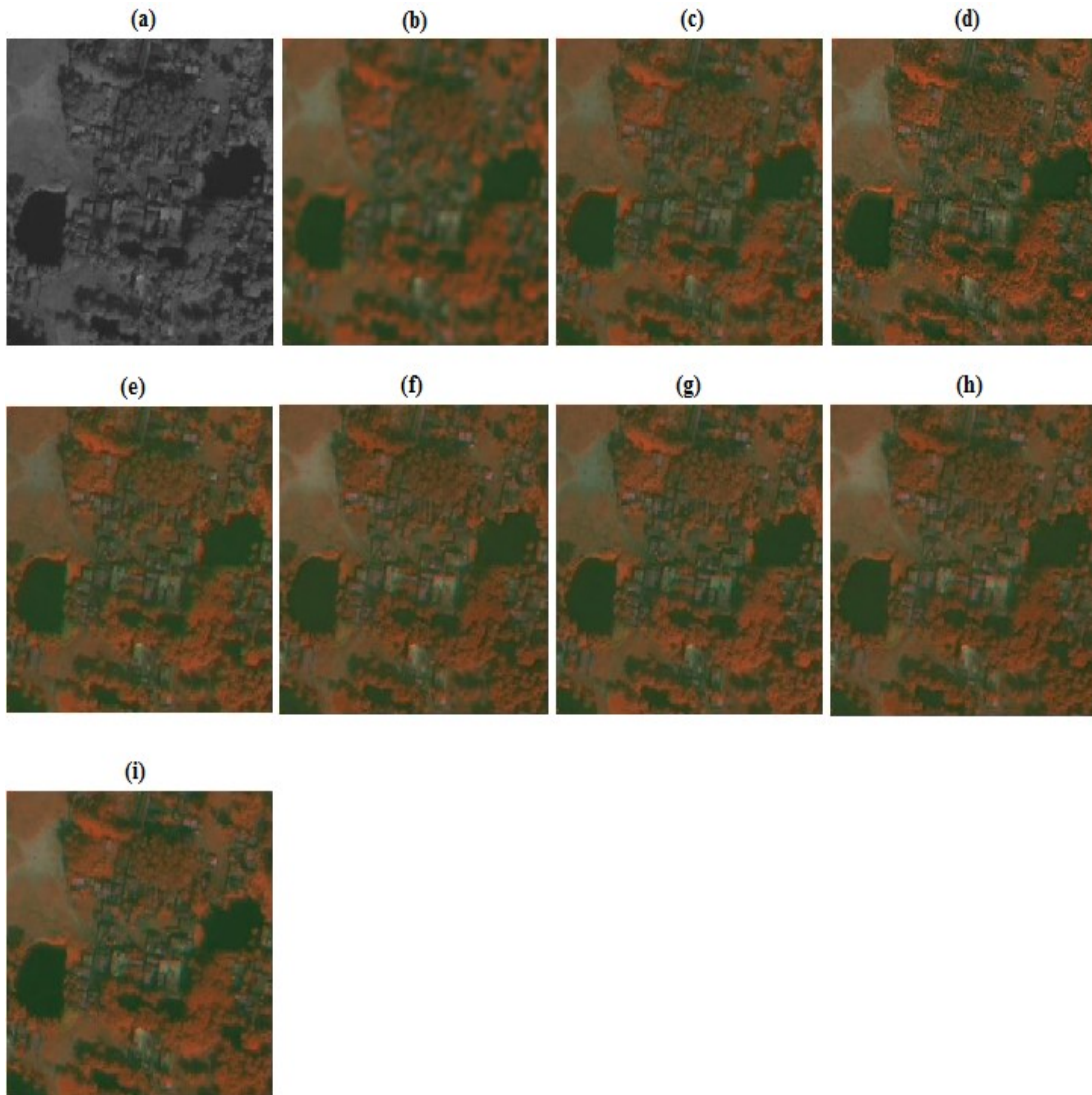


Figure 3.6: Pan-sharpening results evaluated using QNR protocol for QuickBird data (a) PAN image (256×256 , 0.7m) (b) Up-sampled MS image (256×256 , 2.8-m) (c) FIHS (d) AWLP (e) SR-Li (f) SR-TD (g) SR-WT (h) SR-D (i) Proposed method

The observations of visual results reveal that the outcome of FIHS method as shown in Fig.3.6(c) exhibits slight color changes and distortion in the form of blurred spatial details in some parts of the image. The roofs and edges in the fused image experiences

blurring in AWLP method's outcome. The results of SR-based methods are relatively similar, whereas, modest sharp details are visible in the outcome of SR-D method. The outcomes of the proposed method (Fig.3.6(i)) possesses best performance features value of both spatial and spectral details. Table 3.5 reports the quality metrics associated with the visual results presented in Fig.3.6. The highest value for QNR index justifies that the proposed method is superior than the other considered methods.

Table 3.5: Quality metrics for QuickBird data evaluated using QNR protocol

	D_λ	D_s	QNR
FIHS	0.0467	0.1317	0.8277
AWLP	0.0426	0.0815	0.8794
SR-Li	0.0297	0.0795	0.8931
SR-TD	0.0276	0.0515	0.9223
SR-WT	0.0301	0.0423	0.9289
SR-D	0.0317	0.0338	0.9312
Proposed	0.0272	0.0396	0.9343

Further, the proposed method is appraised with Pléiades dataset. The Pléiades dataset represents an urban area from Toulouse city of France. The MS bands maintain 60-cm resolution and the PAN image is produced by simulation. The PAN and up-sampled MS images are shown in Fig.3.7(a) and (b), respectively. The pansharpened outcomes of different methods for Pléiades data at full-scale are presented in Fig.3.7(c)-3.7(i). The quantitative evaluation results are presented in Table 3.6. The visual analysis and quality metrics values illustrate the improved performance of proposed method over the other reported methods.

The percentage improvement of SAM index for the proposed method over the second highest values are: 8.7% for QuickBird and 9.8% for WorldView-2 datasets at reduced-scale evaluation (Table 3.2 and 3.3 respectively). For the two tested datasets at full-scale the proposed method acquires lowest values for spectral distortion index D_λ . These values approve the substantial reduction in spectral distortion achieved by adopting the deblurring operation for up-sampled MS image bands. Similarly, the highest value for the metrics Q4 and QNR for all the tested datasets, manifests the capability of the proposed method in enhancing the overall quality of the pan-sharpened output.

The visual outcomes and quality index values approve that the de-blurring operation results in spectral distortion reduction. Further, the dual-dictionary learning process enhances the spatial quality of fused image.



Figure 3.7: Pan-sharpening results evaluated using QNR protocol for Pléiades data (a) PAN image (256×256 , simulated) (b) Up-sampled MS image (256×256 , 60-cm) (c) FIHS (d) AWLP (e) SR-Li (f) SR-TD (g) SR-WT (h) SR-D (i) Proposed method

Table 3.6: Quality metrics for Pléiades data evaluated using QNR protocol

	D_λ	D_S	QNR
FIHS	0.0371	0.0584	0.9066
AWLP	0.0393	0.0416	0.9207
SR-Li	0.0258	0.0385	0.9367
SR-TD	0.0225	0.0219	0.9561
SR-WT	0.0118	0.0202	0.9682
SR-D	0.0086	0.0163	0.9752
Proposed	0.0079	0.0126	0.9796

3.4 Conclusion

While establishing the striking comparison with the traditional SR-based methods, the proposed method comprises of two phases. During first phase after making usage of interpolated MS bands, these MS bands are deblurred. The MTF deblurring process predominantly reduced the spectral distortion. The most relevant spatial information extricated from the PAN image is injected into these MS bands in the second phase. The dual dictionaries are trained using the patches of high-frequency details extracted from the PAN image and its low-resolution version. A computationally efficient algorithm (PAU-DL) is used for dictionary learning. The essential spatial details are enhanced by exploiting the scale invariance concept in conjunction with a jointly learned dictionary. The prevalence of the proposed method is conspicuously demonstrated by the visual outcomes and quantitative performance measure metrics. The proposed method is evaluated using four different datasets at both reduced-scale and full-scale resolutions. The visual outcomes and quantitative results approve the quality of the fused outcome.

CHAPTER 4

SPARSITY INSPIRED PAN-SHARPENING TECHNIQUE USING MULTI-SCALE LEARNED DICTIONARY

The significant issues in remote sensing image fusion are enhancing the spatial details and preserving the essential spectral information. The classical pan-sharpening methods often incur spectral distortion and still striving to produce the fused images with prominent spatial and spectral attributes. Motivated by the enhanced quality outcomes of sparse representation (SR) theory, a novel pan-sharpening method is developed based on sparse representation of high-frequency (HF) components over a multi-scale learned dictionary (MSLD). MSLD technique acquires the capability of extracting the intrinsic characteristics of images, wherein, it possesses the features of both multi-scale representation and learned dictionaries.

An efficient pan-sharpening (PS) method is developed that effectively combines the ARSIS scheme with sparse representation of spatial details over a multi-scale learned dictionary. Ophir et.al,[\(Ophir et al., 2011\)](#) proposed the multi-scale dictionary (MSD) learning concept by merging the existing two approaches for the choice of dictionary: analytic approach and learning based approaches, in order to gain the advantages of both. [\(Yin, 2015\)](#) employed SR with MSLD (multi-scale learned dictionary) concept for the fusion of medical images, by using a joint dictionary learning approach. The proposed method exploits the hypothesis of direct coherence between the patches over different scales to spatial components enrich with high-frequency details, i.e., the sparse coefficients of spatial details extracted from PAN image and its low-resolution variant images are comparable over the pertinent dictionaries. In contrast to the methods [\(Jiang et al., 2014; Vicinanza et al., 2015\)](#), the proposed PS-MSLD method explores the dictionaries learning in wavelet domain. The dictionaries are trained by the respective high-frequency details extracted from the two versions of available PAN image realized at different spatial resolutions. A quick and efficient method, Atom-Profile Updating Dictionary Learning (APrU-DL) [\(Sadeghi et al., 2014\)](#) is used for training the dictionaries. The APrU-DL algorithm has proven its efficiency over the other dictionary learning methods, not only for the substantial reduction of computational complexity but also in accomplishing the more accurate outputs.

The PS-MSLD method inherits the following novelties.

- Since the dictionaries are learned using the details of the PAN image, it improves adaptiveness of the dictionary and also the robustness of pansharpening process.
- Multi-scale learned dictionaries acquire the features of both analytic and learned dictionaries, yield better results in remote sensing image fusion. Further, the MSLD enhances the representation of inherent features of the images, with efficient sparse coding.
- The APrU-DL algorithm used for training, ameliorate the quality of output with better convergence rate.

4.1 Multi-Scale Learned Dictionary

The learning algorithms essentially attempt to construct an appropriate dictionary, that can provide finest sparse approximation of the given image samples with respect to the trained dictionary.

The multi-scale learned dictionary (MSLD) inherits the features of both analytic and learned versions of dictionaries and has the advantage of easily implementable sparse coding of signals with only local operations. The concept of learning a dictionary from a set of training signals is represented as follows:

$$\operatorname{argmin}_{D,S} \|Y - DS\|_F^2 \quad \text{subject to } \|s_i\|_0 \leq T_0 \quad \forall i \quad (4.1)$$

where $Y = \{y_1, y_2, \dots, y_N\} \in \mathbb{R}^{n \times N}$ is a matrix of columns (y_i) being training signals. $S = \{s_1, s_2, \dots, s_N\} \in \mathbb{R}^{K \times N}$, with s_i as columns is a matrix of sparse representations for y_i . Let $D \in \mathbb{R}^{n \times K}$ ($n \ll K$) be an over complete dictionary. T_0 denotes the admitted sparsity level for each column s_i and $\|\cdot\|_F$ represents the frobenius norm. On amending Eq.(4.1), the multi-scale dictionary learning using wavelets can be expressed as,

$$\operatorname{argmin}_{D,S} \|W_A Y - DS\|_F^2 \quad \text{subject to } \|s_i\|_0 \leq T_0 \quad \forall i \quad (4.2)$$

here, W_A is the wavelet analysis operator.

According to this formulation the data samples can be represented as sparse combination of atoms from a multi-scale core dictionary like wavelets. Since the wavelet domain is a combination of bands at different scales, the multi-scale dictionary learning process can be summarized as learning separate dictionaries for different bands which

in turn are oriented in different directions, namely horizontal, vertical and diagonal. Thereby, the dictionary learning in the transform domain is expressed as:

$$\forall b \underset{\mathbb{D}, \mathbb{S}}{\operatorname{argmin}} \|(W_A Y)_b - \mathbb{D}_b S_b\|_F^2 \quad \text{subject to } \|s_{i,b}\|_0 \leq T_0 \quad \forall i \quad (4.3)$$

here, b denotes index for wavelet sub-bands.

Aharon et al. (2006) proposed a well-established and prominently used algorithm named as K-SVD for dictionary learning, which is flexible and compatible with any sparse coding technique. K-SVD is an iterative method that alternates between two intricate steps called sparse coding and dictionary update. In the sparse coding stage; based on the current dictionary \mathbb{D} , the algorithm solves the Eq.4.1 for sparse coefficient vector (s_i) for each column of the training signal, Y . During the second stage; the columns of dictionary are updated together with the non zero coefficients of the sparse vectors. For the first stage, any sparse coding algorithm can be used. In the dictionary update stage, each atom of the dictionary \mathbb{D} , and the non-zero elements of its corresponding row vector in coefficient matrix S (usually labeled as profile) are modified in each iteration, which leads to the faster algorithm convergence.

Sadeghi et al. (2014) developed an efficient dictionary learning algorithm labeled as APrU-DL, by proposing few modifications for both the phases of K-SVD. To enhance the performance of sparse coding process, debiasing technique is used in the first stage. The dictionary atom along with its entire profile (not only non zero elements) is allowed to change in the second stage. These modifications greatly enhance the global convergence rate and reduce the computational complexity in learning process without compromising the performance.

4.2 Pan-sharpening Using Multi-Scale Learned Dictionary (PS-MSLD)

Let X_k , ($k = 1, 2, \dots, B$) be the MS image with B bands, and P be the high resolution PAN image, used as source images for fusion. Let \tilde{X}_k , ($k = 1, 2, \dots, B$) be the MS image up-sampled to the resolution of PAN image with N number of bands and LRP be the low-resolution (LR) PAN image obtained by down-sampling the PAN image to the resolution of MS image. Presuming \hat{X}_k , ($k = 1, 2, \dots, B$) represents the fused bands of high-resolution (HR) MS image obtained by pan-sharpening.

Wavelet transform is specifically a multi-scale decomposition operator. The HR and LR dictionary formulation strategy suggests that the dictionaries are trained in wavelet domain instead of the image domain. The wavelet analysis domain is a collection of coefficient images. The different coefficient images (or sub-images) contain data at different scales and in different orientations (horizontal, vertical and diagonal). Hence, to represent these coefficient images usage of separate sub-dictionaries is more justifiable. The patches used for dictionary training in image domain assert merely the local correlation between pixels. Whereas, a small patch in the wavelet domain affects large area in the image domain, and this effect increases proportionally with the increase of wavelet decomposition level. The learning process in wavelet domain captures the correlation between adjacent wavelet coefficients. The attractive feature of learning dictionary in transform domain is the ease of implementing the sparse coding process.

The high intensity variations in the images i.e., high-frequency components are more appropriate for sparsity. The Discrete wavelet transform (DWT) divides the original image into four sub images at its first level of decomposition. The detail coefficient images i.e., horizontal, vertical and diagonal component images resolves the edges and curves aligned in respective directions. To construct the dictionary, the training signals are taken from these details images by using patch processing. i.e., images are divided into overlapping patches and then the overlapped patches are converted into column vectors. Hence, the dictionaries used in this method are trained by the vectors that are constructed using the coefficients carrying high-frequency details.

The block diagram of the proposed method is shown in Fig.4.1. The schematic diagram for dictionary construction from the intensity matched images is shown in Fig.4.2.

The implementation of PS-MSLD method consists of the following four stages; HR dictionary construction, LR dictionary construction, sparse coefficients estimation, HRMS image reconstruction.

4.2.1 HR Dictionary Construction

A high-resolution (HR) multi-scale dictionary is learned with the high-frequency (HF) sub-images derived from PAN image, P by applying wavelet transform.

Let HIP_k , ($k = 1, 2, \dots, B$) be the new PAN image obtained by performing the histogram matching of PAN image, P with each band of up-sample MS image, \tilde{X}_k . Towards it, generate one HR sub-dictionary corresponding to every HIP image as follows:

Each HIP image is decomposed by using wavelet transform to obtain three relevant high-frequency sub-images (HIP_H, HIP_V, HIP_D). The letters H,V and D in suffix represents the horizontal, vertical and diagonal directional sub-images, respectively extracted from the corresponding image by applying wavelet transform. Each of these sub-images are arranged as patches of size $(p \times p)$ and are aligned into column vectors. These column vectors are used as training data to obtain the elements of HR dictionary,

$$\mathbb{D}_k^{HR} = \begin{bmatrix} D_{H_k}^{HR} & & \\ & D_{V_k}^{HR} & \\ & & D_{D_k}^{HR} \end{bmatrix} \quad (4.4)$$

where, $k = 1, 2, \dots, B$.

The elements of the above matrix are the dictionaries trained by MSLD with relevant HF data samples as training data. i.e., $D_{H_k}^{HR}$ is the dictionary trained by the data samples taken from horizontal sub-image obtained from HIP_k image.

4.2.2 LR Dictionary Construction

A low-resolution (LR) dictionary is constructed from the LIP image bands. Generate one LR sub dictionary corresponding to every LIP image (LIP_k), that is obtained by performing histogram matching with each band of MS image (X_k) with LRP image. The LR sub-dictionaries can be represented as

$$\mathbb{D}_k^{LR} = \begin{bmatrix} D_{H_k}^{LR} & & \\ & D_{V_k}^{LR} & \\ & & D_{D_k}^{LR} \end{bmatrix} \quad (4.5)$$

where $D_{H_k}^{LR}$ is the dictionary trained by the column vectors extracted from horizontal sub-image patches of size $(p/R \times p/R)$ of LIP_k image (where R is the resolution ratio between MS and PAN images). Similarly, $D_{V_k}^{LR}$ and $D_{D_k}^{LR}$ are the trained dictionaries obtained by vertical and diagonal sub-images of LIP_k , respectively.

The histogram matching process assists in improving the spatial similarity between the PAN and MS images. To enhance the dictionary properties like abundance in training and shift-invariance, the training patches are considered with maximum possible overlapping. The blocking effects are introduced if there is no overlapping. Whereas, redundancy increases with the increase of overlapping between the patches. Hence, the

amount of overlapping should be selected so as to balance the quality of reconstructed fused image and the associated computational complexity.

The pan-sharpening algorithms developed in (Jiang et al., 2014; Vicinanza et al., 2015) justify the scale correspondence between HR and LR image patches. In the proposed PS-MSLD method, the patch similarity is exploited for the high-frequency details that are derived from the PAN image at different scales. The ratio between the HR and LR dictionary patches is equal to the resolution ratio between MS image and PAN image. The two dictionaries, namely \mathbb{D}^{HR} and \mathbb{D}^{LR} used in this method are constructed at PAN scale and MS scale, respectively. If the patch size for LR dictionary construction is selected as $M \times M$ pixels, to cover the same area the corresponding patch size in HR dictionary must be $R^2(p \times p)$. Thereby, the ratio between HR and LR patches is same as the spatial resolution ratio between PAN and MS images. Hence, the HR and LR dictionaries inherit the same size.

4.2.3 Sparse Coefficients Estimation

Owing to the fact that PAN and MS images are captured simultaneously over same area, the details of MS image bands can be sparsely represented using a dictionary learned from the PAN image. The possibility of sparse representation of the hyperspectral image patches over the similar patches extracted from PAN image, has been used for super resolution of hyperspectral images (Zhao et al., 2011, 2014). Though, in this fusion method based on the hypothesis that two images are acquired over the same scene, it is presumed that the spatial information contained in PAN image and MS bands is similar.

With the basis of patch invariance, it is assumed that the high-frequency details of LR and HR image patches admit the similar sparse coefficients over the corresponding dictionaries \mathbb{D}_k^{LR} and \mathbb{D}_k^{HR} , respectively. By using OMP algorithm the sparse coefficients are estimated from MS image patches. Each band of MS image (X_k), ($k = 1, 2, \dots, B$.) is decomposed using wavelet transform to get the three corresponding HF sub-images. Further, on arranging each of the sub-image as overlapping patches of size $(p/R \times p/R)$ and subsequently these patches are represented by column vectors namely, H_{ms_k} , V_{ms_k} and D_{ms_k} . Usage of these column vectors lead to a coefficient matrix as $C_k = [H_{\text{ms}_k}, V_{\text{ms}_k}, D_{\text{ms}_k}]^T$. The sparse coefficients matrix corresponding to k^{th} band of MS image (X_k) can be estimated on solving the following sub problem.

$$\hat{\alpha}_k = \operatorname{argmin} \|\alpha_k\|_0 \quad \text{s.t.} \quad \|C_k - \mathbb{D}_k^{\text{LR}} \alpha_k\|_2^2 \leq \epsilon \quad (4.6)$$

The Eq.4.6 is solved by processing the each element of C_k and D_k^{LR} individually to obtain the sparse coefficient matrix α_k , where $\alpha_k = [\alpha_{H_k}, \alpha_{V_k}, \alpha_{D_k}]^T$. Similarly, the sparse coefficients for all the bands of MS image are obtained by solving the Eq.4.6 for $k = 1, 2, \dots, B$.

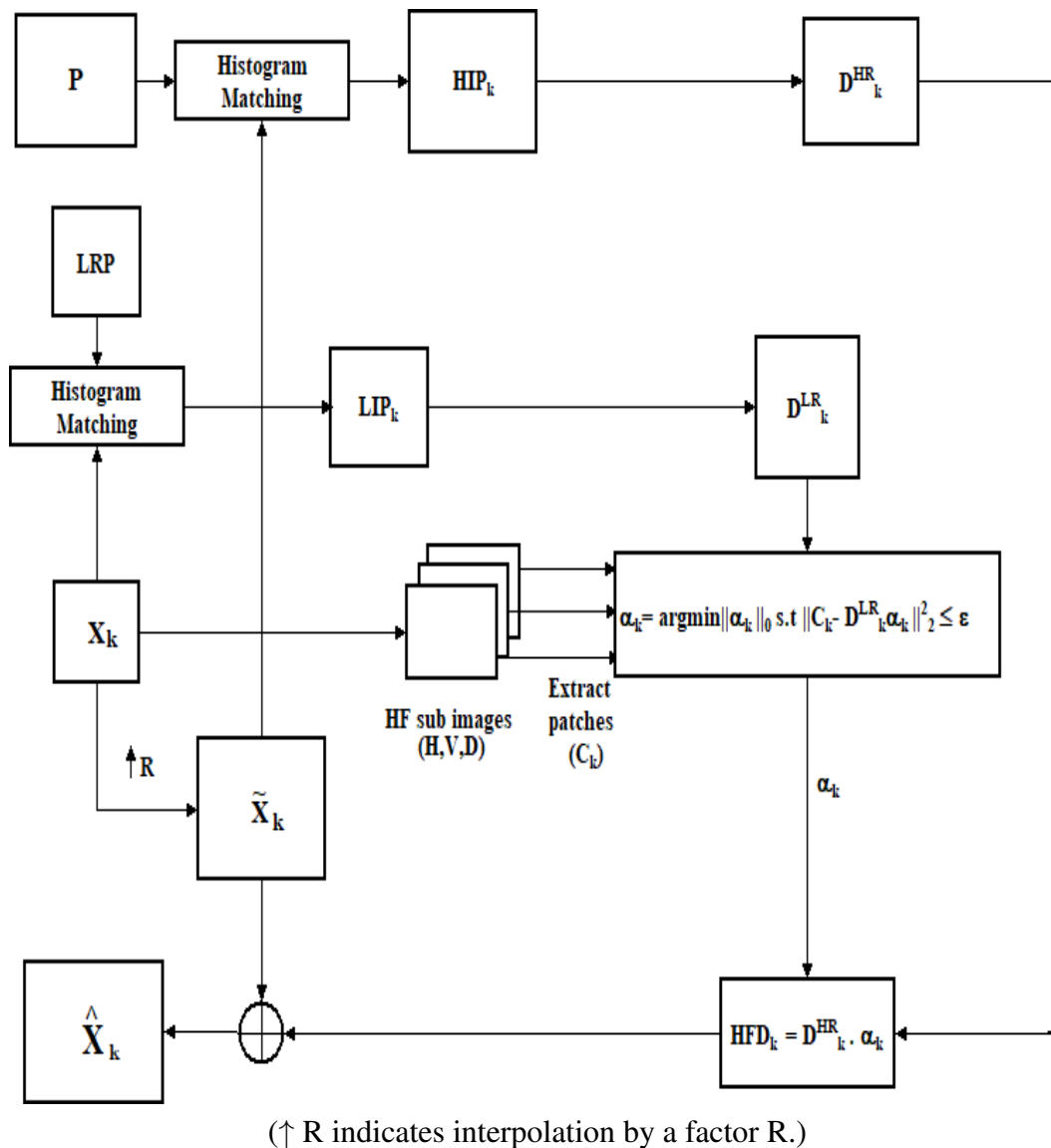


Figure 4.1: The block-diagram of PS-MSLD

4.2.4 Pan-sharpened Image Reconstruction

The patches of missing high-frequency details for each band are obtained on multiplying the sparse coefficients α_k with the corresponding HR dictionary D_k^{HR} .

$$[\text{HFD}_{H_k}, \text{HFD}_{V_k}, \text{HFD}_{D_k}]^T = \mathbb{D}_k^{\text{HR}} \cdot \alpha_k \quad (4.7)$$

Rearranging the patches obtained from Eq.4.7 as images and on performing the inverse wavelet transform on these image patches. This process results into the missing HF details (HFD_k) corresponding to each band. The high resolution MS image, \widehat{X}_k is constructed by injecting the HFD_k into up-sampled MS image bands by using the additive model of the ARSIS scheme as,

$$\widehat{X}_k = \widetilde{X}_k + (\text{HFD}_k), \quad k = 1, 2, \dots, B. \quad (4.8)$$

The four phases described above are implemented for each band of the source MS image and histogram matched PAN image bands, to obtain the final HRMS image.

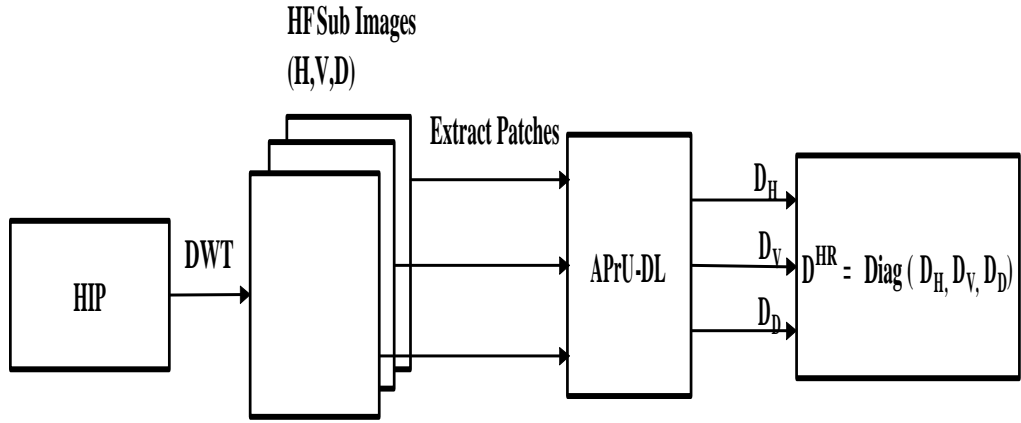


Figure 4.2: Schematic diagram for dictionary learning

4.2.5 Pseudo code for PS-MSLD algorithm

To implement the proposed PS-MSLD algorithm, the working mechanism in the form of pseudo code can be presented as:

Pseudo code: PS-MSLD Algorithm for Pan-sharpening

Input: Panchromatic (P) and N-band MS image (X_k)

Output: High-Resolution MS image (\widehat{X}_k)

for each band of MS image $k = 1 \dots B$ do,

- $\tilde{X}_k = (X_k) \uparrow R$. (Up-sampling the k -th band of MS image by a factor R).
- $HIP_k = \text{Histogram Matching}(P, \tilde{X}_k)$.
- **H1**: $[HIP_H, HIP_V, HIP_D] = \text{DWT}(HIP_k)$.
- For each of the sub-images obtained as given in step **H1**, extract the overlapping patches and arrange the patches as column vectors.
- For each set of training vectors : use APrU-DL algorithm to generate sub-dictionaries D_H^{HR}, D_V^{HR} and D_D^{HR} .
- $\mathbb{D}_k^{HR} = \text{diag}(D_H^{HR}, D_V^{HR}, D_D^{HR})$.
- $LRP = (P) \downarrow R$. (Down-sampling of HRP image by a factor R).
- $LIP_k = \text{Histogram Matching}(LRP, X_k)$.
- **L1**: $[LIP_H, LIP_V, LIP_D] = \text{DWT}(LIP_k)$.
- Obtain three sets of training signals by patch processing using the sub-images obtained as given in step **L1**.
- Invoke APrU-DL algorithm to each training set and generate sub-dictionaries, $D_{H_k}^{LR}, D_{V_k}^{LR}$ and $D_{D_k}^{LR}$.
- $\mathbb{D}_k^{LR} = \text{diag}(D_{H_k}^{LR}, D_{V_k}^{LR}, D_{D_k}^{LR})$.
- $[MS_{H_k}, MS_{V_k}, MS_{D_k}] = \text{DWT}(X_k)$.
- $H_{ms_k}, V_{ms_k}, D_{ms_k}$ are the extracted patches.
- $C_k = [H_{ms_k}, V_{ms_k}, D_{ms_k}]^T$.
- $\hat{\alpha}_k = \text{argmin} \|\alpha_k\|_0 \text{ s.t. } \|C_k - \mathbb{D}_k^{LR} \alpha_k\|^2 \leq \epsilon$.
- $[HFD_{H_k}, HFD_{V_k}, HFD_{D_k}] = \mathbb{D}_k^{HR} \cdot \alpha_k$.
- $HFD_k = \text{IDWT}(HFD_{H_k}, HFD_{V_k}, HFD_{D_k})$.
- $\hat{X}_k = \tilde{X}_k + HFD_k$.

End for

4.3 Experimental Results

The validation of the PS-MSLD method is performed at both reduced-scale and full-scale using the datasets composed of different geographical terrains. The synthesis property of Wald protocol ([Wald et al., 1997](#)) is used to evaluate performance measures at reduced-scale. As per this protocol, the given source images (PAN image and MS image) are decimated by a factor R (which is equal to the resolution ratio between MS and PAN images). Subsequently, fusion is performed on the decimated images, and the fusion scheme outcomes are compared with the original MS image. The original MS image is treated as a reference image for comparison. For the experimentation at full-scale, under the light of no reference image availability, fusion is performed on real dataset. The metric ‘Quality with No Reference’ (QNR) ([Alparone et al., 2008](#)) is used for performance evaluation.

The following pan-sharpening methods are used to perform comparative analysis with the proposed PS-MSLD method.

