

M.Sc. Thesis

# Image Fusion in the Multi-Scale Transforms Domain Using Fuzzy Logic and Particle Swarm Optimization



Author  
**Jamal Saeedi**

Thesis advisor  
**Prof. Karim Faez**



**Amirkabir University of Technology  
(Tehran Polytechnic)**

**Electrical Engineering Department**

A Thesis Submitted in Partial Fulfillment of the Requirement for the  
Degree of Master of Science in Electrical Engineering, Electronics

**Title**

**Image Fusion in the Multi-Scale Transforms Domain  
Using Fuzzy Logic and Particle Swarm Optimization**

**By**

**Jamal Saeedi**

**Supervisor**

**Dr. Karim Faez**

**February 2010**



Copyright © 2010 by Jamal Saeedi. All rights reserved. Reproduction  
by any means or translation of any part of this work is forbidden  
without permission of the copyright holder.

*To my beloved Rozita*

*To my father Saeed, my mother Shahrbanoo*

*and my sisters Zari and Roza.*

## Acknowledgments

- ❖ Many thanks belong to my advisor and mentor, Dr. Karim Faez as well as the students at Tehran Polytechnic in Machine Vision Research Lab (MVRL) who helped so much with this research. Most of all, I wish to thank my amazing wife Rozita for being extremely patient and having so much faith in me. Thank you.
- ❖ The author would like to thank N. Kingsbury for his assistance in providing the code for the DT-DWT. A. Toet of the TNO Human Factors Research Institute kindly supplied the original “UN Camp”, “Dune”, and “Trees” IR and visible images.
- ❖ The author would like to express their honest appreciation to the Iranian Space Agency for providing the Landsat-7+ETM images for this research. In addition, the QuickBird, and Wordview2 imagery were obtained from Space Imaging, DigitalGlobe, and Global Land Cover Facility (GLCF), respectively.

## **Abstract**

The research presented in this thesis is concerned with the problem of multi-sensor pixel-level image fusion. The objective is to introduce reliable methods that represent the visual information, obtained from a number of disparate imaging sensors, in a single fused image without the introduction of distortion or loss of information. Generally, the image fusion's task is used for three different applications consisting of fusing the multi-focus images, fusing the infrared and visible images, and fusing the multi-spectral and panchromatic images. For fusing of multi-focus image, first we use dual-tree discrete wavelet transform for decomposing the source images and then combine wavelet coefficients of them. We formulate the image fusion process as a two-class problem: in focus and out of focus classes, in which the decision map for selecting important coefficients between input images is obtained using two-class fisher classifier. In the proposed method for fusing infrared and visible images, first, the input images are decomposed using dual-tree discrete wavelet transform and then, we use a dissimilarity measure of source images to combine three different fusion rules for selecting high frequency wavelet coefficients between source images. In addition, we propose a new method based on particle swarm optimization for fusing of low frequency coefficients, which produces the fused images with high entropy. Finally, a new method for fusion of remote sensing images is proposed. In this method, the aim is to improve spatial and spectral quality of the fused image, simultaneously. We use the shiftable contourlet transform and multi-objective particle swarm optimization for this purpose. Extensive comparisons with the state-of-the-art image fusion algorithm indicate that our image fusion algorithms have a better performance as subjective and objective evaluations.

**Key Words:** Multi-Focus images; Dual-tree discrete wavelet transform; Fisher classifier; Shiftable Contourlet Transform; and Multi-Objective Particle Swarm Optimization.

# Table of Contents

<b>Acknowledgments.....</b>	<b>v</b>
<b>Abstract.....</b>	<b>vi</b>
<b>Table of Contents.....</b>	<b>vii</b>
<b>List of Figures .....</b>	<b>x</b>
<b>List of Tables .....</b>	<b>xvi</b>
<b>Acronyms .....</b>	<b>xviii</b>
<b>List of Publications .....</b>	<b>xxi</b>
<b>Chapter 1. Introduction .....</b>	<b>1</b>
1.1 Image fusion.....	1
1.2 Motivation for image fusion research.....	4
1.3 Image preprocessing .....	6
1.4 Project objectives and organization of the thesis.....	7
<b>Chapter 2. Literature survey.....</b>	<b>9</b>
2.1 Multi-focus image fusion.....	10
2.1.1 Spatial domain methods .....	10
2.1.2 Optimization methods .....	14
2.1.3 Multi-scale decomposition-based methods .....	14
2.2 Infrared and visible image fusion .....	17
2.3 MS and Pan image fusion .....	19
2.3.1 Color transformation .....	21
2.3.2 Statistical methods.....	21
2.3.3 Numerical methods.....	22
2.3.4 Hybrid methods .....	22
2.4 Summary.....	23
<b>Chapter 3. Multi-scale image representation .....</b>	<b>24</b>

3.1 Discrete wavelet transform .....	25
3.2 Dual-tree discrete wavelet transform.....	27
3.3 Contourlet transform.....	31
3.4 Shiftable Contourlet transform .....	34
3.5 Summary.....	38
<b>Chapter 4. The proposed methods for image fusion .....</b>	<b>39</b>
4.1 Multi-focus image fusion using fuzzy logic and classification .....	39
4.1.1 Feature extraction.....	42
4.1.2 Fisher classifier.....	44
4.1.3 Fusion rule for high frequency coefficients.....	48
4.1.4 Fusion rule for low frequency coefficients.....	54
4.2 IR and visible image fusion using fuzzy logic and particle swarm optimization ...	57
4.2.1 Image segmentation algorithm using watershed transform.....	57
4.2.2 The proposed image fusion method .....	60
4.2.2.1 Generation of pixel and region-based decision maps .....	61
4.2.2.2 Fusion rule for high frequency wavelet coefficients .....	64
4.2.2.3 Fusion rule for low frequency wavelet coefficients .....	67
4.2.3 Optimization algorithm .....	71
4.2.3.1 Particle swarm optimization .....	73
4.2.3.2 Genetic algorithm .....	76
4.2.4 Computation time .....	77
4.3 Remote-sensing image fusion using multi-objective particle Swarm optimization and the shiftable Contourlet transform .....	78
4.3.1 Generation of two initial pan-sharpened results .....	81
4.3.2 Multi-objective optimization .....	87
4.3.3 Multi-objective particle swarm optimization .....	90
4.3.4 Parameters selection for the proposed pan-sharpening algorithm.....	94
4.3.5 Comparisons between different multi-resolution transformations in the proposed algorithm.....	98
4.4 Summary.....	99
<b>Chapter 5. Experimental results .....</b>	<b>103</b>
5.1 The results of multi-focus image fusion .....	105
5.1.1 Fusion results of natural images .....	106
5.1.2 Fusion results of artificial images.....	111

5.2 The results of IR and visible image fusion .....	111
5.3 The results of remote-sensing image fusion .....	118
5.3.1 Experimental datasets.....	119
5.3.2 Comparisons between different state-of-the-arts Pan-sharpening algorithms .....	122
5.3.3 Comparisons between different multi-resolution-based Pan-sharpening algorithms	127
<b>Chapter 6. Conclusions and future works .....</b>	<b>132</b>
<b>References .....</b>	<b>135</b>

## List of Figures

Figure 1.1: An automatic target detection/recognition (ATD/ATR) system using image fusion at different processing levels.	3
Figure 1.2: Basic structure of a multi-sensor system using pixel-level image fusion [6]	4
Figure 2.1: Multi-focus image fusion. Left to right: wright-focus, left-focus, and fused images [16].	10
Figure 2.2: Schematic diagram of the spatial frequency method [21].	13
Figure 2.3: Schematic diagram of the improved spatial frequency method [22].	13
Figure 2.4: Schematic diagram of region-based fusion methods.	13
Figure 2.5: Multi-scale fusion block diagram.	15
Figure 2.6: Fusion of multi-sensor images, from left to right: infrared and visible “UN Camp” images, and fused image.	17
Figure 2.7: A general framework for multi-resolution-based image fusion [8].	18
Figure 2.8: Spectral ranges of different bands of Landsat-7 satellite.	20
Figure 2.9: Remote sensing image fusion. Left to right: Pan, MS with lower resolution, and fused image (2004 Digital Globe).	20
Figure 3.1: Two channel filter bank of DWT	26
Figure 3.2: Analysis filter bank structure of the DWT.	27
Figure 3.3: 2-D DWT. Left: Barbara test image and right: DWT using “db2” mother wavelet at two decomposition levels. Small coefficients are shown in black while large coefficients are shown in white.	27
Figure 3.4: The Q-shift version of the DT CWT, giving real and imaginary parts of complex coefficients from tree $a$ and tree $b$ respectively. Figures in brackets indicate the delay for each filter, where $q = 1/4$ sample period [81].	30

Figure 3.5: Improvement for directional features extraction in DT-DWT, Top: DT-DWT subbands at two decomposition levels, and bottom: DT-DWT transform of Barbara test image. Small coefficients are shown in black while large coefficients are shown in white.	31
Figure 3.6: Discrete contourlet transform (This figure is a modified version of [83]).	33
Figure 3.7: Two-dimensional spectrum partition using quincunx filter banks with fan filters. The black regions represent the ideal frequency supports of each filter. Q is a quincunx sampling matrix [83].	34
Figure 3.8: Example of shearing operation that is used like a rotation operation for DFB decomposition. Left: The “cameraman” image and right: The “cameraman” image after a shearing operation [83].	34
Figure 3.9: Examples of the contourlet transform on the <i>Barbara</i> image. For clear visualization, each image is only decomposed into two pyramidal levels, which are then decomposed into four and eight directional subbands. Small coefficients are shown in black while large coefficients are shown in white.	35
Figure 3.10: The structure of the shiftable complex directional pyramid, analysis (left) and synthesis (right) sides [89].	36
Figure 4.1: Block diagram of proposed image fusion algorithm.	41
Figure 4.2: Top: Training images used in the experiment, and bottom: class labels (black regions indicate out-of-focus areas and white regions indicate in focus areas for the first training image and for the second training image, black regions indicate in of focus areas and white regions indicate out-of-focus areas).	46
Figure 4.3: Classification results for the different multi-focus test images. First and second columns from top to bottom: pairs of multi-focus “Flower”, “Doll”, “Rose”, “Book”, and “Disk” images, third and fourth columns: the classification result, and the handmade classification maps, respectively.	47
Figure 4.4: First column: the classification results for multi-focus “Doll”, “Rose”, “Book”, and “Disk” images, second column: post-processing results using majority filter, third column: the post-processing results using proposed method and fourth column: handmade classification map.	49

Figure 4.5: Fuzzy membership function, $T_1 = 0.1$ and $T_2 = 0.25$ , which are obtained using test images by trial and error.	51
Figure 4.6: Border map for multi-focus “Doll”, “Rose”, “Disk”, and “Book” images (from top left, clockwise).	52
Figure 4.7: Subjective fusion results, Top: from left to right: A part of Multi-focus “Disk” images, results without and with border correction, and bottom: from left to right: A part of Multi-focus “Flower” images, results without and with border correction.	53
Figure 4.8: Segmentation of IR and visible images, from top left, clockwise: IR image, visible image, segmentation map, and the joint gradient image.	59
Figure 4.9: Block diagram of the proposed image fusion algorithm.	61
Figure 4.10: Pixel- and region-based decision maps (first and second row) obtained from “UN Camp” images in the first decomposition level and six directional sub-bands of the DT-DWT, from left to right: -15, +15, -45, +45, -75 and +75 directions, respectively. Black regions indicate that high frequency wavelet coefficients in the DT-DWT sub-bands of the final fused image should be selected from the visible image, and white regions indicate that high frequency wavelet coefficients in the DT-DWT sub-bands of the final fused image should be selected from the IR image.	64
Figure 4.11: Triangular fuzzy membership functions.	68
Figure 4.12: The Petrovic index of fused images versus different thresholds ( $T$ for building fuzzy membership functions) obtained from different multi-sensor datasets, from left to right: “UN Camp”, “Dune”, and “Trees”.	68
Figure 4.13: The Entropy index of fused images versus iteration of the PSO and genetic algorithms obtained from different multi-sensor datasets, from left to right: “UN Camp”, “Dune”, and “Trees”.	72
Figure 4.14: Subjective image fusion results using different low frequency fusion rule in the proposed fusion scheme. (a) “UN Camp” IR image, (b) visible image, and fused images using (c) averaging, (d) weighted averaging [52], (e) Arithmetic Combination [100] and (f) the proposed method based on PSO algorithm.	73
Figure 4.15: The entropy index of fused image obtained from “UN Camp” dataset at (a)	

different iterations of the PSO algorithm using different particles number and (b) different generation of genetic algorithm using different population sizes.	76
Figure 4.16: Flowchart of the proposed pan-sharpening algorithm.	82
Figure 4.17: True-color composites of the two initial pan-sharpened images for different multi-spectral and panchromatic images from the Landsat-7 dataset.	87
Figure 4.18: (a) The particles at first and second iterations, (b) the dominant (white circles) and Pareto front solutions (black circles), and (c) the true Pareto front, the Euclidian distances between Pareto front set and the true Pareto front, and the <i>leader</i> .	94
Figure 4.19: (a) Overall CC to the Pan index, (b) ERGAS index, and (c) computation time of the proposed algorithm versus the size of the sliding window for local parameter estimation.	96
Figure 4.20: Visual comparison between different multi-resolution transformations in the proposed algorithm. Top row, left to right: Low resolution MS image, Pan image with two pixels shift, and result of DT-DWT+MOPSO. Bottom row: Results of CT+MOPSO, SCT+MOPSO, and SCT+MOPSO without shifting.	100
Figure 5.1: (a) Subjective and (b) objective, and fusion parameter optimization	104
Figure 5.2: Multi-focus source images, from top left, clockwise: “Pepsi”, “Clock”, “Jug”, and “Lab” images.	107
Figure 5.3: PSNR and Petrovic indexes for different image fusion methods.	109
Figure 5.4: Subjective fusion results. Top row, left to right: a part of multi-focus “Disk” images, the fused result using PCA, SF methods. Bottom row, left to right: the fused result using MS, WA, WBV, and the proposed methods.	110
Figure 5.5: Subjective fusion results. Top row, left to right: a part of multi-focus “Lab” images, the fused results using PCA, SF methods. Middle row, left to right: the fused results using MS, WA, WBV, and the proposed methods. Bottom row, left to right: magnified head regions of the fused results produced with PCA, SF, MS, WA, WBV, and the proposed method.	110
Figure 5.6: Ten ground truth images.	112

Figure 5.7: PSNR and Petrovic indexes for different image fusion methods.	112
Figure 5.8: Subjective fusion results of “UN Camp” images: Original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [34], MS, WA [45] methods (middle-row), and fused images using WBV [47], Lewis [49], the proposed methods (bottom-row).	115
Figure 5.9: Subjective fusion results of “Dune” images: Original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [34], MS, WA [45] methods (middle-row), and fused images using WBV [47], Lewis [49], the proposed methods (bottom-row).	116
Figure 5.10: Subjective fusion results of “Trees” images: Original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [34], MS, WA [45] methods (middle-row), and fused images using WBV [47], Lewis [49], the proposed methods (bottom-row).	117
Figure 5.11: True-color composites of multi-spectral data used in the experiment. (a) Lansat-7 ETM+ scene from Iran, (b) Quickbird scene from Boulder, CO, USA, (c) Quickbird scene from Kokilai-Lagoon, Srilanka, and (d), Wordview2 scene from Rome, Italy.	121
Figure 5.12: Subjective results of different pan-sharpening methods. True-color composites of Low Resolution Multi-spectral image form the Wordview2 dataset, corresponding panchromatic image, and the result of IHS method (top-row, left to right). The results of Brovey, PCA, and P+XS methods (middle-row). The results of Ehlers, BDF, and the proposed methods, (bottom-row).	126
Figure 5.13: Subjective results of different pan-sharpening methods. True-color composites of Low Resolution Multi-spectral image form the Quickbird dataset, corresponding panchromatic image, and the result of IHS method (top-row, left to right). The results of Brovey, PCA, and P+XS methods (middle-row). The results of Ehlers, BDF, and the proposed methods, (bottom-row).	127
Figure 5.14: Subjective results of different multi-resolution-based pan-sharpening methods. True-color composites of Low Resolution Multi-spectral image from the Wordview2 dataset, corresponding panchromatic image, and the result of the DT-DWT+A method (top-row, left to right). The results of the DT-DWT+S, IHS+DT-	131

DWT+S, and CT+A methods (middle-row). The results of CT+S, PCA+CT, and the proposed methods (bottom-row).

## List of Tables

Table 3.1: Comparison between different multi-scale transformations (N is decomposition level, and K is number of direction in each level)	38
Table 4.1: Classification Accuracy using different local features.	46
Table 4.2: Classification Accuracy after post-processing.	49
Table 4.3: The best value of the PSNR <sub>dB</sub> index with respect to decomposition levels obtained for different multi-focus images and different fusion rules.	55
Table 4.4: PSNR <sub>dB</sub> index for different low frequency fusion rules.	56
Table 4.5: The entropy index of fused images obtained from different datasets using different LF fusion rules in the proposed fusion scheme.	71
Table 4.6: The PSO parameters selected for the optimization.	76
Table 4.7: Relative computation time of different image fusion techniques.	78
Table 4.8: Different objective measures for the two initial pan-sharpened results from different datasets shown in Fig. 4.17.	88
Table 4.9: The PSO parameters selected for the multi-objective optimization.	95
Table 4.10: Average performance results obtained from different MS and Pan images in the datasets using different decomposition levels of the SCT in the proposed algorithm.	97
Table 4.11: Performance results obtained from 20 MS and Panchromatic images using different multi-resolution transformations in the proposed algorithm.	100
Table 5.1: The objective results of final fused images using different decomposition levels in the proposed algorithm.	117
Table 5.2: Average performance results of different image fusion methods.	118

Table 5.3: Satellite datasets characteristics.	122
Table 5.4: Datasets' Summary	122
Table 5.5: Average Performance of different state-of-the-arts Pan-sharpening algorithms against the proposed method for different datasets.	124
Table 5.6: Average Performance of different multi-resolution-based Pan-sharpening algorithms against the proposed method for different datasets.	130

## Acronyms

<b>AG</b>	Average Gradient
<b>CM</b>	Choose Max
<b>CT</b>	Contourlet Transform
<b>DM</b>	Decision Map
<b>DFB</b>	Directional Filter Bank
<b>DWT</b>	Discrete Wavelet Transform
<b>DIS</b>	Dissimilarity Measure
<b>DTDWT</b>	Dual-Tree Discrete Wavelet Transform
<b>HF</b>	High Frequency
<b>IHS</b>	Intensity-Hue-Saturation
<b>LF</b>	Low Frequency
<b>MAS</b>	Maximum Absolute Selection
<b>ML</b>	Maximum Likelihood
<b>MSE</b>	Mean Square Error
<b>MF</b>	Membership Function
<b>MOPSO</b>	Multi-Objective Particle Swarm Optimization

<b>MG</b>	Multi-Scale Grouping
<b>MST</b>	Multiscale Transform
<b>MS</b>	Multi-spectral
<b>NN</b>	Neural Network
<b>NG</b>	No-Grouping
<b>NSCT</b>	Nonsubsampled Contourlet Transform
<b>Pan</b>	Panchromatic
<b>Pf</b>	Pareto-front
<b>PSNR</b>	Peak Signal to Noise Ratio
<b>PR</b>	Perfect Reconstruction
<b>PBA</b>	Pixel-based Activity
<b>PDM</b>	Pixel-based Decision Map
<b>PCA</b>	Principal Component Analysis
<b>PDTDFB</b>	Pyramidal Dual-Tree Directional Filter Bank
<b>RBA</b>	Region-based Activity
<b>RDM</b>	Region-based Decision Map
<b>SCT</b>	Shiftable Contourlet Transform
<b>SAV</b>	Sign-based Averaging
<b>SOPSO</b>	Single Objective Particle Swarm Optimization

<b>SG</b>	Single-Scale Grouping
<b>SF</b>	Spatial Frequency
<b>SSIM</b>	Structural Similarity Index Measure
<b>SVM</b>	Support Vector Machine
<b>UDWT</b>	Undecimated Discrete Wavelet Transform
<b>VI</b>	Visibility
<b>WA</b>	Weighted Average
<b>WBA</b>	Window-based Activity

# List of Publications

The following is a list of publications that is a result of the research presented in this dissertation:

## Journal papers

- J1. **J. Saeedi**, K. Faez, “Infrared and visible image fusion using fuzzy logic and population-based optimization,” *Elsevier Applied Soft Computing Journal* 12 (3), pp. 1041–1054, March 2012. [[pdf](#)] [[dataset](#)]
- J2. **J. Saeedi**, K. Faez, “A new pan-sharpening method using multiobjective particle swarm optimization and the shiftable contourlet transform,” *Elsevier Photogrammetry & Remote Sensing Journal* 66, pp. 365-381, 2011. [[pdf](#)] [[dataset](#)]
- J3. **J. Saeedi**, K. Faez, “A classification and fuzzy-based approach for digital multi-focus image fusion,” *Pattern Analysis and Application Journal* 16 (3), pp. 365-379, July 2013. [[pdf](#)] [[dataset](#)]

## Conference papers

- C1. **J. Saeedi**, K. Faez, and S. Mozaffari, “Multi-focus Image Fusion Based on Fuzzy and Wavelet Transform,” *CIARP 2009, LNCS 5856, Springer-Verlag Berlin Heidelberg*, pp. 970–977, 2009. [[pdf](#)]
- C2. **J. Saeedi**, K. Faez, “Fisher Classifier and Fuzzy Logic Based Multi-Focus Image Fusion,” *IEEE International Conference on Intelligent Computing and Systems Intelligent*, pp. 420-425, 2009. [[pdf](#)]
- C3. **J. Saeedi**, K. Faez, “The New Segmentation and Fuzzy Logic based Multi-Sensor Image Fusion,” *24<sup>th</sup> IEEE International Conference Image and Vision Computing New Zealand*, pp. 328-333, 2009. [[pdf](#)] [[ppt](#)]

# Chapter 1. Introduction

This Chapter describes definitions, motivation, image preprocessing, and project objectives and the thesis' organization.

## 1.1 Image fusion

Image fusion defines the process of fusing visual information from a number of registered images into a single fused image. It is part of the much broader subject of multi-sensor information fusion, which has attracted a considerable amount of research attention in the last two decades [1-4].

Multi-sensor information fusion utilizes information obtained from a number of different sensors surveying an environment. The aim is to achieve better situation assessment and/or more rapid and accurate completion of a pre-defined task than would be possible using any of the sensors individually. In the literature, it has been defined as the synergistic combination of different sources of sensory information into a single representational format [5-7]. The only formal definition of information fusion (data fusion) to date is that given by the U.S. Department of Defense, Joint Directors of Laboratories Data Fusion Subpanel that represents the first formal body explicitly dealing with the process of data fusion. Their definition can be found in [8] as “*a multilevel, multifaceted process dealing with the automatic detection,*

*association, correlation, estimation and combination of data and information from multiple sources.”*

Image fusion represents a specific case of multi-sensor information fusion in which all the information sources used represent imaging sensors. Information fusion can be achieved at any level of the image information representation. Analogous to other forms of information fusion, image fusion is usually performed at one of the three different processing levels: signal, feature and decision. Signal level image fusion, also known as pixel-level image fusion, represents fusion at the lowest level, where a number of raw input image signals are combined to produce a single fused image signal. Object level image fusion, also called feature level image fusion, fuses feature and object labels and property descriptor information that have already been extracted from individual input images. Finally, the highest level, decision or symbol level image fusion represents fusion of probabilistic decision information obtained by local decision makers operating on the results of feature level processing on image data produced from individual sensors. Fig. 1.1 illustrates an example of an automatic target detection/recognition (ATD/ATR) system using image fusion at different processing levels.