- EXP: Interpolated MS image without any details injection. The Matlab function ‘imresize’ is used in conjunction with bi-cubic interpolation method.
- GIHS: Generalized Intensity Hue Saturation method ([Tu et al., 2001](#)).
- SW: The Wavelet fusion scheme with substitution model ([Amolins et al., 2007a](#)).
- GSA: Gram-Schmidt Adaptive method ([Aiazzi et al., 2007](#)).
- MGC: (MTF-GLP-CBD) Modulation Transfer Function matched filter with Generalized Laplacian Pyramid. The regression based injection model is used in this method ([Alparone et al., 2007](#)).
- AWLP: Additive Wavelet Luminance Proportional method ([Otazu et al., 2005](#)).
- SR-TD : Sparse Representation based pansharpening using Trained Dictionary ([Cheng et al., 2014](#)).
- SR-WT: Fusion via Wavelet Transform and Sparse Representation ([Cheng et al., 2015](#)).
- SR-D : Sparse Representation of injected Details ([Vicinanza et al., 2015](#)).

The implementation details of GIHS, GSA, MGC and AWLP methods are taken from the software package developed by ([Vivone et al., 2014](#)), in that the decomposition level for AWLP is set as two. For the SR-TD method, the dictionary is trained over 50 iterations by using 10000 random HRMS patches together with the corresponding

HR PAN patches. The initial dictionary consists of 2500 random HRMS patches. K-SVD algorithm is used for training purpose and the OMP algorithm is used for sparse coding. The remaining parameters are considered as given in (Cheng et al., 2014). The fusion methods presented in SW (Amolins et al., 2007b) and SR-WT (Cheng et al., 2015) are extended for four bands in order to have consistent comparison with the other considered methods. For SR-WT, the parameters are chosen as : α_i in IHS is 1/4, in SFIM the averaging filter size is 5×5 , OMP algorithm for sparse coding and the training dictionary size is 64×256 . For SR-D method the the dictionary sizes are 64×256 at high-resolution and 4×256 at low-resolution, respectively. The sparse coefficients are estimated using OMP algorithm. In the proposed PS-MSLD method, Haar wavelet transform with single level decomposition is used to extract HF coefficients. Haar is the first and simplest wavelet orthonormal basis. Haar wavelet is exactly reversible without introducing edge effects. On the contrary; other bases like Daubechies and its derivatives are smoother than Haar. Since high-frequency details are required to construct the dictionaries, Haar is relatively better compared to the other wavelet bases that yield smoothing effect. During the training stage of dictionaries, the number of iterations in APrU-DL is set as 50. OMP algorithm with error threshold 1 is used for sparse coefficients estimation. Since the training data is chosen directly from the available PAN image, the learning mechanism improves the adaptability of dictionaries \mathbb{D}^{HR} and \mathbb{D}^{LR} . The size of HR and LR sub-dictionaries are 64×512 and 4×512 , respectively. The up-sampled MS image, \tilde{X}_k is obtained from the original MS image, X_k by invoking bi-cubic interpolation method (The MATLAB Function ‘imresize’ is used to perform the interpolation operation).

For the consistent comparison, the patch size of 8×8 with 4 pixels overlapping is considered for all SR based methods. In the proposed PS-MSLD scheme, to construct LR dictionary the patch size is 2×2 with 1 pixel overlapping and the corresponding HR dictionary patch size is 8×8 with 4 pixels overlapping is considered. The resolution ratio (R) between MS and PAN images is 4 for all datasets considered for the evaluation of PS-MSLD scheme. All the methods are implemented in MATLAB on pentium 3.1GHz, i5- processor, windows 7 PC with 8GB memory.

4.3.1 Experimental Results at Reduced-Scale

The datasets acquired from three different sensors, namely, Pléiades, QuickBird and IKONOS are used to estimate the performance of the proposed PS-MSLD method at reduced-scale. The specifications of datasets are given as:

QuickBird dataset: The QuickBird sensor offers PAN image with 0.7-m resolution and 4-band (R,G,B and NIR) MS image with 2.8-m resolution. These images include a forest area of the region named Surdarbans from eastern part of India. The size of the images used are 512×512 for PAN and $128 \times 128 \times 4$ for MS images.

IKONOS dataset: The IKONOS dataset is composed of mountainous area with trees of sichuan from China. The spatial resolution of PAN and MS images are 1-m and 4-m, respectively. The size of the images are 256×256 pixels for PAN image and $64 \times 64 \times 4$ pixels for MS image.

Pléiades dataset: The size of dataset is 1024×1024 and the resolution ratio is considered as 4. Since no PAN image is available for this dataset, the same is synthesized from the available MS bands. This dataset is available from the software package provided by (Vivone et al., 2015), consists of urban area from Toulouse (France).

For all the datasets, the source images are pre-registered.

The following Quality indices are considered for evaluation at reduced-scale.

- CC: Correlation Coefficient whose optimal value is one.
- SSIM: Structural Similarity between two images. Larger the value of SSIM higher the similarity between fused and source image.
- PSNR: Peak Signal to Noise Ratio, higher value indicates less noise ingradient.
- ERGAS: Relative dimension less global error in synthesis, Low value of ERGAS indicates similarity between multispectral bands.
- SAM: Spectral Angle Mapper, generally expressed in degrees and is equal to zero if two images are spectrally identical.
- Q4: Measures the similarity between two images, and is suitable for images with four spectral bands. Q4 values ranges from 0 to 1, and is equal to 1 if both images are equal. (For calculating Q4, block size is chosen as 32×32 pixels).

The values of CC, SSIM and PSNR are represented as an average of four bands. For all the experiments at reduced-scale, a Gaussian filter matched with the MTF of the sensor has been used for all the bands to avoid aliasing. Usually, the sensor gain at Nyquist frequency is provided by the manufacturer. The original PAN and MS images are decimated by a factor 4. To emphasize the finest spatial details the pansharpened outcomes of various methods are presented with a magnified square region on using the datasets at reduced-scale.

4.3.1.1 Experimental Results Using QuickBird Dataset

The outcomes are obtained by performing the fusion between degraded PAN image with 2.8-m resolution and the degraded MS image bands having 11.2-m resolution. Subsequently, the obtained results are compared with reference MS image having 2.8-m resolution. Fig.4.3(a) and 4.3(b) shows the degraded PAN image with 2.8-m resolution and the reference MS image (false color image: green-NIR-blue) with 2.8-m resolution. The up-sampled MS image (EXP) maintaining 11.2-m resolution is shown in Fig.4.3(c). The fused images on using different methods are depicted in Fig.4.3(d)-(l). In addition

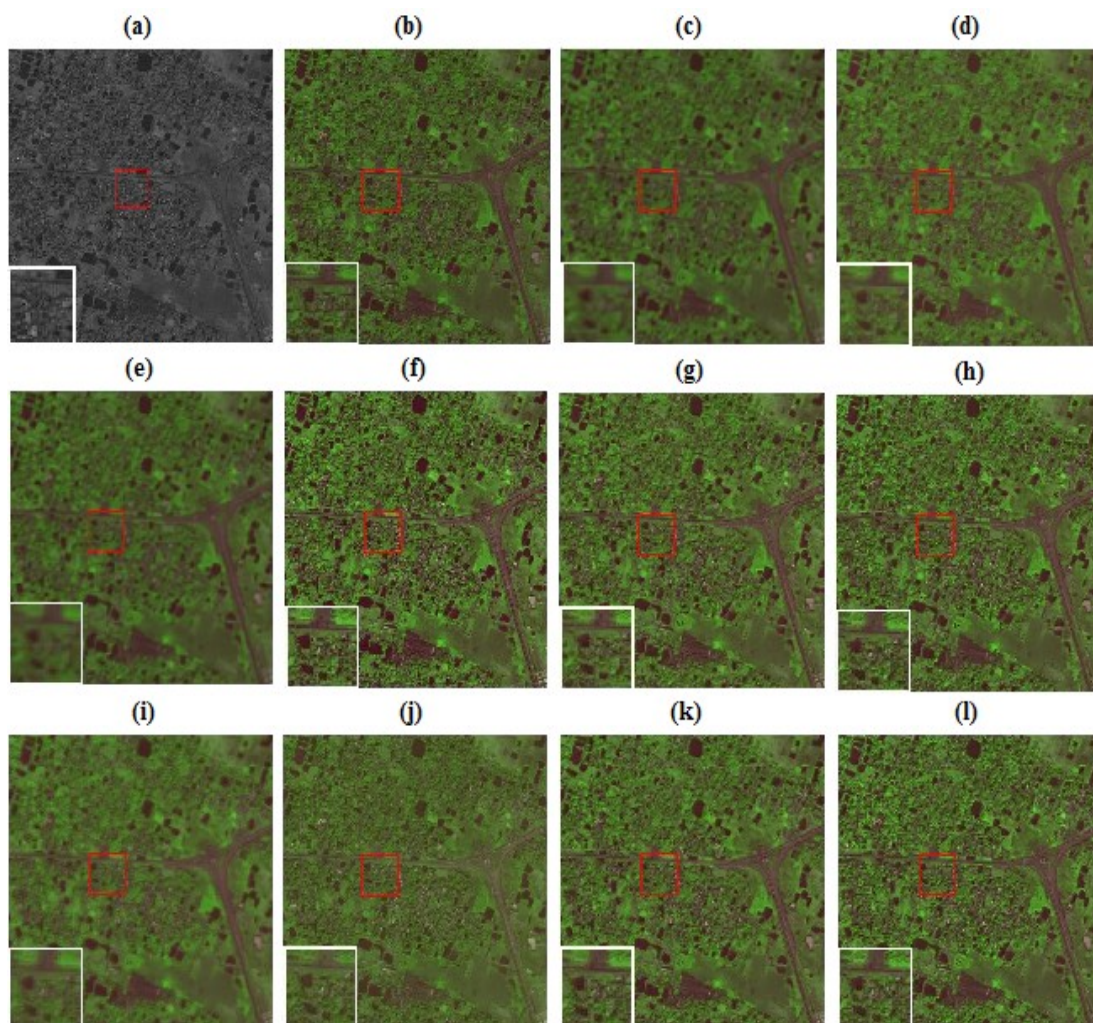


Figure 4.3: Pan-sharpening results with QuickBird data at reduced- scale (a) Degraded PAN image (512x512, 2.8-m) (b) original MS image (2.8-m) (c) EXP (up-sampled MS image) (d) GIHS (e) SW (f) GSA (g) MGC (h) AWLP (i) SR-TD (j) SR-WT (k) SR-D (l) PS-MSLD

to fused outcomes, for each image a square sub-region (marked in red color) comprising a part of highway and trees is zoomed and shown at the left bottom corner of the images.

Observation of visual results as presented in Fig.4.3 reveals that, the proposed method produces relatively less spectral distortion compared to the fused outcomes obtained using other considered methods. The result of GSA (Fig.4.3(f)) method appears to be sharper and with clear details among the classical methods. On analyzing the result of the proposed method as shown in Fig.4.3(l); it is evident that the high-frequency spatial details are effectively extracted. In terms of visual perception of enlarged sub-region, the proposed method is comparable with other methods such as SR-D, AWLP and GSA. The quantitative measures for the QuickBird dataset at reduced-scale are reported in Table 4.1.

Table 4.1: Quality indices for QuickBird (India-Surderban) dataset at reduced-scale

		CC	SSIM	PSNR	ERGAS	SAM	Q4
	EXP	0.7308	0.4012	20.7231	3.2031	3.7119	0.6834
Classical Methods	IHS	0.6313	0.2895	19.6576	3.0620	3.9137	0.4066
	SW	0.7212	0.3813	20.7145	3.1356	3.7487	0.6917
	GSA	0.8186	0.7614	21.7029	3.0177	3.5914	0.7061
	MGC	0.8176	0.4198	21.9732	2.8317	3.4713	0.7026
	AWLP	0.8247	0.7878	<u>22.1430</u>	2.9347	3.6392	0.7216
	SR-based Methods	SR-TD	0.8329	0.7863	21.3452	2.8435	3.9045
	SR-WT	0.8341	0.8091	21.8764	2.9871	3.7161	0.7545
	SR-D	<u>0.8476</u>	<u>0.8125</u>	22.3653	<u>2.8325</u>	<u>3.3256</u>	<u>0.7892</u>
Proposed	PS-MSLD	0.8508	0.8321	22.0231	2.9429	3.3211	0.7912

The proposed method achieves optimal performance for CC, SSIM, SAM and Q4 measures. Although, the other two measures namely PSNR and ERGAS are slightly compromised. The best value for SAM indicates the fused outcome of proposed method is almost similar to the reference image compared to the results obtained using the other reported methods. It is evident from the visual observation and quantitative measures that the SR-D method and the proposed PS-MSLD method better preserves the spectral fidelity than the other methods. The optimal value for each quality metric is indicated in bold and the second best value is highlighted with an underline.

4.3.1.2 Experimental Results With IKONOS Dataset

Fig.4.4(a) and 4.4(b) shows the degraded PAN image with 4-m resolution and original MS image (false color image: NIR-blue-green) with 4-m resolution. The up-sampled MS image maintained at 4-m resolution, is shown in Fig.4.4(c). The fusion outcomes of different methods are depicted in Fig.4.4(d)-(l). Further, a smaller sub-region of the forest area (enclosed by a red color square box) is enhanced and displayed at the left

bottom corner of each image.

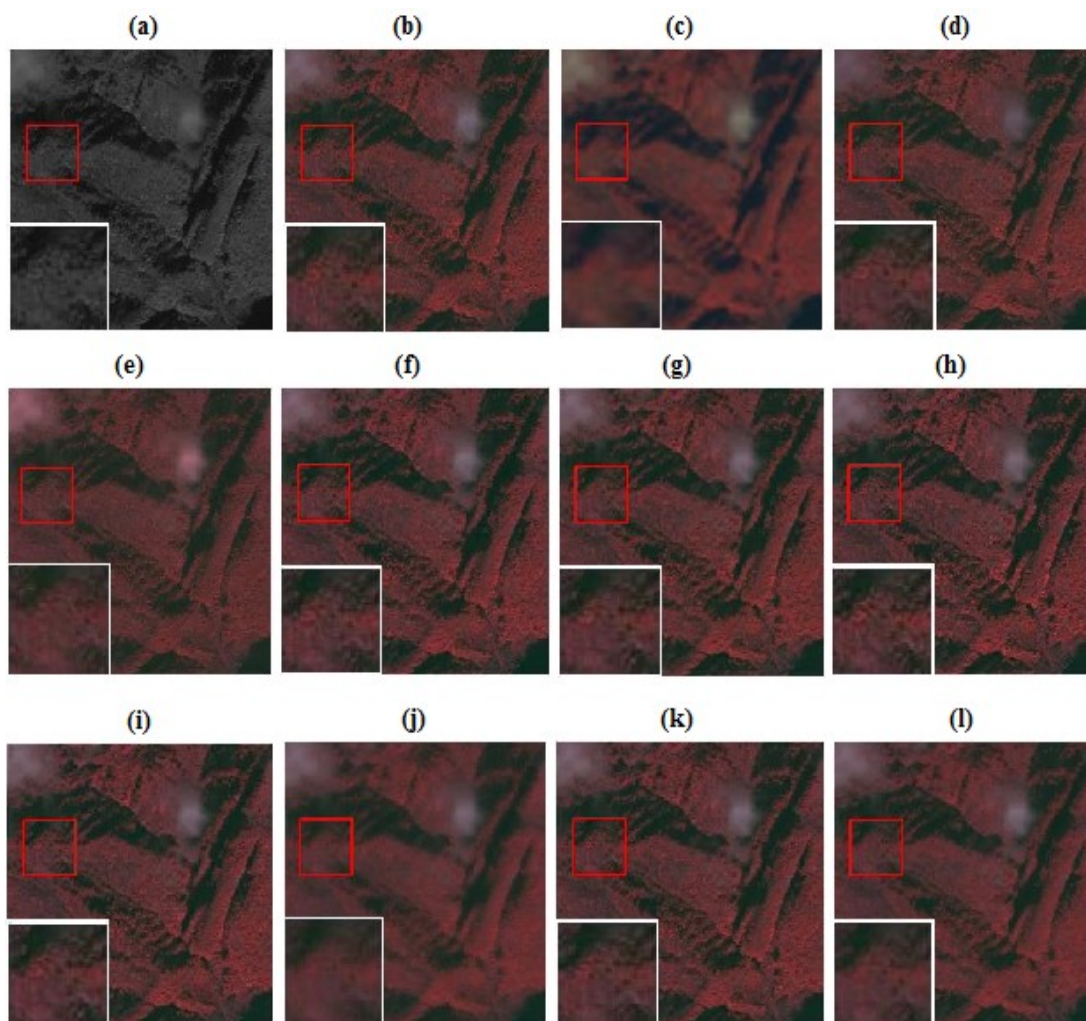


Figure 4.4: Pan-sharpening results with IKONOS data at reduced-scale (a) PAN image (256x256, 4-m) (b) Reference MS image (256x256,4-m) (c) EXP (upsampled MS image) (d) GIHS (e) SW (f) GSA (g) MGC (h) AWLP (i) SR-TD (j) SR-WT (k) SR-D (l) PS-MSLD

On comparing the fused outcomes obtained by GIHS as shown in Fig.4.4(d) and that of SW method shown in Fig.4.4(e) with the reference image shown in Fig.4.4(b), modest spectral distortions are clearly visible. The similar effect with comparatively less color distortion also appears in the fused outcome of SR-WT method (Fig.4.4(j)). The proposed method leads to qualitatively improved outcomes in terms of maintaining finer spatial details and rich spectral contents and the same is reflected in the extracted enlarged image segment shown in the lower left square box.

The quantitative performance measures corresponding to the visual results as shown in Fig.4.4 are reported in Table 4.2. The proposed method accomplishes the best values

for CC, SSIM, ERGAS and Q4. The visual fused outcomes and quality indices are consistent with each other. The optimal value for each quality metric is indicated in bold and the second best value is highlighted with an underline. On comparing the performance measures presented in Table 4.2; it is evident that the performance of AWLP matches with the outcomes obtained using the SR-based methods. The proposed PS-MSLD method outperforms almost in all the quality metrics compared to other reported methods.

The quantitative assessment results demonstrate the effectiveness of the proposed PS-MSLD algorithm in terms of its ability to extract the desired spatial details.

Table 4.2: Quality indices for IKONOS (China-Sichuan) dataset at reduced-scale

		CC	SSIM	PSNR	ERGAS	SAM	Q4
	EXP	0.9233	0.7129	22.2870	3.9785	4.6243	0.7436
Classical Methods	GIHS	0.8744	0.6405	21.1568	3.3592	4.9417	0.7210
	SW	0.9467	0.6969	24.3324	2.7690	4.1590	0.8166
	GSA	0.9538	0.7556	24.3424	2.5244	3.3271	0.8529
	MGC	0.9483	0.7483	24.0361	2.5134	3.4034	0.8591
	AWLP	0.9597	0.7772	25.1861	<u>2.3323</u>	<u>2.9329</u>	0.8697
SR-Based Methods	SR-TD	0.9584	0.7801	24.8137	2.4142	3.0145	0.8711
	SR-WT	0.9607	0.8194	24.4113	2.4652	3.1126	0.8843
	SR-D	<u>0.9623</u>	<u>0.8219</u>	24.8267	2.3502	2.8158	<u>0.8879</u>
Proposed	PS-MSLD	0.9711	0.8278	<u>25.1247</u>	2.3131	2.9399	0.8931

4.3.1.3 Experimental Results Using Pléiades Dataset

The PAN image for this dataset is simulated from the available MS bands of size 1024x1024 (Vivone et al., 2015). The visual outcomes of Pléiades dataset at reduced-scale are presented in Fig.4.5. analysis of it clearly reveal that the color distortion is present in the outcomes of GIHS and SW schemes as shown in Fig.4.5(d) and (e), respectively. In the family of classical methods, GSA and AWLP outcomes are slightly sharper as compared to the reference image (Fig.4.5(b)). The outcomes of SR-based methods manifest the sharpness in appearance and the same can be observed in quantitative results. An area surrounded by buildings is magnified and shown at the top left corner in each image.

The Quantitative performance measures at reduced-scale corresponding to visual observations shown in Fig.4.5 are reported in Table 4.3. Analysis of quality metrics reported in Table 4.3 reveal that, for the urban dataset the performance of classical methods namely, GSA, AWLP and MGC are moderately good for the performance metrics

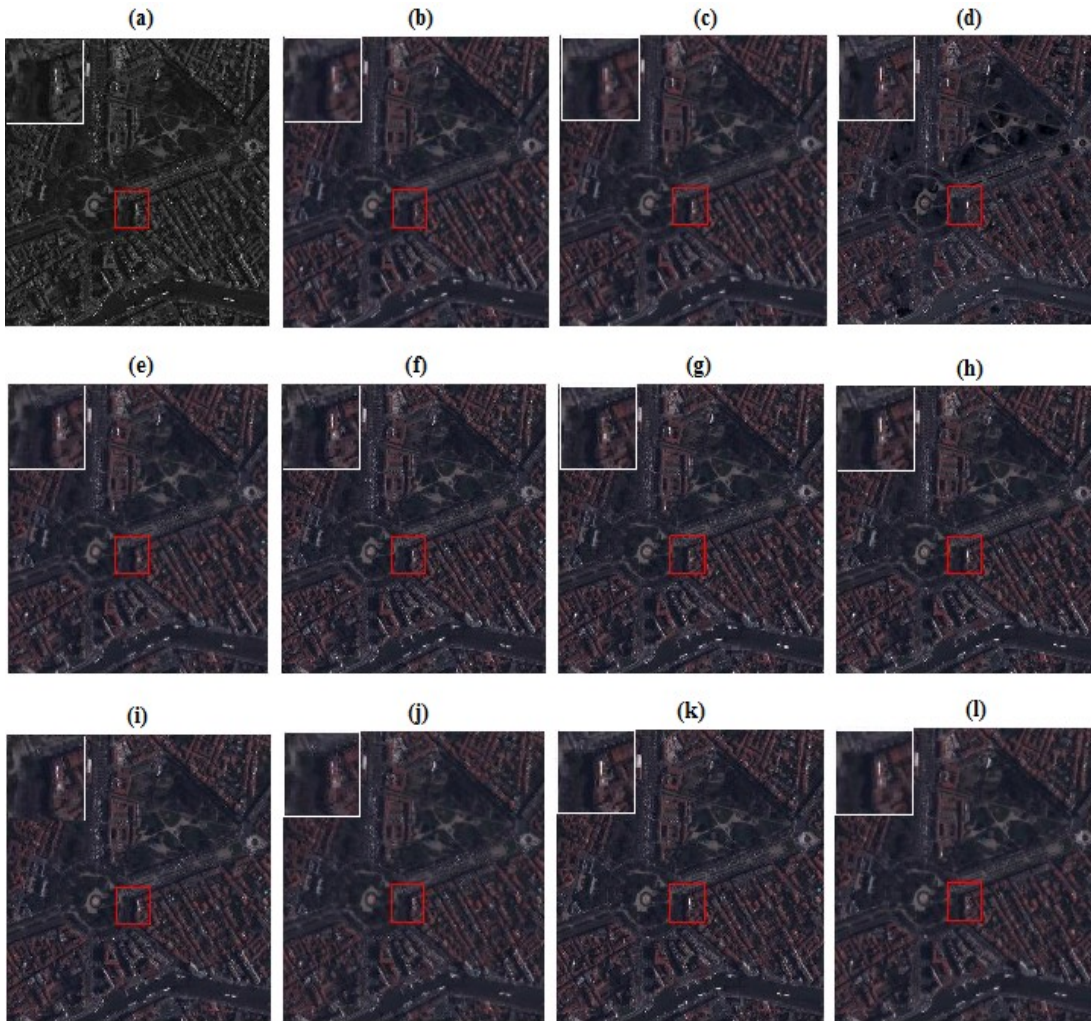


Figure 4.5: Pan-sharpening results with pléiades data at reduced-scale (a) PAN image (b) MS image (c) EXP (d) GIHS (e) SW (f) GSA (g) MGC (h) AWLP (i) SR-TD (j) SR-WT (k) SR-D (l) PS-MSLD

CC, SSIM and Q4. The proposed PS-MSLD method achieves the optimal values for CC, PSNR and Q4. The quantitative results evidence the competitive performance of SR-D method with the proposed method for the urban dataset.

Three different datasets with different geographical compositions like urban, vegetation and mountains with trees are used for the analysis at reduced-scale. The detailed analysis of visual and quantitative results lead to the following conclusions: The color distortion is exhibited by the classical methods like GIHS and SW for all the considered datasets. It is clearly justified by the higher values of SAM index and also from the visual outcomes. For all the considered datasets AWLP and GSA schemes manifest superior performance among the classical methods. For the images comprises of vegetated area the MTF-GLP method achieves relatively better performance within the

Table 4.3: Quality indices for Pléiades (Toulouse, France) dataset at reduced-scale

		CC	SSIM	PSNR	ERGAS	SAM	Q4
	EXP	0.8821	0.9048	14.9729	6.1045	4.6951	0.7771
Classical Methods	GIHS	0.9264	0.8542	16.5606	5.0658	5.2064	0.8395
	SW	0.9073	0.8490	16.9707	5.4383	4.6317	0.8343
	GSA	0.9627	0.9665	22.1795	3.1023	4.4682	<u>0.9582</u>
	MGC	0.9654	0.8315	21.9842	3.3392	4.2935	0.9507
	AWLP	0.9661	<u>0.9635</u>	20.2386	3.5483	4.3505	0.9417
SR-based Methods	SR-TD	0.9672	0.9611	20.8765	3.5764	4.4765	0.9487
	SR-WT	0.9711	0.9634	22.1881	3.4623	4.3689	0.9511
	SR-D	<u>0.9724</u>	0.9614	22.1138	3.5232	4.4123	0.9572
Proposed	PS-MSLD	0.9817	0.9573	22.2127	<u>3.2549</u>	<u>4.3214</u>	0.9606

classical algorithms family. On analyzing the performance measures obtained using the class of the SR-based methods; the outcomes of SR-D method are comparable with the proposed PS-MSLD method. SR-WT method outcomes appear smoother particularly for vegetated and mountainous datasets. The detailed visual and quantitative results at reduced-scale illustrate the effectiveness of the proposed PS-MSLD in reducing the distortions attribute and improving the overall quality of the fused outcomes.

4.3.2 Experimental Results at Full-Scale

In this section the proposed method is analyzed on WorldView-2 and another set of QuickBird images at full-scale. The metrics at full-scale based on QNR protocol give a more decisive measure of the quality of pansharpened images. As per QNR protocol, the fusion is performed on original PAN image and the MS image up-sampled to the scale of PAN image. The MATLAB function 'imresize' is used for up-sampling of the MS image (bi-cubic interpolation method). The PAN image of size 512×512 pixels and MS image of size $128 \times 128 \times 4$ pixels are used for full-scale evaluation.

The WorldView-2 sensor provides the PAN and MS images at the spatial resolution of 0.5-m and 2-m, respectively. The MS image of the WorldView-2 dataset contains 8 bands out of which, only 4 bands (Blue, Green, Red and NIR1 (2,3,5 and 7)) are considered for experimentation. This dataset is composed of different buildings, roads and trees. Fig.4.6 shows the visual fused outcomes of various pansharpening algorithms on WorldView-2 images. Fig.4.6(a) and (b) shows PAN image at 0.5-m and up-sampled MS image (EXP) at 2-m resolution.

The visual perception of GIHS method as shown in Fig.4.6(c) exhibits a perceptible

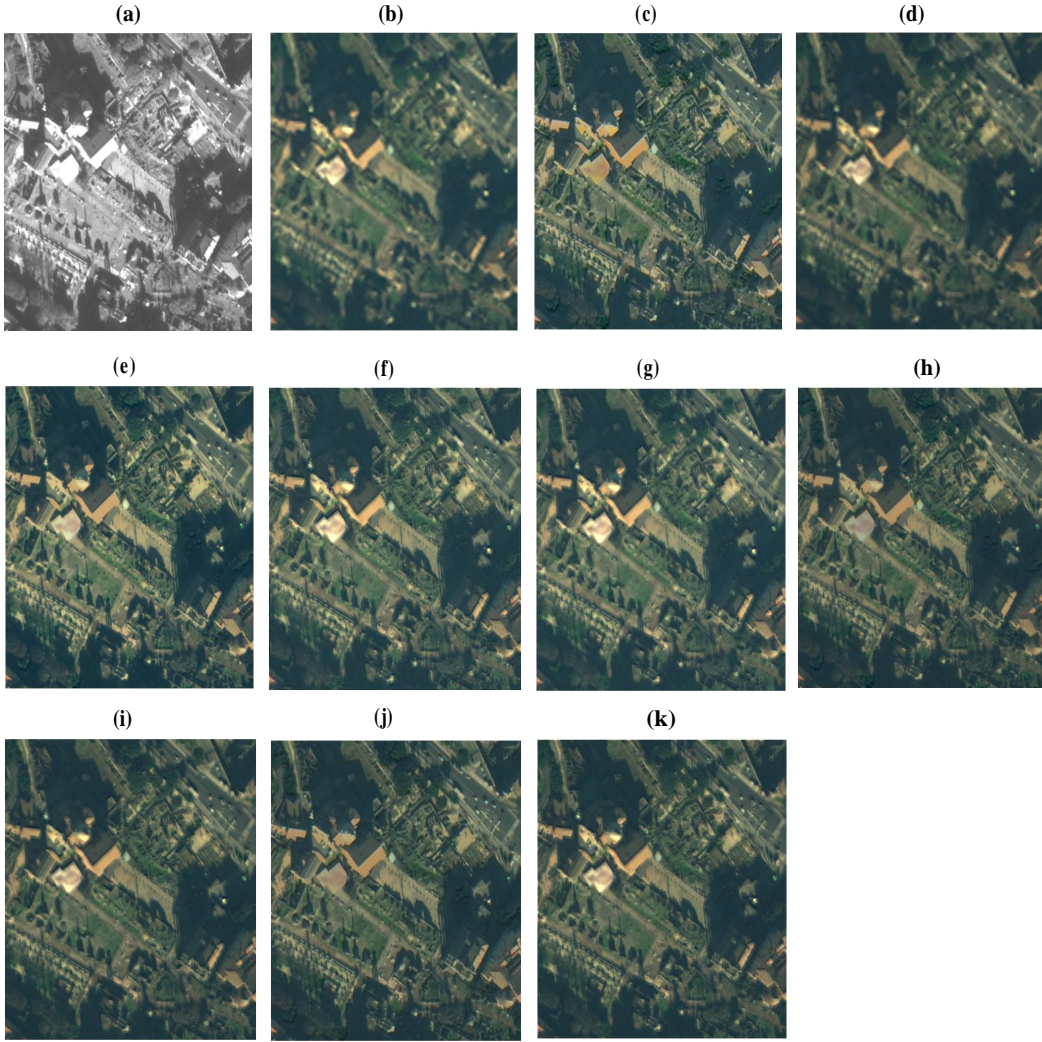


Figure 4.6: Pan-sharpening results with WorldView-2 data at full-scale (true color images (5-3-2)) a) PAN image b) Up-sampled MS image (EXP) (c) GIHS (d) SW (e) GSA (f) MGC (g) AWLP (h) SR-TD (i) SR-WT (j) SR-D (k) PS-MSLD

Table 4.4: Quality indices for WorldView-2 dataset at full-scale

	EXP	GIHS	SW	GSA	MGC	AWLP	SR-TD	SR-WT	SR-D	PS ¹
D_λ	0	0.0545	0.0305	0.0280	0.0252	0.0147	0.0164	0.0192	0.0059	<u>0.0064</u>
D_s	0.0952	0.0538	0.0550	0.0540	0.0484	0.0486	0.0463	0.0411	<u>0.0351</u>	0.0342
QNR	0.9048	0.8946	0.9162	0.9195	0.9276	0.9375	0.9381	0.9404	<u>0.9592</u>	0.9595

1.PS-MSLD

color change in the green area. The result of SW method (Fig.4.6(d)) appears to be slightly blurred. This effect is justified by the spatial distortion index estimate for SW method as reported in Table 4.4. The quality indices for the full-scale evaluation are reported in Table 4.4. The proposed PS-MSLD method yields best values for spatial

distortion index (D_s) and for QNR. From quantitative results obtained at full-scale, in relative sense, the performance of AWLP method is optimum among the other classical methods. The visual outcomes and value of quality indices exhibit the superiority of the proposed PS-MSLD method.