The subject matter of this dissertation is pixel-level image fusion. It represents fusion of visual information of the same scene, from any number of registered image signals, obtained using different sensors. The goal of pixel-level image fusion can broadly be defined as: to represent the visual information present in any number of input images, in a single fused image without the introduction of distortion or loss of information.

In another words, the main condition for successful fusion is that all visible information in the input images should also appear visible in the fused image. However, although theoretically possible, due to the redundant nature of multi-sensor information (e.g., slightly different signatures of the same object in different sensor modalities), the complete representation of all of the visual information from a number of input images into a single one is scarcely achieved in practice. Thus, the practical objective of pixel-level image fusion is modified to the fusion, or

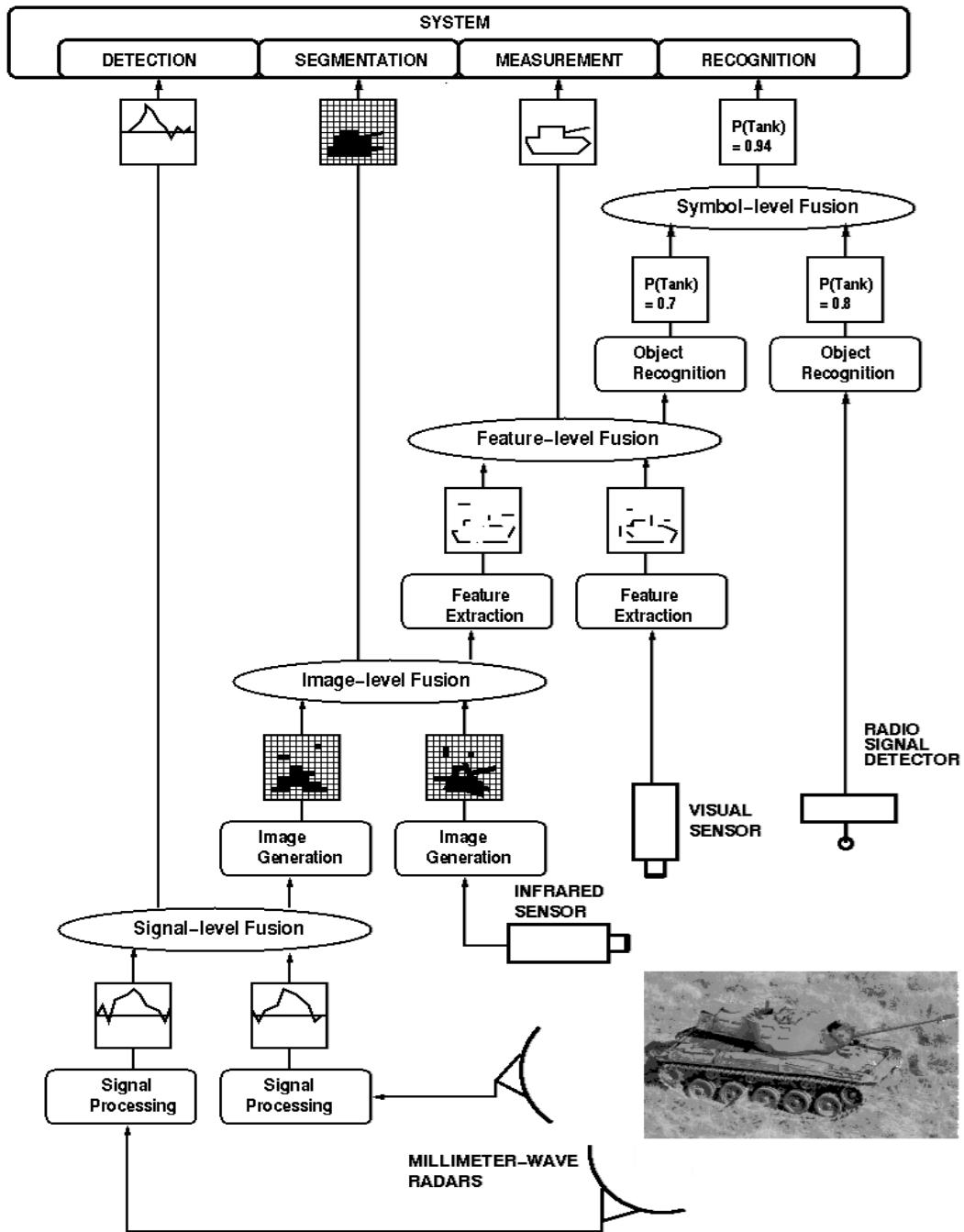


Figure 1.1: An automatic target detection/recognition (ATD/ATR) system using image fusion at different processing levels.

preservation in the output-fused image, of the “most important” visual information that exists in the input image set. The main requirement of the fusion process then, is to identify the most significant features in the input images and to transfer them without loss into the fused image. What defines important visual information is

generally application dependent. In most applications and in image fusion for display purposes in particular, it means perceptually important information.

A simple diagram of a system using pixel-level image fusion is shown in the block diagram in Fig 1.2. For simplicity, only two imaging sensors survey the environment, producing two different representations of the same scene. The representations of the environment are, again, in the form of image signals, which are corrupted by noise arising from the atmospheric aberrations, sensor design, quantization, etc.

The image signals produced by the sensors are directed into a registration process, which ensures that the input images to the fusion process correspond spatially, by geometrically warping one of them, e.g. Image A. Multi-sensor image registration is another widely researched area and more details can be found in [9-10]. In Fig. 1.2, the registered input images are fused and the resulting fused image can then be used directly for display purposes or can be passed on for further processing.

## 1.2 Motivation for image fusion research

Motivation for image fusion research is mainly the result of recent technological advances in the fields of sensing methods and sensor design. Improved robustness and increased resolution of modern imaging sensors and, more significantly, availability at a lower cost, have made the use of multiple sensors common in a range of imaging applications.

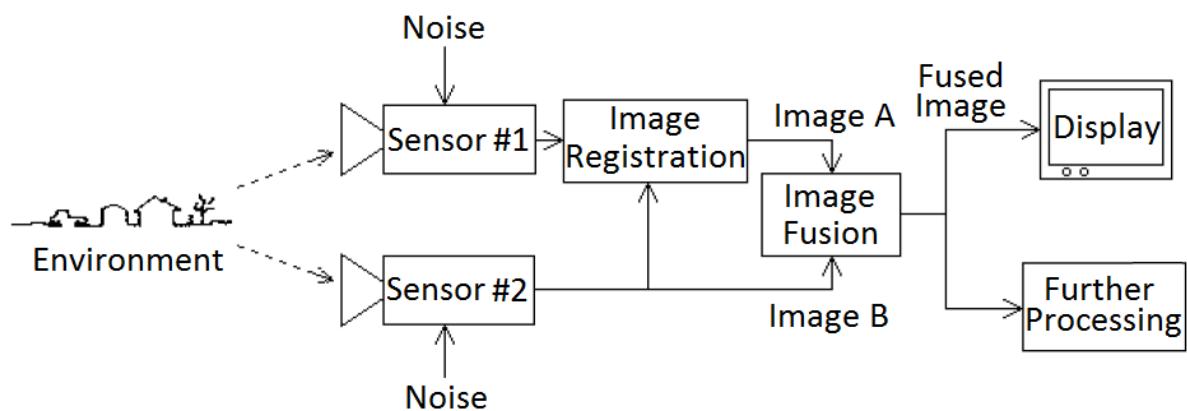


Figure 1.2: Basic structure of a multi-sensor system using pixel-level image fusion [6]

In the past decade, medical imaging, night vision, military and civilian avionics, autonomous vehicle navigation, remote sensing, concealed weapons detection and various security and surveillance systems are only some of the applications that have benefited from such multi-sensor arrays. Increased spatial, spectral resolution and faster scan rates offered by such modern sensor suites provide a more reliable and complete picture of the scanned environment.

This in turn can lead to improved performance of dedicated imaging systems. However, potential performance gains come at the cost of a large increase on the raw amount of sensor data that has to be processed. Thus, an increase in the number of sensors used in a particular application leads to the proportional increase for image data. Furthermore, a linear increase in the size of the imaging arrays has an even more dramatic effect resulting in an exponential (specifically quadratic) increase for data appearing at the sensor array output. This means that, if system performance improvements are to be realized, deployment of additional sensors must be accompanied by a corresponding increase in the processing power of the system.

In automated task applications, such as autonomous vehicle navigation, this demand can be met by a corresponding increase in the number of processing units, using faster digital signal processing and larger memory devices. This solution however, can be quite expensive. In addition, when humans operate imaging systems, presenting multiple images to an individual operator is both cumbersome and places an unreasonable demand on the operator resulting in diminishing performance. Furthermore, in systems, which use a group of operators, integrating visual information across the group is almost impossible [11].

Pixel-level (PL) image fusion algorithms represent an efficient solution to this problem of operator related information overload. By fusing, input signals containing redundant information are condensed to a single representation by choosing the most important features at each image spatial position. Thus, PL fusion effectively reduces the amount of data that needs to be processed without any significant loss of useful information. Additionally, image fusion provides an effective way of integrating visual information across different sensors. To accurately determine spatial alignment

of features in different input images is a near impossible task for human observers either by viewing the images simultaneously or in a sequence. By preserving the spatial characteristics of image features during fusion, the integration of spatial information is achieved by displaying the features in a single fused image. Having to consider only one displayed image at one time significantly reduces the workload of the operator.

Using novel signal processing techniques it is possible to achieve fusion, which results in a reduced loss of input information and a significant reduction in the amount of distortion present in the fused image. In this dissertation, new methods have been proposed for different image fusion application to improve the performance of existing fusion methods towards the theoretical aim of pixel-level image fusion.

### 1.3 Image preprocessing

There are very often some issues that have to be dealt with before the fusion can be performed. Most of the time the images are misaligned. Registration is used to establish a spatial correspondence between the sensor images and to determine a spatial geometric transformation, called warping, which aligns the images. Misalignment of image features is caused by several factors including the geometries of the sensors, different spatial positions of the sensors, different temporal capture rates of the sensors and the inherent misalignment of the sensing elements. Registration techniques align the images by exploiting the similarities between sensor images [10].

The mismatch of image features in multi-sensor images reduces the similarities between the images and makes it difficult to establish the correspondence between the images. There is often a difference in spatial resolution between the images produced by different sensors. There are several techniques to overcome this issue such as the super-resolution techniques [12-13]. The aim of these techniques is to improve the resolution, when possible. Another approach is to use multi-resolution image representations so that the lower resolution imagery does not adversely affect the

higher resolution imagery. This final issue is beyond of the scope of our research, although some of these techniques have been developed to achieve other objectives.

#### **1.4 Project objectives and organization of the thesis**

The objectives of the research work described in this thesis consisting of the following:

- ❖ Proposing new technique for multi-focus image fusion, which improve the performance of existing fusion methods in terms of objective and subjective measures.
- ❖ Introducing new technique for infrared (IR) and visual image fusion, which improve the performance of existing fusion methods in terms of objective and subjective measures.
- ❖ Presenting new technique for multi-spectral (MS) and panchromatic (Pan) image fusion, which improve the performance of existing fusion methods in terms of objective and subjective measures.

The chapter immediately following this introduction provides background information on pixel-level image fusion and discusses the work reported in the past years in various publications in the fields of multi-focus, IR and visual, and remote sensing image fusion.

Chapters 3 is devoted to multi-scale and multi-resolution image transformations, which are one of the tools for image analysis in the novel fusion methods in this thesis. The transformations including the discrete wavelet transform (DWT), dual-tree discrete wavelet transform (DT-DWT), contourlet transform (CT), and shiftable contourlet transform (SCT) that are shortly presented.

Chapter 4 is devoted to the objectives of this project including: 1) a new method for multi-focus image fusion using Fisher classifier and fuzzy logic which is wavelet-based method. 2) A new method for IR and visual image fusion based on fuzzy logic and particle swarm optimization (PSO). 3) A new method for remote sensing MS and

Pan image fusion using SCT and multi-objective particle swarm optimization (MOPSO).

The performance of the proposed image fusion methods compared to the state-of-the-art algorithms will be presented in Chapter 5. Finally, conclusions and recommendations for further work are discussed in chapter 6

## Chapter 2. Literature survey

Multi-sensor image fusion has attracted a considerable amount of research attention in the last ten years. Soon after the introduction of the first multi-sensor arrays in image dependent systems, researchers began considering image fusion as a necessity to solve the growing problem of information overload. Since the end of the 1980s and throughout the 1990s image, and in particular pixel-level, fusion was established as a subject through a stream of publications presenting fusion algorithms.

In this Chapter, due to different applications of image fusion that are in line with the objectives of this dissertation, three separate subsections are considered. In subsection 2.1 the subject of image fusion with different focuses and the literatures are published in this area will be reviewed. Then, subsection 2.2 discuss the subject of IR and visual image fusion as well as the literatures in this area of research. Finally, in subsection 2.3 MS and Pan image fusion and state-of-the-art methods in this field are discussed.

It should be noted that there are similar methods to solve different applications considered in this thesis, but because of the diversity and importance of each of these applications as well as individual goals that we are looking at these applications, separate subsections are considered with possible correlations between them.

## 2.1 Multi-focus image fusion

Due to the limited depth-of-focus of optical lenses in CCD devices, it is often impossible to take an image that contains all objects “in focus”. The depth-of-focus is the range of distance from a camera that is reasonably sharp in the image obtained by that camera. One method to conquer this problem is to take several pictures with different focus points and combine them into a single composite, which contains the focused regions of all input images. This could be useful, for example, in digital camera design or in industrial inspection applications where the need to visualize objects at very short distances makes difficult the preservation of the depth-of-focus [14-15]. Fig. 2.1 shows how combining images with different focuses can effectively extend the depth of field in a visual system.

There are various methods for fusing multi-focus images. We have grouped them into three major categories; however, these categories may cover each other in various ways: 1) spatial domain methods, 2) optimization-based methods, and 3) multi-scale decomposition-based methods.

### 2.1.1 Spatial domain methods

Synthesizing the composite image by averaging the corresponding pixels of the source images is a simple method for image fusion. In addition to simplicity, this method usually causes many undesirable effects on the fused images such as low



Figure 2.1: Multi-focus image fusion. Left to right: wright-focus, left-focus, and fused images [16].

contrast. An optimal weighting can be determined, for example, by a principal component analysis of the correlation or covariance matrix of the source images. The weights for each input image are obtained from the eigenvector corresponding to the largest eigenvalue [17-18]. Variations of this technique and other arithmetic image combinations are numerous [19-20].

Shutao et al. in [21] developed a method based on spatial frequency, which is computationally simple and can be used in real-time applications. The spatial frequency measures the overall activity level in an image. For an  $M \times N$  image, with the gray value at pixel position  $(m, n)$  denoted by  $F(m, n)$ , its spatial frequency is defined as:

$$SF = \sqrt{RF^2 + CF^2} \quad (2.1)$$

where  $RF$  and  $CF$  are the row and column frequencies:

$$CF = \sqrt{\frac{1}{NM} \sum_{i=1}^M \sum_{j=1}^N (F(i, j) - F(i-1, j))^2} \quad RF = \sqrt{\frac{1}{NM} \sum_{i=1}^M \sum_{j=1}^N (F(i, j) - F(i, j-1))^2} \quad (2.2)$$

Fig. 2.2 represents a block diagram for this fusion method, in which the main steps are as follows:

- Decompose the source images into blocks of size  $M \times N$ .
- Compute the spatial frequency for each block.
- Compare the spatial frequencies of two corresponding blocks  $A_i$  and  $B_i$ , and construct the  $i$ th block  $F_i$  of the fused image as:

$$F_i = \begin{cases} A_i & \text{if } SF_i^A > SF_i^B + TH \\ B_i & \text{if } SF_i^A < SF_i^B - TH \\ \frac{(A_i + B_i)}{2} & \text{otherwise} \end{cases} \quad (2.3)$$

- Verify and correct the fusion result in step three with saliency checking. In this case, the purpose of this process is to avoid the isolated blocks instead of isolated points.

An improved version of this technique developed by the same research group based on the use of image blocks and artificial neural networks was suggested in [22]. A block diagram of this improved multi-focus image fusion method is shown in Fig. 2.3. First, the source images are decomposed into blocks. Then a neural network is trained to determine which of the two blocks given, one from each source image, is clearer. Finally, fusion proceeds by selecting the clearer block in constructing the final image. This method uses three features extracted from each block image, which reflect image clarity. The first of the features is the Spatial Frequency, which is, as previously described, used to measure the overall activity level of an image. The second feature is called visibility (VI) and is inspired by the human visual system. Its mathematical formulation is:

$$VI = \sum_{i=1}^M \sum_{j=1}^N \frac{|F(i, j) - \mu|}{\mu^{\alpha+1}} \quad (2.4)$$

where  $\mu$  is the mean intensity value of the image, and  $\alpha$  is a visual constant ranging from 0.6 to 0.7.

Finally, the third and last feature is based on the number of edges extracted from the image. For images of similar complexity, a clearer image is supposed to have more edges. The proposed edge detector method is the Canny algorithm. The total number of one's in the resultant binary image block is then taken as the edge feature.

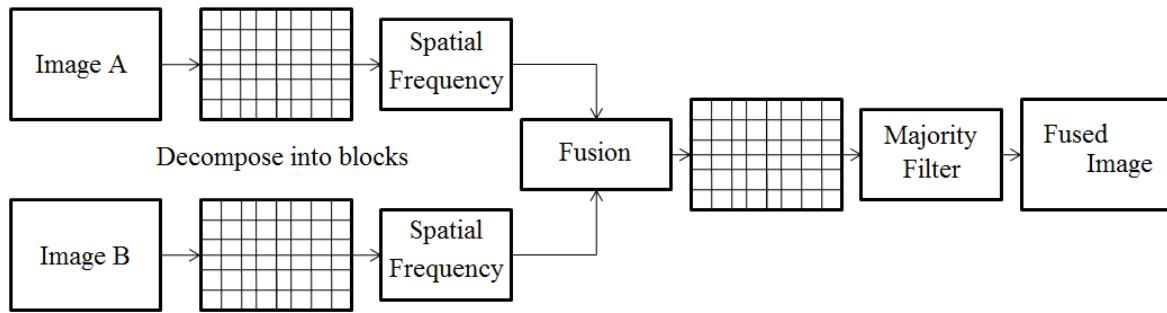


Figure 2.2: Schematic diagram of the spatial frequency method [21].

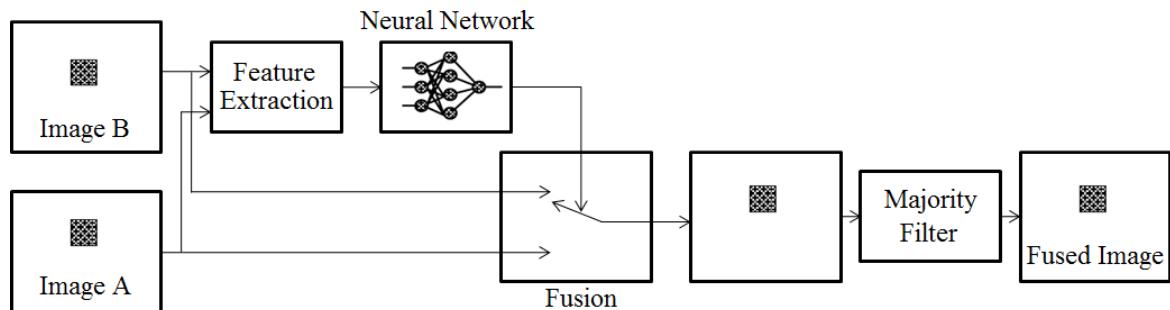


Figure 2.3: Schematic diagram of the improved spatial frequency method [22].

In addition, there are many region-based methods, which combine regions instead of pixels. The region-based methods are reported in [23-24], in which a segmentation map is optioned from source images, and some features in the labeled regions are extracted. By comparing the extracted features between source images, the corresponding regions in the final fused image are obtained. Fig. 2.4 shows a block diagram of the region-based multi-focus image fusion methods.

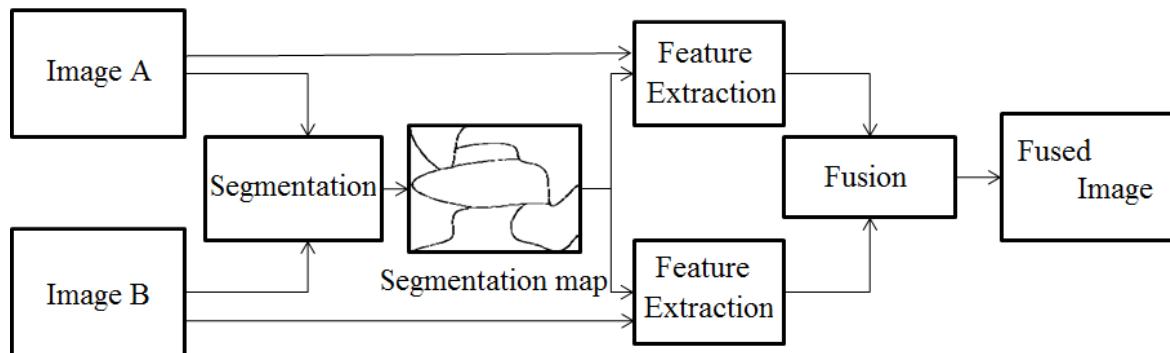


Figure 2.4: Schematic diagram of region-based fusion methods.

### 2.1.2 Optimization methods

This method is based on an *a priori* model of the real scene and the fusion process is stated as an optimization problem [25-27]. In Bayesian optimization, the goal is to find the fused image, which maximizes the *a posteriori* probability. Common procedures contain estimation of the maximum *a posteriori* (MAP) and the maximum likelihood (ML). Let  $a$  denote the set of sensor images and  $s$  the underlying true scene to be estimated. Furthermore, assume  $p(s|a)$  and  $p(a|s)$  denote the conditional probabilities, while  $p(s)$  and  $p(a)$  denote the corresponding marginal probabilities. The MAP estimate chooses to maximize  $p(s|a)$ , and the ML estimate chooses to maximize  $p(a|s)$ . The Bayes formula is as follows:

$$p(s|a) = \frac{p(a|s)p(s)}{p(a)} \quad (2.5)$$

which indicates that the MAP and the ML estimates are closely related [17].

Some examples of probabilistic fusion schemes can be found in [28-29]. In addition, in the Markov random field methods, the input images are first modeled as Markov random fields to define a cost function, which describes the fusion goal. Then, a global optimization strategy such as simulated annealing can be employed to minimize this cost function [30-32].

### 2.1.3 Multi-scale decomposition-based methods

In recent years, many researchers have recognized that multi-scale transforms (MST) are very helpful for analyzing the information content of images for the purpose of image fusion [33-35]. Multi-scale decomposition-based (MSD) image fusion is a biologically motivated method, which fuses multiple images at different spatial resolutions. MSD-based image fusion includes three steps [36]:

- The input images are decomposed into a resolution pyramid of numerous levels. Each level contains one or more bands representing low and high frequency

information.

- Following this decomposition, the coefficients or samples in each band of the source images are fused based on some fusion rules.
- The fused pyramid is then reconstructed to generate the final fused image.

Fig. 2.5 depicts the MSD-based image fusion process. Up to now several algorithms based on multi-scale transforms have been proposed, which are used multi-resolution transformations consisting of the Laplacian pyramid [34], gradient pyramid [38], morphological pyramid [39], and wavelet transform [40]. Research results disclose that the wavelet transform schemes have more advantages over pyramid schemes such as increased directional information, no blocking artifacts, better signal-to-noise ratio, improved perception, and so forth [41]. In recent years, several fusion schemes based on improved wavelet transforms have been developed [42-44]. Hill et al. applied the shift invariant and directionally selective dual tree complex wavelet transform (DT-CWT) to image fusion [44]. The DT-CWT is an over-complete wavelet, which provides both good shift invariance and directional selectivity over the DWT.