Another set of images from QuickBird sensor, consists of trees and roads is used for the full-scale assessment. Fig.4.7 presents the QuickBird images at full-scale and the corresponding pan-sharpened images. Fig.4.7(a) and (b) shows the PAN image at 0.7-m resolution and MS image at 2.8-m resolution.

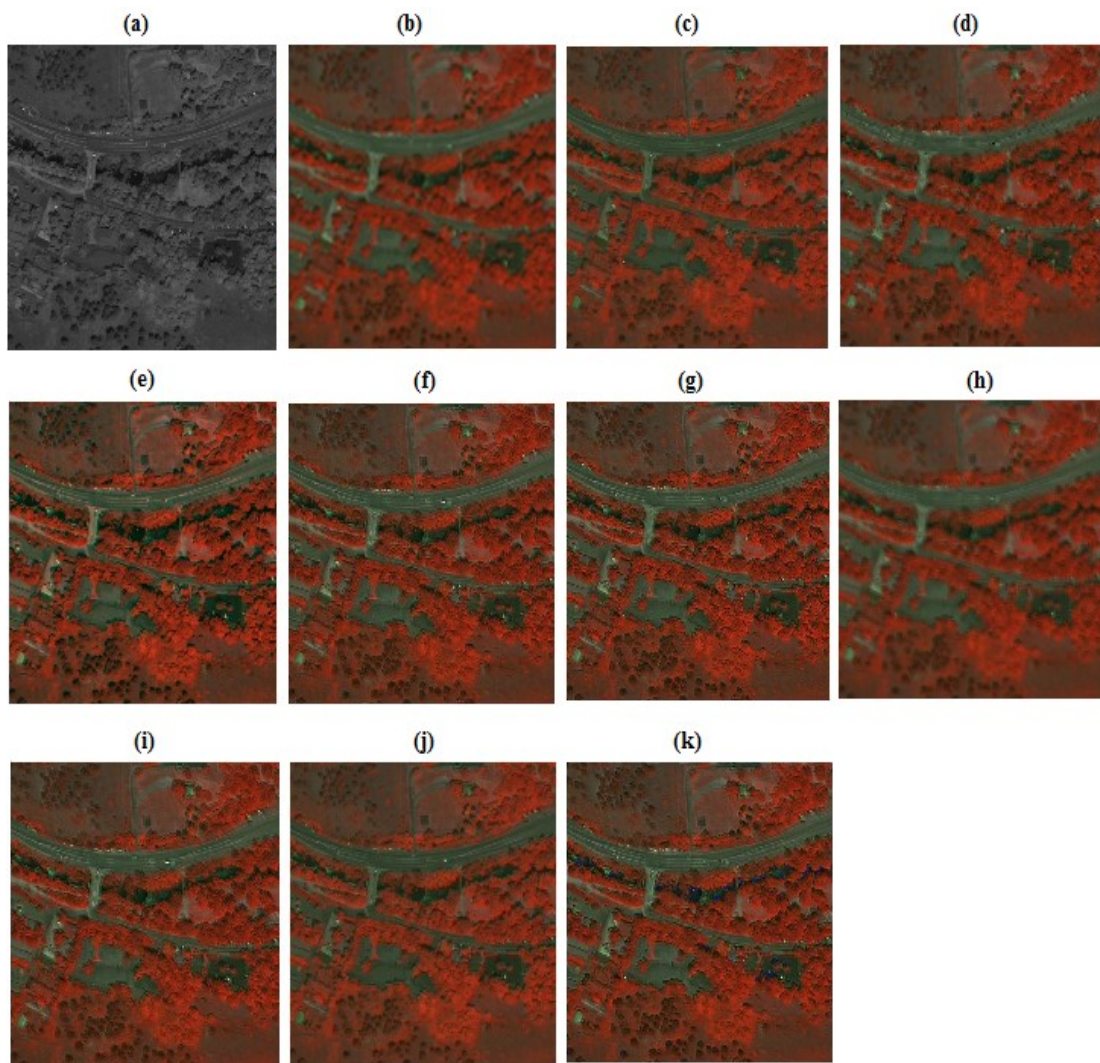


Figure 4.7: Pan-sharpening results with QuickBird data at full-scale (false color images(4-2-3)) a) PAN image (512x512, 0.7-m) b) up-sampled MS image (EXP) (512x512, 2.8-m) (c) GIHS (d) SW (e) GSA (f) MGC (g) AWLP (h) SR-TD (i) SR-WT (j) SR-D (k) PS-MSLD

For the better appearance of details;the images are displayed by false color compo-

sition with NIR-green-blue (4-3-2) bands. The experimental outcomes of all SR-based methods surmounts in spatial and spectral preservation compared to the classical methods. The quality metrics comprising spectral, spatial distortion index and QNR values are reported in Table 4.5.

Table 4.5: Quality indices for QuickBird dataset at full-scale

	EXP	GIHS	SW	GSA	MGC	AWLP	SR-TD	SR-WT	SR-D	PS ¹
D_λ	0	0.0754	0.0735	0.908	0.0704	0.0737	0.0664	0.0592	0.0346	<u>0.0352</u>
D_s	0.2965	0.3201	0.3124	0.2455	0.2178	0.2286	0.1989	0.1811	<u>0.1435</u>	0.1379
QNR	0.7035	0.6286	0.6370	0.7051	0.7271	0.7145	0.7479	0.7704	<u>0.8269</u>	0.8318

1.PS-MSLD

The proposed PS-MSLD method yields optimal values for QNR index for both the datasets at full-scale evaluation. It demonstrates that the proposed PS-MSLD method maintains good spatial characteristics and preserves the desired spectral information comprehensively.

4.4 Conclusion

The notion of multi-scale learned dictionary (PS-MSLD) based sparse regularization is investigated for pan-sharpening in this work. Wavelets as basis is used to impart the learning of dictionaries. Multi-scale learned dictionary is attractive because of its ease in performing the sparse coding. The proposed PS-MSLD successfully combines sparse representation theory and the concept of similarity between patches at different scales with corresponding dictionaries. The experimental results demonstrate the capability of the PS-MSLD towards enhancing the spatial resolution thus, maintaining the quality spatial details and accompanied significant reduction in spectral distortion. Although, separate dictionaries are learned for each band, the proposed method requires moderate computational efforts compared to the traditional dictionary learning based methods. The APrU-DL algorithm improves the quality of the fused outcomes with significant reduction in learning complexity and better convergence rate. A further study may be carried out to develop a structured multi-scale dictionary to enhance properties like spatial invariance, adaptivity and over-completeness. It may lead to much improved quality of fusion process.

CHAPTER 5

PAN-SHARPENING ALGORITHMS BASED ON CONVOLUTIONAL SPARSE REPRESENTATION

Pan-sharpening (PS) is a prominent remote sensing image fusion technique. It yields high-resolution multispectral (HRMS) images, which are imperative for the applications such as recognition, detection etc. PS methods based on conventional sparse representation induce blurring effects and are unable to preserve the essential spatial details in the fused image. A review of literature infers that the pan-sharpening methods can be categorized as component substitution (CS), multi resolution analysis (MRA) based methods and sparse representation (SR) based methods. CS methods are capable of retaining the spatial details, however, these methods lead to spectral distortion. MRA methods are capable to preserve the inherent spectral information though, these methods are not as promising as the CS methods in terms of extracting pertinent spatial details. With an intention of improving the quality of the fused outcome, the principles of CS and MRA are combined and a generalized version of band dependent spatial details (BDSD) scheme is developed by (Zhong et al., 2017). This method achieves considerable reduction in spectral distortion compared with the other conventional PS methods. However, it suffers from few drawbacks such as spatial artifacts present in the fused image.

Sparse Representation (SR) of image patches has become prevalent (Candes et al., 2006; Elad et al., 2010), and applied to remote sensing image fusion due to its effectiveness in enhancing the spectral and spatial details in the resultant fused image. The two major categories of SR based pan-sharpening methods are; the dictionary learned from the patches of PAN and MS images and the dictionary learned from PAN image and its low-resolution version. In contrast to the existing PS methods, pansharpening based on a single compact dictionary generated using HRMS images is proposed in (Ayas et al., 2018), to scale-down the time complexity of the algorithm. Apart from the various modifications being proposed, the PS algorithms are still striving to achieve a balanced trade-off, between the spatial and spectral contents in the fused image.

However, the dictionary construction and subsequent update process, in most of these SR based methods are computationally expensive and so obtained pan-sharpened

images incur spectral distortion. The dictionary learned from PAN image may not be adequate to sparsely represent the MS image patches. Moreover, the phenomenal difference in spectral response between the MS and PAN images results in an unwarranted degradation of fusion performance. Most of the SR based pan-sharpening methods adapted patch-partition based processing strategy. In that, the image is decomposed into overlapped patches and usually, the sparse coding is performed on these individual patches. The overlapped pixels are averaged and rearranged to reconstruct the fused image. An increase in amount of overlapping between adjacent patches increases the smoothing of some details in the resultant image thereby it results in redundant representation. In multi-sensor image fusion, shift-invariance or translation-invariance is an issue of significant concern and must be dealt with precise care. In scenario, wherein the representation scheme is not shift-invariant, the fused image incurs ringing artifacts for misregistered regions. To overcome the aforementioned featured limitations, the mechanisms of representation and averaging are applied for the entire image, instead of exercising it for the individual patches. Using this approach, it is possible to design a fusion scheme, which is robust to misregistration.

The fusion methods proposed in this chapter attempt to mitigating the shortcomings of patch based pan-sharpening methods using an alternative representation format known as convolutional sparse coding (CSC)/convolutional sparse representation (CSR) (Wohlberg, 2016a). In CSC, the sparse coding is applied to the entire image using a set of convolutional filters.

This chapter is organized into two parts that discuss two different pan-sharpening methods, which operate on the CSC mechanism to accomplish the HRMS images with rich spatial and spectral features. The first part presents a PS algorithm in which the source images are decomposed as cartoon and texture components. Distinct fusion models are implemented on the decomposed componets to obtain the pan-sharpened image. In contrast to the first method, the second part discusses a pan-sharpening method in which the fusion mechanism is executed in wavelet domain. The preliminary concepts such as convolutional sparse coding (CSC) and cartoon plus texture decomposition (CPT) are presented in section 5.1. A robust pan-sharpening algorithm based on convolutional sparse representation for spatial enhancement is presented in part I of this chapter. Part II of this chapter comprises a pan-sharpening method based on discrete wavelet decomposition and convolutional sparse representation. The outcomes of these two schemes are presented and analysed using visual results and quantitative performance measures.

5.1 Preliminaries

The basic principles and mathematical framework used in the proposed pan-sharpening algorithms, convolutional sparse coding (CSC) and cartoon plus texture (CPT) decomposition are presented in this section. In addition, the basic difference in implementing the sparse representation (SR) and convolutional sparse representation (CSR) mechanisms is also presented in this section.

5.1.1 Convolutional Sparse Coding Mechanism

The Convolutional Sparse Coding (CSC) is a relatively recent development in many image processing applications, as an alternative to the generalized sparse representation. The basic idea of CSC is to obtain a convolutional decomposition of an image under sparsity constraint. The CSC model makes use of a set of dictionary filters $\{d_m, m = 1 \dots M\}$, in place of conventional dictionary \mathbb{D} . A set that comprises sum of convolutions of d_m with the corresponding representations $\{x_m, m = 1 \dots M\}$ reconstructs the entire image 'I'. Here, x_m is a set of unknown coefficients maps. By regularizing x_m with sparsity prior, the CSC model can be formulated as:

$$\operatorname{argmin}_{\{x_m\}} \frac{1}{2} \|d_m * x_m - \mathbf{I}\|_2^2 + \lambda \sum_m \|x_m\|_1 \quad (5.1)$$

where λ is a regularizing parameter that maintains balance between the sparsity and reconstruction error. The dictionary filters are usually learned from a number of training images using the following model:

$$\operatorname{argmin}_{\{d_m\}\{x_{n,m}\}} \frac{1}{2} \sum_k \left\| \sum_m d_m * x_{n,m} - \mathbf{I}_i \right\|_2^2 + \lambda \sum_n \sum_m \|x_{n,m}\|_1 \text{ s.t. } \|d_m\|_2 = 1. \quad (5.2)$$

Here, \mathbf{I}_i is a set of training images and the constraint on d_m is used to prevent the scaling ambiguity and the index n represents the arbitrary number of training images.

Eq.5.1 is the convolution form of the basis pursuit denoising (BPDN) problem, known as convolutional BPDN (CBPDN). In this work, alternating direction method of multipliers (ADMM) is used to solve the CBPDN problem. The sub-problem involving convolution in ADMM algorithm is calculated in Fourier domain (Bristow et al., 2013). An extensive study has been conducted on the performance of ADMM for different values of the balancing parameter λ in (Wohlberg, 2015). Usually, the ADMM is

faster with much smaller values of λ (less than 0.1) for image reconstruction problems. Smaller value of λ emerges as a good choice for image processing applications that demand low reconstruction error. As per (Wohlberg, 2015), in the proposed method, λ value is chosen as 0.05. The comprehensive learning mechanism for the dictionary filters and the strategy to implement CBPDN is same as given in (Wohlberg, 2016b).

The convolutional model is acquired from the shift-invariance paradigm, which is an important requisite property for image fusion. In contrast to conventional SR, the CSC model computes the sparse representation for the entire image, hence the obtained representation is single-valued and optimized over the entire image (Liu et al., 2016). To overcome the drawbacks of conventional patch based SR algorithms, CSC model is invoked for pan-sharpening of multispectral images in this work.

5.1.2 Sparse Representation Versus Convolutional Sparse Representation

The sparse representation (SR) mechanism estimates the sparse coefficient vector, s for an image patch, y using an over-complete dictionary, \mathbb{D} , as

$$s = \operatorname{argmin} \|s\|_1 \quad \text{subject to} \quad \|y - \mathbb{D}s\|_2^2 \leq \varepsilon \quad (5.3)$$

here, ‘ ε ’ is the tolerable reconstruction error. The schematic of SR mechanism is shown in Fig.5.1. An image, \mathbf{I} composed of n number of patches, can be expressed as a linear

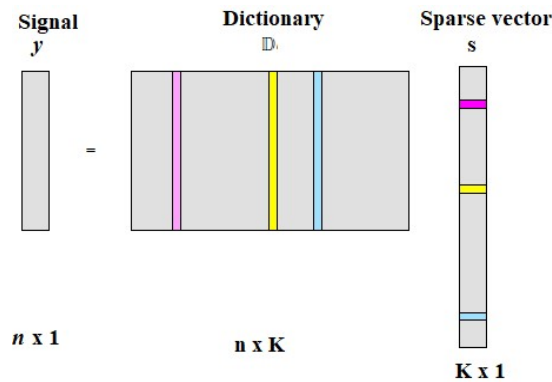


Figure 5.1: Sparse representation schematic.

combination of dictionary atoms d_i , and the sparse coefficient vectors, s_i corresponding to the image patches. i.e., $\mathbf{I} = \sum_i d_i s_i, i = 1, 2, \dots, n$. With n number of patches extracted from the image \mathbf{I} , and y, s are the patches and coefficients vectors, respectively.

In CSR mechanism, the image \mathbf{I} can be expressed as sum of convolutions of dictionary elements, d_m and each feature maps, x_m . i.e., $\mathbf{I} = \sum_m d_m * x_m$. The size and number of the dictionary filters can be chosen arbitrarily. The size of the coefficients maps, x_m are same as that of the image \mathbf{I} , and the index, m is same as that of the dictionary filters, d_m . The schematic of CSR mechanism is presented in Fig.5.2.

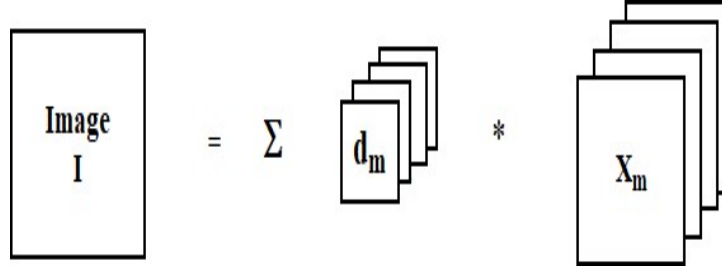


Figure 5.2: Convolutional sparse representation schematic.

5.1.3 Cartoon and Texture Decomposition

Each image can be characterized as a combination/mixture of its texture (T) and cartoon (C) components. The geometric and smoothly varying component, composed of image hues and boundaries is termed as cartoon component. The texture element is an oscillatory or small scale spatial patterns of the image. Therefore, the spectral information and spatial details of an image can be conveniently mapped to its cartoon and texture components, respectively.

Meyer (Meyer, 2001) presented the conventional mechanism to decompose an image \mathbf{I} into cartoon plus texture components and is given by :

$$\inf_{C, T \in X_1 \times X_2} F_1(C) + \gamma F_2(T) : \mathbf{I} = C + T \quad (5.4)$$

where, F_1 and F_2 are functionals and X_1 and X_2 are spaces of functionals such that $X_1 = \{C : F_1(C) \leq \infty\}$ and $X_2 = \{T : F_2(T) \leq \infty\}$.

The decomposition of an image as $\mathbf{I} = u + v$, where u (cartoon component) and v (texture component) are the solutions of an optimization problem of the form :

$$\min_u T(\mathbf{I} - u) + \gamma \mathbb{J}(u) \quad (5.5)$$

where, $\mathbb{J}(u)$ is cartoon image prior, $T(v)$ is texture image prior and γ is a regularization parameter, that controls the trade-off between cartoon and texture components.

In this work, \mathbb{J} is considered as Total Variation (TV) prior. Hilbert framework (Aujol and Gilboa, 2006) is used to define the texture prior T. The attributes suggest that Total Variation (TV) is the relevant space to represent cartoon component. The TV of a smooth image f is defined as

$$\mathbb{J}(f) = \int \|\nabla f(x)\| dx \quad (5.6)$$

The discrete version of a gradient operator is defined as :

$$(\nabla \mathbf{u})_{i,j} = ((\nabla \mathbf{u})_{i,j}^1, (\nabla \mathbf{u})_{i,j}^2) \quad (5.7)$$

where,

$$(\nabla \mathbf{u})_{i,j}^1 = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < p. \\ 0 & \text{if } i = p. \end{cases} \quad (5.8)$$

and,

$$(\nabla \mathbf{u})_{i,j}^2 = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < q. \\ 0 & \text{if } j = q. \end{cases} \quad (5.9)$$

where the size of the image is $p \times q$.

The discrete total variation of \mathbf{u} is defined as:

$$\mathbb{J}(\mathbf{u}) = \sum_{1 \leq i \leq p, 1 \leq j \leq q} |(\nabla \mathbf{u})_{i,j}| \quad (5.10)$$

The gradient of the TV norm is

$$\text{Grad } \mathbb{J}(\mathbf{u}) = \text{div} \left(\frac{\nabla \mathbf{u}}{\|\nabla \mathbf{u}\|} \right) \quad (5.11)$$

The discrete version of the divergence operator can be defined as $\text{div} = -\nabla^*$, where ∇^* is adjoint of ∇ .

Note: The gradient of the TV norm is not defined at pixel 'x' with $\nabla f(x) = 0$. The gradient flow of TV norm is not well defined. Hence, the gradient flow is defined for a

smooth TV norm as:

$$J_{\varepsilon}(f) = \int \sqrt{(\varepsilon)^2 + (\|\nabla f(x)\|)^2} \quad (5.12)$$

where, ε is a regularization parameter.

The texture norm is defined using the Fourier transform F as:

$$T(v) = \frac{1}{2} \|\text{diag}(W) F(u)\|^2 \quad (5.13)$$

W is the weight associated with frequency. The selection of W is large for low frequencies, and small for high frequencies to ensure the pertinent features of texture component.

The gradient of texture norm is expressed as:

$$\text{Grad } T(v) = \dot{H}v, \quad \dot{H} = F^* \text{diag}(W^2)F \quad (5.14)$$

F^* is the inverse Fourier transform. The Hilbert, H^{-1} model is intended to capture the high-frequency textures effectively. In this work, Total Variation - Hilbert $\{\text{TV-H}^{-1}\}$ method is implemented for image decomposition (Osher et al., 2003). The TV-H^{-1} decomposition solves,

$$\min_u \zeta(u) = \frac{1}{2} \|\text{WF}(\mathbf{I} - u)\|^2 + \gamma \mathbb{J}(u) \quad (5.15)$$

The mapping $u \mapsto \zeta(u)$ is a smooth function and can be minimized using a gradient descent (with the assumption of initial cartoon layer $u = \mathbf{I}$) as:

$$u = u - \tau(\text{Grad } T(u - \mathbf{I}) + \gamma \text{Grad } \mathbb{J}(u)) \quad (5.16)$$

where $\text{Grad } \mathbb{J}(u)$ is gradient of $\mathbb{J}(u)$, γ is the trade-off between cartoon and texture components. Usually, the H^{-1} texture model is intended to chose very high-frequency and thus, perform poorly for medium frequency textures. In order to retain the desired textures the trade-off parameter γ need to be chosen meticulously.

The gradient of texture norm is defined as: τ is the gradient descent step size and it should satisfy the condition

$$\tau < \frac{2}{\max_{\omega} W_{\omega}^2 + \gamma \cdot \varepsilon / 8} \quad (5.17)$$

where, W_{ω} is the weight associated with the operating frequency ω and ε is a regular-

ization parameter chosen as a small value ($1e^{-2}$) in this work.

The gradient descent operation is performed to find out the steady and converged cartoon component. The texture component is the difference between the given original image and the associated cartoon component.

$$v = \mathbf{I} - u. \quad (5.18)$$

PART I

A Robust Pan-Sharpening Algorithm Based on Convolutional Sparse Coding for Spatial Enhancement

An effective PS method is proposed based on cartoon and texture decomposition of each of the source images. The motivation for cartoon plus texture (CPT) decomposition of each source image can be illustrated as: The texture component of an image contains meaningful and repeated structures and oscillating patterns, whereas the cartoon component is piece-wise smooth, and consists of geometric structures corresponding to the sharpe edges and object hues of image (Buades et al., 2010). The spectral information predominantly resides in the cartoon component, and it's orthogonal complement, texture component manifests spatial details of remote sensing images. An effective CPT decomposition method can discriminate spatial and spectral features of remote sensing images comprehensively. The quality of fused outcome greatly affected by the amount of details injected. In this context, appropriate balance must be maintained as excessive injection causes spectral distortion, whereas insufficient quantum of injection leads to loss of finer details.

The pan-sharpening method based on CPT decomposition is proposed in (Lotfi and Ghassemian, 2018b), in that only the texture component of PAN image is exploited for the spatial details extraction. Limitations of this scheme are moderate spectral distortion and marginal spatial enhancement. These observations navigate to devise an appropriate pan-sharpening mechanism to enhance the quality of the fused outcome. It can be attained by choosing proper fusion models for the cartoon and texture components.

The proposed scheme combines the CSC and cartoon plus texture based image decomposition. The primary motive is to enhance the spatial quality of the fused image while preserving the necessary spectral information. The main contributions of the proposed work can be summarized as:

- The two major shortcomings of the existing SR based PS methods are (a) limited

ability in detail preservation and (b) high sensitivity to misregistration. Impact of both of these drawbacks can be reduced by adapting CSC for pansharpening.

- A modified CPT decomposition procedure is proposed, to separate the spatial and spectral components of source PAN and MS images precisely. The separation mechanism assists the balanced detail injection, which further provides the legitimate trade-off between spatial details and spectral information in the pan-sharpened image.
- Appropriate fusion rules for cartoon and texture components are designed based on the components intrinsic characteristics.
- To capture the maximum features in the fused image, an efficient gradient based fusion model is proposed to merge the spatial and spectral components holistically which in turn yields the pan-sharpened image.

5.2 A Hybrid Pan-Sharpener Algorithm

Let $X = \{X_k\}_{k=1,2,\dots,B}$ be the source MS image having B spectral bands, P be the PAN image and $\tilde{X} = \{\tilde{X}_k\}_{k=1,2,\dots,B}$ be the MS image up-sampled to the size of PAN image. The pan-sharpened image $\{\hat{X}_k\}_{k=1,2,\dots,B}$ is synthesized by the superposition of its texture and cartoon components, estimated from the source images (PAN and MS images, P and \tilde{X}). Ideally, the reconstructed pan-sharpened image should possess the high spatial resolution of PAN image and the precise spectral information of the MS image.

A preliminary equalization is performed between the PAN image (P) and MS image bands (X_k) to generate N -bands for PAN image ($\{P_k\}_{k=1,\dots,B}$). Objective of the preliminary equalization is to reduce the inhomogenities between PAN image and MS image bands. The PAN image is preprocessed, as it is histogram matched with each band of MS image. The purpose is to maintain the same mean and variance for PAN and MS images. The histogram matching process results in four PAN images ($P_k, k = 1, 2, 3, 4$) corresponding to four MS image bands and assists in improving the spatial similarity between the PAN and MS images.

The proposed pan-sharpening approach for each band $k= 1,2,\dots,B$. is described as:

Decompose each image P_k and \tilde{X}_k into its corresponding texture (T) and cartoon (C) components as:

$$P_k = T_p^k + C_p^k \quad (5.19)$$

$$\tilde{\mathbf{X}}_k = \mathbf{T}_{\tilde{\mathbf{X}}}^k + \mathbf{C}_{\tilde{\mathbf{X}}}^k \quad (5.20)$$

The notion of CPT decomposition is adapted to separate spatial details and band specific spectral information efficiently and further, to perform the controlled injection of details at the appropriate pixel regions. The legitimate fusion rules are adapted for cartoon and texture components in order to (i) reduce the shortcomings induced by the patch based fusion methods and (ii) enhance the visual quality of the pan-sharpened image extensively.

Fused texture component is obtained by adopting CSC model and subsequently, the activity level measurement of the resultant coefficients obtained from CSC. Fused cartoon component is accomplished by using averaging strategy. The HRMS image is reconstructed by linearly combining the fused texture and cartoon components based on the constituent gradient information.

5.2.1 Fusion of Texture Components

The texture components are predominantly composed of the structural information of the source images. Thus, the spatial details of PAN and MS images can be characterized by its constituent texture components. The specific details injection refrains the spectral distortion in the fused image. Hence, the fusion rules for texture component are developed to strengthen the boundary information and to impart the essential details into the pansharpened image.

Two key issues for the fusion process include, activity level measurement and combining coefficients. The larger activity level implies more information. The appropriate choice of combining coefficients transfers the most useful information into the fused image (Dogra et al., 2017).

The CSC model presented in (Wohlberg, 2016b) is invoked to determine the sparse coefficient maps for the texture components \mathbf{T}^k , $\{ \mathbf{T}_P^k \text{ and } \mathbf{T}_{\tilde{\mathbf{X}}}^k \}$ and is given as:

$$\operatorname{argmin}_{\{\mathbf{M}_{k,j}\}} \frac{1}{2} \left\| \sum_{j=1}^J d_j * \mathbf{M}_{k,j} - \mathbf{T}^k \right\|_2^2 + \lambda \sum_{j=1}^J \|\mathbf{M}_{k,j}\|_1 \quad (5.21)$$

where, $\mathbf{M}_{k,j}, j = 1, 2, \dots, J$ are sparse coefficients maps for the image components \mathbf{T}^k and $\mathbf{M}_{k,1:J}(x, y)$ denotes a J -dimensional vector with the elements of $\mathbf{M}_{k,j}$ at the spatial position (x, y) . The l_1 - norm for each of the vector, $\mathbf{M}_{k,1:J}(x, y)$, is measured and validated

as an activity level measurement.

It is trivial that, there is a misregistration between the images of same scene produced by remote sensing acquisition systems. Owing to misregistration, the further processing like interpolation and fusion between the images of different spatial resolutions do not yield the desired results (Alparone et al., 2015). The performance degradation caused by misregistration can be alleviated by implementing an averaging mechanism based on a window of size $r \times s$. The resultant activity level pattern is obtained as:

$$\bar{L}_k(x, y) = \frac{1}{(2n+1)^2} \sum_{r,s=-n}^n \|\mathbf{M}_{k,1:J}(x+r, y+s)\|_1 \quad (5.22)$$

The value of 'n' decides the size of the window and affects the preservance of finer details. Usually, smaller values are chosen for multi-sensor image fusion and is considered as three in this work.

To ensure that all the important features of source images remain intact, the fused coefficients maps for cartoon component is obtained by invoking choose-max (CM) fusion rule (Yang and Li, 2010). Essentially, under this specific fusion rule, the sparse coefficients maps having the highest corresponding activity level pattern is selected based on the following hypothesis:

$$\mathbf{M}_{f,j}(x, y) = \begin{cases} \mathbf{M}_{u,j}(x, y) & \text{if } \bar{L}_u(x, y) \geq \bar{L}_v(x, y) \\ \mathbf{M}_{v,j}(x, y) & \text{otherwise} \end{cases} \quad \text{for } j = 1, \dots, J. \quad (5.23)$$

Subsequently, the fused texture component, T_f^k is estimated as:

$$T_f^k = \sum_{j=1}^J \mathbf{d}_j * \mathbf{M}_{f,j}. \quad (5.24)$$

Where, * indicates the convolution operation.

5.2.2 Fusion of Cartoon Components

The detailed information about geometric shapes of the source images is constituted by its associated cartoon components. Cartoon component is the piecewise-smooth segment and it possesses the global structural information of the image.

The cartoon component for the fused image is obtained by performing average estimation of the individual components. The inherent variations in the characteristics

of PAN and MS images lead to different intensity values at the same pixel locations. Averaging exercise is simple and the most frequently used strategy to reduce the visual disparity in the resultant image (Liu et al., 2016).

The fused cartoon component C_f^k is obtained as:

$$C_f^k = \frac{1}{2}(C_P^k + C_X^k) \quad (5.25)$$

The rationale in favour of averaging operation is that it preserves the underlying geometric structures of the source images. Similar degree of geometric structure preservice is not possible on using the 'absolute-max' fusion rule.

5.2.3 Pan-sharpened Image Reconstruction

Image gradient information has proven its effectiveness in many image processing applications like, texture analysis, image de-noising etc. For multi modality image fusion, gradient based fusion model has been successfully used in (Zhu et al., 2018).

For an image, \mathbf{I} , at the spatial location (c, d) the gradient information $g^I(c, d)$ can be expressed as :

$$g^I(c, d) = \sqrt{g_x^I(c, d)^2 + g_y^I(c, d)^2} \quad (5.26)$$

The symbols $g_x^I(c, d)$ and $g_y^I(c, d)$ denote the gradients in x and y directions, respectively. The gradient strength for the entire image, \mathbf{I} , of size (C, D) can be calculated as:

$$G^I = \sum_{c=1}^C \sum_{d=1}^D g^I(c, d) \quad (5.27)$$

To preserve the texture information and to overcome the over-smoothing effect in the pan-sharpened image, a fused weight is determined based on gradient strength. HRMS image is reconstructed by the integration of fused texture and cartoon components as:

$$\hat{X}_k = C_f^k + \sqrt{\frac{G^{T_1} + G^{T_2}}{G^{T_f}}}.T_f^k \quad (5.28)$$

G^{T_1} = Gradient strength of T_Y^k , G^{T_2} =Gradient strength of T_X^k and G^{T_f} = Gradient strength of T_f^k

A schematic diagram for the proposed pan-sharpening algorithm is depicted in Figure 5.3.

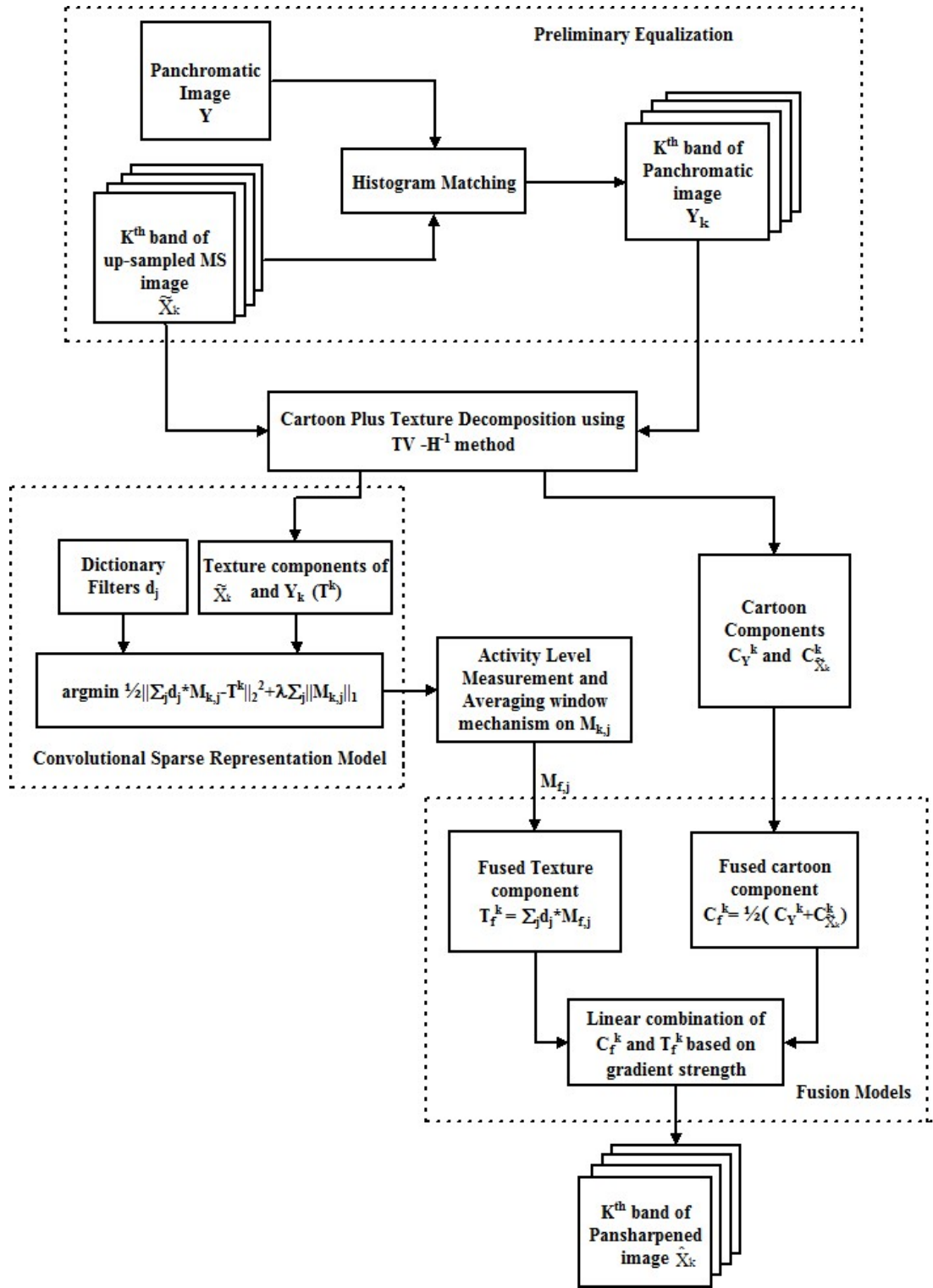


Figure 5.3: The proposed pan-sharpening algorithm schematic.

5.3 Results and Discussions

The performance of the proposed hybrid algorithm is evaluated using datasets with diverse geographical contents obtained from the sensors: QuickBird, IKONOS, Pléiades and WorldView-2. The PAN and MS images having different sizes are used for experimentation. The scale ratio is four for all the considered datasets. The proposed algorithm does not perform any decomposition of the PAN and MS images in different scales and orientations. The dictionary filters are applied for the entire image. Hence, there is no constraint on the resolution ratio between MS and PAN images. Further, the proposed algorithm is insensitive to resolution ratio between source images, i.e., PAN and MS images.

To demonstrate the efficiency of the proposed algorithm in a comprehensive manner, it is tested under various scenarios: Reduced-scale evaluation using synthesis property of Wald's protocol (Wald et al., 1997), Full-scale evaluation using consistency property of Wald's protocol and QNR protocol (Alparone et al., 2008). The proposed method is compared with seven other state-of-the-art PS methods IHS(Carper, 1990), AWLP (Otazu et al., 2005), MTF-GLP (Aiazzi et al., 2006), SR-Li (Li and Yang, 2011), SR-TD(Cheng et al., 2014), SR-D (Vicinanza et al., 2015), SR-LD (Li et al., 2013b) and SR-CD (Ayas et al., 2018). In addition, a resampled MS image is also included during the comparison and is referred as EXP. For quantitative evaluation, five quality/performance metrics are adapted, namely correlation coefficient (CC), spectral angle mapper (SAM), root mean square error (RMSE), relative dimensionless global error in synthesis (ER-GAS), and universal image quality index (Q4).

In the proposed algorithm, $TV - H^{-1}$ approach is used to decompose a given image into cartoon and texture components. The cartoon/texture trade-off parameter (γ) is selected as 4. To impart learning behaviour for the dictionary filters and to estimate sparse coefficients maps, ADMM algorithm is used. In order to reduce the computational complexity and to enhance the reconstruction quality, the large linear system in the subproblem of convolution is solved by adopting Sherman -Morrison (SM) method. The number of iterations for learning filters is considered as 50. The dictionary filters of size $8 \times 8 \times 64$ and 64 sparse coefficient filters with the size equivalent to the corresponding image are implemented for CSC. The execution of CSC algorithm is done as deliberated in (Wohlberg, 2016b).

All the methods are implemented using MATLAB (R2013a), on a computing machine with 8 GB RAM, i5-2400 CPU @ 3.10GHz.

5.3.1 Evaluation Using Synthesis Property

The evaluation of the proposed method at reduced-scale is exercised using Pléiades and IKONOS datasets. For the Pléiades dataset the PAN image is simulated from the available MS bands having 60-cm resolution. This dataset represents the urban area of Toulouse from France. The IKONOS dataset represents a mountainous area of Sichuan from China. IKONOS satellite produces 4-m MS images (4 bands) and 1-m panchromatic image. The images of size 256×256 for PAN image and $64 \times 64 \times 4$ for MS image are used for both the datasets during the reduced-scale experimentation.

The visual results obtained on using different pan-sharpening algorithms for the Pléiades dataset are presented in Fig.5.4. Fig.5.4(a) shows the degraded PAN image. Fig. 5.4(b) shows the MS image up-sampled to the size of PAN image and termed as EXP. Fig.5.4(c) is the reference MS image used for comparison. Fig.5.4(d)-(l) presents the outcomes of different pan-sharpening schemes.

The quantitative results for the Pléiades dataset corresponding to the visual results (Fig.5.4) are presented in Table 5.1. The proposed method attains the optimal values for all the quality metrics except for ERGAS on Pléiades dataset. In comparative terms, the analysis of visual results as shown in Fig.5.4(d)-(l) reveal that on assessing the performance about preserving the spatial details, the proposed method performs much better than the other reported methods. Visual outcomes obtained on using the proposed method indicate that, the roof edges appearance is much sharper than that acquired from the other methods. The MRA based method, AWLP retains moderately good spectral information, however this is achieved at the cost of spatial details. Moreover, the SR based methods utilizing different dictionaries are simulated for the comparison purpose. The efficacy of the proposed method in terms of offering significant enhancement of fusion quality over the SR based methods is evidenced by the visual and quantitative results.

The reduced-scale outcomes of IKONOS dataset are presented in Fig.5.5. Fig.5.5(a) shows the degraded PAN, whereas Fig.5.5(b) is the MS image up-sampled to the resolution of PAN image. Fig.5.5(c) is treated as a reference MS image. The visual outcomes of different PS methods are presented in Fig.5.5(d)-(l). As can be seen from Fig.5.5 the fused image obtained by IHS method (Fig.5.5(d)) suffers from spectral distortion. The outcomes of classical methods, AWLP and MTF-GLP, and SR-Li (Fig.5.5(e), (f), (g)) exhibit good spectral preservation but possess slight blurring artifacts. Besides this the result of SR-TD method (Fig.5.5(h)) is accompanied by slight color distortion at bottom left corner. The fused outcomes obtained on using the SR-LD, SR-D, SR-CD and the

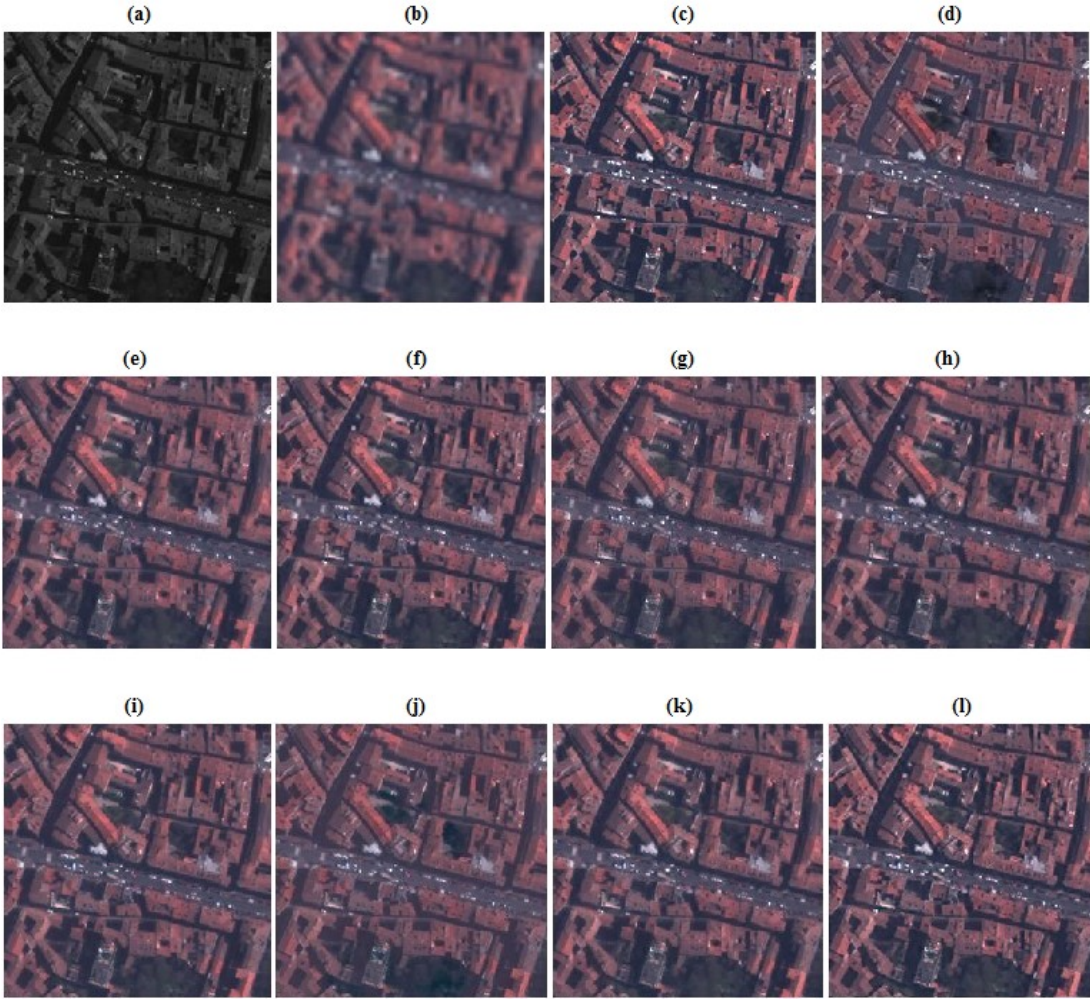


Figure 5.4: Visual results of pan-sharpening methods for Pléiades data at reduced-scale
 (a) Degraded PAN image (b) Up-sampled MS image (EXP) (c) Reference MS image (d) IHS (e) AWLP (f) MTF-GLP (g) SR-Li (h) SR-TD (i) SR-D (j) SR-LD (k) SR-CD (l) Proposed Method

Table 5.1: Quantitative results for Pléiades data using synthesis property

	Q4	ERGAS	SAM	CC	RMSE
EXP	0.7781	5.8893	4.5972	0.6983	16.4325
IHS	0.8629	4.3284	4.7326	0.9116	14.2918
AWLP	0.9463	3.2753	3.4327	0.9283	12.3864
MTF-GLP	0.9515	3.2284	3.4161	0.9311	12.4108
SR-Li	0.9341	3.3587	3.7618	0.9134	13.8942
SR-TD	0.9481	2.8758	3.3297	0.9358	12.9853
SR-D	0.9385	2.9426	3.3286	0.9387	12.2302
SR-LD	0.9479	2.9654	3.2961	0.9323	12.2336
SR-CD	0.9392	2.9651	3.2853	0.9342	12.2351
Proposed	0.9573	2.8761	3.2837	0.9431	11.9727

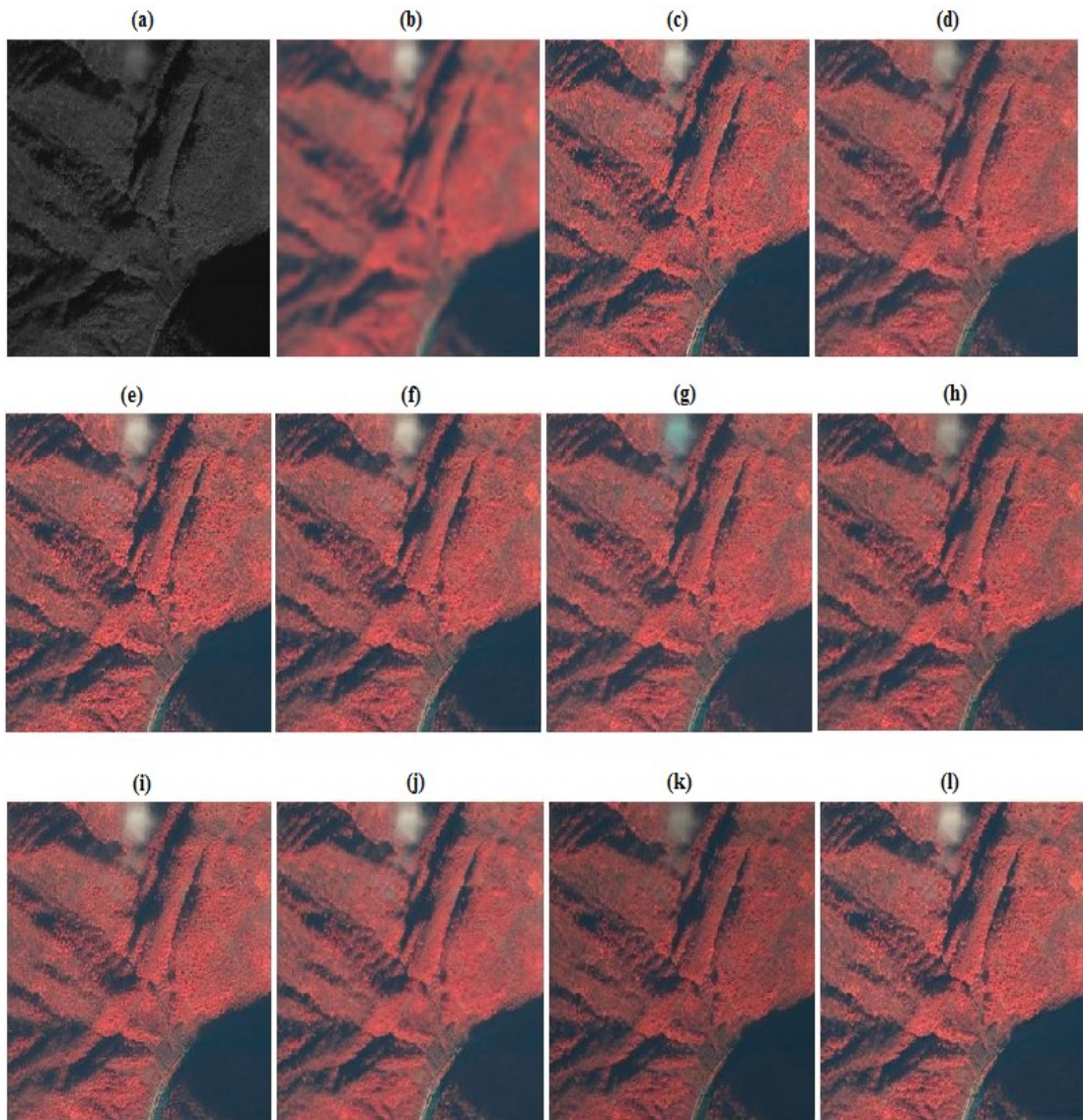


Figure 5.5: Visual results of pan-sharpening methods for IKONOS data at reduced-scale (a) Degraded PAN image (b) Up-sampled MS image (EXP) (c) Reference MS image (d) IHS (e) AWLP (f) MTF-GLP (g) SR-Li (h) SR-TD (i) SR-D (j)SR-LD (k) SR-CD (l) Proposed Method

proposed method, manifest slightly better preservance of spatial details. The proposed method (Fig.5.5(l)) shows the overall improved visual quality and the presence of much sharper details. The quantitative results for IKONOS dataset are presented in Table 5.2. The proposed method achieves the optimal values for performance measures namely, Q4, ERGAS, SAM and RMSE. Although, the performance measures obtained from SR-LD and SR-CD schemes are in close match with that obtained on using the proposed PS scheme, computationally SR-LD is more intensive than the proposed scheme.

The proposed method accomplishes the optimal values for Q4, SAM and RMSE

Table 5.2: Quantitative results for IKONOS data using synthesis property

	Q4	ERGAS	SAM	CC	RMSE
EXP	0.7536	3.9632	4.5362	0.7842	14.6846
IHS	0.7524	3.7653	4.6583	0.8115	13.8975
AWLP	0.8987	2.9243	3.1658	0.9153	12.3156
MTF-GLP	0.8835	3.1572	3.2617	0.9241	12.3426
SR-Li	0.8932	3.4378	3.3476	0.9023	12.6748
SR-TD	0.9287	3.2294	3.2686	0.9226	12.1136
SR-D	0.9274	3.1641	3.2574	0.9247	11.9432
SR-LD	0.9389	2.9534	3.1215	0.9328	11.8361
SR-CD	0.9345	2.9642	3.2453	0.9311	11.9147
Proposed	0.9438	2.9242	3.1157	0.9324	10.9204

for both the datasets. It is clearly evident from the reduced-scale evaluation, that the quality of fused outcomes achieved using the proposed method is significantly enhanced compared to the reported SR based methods.

5.3.2 Evaluation Using Consistency Property

According to the consistency property of Wald's protocol, the pan-sharpening methods are implemented on the source images without encountering the degradation process. Further, the fused image is spatially degraded to the resolution of original MS image, which is considered as a reference image for the comparison.

The dataset acquired from QuickBird sensor represent Sunderbans region located in India, is used for the evaluation using consistency property. QuickBird provides PAN image of 0.7-m resolution and the corresponding 4-band MS image with 2.8-m resolution. The size of PAN and MS images used for experimentation is 256×256 pixels and $64 \times 64 \times 4$ pixels, respectively. Fig.5.6(a) and (b) shows the considered PAN and up-sampled MS images. The corresponding pan-sharpened results are presented in Fig.5.6 (c)-(k).

On analyzing the visual results, it is clearly evident that the fused outcomes obtained on using the IHS and SR-Li methods, are unable to preserve the desired spectral information and this artifact emerges in a form of mild color change in the regions containing trees. The fused outcomes obtained using AWLP, MTF-GLP and SR-D methods display modest blurring of spatial details. On observing the magnified regions of the visual outcomes of schemes SR-TD, SR-LD, SR-CD and the proposed method it can be concluded that, the proposed PS method is able to retain sharp edges in the fused

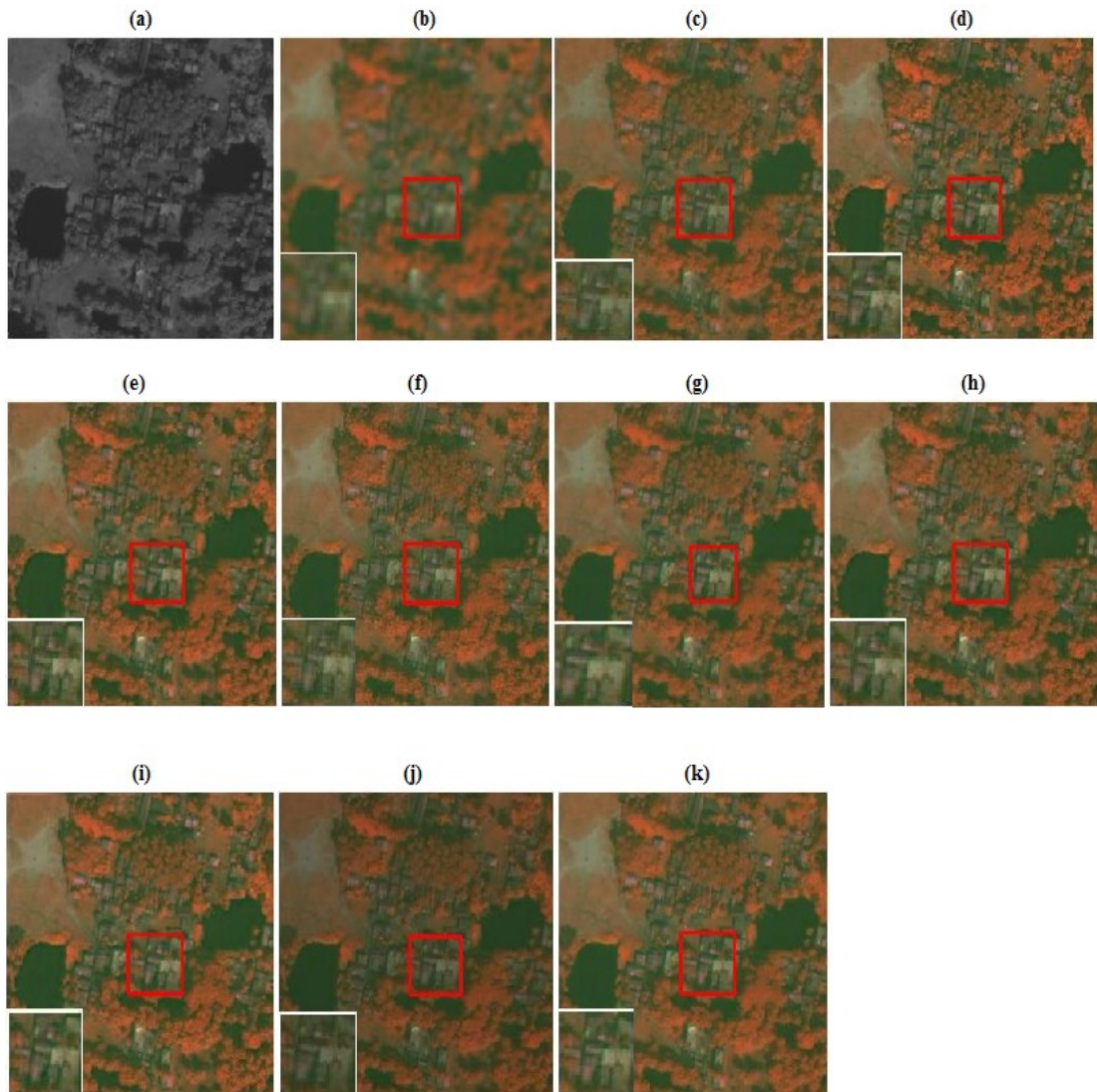


Figure 5.6: Visual results of pan-sharpening methods for QuickBird data at full-scale (a) PAN image (b) Up-sampled MS image (EXP) (c) IHS (d) AWLP (e) MTF-GLP (f) SR-Li (g) SR-TD (h) SR-D (i)SR-LD (j) SR-CD (k) Proposed Method

outcome. The proposed method exhibits comparatively less spectral distortion and preserves the spatial information effectively. For the proposed method the zoomed portion of the image clearly indicates the spatial details preservation. The associated quality metrics values are furnished in Table 5.3. Proposed method achieves the optimal value for performance measures ERGAS, SAM and CC. The SR-CD method results slightly better value than the proposed method for Q4. The distortion indices ERGAS, SAM and the spatial quality measure CC values are optimal for the proposed method. The visual outcomes and the quality metrics validate the superiority of the proposed method over the reported SR based methods.

Table 5.3: Quantitative results for QuickBird data using consistency property

	Q4	ERGAS	SAM	CC	RMSE
EXP	0.9352	1.8467	1.9561	0.9461	15.2419
IHS	0.9273	1.7593	1.9862	0.9317	16.2976
AWLP	0.9521	1.5386	1.7535	0.9592	10.1308
MTF-GLP	0.9532	1.5293	1.6935	0.9882	10.1316
SR-Li	0.9427	1.7462	1.8213	0.9534	11.8675
SR-TD	0.9583	1.4876	1.6392	0.9739	11.6574
SR-D	0.9624	1.5621	1.6427	0.9715	11.8749
SR-LD	0.9664	1.4369	1.5372	0.9747	10.2152
SR-CD	0.9672	1.4371	1.5483	0.9725	10.3927
Proposed	0.9633	1.3287	1.4238	0.9791	10.3542

5.3.3 Evaluation Using QNR Protocol

The Quality with No Reference (QNR) protocol (Alparone et al., 2008) appraises the quality of the pan-sharpened outcomes in the absence of any reference image. QNR protocol consists of spectral distortion index D_λ , spatial distortion index D_s , and a similarity measure QNR.

The dataset acquired from WorldView-2 sensor that represents an area from Sydney, Australia is used for evaluation at full-scale resolution. WorldView-2 sensor provides 0.5-m PAN image and 8-band MS image (In this work only 4 bands are considered for experimentation, Red (R), Green (G), Blue (B) and NearInfraRed (NIR)) with each spectral band maintaining 2-m resolution. The size of the PAN and MS images is 512×512 and $128 \times 128 \times 4$, respectively.

The visual pan-sharpened outcomes for the assessment at full-scale with WorldView-2 data are shown in Fig.5.7. Associated quantitative results are presented in Table 5.4. The proposed method attains optimal values for the spatial distortion index (D_s) and for the overall quality measure (QNR). The quantitative results indicate that the pan-sharpened image obtained on using the proposed method shows finest spatial details along with the overall improved quality. Fig.5.7(a) and (b) shows the PAN and up-sampled MS images, respectively. Fig.5.7(c)-(k) represents the pansharpened outcomes obtained on executing the different methods. In Fig.5.7, the spatial details like roofs of the buildings are blurred in the outcomes obtained from AWLP (Fig.5.7(d)) and SR-Li (Fig.5.7(f)) schemes. A slight color change appears at the green areas in the fused outcome resulting from IHS method as shown in Fig.5.7(c). The outcomes of SR-TD (Fig.5.7(g)), SR-D (Fig.5.7(h)) and SR-LD (Fig.5.7(i)) are oversharpened in some regions. The visual outcome of SR-CD (Fig.5.7(j)) shows better consistency in both

spatial and spectral aspects compared with the other SR based methods used for evaluation. The proposed method (Fig.5.7(k)) retains the essential spectral details and shows moderately good enhancement in spatial components appearance.

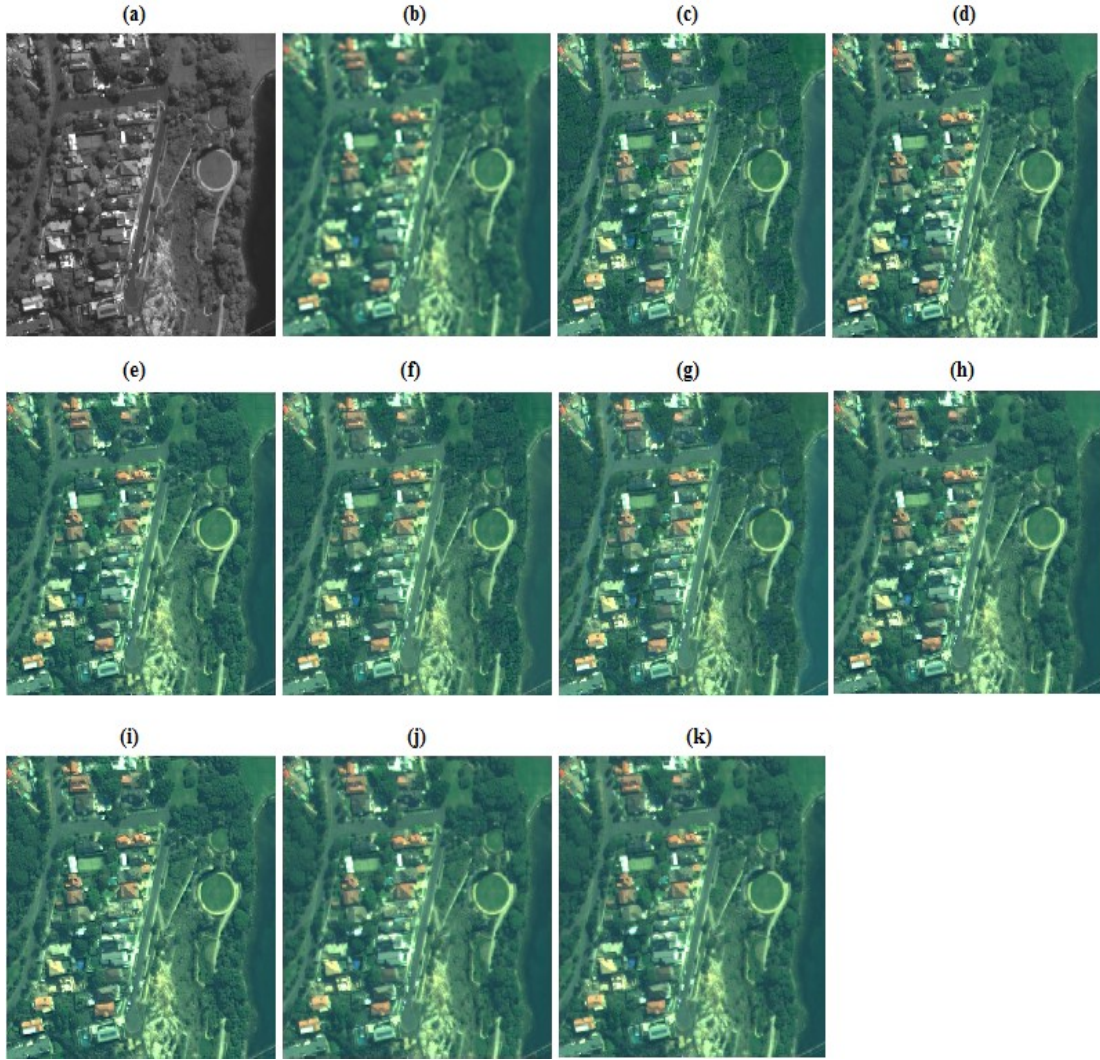


Figure 5.7: Visual results of pan-sharpening methods for WorldView-2 data at full-scale (a) Degraded PAN image (b) Up-sampled MS image (EXP) (c) IHS (d) AWLP (e) MTF-GLP (f) SR-Li (g) SR-TD (h)SR-D (i) SR-LD (j) SR-CD (k) Proposed Method

Table 5.4: Quantitative results for WorldView-2 data using QNR protocol

	EXP	IHS	AWLP	MTF-GLP	SR-Li	SR-TD	SR-D	SR-LD	SR-CD	Proposed
D_λ	0	0.0463	0.0387	0.0368	0.0356	0.0319	0.0271	0.0278	0.0276	0.0274
D_S	0.0968	0.0581	0.0559	0.0514	0.0528	0.0476	0.0356	0.0399	0.0384	0.0358
QNR	0.9032	0.8983	0.9076	0.9137	0.9135	0.9220	0.9322	0.9334	0.9351	0.9378

The comprehensive analysis of visual and quantitative results indicate that, the proposed method is comparable with the SR based pan-sharpening methods, such as SR-D,

SR-LD and SR-CD for all the datasets considered in this work. However, the proposed method does not accomplish the optimal values for all the quality metrics, it is computationally less complex than the SR based methods. The full scale and reduced scale assessments obtained on using the proposed algorithm clearly demonstrates the enhancement in spatial quality, as compared to fused outcomes obtained using the reported SR based methods.