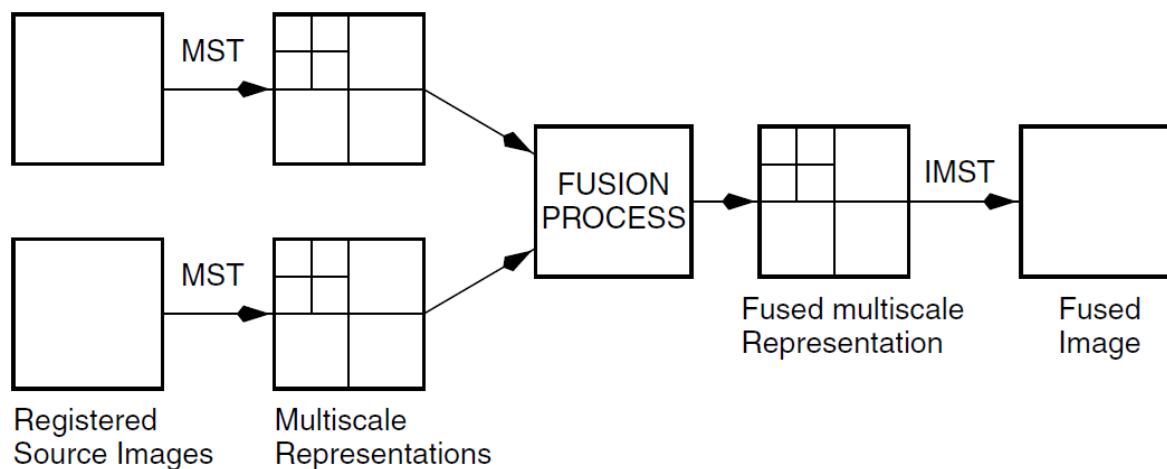


Figure 2.5: Multi-scale fusion block diagram.

In this study, we have focused on the wavelet-based approach, which is a subset of multi-scale decomposition-based methods. There are many reasons for selecting wavelet transform as a tool for image fusion, e.g. low computational complexity and the ability of the wavelet transform to capture important features in a picture. The most important reason is the time-frequency analysis of wavelet transform. Since in focus pixels of an image contain high frequency information, therefore we can simply use this ability of wavelet transform to determine in focus pixels.

In this thesis, we propose a new multi-focus image fusion method, which is also based on multi-scale decomposition. Specifically, we present new fusion rules to merge high and low frequency wavelet coefficients in an appropriate way in order to obtain the best quality in the fused image. Since, the key step in the MSD-based image fusion is the coefficient combination step or the fusion rules, we have only focused on this issue. For this purpose, we have studied different fusion rules and their disadvantages. The three previously important fusion rules, are as follows [44]: The maximum selection (MS), which just picks the coefficients in each sub-band with the largest magnitude; The weighted average (WA), which is proposed by Burt and Kolczynski [45] and used a normalized correlation between the two image sub-bands over a small local area. The resulting coefficients for reconstruction are calculated from this measure via a weighted average of the two images coefficients; The window-based verification (WBV), which is developed by Li et al. [39] and creates a binary decision map to choose between each pair of coefficients using a majority filter.

These fusion rules ignore some useful information and are sensitive to noise. Selective operation renders the fused coefficients completely dependent on the coefficients with larger average of local area energy and ignores the other corresponding coefficients. In the weighted average scheme, the weights were computed by a linear function, which cannot describe the uncertainty of each source image contributions. In addition, in coarser level of decomposition, because of passing through low frequency filter banks, source images become smoother and therefore there are not enough differences between their wavelet coefficients for selecting one of them as in focus coefficients.

## 2.2 Infrared and visible image fusion

In this paper, we have concentrated on the fusion of visible and infrared images. The most important motivation for integrating visible and infrared (IR) images to produce a fused image, which is constructed by a combination of features, and allows improved detection and directly recognizable localization of a target in the IR image with respect to its background in the visible image [46]. Therefore, a human operator using a suitably fused representation of visible and IR images may be able to construct a more complete and accurate mental representation of the perceived scene, and results in a larger degree of situation awareness [6]. Fig. 2.6 illustrates how information from visible and IR images can improve situational awareness in a military campus.

In general, visual and infrared image fusion methods are similar to those for multi-focus image fusion, which were mentioned in the previous subsection. As it has mentioned several algorithms based on multi-scale transforms have been proposed, which use multi-resolution transformations for image fusion. Research results revealed that the wavelet transform schemes have more advantages over pyramid schemes such as increased directional information, no blocking artifacts, better signal-to-noise ratio, improved perception, and so forth [40].

The key step in the MSD-based image fusion is the coefficients combining step or fusion rule. The fusion rule is the process of merging the wavelet coefficients in



Figure 2.6: Fusion of multi-sensor images, from left to right: infrared and visible “UN Camp” images, and fused image.

an appropriate method in order to achieve the best quality in the final fused image. Some general alternatives to construct a fusion rule are illustrated in Fig. 2.7. As it can be seen in Fig. 2.7, the activity level measurement is used to determine the quality of each source image in the transform domain. Grouping and combining methods are also used to achieve the composite multi-scale representation of the fused image. The objective of the consistency verification process is generation of a composite MSD coefficient in a completely similar way from all its neighbors [8]. There are several alternatives in the processes distinguished by the dashed boxes in Fig. 2.7, and different combinations of these alternatives produce different fusion schemes.

Several pixel-based fusion methods exist in the literature [40, 47-48], wherein each individual coefficient of the MR decomposition (or may be the coefficients in a small fixed neighborhood) is treated more or less independently. It is followed by many region-based fusion methods to combine objects instead of pixels [49-52]. For the fusion of IR and visible images, it could be more significant to combine objects rather than pixels, because there are some regions or objects in the IR image that are not visible in the visual image, and vice versa. Additional advantage of the region-based approaches is that the fusion process becomes more robust and this may help avoid some of the well-known disadvantages of pixel-based methods, such as blurring effects, high sensitivity to noise, and image misregistration [50].

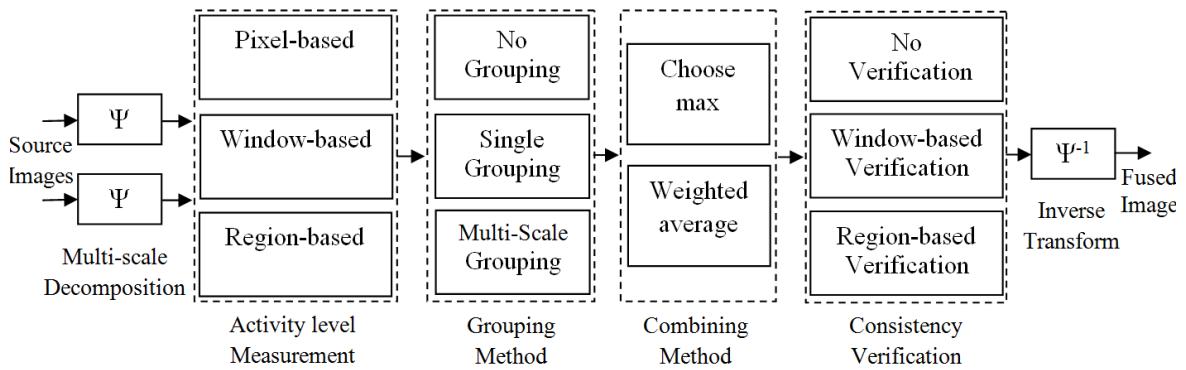


Figure 2.7: A general framework for multi-resolution-based image fusion [8].

Commonly, there are two ways for the region-based fusion methods: maximum selection (MS) and weighted averaging (WA). In the MS scheme, a region from one of the source images is selected based on the activity level measurement, and the other corresponding regions in the remaining images are neglected (e.g. in [49], normalized Shannon entropy is used for activity level measurement). In the WA scheme, first, the activity measurements of similar regions in different source images are obtained, and then the resultant coefficients for reconstruction are calculated from these measures via a weighted average of the coefficients in the similar regions of different sources (e.g. in [52], the weights are obtained based on Alpha-Stable modeling of the wavelet coefficients).

These fusion rules ignore some useful information. Selective operation makes the fused coefficients completely dependent on the coefficients with larger activity in the image regions and ignores other corresponding coefficients from other source images. Besides, in the weighted averaging scheme, choosing one weight as the activity level for all of the coefficients in a region cannot describe the uncertainty of each source image contributions, and results in a low-contrast fused image.

### 2.3 MS and Pan image fusion

Pan-sharpening is a shorthand for panchromatic (Pan) sharpening. It indicates using a Pan image to sharpen a multispectral (MS) image. In this sense, to sharpen means to increase the spatial resolution of an MS image. Therefore, the goal of pan-sharpening is to combine the high spatial resolution of the Pan image with the precise spectral information of the MS image.

An MS image is an image, which contains more than one spectral band. A color image is a very simple example of an MS image, which contains three bands. In this case, these bands correspond to the blue, green, and red wavelengths of the electromagnetic spectrum. In addition, many satellite imaging systems, such as the Quickbird and Landsat-7 satellites produce a Pan image to accompany the MS imagery. This Pan image has higher spatial resolution than the MS data, but the

spectral response spans in a wide range, such as visible or thermal infrared. For example, wavelength of different bands of Landsat-7 satellite is depicted in Fig. 2.8.

Pan-sharpening algorithms depend on the input images being co-registered, because they all perform operations on corresponding pixels in both images. They all do something with the MS pixel and the Pan pixels to create new pixels. If the images are not co-registered, the processing will use the wrong pixels, not the corresponding ones and the result will not look natural. In addition, the MS data is resampled into the same spatial reference and grid as the Pan data, using the nearest neighbor, bilinear, or cubic convolution techniques.

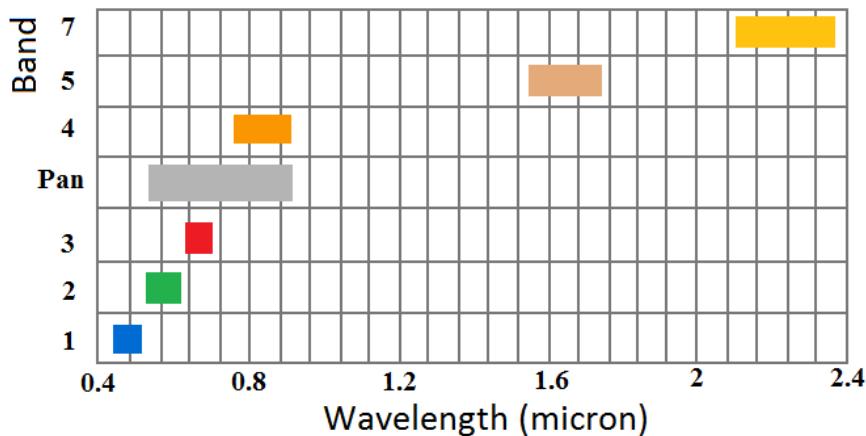


Figure 2.8: Spectral ranges of different bands of Landsat-7 satellite.



Figure 2.9: Remote sensing image fusion. Left to right: Pan, MS with lower resolution, and fused image (2004 Digital Globe).

A ground-truth usually is not available to evaluate the efficiency of a pan-sharpening technique. Therefore, quality assessment should refer to the goals of the fusion process [53]. The most straightforward objective is image-analysis, but also such automated tasks as features extraction and segmentation or classification have been found, which have benefited from the pan-sharpening methods [54-56]. Fig. 2.9 shows an example of MS and Pan image fusion. Here, we categorize pan-sharpening algorithms into the four groups:

### 2.3.1 Color transformation

Intensity-hue-saturation (IHS) is the most common image fusion method for remote sensing applications, and is used in commercial pan-sharpening software. This method converts a color MS image from RGB space to the IHS color space. Here, the I (intensity) band is replaced by the Pan image. Before fusing the images, the MS and the Pan image are histogram matched. HIS-based methods in the literature is too diverse. Tu et al. in [57] proposed a fast method based on IHS transform which reduce color or spectral distortion of fused image. Choi in [58] proposed a similar method, which uses a parameter in order to balance between spatial and spectral resolutions in the final fused image.

### 2.3.2 Statistical methods

Principal component analysis (PCA)-based method is one of well-known technique in this category. In the PCA-based methods, the PCA transform converts intercorrelated MS bands into a new set of uncorrelated components. It is assumed that the first PC image with the highest variance contains the highest amount of information from the original image and will be the ideal choice to replace the high spatial resolution Pan image. All the other new MS bands are unchanged. An inverse PCA transform is performed on the modified Pan and MS images to obtain a high-resolution pan-sharpened image [59]. In addition, a method based on independent component analysis (ICA) is reported in [60].

### 2.3.3 Numerical methods

Simple method of this category includes the Brovey algorithm. The basic process of the Brovey transform is as follows: first, it multiplies each MS band by the high-resolution Pan band, and then divides each product by the sum of the MS bands [61]. More complicated methods of this group are the wavelet-based sharpening algorithms. In the wavelet-based schemes, the detailed information is extracted from the Pan image using wavelet transforms, and will be injected into the MS image. The low frequency sub-band of the pan-sharpened image is usually selected from the low-resolution MS image. There are three ways for fusing high frequency (HF) sub-bands: Substitutive, additive, and weighted methods [62]. In the substitutive scheme, the HF sub-bands of the MS image are replaced with HF sub-bands of the Pan image. In the additive scheme, the HF sub-bands of the Pan image are added to HF sub-bands of the MS image, and finally in the weighted scheme, the HF sub-band of the pan-sharpened image is obtained by a weighted combination of HF sub-bands from the Pan and MS images [63-64].

Ballester et al. in [65] proposed a method called P+XS, which is a variational method, which calculates the pan-sharpened image by minimizing an energy function. It obtains the edge information of the Pan image, using the gradient operator. The spectral information is obtained by approximating the Pan image as a linear combination of the MS bands.

### 2.3.4 Hybrid methods

The Ehlers fusion is the most well-known method of this group that is based on an IHS transform joined with a Fourier domain filtering [66]. First, the IHS transform is applied to the MS bands. Using the fast Fourier transform methods, the spatial components to be enhanced can be directly obtained. The intensity spectrum is filtered with a low pass filter whereas the Pan spectrum is filtered with an inverse high pass filter. After filtering, the images are transformed back into the spatial domain with an inverse FFT and added together to form a fused intensity

component with the low-frequency information from the low-resolution MS image and the high-frequency information from the Pan image. This new intensity component and the original hue and saturation components of the MS image form a new IHS image. Finally, the pan-sharpened image is obtained via an inverse IHS transform.

Similar to the Ehlers method, the hybrid method based on MST, in the first step the IHS transform is applied to extract intensity image from the MS bands. Then it uses the wavelet-based sharpening method for extracting the detailed information from the Pan image to inject into the intensity image. This new intensity component and the original hue and saturation components of the MS image form a new IHS image. Finally, the pan-sharpened image are obtained via an inverse IHS transform [67]. In addition, other combinations of PCA and MST-based method have been proposed [59, 68].

## 2.4 Summary

In the chapter just finished, a review of previous image fusion methods in different application areas was carried out. In short, to obtain images that are all objects within focus, first images of the same scene with different focuses produced, and then these images are combined into one image so that focus areas are preserved. In IR and visual image fusion, the human operator can have more accurate and more complete view from the fused image. To combine images with different focuses and visual and infrared images, the previous algorithms in three different groups were studied: 1) spatial domain methods, 2) optimization methods, and 3) methods based on multi-scale decomposition.

It was also noted that the objective of combining MS and Pan images is enhancing spatial resolution of MS image. The methods in this field were divided into four groups: 1) methods based on color transformation, 2) statistical methods, 3) numerical methods, and 4) hybrid methods.

## Chapter 3. Multi-scale image representation

This Chapter presents a review about multi-scale transformations, which are a basic tool for the research presented in this dissertation. In recent years, multi-resolution transformations have been recognized as a very useful approach to analyze the information content of images for different purposes such as noise reduction [69-70], compression [71], and the subject of this thesis, image fusion. The notion of multi-resolution analysis was initiated by Burt and Adelson [34] who introduced a multi-resolution image representation, called Gauss-Laplacian pyramid. Their under-lying idea is to decompose an image into a set of band-pass filtered component images, each of which represents a different band of spatial frequency. This idea was further elaborated by other researchers such as Mallat [72] and Meyer [73], to establish a multi-resolution analysis for continuous functions in connection with wavelet transformation.

In the following of this Chapter, discrete wavelet transform (DWT) is briefly introduced in subsection 3.1. In subsection 3.2 problems related to DWT and an improved version of it called a dual-tree discrete wavelet transform (DT-DWT) is introduced. Contourlet transform (CT) is explained in subsection 3.3 and shiftable CT is described in subsection 3.4. Finally, conclusions will be provided in section 3.5.

### 3.1 Discrete wavelet transform

The wavelet transform, originally developed in the mid 80's, is a signal analysis tool that provides a multi-resolution decomposition of an image in a biorthogonal basis and results in a non-redundant image representation. This basis are called wavelets, and they are functions generated from one single function, called mother wavelet, by dilations and translations. Although this is not a new idea, what makes this transformation more suitable than other transformations such as the Fourier Transform or the Discrete Cosine Transform, is the ability of representing signal features in both time and frequency domain.

Fig. 3.1 shows an implementation of the discrete wavelet transform (DWT). In this filter bank (FB), the input signal goes through two one-dimensional digital filters. One of them,  $H_0$ , performs a high pass filtering operation and the other  $H_1$  a low pass one. Each filtering operation is followed by subsampling by a factor of two. Then, the signal is reconstructed by first upsampling, then filtering and summing the subbands.

The synthesis filters  $F_1$  and  $F_2$  must be specially adapted to the analysis filters  $H_0$  and  $H_1$  to achieve perfect reconstruction [74]. By considering the z-transfer function of the 2-channel filter bank shown in Fig. 3.1 it is easy to obtain the relationship that those filters need to satisfy. After analysis, the two subbands are:

$$\frac{1}{2} \left[ H_0\left(z^{\frac{1}{2}}\right)X\left(z^{\frac{1}{2}}\right) + H_0\left(-z^{\frac{1}{2}}\right)X\left(-z^{\frac{1}{2}}\right) \right] \quad (3.1)$$

$$\frac{1}{2} \left[ H_1\left(z^{\frac{1}{2}}\right)X\left(z^{\frac{1}{2}}\right) + H_1\left(-z^{\frac{1}{2}}\right)X\left(-z^{\frac{1}{2}}\right) \right] \quad (3.2)$$

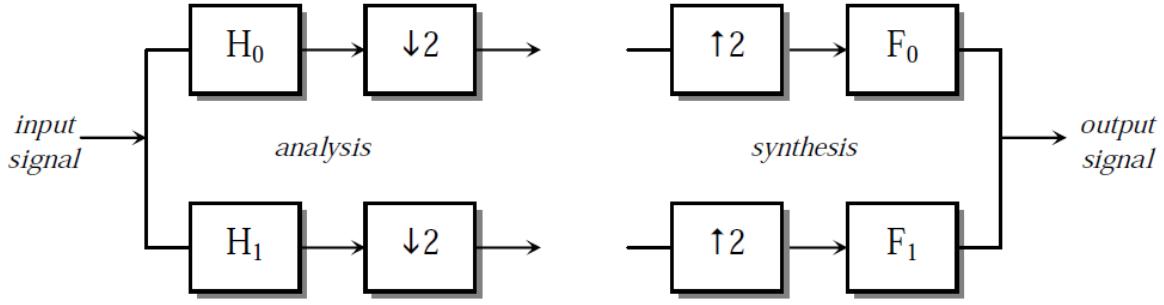


Figure 3.1: Two channel filter bank of DWT

Then, the filter bank combines the channels to get  $\hat{x}(n)$ . In the  $z$ -domain this is  $\hat{X}(z)$ . Half of the terms involve  $X(z)$  and half involve  $X(-z)$ .

$$\hat{X}(z) = \frac{1}{2} [F_0(z)H_0(z) + F_1(z)H_1(z)]X(z) + \frac{1}{2} [F_0(z)H_0(-z) + F_1(z)H_1(-z)]X(-z) \quad (3.3)$$

There are two factors to eliminate including aliasing and distortion. For alias cancellation, choose:

$$F_0(z) = H_1(-z), F_1(z) = -H_0(-z) \quad (3.4)$$

The distortion must be reduced to a delay term, to achieve this Smith and Barnwell suggested [75]:

$$H_1(z) = -z^{-N} H_0(-z^{-1}) \quad (3.5)$$

With these restrictions the final filtering equation is:

$$\hat{X}(z) = \frac{1}{2} z^{-N} [H_0(z)H_0(z^{-1}) + H_0(-z^{-1})H_0(z)]X(z) \quad (3.6)$$

Fig. 3.2 represents one-step in a multi-scale pyramid decomposition of an image [23]. The algorithm applies a one-dimensional high and low pass filtering step to the rows and columns separately in the input image. Successive application of this decomposition to the LL subband gives rise to a pyramid decomposition where the sub-images correspond to different resolution levels and orientations. A sample image decomposed with the wavelet transform is shown in Fig. 3.3.

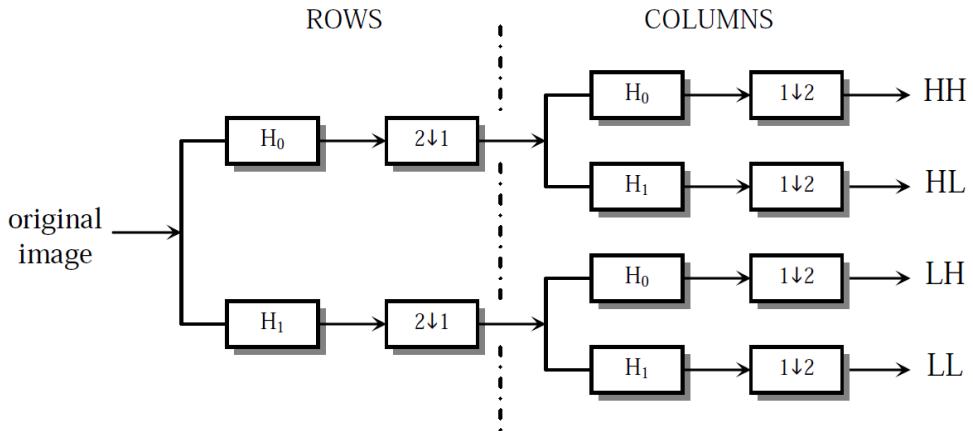


Figure 3.2: Analysis filter bank structure of the DWT.



Figure 3.3: 2-D DWT. Left: Barbara test image and right: DWT using “db2” mother wavelet at two decomposition levels. Small coefficients are shown in black while large coefficients are shown in white

### 3.2 Dual-tree discrete wavelet transform

DWT provides a good time frequency analysis of the signal, with a non-redundant signal representation and an optimal representation of singularities. However, DWT suffers from five fundamental shortcomings [76]: oscillations, aliasing, shift variance, poor directionality, and absence of phase information. Shift invariance and directional selectivity are essential to the quality of wavelet-based

image fusion. Because of the down-sampling operation in the DWT FB, the DWT is not shift invariant, and causes some visual artifacts in the pan-sharpened image. The human visual system is primarily sensitive to moving light stimulus; therefore, moving artifacts are highly distracting to the human observer [77]. In addition, if the directional selectivity of a FB is defined as the ability to extract directional features into separate images, the 2-D DWT has very poor directional selectivity, because 2-D DWT has four sub-images, which are usually referred to as LL, LH, HL, and HH images.

A simple solution to the problem of shift-invariance is to omit the down-sampling blocks in wavelet filter bank. In this case, the sub-bands are completely invariant to shift (undecimated discrete wavelet transform (UDW)). The properties of this new transformation are redundant signal representation, more memory, and high computational cost as compared to the DWT. Another form of shift-invariance called energy- shift-invariance or shiftability is also proposed in [78].