Further, for a broad regime of applications, the performance analysis of algorithm implementation is usually evaluated in terms of parametric variations, time complexity, memory storage requirements etc. In the following two subsections; analysis of parametric variations (varying number of filters) and estimation of the proposed algorithm computational time is investigated and compared with the reported pan-sharpening methods.

5.3.4 Analyzing Impact of Number of Filters

The quality of fusion outcome using CSC is influenced by two crucial factors: the size of filter and number of dictionary filters. In order to learn the necessary details (spatial and spectral features of source images) and to impart precise care for the inherent structures, the filter size is considered as 8×8 pixels. Apart from the size of filters employed, the other parameters values considered is same as given in section 4. The proposed method is validated for 16, 32, 64 and 128 number of dictionary filters, to analyze its impact on three quality metrics SAM, ERGAS and Q4.

On analyzing these characteristics, it can be observed that, for the number of filters equal to or greater than 64, all these three quality metrics attain the steady state value. Since for the number of filters greater than 64, a marginal improvement in the quality metrics with respect to attained steady state value is hard to discriminate, it drives to fix the number of filters parameter as 64. Thus, to maintain the striking balance between the quality metrics and the computational complexity of algorithmic implementation; in the proposed pan-sharpening method the number of filters are considered as 64.

5.3.5 Proposed Algorithm's Execution Time Analysis

The computational time of the proposed method is compared with the other reported pansharpening methods. The execution time estimation with IKONOS data set of size 256×256 pixels is furnished in Table 5.5. All the values of timings reported in Table

5.5 are measured in units of seconds.

Table 5.5: Computational time of different methods for IKONOS dataset

Method	IHS	AWLP	MTF-GLP	SR-Li	SR-TD	SR-D	SR-LD	Proposed
Time(sec.)	0.3	0.9	1.2	1425.3	1865.2	81.3	2869.1	127.8

All these methods are implemented in MATLAB R2013a, on a computing system with 3.1GHz, 8-GB RAM. From the statistics furnished in Table 5.5, IHS, AWLP and MTF-GLP methods consume less time. Although, from the quantitative and qualitative perspectives IHS method exhibits poor performance. AWLP and MTF-GLP methods show considerable improvement over IHS, however, deficient compared with the proposed method. All the SR based methods used for the comparison comprises of dictionary training stage as well. The training phase execution is done in iterative manner on using the image patches. Therefore, the class of SR methods requires more computation time compared to CS and MRA methods. Though, SR-Li method doesn't involve any training phase, however, the relatively higher computation time owes to a large size dictionary synthesized using raw patches.

For the given implementation specifications, the performance of SR-Li is also not competitive with the remaining SR based methods. SR-TD and SR-LD methods involve dictionary which is trained using the patches of PAN and MS images. This task is obviously computationally intensive thus, demands higher execution time. Compared to all these methods, SR-D scheme is comparatively faster since the dictionary is constructed by means of direct filtering operations. Although, the accompanying limitation of these SR based methods is that the trade-off between the spatial details and spectral preservation is unbalanced and thus inadequate. SR-CD method also involves the dictionary training process. Compared with the other SR based methods, in SR-CD method, fewer samples are used in training process to acquire the comparable quality for fused outcome. Obviously, the increase size of the dictionary increases the quality of the pansharpened image as well as computational complexity.

The computation time of the proposed method is given in Table 5.5, for the images of size 256×256 pixels, filter size 8×8 pixels and 64 number of dictionary filters. The CSC based pan-sharpening method with specified parameters described in this work, is attractive in terms of visual quality, quantitative performance metrics and computational cost. The proposed method is computationally superior in terms of supportive execution time when compared to the most of the state-of-the-art SR based methods.

PART II

Pan-sharpening With Discrete Wavelet Decomposition and Convolutional Sparse Representation

To overcome the drawbacks of traditional PS methods, an efficient pansharpening (PS) method is proposed in order to attain the prevalence of discrete wavelet transform (DWT) and CSR paradigms. Notably, in this method the PAN and MS images are decomposed into approximate and detail coefficient bands using discrete wavelet transform. Whereas, in the method presented in PART I ([Gogineni and Chaturvedi, 2019](#)), cartoon and texture decomposition is implemented. The CSR mechanism maintains the shift-invariance feature, which is an essential component for image fusion. The traditional fusion models such as 'max-absolute' and 'averaging components' yields smoothing effects in the fused image. The proposed PS technique implements the fusion models that are designed to transmit the important details from source images into the fused image. The CSR paradigm is invoked to merge the approximate coefficient bands obtained by wavelet decomposition. The detail coefficient bands are fused based on the relative wavelet energy estimated over a specified region. Finally, the fused approximate and detail coefficient bands are merged to synthesize the HRMS image. The qualitative and quantitative outcomes approve that the explored fusion models in this method, effectively impart the most consistent features from PAN and MS images to the HRMS image.

5.4 Pan-Sharpener Framework With Combined DWT and CSR

The low-resolution MS (LRMS) image, X_k , $k = 1, 2 \dots B.$, is up-sampled to get the resolution of the PAN image, is represented as \tilde{X}_k . The histogram equalization operation is performed between PAN image, P , and the up-sampled LRMS image, \tilde{X}_k , to generate the corresponding equalized PAN images, P_k . A pansharpening technique that exploits the features of discrete wavelet transform (DWT) and convolutional sparse representation (CSR) is presented in this section.

The convolutional sparse coding based fusion rules are implemented in the wavelet domain in order to overcome the shortcomings of traditional pansharpening methods. In contrast to earlier approaches, the fusion process is performed in the wavelet domain

but not in image domain. For each band, i.e., $k=1,2,\dots,B$; the proposed pansharpening framework is presented in Fig. 5.8.

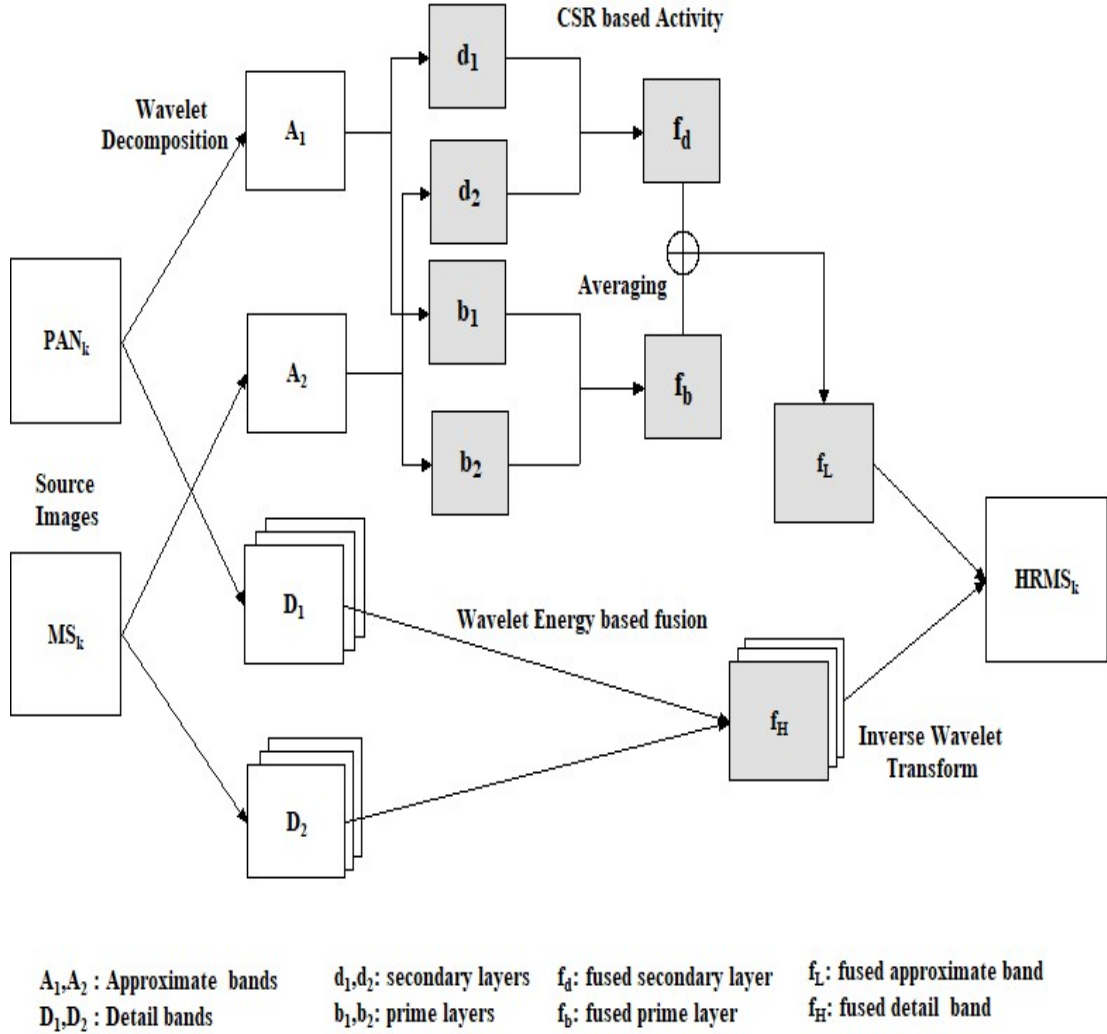


Figure 5.8: The combined DWT and CSR based pan-sharpening schematic

5.4.1 Decomposition Using Discrete Wavelet Transform

The pansharpening outcome is considerably affected by the level of decomposition of the wavelet transform. For the higher decomposition levels, a single coefficient in the transform domain influences a relatively large area of the image. Hence, a small error in decomposed bands causes primitive artifacts in the resultant image obtained by the inverse transformation. To make the process robust to noise and misregistration, the wavelet decomposition level is selected as two in the proposed pansharpening method between PAN and MS images.

The source images; interpolated MS image, \tilde{X}_k and PAN image P_k are decomposed using wavelet transform to generate the corresponding approximate coefficient bands $\{A_{\tilde{X}_k}, A_{P_k}\}$ and detail coefficient bands $\{D_{\tilde{X}_k}, D_{P_k}\}$.

5.4.2 Fusion of Approximate Coefficient Bands

In contrast to the conventional patch-partition based methodology used in SR schemes, CSR is an efficient coding paradigm to represent the entire image using sparse coefficient maps. In CSR mechanism, any image, \mathbf{I} is represented as a sum of set of convolutions between dictionary filters \mathbf{d}_i and the unknown sparse coefficient maps \mathbf{x}_i .

$$\arg \min_{\{\mathbf{x}_i\}} \frac{1}{2} \left\| \sum_i \mathbf{d}_i * \mathbf{x}_i - \mathbf{I} \right\|_2^2 + \lambda \sum_i \|\mathbf{x}_i\|_1 \quad (5.29)$$

Where, λ is a regularization parameter and '*' represents convolution operator.

Various image fusion methods (Li et al., 2013a; Liu et al., 2016) have been developed based on the decomposition of source images into its constituent prime and secondary layers. The prime layer composed of large-scale intensity variations and the secondary layer consists of small-scale variations of the source image. Since the approximate coefficient bands also consist of small intensity variations that reflect the geometric features, the prime and secondary layer decomposition strategy is deployed.

The approximate coefficient bands of MS and PAN images are further decomposed into its constituent prime and secondary layers. The prime layer, $A_{\tilde{X}_k}^b$ of approximate coefficient band $\{A_{\tilde{X}_k}\}$ of MS image is estimated by solving the optimization problem formulated in Eq.5.30. Similarly, the solution of Eq.5.31 yields the prime layer, $A_{P_k}^b$ of approximate coefficient band, $\{A_{P_k}\}$ of the PAN image.

$$\arg \min_{A_{\tilde{X}_k}^b} \|A_{\tilde{X}_k} - A_{\tilde{X}_k}^b\|_F^2 + \eta (\|g_x * A_{\tilde{X}_k}^b\|_F^2 + \|g_y * A_{\tilde{X}_k}^b\|_F^2) \quad (5.30)$$

$$\arg \min_{A_{P_k}^b} \|A_{P_k} - A_{P_k}^b\|_F^2 + \eta (\|g_x * A_{P_k}^b\|_F^2 + \|g_y * A_{P_k}^b\|_F^2) \quad (5.31)$$

The parameters $g_x = [-1 \ 1]$ and $g_y = [-1 \ 1]^T$ are horizontal and vertical gradient operators, respectively. The notion $\|\cdot\|_F$ represents Frobenius norm. The regularization parameter η is selected as 5 for the proposed CSR based pansharpener scheme. The

Tikhonov regularization (Eq.5.30, Eq.5.31) based fast Fourier transform computes the prime layer. Further, the secondary layers are be estimated as

$$A_{\tilde{X}_k}^d = A_{\tilde{X}_k} - A_{\tilde{X}_k}^b. \quad (5.32)$$

$$A_{P_k}^d = A_{P_k} - A_{P_k}^b. \quad (5.33)$$

Here, $A_{\tilde{X}_k}^d$ and $A_{P_k}^d$ are the secondary layers decomposed from $A_{\tilde{X}_k}$ and A_{P_k} , respectively.

A set of dictionary filters $\mathbf{d}_m, m = 0, 1 \dots M.$ are learned from the approximate coefficient bands by using the ADMM dictionary learning algorithm. The implementation details and complete description of ADMM algorithm to solve the CBPDN problem are given in (Wohlberg, 2015).

The sparse coefficients maps $x_{i,m}$ for each of the secondary layer $A^d = \{A_{\tilde{X}_k}^d, A_{P_k}^d\}$ are obtained by solving the CSR model.

$$\arg \min_{\{x_{i,m}\}} \frac{1}{2} \left\| \sum_{m=1}^M \mathbf{d}_m * x_{i,m} - A^d \right\|_2^2 + \lambda \sum_{m=1}^M \|x_{i,m}\|_1. \quad (5.34)$$

With $i = 1, 2.$ and $m = 1 \dots M.$ Where $i = 1$ corresponds to the sparse coefficient maps of $A_{\tilde{X}_k}^d$ and $i = 2$ corresponds to the sparse coefficients maps of $A_{P_k}^d$.

$\|L(c, d)\|_1$ is the l_1 norm for the sparse coefficient maps $x_{1:M}(c, d)$, which is considered as activity level measure. Here, c and d represents the spatial coordinates. The activity level pattern for the sparse coefficient maps, $x_{i,m}$ is evaluated with a window of size (r, s) as

$$\bar{L}_i(x, y) = \frac{1}{(2n+1)^2} \sum_{r,s=-n}^n \|x_{1:M}(c+r, d+s)\|_1 \quad (5.35)$$

The value of 'n' determines the size of the window. Further, to sustain the fine details of the source images, the value of n is considered as 3 for the pansharpening of multispectral images.

The fused coefficient maps, $x_{f,m}$ are evaluated by executing the following decision

rule.

$$x_{f,m}(c, d) = \begin{cases} x_{1,m}(c, d) & \text{if } \bar{L}_1(c, d) \geq \bar{L}_2(c, d) \\ x_{2,m}(c, d), & \text{otherwise} \end{cases} \quad \text{for } m = 1 \dots M. \quad (5.36)$$

Subsequently, the band specific fused secondary layers are obtained as

$$A_{f_k}^d = \sum_{m=1}^M d_m * x_{f,m}. \quad (5.37)$$

The fused prime layer is obtained by an averaging mechanism

$$A_{f_k}^b = \frac{1}{2}(A_{\tilde{X}_k}^b + A_{P_k}^b) \quad (5.38)$$

The aggregate approximate fused band, A_{f_k} is generated as

$$A_{f_k} = A_{f_k}^d + A_{f_k}^b \quad (5.39)$$

5.4.3 Fusion of Detail Coefficient Bands

The associated Wavelet energy is used to fuse the detail coefficient bands, and is defined as

$$\check{E}(w) = \frac{1}{w_1 \times w_2} \sum_{c=1}^{w_1} \sum_{d=1}^{w_2} g(c, d)^2 \quad (5.40)$$

Where, $g(c, d)$ is the wavelet coefficient at the spatial location (c, d) and w is a local window of size $w_1 \times w_2$ pixels. The fused detail coefficient band, D_{f_k} is produced by imposing the following rule.

$$D_{f_k}(c, d) = \begin{cases} D_{\tilde{X}_k}(c, d) & \text{if } \check{E}(w_{\tilde{X}_k}(c, d)) > \check{E}(w_{P_k}(c, d)) \\ D_{P_k}(c, d), & \text{otherwise} \end{cases} \quad (5.41)$$

Where, $w_{\tilde{X}_k}(c, d)$ and $w_{P_k}(c, d)$ are the windows centered at the pixel locations (c, d) of the bands $D_{\tilde{X}_k}$ and D_{P_k} , respectively and $\check{E}(\cdot)$ represents the wavelet energy of the local window in an image, centered at the spatial location (c, d) .

Finally, the k^{th} band of fused high-resolution MS image, \hat{X}_k is reconstructed by performing the inverse wavelet transform on fused approximate and detail components A_{f_k} and D_{f_k} , respectively.

5.5 Results and Analysis

In this section, the proposed pan-sharpening method is validated using different satellites datasets including QuickBird, IKONOS and Pléiades. The proposed method is compared with ten state-of-the-art methods: FIHS (Tu et al., 2004) from CS category, AWLP (Otazu et al., 2005), and MTF-GLP (Aiazzi et al., 2006) from MRA category, SR-LD (Li et al., 2013b), SR-D (Vicinanza et al., 2015), random sampling-based SR method termed as SR-RS (Yang et al., 2018) and SR-CD (Ayas et al., 2018) from sparse representation, SR-WT (Cheng et al., 2015) from wavelet decomposition based methods, and CSR-D (Fei et al., 2019) and CSR-R (Gogineni and Chaturvedi, 2019) from the convolutional sparse coding based PS methods. Further, the interpolated MS image by using a 23 coefficient polynomial kernel (Aiazzi et al., 2002) (termed as EXP) is also used for the comparison.

The quality metrics defined by authorized protocols such as correlation coefficient (CC), spectral angle mapper (SAM), erreur relative global adimensionnelle de synthese (ERGAS), root mean square error (RMSE) and Q4 (4-band extension of universal image quality index (UIQI)) are used to validate the efficiency of the proposed method. The implementation details for the conventional FIHS, AWLP, and MTF-GLP methods are considered from the software package provided by (Vivone et al., 2014). The parameters for SR based methods used in the comparison are manually tuned to achieve the optimal performance. For the implementation of the proposed method, DWT with two-level decomposition is used for the decomposition of source images. The regularization parameter λ in CSR (Eq.5.29) is selected as 0.12 to maintain a balance between the computational effort and performance outcomes. The local window size to measure the wavelet energy is set as 3×3 pixels.

All the experiments are conducted using MATLAB2015a, on a personal computer with CPU intel core i5 @ 3.10 GHz, 8 GB RAM.

5.5.1 Experimental Results With QuickBird Dataset

QuickBird sensor produce a 0.7-m PAN image and a 2.8-m MS image having four bands. The synthesis property of Wald's protocol (Wald et al., 1997) is adapted to test the effectiveness of the proposed method at a reduced-scale, where the source images are filtered with 7×7 Gaussian filter and down-sampled by a factor four. The pansharpening process is performed on degraded images. The fused outcome is compared with

the original MS image, which is treated as a reference image.

The PAN image, reference MS image and the interpolated MS image (termed as EXP) are presented in Fig.5.9(a), (b) and (c), respectively. The visual outcomes of the

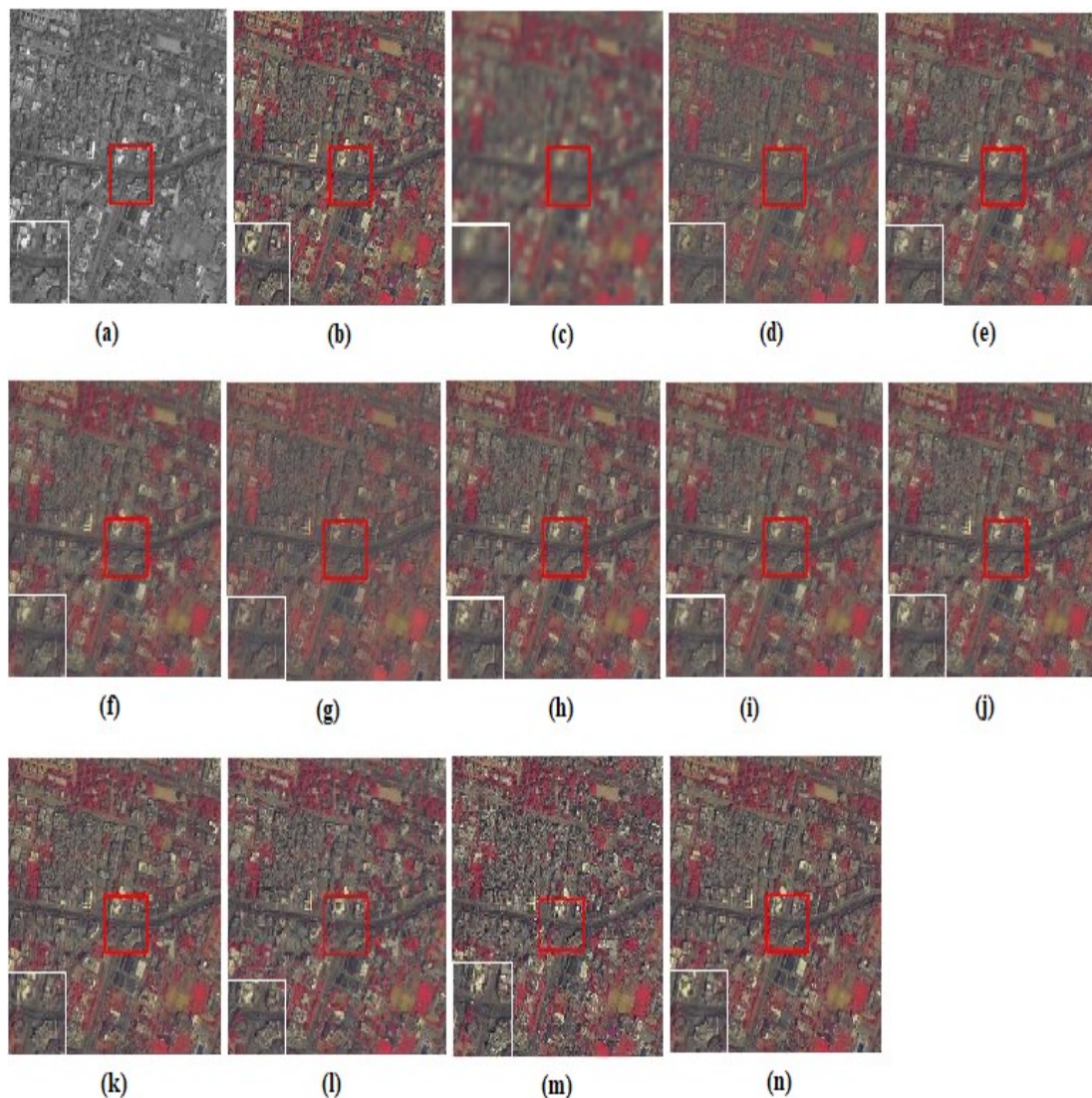


Figure 5.9: Visual results of QuickBird data (a) PAN image (b) Reference MS image (c) Up-sampled MS image (EXP) (d) FIHS (e) AWLP (f) MTF-GLP (g) SR-LD (h) SR-WT (i) SR-D (j) SR-RS (k) SR-CD (l) CSR-D (m) CSR-R (n) Proposed method

CS and MRA methods; FIHS, AWLP and MTF-GLP are presented in Fig.5.9(d), (e) and (f), respectively. The pan-sharpened images obtained from the five SR based methods are presented in Fig.5.9(g)-(k). The two CSR based methods and the proposed method's outcome are presented in Fig.5.9(l)-(n). The corresponding quantitative measures are given in Table 5.6. The optimal value for each metric is indicated in boldface. The second best optimal value is highlighted by an underline. A section of the image that

covers a part of the highway region (enclosed in red colored square box) is zoomed and shown separately at the left bottom corner of every image.

Table 5.6: Quantitative measures for QuickBird dataset

	CC	ERGAS	RMSE	SAM	Q4
EXP	0.8531	4.1436	17.1492	4.1127	0.8253
FIHS	0.8718	3.2146	16.2178	3.6403	0.8559
AWLP	0.8904	2.6168	15.7317	3.5371	0.8716
MTF-GLP	0.9017	2.5361	15.6193	3.5411	0.8738
SR-LD	0.9161	2.6273	15.7145	3.5281	0.8757
SR-WT	0.9247	2.6384	15.8875	3.3874	0.8846
SR-D	0.9265	2.6196	14.6329	3.5673	0.8891
SR-RS	0.9345	2.5127	14.3127	3.2115	0.9286
SR-CD	0.9376	<u>2.4683</u>	13.9638	3.2143	0.9245
CSR-D	0.9379	2.4623	13.7431	<u>3.1519</u>	0.9316
CSR-R	<u>0.9380</u>	2.4762	<u>13.6147</u>	3.2218	<u>0.9288</u>
Proposed	0.9382	2.4759	13.2618	3.1473	0.9251

It can be observed from Fig.5.9 (d) and Fig.5.9(g) that, the outcomes of FIHS, and SR-LD are exhibiting spectral distortion and appears to be blurred compared with the reference image (Fig.5.9(b)). The wavelet based method, SR-WT (Fig.5.9(h)) suffers from minimal blocking artifacts particularly at the roofs of the buildings. Fig.5.9(e), Fig.5.9(f), Fig.5.9(j) and Fig.5.9(k) shows the outcomes of AWLP, MTF-GLP, SR-RS, and SR-CD, respectively, appears to have better visual quality. The edges are not well sharp enough in the outcome of SR-D (Fig.5.9(i)).

It is evident from the visual outcomes that the CSR based methods (Fig.5.9(l) and (m)) are relatively better in preserving the spectral information and retaining the geometric details compared with the other reported methods. The outcome of the proposed method (Fig.5.9(n)) is relatively in a close match with the reference MS image. It is obvious from the zoomed portion that, the proposed method better preserves the spatial structures and yields the minimum possible spectral distortion.

From the quantitative measures presented in Table 5.6, it is evident that the proposed method is effective in maintaining the reasonable balance between the spatial and spectral features in the pansharpened image.

5.5.2 Experimental Results With IKONOS Dataset

The performance of the proposed method is estimated at a full-scale using the consistency property of Wald's protocol. The source images are 1-m PAN image and 4-m MS image with four bands, captured by the IKONOS sensor representing a mountainous area of Sichuan from China. The pan-sharpening process is executed on the original data set further, for comparison purpose the fused outcome is down-sampled to the size of the original MS image. The visual outcomes of different methods used for comparison are shown in Fig.5.10.

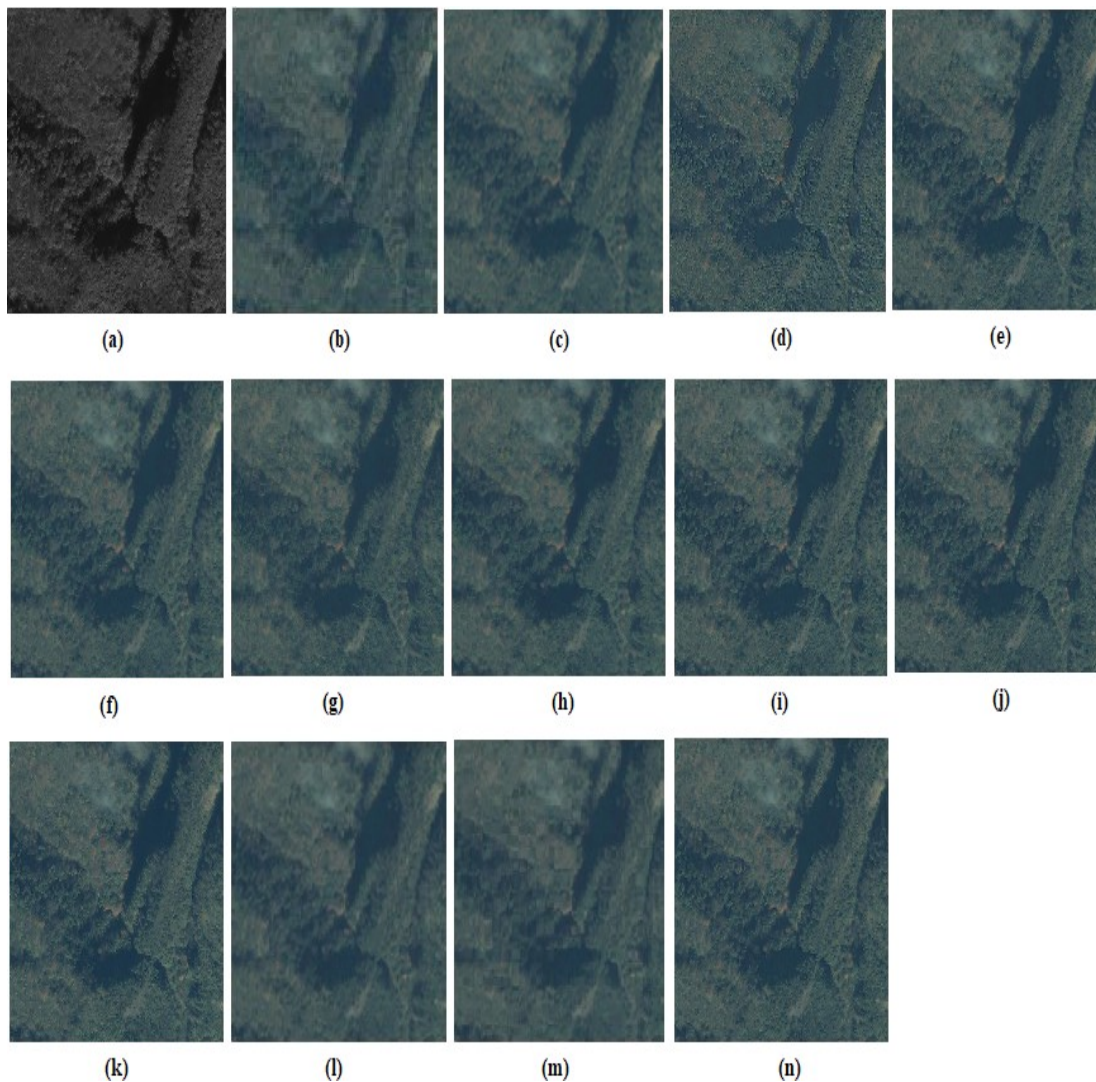


Figure 5.10: Visual results of IKONOS data (a) PAN image (b) Reference MS image (c) Up-sampled MS image (EXP) (d) FIHS (e) AWLP (f) MTF-GLP (g) SR-LD (h) SR-WT (i) SR-D (j) SR-RS (k) SR-CD (l) CSR-D (m) CSR-R (n) Proposed method

The PAN image, reference MS image, and the up-sampled MS image are shown in

Fig.5.10 (a), (b) and (c), respectively. The outcome of the FIHS (Fig.5.10(d)) method suffers from intensity distortion. The consistent details appear in the outcomes of AWLP and MTF-GLP methods are shown in Fig.5.10 (e) and (f), respectively. The fused outcome of SR-LD, SR-WT and SR-D methods are unable to effectively preserve the spectral details. It is difficult to compare the performance of SR-RS, SR-CD and the proposed method visually. The pansharpened images produced by the CSR based methods are quite similar to the reference MS image. To validate the effectiveness, the difference between reference MS image and the outcomes of the reported methods are presented in Fig.5.11. The quantum of difference appeared as a mix of bright region

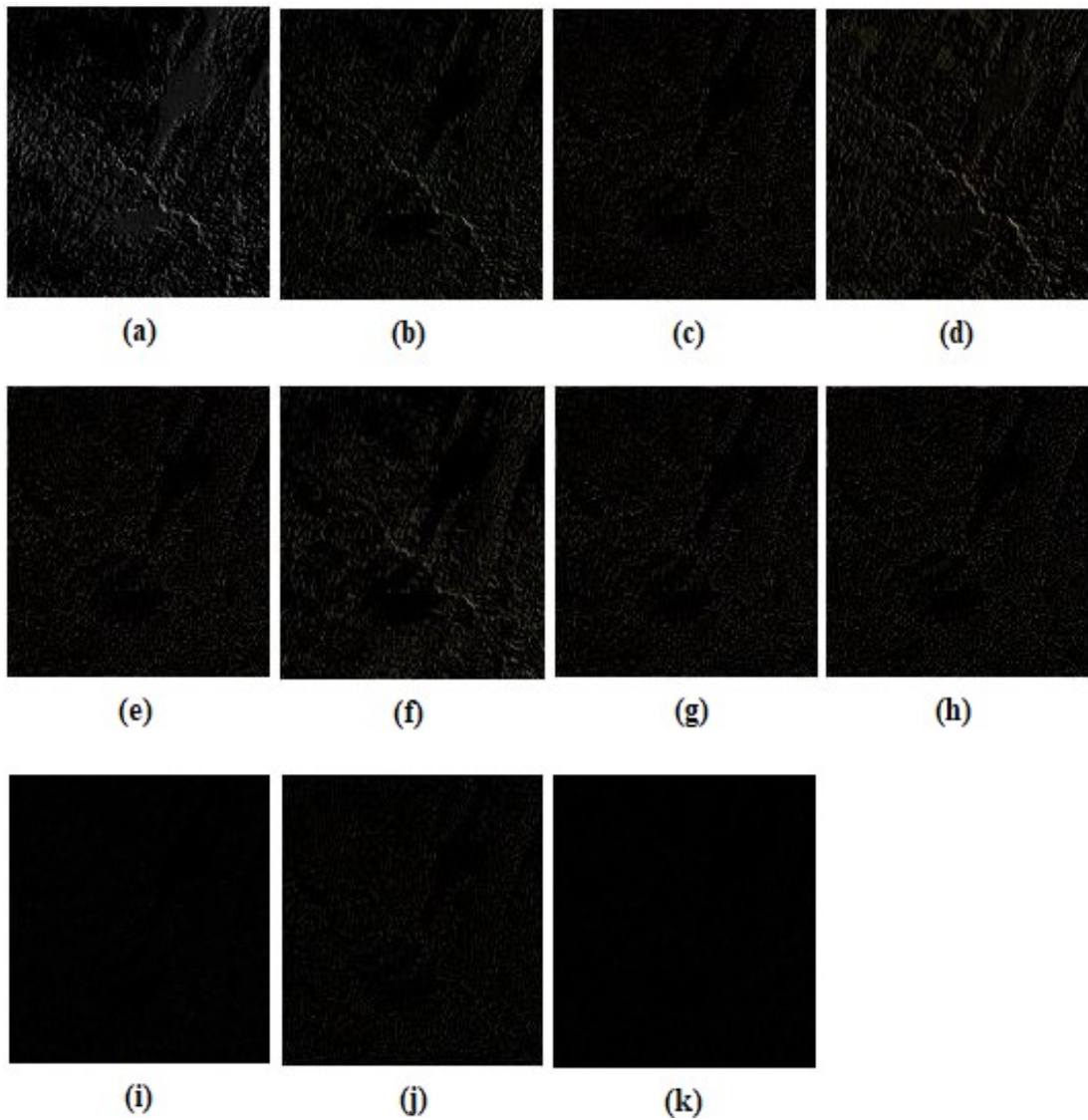


Figure 5.11: Difference between the fused outcomes and the reference MS image for IKONOS data (a) FIHS (b) AWLP (c) MTF-GLP (d) SR-LD (e) SR-WT (f) SR-D (g) SR-RS (h) SR-CD (i) CSR-D (j) CSR-R (k) Proposed method

and the dark regions. Further, the dark regions indicate that there is no substantial difference. It can be observed from the absolute difference images that the outcomes of CS and MRA based methods exhibit considerable difference and appears as bright regions. The difference images for CSR based outcomes and the proposed method (Fig.5.11(i)-(k)) show maximum similarity with the reference image, which is evidenced by the dark regions. Besides this the quantitative measures are presented in Table 5.7. The proposed method yields optimal values for CC, RMSE, Q4 and ERGAS and the second best value for SAM.

Table 5.7: Quantitative measures for IKONOS dataset

	CC	ERGAS	RMSE	SAM	Q4
EXP	0.8853	4.8329	15.1183	4.5736	0.8517
FIHS	0.9136	4.3163	14.9375	3.9537	0.8847
AWLP	0.9468	4.1461	14.8760	3.4817	0.9284
MTF-GLP	0.9481	4.1527	14.5397	3.3295	0.9279
SR-LD	0.9511	4.1731	14.2749	3.3417	0.9273
SR-WT	0.9542	3.9426	14.1261	3.2943	0.9268
SR-D	0.9558	3.9257	13.8672	3.2397	0.9313
SR-RS	<u>0.9617</u>	3.4621	13.4293	3.2104	0.9391
SR-CD	0.9587	3.5738	13.2841	3.1463	0.9485
CSR-D	0.9591	3.4179	13.1164	3.2018	0.9496
CSR-R	0.9573	<u>3.3884</u>	13.2175	3.2183	<u>0.9514</u>
Proposed	0.9624	3.3386	<u>13.2103</u>	<u>3.1792</u>	0.9536

5.5.3 Experimental Results With Toulouse (Pléiades) Dataset

The assessment at the full-scale is estimated by quality with no reference (QNR) protocol (Alparone et al., 2008) and is comprises of two metrics: spectral distortion index D_λ and spatial distortion index D_s . However, the index QNR itself is a quality metric and infers the similarity measure. The fusion mechanism is performed on the original dataset without decimating the source images, and the MS image is interpolated to the size of the PAN image.

The PAN image is shown in Fig.5.12(a) and interpolated MS image is shown in Fig.5.12(b). The visual outcomes of different CS and MRA methods used for comparison are presented in Fig.5.12(c)-(e). The visual outcomes of different SR based methods

used for comparison are shown in Fig.5.12(f)-(j). The outcomes of CSR based methods and the proposed method are shown in Fig.5.12(k)-(m). The dataset is composed of an

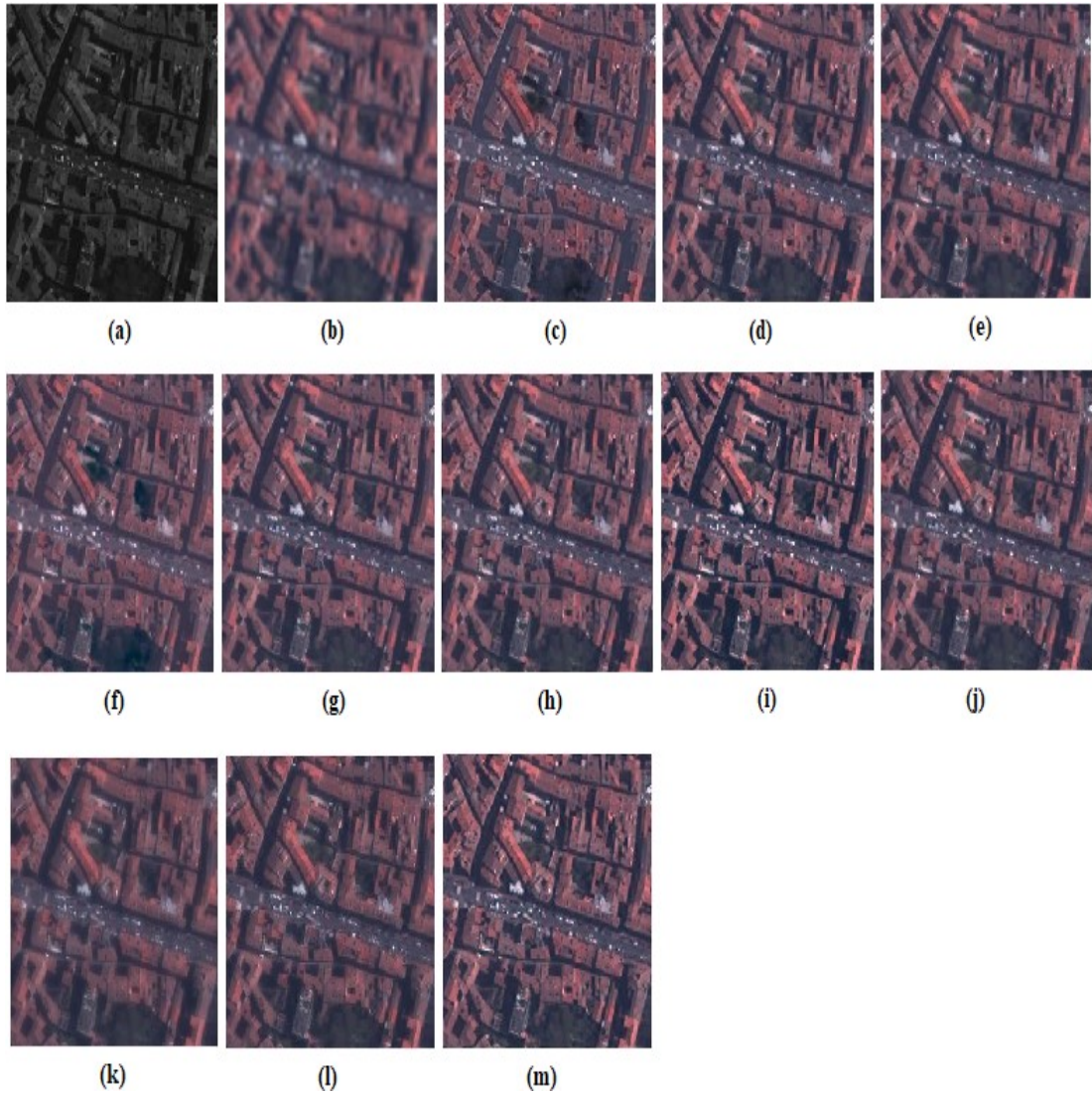


Figure 5.12: Visual results of Pléiades data (a) PAN image (b) Up-sampled MS image (EXP) (c) FIHS (d) AWLP (e) MTF-GLP (f) SR-LD (g) SR-WT (h) SR-D (i) SR-RS (j) SR-CD (k) CSR-D (l) CSR-R (m) Proposed method

urban area, buildings, and roads from Toulouse city, France. It can be observed that the outcomes of FIHS and SR-LD suffers from distortion at the building roofs and roads segment. The spatial distortion can be easily evidenced in some other segments for the outcomes of SR-WT and MTF-GLP. According to the visual results, AWLP, SR-RS, SR-CD, and the proposed method outcomes appear to be relatively sharper.

The corresponding quantitative measures are presented in Table 5.8. From the quality metrics, the CSR-D and CSR-R methods outperforms and yield the second best

optimal value in terms of overall quality measure QNR. For the remaining two metrics, D_λ and D_s , the proposed method leads to slightly better values than the CSR based methods. The optimal values are obtained for spatial distortion index, spectral distortion index and the global quality metric (QNR) using the proposed method.

The visual and quantitative results at the full-scale evaluation approve the effectiveness of the proposed method in preserving the spatial details and maintaining the requisite spectral information.

Table 5.8: Quantitative measures for Pléiades data

Method	D_λ	D_s	QNR
EXP	0	0.0934	0.9066
FIHS	0.0351	0.0473	0.9193
AWLP	0.0343	0.0417	0.9254
MTF-GLP	0.0397	0.0386	0.9232
SR-LD	0.0375	0.0411	0.9229
SR-WT	0.0368	0.0394	0.9252
SR-D	0.0281	0.0359	0.9370
SR-RS	0.0286	0.0347	0.9377
SR-CD	0.0279	0.0336	0.9394
CSR-D	<u>0.0265</u>	0.0332	<u>0.9412</u>
CSR-R	0.0268	<u>0.0329</u>	<u>0.9412</u>
Proposed	0.0246	0.0327	0.9435

The comprehensive experimental results obtained at reduced-scale and full-scale evaluation validate that the proposed method effectively overcome the drawbacks of conventional SR based methods and competitive with the other CSR based methods. The proposed method yields 2.31% improvement compared with the second best results in spectral distortion index (SAM) for QuickBird dataset and second optimal value for IKONOS dataset. In addition at full-scale evaluation the proposed method accomplishes the optimal value for spectral distortion index. Also the proposed method accomplishes optimal values for the overall quality measures CC and RMSE. For the full-scale evaluation, the proposed method shows better performance in terms of spatial and spectral distortion reduction as well as quality of the overall outcomes. The visual and quantitative results confirmed that the proposed method effectively preserves all the requisite details in the fused image.

5.6 Conclusion

Convolutional sparse coding (CSC) is an effective mechanism developed in recent past, to deal with the shortcomings of traditional patch-partition based SR fusion methods. Further, CSR based pansharpening methods yield a single-valued output by exploiting the shift-invariance feature.

In the first part of this work, a novel pan-sharpening algorithm based on cartoon and texture decomposition and the convolutional sparse coding is developed. High-frequency textures are extracted from the PAN and upsampled MS images by means of TV-Hilbert decomposition. The fused texture component is acquired by measuring the activity level of sparse coefficient maps obtained using CSC mechanism. To preserve the underlying geometric structures of the source images, the cartoon components are averaged, which leads to the fused cartoon component. In addition, the fusion rules are appropriately chosen in order to holistically transfer spatial and spectral features into the resultant fused image. In the reconstructed HRMS image, to ensure the preservation of essential spatial details, gradient-based fusion rule combines fused cartoon and texture components. Furthermore, to analyze the impact of the number of filters used in conjunction with CSC, it has been treated as a parameter during quality metrics determination.

The second part of this chapter presents a pan-sharpening method based on the combination of discrete wavelet transform and convolutional sparse representation. The source images are decomposed by using discrete wavelet transform, into corresponding approximate and detail coefficient bands. The approximate coefficient bands are fused using CSR based activity level measures. The wavelet energy constituted by a window of specified size is used to fuse the detail coefficient bands. Finally, the desired high-resolution MS (HRMS) image is synthesized by applying the inverse wavelet transform over fused approximate and detail coefficient bands.

The experimental results obtained at reduced-resolution and full-resolution are used to evaluate the proposed method. The pan-sharpened images with promising spectral quality are more suitable for applications like land-use classification and object detection and extraction. The classification accuracy increases with the decrease of spectral distortion. The visual outcomes, particularly the difference images specify the enhancement of geometric structures in the fused outcome by the proposed method. The specified quality makes the HRMS image more suitable for a particular application like change detection. The visual results and quality metrics confirm the eminence of the

CSR based pan-sharpening methods developed in this work, in maintaining a balanced trade-off between geometric details enhancement and spectral richness in the fused image. A strong structural correlation usually exists among the MS image bands. Thus, in future investigations, the exploitation of correlation information can improve the spectral quality and hence, the overall performance of the fused image.

CHAPTER 6

A VARIATIONAL PAN-SHARPENING ALGORITHM TO ENHANCE THE SPECTRAL AND SPATIAL DETAILS

Various kinds of pan-sharpening methods have been developed over the past few decades to produce high-resolution multispectral (HRMS) images (Vivone et al., 2015). The component substitution (CS) based methods yield the images with reasonably good visual quality but suffer from the spectral distortion. The well-known multi resolution analysis (MRA) based methods in which spatial details are injected into MS image bands that are extracted from the panchromatic (PAN) image via multi resolution decomposition. The HRMS images produced by the MRA based methods incur spatial degradation in the form of blocky and aliasing artifacts, though the spectral distortion is considerably reduced in comparison with the CS based methods. The sparse representation (SR) based methods are associated with patch based processing and intricate dictionary training phases. The amount of overlapping of patches and the size of the designed dictionary greatly affects the quality of the fused image. The general conclusion reveals that there is a trade-off between spatial and spectral quality in the fused image produced by the aforementioned classes of pan-sharpening (PS) methods.

Recently, variational methods (Duran et al., 2017) have been evolved as a prominent and attractive alternative for pan-sharpening of multispectral images. The primary concern in the variational model based pansharpening mechanism is to construct an objective energy functional with efficient and reliable prior terms. Further, the pan-sharpened outcome is obtained using the energy functional minimization approach. The HRMS images are produced with high-spatial and spectral qualities by combining the mechanisms of various PS methods into a single variational framework.

The first variational pan-sharpening method named as pan plus multispectral images (P+XS) was proposed by Ballester et al. (Ballester et al., 2006). The energy function is composed of three prior terms, two of these prior terms are associated with the remote sensing image formation model. The third term is based on the geometric information contained in the panchromatic image. However, P+XS method results in blurring artifacts. The formulation of P+XS has motivated the development of many variational pan-

sharpening methods. An alternate variational pansharpening method (AVWP) (Möller et al., 2012) is presented, in this scheme wavelet fusion and P+XS method are combined with a new spectral consistency term. Although, the AVWP method produce blocky artifacts in the pan-sharpened image. A non-local regularization term based on self-similarity principle applied to PAN image is used for pansharpening in a novel variational model (NVM) (Duran et al., 2014). In this approach, two fidelity terms are used to establish the relation between source images and the HRMS image. In (Lotfi and Ghassemian, 2018a), a pansharpening model is proposed based on the decomposition of source images into the constituent cartoon and texture components. These methods suffer more degradation in reconstructed image quality.

Sparse optimization based continuous modelling is proposed for pansharpening of multispectral images in (Deng et al., 2018). In this work, the correlation between the MS image bands is enforced by a Toeplitz sparse prior term. The alternating direction method of multipliers (ADMM) based operator splitting framework is used to solve the formulated variational model. In discrete formulation paradigm, pan-sharpening methods have been proposed based on total variation (TV) as a regularizer (Palsson et al., 2014; He et al., 2014). Recently, dynamic gradient based sparsity is exploited for geometric consistency and the method is named as dynamic TV (DTV) (Chen et al., 2014). To instigate the feature of sparse representation and to enhance the fidelity of fused image, l_1 based regularized term is used for pan-sharpening in (Chen et al., 2018). Most of these variational methods are able to mitigate the spectral distortion. However, they induce spatial artifacts like blocks and blurring in the resultant images. In addition to spectral enhancement, a holistic model is required to preserve the essential spatial information as well in the fused image.

The objective of the pan-sharpening process is to maintain a balanced trade-off between the spatial and spectral information in the fused image. The variational model-based methods treat the fusion process as an ill-posed inverse optimization problem and construct the energy function based on the high-resolution PAN image, the low-resolution MS image, and the ideal fused image. In this chapter, a new variational PS method is proposed to inject the most vital spatial features of PAN image into HRMS image, while preserving the essential spectral details. The main contributions of the proposed variational model for pan-sharpening are listed as follows:

- A new total generalized variation (TGV) based prior term is proposed to precisely inject the geometric features of PAN image like edges and texture information into the pan-sharpened image. Further, TGV reduces the artifacts and preserves the higher-order smoothness in the fused image.

- By exploiting the correlation between MS image bands, a new spectral distortion minimization term is designed.
- A data fidelity term is adapted from the image formation model to efficiently inject the spectral details from the source MS image to the pan-sharpened image. According to the image formation model, the LRMS image is considered as a decimated and blurred version of the HRMS image. Under these baseline assumptions, a data generative term is formulated.
- Based on the alternating direction method of multipliers (ADMM), an efficient operator splitting framework is formulated to solve the proposed model.

6.1 Preliminaries

This section presents a necessary mathematical framework of total generalized variation (TGV) and spectral angle mapper (SAM) concepts.

6.1.1 Total Generalized Variation (TGV)

Total Variation (TV) has been extensively used as a regularizer in image processing applications such as denoising, restoration, and reconstruction. Since TV deals with only first-order derivatives, it leads to undesirable blocky and oil painting artifacts in the resultant image. Total Generalized Variation (TGV) is a generalized version of TV and possesses the higher-order derivatives of greater than or equal to two. TGV better preserves higher-order smoothness, edges and eliminates the artifacts in the reconstructed image (Bredies et al., 2010).

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, $C_c^k(\Omega, \text{Sym}^k(\mathbb{R}^d))$ be the space of compactly supported symmetric tensor fields and $\kappa = (\kappa_0, \kappa_1, \dots, \kappa_{k-1}) > 0$ is a vector of fixed positive real valued parameters. Using these, TGV of order k is defined as:

$$\text{TGV}_{\kappa}^k(u) = \sup \left\{ \int_{\Omega} u \operatorname{div}^k v \, dx \mid v \in C_c^k(\Omega, \text{Sym}^k(\mathbb{R}^d)), \right. \\ \left. \left\| \operatorname{div}^l v \right\|_{\infty} \leq \kappa_l, \quad l = 0, \dots, k-1 \right\}, \quad (6.1)$$

where, $\text{Sym}^k(\mathbb{R}^d)$ is assumed as the space of symmetric tensors on \mathbb{R}^d and v are defined as bounded vector fields. TGV_{κ}^k is referred as total generalized bounded variation of order k with a weight vector $\kappa \in \mathbb{R}^k$. For $k=1$, $\kappa_0 = 1$ the seminorm TGV_{κ}^k coincides with the bounded variation seminorm.

The space bounded generalized variation (BGV) can be defined as

$$\begin{aligned} \text{BGV}_{\kappa}^k(\Omega) &= \left\{ u \in L^1(\Omega) \mid \text{TGV}_{\kappa}^k(u) < \infty \right\}, \\ \|u\|_{\text{BGV}_{\kappa}^k} &= \|u\|_1 + \text{TGV}_{\kappa}^k(u). \end{aligned} \quad (6.2)$$

BGV spans the set of functions of order k with a weight vector κ , which are generalized version of bounded variations. TGV_{κ}^k is a seminorm on the normed space BGV_{κ}^k , and the space $\text{BGV}_{\kappa}^k(\Omega)$ is independent of κ . For $k = 2$, $\text{Sym}^2(\mathbb{R}^d)$ is the space $\mathbb{S}^{d \times d}$ that spans all symmetric $d \times d$ matrices and models a set of bilinear forms which are usually symmetric. Specifically, the second-order TGV (with $k=2$) can be expressed as:

$$\text{TGV}_{\kappa}^2(u) = \sup \left\{ \int_{\Omega} u \operatorname{div}^2 w \, dx \mid w \in C_c^2(\Omega, \mathbb{S}^{d \times d}), \|w\|_{\infty} \leq \kappa_0, \|\operatorname{div} w\|_{\infty} \leq \kappa_1 \right\}, \quad (6.3)$$

where, the divergences can be defined as:

$$(\operatorname{div} w)_h = \sum_{j=1}^d \frac{\partial w_{hj}}{\partial x_j}, \quad 1 \leq h \leq d, \quad (6.4)$$

and

$$\operatorname{div}^2 w = \sum_{h,j=1}^d \frac{\partial^2 w_{hj}}{\partial x_h \partial x_j}. \quad (6.5)$$

TGV is a convex functional and the polynomials of the order less than $k-1$. In such circumstances the value of the semi-norm TGV_{κ}^k is zero.

6.1.2 Spectral Angle Mapper (SAM)

Given two spectral vectors, in which $\mathbf{v} = \{v_1, v_2, \dots, v_N\}$ be the pixel vector of multi-spectral bands associated with an original image and $\hat{\mathbf{v}} = \{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_N\}$ be the pixel vector of the corresponding fused bands, respectively. Let N be the number of bands present in the image. The spectral angle mapper, SAM is determined as the spectral angle between the two vectors (Alparone et al., 2015) as

$$\text{SAM}(\mathbf{v}, \hat{\mathbf{v}}) = \arccos \left(\frac{\langle \mathbf{v}, \hat{\mathbf{v}} \rangle}{\|\mathbf{v}\|_2 \|\hat{\mathbf{v}}\|_2} \right) \quad (6.6)$$

SAM is generally estimated on performing average operation over the entire image. If the two images used for comparison are spectrally equal then the SAM value is zero, this optimal value is true in an idealistic scenario and it indicates the absence of spectral distortion, however, there exists a possibility of radiometric distortion.

6.2 Proposed Variational Model

The proposed variational pan-sharpening model consists of three terms to address the following objectives: (i) To preserve the spectral information, (ii) To enhance the spatial details and (iii) To reduce the spectral distortion. The terms are designed to attain a pan-sharpened image having desired spatial and spectral features with reference to the given PAN and MS images. The First term is adapted from the conventional remote sensing image formation model, which can be operated as a data synthesizing fidelity term. The fidelity term enforces spectral information preservation. To retain the requisite geometric structures and to impart the spatial information in a comprehensive manner from PAN image to HRMS image, TGV based spatial details extracting term is designed. To reduce the spectral distortion in the fused image, the correlation information between MS bands is exploited. The spectral distortion index namely, SAM is utilized to design an inter-band correlation term.

6.2.1 Data Synthesizing Fidelity Term

The remote sensing image formation model is presented abstractly that prepares underlying theme for the data synthesizing fidelity term:

Let $P \in \mathbb{R}^{M \times N}$, be a PAN image and $X_i, (i = 1, 2..B)$ be the LRMS image with B bands with each band of size $m \times n$ where, $m = M/4$ and $n = N/4$. The spatial resolution of PAN image is four times that of the MS image. The pan-sharpened image obtained post fusion process is 'B' band HRMS image ($\hat{X} = (\hat{X}_1, \hat{X}_2, \dots, \hat{X}_B)$) maintaining the spatial resolution of the PAN image with the size of $M \times N$ for each band. Let $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_B)$ be the up-sampled LRMS image and has the same size as PAN image, P .

The source images P , X , and the pan-sharpened image \hat{X} are represented as vectors for computational purposes. Following the well-established remote sensing image formation model, the low-resolution MS image bands can be treated as decimated and

blurred variants of the corresponding HRMS image bands.

$$\mathbf{X}_i = \mathbf{G}\widehat{\mathbf{X}}_i + \mathbf{v}_i \quad i = 1, 2, \dots, B. \quad (6.7)$$

Where, $\mathbf{G} = \gamma_i \mathbf{H}$. Let γ_i is the blur filter for i^{th} band, and \mathbf{H} is the decimation matrix. The blur filter for each band is realized based on the modulation transfer function (MTF) of the band with its respective cutoff frequency. The decimation matrix, $\mathbf{H} = \frac{1}{16} \cdot \mathbf{I}_4 \otimes ((\mathbf{I}_n \otimes \mathbf{1}_{4 \times 1}^T) \otimes (\mathbf{I}_m \otimes \mathbf{1}_{4 \times 1}^T))$ is of $(4mn \times 4MN)$ size to perform the down-sampling operation and \mathbf{v}_i is an additive Gaussian noise matrix for the i^{th} band. The operator \otimes denotes Kronecker product. \mathbf{I}_4 is an identity matrix of size 4×4 and $\mathbf{1}_{4 \times 1}$ is a 4×1 vector with all entries as unity.

To preserve the spectral details of the LRMS image, the data synthesizing fidelity term is formulated as:

$$J_1(\widehat{\mathbf{X}}) = \frac{1}{2} \sum_{i=1}^B \|\mathbf{X}_i - \mathbf{G}\widehat{\mathbf{X}}_i\|_2^2 \quad (6.8)$$

Where B denotes the number of constituent spectral bands of LRMS/HRMS images.

6.2.2 Spatial Details Preserving Term

To inject the vital geometric features like edges and rich texture information of the PAN image into the pansharpened image, TGV is adapted as a regularizer in the proposed algorithm.

The TGV has been used in the reconstruction of images belonging to different modalities like medical images, fusion of visible and infrared images etc. In addition to the features that are recovered by the TV, the regularization using TGV predominantly refrains the typical artifacts like oil painting effects.

TGV² (Bredies and Valkonen, 2011) can be reformulated as:

$$\text{TGV}_{\kappa}^2(\mathbf{u}) = \min_{\mathbf{u} \in \text{BGV}_{\kappa}^2(\Omega), \mathbf{r} \in \text{BD}(\Omega)} \kappa_1 \int_{\Omega} |\nabla \mathbf{u} - \mathbf{r}| + \kappa_0 \int_{\Omega} |\boldsymbol{\varepsilon}(\mathbf{r})|, \quad (6.9)$$

With $\text{BGV} = \{\mathbf{u} \in \Omega / \text{TGV}_{\kappa}^k(\mathbf{u}) < \infty\}$ is called the space of bounded generalized variation of order k , with weight κ . Where, \mathbf{r} are the vector fields of bounded deformation, i.e. their distributional symmetrized derivative $\boldsymbol{\varepsilon}(\mathbf{r})$ is a measure and $\nabla \mathbf{u}$ represents the gradient of vector \mathbf{u} .