The dual-tree discrete wavelet transform (DT-DWT) is a modified version of the DWT, and was proposed to conquer shift variance and directionality limitations of the DWT while maintaining the perfect reconstruction property with limited redundancy [79-80]. The DT-DWT is composed of two parallel DWT filter bank trees. The wavelet and scaling functions used in one tree can be defined as approximate Hilbert transforms of the functions in the other tree. The filters used in both trees are real, but the combined filters are referred to as analytic.

The DT-DWT employs two real DWTs, the first one (tree *a*) gives the real part of the transform while the second one (tree *b*) gives the imaginary part. The successful operation of the DT-DWT is owed to the differences between the filters in the two trees, in which the DT-DWT is achieved by a simple delay of one sample between the level 1 filters in each tree, and for subsequent levels, by the use of alternate odd-length and even-length linear-phase filters. Unfortunately, the odd/even filter approach suffers from certain problems [80]: (a) the sub-sampling structure is not very symmetrical, (b) the two trees have slightly different frequency responses, and (c) the filter sets must be bi-orthogonal. Kingsbury proposed a Q-

shift dual tree to overcome all problems above [81]. The key to designing filters for the Q-shift version of the DT-DWT lies in finding a good even length lowpass filter with a delay of  $1/4$  sample which also satisfies the standard orthonormal perfect reconstruction condition of two-band filter banks [80].

As shown in Fig. 3.4 all the filters beyond level one are even length, but they are no longer strictly linear phase. They are designed to have a group delay of approximately  $1/4$  sample ( $q$ ). The required delay difference of  $1/2$  sample ( $2q$ ) is then achieved by using the time reverse of the tree a filters in tree b. Furthermore, all filters beyond level 1 are derived from the same orthonormal prototype set. This leads to a more symmetric sub-sampling structure, but which preserves the key advantages of DT-DWT that are approximate shift invariance and good directional selectivity. The symmetry of the sub-sampling process is important for hierarchical algorithms, which relate wavelet coefficients at one level to those at the same spatial location at levels above or below. Therefore, we decompose the input multi-focus images with Q-shift CWT. Here, The (9,7) filters were chosen for level 1 as a good compromise between smoothness and complexity (a much closer approximation to orthonormality at level 1 may be obtained from the (13,19)-tap filters) and Q-shift (14,14)-tap filters as the levels beyond.

This combination led to a complex extension of real signals. As the complex wavelets can distinguish between positive and negative frequencies, the diagonal sub-bands can be discriminated from horizontal and vertical sub-bands. Later on, horizontal and vertical sub-bands are divided giving six distinct sub-bands at each scale (at orientation  $\pm 15^\circ$ ,  $\pm 45^\circ$ , and  $\pm 75^\circ$ ).

Directional subbands of DT-DWT are illustrated in Fig. 3.5 for a sample image, in which the improvement for directional feature extractions is evident compared to DWT (Fig. 3.3)).

In this thesis, the DT-DWT of an image  $x$  is denoted by  $y$  and is assumed in the different scales to be of the form:

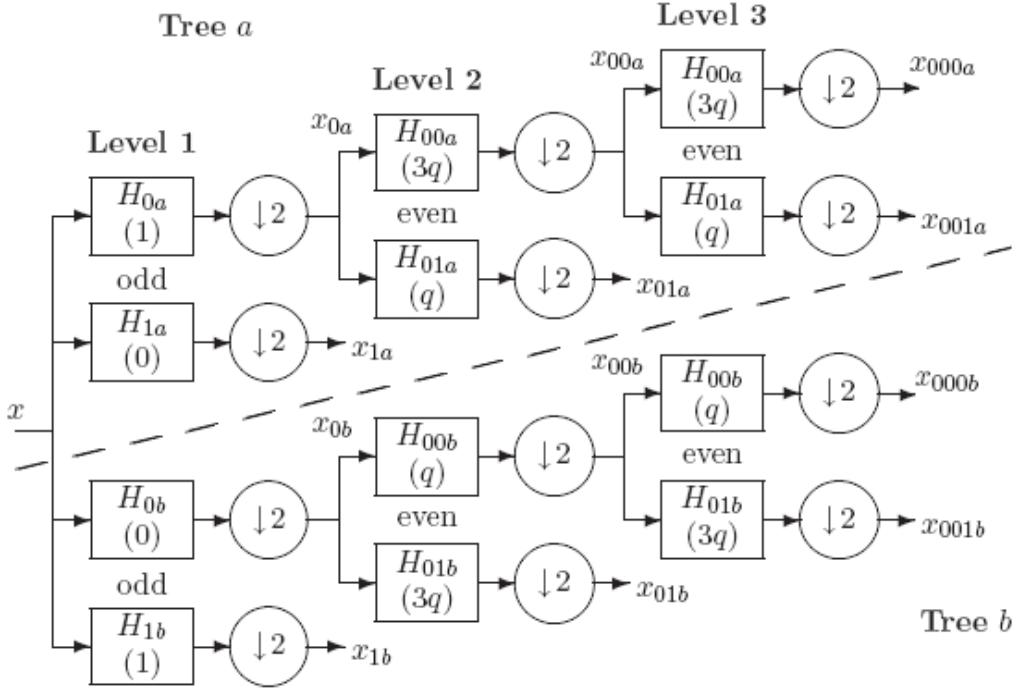


Figure 3.4: The Q-shift version of the DT CWT, giving real and imaginary parts of complex coefficients from tree  $a$  and tree  $b$  respectively. Figures in brackets indicate the delay for each filter, where  $q = 1/4$  sample period [81].

$$y = \{y^1, y^2 \dots y^L, x^L\} \quad (3.7)$$

where  $x^L$  represents the approximation or low frequency sub-bands at the last decomposition level, while  $y^l$  represents the details or high frequency sub-bands at level  $l$ . In addition,  $y^l$  is composed of twelve directional sub-bands, six of which are real, and six are imaginary:

$$y^l = \begin{cases} \text{real} & y_{\text{real}}^l(i, j | 1), y_{\text{real}}^l(i, j | 2), \dots, y_{\text{real}}^l(i, j | 6) \\ \text{imaginary} & y_{\text{imag}}^l(i, j | 1), y_{\text{imag}}^l(i, j | 2), \dots, y_{\text{imag}}^l(i, j | 6) \end{cases} \quad (3.8)$$



Figure 3.5: Improvement for directional features extraction in DT-DWT, Top: DT-DWT subbands at two decomposition levels, and bottom: DT-DWT transform of Barbara test image. Small coefficients are shown in black while large coefficients are shown in white.

We use the coordinates  $(i, j)$  or the shorthand notation of  $(.)$ , to index the spatial position of the coefficients. It should be mentioned that we have used the DT-DWT toolbox version 4.3 provided by Kingsbury<sup>1</sup> for implementation.

### 3.3 Contourlet transform

Natural images are not simply stacks of 1-D piecewise smooth scan-lines; discontinuity points (i.e. edges) are typically located along *smooth* curves (i.e. contours) owing to smooth boundaries of physical objects. Thus, natural images contain *intrinsic geometrical structures* that are key features in visual information. As a result of a separable extension from 1-D bases, wavelets in 2-D are good at

<sup>1</sup> <http://www.eng.cam.ac.uk/~ngk/>

isolating the discontinuities at *edge points*, but will not “see” the smoothness along the *contours*. In addition, separable wavelets can capture only limited *directional* information – an important and unique feature of multidimensional signals. These disappointing behaviors indicate that more powerful representations are needed in higher dimensions.

The idea of combining the directional filter bank (DFB) and a multi-resolution decomposition to create a multi-resolution multidirectional decomposition is suggested in [82]. In particular, Do and Vetterli built the theory of the Contourlet transform from the pyramidal DFB in [83].

Contourlets provide a new system representation for image analysis. The contourlet name is so called because of its ability to capture and link the point of discontinuities to form a linear structure (contours). The two-stage process used to derive the contourlet coefficients involves a multi-scale transform and a local directional transform. The point of discontinuities and multi-scale transformation is obtained via the Laplacian pyramid [34]. The local directional filter bank is used to group these wavelet-like coefficients to obtain a smooth contour. Contourlets provide  $2l$  directions at each scale, where  $l$  is the number of required orientation. This flexibility of having different numbers of direction at each scale makes contourlets different from other available multi-scale and directional image representation, which includes 2-D Gabor wavelets [78], the cortex transforms [84], and the steerable pyramid [85]. Similar to wavelets, contourlets also have different implementation of the subsampled and nonsubsampled transforms.

The discrete CT is developed in the discrete domain using the fast-iterated nonseparable filter banks having an order of  $N$  operations for  $N$ -pixel images. The transformation stage includes two filter banks: the Laplacian pyramid to generate multi-scale decomposition and the directional filter bank to reveal directional details at each decomposition level as illustrated in Fig. 3.6.

The subsampled CT consists of down-sampling the image by a factor of  $M$  at each level of transformation. The 2-D filters are obtained from mapping 1-D filters. For perfect reconstruction, the 1-D filters satisfy the condition [86]:

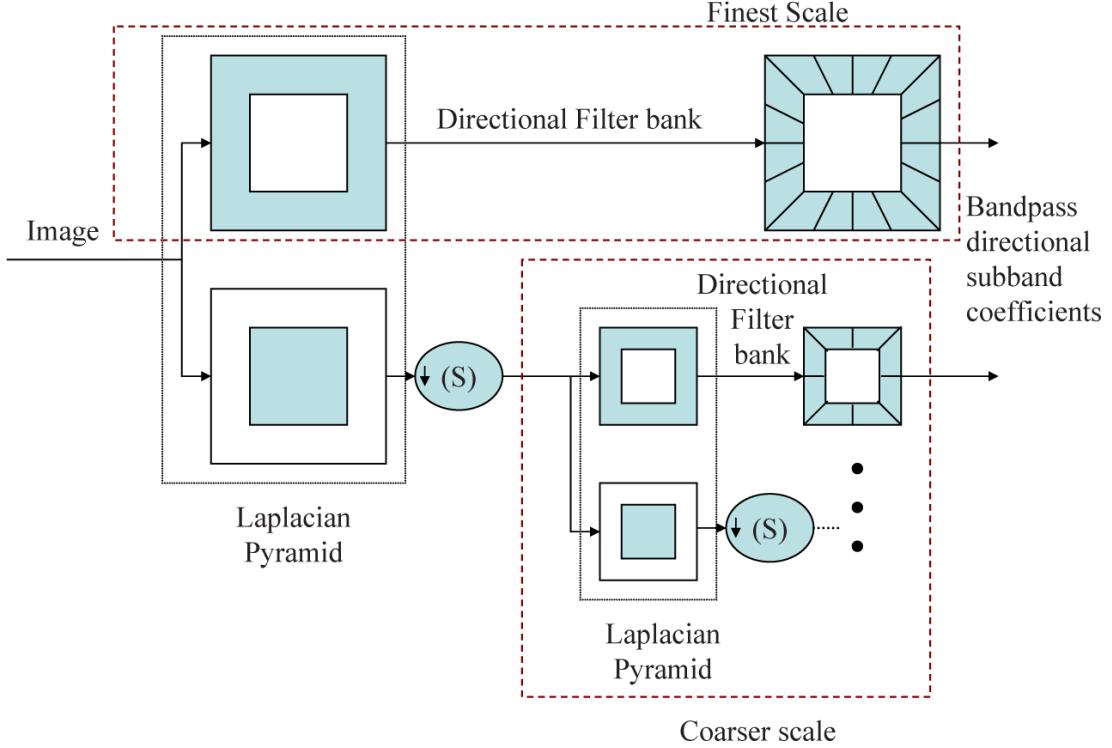


Figure 3.6: Discrete contourlet transform (This figure is a modified version of [83])

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 2 \quad (3.9)$$

where  $H_0(z)$  and  $H_1(z)$  represent the low-pass and high-pass analysis filters, whereas  $G_0(z)$  and  $G_1(z)$  are the low-pass and high-pass synthesis filters.

The 2-D perfect reconstruction is achieved by choosing the mapping filter  $M(z)$  to satisfy the condition:

$$H_0(M(z))G_0(M(z)) + H_1(M(z))G_1(M(z)) = 2 \quad (3.10)$$

$M(z)$  has a separable property in the polyphase domain, which reduces the filter complexity from  $O(N^2)$  to  $O(N)$  [86].

The 2-D directional filter bank (DFB) is constructed in two steps. In the first step, the frequency spectrum is divided into horizontal and vertical directions using a two-channel quincunx filter bank. This is followed by the shearing operator. An appropriate choice of shearing operation and the two-direction partition of quincunx filter banks provide the required directional division of the 2-D spectrum

(Figs. 3.7 and 3.8). An example of CT is shown in Fig. 3.9. As it can be seen, only contourlets that match with both location and direction of image contours produce significant coefficients.

### 3.4 Shiftable Contourlet transform

Similar to the DWT, the CT is not shift-invariant and results in artifacts along the edges to some extent. The Nonsubsampled Contourlet Transform (NSCT) was proposed to meet the shift invariance. The NSCT is the shift-invariant version of the CT and is built based on nonsubsampled pyramid decomposition and nonsubsampled filter banks, to obtain the shift invariance [87-88]. But, the NSCT has the overcomplete ratio of  $N \times K$ , where  $N$  is the number of resolution levels and  $K$  is the number of directions, which is very high compared to the overcomplete

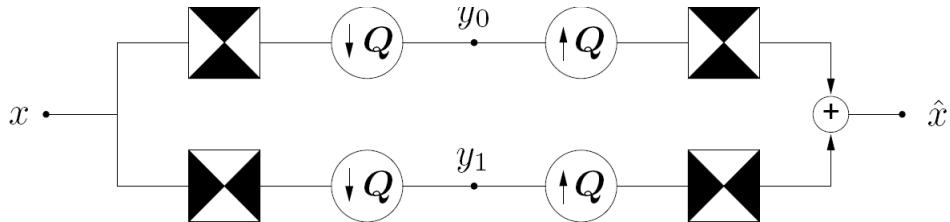


Figure 3.7: Two-dimensional spectrum partition using quincunx filter banks with fan filters. The black regions represent the ideal frequency supports of each filter.  $Q$  is a quincunx sampling matrix [83].

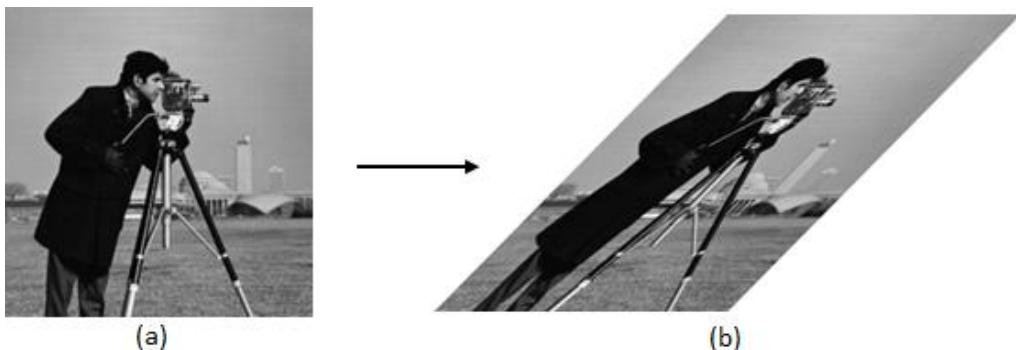


Figure 3.8: Example of shearing operation that is used like a rotation operation for DFB decomposition. Left: The “cameraman” image and right: The “cameraman” image after a shearing operation [83].

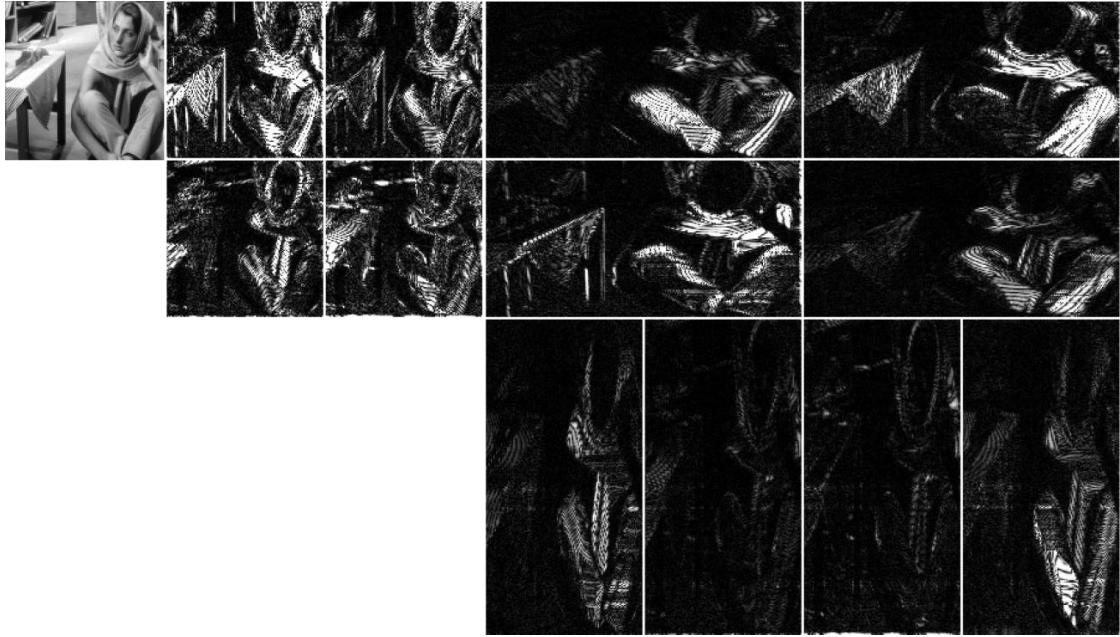


Figure 3.9: Examples of the contourlet transform on the *Barbara* image. For clear visualization, each image is only decomposed into two pyramidal levels, which are then decomposed into four and eight directional subbands. Small coefficients are shown in black while large coefficients are shown in white.

ratio of the CT (4/3). However, a reduced form of translation invariance exists, namely, energy shift-invariance or “shiftability” [78], which means that the energy of the output signal is shift invariant. As for these reasons, we use the SCT, which is shown to have a number of desirable properties for image analysis including shiftable sub-band, arbitrarily high directionality, and low redundancy (11/3) [89].

The shiftable contourlet transform or shiftable complex directional pyramid is a novel image transform, which provides a directional and shiftable image representation, as shown by its ideal construction in Fig. 3.10. The forward and inverse transforms are carried out by the analysis and synthesis pyramidal dual-tree directional filter bank (PDTDFB). The PDTDFB is a combination of an iterative multi-resolution FB and a dual-tree of DFB at each resolution, as presented in Fig. 3.10.

In order to construct a shiftable multi-scale and multidirectional decomposition, a combination of a multi-resolution FB with the dual-tree DFB at high frequency is

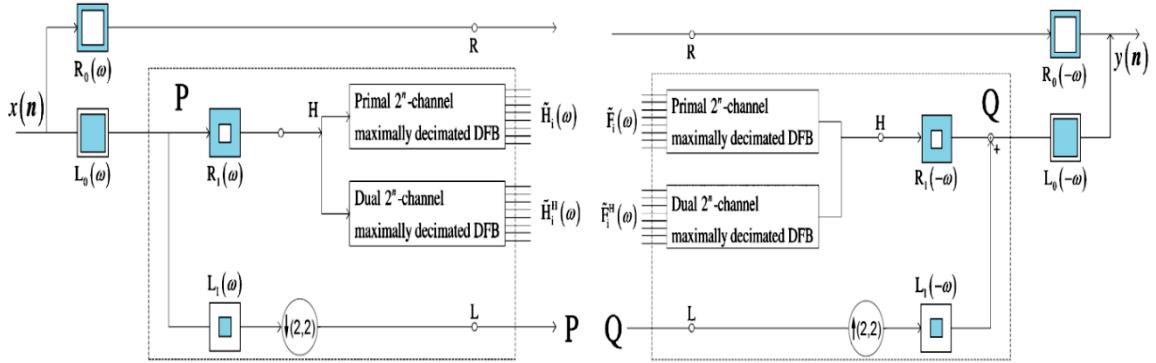


Figure 3.10: The structure of the shiftable complex directional pyramid, analysis (left) and synthesis (right) sides [89].

proposed by Nguyen and Oraintara [89]. The combination of the multi-resolution FB and the dual-tree DFB has a double objective. First, the multi-scale FB provides a multi-resolution image decomposition. Secondly, it can also remove the remaining aliasing components in the dual-tree DFB. In order to keep the complexity low, the FB should be efficiently implemented using separable filtering. The proposed multi-scale FB consists of an undecimated two channel FB, and an iterated 2-D multi-resolution FB having decimation at the low-pass branch. Consider the construction in Fig. 3.10. At the front end, an undecimated two-channel FB [ $L_0(z)$  and  $R_0(z)$ ] is used to filter out the high frequency components near  $(\pm\pi, 0)$  and  $(0, \pm\pi)$ , which can potentially cause aliasing in the dual-tree. The high-pass filter produces a 'residual' image similar to that in the steerable pyramid. It is clear that, for this undecimated FB to be PR, the filters must satisfy:

$$|R_0(\omega)|^2 + |L_0(\omega)|^2 = 1 \quad (3.11)$$

The output of the wideband low-pass filter  $L_0(z)$  is then fed into the first stage of the multi-resolution pyramidal FB where the signal is divided into two parts: the coarse approximation [point L in Fig. 3.10] and the high frequency component [point H in Fig. 3.10]. This high frequency component is then further decomposed by a dual-tree of DFBs to produce the real and imaginary parts of the  $2^n$  complex directional sub-bands. This decomposition (block P) is also repeated at the low

frequency output (point L), forming a pyramid. The filters in blocks P and Q are designed to satisfy the PR and non-aliasing conditions:

$$|R_1(\omega)|^2 + \frac{1}{4}|L_1(\omega)|^2 = 1, L_1(\omega_1, \omega_2) = 0 \quad (3.12)$$

where  $\min(|\omega_1|, |\omega_2|) > (\pi/2)$ , and the factor 1/4 in the first equation is to ensure that the two-channel FB is PR.

The PDTDFB takes the shiftable pyramidal FB discussed above and applies a dual-tree  $2^n$ -channel DFB at the high-pass output of each level. The first level of the PDTDFB is illustrated in Fig. 3.10. The blocks  $P$  and  $Q$  are iterated to provide a multi-scale decomposition and synthesis. Based on the construction of the PDTDFB, the image representation produced by the structure in Fig. 3.10 is shiftable. The DFBs employed in the PDTDFB are constructed using the proposed structure by Do and Vetterli [83]. The primal and dual DFBs of the dual-tree are identical at every level except for the second one where the filters of the two-channel fan FBs in the dual DFB have to satisfy the phase conditions. By this construction, the resulting directional filters in the primal and dual DFBs form Hilbert transform pairs. The conventional DFB is constructed by using a binary tree of two-channel FBs [90]. These two-channel FBs can be obtained from one prototype fan FB if appropriate resampling blocks, which are up or down-sampling blocks having determinant one, are used in the tree [91]. The  $2^n$ -channel dual-tree DFB is constructed from four-channel dual-tree DFB by cascading two-channel FBs, using the same prototype fan FBs and resampling blocks. Therefore, each pair of directional filters in the primal and dual  $2^n$ -channel DFBs still form a Hilbert transform pair.

The image decomposition by the PDTDFB is thus a multi-resolution and directional transform whose basis functions are the sub-band filters' impulse responses. Therefore, the PDTDFB can be regarded as a shiftable Contourlet transform (SCT). The SCT is shown to have a number of desirable properties for

image analysis including shiftable sub-band, arbitrarily high directionality, and low redundancy. Since the decimation ratio after  $L_1(s)$  in each level is  $(1/4)$ , the upper bound of the overcomplete ratio for the case of single (maximally decimated) DFB is  $1+(1/4)+(1/6) = (4/3)$ . Since there are two DFBs in each level, the overcomplete ratio is then bounded by  $2\times(4/3) = (8/3)$ . Taking into account the residual sub-band at the output of filter  $R_0(z)$ , the total overcomplete ratio of the decomposition is  $1+(8/3) = (11/3)$  [89].

### 3.5 Summary

In the Chapter just finished, we discussed about some of important multi-scale image transformations. As we have stated DT-DWT is an improved version of DWT, which has higher angular resolutions and it is also shift invariant. Then, CT is studied to extract the geometry of the image, i.e. the contours and its shiftable version, i.e. SCT were described. Table 3.1 provides a comparison between various transformations in terms of shiftability and data redundancy.

Table 3.1: Comparison between different multi-scale transformations (N is decomposition level, and K is number of direction in each level)

Filter Bank	data redundancy	shiftability
<b>DWT</b>	1	Variant
<b>UDWT</b>	$3\times N+1$	Linear invariant
<b>DT-DWT</b>	4	Approximate invariant
<b>CT</b>	$4/3$	Variant
<b>NSCT</b>	$N\times K$	Linear invariant
<b>SCT</b>	$11/3$	Approximate invariant

# Chapter 4. The proposed methods for image fusion

In this Chapter, the proposed algorithms for different applications of image fusion will be described. In previous Chapter, different multi-scale transformations for image representation were briefly reviewed that are the basis for the proposed algorithms in this thesis. In fact, in this study, we have used the DT-DWT for fusing multi-focus and IR and visible image fusion. In addition, the SCT is used for MS and Pan image fusion or pansharpening. The choice of multi-scale transform is related to the proposed algorithm and the objectives of respected application.

In the following, in subsection 4.1, a new method based on Fisher classifier and fuzzy logic is presented to fuse images with different focuses. In subsection 4.2, a new algorithm based on fuzzy logic and particle swarm optimization is proposed for IR and visible image fusion. In subsection 4.3, multi-objective particle swarm optimization is used for remote-sensing image fusion. Finally, a summary is provided in subsection 4.4.

## 4.1 Multi-focus image fusion using fuzzy logic and classification

In this study, we aim to reach best performance that a wavelet-based method can produce. The best performance is related to manually picking in focus wavelet coefficients from the transformed source images using hand-made decision map to generate best possible fused image in the wavelet domain. Therefore, we present a new method for merging high and low frequency wavelet coefficients as a fusion

rule using Fisher classifier and fuzzy logic. We use the inter-scale dependency between wavelet coefficients in the DT-DWT sub-bands to obtain a reliable decision map. First, a twelve dimensional feature vector is obtained using six directional sub-bands of DT-DWT in the first decomposition level of the source images. Then, this feature vector is projected to one-dimensional space using a trained Fisher classifier. We use the classifier output as a decision map for selecting high frequency wavelet coefficients between two source images in the different directions and decomposition levels, equally [92-94].

In addition, there is an uncertainty about selecting high frequency wavelet coefficients in the smooth regions of two images, which causes some misclassified regions in the classifier output. In order to solve this uncertainty and integrate as much information of each source image as possible into the fused image, we propose an algorithm based on fuzzy logic, which combines the outputs of two different fusion rules: Selection based on decision map, and weighted averaging. The classifier output is also used for selecting low frequency wavelet coefficients between the source images instead of simple averaging.

For testing our new fusion rule, we use DT-DWT, which introduces limited redundancy and allows the transform to provide approximate shift invariance and directionally selective filters while preserving the usual properties of perfect reconstruction and computational efficiency. In the following, the proposed fusion algorithm using the Fisher classifier and fuzzy logic is presented.

Fig. 4.1 shows the block diagram of the proposed method, which consists of some essential stages:

- 1) The source images are decomposed into different directions and scales using DT-DWT.
- 2) A twelve dimensional feature vector is extracted using the directional sub-bands of DT-DWT in the first decomposition level.
- 3) A reliable decision map is obtained using a trained Fisher classifier.
- 4) The high frequency wavelet coefficients of the source images are integrated

using fuzzy logic, which combines outputs of two different fusion rules (selection based on decision map, and weighted averaging), based on a dissimilarity measure of the source images.

- 5) The low frequency wavelet coefficients of the final fused image are selected from the source images using a down sampled version of the decision map, which is obtained in step 3.
- 6) The inverse DT-DWT of the new low and high frequency wavelet coefficients generates the final fused image.

In the following subsections, we have provided more detailed explication of the image fusion process.

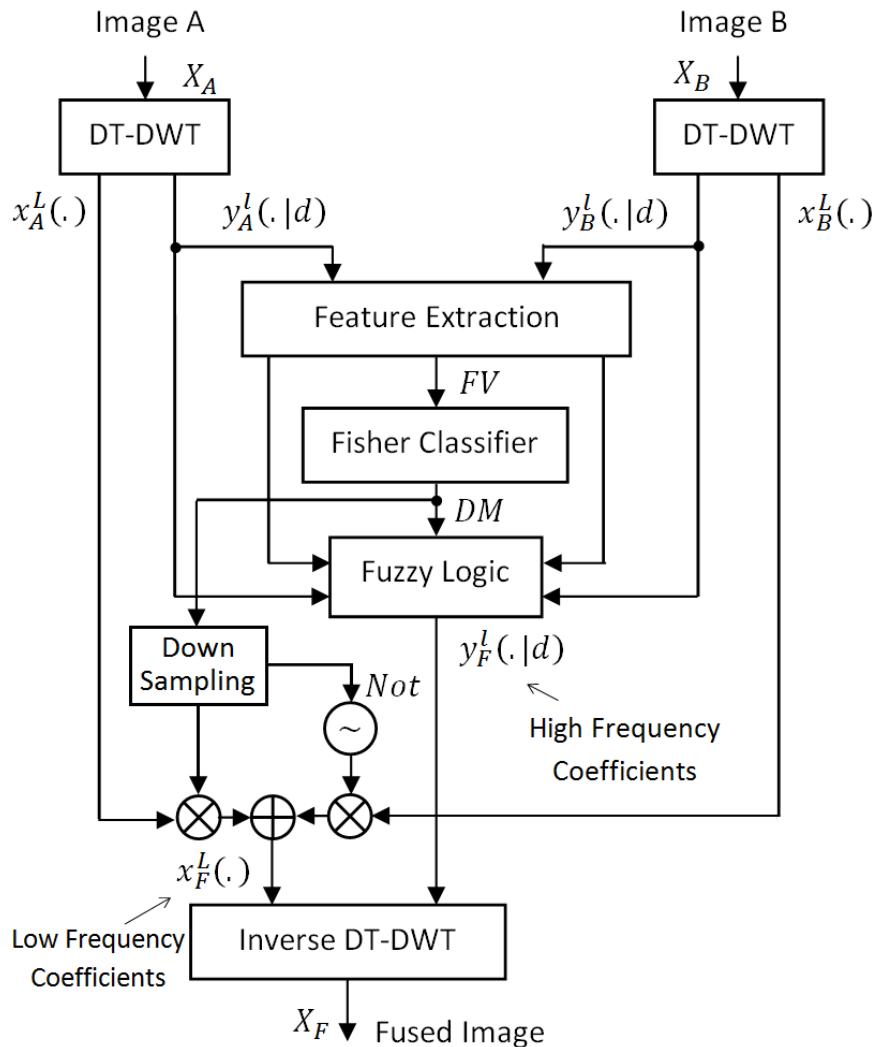


Figure 4.1: Block diagram of proposed image fusion algorithm.

### 4.1.1 Feature extraction

The high frequency wavelet coefficients reflect the image edges and detailed information. Based on the mechanism of imaging in the optical system, the bandwidth of system function for in focus images is wider for out-of-focus images [95]. Therefore, the absolute values of high frequency wavelet coefficients of clear images are larger than blurred images.

Using the information of high frequency wavelet coefficients is a straightforward approach to distinguish between in and out of focus regions; however, we can use some local features (such as energy, mean, standard deviation, and normalized Shannon entropy), to enhance the information of wavelet coefficients. Commonly, absolute values of wavelet coefficients are used for calculating local features as follows:

$$\text{energy}(i, j) = \frac{1}{W} \sum_{n=-N}^N \sum_{m=-M}^M |y(i+n, j+m)|^2 \quad (4.1)$$

$$\text{mean}(i, j) = \frac{1}{W} \sum_{n=-N}^N \sum_{m=-M}^M |y(i+n, j+m)| \quad (4.2)$$

$$\text{std}(i, j) = \left[ \frac{1}{W} \sum_{n=-N}^N \sum_{m=-M}^M (|y(i+n, j+m)| - |\bar{y}|)^2 \right]^{1/2} \quad (4.3)$$

$$\text{entropy}(i, j) = -\sum_{l=0}^L h(l) \times \log_2(h(l)) \quad (4.4)$$

where  $y(i, j)$ 's are the wavelet coefficients at the position of  $(i, j)$ ,  $(2N + 1, 2M + 1)$  is the size of local window,  $W$  is the number of coefficients in the local window,  $h$  is normalized histogram of absolute values of wavelet coefficients in a local window at the position of  $(i, j)$ , and  $L$  is the number of levels.

Here, we use the magnitude of complex wavelet coefficients at the first decomposition level for calculating the local features:

$$y_{abs}^1(\cdot | d) = \sqrt{y_{real}^1(\cdot | d)^2 + y_{imag}^1(\cdot | d)^2} \quad (4.5)$$

where  $y_{real}^1(\cdot | d)$  and  $y_{imag}^1(\cdot | d)$  are the real and imaginary wavelet coefficients of the DT-DWT at the first level and orientation  $d$ , which can be  $\pm 15^\circ$ ,  $\pm 45^\circ$ , and  $\pm 75^\circ$ .

After extracting one of the mentioned local features using the magnitude of complex wavelet coefficients, it is followed by nonlinear averaging in the local window for taking into account neighbor dependency and smoothing on the feature space:

$$NF(i, j) = \sum_{n=-N}^N \sum_{m=-M}^M G(n, m) \times F(i+n, j+m) \quad (4.6)$$

where  $F$  is the local feature,  $G(n, m)$  is the local window's weight, which is obtained by a Gaussian filter.

We use two different local windows ( $3 \times 3$ , and  $7 \times 7$ ), to calculate the local features in the six sub-bands of DT-DWT of source images. Calculating the features using two different local windows can reduce the existent uncertainty in the border of in and out of focus regions. Therefore, twelve feature images are obtained for each source image using two local windows in the six directional sub-bands of DT-DWT. Then, the differences between extracted features of the source images are used as a feature vector for the classification problem:

$$FV_{LW}^d = NF_{LW}^d(A) - NF_{LW}^d(B) \quad (4.7)$$

where  $A$  and  $B$  are the source images,  $NF$  is obtained using (4.6),  $d$ 's are the six high frequency directional sub-bands of DT-DWT, and  $LW$  are the two local windows.

Finally, the twelve dimensional features vector is defined as follows:

$$FV = (FV_{3 \times 3}^1, FV_{3 \times 3}^2, \dots, FV_{3 \times 3}^6, FV_{7 \times 7}^1, FV_{7 \times 7}^2, \dots, FV_{7 \times 7}^6) \quad (4.8)$$

Thus, at the end of feature extraction step we have a stack of twelve transformed images, which is obtained using six high frequency directional sub-bands of DT-DWT of two source images with different focus points.

#### 4.1.2 Fisher classifier

Having the feature vector, for classification of wavelet coefficients as either in focus or out of focus, we use the Fisher classifier. Compared with the Neural Network (NN) and the Support Vector Machine (SVM), the Fisher classifier is easier to train, faster to classify, needs fewer training samples, and does not suffer from overtraining problems [96-97].

For a feature vector  $X$ , the Fisher classifier projects  $X$  into one dimension  $Y$  in direction  $W$  using:

$$Y = W^T X \quad (4.9)$$

The Fisher criterion finds the optimal projection direction  $W_0$  by maximizing the ratio of the between-class scatter to the within-class scatter. Let  $S_w$  and  $S_b$  be the within- and between-class scatter matrices respectively,

$$S_w = \sum_{k=1}^K \sum_{x \in class K} (x - u_k)(x - u_k)^T \quad (4.10)$$

$$S_b = \sum_{k=1}^K (u_k - u_0)(u_k - u_0)^T \quad (4.11)$$

$$u_o = \sum_{k=1}^K u_k \quad (4.12)$$

where  $u_k$  is the mean vector of the  $k$ th class,  $u_0$  is the global mean vector, and  $K$  is the number of classes.

The optimal projection direction is the eigenvector of  $S_w^{-1}S_b$  corresponding to its largest eigenvalue [98]. For a two-class classification problem, we do not need to

calculate the eigenvectors of  $S_w^{-1}S_b$ . It is shown that the optimal projection direction is [98]:

$$W_o = S_w^{-1}(u_1 - u_2) \quad (4.13)$$

Fig. 4.2 shows train images and their corresponding class labels for training Fisher classifier (black regions indicate out of focus area and white regions indicate in focus area for the first source image and vice versa). Let  $Y_1$  and  $Y_2$  be the projections of training images for in and out-of-focus classes to the optimal projection direction  $W_0$ , and let  $E[Y_1]$  and  $E[Y_2]$  be the means of  $Y_1$  and  $Y_2$ , respectively. Suppose  $E[Y_1] > E[Y_2]$ , then for a pair of test images with feature vector  $X$ , the decision can be made as:

$$Y = W_0^T X \quad (4.14)$$

Then,

$$C(X) = \begin{cases} \text{class 1} & \text{if } Y > \frac{E[Y_1] + E[Y_2]}{2} \\ \text{class 2} & \text{otherwise} \end{cases} \quad (4.15)$$

It is known that, if the feature vector  $X$  is jointly Gaussian distributed and the two classes have the same covariance matrices, then the Fisher classifier is optimal in a minimum classification error sense [98].

Table 4.1 shows classification accuracy using different local features (Mean, Energy, Normalized Shannon Entropy, and Standard Deviation) for test images. It should be mentioned that for calculating the classification accuracy, the first handmade classification maps are extracted from different multi-focus images using cut and paste method. Then, these handmade classification maps are compared with the classification outputs for different test images. As it can be seen in Table 1, the best result is related to the standard deviation as the local feature for the classification problem. Also using two local features, while increasing the

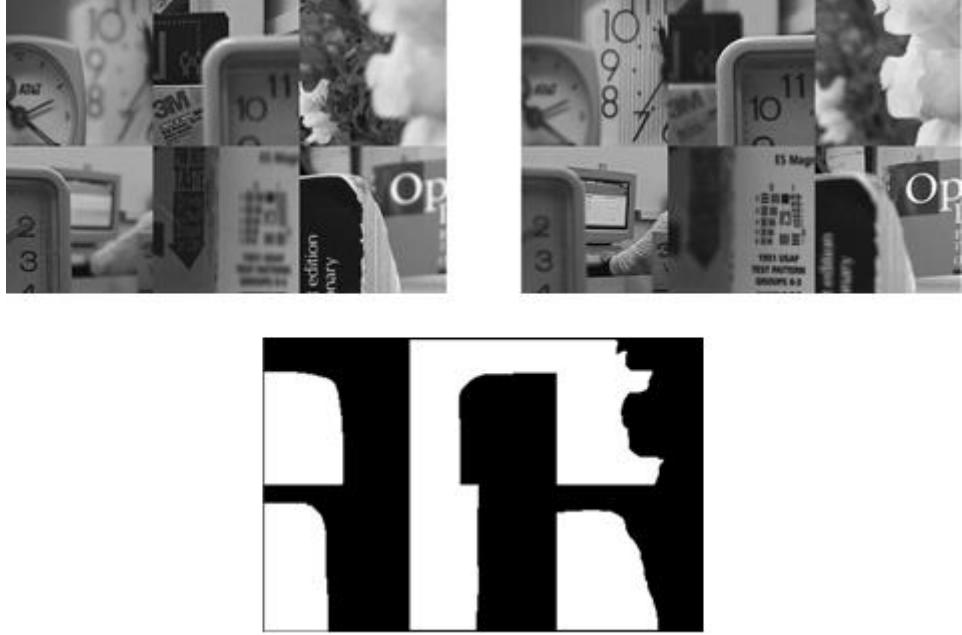


Figure 4.2: Top: Training images used in the experiment, and bottom: class labels (black regions indicate out-of-focus areas and white regions indicate in focus areas for the first training image and for the second training image, black regions indicate in of focus areas and white regions indicate out-of-focus areas).

complexity, is not effective for the classification accuracy. The handmade classification maps and the classification results using the standard deviation as the local feature, for a number of multi-focus images are shown in Fig. 4.3.

After completion of classification, there may existed some misclassified pixels. In order to solve this problem, we have used post-processing. Li in [39] applied a majority filter for this purpose. In the Li method, if the number of 1's pixels in a local window (e.g.  $5 \times 5$ ) around a central pixel in the binary decision map are more than 0's pixels, then the value of central pixel is set to 1, and if the number of 0's pixels are more than 1's pixels, then the value of central pixel is set to 0.

Table 4.1: Classification Accuracy using different local features.

Feature	Mean	Energy	Std.	Entropy	Mean+Std.	Energy+Std.
<b>Classification Accuracy</b>	88.11 %	81.83 %	<b>88.53 %</b>	85.67 %	87.50%	86.49 %

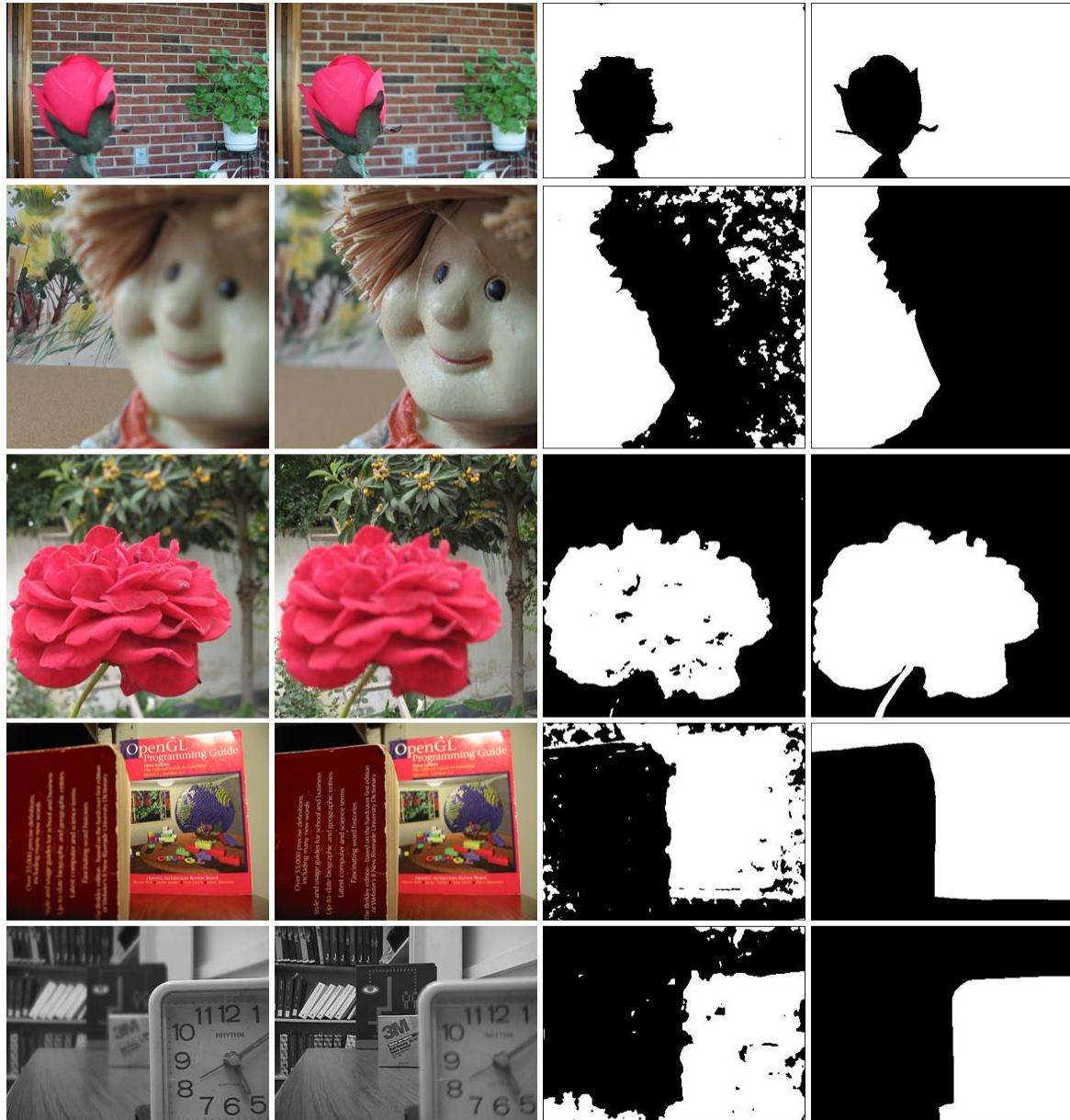


Figure 4.3: Classification results for the different multi-focus test images. First and second columns from top to bottom: pairs of multi-focus “Flower”, “Doll”, “Rose”, “Book”, and “Disk” images, third and fourth columns: the classification result, and the handmade classification maps, respectively.

In this study, we have proposed a new method based on area of separate regions in the binary classification map. First, each separate region in the binary classification map is labeled (using the “bwlabel” function in the Matlab software). Then, the area of each labeled region is calculated (using the “regionprops” function). Finally, the value of each labeled region in the decision map, which its

area is smaller than a specific threshold, is reversed (i.e. from one to zero and vice versa). In order to determine the threshold value, the calculated areas of the isolated regions in the decision map are sorted. Then, the second value from top is selected for the threshold. It is supposed that there are two dominant regions in the classification map (i.e. in and out of focus regions), and isolated regions with smaller areas are misclassified. Fig. 4.4 shows the results of post-processing using majority filter and the proposed method in this paper. In addition, Table 4.2 shows the classification accuracy after post-processing using these methods. As it can be observed in Table 4.2, the improvement using our post-processing method is promising (more than 6.5 %)

#### 4.1.3 Fusion rule for high frequency coefficients

The classifier output obtained in the previous subsection is used as a decision map (DM) for selecting high frequency wavelet coefficients between the multi-focus source images in different directions and decomposition levels of the DT-DWT. In fact, the fusion rule is defined as follows:

$$y_1^l(\cdot | d) = y_A^l(\cdot | d) \times DM(\cdot) + y_B^l(\cdot | d) \times (1 - DM(\cdot)) \quad (4.16)$$

where  $y_A^l(\cdot | d)$  and  $y_B^l(\cdot | d)$  are the high frequency wavelet coefficients of the source images  $A$  and  $B$ , at level  $l$  and orientation  $d$ ,  $l = 1, 2 \dots L$ , and  $d = 1, 2 \dots 6$  which can be  $\pm 15^\circ$ ,  $\pm 45^\circ$ , and  $\pm 75^\circ$ . Also in the coarser decomposition level, an estimation of DM using down sampling is used.