In order to solve the energy function efficiently using the minimizer called Alter-

nating direction method of multiplier (ADMM), the discretized version of TGV² is developed in (Bredies et al., 2010).

$$\text{TGV}^2(\mathbf{u}) = \min_{\mathbf{r}} \kappa_1 \|\nabla \mathbf{u} - \mathbf{r}\|_1 + \alpha_0 \|\boldsymbol{\varepsilon}(\mathbf{r})\|_1 \quad (6.10)$$

Here, $\nabla \mathbf{u} = \begin{bmatrix} \partial_x \mathbf{u} \\ \partial_y \mathbf{u} \end{bmatrix}$ and $\boldsymbol{\varepsilon}(\mathbf{r}) = \frac{1}{2}[\nabla \mathbf{r} + \nabla \mathbf{r}^T]$ denotes symmetrized derivative. In the proposed method, \mathbf{r} represents the processed image. In order to efficiently solve Eq.6.10, the directional derivative $\nabla \mathbf{u}$ is approximated with $D\mathbf{u}$. Where $D = (D_1; D_2)$. Hence,

$$\boldsymbol{\varepsilon}(\mathbf{r}) = \begin{bmatrix} D_1 \mathbf{r}_1 & \frac{1}{2}(D_2 \mathbf{r}_1 + D_1 \mathbf{r}_2) \\ \frac{1}{2}(D_2 \mathbf{r}_1 + D_1 \mathbf{r}_2) & D_2 \mathbf{r}_2 \end{bmatrix} \quad (6.11)$$

where the finite forward differences in x and y directions are measured by the circulant matrices D_1 and D_2 , respectively.. Based on the reformulation of TGV, the spatial difference between HRMS image and PAN image is expressed as:

$$J_2(\widehat{\mathbf{X}}) = \kappa_1 \sum_{i=1}^B \|\mathbf{D}(\widehat{\mathbf{X}}_i - \mathbf{P}) - \mathbf{r}\|_1 + \alpha_0 \|\boldsymbol{\varepsilon}(\mathbf{r})\|_1 \quad (6.12)$$

The performance eminence of TGV mechanism over TV in preserving textures, edges and reducing staircase effects has been presented experimentally for the different imaging modalities in (Knoll et al., 2011; Guo et al., 2014a).

6.2.3 Inter-Band Correlation Preserving Term

The prime objective behind acquiring pansharpened image is to make it pragmatic for applications such as classification, recognition, and detection. Ideally, the distortion index called Spectral Angle Mapper (SAM) value should be close to zero so as to attain the optimum correlation between spectral bands of MS image. In vector analogy, SAM is zero if the two spectral vectors corresponding to the images $\widehat{\mathbf{X}}$ and $\widetilde{\mathbf{X}}$ are parallel.

Under this assumption; to preserve the correlation between the MS image bands, a hypothesis is presented in the proposed method. The ratio of any two different spectral bands of HRMS image ($\widehat{\mathbf{X}}$) should be equal to that of up-sampled MS bands ($\widetilde{\mathbf{X}}$),

$$\frac{\widehat{\mathbf{X}}_i}{\widehat{\mathbf{X}}_j} = \frac{\widetilde{\mathbf{X}}_i}{\widetilde{\mathbf{X}}_j}, \quad 1 \leq i, j \leq B., \quad i \neq j. \quad (6.13)$$

This constraint can be formulated as:

$$\widehat{X}_i \widetilde{X}_j - \widehat{X}_j \widetilde{X}_i = \mathbf{0} \quad i, j = 1, 2 \dots B., \quad i \neq j. \quad (6.14)$$

An energy term, $J_3(\widehat{X})$ is defined to preserve the desirable correlation information between multispectral bands. Further, $J_3(\widehat{X})$ also reduces spectral distortion.

$$J_3(\widehat{X}) = \sum_{i=1}^B \sum_{j=1}^B \|\widehat{X}_i \widetilde{X}_j - \widehat{X}_j \widetilde{X}_i\|_2^2. \quad (6.15)$$

On combining terms, $J_1(\widehat{X})$, $J_2(\widehat{X})$ and $J_3(\widehat{X})$, the cost function for the proposed pan-sharpening algorithm is formulated as:

$$\begin{aligned} J(\widehat{X}) = & \frac{1}{2} \sum_{i=1}^B \|X_i - G\widehat{X}_i\|_2^2 + \kappa_1 \sum_{k=1}^B \|D(\widehat{X}_i - P) - r\|_1 + \kappa_0 \|\varepsilon(r)\|_1 + \\ & \frac{\lambda}{2} \sum_{i=1}^B \sum_{j=i}^B \|\widehat{X}_i \widetilde{X}_j - \widehat{X}_j \widetilde{X}_i\|_2^2 \end{aligned} \quad (6.16)$$

Where, λ is a regularization parameter and its value determines the relative contribution of inter-band correlation term.

6.2.4 Optimization Method

The proposed formulation given in Eq.6.10 can be efficiently solved by alternating direction method of multipliers (ADMM) (Gabay and Mercier, 1975). ADMM solves the linearly constrained separable convex function of the form

$$\min [\theta_1(x_1) + \theta_2(x_2)] \quad \text{subject to} \quad A_1 x_1 + A_2 x_2 = b \quad (6.17)$$

With $x_1 \in \phi_1$ and $x_2 \in \phi_2$. $\theta_1 : \mathfrak{R}^{n_1} \rightarrow \mathfrak{R}$ and $\theta_2 : \mathfrak{R}^{n_2} \rightarrow \mathfrak{R}$ are closed proper convex functions; $\phi_1 \subset \mathfrak{R}^{n_1}$ and $\phi_2 \subset \mathfrak{R}^{n_2}$ are closed convex sets; $A_1 \in \mathfrak{R}^{l \times n_1}$ and $A_2 \in \mathfrak{R}^{l \times n_2}$ are given matrices and $b \in \mathfrak{R}^l$ is a given vector. The Lagrangian is defined as:

$L(x_1, x_2; t) = \theta_1(x_1) + \theta_2(x_2) + \frac{\beta}{2} \|A_1 x_1 + A_2 x_2 - b - t\|_2^2$, In that t is the scaled Lagrange multiplier and β is a positive parameter. ADMM solves Eq.6.17 in an iterative manner, the process initialize with $x_2^0 = 0$ and $t^0 = 0$ as follows:

$$x_1^{k+1} = \operatorname{argmin}_{x_1} L(x_1, x_2^k; t^k)$$

$$\begin{aligned} \mathbf{x}_2^{k+1} &= \operatorname{argmin}_{\mathbf{x}_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2; \mathbf{t}^k) \\ \mathbf{t}^{k+1} &= \mathbf{t}^k + \beta(\mathbf{b} - (\mathbf{A}_1 \mathbf{x}_1^{k+1} + \mathbf{A}_2 \mathbf{x}_2^{k+1})) \end{aligned}$$

The pansharpening model described in Eq.6.16 consists of a non-smooth l_1 term, and can be amended by introducing auxiliary variables as

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad \mathbf{s} = \begin{bmatrix} s_1 & s_3 \\ s_3 & s_2 \end{bmatrix}$$

The model presented in Eq.6.16 is solved for each $\widehat{\mathbf{X}}_i$ value and can be presented as

$$\min_{\widehat{\mathbf{X}}_i, \mathbf{P}, \mathbf{v}, \mathbf{s}} \frac{1}{2} \|\mathbf{X}_i - \mathbf{G}\widehat{\mathbf{X}}_i\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^B \sum_{j=i}^B \|\widehat{\mathbf{X}}_i \widetilde{\mathbf{X}}_j - \widehat{\mathbf{X}}_j \widetilde{\mathbf{X}}_i\|_2^2 + \kappa_1 \|\mathbf{v}_i\|_1 + \kappa_0 \|\mathbf{s}\|_1 \quad (6.18)$$

with the change of variables $\mathbf{v}_i = \mathbf{D}(\widehat{\mathbf{X}}_i - \mathbf{P}) - \mathbf{r}$, $\mathbf{s} = \boldsymbol{\varepsilon}(\mathbf{r})$. The terms $\|\mathbf{v}\|_1$ ($\|\mathbf{s}\|_1$) is the sum of l_2 -norms (Frobenius norm) of all 2×1 vectors (2×2 matrices).

The application of ADMM (Guo et al., 2014b) results in the following parameters estimation:

$$\mathbf{v}^{n+1} = \operatorname{argmin}_{\mathbf{v}} \|\mathbf{v}\|_1 + \frac{\mu_1}{2} \|\mathbf{v} - (\mathbf{D}(\mathbf{X} - \mathbf{P})^n - \mathbf{r}^n) - (\widetilde{\mathbf{v}}^n)\|_2^2 \quad (6.19)$$

$$\mathbf{s}^{n+1} = \operatorname{argmin}_{\mathbf{s}} \|\mathbf{s}\|_1 + \frac{\mu_2}{2} \|\mathbf{s} - \boldsymbol{\varepsilon}(\mathbf{r}^n) - (\widetilde{\mathbf{s}}^n)\|_2^2 \quad (6.20)$$

$$\begin{aligned} (\widehat{\mathbf{X}}_i^{n+1}, \mathbf{P}^{n+1}) &= \operatorname{argmin}_{\widehat{\mathbf{X}}_i, \mathbf{P}} \frac{1}{2} \|\mathbf{X}_i - \mathbf{G}\widehat{\mathbf{X}}_i\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^B \sum_{j=i}^B \|\widehat{\mathbf{X}}_i \widetilde{\mathbf{X}}_j - \widehat{\mathbf{X}}_j \widetilde{\mathbf{X}}_i\|_2^2 \\ &\quad + \kappa_1 \frac{\mu_1}{2} \|\mathbf{v}^{n+1} - (\mathbf{D}(\widehat{\mathbf{X}}_i - \mathbf{P}) - \mathbf{r}) - (\widetilde{\mathbf{v}}^n)\|_2^2 \\ &\quad + \kappa_0 \frac{\mu_2}{2} \|\mathbf{s}^{n+1} - \boldsymbol{\varepsilon}(\mathbf{r}) - (\widetilde{\mathbf{s}}^n)\|_2^2 \end{aligned} \quad (6.21)$$

$$\widetilde{\mathbf{v}}^{n+1} = \widetilde{\mathbf{v}}^n + \mu(\mathbf{D}(\mathbf{X}_i - \mathbf{P})^{n+1} - \mathbf{v}^{n+1}) \quad (6.22)$$

$$\widetilde{\mathbf{s}}^{n+1} = \widetilde{\mathbf{s}}^n + \mu(\boldsymbol{\varepsilon}(\mathbf{r}^{n+1}) - \mathbf{s}^{n+1}) \quad (6.23)$$

During each cycle of iterations involving Eq.6.19 to Eq.6.23, Eq.6.21 is a differentiable

optimization problem.

The v and s subproblems specified in Eq.6.22 and Eq.6.23 can be solved using shrinkage operators as:

v -subproblem:

$$v^{n+1} = \text{shrink}_1(D(X_i - P)^n - r^n + \tilde{v}^n, \frac{1}{\mu_1}) \quad (6.24)$$

Where,

$$\text{shrink}_a(a, \frac{1}{\mu_1}) = \frac{a}{\|a\|_2} \cdot \max(\|a\|_2 - \frac{1}{\mu_1}, 0) \quad (6.25)$$

s -subproblem:

$$s^{n+1} = \text{shrink}_b(\varepsilon(r^n) + \tilde{s}^n, \frac{1}{\mu_2}) \quad (6.26)$$

Where,

$$\text{shrink}_b(b, \frac{1}{\mu_2}) = \frac{b}{\|b\|_F} \cdot \max(\|b\|_F - \frac{1}{\mu_2}, 0) \quad (6.27)$$

To obtain the solution of the differentiable part and the convergence analysis of ADMM algorithm with comprehensive investigation, details are reported in (Guo et al., 2014b). The solution of the formulated constrained optimization problem yields the requisite high-resolution MS image. The parameters selected to implement the ADMM algorithm and the convergence details are discussed in the following section.

6.3 Results and Analysis

In this section, the developed variational model for pan-sharpening is compared with the following state-of-the-art methods at reduced-scale and full-scale: IHS(Tu et al., 2001) as the classical CS based method; AWLP (Otazu et al., 2005), MTF-GLP (Aiazzi et al., 2006) as the classical MRA based methods; SR-LD (Li et al., 2013b), SR-CD (Ayas et al., 2018) as the sparse representation based methods; AVWP (Möller et al., 2012), DTV (Chen et al., 2014) and V-L1 (Chen et al., 2018) as the representative variational methods.

To validate the effectiveness of the proposed method, both the visual and quantitative assessments are performed on datasets namely IKONOS, Pléiades, and QuickBird sensors. The resolution ratio between MS and PAN images is four for all the experimental datasets. All the methods are implemented in MATLAB R2013a, run on a personal computer with intel CPU @3.10-GHz and 8-GB RAM.

6.3.1 Parameters Selection

Implementation of the proposed algorithm involves several parameters, which are to be selected carefully. The regularization parameters for TGV are selected so as to maintain an appropriate balance with the data synthesizing fidelity term and inter-band correlation term. The image features are lost if more priority is given to the TGV term, whereas less priority results in residual noise in the fused image. The crucial regularization parameter λ is tuned to give the best adaptive outcome between SAM and ERGAS. The value of λ influences the spectral quality of the fused image. For the reduced-scale dataset the performance characteristics of SAM and ERGAS with respect to variation in parameter λ are shown in Fig.6.1. Fig.6.2 presents the variation of Q4 and full-scale metrics D_λ and D_s with parameter λ .

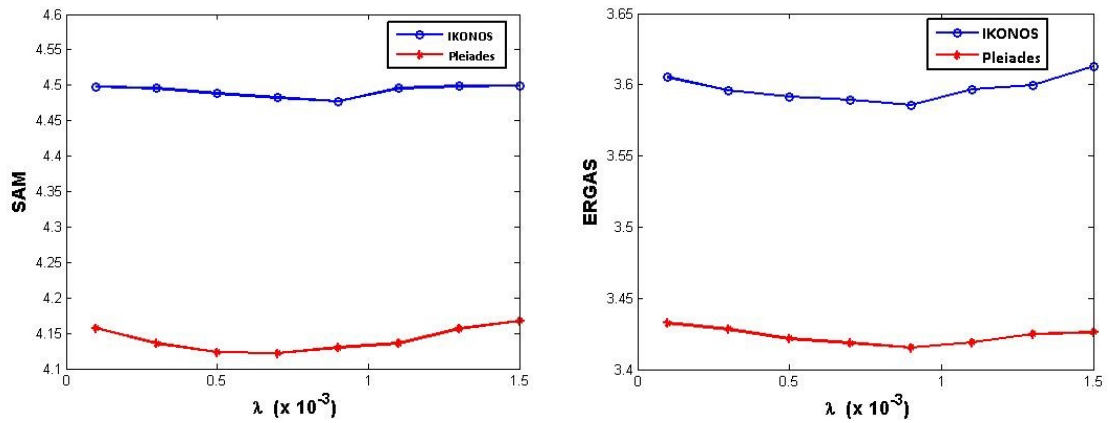


Figure 6.1: Performance characteristics of SAM and ERGAS with regularization parameter λ variation (a) SAM Vs. λ (b) ERGAS Vs. λ

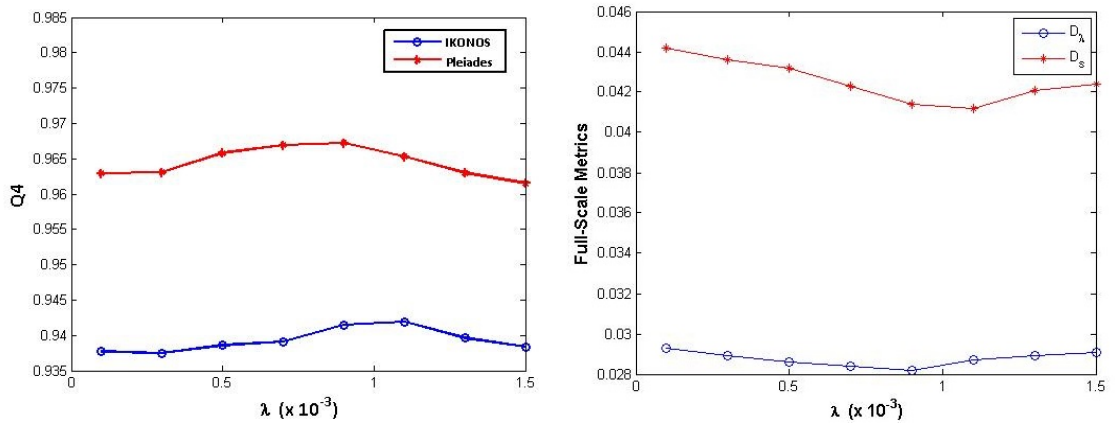


Figure 6.2: Performance characteristics of Q4, full-scale metrics with regularization parameter λ variation (a) Q4 Vs. λ (b) D_λ , D_s Vs. λ

Based on these observations λ value is tuned to 0.9×10^{-3} for all the experiments. The quantitative results do not show any noticeable change for the variations in α_0 and α_1 . Hence, the other regularization parameters are selected as $\alpha_0 = 10^{-2}$, $\alpha_1 = 10^{-3}$. The remaining parameters are set as $\mu_1 = 10^{-3}$, $\mu_2 = 10^{-5}$, since these parameters do not effect the spatial and spectral indices much. To maintain the balance between performance and computational complexity, the number of iterations for the algorithm is selected as 25.

6.3.2 Reduced-Resolution Assessment

Since the reference high-resolution MS image is not available, Wald's protocol (Wald et al., 1997) is considered to evaluate the quality of the fused image. The source images are degraded by a factor of four, which is equal to the resolution ratio between PAN and MS images and the original MS image is treated as a reference image. Several indices have been developed for the quantitative assessment of spatial and spectral distortions of the pansharpened image. In this work, Root mean square error (RMSE), spectral angle mapper (SAM), Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), correlation coefficient (CC) and Universal Image Quality Index (UIQI) or Q-index (Q4) are used as quality metrics.

The PAN and MS images produced by IKONOS dataset are of 1-m and 4-m resolutions, respectively. The size of images for experimentation is considered as 256×256 pixels. The visual outcomes of different methods for the IKONOS dataset at reduced-scale are presented in Fig.6.3. The corresponding quality metrics are reported in Table 6.1. Fig.6.3(a) and 6.3(b) show the PAN and MS images used for the experimental purpose. Fig.6.3(c) presents the up-sampled MS image and termed as EXP. From the pansharpened images, it is observed that the IHS method yields an outcome with spectral distortion in the form of inconsistent color. The AVWP method results in spatial and spectral distortions. The SR-CD, MTF-GLP and AWLP methods display fine spatial details. The precise observation of the outcomes of variational methods indicates a slight color distortion. The outcomes of DTV and V-L1 methods are moderately good as these schemes are able to preserve edges. The comprehensive perception manifests that the proposed method preserves more spatial details than the other considered methods. For improved visualization of spatial details, a part of the image is zoomed (encircled in a red box) and shown at the bottom left corner for every image. From Table 6.1, comparative analysis infer that the proposed method outperforms over all the reported methods on all the quality metrics considered.

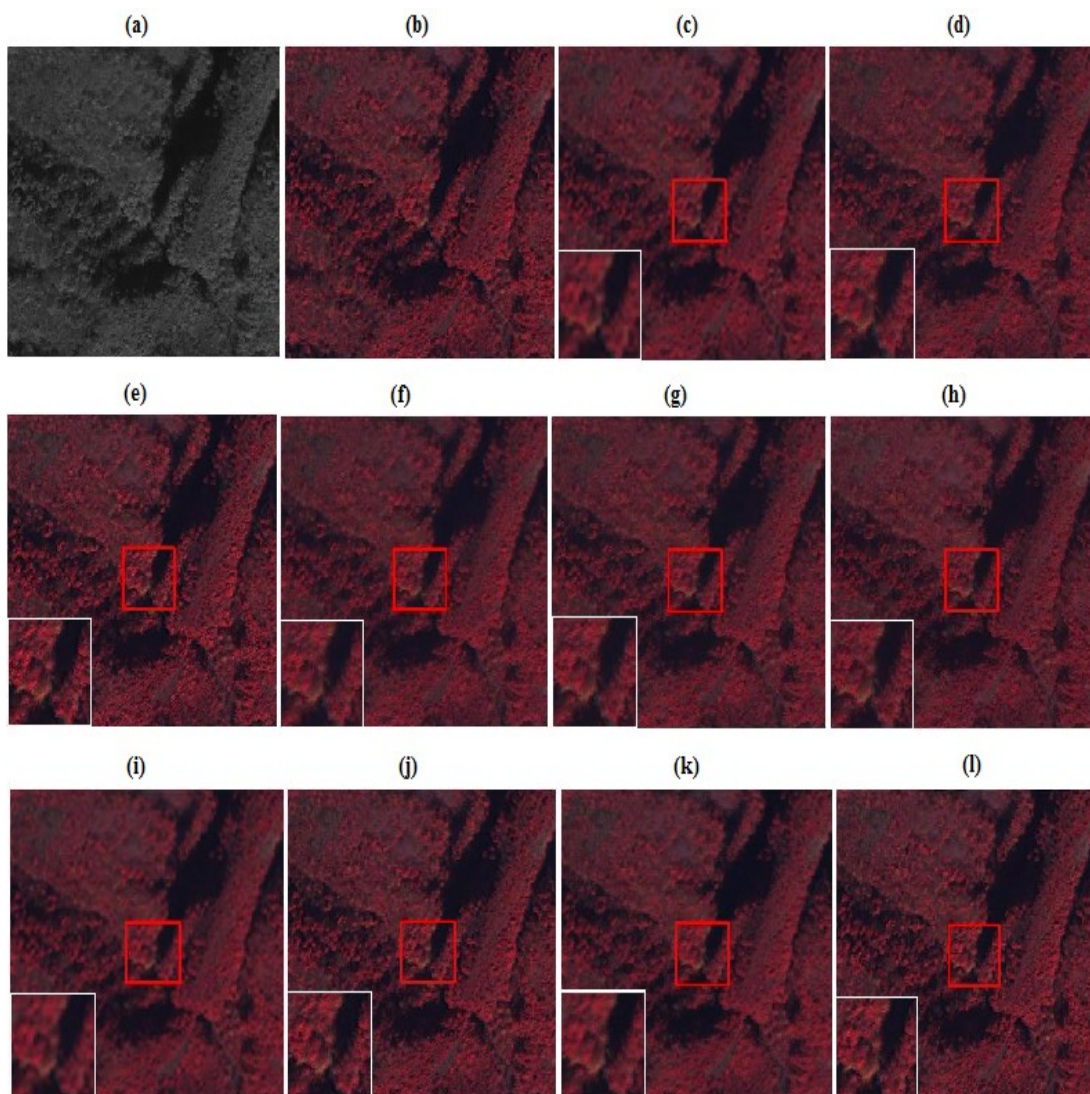


Figure 6.3: IKONOS dataset and experimental results by different methods (a) PAN image (b) Reference MS image (c) EXP (up-sampled MS image) (d) IHS (e) AWLP (f) MTF-GLP (g) AVWP (h) SR-LD (i) DTV (j) V-L1 (k) SR-CD (l) Proposed method

The Pléiades dataset was collected by an aerial platform of an urban area of Toulouse (France). The resolution of the four MS bands is 60 cm and the corresponding high resolution PAN image was simulated from the available green and red channels. The size of the images used for the implementation of pan-sharpening methods at the reduced resolution is 256×256 pixels.

Fig.6.4(a) shows the MS image covering an urban area, resampled to the size of the PAN image. Fig.6.4(b) shows the corresponding PAN image. The visual outcomes of EXP, IHS, AWLP, MTF-GLP, AVWP, SR-LD, DTV, V-L1, SR-CD, and the proposed method are shown in Fig.6.4(c)-(l), respectively. The IHS method suffers from spectral

Table 6.1: Reduced-resolution quality Metrics for IKONOS data set

Method	Q4	SAM	RMSE	ERGAS	CC
EXP	0.8251	5.6492	16.3371	5.1386	0.7917
IHS	0.8627	5.3537	15.6392	4.4137	0.8458
AWLP	0.9158	5.0845	15.1156	3.6324	0.9214
MTF-GLP	0.9231	5.1427	15.1093	3.5937	0.9276
AVWP	0.8885	5.2175	15.9173	4.5273	0.8721
SR-LD	0.8754	5.5146	16.1183	4.6259	0.8735
DTV	0.9237	4.6842	14.4131	<u>3.6521</u>	0.9283
V-L1	0.9314	4.7163	<u>14.3275</u>	3.7128	0.9351
SR-CD	<u>0.9342</u>	<u>4.5739</u>	14.9748	3.8862	<u>0.9363</u>
Proposed	0.9418	4.4753	14.1126	3.5861	0.9389

distortion in the regions containing grass. It can be seen from Fig.6.4(g) that the AVWP method is not good in spatial resolution. The AWLP method and SR based methods better preserve the color in various regions of the fused image. It is observed from the results that AWLP, MTF-GLP, DTV, and SR-CD produce comprehensive visual quality. The proposed method is superior in reducing the color distortion and retaining the sharp spatial details than the other reported methods. Besides, the quantitative results for the Pléiades dataset are presented in Table 6.2. Quantitative values of the metrics validate the efficiency of the proposed method.

Table 6.2: Reduced-resolution quality Metrics for Pléiades data set

Method	Q4	SAM	RMSE	ERGAS	CC
EXP	0.7841	4.6853	13.7956	5.9562	0.8257
IHS	0.8495	4.9357	13.4531	5.1048	0.8573
AWLP	0.9413	4.4138	12.4176	3.5182	0.9587
MTF-GLP	0.9478	4.1735	12.3362	3.3174	0.9564
AVWP	0.9247	5.1123	12.4175	3.7293	0.9381
SR-LD	0.9483	4.3352	11.9731	3.4832	0.9653
DTV	0.9571	4.2673	10.9318	3.4536	0.9617
V-L1	0.9583	4.2759	10.6572	<u>3.4369</u>	<u>0.9776</u>
SR-CD	<u>0.9654</u>	4.0381	<u>10.4729</u>	3.4371	0.9743
Proposed	0.9673	<u>4.1736</u>	10.2358	3.4152	0.9782

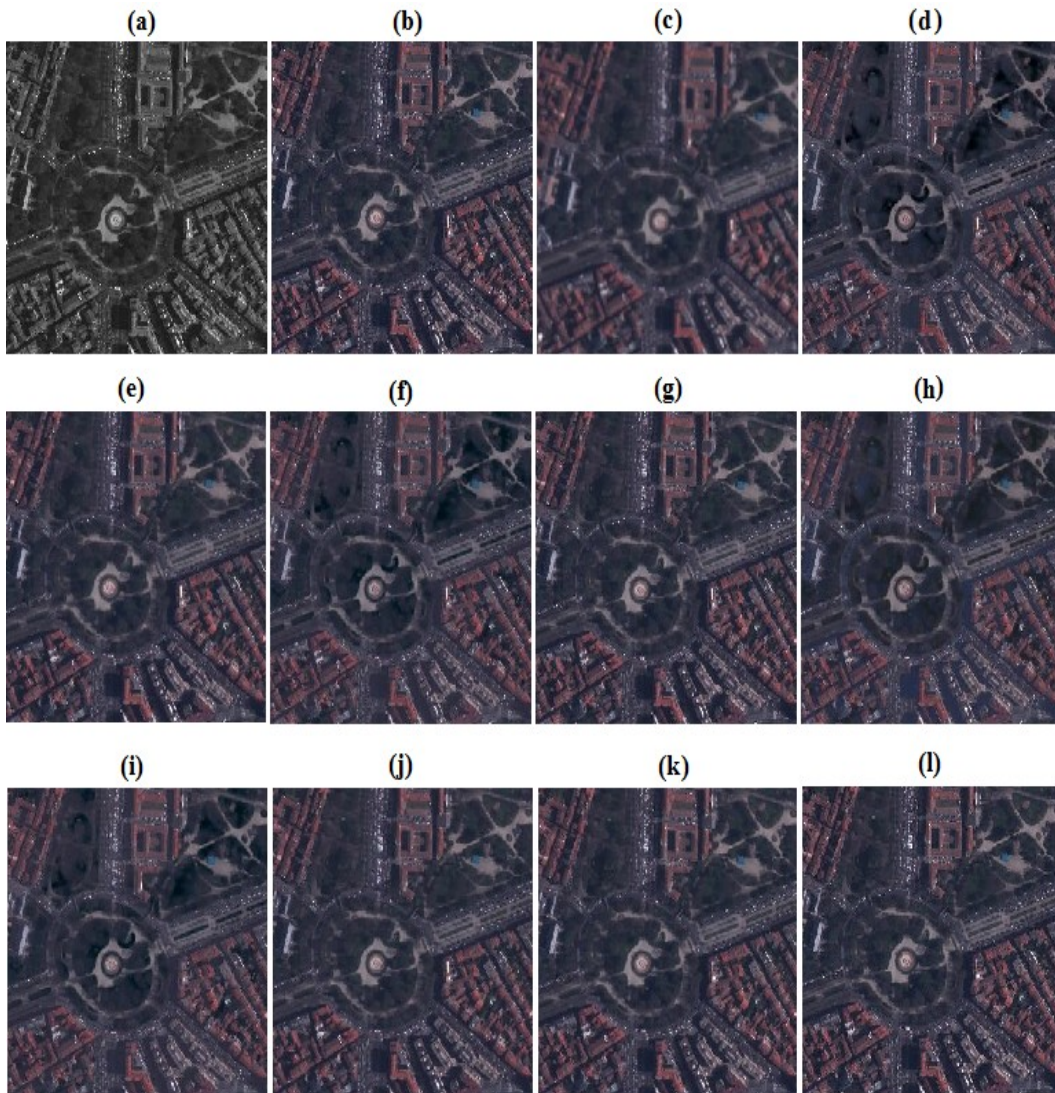


Figure 6.4: Pléiades dataset and experimental results by different methods (a) PAN image (256×256) (b) Reference MS image (c) EXP (up-sampled MS image) (d) IHS (e) AWLP (f) MTF-GLP (g) AVWP (h) SR-LD (i) DTV (j) V-L1 (k) SR-CD (l) Proposed method

6.3.3 Full-Resolution Assessment

QuickBird dataset produces the PAN and 4-band MS images at 0.7-m and 2.8-m spatial resolution, respectively. The QNR protocol (Alparone et al., 2008) is used for the quantitative evaluations. Fig. 6.5 presents pansharpening results at full-resolution for the QuickBird dataset. The estimated value of quality metrics in terms of spectral distortion index D_λ , spatial distortion index D_s and QNR are reported in Table 6.3.