Although the decision maps obtained using the Fisher classifier are very accurate, it does not result in an optimal fused image. Indeed, there are many smooth regions in the source images, and the extracted features could not distinguish them enough to classify them as in or out of focus pixels. Therefore, some pixels are misclassified in the decision map. In order to solve this uncertainty problem, a weighted averaging rule is used for these regions:

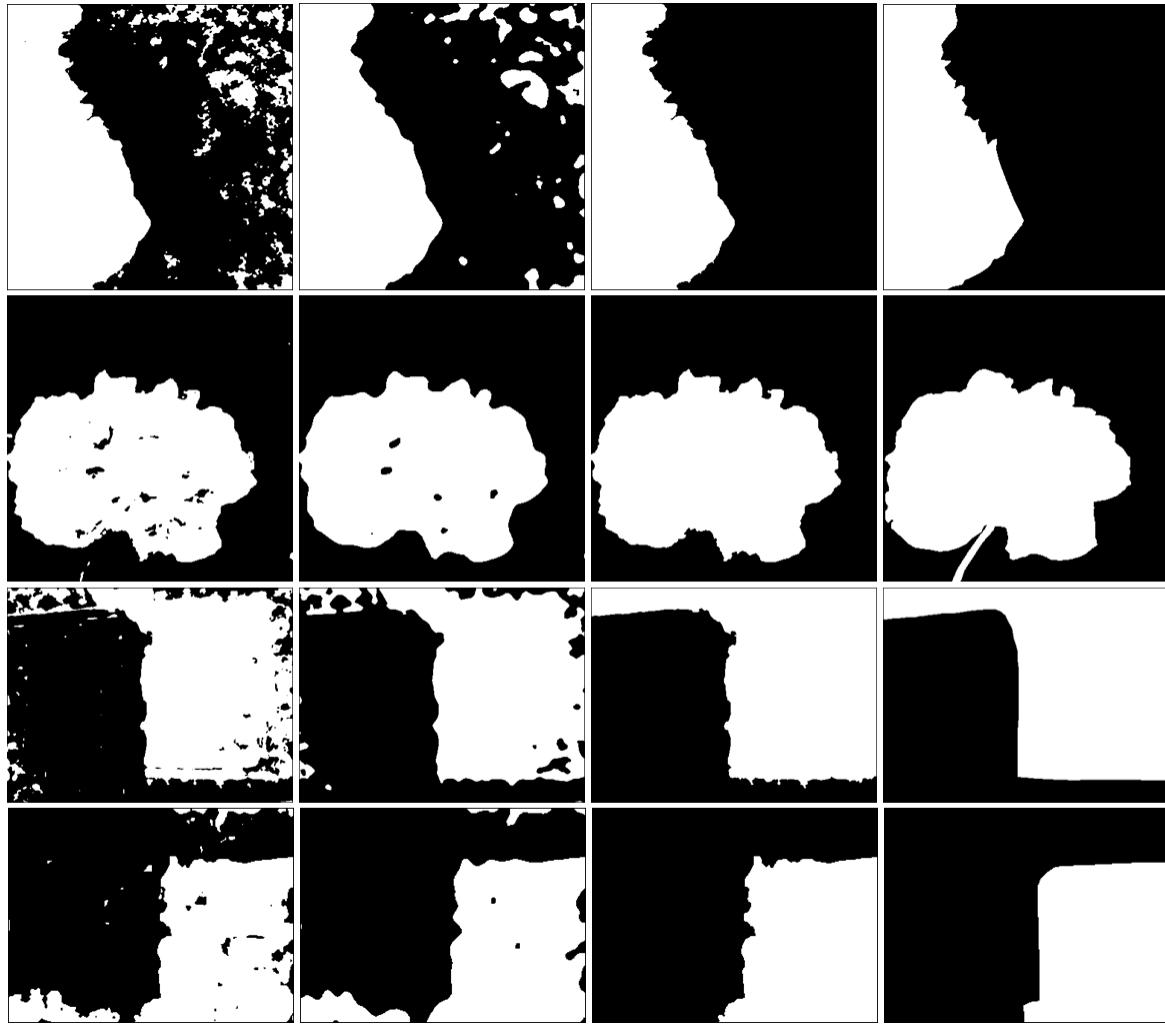


Figure 4.4: First column: the classification results for multi-focus “Doll”, “Rose”, “Book”, and “Disk” images, second column: post-processing results using majority filter, third column: the post-processing results using proposed method and fourth column: handmade classification map.

Table 4.2: Classification Accuracy after post-processing.

Method	Before Post-Processing	Majority Filter [39]	Proposed
<b>Classification Accuracy</b>	88.53 %	90.15 %	<b>95.07 %</b>

$$y_2^l(\cdot | d) = \frac{y_A^l(\cdot | d) \times W_A(\cdot) + y_B^l(\cdot | d) \times W_B(\cdot)}{W_A(\cdot) + W_B(\cdot)} \quad (4.17)$$

where  $W_A(\cdot)$  and  $W_B(\cdot)$  are obtained using the following formula:

$$W_S(i, j) = \frac{1}{12} \sum_{LW=3 \times 3, 7 \times 7} \sum_{d=1}^6 NF_{LW}^d(i, j) \quad (4.18)$$

where  $NF$  is the extracted feature for each source image using (4.6).

We plan to design a good fusion method for combining two fusion rules using (4.16) and (4.17) to integrate as much information as possible into the fused image. We define a dissimilarity measure (DIS) for this purpose. The DIS is intended to measure the degree of ‘dissimilarity’ from the source images. In the following expression, this measure is defined as:

$$AV(i, j) = \frac{1}{12} \sum_{LW=3 \times 3, 7 \times 7} \sum_{d=1}^6 |NF_{LW}^d(A) - NF_{LW}^d(B)| \quad (4.19)$$

Then,

$$DIS(i, j) = \left[ \sin\left(\frac{\pi \times AV(i, j)}{2 \times T}\right) \right]^{1/2} \quad (4.20)$$

where  $T = \max_{i,j}(AV(i, j))$ .

By analyzing the DIS measure, we can determine where and to what extent the source images differ, and use this information to combine the fusion rules. First, we define the following linguistic rules for the proposed fusion rule:

**IF** the DIS measure at a given position is high (i.e. the sources are different at that position) **THEN** we use the first fusion rule (selection using DM)

**IF** the DIS measure at a given position is low (i.e. the sources are similar at that position) **THEN** we use the third fusion rule (weighted averaging)

Then, for constructing standard rules from Linguistic ones, we define small and large membership functions. Fuzzy membership function is a curve, which defines how each point in the input space is mapped to a membership value (or degree of

membership) between zero and one. MF is often given the designation of  $\mu$ . Here, we use the trapezoidal membership function, which is shown in Fig. 4.5. Finally, the fusion rule for high frequency coefficients is defined as follows:

$$y_F^l(\cdot | d) = \sum_{i=1}^2 y_i^l(\cdot | d) \times \mu_i(DIS(\cdot)) \quad (4.21)$$

where  $y_1^l$  and  $y_2^l$  are obtained using (4.16) and (4.17), and  $\mu_1$  and  $\mu_2$  are the large and small membership functions, respectively.

The proposed fusion rule for the high frequency coefficients causes blurring effect around the border area between the in and out of focus regions in the fused image. Although the decision maps obtained using the Fisher classifier are very accurate, but the border points between the in and out of focus regions are not extracted, appropriately. In order to solve this problem, we change the first fusion rule expressed in (4.16). First, the border of in and out of focus regions is obtained using the decision map. This border is then widened (using the “imdilate” function of Matlab software). Fig. 4.6 shows the extracted border maps (BM) using the decision maps of different multi-focus images. Finally, the BM is used for solving the blurring effect around the border regions as:

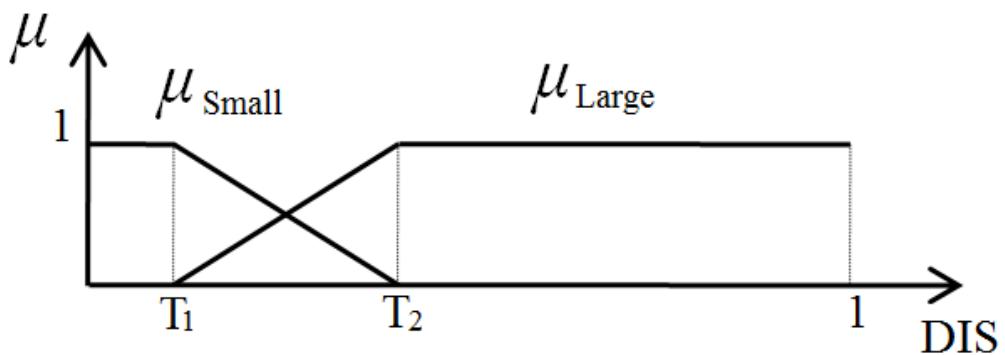


Figure 4.5: Fuzzy membership function,  $T_1 = 0.1$  and  $T_2 = 0.25$ , which are obtained using test images by trial and error.

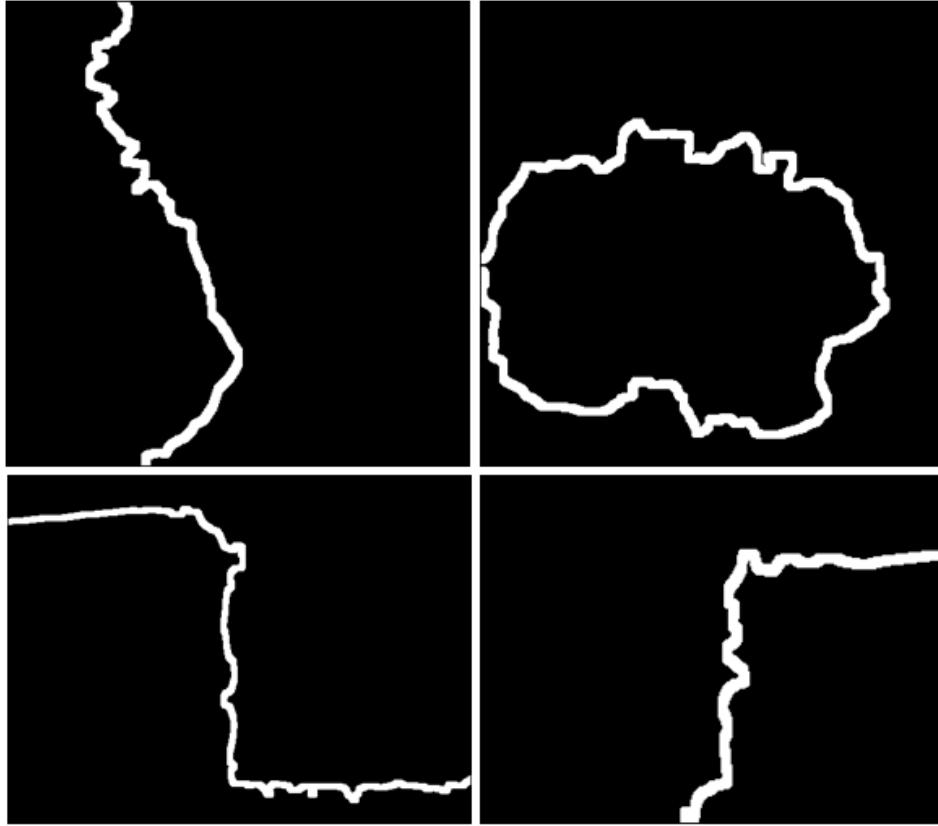


Figure 4.6: Border map for multi-focus “Doll”, “Rose”, “Disk”, and “Book” images (from top left, clockwise).

$$y_3^l(\cdot | d) = \begin{cases} y_A^l(\cdot | d) & \text{if } F_{3 \times 3}(y_A^l(\cdot | d)) > F_{3 \times 3}(y_B^l(\cdot | d)) \\ y_B^l(\cdot | d) & \text{otherwise} \end{cases} \quad (4.22)$$

where  $F_{3 \times 3}$  is the standard deviation using (4.3) in a  $3 \times 3$  local window. Then, the first fusion rule is changed as follows:

$$\tilde{y}_1^l(i, j | d) = \begin{cases} y_3^l(i, j | d) & \text{if } BM(i, j) = 1 \\ y_1^l(i, j | d) & \text{otherwise} \end{cases} \quad (4.23)$$

Fig. 4.7 shows the subjective results using this correction for the different multi-focus images. As it can be seen in Fig. 4.7, the border regions are appropriately retrieved.

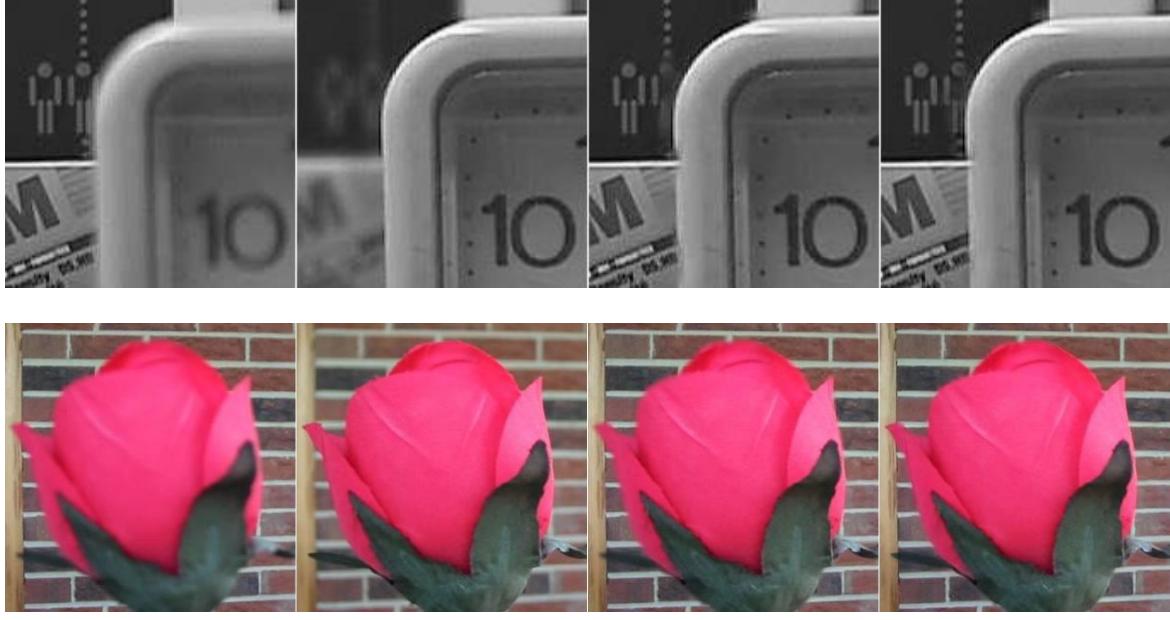


Figure 4.7: Subjective fusion results, Top: from left to right: A part of Multi-focus “Disk” images, results without and with border correction, and bottom: from left to right: A part of Multi-focus “Flower” images, results without and with border correction.

It should be mentioned that we use an estimation of the DM, DIS and BM using down sampling for coarser decomposition levels for the fusion rule. In fact, the proposed method is based on multi-scale grouping, which uses a similar approach for fusion of high frequency wavelet coefficients in different directions and scales of dual-tree discrete wavelet transform.

Usually, using more decomposition levels for wavelet-based image fusion results in a better fused image. In the previous wavelet-based methods, the decision map is obtained by comparing the activity level measurement of wavelet coefficients between source images in different directions and scales. In addition to lower accuracy of the decision maps in finer scales, also in coarser scales due to loss of high frequency information of images (because of passing through low frequency filter banks), the decision maps were wrongly extracted for selecting more important coefficients. Therefore, a certain levels numbers as the best level in the wavelet analysis was selected. Even in some literature, algorithms for determining the optimal decomposition level have been suggested [99].

In the proposed method, because of using an accurate and reliable decision map in different wavelet decomposition levels, the problem of selecting appropriate levels is solved. In fact, the experimental results show that for N-2 decomposition levels, the best results are obtained, where N is the possible decomposition levels for an image (e.g. for a  $512 \times 512$  image,  $N = \log_2 512 = 8$ ).

Table 4.3 shows a comparison between previous fusion rules for high frequency wavelet coefficients consisting of maximum selection (MS), Weighted average (WA) [45], and window based verification (WBV) [39], and the proposed method in this paper. In this experiment, which has been performed on different multi-focus images, the best value of the PSNR index with respect to the decomposition levels for the fused images (compared to the handmade or reference image) are obtained for different methods. Handmade or reference image is obtained by transferring in focus regions of source images into one image. As it can be observed in Table 4.3, the best results for the proposed method are obtained using N-2 decomposition levels, while for other fusion rules the best results are obtained in different decomposition levels. The purpose of this experiment is determining optimal decomposition levels for image fusion process, and comparisons between different methods will be fully discussed in experimental results Chapter.

#### 4.1.4 Fusion rule for low frequency coefficients

The information contained in the low frequency (LF) coefficients is very important for the natural appearance of an image, because LF coefficients contain the bulk of the energy of an image. Commonly, the averaging is an accepted method to fuse approximation or low frequency coefficients of the source images in the wavelet-based fusion methods:

$$x_F^L(.) = \frac{x_A^L(.) + x_B^L(.)}{2} \quad (4.24)$$

Table 4.3: The best value of the PSNR<sub>dB</sub> index with respect to decomposition levels obtained for different multi-focus images and different fusion rules.

Test Image	MS	WA [45]	WBV [39]	Proposed	N
“Disk”	35.29	37.41	37.76	42.56	5
	L = 4	L = 5	L = 5	L = <b>3</b>	
“Lab”	35.83	37.64	37.85	41.62	5
	L = 5	L = 5	L = 5	L = <b>3</b>	
“Pepsi”	39.37	40.77	41.57	44.34	8
	L = 5	L = 5	L = 5	L = <b>6</b>	
“Flower”	33.42	34.92	34.80	41.36	8
	L = 4	L = 7	L = 7	L = <b>6</b>	
“Rose”	32.54	33.95	33.77	37.54	8
	L = 4	L = 5	L = 5	L = <b>6</b>	
“Doll”	38.22	38.56	39.14	42.75	8
	L = 6	L = 7	L = 7	L = <b>6</b>	
“Book”	37.97	39.56	39.98	44.87	8
	L = 3	L = 4	L = 4	L = <b>6</b>	

where  $x_F^L()$ ,  $x_A^L()$  and  $x_B^L()$  are the fused and input LF coefficients, and  $L$  represents the coarsest resolution level.

However, equally weighing the input approximation or LF sub-bands in the final fused image causes the problem of contrast reduction, because the source images have different contrast conditions. In [100], LF fusion rule is performed using arithmetic combinations of input LF coefficients. Specifically, the offset zero-mean addition is used, which is defined as:

$$x_F^L() = x_A^L() + x_B^L() - \frac{\mu_A + \mu_B}{2} \quad (4.25)$$

where  $\mu_A$  and  $\mu_B$  are the mean values of the two input LF sub-bands, and  $L$  represents the coarsest resolution level. Like other arithmetic fusion methods, the

fusion defined in (4.24) is weak to the destructive superposition, especially if the input LF sub-bands have opposing illumination levels.

In this thesis, we have proposed a new method for low frequency fusion rule. Since, there are not enough differences between LF coefficients of the source images to distinguish them as in or out of focus coefficients, we have used an estimation of the decision map obtained by the Fisher classifier for LF fusion rule, which is defined as:

$$x_F^L(.) = x_A^L(.) \times \hat{DM}(.) + x_B^L(.) \times \left(1 - \hat{DM}(.)\right) \quad (4.26)$$

where  $\hat{DM}(.)$  is obtained via down sampling of the DM, which is the classification output.

Finally, after combining low and high frequency dual-tree discrete wavelet coefficients of the source images, the final fused image is obtained using inverse DT-DWT:

$$X_F = \psi^{-1}(y_F, x_F^L) \quad (4.27)$$

where  $\psi^{-1}$  is the inverse DT-DWT,  $y_F$  are the high frequency coefficients using (4.21), and  $x_F^L$  are the low frequency coefficients using (4.26).

Table 4.4 shows the PSNR index for three different alternative low frequency fusion rules in our fusion scheme. This experiment was performed on the multi-focus images dataset, and average PSNRs for all test images are shown in Table 4.4. Unlike many authors who believe that the LF fusion rules have little influence on the overall fusion performance [100], it can be seen in Table 4 that the proposed LF fusion rule is very effective (more than +1.46 dB on average).

Table 4.4: PSNR<sub>dB</sub> index for different low frequency fusion rules.

Low Frequency Fusion rule	Averaging	Arithmetic combinations [100]	Proposed
<b>PSNR<sub>dB</sub></b>	40.46	40.87	<b>42.33</b>

## 4.2 IR and visible image fusion using fuzzy logic and particle swarm optimization

In this subsection, we propose a new wavelet-based image fusion method, which combines pixel and region-based fusion algorithms using fuzzy logic. We specifically present new fusion rules to merge high and low frequency wavelet coefficients in an appropriate way in order to achieve the best quality in the fused image [101-102]. The basic idea is to compute a dissimilarity measure of the source images to combine three different fusion rules to fuse high frequency wavelet coefficients, which are the weighted averaging, selection using pixel-based decision map, and selection using region-based decision map. In addition, we use an optimization-based method for low frequency fusion rule instead of simple averaging, because equal weighing of the input approximation sub-bands leads to the problem of contrast reduction. To evaluate the new fusion algorithm with other wavelet-based fusion methods, we use DT-DWT, which introduces limited redundancy and allows the transform to provide approximate shift invariance and directionally selective filters, while preserving the usual properties of perfect reconstruction and computational efficiency [103, 76].

In the following, an image segmentation algorithm using watershed transform is described. Then, the proposed image fusion method will be presented.

### 4.2.1 Image segmentation algorithm using watershed transform

In this subsection, we describe an algorithm based on watershed segmentation for multi-valued images segmentation, especially for IR and visible images. In the immersion-based watershed algorithm of [104], the gradient magnitude of an image is firstly calculated. This image is considered as a topographic relief in where the brightness value of each pixel corresponds to a physical elevation. The technique can simply be described by figuring that holes are pierced in each local minimum of the topographic relief. Eventually, the surface is slowly immersed into a 'lake', through that filling all the catchment basins, starting from the basin which is

associated with the global minimum. As soon as two catchment basins tend to join, a dam is constructed. The procedure results in a partitioning of the image in many catchment basins of which the borders describe the watershed lines [105].

For multi-valued images, a single-valued method can be adopted by segmenting each band, separately, but jointly segmented images work better for the region-based image fusion algorithms. This is because the segmentation map will contain a smaller number of regions to represent all features in the both IR and visible images, and therefore reduces the computational cost.

Here we use a new method to obtain a joint gradient image from IR and visible images. First, a Gaussian derivative function is used to generate the gradient magnitude from the source images as:

$$G_S(x, y) = \sqrt{(S(x, y) * G'_x)^2 + (S(x, y) * G'_y)^2} \quad (4.28)$$

where  $S(x, y)$  is the gray scale image,  $G'_x$  and  $G'_y$  are the Gaussian partial derivative filters in the x and y directions, and  $*$  denotes convolution.

It is followed by a morphological opening, which operates similar to the local median filtering to reduce step noise in the gradient image. The opening is based on a min(.) followed by a max(.) operation on a local neighborhood around the pixels. Then a joint gradient image (JG) is obtained from the source images using the following formula:

$$JG(x, y) = \frac{1}{2} \left( \frac{\tilde{G}_A(x, y)}{\max_{x, y} \tilde{G}_A(x, y)} + \frac{\tilde{G}_B(x, y)}{\max_{x, y} \tilde{G}_B(x, y)} \right) \quad (4.29)$$

where  $\tilde{G}_A$  and  $\tilde{G}_B$  are the filtered gradient images from the IR and visible images.

Finally, the joint gradient image is used to obtain segmentation map via watershed algorithm. The watershed algorithm is known to suffer from the over segmentation problem due to noise and other local irregularities of the gradient

image. A practical solution to this problem is to limit the number of allowable

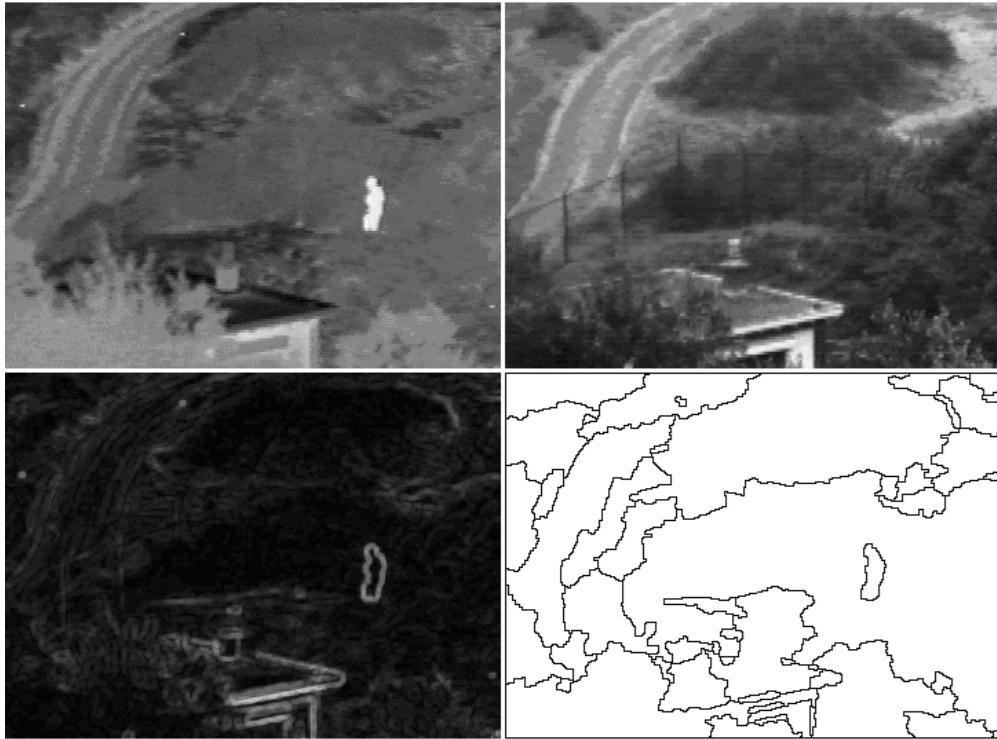


Figure 4.8: Segmentation of IR and visible images, from top left, clockwise: IR image, visible image, segmentation map, and the joint gradient image.

regions by incorporating a preprocessing step, which is designed to bring additional knowledge into the segmentation process. An approach used to control over segmentation is based on the concept of markers. A *marker* is a connected component belonging to an image. There are *internal* markers, associated with the object of interest, and *external* markers, associated with the background [106].

Here, we have used a binary version of the joint gradient (JG) image as the external marker (EM), and its edges as the internal marker (IM). Then we have used a morphological reconstruction based method to limit the number of regional minima in the JG image to control the over segmentation problem as follows:

$$\bar{JG} = \text{imimposemin} (\text{JG}, \text{EM} \& \text{IM}) \quad (4.30)$$

where the “imimposemin” function in the MATLAB software modifies the joint

gradient image using morphological reconstruction, so it only has regional minima wherever  $(EM \& IM)$  is nonzero [106], and  $\&$  is the logical AND operator.

Now, the watershed algorithm is applied to the modified JG image. Thus, at the end, we have a joint segmentation map for the both IR and visible images (see Fig. 4.8).

#### 4.2.2 The proposed image fusion method

In this subsection, we present the proposed multi-sensor image fusion algorithm. Fig. 4.9 shows the block diagram of the proposed method, which consists of a number of essential stages:

- 1) A segmentation map is obtained using the marker-based watershed segmentation algorithm based on all the source images.
- 2) The source images are decomposed into different directions and scales using the DT-DWT.
- 3) The pixel-based decision maps are obtained through activities measurement of high frequency wavelet coefficients in each scale and direction of the DT-DWT.
- 4) Having the segmentation map, the region-based decision maps are generated from pixel-based decision maps.
- 5) High frequency wavelet coefficients of the source images are integrated using fuzzy logic, which combines outputs of three different fusion rules based on a dissimilarity measure of the source images: selection using pixel-based decision map, selection using region-based decision map, and weighted averaging.
- 6) Low frequency wavelet coefficients of the final fused image are obtained via weighted averaging, in which the weights are optimally estimated using the Particle Swarm Optimization algorithm.
- 7) The Inverse DT-DWT of the new low and high frequency wavelet coefficients generates the final fused image.

In the following subsections, we have provided more detailed explications of the image fusion process.

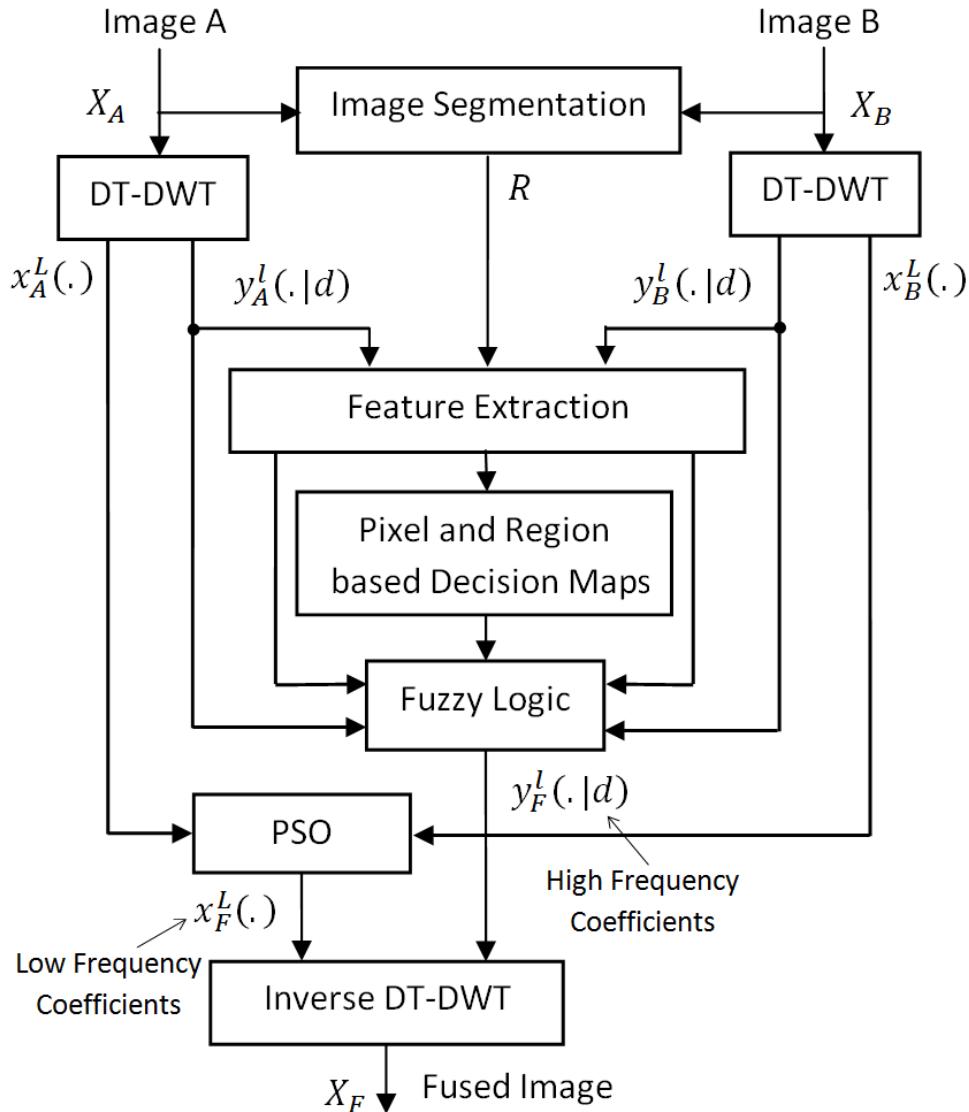


Figure 4.9: Block diagram of the proposed image fusion algorithm.

#### 4.2.2.1 Generation of pixel and region-based decision maps

In the infrared images, usually the regions do not contain texture information, and the information within them is borders of different areas. This is because of temperature differences between adjacent areas, which are measured by the infrared sensors. On the other hand, the regions in the visual images contain textures, and in

special conditions--such as poor light condition and foggy weather--the borders of different areas are not correctly detected.

Wavelet-based image fusion methods are motivated by the fact that the human visual system is primarily sensitive to the contrast changes or edges. Based on this fact, the information of the infrared image (borders of areas) with the information of the visual image (texture of areas) can be combined in the wavelet transform domain. Generation of decision map (DM) for selecting the more important coefficients in the wavelet transform domain between the infrared and visual images is the main idea.

The activity level measurement is used to judge the quality of a given part of each source image in the transform domain. Here, we use the local standard deviation as the window-based activity, which is obtained using the magnitude of complex wavelet coefficients of the source images:

$$y_{abs}^l(\cdot | d) = \sqrt{y_{real}^l(\cdot | d)^2 + y_{imag}^l(\cdot | d)^2} \quad (4.31)$$

Then,

$$F^l(i, j | d) = \sqrt{\frac{1}{W} \sum_{n=-N}^N \sum_{m=-M}^M \left( |y_{abs}^l(i+n, j+m | d)| - |\bar{y}_{abs}^l(\cdot | d)| \right)^2} \quad (4.32)$$

where  $y_{imag}^l(\cdot | d)$  and  $y_{real}^l(\cdot | d)$  are the imaginary and real high frequency wavelet coefficients,  $l$  is the decomposition level, and  $d = 1, 2..6$  is the direction of high frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ).  $W$  is the number of pixels in the local window  $[2N+1, 2M+1]$ , and  $(\cdot)$  is the shorthand notation for spatial position.

After extracting the local feature using the magnitude of complex wavelet coefficients, it is followed by nonlinear averaging in the local window for taking into account the neighbor dependency and smoothing on the feature space:

$$NF^l(i, j | d) = \sum_{n=-N}^N \sum_{m=-M}^M G(n, m) \times F^l(i+n, j+m | d) \quad (4.33)$$

where  $F$  is the local feature,  $G(n, m)$  is the local window's weight, which is obtained by the Gaussian filter.

After extracting local features, pixel-based decision map (PDM) in each decomposition level and direction is obtained using the following formula:

$$PDM^l(\cdot | d) = \begin{cases} 1 & \text{if } NF_A^l(\cdot | d) > NF_B^l(\cdot | d) \\ 0 & \text{otherwise} \end{cases} \quad (4.34)$$

where  $NF_A^l(\cdot | d)$  and  $NF_B^l(\cdot | d)$  are the extracted features using (4.33) for the input source images,  $l$  is the decomposition level,  $d = 1, 2..6$  is the direction of high frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ), and  $(\cdot)$  is the shorthand notation for spatial position.

Region-based activity measurements are also used to obtain the decision map for selecting more important high frequency wavelet coefficients between source images, but it is very sensitive to the large coefficients in the region and is not accurate [49-50]. Therefore, we use a new method to generate the region-based decision map (RDM). First, a joint segmentation map based on all source images is obtained, which is explained in previous subsection. Then, by looking into the labeled region, if numbers of one in the corresponding region within the PDM are more than numbers of zero, then the same region in the RDM is set to one, and otherwise the corresponding region in the RDM is set to zero. This is expressed in the following equation:

$$RDM^l(R | d) = \begin{cases} 1 & \text{if } \left( \sum_{n \in R} PDM^l(n | d) \right) > \frac{|R|}{2} \\ 0 & \text{otherwise} \end{cases} \quad (4.35)$$

where  $R$  is the region with size  $|R|$ , PDM is the pixel-based decision map using (4.34),  $l$  is the level of decomposition, and  $d = 1, 2..6$  is the direction of high

frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ).

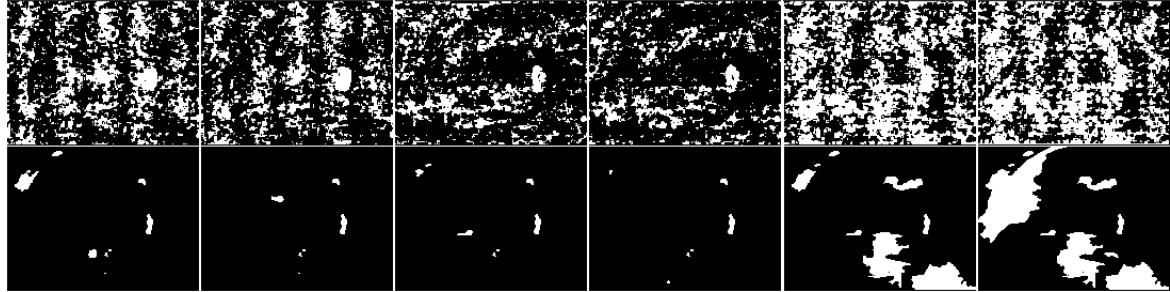


Figure 4.10: Pixel- and region-based decision maps (first and second row) obtained from “UN Camp” images in the first decomposition level and six directional sub-bands of the DT-DWT, from left to right:  $-15, +15, -45, +45, -75$  and  $+75$  directions, respectively. Black regions indicate that high frequency wavelet coefficients in the DT-DWT sub-bands of the final fused image should be selected from the visible image, and white regions indicate that high frequency wavelet coefficients in the DT-DWT sub-bands of the final fused image should be selected from the IR image.

Fig. 4.10 shows pixel- and region-based decision maps obtained from IR and visible “UN camp” images in the first decomposition level and six directional sub-bands of DT-DWT.

#### 4.2.2.2 Fusion rule for high frequency wavelet coefficients

Having the pixel and region-based decision maps, there are two strategies for fusion of high frequency wavelet coefficient. Indeed, we can use either the PDM or the RDM for selecting more important high frequency wavelet coefficients between the source images using the following:

$$y_1^l(\cdot | d) = PDM^l(\cdot | d) \times y_A^l(\cdot | d) + (1 - PDM^l(\cdot | d)) \times y_B^l(\cdot | d) \quad (4.36)$$

and,

$$y_2^l(\cdot | d) = RDM^l(\cdot | d) \times y_A^l(\cdot | d) + (1 - RDM^l(\cdot | d)) \times y_B^l(\cdot | d) \quad (4.37)$$

where  $y_A^l(\cdot | d)$  and  $y_B^l(\cdot | d)$  are the high frequency wavelet coefficients of source images,  $l$  is the level of decomposition,  $d = 1, 2..6$  is the direction of high frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ), PDM and RDM are the pixel and region-based decision maps using (4.34) and (4.35), and  $(\cdot)$  is the shorthand notation for spatial position.

As it can be seen in Fig. 4.10, using region-based decision map (RDM) for fusing high frequency wavelet coefficients using (4.37) will eliminate a lot of information from the infrared image in the final fused image. On the other hand, using pixel-based decision map (PDM) for fusion of high frequency wavelet coefficients using (4.35) affected by problems such as noise, image misregistration, and usually results in artifacts in the fused image. In addition, in some regions, we cannot make a good decision for selecting more important wavelet coefficients between the source images, because there is not enough difference between the extracted features from these regions. Therefore, a weighted averaging rule is proposed for these pixels:

$$y_3^l(\cdot | d) = \frac{NF_A^l(\cdot | d) \times y_A^l(\cdot | d) + NF_B^l(\cdot | d) \times y_B^l(\cdot | d)}{NF_A^l(\cdot | d) + NF_B^l(\cdot | d)} \quad (4.38)$$

where  $y_A^l(\cdot | d)$  and  $y_B^l(\cdot | d)$  are the high frequency wavelet coefficients of source images,  $l$  is the level of decomposition,  $d = 1, 2..6$  is the direction of high frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ),  $NF_A^l(\cdot | d)$  and  $NF_B^l(\cdot | d)$  are the extracted features using (4.33) for the input source images  $A$  and  $B$ , and  $(\cdot)$  is the shorthand notation for spatial position.

None of the three fusion rules expressed in (4.36), (4.37), and (4.38) results in a good fused image, individually. Indeed, we use fuzzy logic to design a good fusion method by combining the three fusion rules to transfer as much information as possible from the source images into the fused image. To this end, a dissimilarity measure (DIS) of the source images is defined. The DIS is intended to quantify the

degree of ‘dissimilarity’ between the source images. In the following expression, this measure is defined as:

$$DNF^l(i, j | d) = |NF_A^l(i, j | d) - NF_B^l(i, j | d)| \quad (4.39)$$

Then,

$$DIS^l(i, j | d) = \sqrt{\sin\left(\frac{\pi \times DNF^l(i, j | d)}{2 \times T}\right)} \quad (4.40)$$

where  $NF_A^l(\cdot | d)$  and  $NF_B^l(\cdot | d)$  are the extracted features using (4.33) for the input source images  $A$  and  $B$ ,  $l$  is the level of decomposition,  $d = 1, 2..6$  is the direction of high frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ), and  $T = \max_{i,j} DNF^l(i, j | d)$ .

Indeed, we have used sinusoid function in (4.40), in order to normalize the dissimilarity measure between zero and one. By analyzing the DIS measure, we can determine where and to what extent the source images differ, and use this information to combine the three different fusion rules expressed in (4.36), (4.37), and (4.38). First, we define the following linguistic rules for fusion process, which are obtained based on our observation:

**IF** the DIS measure at a given position is *high* (i.e. the sources are distinctly different at that position) **THEN** we use the first fusion rule (selection using the pixel-based decision map)

**IF** the DIS measure at a given position is *medium* (i.e. the sources are different at that position) **THEN** we use the second fusion rule (selection using the region-based decision map)

**IF** the DIS measure at a given position is *low* (i.e. the sources are similar at that position) **THEN** we use the third fusion rule (weighted averaging)

Then, for constructing standard rules from Linguistic ones, we define small, medium, and large membership functions. Fuzzy membership function (MF) is a

curve that defines how each point in the input space is mapped to a membership value or degree of membership between zero and one. MF is often given the designation of  $\mu$ . Here, we use the triangular membership function, which is shown in Fig. 4.11. In order to construct the three fuzzy membership functions, one threshold ( $T$  in Fig. 4.11) must be defined. The threshold should be set to partition the feature space (DIS) into the three regions of linguistic variables (i.e. small, medium, and large) in an appropriate way. In order to find the best value for the threshold, we have performed an experiment using different datasets. Fig. 4.12 shows Petrovic index of fused image versus different thresholds obtained from different multi-sensor datasets [107]. The Petrovic index considers the amount of edge information transferred from the input images to the fused image using the Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images. Based on the experiment, the threshold is selected which maximizes the Petrovic index of final fused image for different datasets ( $T \approx 0.2$ ). It should be mentioned that the averaging fusion rule is used to fuse low frequency wavelet coefficients of the source images in this experiment.

Finally, the fusion rule for integrating high frequency wavelet coefficients from the source images is defined as follows:

$$y_F^l(\cdot | d) = \sum_{i=1}^3 y_i^l(\cdot | d) \times \mu_i(DIS^l(\cdot | d)) \quad (4.41)$$

where  $y_1^l(\cdot | d)$ ,  $y_2^l(\cdot | d)$  and  $y_3^l(\cdot | d)$  are obtained using (4.36), (4.37) and (4.38),  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  are the large, medium and small membership functions, respectively,  $l$  is the level of decomposition,  $d = 1, 2, \dots, 6$  is the direction of high frequency wavelet coefficients ( $\pm 15, \pm 45, \pm 75$ ), and  $(\cdot)$  is shorthand notation for spatial position.

#### 4.2.2.3 Fusion rule for low frequency wavelet coefficients

The information contained in the low frequency (LF) coefficients is vital for the natural appearance of an image, because LF coefficients contain the bulk of the

energy of an image. Commonly, averaging is a popular method to fuse LF subbands of the source images in most wavelet-based methods:

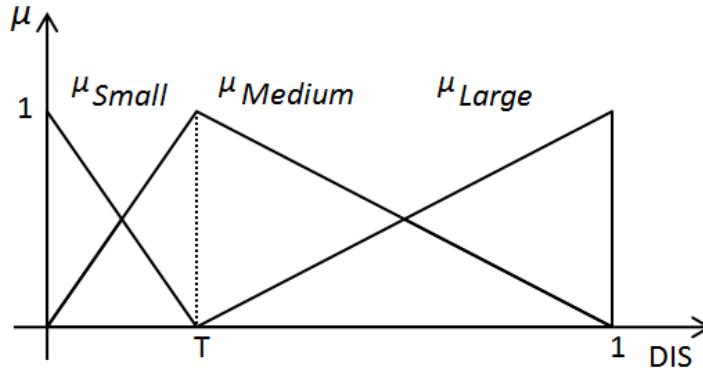


Figure 4.11: Triangular fuzzy membership functions.

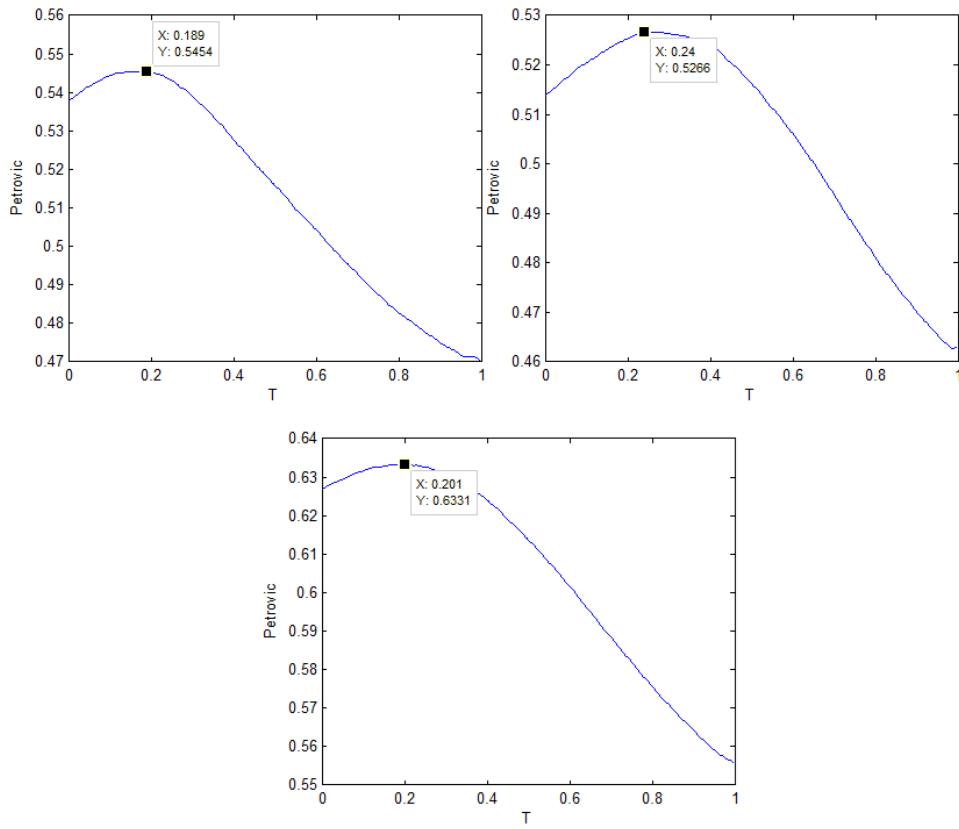


Figure 4.12: The Petrovic index of fused images versus different thresholds (T for building fuzzy membership functions) obtained from different multi-sensor datasets, from left to right: “UN Camp”, “Dune”, and “Trees”.

$$x_F^L(i, j) = \frac{x_A^L(i, j) + x_B^L(i, j)}{2} \quad (4.42)$$

where  $x_F^L(i, j)$ ,  $x_A^L(i, j)$  and  $x_B^L(i, j)$  are the fused and input LF wavelet coefficients, and  $L$  represents the coarsest resolution level.

However, equal weighing of the input approximation sub-bands leads to the problem of contrast reduction, because IR and visible images have different contrast conditions. Petrovic and Xydeas proposed a LF fusion rule, which is performed using an arithmetic combination of the input LF coefficients [100]. Specifically, the offset zero-mean addition is used, which is defined as:

$$x_F^L(i, j) = x_A^L(i, j) + x_B^L(i, j) - \frac{\mu_A + \mu_B}{2} \quad (4.43)$$

where  $\mu_A$  and  $\mu_B$  are the mean values of the two input LF sub-bands, and  $L$  represents the coarsest resolution level. Like other arithmetic fusion methods, the fusion defined in (4.38) is weak to the destructive superposition, especially if the input LF sub-bands have opposing illumination levels [100].