Fig. 6.5(a) and 6.5(b) shows the full-resolution PAN and up-sampled MS images of size 512×512 pixels. The outcomes of methods IHS, AVWP, and V-L1 exhibit

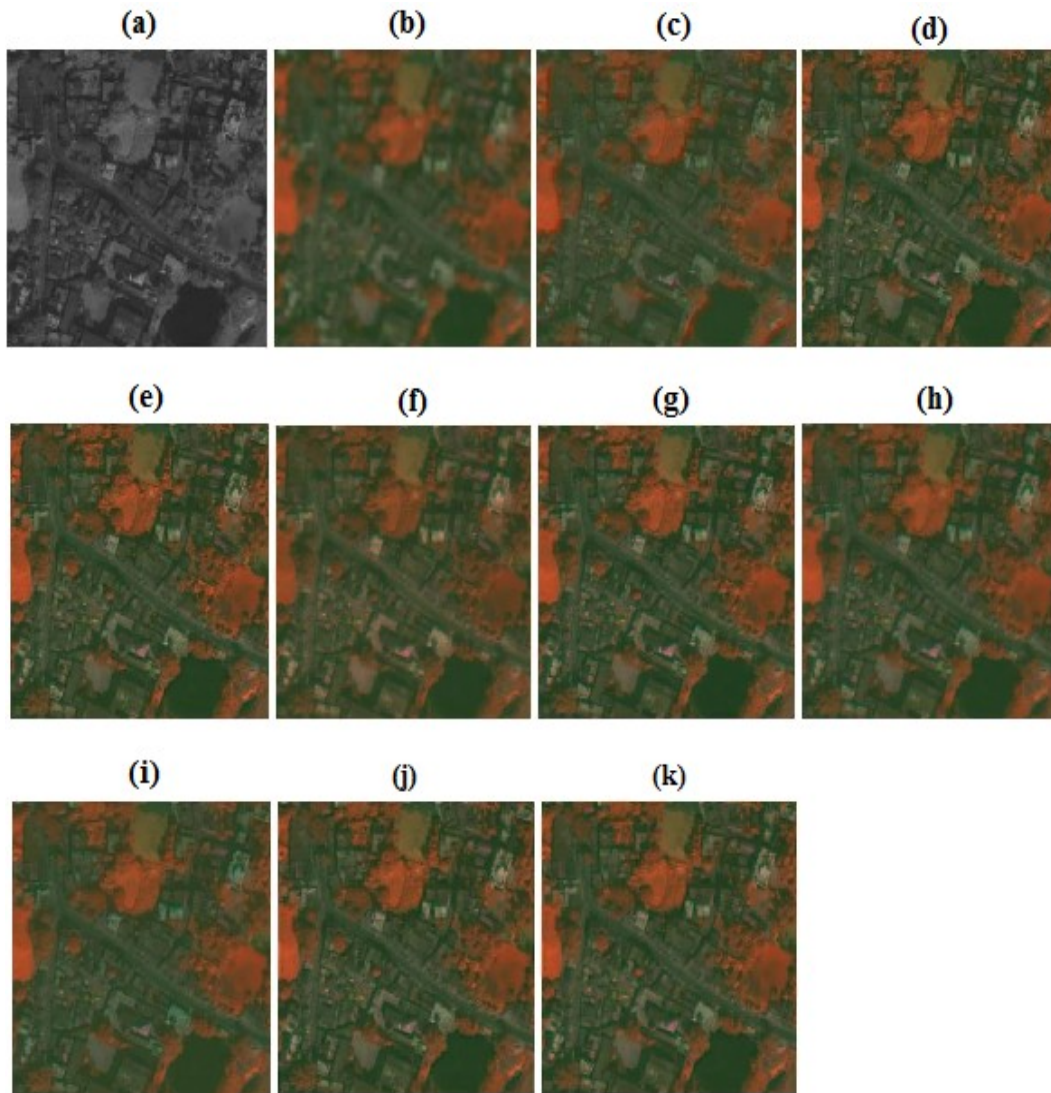


Figure 6.5: QuickBird dataset and experimental results by different methods (a) PAN image (400×400) (b) up-sampled MS image (c) IHS (d) AWLP (e) MTF-GLP (f) AVWP (g) SR-LD (h) DTV (i) V-L1 (j) SR-CD (k) Proposed method

slight color change and blurring artifacts at the red colored soil and highway portion of the image. The conventional methods AWLP and MTF-GLP outcomes preserve the spectral information. The outcome of the DTV method shows tiny blocking artifacts. The pansharpened outcomes of SR-LD and SR-CD methods exhibit acceptable sharpness in preserving spatial details. Moreover, the proposed method yields less spectral distortion and retains sharp edges compared to the other reported methods. The smaller value of D_λ and D_s implies less spectral distortion and spatial distortion, respectively obtained by the proposed pan-sharpening method. The higher value of QNR indicates enhanced global quality of the fused image. The proposed method gives the optimal

values for D_λ , D_s and QNR as reported in Table 6.3. Hence, the overall performance and efficacy of the proposed method is demonstrated by visual outcomes as well as quantitative results.

Table 6.3: Full-resolution quality Metrics for QuickBird dataset

Method	D_λ	D_s	QNR
EXP	0	0.1457	0.8544
IHS	0.0410	0.0576	0.9037
AWLP	0.0406	0.0480	0.9135
MTF-GLP	0.0372	<u>0.0414</u>	0.9229
AVWP	0.0428	0.0507	0.9087
SR-LD	0.0356	0.0513	0.9149
DTV	0.0338	0.0481	0.9197
V-L1	0.0315	0.0489	0.9211
SR-CD	<u>0.0298</u>	0.0453	<u>0.9262</u>
Proposed	0.0283	0.0409	0.9319

6.3.4 Comparative Analysis of Computational Cost

The efficiency of all the considered methods is evaluated in terms of algorithm execution time measured in seconds. The average execution time is measured for the test datasets presented in Fig.6.3- Fig.6.5. The CS and MRA methods are recognized for the minimal possible computational cost. The execution time for all the compared methods is reported in Table 6.4. It can be observed from Table 6.4, that IHS, AWLP, and MTF-GLP methods consume the least execution time among all the reported methods. Since the dictionary training process is laborious, obviously the SR based methods consume exceptionally longer time. The proposed method is not as computationally efficient as some of the the variational methods like AVWP, DTV, and V-L1. However, the proposed method reduces the blocking artifacts and preserves spectral information at the cost of execution time. From the perspective of accuracy and convergence speed, Fig.6.6 presents the relative error versus the iteration count for the proposed method. The relative error is evaluated as

$$\text{relative error} = \frac{\|\hat{\mathbf{X}}^i - \mathbf{X}^r\|}{\|\mathbf{X}^r\|} \quad (6.28)$$

Table 6.4: Average execution time comparison of different methods (in seconds)

Method	Time(sec.)	Method	Time(sec.)
EXP	0	SR-LD	1486.0617
IHS	0.0119	DTV	39.7352
AWLP	0.1935	V-L1	116.2947
MTF-GLP	0.1385	SR-CD	652.3795
AVWP	95.3526	Proposed	218.6472

where, \hat{X}^i is the pan-sharpened image obtained at i^{th} iteration and X^r is the reference MS image. The relative error attains to the least possible value for all the datasets on completing 25 iterations. The visual and quantitative results validate that the proposed method is superior to the reported methods namely, AVWP, DTV, and V-L1 in terms of maintaining a trade-off between spatial and spectral qualities. From the execution time perspective, the proposed method is efficient than the considered SR based methods.

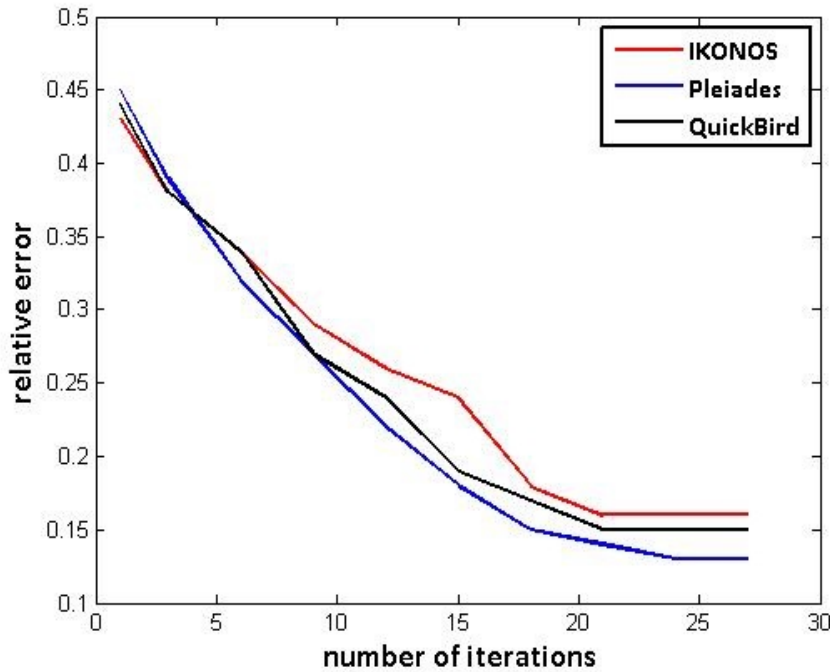


Figure 6.6: Convergence speed of the proposed method for the datasets utilized in experimentation

6.4 Conclusion

In recent years, the variational method emerge as an attractive alternative for the pan-sharpening of multispectral images. This work proposed a variational pan-sharpening model based on total generalized variation (TGV) and inter-band correlation. The conventional total variation based geometric terms result in a staircase effect in the pansharpened image. This work investigates a second-order TGV based spatial difference term to reduce the geometric structural difference between HRMS image and PAN image. Further, the inter-band correlation term, explored from spectral angle mapper (SAM) is used to reduce the spectral distortion. Moreover, an efficient optimization algorithm called ADMM with an operator splitting framework is utilized to solve the proposed pan-sharpening model. For comprehensive performance analysis of the proposed method, the experiments are performed on reduced and full-resolution data. Furthermore, the proposed method is competitive with most of the SR and variational methods and leads to satisfactory results compared to AVWP, DTV, and V-L1 methods. The proposed work can be further extended to reduce the execution time and to exploit more reliable terms to efficiently characterize the relation between source and fused images.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

Remote sensing image data captured by the satellite sensors are having phenomenal importance for applications such as urban planning, land cover classification, change detection, and Google Maps. The hardware constraints limit the sensors in producing images with high-spatial and spectral resolution characteristics. This thesis explores the problem of pan-sharpening (PS), a remote sensing image fusion technique that combines the spatial characteristics of panchromatic (PAN) image and spectral features of multispectral (MS) image bands into a high-resolution multispectral (HRMS) image.

The literature review indicates that the existing pan-sharpening methods are still striving to produce the fused image with an adequate balance between spatial and spectral qualities. The component substitution (CS) methods are global, i.e. these methods do not consider the local discrepancies between source images used for the fusion process. Hence, the spectral mismatch between PAN and MS images results in color distortion in the pan-sharpened image. In most of the multi resolution analysis (MRA) based methods, the mechanism used to extract the spatial details affects the quality of fused outcome. Further, a set of inappropriate coefficients used to inject spatial details into MS bands induces blurring artifacts in the pan-sharpened outcome. The amount of overlap between the image patches, the dictionary construction mechanism, and the size of the patch used in sparse coding process are the critical issues that regulate the performance of sparse representation (SR) based PS techniques.

In this thesis, the pan-sharpening problem is investigated in the perspective of balancing the trade-off between spatial features enhancement and spectral details preservation in the fused image. The proposed work explores SR based methods and variational models in order to overcome the drawbacks of conventional PS methods. Two SR based techniques using multi-scale learned dictionary (MSLD) and dual dictionaries are proposed. The concept of scale-invariance between the PAN image patches at two different resolutions is deployed to extract the spatial details that are missing in MS image bands. A multi-scale learned dictionary is attractive because of its ease in performing the sparse

coding. The experimental results demonstrate the capability of the MSLD and dual dictionary based PS methods towards enhancing the spatial features, thus maintaining the quality spatial details and accompanied a significant reduction in spectral distortion. However, the dictionary construction and subsequent update process, in most of these SR based methods are computationally intensive.

Most of the SR based pan-sharpening methods adapted patch-partition based processing strategy. An increase in the amount of overlapping between adjacent patches increases the smoothing of some details in the resultant fused image, thereby it results in redundant representation. To mitigate some of these drawbacks of SR based methods, a pan-sharpening technique based on an alternative framework, namely convolutional sparse representation (CSR), is proposed in this thesis. A novel pan-sharpening algorithm based on cartoon plus texture decomposition (CPT) and the convolutional sparse representation is developed. The fusion rules are appropriately chosen to holistically transfer spatial and spectral features into the resultant fused image. In the reconstructed HRMS image, to ensure the presence of essential spatial details, gradient based fusion rule is deployed to fuse cartoon and texture components. As an extension of this approach, the thesis presents another pan-sharpening method that combines discrete wavelet transform (DWT) and CSR framework. The source images are decomposed using DWT, and the resultant components are fused based on CSR based models. The CSR based pan-sharpened images possess a promising spectral quality that is more suitable for the applications like land-use classification, object detection and extraction. Further, a variational pan-sharpening model based on total generalized variation (TGV) and inter-band correlation is proposed to reduce the geometric structural difference between HRMS image and PAN image and to reduce the spectral distortion. On formulating optimization framework driven objective function, an efficient algorithm called alternating direction method of multipliers (ADMM) with an operator splitting framework is utilized to solve the proposed pan-sharpening model.

For the comprehensive analysis of all the proposed methods, the experimentation is performed on the datasets acquired from different geographical terrains. The visual analysis is performed on reduced and full-resolution data. The quality metrics prescribed by the Wald's protocol and quality with no reference (QNR) protocols are used for quantitative assessment. The experimental results obtained using the proposed SR based methods lead to moderate reduction in computational cost compared to traditional SR based methods. The pan-sharpening methods developed using CSR mechanism yields the fused outcomes with an enhanced spatial and spectral quality.

7.2 Future Work

With a motivation to enhance the quality of HRMS image, the pan-sharpening approaches presented in this thesis can be further explored while investigating the following issues.

- Recently, deep learning (DL) has achieved great success in various fields, such as image processing, pattern recognition, and computer vision. Deep learning has become a potentially interesting topic for remote sensing data analysis. A further investigation can be carried out to explore the application of DL for pan-sharpening.
- The spectral range mismatch and unavoidable misregistration between PAN and MS images are the critical problems in remote sensing imagery. The robust pan-sharpening methods need to be designed that minimize the effects of these unwarranted features.
- The learned dictionaries usually consist of a large number of atoms to accurately reconstruct the image patches. Various existing dictionaries are analyzed during this research work in terms of the number of atoms, adaptability, and computational complexity. However, learning a dictionary having acceptable representation capability with a small fixed number of atoms is desirable for sparse representation based pan-sharpening methods and still a challenging problem.
- Most of the existing algorithms are more suitable for the pan-sharpening of multispectral images. The possible extension of these algorithms with appropriate modifications can be investigated for the pan-sharpening of hyperspectral images.

REFERENCES

- Aharon, Michal, Michael Elad, and Alfred Bruckstein (2006), “*rm* K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.” *IEEE Transactions on signal processing*, 54, 4311–4322.
- Aiazzi, B, L Alparone, S Baronti, A Garzelli, and M Selva (2006), “MTF-tailored multiscale fusion of high-resolution MS and Pan imagery.” *Photogrammetric Engineering & Remote Sensing*, 72, 591–596.
- Aiazzi, Bruno, Luciano Alparone, Stefano Baronti, and Andrea Garzelli (2002), “Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis.” *IEEE Transactions on geoscience and remote sensing*, 40, 2300–2312.
- Aiazzi, Bruno, Stefano Baronti, and Massimo Selva (2007), “Improving component substitution pansharpening through multivariate regression of MS + Pan data.” *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3230–3239.
- Alparone, Luciano, Bruno Aiazzi, Stefano Baronti, and Andrea Garzelli (2015), *Remote sensing image fusion*. Crc Press.
- Alparone, Luciano, Bruno Aiazzi, Stefano Baronti, Andrea Garzelli, Filippo Nencini, and Massimo Selva (2008), “Multispectral and panchromatic data fusion assessment without reference.” *Photogrammetric Engineering & Remote Sensing*, 74, 193–200.
- Alparone, Luciano, Lucien Wald, Jocelyn Chanussot, Claire Thomas, Paolo Gamba, and Lori-Man Bruce (2007), “Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data fusion contest.” *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3012–3021.
- Amolins, Krista, Yun Zhang, and Peter Dare (2007a), “Wavelet based image fusion techniques — an introduction, review and comparison.” *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 249 – 263.
- Amolins, Krista, Yun Zhang, and Peter Dare (2007b), “Wavelet based image fusion techniques—an introduction, review and comparison.” *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 249–263.

Aujol, Jean-François and Guy Gilboa (2006), “Constrained and snr-based solutions for tv-hilbert space image denoising.” *Journal of Mathematical Imaging and Vision*, 26, 217–237.

Ayas, Selen, Esra Tunc Gormus, and Murat Ekinci (2018), “An efficient pan sharpening via texture based dictionary learning and sparse representation.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Ballester, Coloma, Vicent Caselles, Laura Igual, Joan Verdera, and Bernard Rougé (2006), “A variational model for p+ xs image fusion.” *International Journal of Computer Vision*, 69, 43–58.

Bredies, Kristian, Karl Kunisch, and Thomas Pock (2010), “Total generalized variation.” *SIAM Journal on Imaging Sciences*, 3, 492–526.

Bredies, Kristian and Tuomo Valkonen (2011), “Inverse problems with second-order total generalized variation constraints.” *Proceedings of SampTA*, 201.

Bristow, Hilton, Anders Eriksson, and Simon Lucey (2013), “Fast convolutional sparse coding.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 391–398.

Buades, Antoni, Triet M Le, Jean-Michel Morel, and Luminita A Vese (2010), “Fast cartoon+ texture image filters.” *IEEE Transactions on Image Processing*, 19, 1978–1986.

Candes, Emmanuel J, Justin K Romberg, and Terence Tao (2006), “Stable signal recovery from incomplete and inaccurate measurements.” *Communications on pure and applied mathematics*, 59, 1207–1223.

Carper, W Joseph (1990), “The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multi-spectral image data.” *Photogramm. Eng. Remote Sens.*, 56, 457–467.

Chavez, Pats, Stuart C Sides, Jeffrey A Anderson, et al. (1991), “Comparison of three different methods to merge multiresolution and multispectral data- Landsat TM and SPOT panchromatic.” *Photogrammetric Engineering and remote sensing*, 57, 295–303.

Chen, Chaoqian, Yong Meng, Qixiang Luo, and Zeming Zhou (2018), “A novel variational model for pan-sharpening based on l1 regularization.” *Remote Sensing Letters*, 9, 170–179.

- Chen, Chen, Yeqing Li, Wei Liu, and Junzhou Huang (2014), “Image fusion with local spectral consistency and dynamic gradient sparsity.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2760–2765.
- Cheng, Jian, Haijun Liu, Ting Liu, Feng Wang, and Hongsheng Li (2015), “Remote sensing image fusion via wavelet transform and sparse representation.” *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 158–173.
- Cheng, Ming, Cheng Wang, and Jonathan Li (2014), “Sparse representation based pansharpening using trained dictionary.” *IEEE Geoscience and Remote Sensing Letters*, 11, 293–297.
- Deng, Liang-Jian, Gemine Vivone, Weihong Guo, Mauro Dalla Mura, and Jocelyn Chanussot (2018), “A variational pansharpening approach based on reproducible kernel hilbert space and heaviside function.” *IEEE Transactions on Image Processing*, 27, 4330–4344.
- Dogra, Ayush, Bhawna Goyal, and Sunil Agrawal (2017), “From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications.” *IEEE Access*, 5, 16040–16067.
- Donoho, David L (2006), “For most large underdetermined systems of equations, the minimal l_1 -norm near-solution approximates the sparsest near-solution.” *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59, 907–934.
- Duran, Joan, Antoni Buades, Bartomeu Coll, and Catalina Sbert (2014), “A nonlocal variational model for pansharpening image fusion.” *SIAM Journal on Imaging Sciences*, 7, 761–796.
- Duran, Joan, Antoni Buades, Bartomeu Coll, Catalina Sbert, and Gwendoline Blanchet (2017), “A survey of pansharpening methods with a new band-decoupled variational model.” *ISPRS Journal of Photogrammetry and Remote Sensing*, 125, 78–105.
- Elad, Michael, Mario AT Figueiredo, and Yi Ma (2010), “On the role of sparse and redundant representations in image processing.” *Proceedings of the IEEE*, 98, 972–982.
- Fei, Rongrong, Jianshe Zhang, Junmin Liu, Fang Du, Peiju Chang, and Junying Hu (2019), “Convolutional sparse representation of injected details for pansharpening.” *IEEE Geoscience and Remote Sensing Letters*, 16, 1595–1599.

Gabay, Daniel and Bertrand Mercier (1975), *A dual algorithm for the solution of non linear variational problems via finite element approximation*. Institut de recherche d'informatique et d'automatique.

Ghassemian, Hassan (2016), "A review of remote sensing image fusion methods." *Information Fusion*, 32, 75–89.

Gogineni, Rajesh and Ashvini Chaturvedi (2019), "A robust pansharpening algorithm based on convolutional sparse coding for spatial enhancement." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12, 4024–4037.

Gonzalez, Rafael C (2016), "Digital image processing."

Guo, Weihong, Jing Qin, and Wotao Yin (2014a), "A new detail-preserving regularization scheme." *SIAM journal on imaging sciences*, 7, 1309–1334.

Guo, Weihong, Jing Qin, and Wotao Yin (2014b), "A new detail-preserving regularization scheme." *SIAM journal on imaging sciences*, 7, 1309–1334.

He, Xiyan, Laurent Condat, José M Bioucas-Dias, Jocelyn Chanussot, and Junshi Xia (2014), "A new pansharpening method based on spatial and spectral sparsity priors." *IEEE Transactions on Image Processing*, 23, 4160–4174.

Jiang, Cheng, Hongyan Zhang, Huanfeng Shen, and Liangpei Zhang (2012), "A practical compressed sensing-based pan-sharpening method." *IEEE Geoscience and Remote Sensing Letters*, 9, 629–633.

Jiang, Cheng, Hongyan Zhang, Huanfeng Shen, and Liangpei Zhang (2014), "Two-step sparse coding for the pan-sharpening of remote sensing images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7, 1792–1805.

Khan, Muhammad Murtaza, Luciano Alparone, and Jocelyn Chanussot (2009), "Pansharpening quality assessment using the modulation transfer functions of instruments." *IEEE transactions on geoscience and remote sensing*, 47, 3880–3891.

Knoll, Florian, Kristian Bredies, Thomas Pock, and Rudolf Stollberger (2011), "Second order total generalized variation (tgv) for mri." *Magnetic resonance in medicine*, 65, 480–491.

Laben, Craig A and Bernard V Brower (2000), "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening." US Patent 6,011,875.

- Li, Shutao, Xudong Kang, and Jianwen Hu (2013a), “Image fusion with guided filtering.” *IEEE Transactions on Image processing*, 22, 2864–2875.
- Li, Shutao and Bin Yang (2011), “A new pan-sharpening method using a compressed sensing technique.” *IEEE Transactions on Geoscience and Remote Sensing*, 49, 738–746.
- Li, Shutao, Haitao Yin, and Leyuan Fang (2013b), “Remote sensing image fusion via sparse representations over learned dictionaries.” *IEEE Transactions on Geoscience and Remote Sensing*, 51, 4779–4789.
- Liu, Yu, Xun Chen, Rabab K Ward, and Z Jane Wang (2016), “Image fusion with convolutional sparse representation.” *IEEE signal processing letters*, 23, 1882–1886.
- Lotfi, Meysam and Hassan Ghassemian (2018a), “A new variational model in texture space for pansharpening.” *IEEE Geoscience and Remote Sensing Letters*, 15, 1269–1273.
- Lotfi, Meysam and Hassan Ghassemian (2018b), “A pansharpening method based on modified cartoon plus texture decomposition.” *Remote Sensing Letters*, 9, 209–217.
- Meyer, Yves (2001), *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, 22. American Mathematical Soc.
- Möller, Michael, Todd Wittman, Andrea L Bertozzi, and Martin Burger (2012), “A variational approach for sharpening high dimensional images.” *SIAM Journal on Imaging Sciences*, 5, 150–178.
- Nencini, Filippo, Andrea Garzelli, Stefano Baronti, and Luciano Alparone (2007), “Remote sensing image fusion using the curvelet transform.” *Information fusion*, 8, 143–156.
- Nunez, Jorge, Xavier Otazu, Octavi Fors, Albert Prades, Vicenc Pala, and Roman Arbiol (1999), “Multiresolution-based image fusion with additive wavelet decomposition.” *IEEE Transactions on Geoscience and Remote sensing*, 37, 1204–1211.
- Ophir, Boaz, Michael Lustig, and Michael Elad (2011), “Multi-scale dictionary learning using wavelets.” *IEEE Journal of Selected Topics in Signal Processing*, 5, 1014–1024.
- Osher, Stanley, Andrés Solé, and Luminita Vese (2003), “Image decomposition and restoration using total variation minimization and the h.” *Multiscale Modeling & Simulation*, 1, 349–370.

Otazu, Xavier, María González-Audícana, Octavi Fors, and Jorge Núñez (2005), “Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods.” *IEEE Transactions on Geoscience and Remote Sensing*, 43, 2376–2385.

Palsson, Frosti, Johannes R Sveinsson, and Magnus O Ulfarsson (2014), “A new pansharpening algorithm based on total variation.” *IEEE Geoscience and Remote Sensing Letters*, 11, 318–322.

Palsson, Frosti, Johannes R Sveinsson, Magnus Orn Ulfarsson, and Jon Atli Benediktsson (2016), “MTF-based deblurring using a wiener filter for CS and MRA pansharpening methods.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9, 2255–2269.

Pohl, Christine and John van Genderen (2015), “Structuring contemporary remote sensing image fusion.” *International Journal of Image and Data Fusion*, 6, 3–21.

Pohl, Christine and John Van Genderen (2016), *Remote sensing image fusion: A practical guide*. Crc Press.

Ranchin, Thierry, Bruno Aiazzi, Luciano Alparone, Stefano Baronti, and Lucien Wald (2003), “Image fusion — the ARSIS concept and some successful implementation schemes.” *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, 4–18.

Rish, Irina and Genady Grabarnik (2014), *Sparse modeling: theory, algorithms, and applications*. CRC press.

Sadeghi, Mostafa, Massoud Babaie-Zadeh, and Christian Jutten (2014), “Learning overcomplete dictionaries based on atom-by-atom updating.” *IEEE Transactions on Signal Processing*, 62, 883–891.

Thomas, Claire, Thierry Ranchin, Lucien Wald, and Jocelyn Chanussot (2008), “Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics.” *IEEE Transactions on Geoscience and Remote Sensing*, 46, 1301–1312.

Tropp, Joel A (2004), “Greed is good: Algorithmic results for sparse approximation.” *IEEE Transactions on Information theory*, 50, 2231–2242.

Tu, Te-Ming, Ping Sheng Huang, Chung-Ling Hung, and Chien-Ping Chang (2004), “A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery.” *IEEE Geoscience and Remote sensing letters*, 1, 309–312.

Tu, Te-Ming, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping S Huang (2001), “A new look at IHS-like image fusion methods.” *Information fusion*, 2, 177–186.

Vicinanza, Maria Rosaria, Rocco Restaino, Gemine Vivone, Mauro Dalla Mura, and Jocelyn Chanussot (2015), “A pansharpening method based on the sparse representation of injected details.” *IEEE Geoscience and Remote Sensing Letters*, 12, 180–184.

Vivone, Gemine, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio Licciardi, Rocco Restaino, and Lucien Wald (2014), “A critical comparison of pansharpening algorithms.” *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 191–194, IEEE.

Vivone, Gemine, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald (2015), “A critical comparison among pansharpening algorithms.” *IEEE Transactions on Geoscience and Remote Sensing*, 53, 2565–2586.

Wald, Lucien (1999), “Some terms of reference in data fusion.” *IEEE Transactions on geoscience and remote sensing*, 37, 1190–1193.

Wald, Lucien, Thierry Ranchin, and Marc Mangolini (1997), “Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images.” *Photogrammetric engineering and remote sensing*, 63, 691–699.

Wang, Jun, Jinye Peng, Xiaoyue Jiang, Xiaoyi Feng, and Jianhong Zhou (2017), “Remote-sensing image fusion using sparse representation with sub-dictionaries.” *International Journal of Remote Sensing*, 38, 3564–3585.

Wohlberg, Brendt (2015), “Efficient algorithms for convolutional sparse representations.” *IEEE Transactions on Image Processing*, 25, 301–315.

Wohlberg, Brendt (2016a), “Efficient algorithms for convolutional sparse representations.” *IEEE Transactions on Image Processing*, 25, 301–315.

Wohlberg, Brendt (2016b), “Sparse optimization research code (sporco).” *Software library available from <http://purl.org/brendt/software/sporco>*.

Yang, Bin and Shutao Li (2010), “Multifocus image fusion and restoration with sparse representation.” *IEEE Transactions on Instrumentation and Measurement*, 59, 884–892.

Yang, Xiao-Hui and Li-Cheng Jiao (2008), “Fusion algorithm for remote sensing images based on nonsubsampling contourlet transform.” *Acta Automatica Sinica*, 34, 274–281.

Yang, Xiaomin, Lihua Jian, Binyu Yan, Kai Liu, Lei Zhang, and Yiguang Liu (2018), “A sparse representation based pansharpening method.” *Future Generation Computer Systems*, 88, 385–399.

Yin, Haitao (2015), “Sparse representation based pansharpening with details injection model.” *Signal Processing*, 113, 218–227.

Zhao, Yongqiang, Jingxiang Yang, and Jonathan Cheung-Wai Chan (2014), “Hyperspectral imagery super-resolution by spatial–spectral joint nonlocal similarity.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7, 2671–2679.

Zhao, Yongqiang, Jinxiang Yang, Qingyong Zhang, Lin Song, Yongmei Cheng, and Quan Pan (2011), “Hyperspectral imagery super-resolution by sparse representation and spectral regularization.” *EURASIP Journal on Advances in Signal Processing*, 2011, 87.

Zhong, Shengwei, Ye Zhang, Yushi Chen, and Di Wu (2017), “Combining component substitution and multiresolution analysis: A novel generalized BSD3D pansharpening algorithm.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 2867–2875.

Zhou, J, DL Civco, and JA Silander (1998), “A wavelet transform method to merge landsat tm and spot panchromatic data.” *International journal of remote sensing*, 19, 743–757.

Zhu, Xiao Xiang and Richard Bamler (2013), “A sparse image fusion algorithm with application to pan-sharpening.” *IEEE transactions on geoscience and remote sensing*, 51, 2827–2836.

Zhu, Xiao Xiang, Claas Grohnfeldt, and Richard Bamler (2016), “Exploiting joint sparsity for pansharpening: the J-SparseFI algorithm.” *IEEE Transactions on Geoscience and Remote Sensing*, 54, 2664–2681.

Zhu, Zhiqin, Hongpeng Yin, Yi Chai, Yanxia Li, and Guanqiu Qi (2018), “A novel multi-modality image fusion method based on image decomposition and sparse representation.” *Information Sciences*, 432, 516–529.

Publications Based on the Thesis

Journals :


1. Gogineni, R., & Chaturvedi, A. (2018). Sparsity inspired pan-sharpening technique using multi-scale learned dictionary. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 360-372.
2. Gogineni, R., & Chaturvedi, A. (2019). A Robust Pansharpening Algorithm Based on Convolutional Sparse Coding for Spatial Enhancement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10), 4024-4037.
3. Rajesh Gogineni, Ashvini Chaturvedi & Daya Sagar B S (2020) A variational pan-sharpening algorithm to enhance the spectral and spatial details, *International Journal of Image and Data Fusion*, DOI: 10.1080/19479832.2020.1838629

Bio-data

NAME : Rajesh Gogineni

CONTACT DETAILS

Address : S/o G. Hanumantha Rao, D.No: 4-37,
Naguluppalapadu (post) (village)
Prakasam Dist., Andhra Pradesh-523183

 : 9010303888

 : rgogineni9@gmail.com

EDUCATIONAL QUALIFICATIONS

Doctor of Philosophy (Ph.D)

National Institute of Technology Karnataka, Surathkal 2015–Till date

Master of Technology (M.Tech)

V R Siddhartha Engineering College, Vijayawada, Andhra Pradesh 2010–2012
Branch : Communication and Signal Processing

Bachelor of Technology (B.Tech)

Sri Sarathi Institute of Engineering and Technology, Nuzvid, A.P. 1998–2002
Branch : Electronics and Communication Engineering