An alternative approach is a weighted averaging proposed in [52], in which the weights are obtained using a region-based activity measurement (the normalized Shannon entropy) of the low frequency wavelet coefficients:

$$W_S(R) = \frac{1}{|R|} \sum_{n \in R} x_S^L(n) \times \log_2 x_S^L(n) \quad (4.44)$$

where  $R$  is the region with size  $|R|$ , and  $x_S^L$  are the input LF wavelet coefficients, and  $L$  represents the coarsest resolution level. Hence, the composite LF sub-band is generated using:

$$x_F^L(i, j) = \frac{W_A(i, j) \times x_A^L(i, j) + W_B(i, j) \times x_B^L(i, j)}{W_A(i, j) + W_B(i, j)} \quad (4.45)$$

where  $x_F^L(i, j)$ ,  $x_A^L(i, j)$  and  $x_B^L(i, j)$  are the fused and input LF wavelet coefficients,  $W_A(i, j)$  and  $W_B(i, j)$  are obtained using (4.44), and  $L$  represents the coarsest resolution level.

Averaging, arithmetic combination [100], and weighted averaging [52] fusion rules cannot lead to an optimal result. In this paper, we have proposed a new weighted averaging rule for the fusion of the LF wavelet coefficients, in which the weights are estimated using an optimization algorithm:

$$x_F^L(i, j) = W_1(i, j) \times x_A^L(i, j) + W_2(i, j) \times x_B^L(i, j) \quad (4.46)$$

where  $x_F^L(i, j)$ ,  $x_A^L(i, j)$  and  $x_B^L(i, j)$  are the fused and input LF wavelet coefficients,  $W_1$  and  $W_2$  are the values in the range  $[0,1]$ , and  $W_1+W_2=1$ .

An optimization algorithm finds the optimal weights with maximizing the Entropy index from the final fused image. The entropy metric measures the information content in an image. An image with high information content will have high entropy. The entropy index is defined as:

$$H = -\sum_{i=0}^K h(i) \times \log_2(h(i)) \quad (4.47)$$

where  $h$  is the normalized histogram of reconstructed or fused image, and  $K$  is the number of gray levels in the image. In addition, the final fused image is obtained using inverse DT-DWT:

$$X_F = \psi^{-1}(x_F^L, y_F) \quad (4.48)$$

in which  $\psi$  is the inverse DT-DWT,  $y_F$  and  $x_F^L$  are the composite high and low frequency wavelet coefficients, which are obtained using (4.41) and (4.46).

### 4.2.3 Optimization algorithm

Over the years, many real-parameter optimization algorithms have been developed by using point-by-point [108-109] as well as multi-point approaches [110-111]. While a point-by-point approach begins with one guess solution and updates the solution iteratively in the hope of reaching near the optimum solution, a multi-point approach deals with a number of solutions in each iteration. Starting with a number of guess solutions, the multi-point algorithm updates one or more solutions in a synergistic manner in the hope of steering the population towards the optimum. In this paper, we focus our attention to the multi-point optimization algorithms, which we shall refer here as the population-based optimization algorithms.

Here, we have considered two population-based search algorithms (particle swarm optimization and genetic algorithm (GA)) for the optimization problem. Fig. 4.13 shows the entropy index of fused images obtained from different IR and visible datasets at different iterations of PSO and different generation of genetic algorithms for the optimization problem. In addition, Table 4.5 shows the entropy index for comparison between different alternative LF fusion rules in the proposed fusion scheme.

Table 4.5: The entropy index of fused images obtained from different datasets using different LF fusion rules in the proposed fusion scheme.

Low Frequency Fusion rule	UN Camp	Dune	Trees
<b>Averaging</b>	6.63	6.85	6.43
<b>Weighted Averaging [52]</b>	6.85	7.11	6.61
<b>Arithmetic Combination [100]</b>	6.97	7.21	6.71
<b>PSO</b>	7.30	7.45	7.01
<b>Genetic</b>	<b>7.41</b>	<b>7.56</b>	<b>7.11</b>

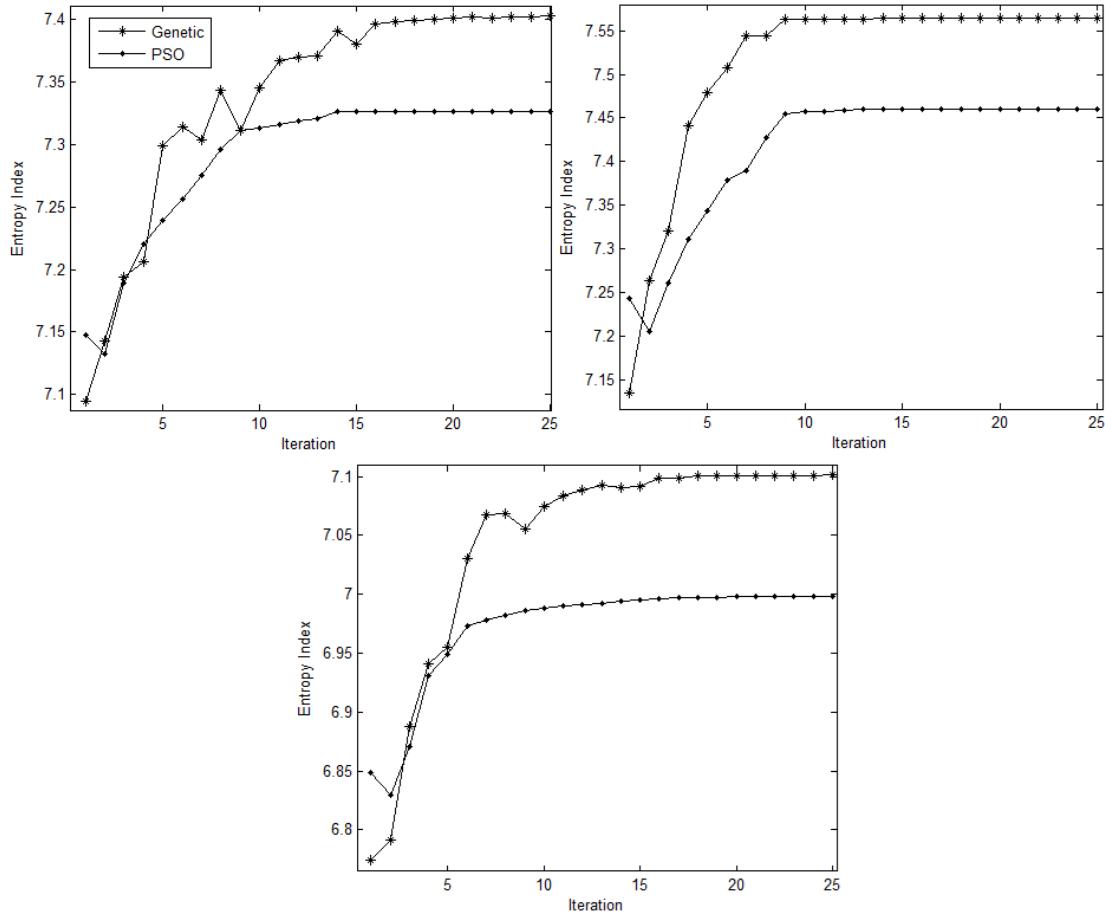


Figure 4.13: The Entropy index of fused images versus iteration of the PSO and genetic algorithms obtained from different multi-sensor datasets, from left to right: “UN Camp”, “Dune”, and “Trees”.

The high amount of the entropy index for the proposed method indicates that the fused image contains higher information compared to other methods. In addition to objective evaluation, a sample image has been selected for comparing different LF fusion rules, subjectively. As it can be seen in Fig. 4.14, the image contrast in the proposed method is far better compared to other methods.

Overall, the results indicate that both GAs and PSO can be used in the optimization of parameters during image fusion. In terms of computational effort, the GA approach is faster, although it should be noted that neither algorithm takes what can be considered an unacceptably long time to determine their results. With respect to accuracy of model parameters, the GA determines values, which are

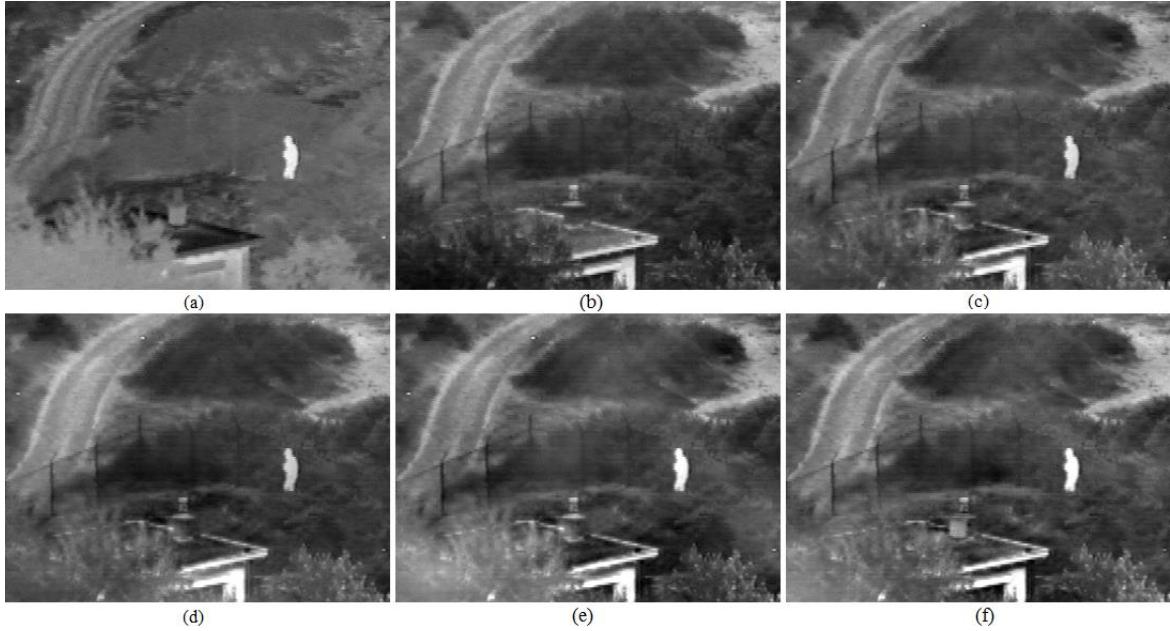


Figure 4.14: Subjective image fusion results using different low frequency fusion rule in the proposed fusion scheme. (a) “UN Camp” IR image, (b) visible image, and fused images using (c) averaging, (d) weighted averaging [52], (e) Arithmetic Combination [100] and (f) the proposed method based on PSO algorithm.

better than does the PSO. Thus, it must be concluded that for the process of optimization, the GA approach is superior to the PSO approach. Techniques such as PSO and GA are inspired by nature, and have proved themselves to be effective solutions to optimization problems. There are control parameters involved in these meta-heuristics, and appropriate setting of these parameters is a key point for success. In general, some form of trial-and-error tuning is necessary for each particular instance of optimization problem. In the following subsections, we briefly describe PSO and genetic algorithms.

#### 4.2.3.1 Particle swarm optimization

James Kennedy and Russell C. Eberhart [110] firstly proposed the PSO algorithm for optimization. PSO is a population-based search algorithm based on the simulation of the social behavior of birds within a flock. In the PSO, particles are flown through hyper-dimensional search space. Changes to the position of the

particles within the search space are based on the social-psychological tendency of individuals to emulate the success of other individuals [110].

The position of each particle is changed according to its own experience and that of its neighbors. Let  $x_i(t)$  denotes the position of particle  $p_i$ , at iteration  $t$ . The position of  $p_i$  is then changed by adding a velocity  $v_i(t)$  to the current position, i.e.:

$$x_i(t) = x_i(t-1) + v_i(t) \quad (4.49)$$

The velocity vector reflects the socially exchanged information, and is generally defined in the following way:

$$v_i(t) = W \times v_i(t-1) + C_1 r_1 \times (x_{pbest_i}(t) - x_i(t)) + C_2 r_2 \times (x_{leader}(t) - x_i(t)) \quad (4.50)$$

where  $W$  is the inertia weight employed to control the impact of the previous history of velocities on the current velocity of a given particle, and  $x_{pbest}$  is the personal best position of a given particle, so far. That is, the position of the particle that has provided the greatest success.  $x_{leader}$  is position of the best particle of the entire swarm.  $C_1$  and  $C_2$  are learning factors and represent the attraction that a particle has toward either its own success or that of its neighbors, and  $r_1, r_2 \in [0,1]$  are random values.

The role of the inertia weight  $W$ , in (4.50), is critical for the PSO's convergence behavior. The inertia weight is employed to control the impact of the previous history of velocities on the current one. Accordingly, the parameter  $W$  adjusts the trade-off between the global and local exploration abilities of the swarm. A large inertia weight facilitates global exploration (searching new areas), while a small one tends to facilitate local exploration, i.e., fine-tuning the current search area. A suitable value for the inertia weight  $W$  usually provides balance between global and local exploration abilities and consequently results in a reduction of the number of iterations required to locate the optimum solution. Initially, the inertia weight was constant. However, experimental results indicated that it is better to initially set the inertia to a large value, in order to promote global exploration of the search space,

and gradually decrease it to get more refined solutions. Thus, an initial value around 1.2 and a gradual decline towards 0 can be considered as a good choice for  $W$ . In [112-113] a decaying inertia weight is proposed and tested, with the aim of favoring global search at the start of the algorithm and local search later.

The parameters  $C_1$  and  $C_2$ , in (4.50), are not critical for PSO's convergence. However, proper fine-tuning may result in faster convergence and alleviation of local minima. An extended study of the acceleration parameter in the first version of PSO is given in [114]. As default values,  $C_1 = C_2 = 2$  were proposed, but experimental results indicate that  $C_1 = C_2 = 0.5$  might provide even better results. Recent work reports that it might be even better to choose a larger cognitive parameter,  $C_1$ , than a social parameter,  $C_2$ , but with  $C_1 + C_2 = 4$  [115]. The parameters  $r_1$  and  $r_2$  are used to maintain the diversity of the population, and they are uniformly distributed in the range  $[0, 1]$ .

Here, default values for the parameters  $C_1$  and  $C_2$  were used:  $C_1 = 2.4$ ,  $C_2 = 1.6$ . Although the choice of the parameter values seems not to be critical for the success of the methods, it appears that faster convergence can be obtained by proper fine-tuning. The balance between the global and local exploration abilities of the PSO is mainly controlled by the inertia weight, since the positions of the particles are updated according to the classical PSO strategy. A time decreasing inertia weight value, i.e., start from 1.2 and gradually decrease towards 0.4, proved to be superior than a constant value. This is because the larger inertia weights at the beginning help to find good seeds and the later small inertia weights facilitate a fine search.

In addition, the number of iterations and particles is set based on convergence of the PSO algorithm. More particles could speed up the convergence of the PSO algorithm; however, it increases its computation time. In addition, using too much particles and iterations is ineffective on the convergence of PSO algorithm. Fig. 4.15 (a) shows the entropy index of fused image obtained from “UN Camp” IR and visible dataset at different iterations of the PSO algorithm using different particles number. Based on this experiment, we have chosen 32 particles in the PSO

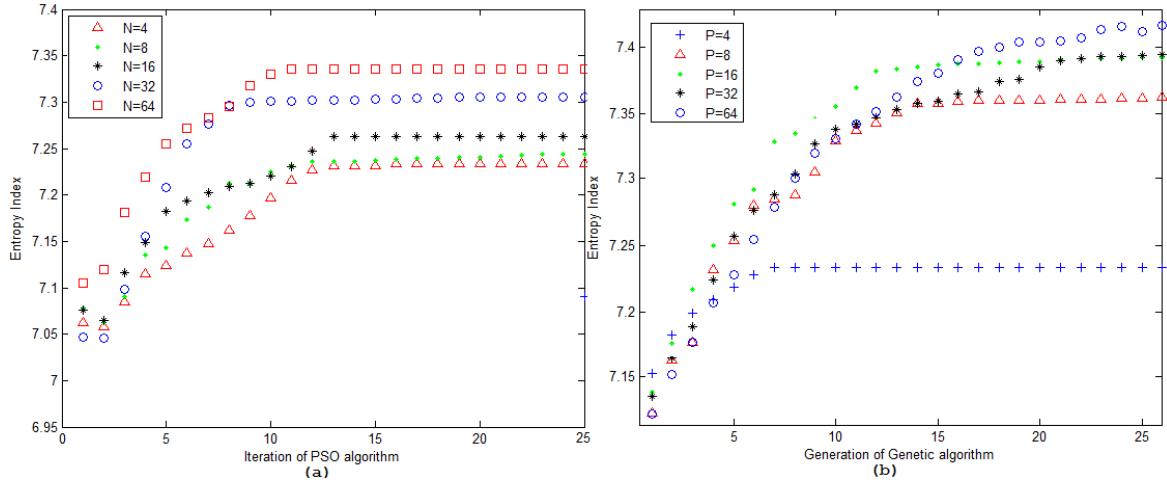


Figure 4.15: The entropy index of fused image obtained from “UN Camp” dataset at (a) different iterations of the PSO algorithm using different particles number and (b) different generation of genetic algorithm using different population sizes.

Table 4.6: The PSO parameters selected for the optimization.

Operation	Parameters
Particle Number	32
Inertia Weight	$W_{init} = 1.2$ , $W_{final} = 0.4$
Learning Factors	$C_1 = 2.4$ , $C_2 = 1.6$
Maximum Iteration	30

algorithm, as a good compromise between accuracy and complexity. In summary, the PSO parameters selected for the optimization problem are listed in Table 4.6.

#### 4.2.3.2 Genetic algorithm

In genetic algorithm (GA) [111], a candidate solution for a specific problem is called an individual or a chromosome and consists of a linear list of genes. Each individual stands for a point in the search space, and therefore a possible solution to the problem. A population consists of a finite number of individuals. Each individual is decided by an evaluating system to obtain its fitness value. Based on

this fitness value and undergoing genetic operators, a new population is iteratively generated with each successive population referred to as a generation.

Three basic genetic operators are sequentially applied to each individual with certain probabilities during each generation, i.e. selection, crossover (recombination), and mutation. First, a number of best-fit individuals are selected based on a user-defined fitness function. The remaining individuals are discarded. Next, a number of individuals are selected and paired with each other. Each individual pair produces one offspring by partially exchanging their genes around one or more randomly selected crossing points. At the end, a certain number of individuals are selected and the mutation operations are applied, i.e. a randomly selected gene of an individual abruptly changes its value. The GA is called a population-based technique because instead of operating on a single potential solution, it uses a population of potential solutions. The larger the population, the greater the diversity of the members of the population, and the larger the area searched by the population. Here, we have used Matlab optimization toolbox for implementation of genetic algorithm. Based on the experimental results (see Fig. 4.15 (b)) the population size is set to 16 and generation size is set to 20 for genetic optimization as a good compromise between accuracy and complexity.

#### 4.2.4 Computation time

It is interesting to evaluate the proposed method from a practical point of view that is the computation time. The proposed method consists of two steps with different computation times. In the first step, high frequency wavelet coefficients of the fused image are obtained using a simple fuzzy rule, and therefore it has low computation time. In the second step, a population-based optimization algorithm is used to generate low frequency wavelet coefficients of the fused image, and consequently the optimization algorithm increases the computation time.

Table 4.7 summarizes the relative computation time of the various image fusion methods considered in this paper. All computations have been performed on a

Table 4.7: Relative computation time of different image fusion techniques.

Method	Unit of time
Averaging	0.332
LT [34]	0.771
MS	0.739
WA [45]	0.821
WBV [47]	0.984
Lewis [49]	1.703
Proposed	2.253

Pentium IV personal computer, using a 2.00-GHz processor, running Windows Vista. The low computation times of different image fusion methods listed in Table 4.7 are because of their simplicity. In addition, the computation time of the proposed method can be better in some ways.

Since the proposed method is fully implemented using Matlab software, it is possible to use an optimum programming language such as C++ to reduce the computation time. In addition, a population-based algorithm can be used in parallel optimization form or synchronous algorithm, which speeds up the algorithm. A parallel optimization algorithm allows higher throughput (solving more complex problems in the same time span), more sophisticated finite element formulations, and higher accuracy.

### **4.3 Remote-sensing image fusion using multi-objective particle Swarm optimization and the shiftable Contourlet transform**

In this subsection, we have proposed a new hybrid method, in which the objective is to maintain spatial similarity of the pan-sharpened image to the Pan image and radiometric and quality in each band. The Pan and MS images are locally different because of different resolutions, and therefore we cannot directly

combine them in the spatial domain. For this reason, we have proposed a two steps algorithm, in which two initial results of the high-resolution MS images are generated in the first step. Then, in the second step, the final pan-sharpened image is obtained through a weighted combination of the two initial results using an optimization algorithm. Here, we have used a multi-resolution-based algorithm to generate the two initial results of the high-resolution MS or pan-sharpened images. The objective is to produce two modified high-resolution MS images, in which one has high spatial similarity to the Pan image and the other one has high radiometric quality. For this purpose, we have used the SCT, which is a new shiftable and modified version of the Contourlet transform and was proposed by Nguyen and Oraintara [89].

There are several ways for spatial details extraction of the Pan image. In the literatures, DWT is the common way for this purpose [62-63]. DWT provides a good time frequency analysis of the signal, with a non-redundant signal representation and an optimal representation of singularities. However, DWT suffers from five fundamental shortcomings [76]: oscillations, aliasing, shift variance, poor directionality, and absence of phase information. Shift invariance and directional selectivity are essential to the quality of wavelet-based image fusion. Because of the down-sampling operation in the DWT FB, the DWT is not shift invariant, and causes some visual artifacts in the pan-sharpened image. The human visual system is primarily sensitive to moving light stimulus; therefore, moving artifacts are highly distracting to the human observer [77]. In addition, if the directional selectivity of a FB is defined as the ability to extract directional features into separate images, the 2-D DWT has very poor directional selectivity, because 2-D DWT has four sub-images, which are usually referred to as LL, LH, HL, and HH images.

The CT is an alternative multi-resolution method, which provides an efficient directional representation and is efficient in capturing intrinsic geometrical structures of the natural image along the smooth contours [83]. Similar to the DWT, the CT is not shift-invariant and results in artifacts along the edges to some

extent. The NSCT was proposed to meet the shift invariance. The NSCT is the shift-invariant version of the CT and is built based on nonsubsampled pyramid decomposition and nonsubsampled filter banks, to obtain the shift invariance [87-88]. But, the NSCT has the overcomplete ratio of  $N \times K$ , where  $N$  is the number of resolution levels and  $K$  is the number of directions, which is very high compared to the overcomplete ratio of the CT ( $4/3$ ). However, a reduced form of translation invariance exists, namely, energy shift-invariance or “shiftability” [78], which means that the energy of the output signal is shift invariant. As for these reasons, we use the SCT, which is shown to have a number of desirable properties for image analysis including shiftable sub-band, arbitrarily high directionality, and low redundancy ( $11/3$ ) [89].

The SCT is applied for decomposing the histogram matched Pan and MS images into different resolutions and directions. In this step, two initial pan-sharpened images are generated using the maximum absolute selection and a new sign-based averaging fusion rules. Finally, the pan-sharpened image is obtained via a weighted combination of the two initial results, in which the weights are locally estimated via a multi-objective particle swarm optimization algorithm to generate a pan-sharpened image with high spatial and radiometric qualities. In the literature, the optimization problem is used for pan-sharpening, and the objective index is usually the mean square error (MSE). However, having only one objective is too simple to meet the real demands [116-117]. In fact, there are various kinds of evaluation indices, and different indices may be compatible or incompatible with one another, so a good evaluation index system of pan-sharpening must balance the advantages of different indexes. For this reason, we use the MOPSO for the optimization problem [118-119]. The MOPSO algorithm is adopted to generate a pan-sharpened image with high spatial and radiometric qualities based on global image fusion metrics. In this study, we have used image data from different satellite imagery consisting of Landsat-7 ETM + (28.5 m MS; 14.25 m Pan), Quickbird (2.44 m MS; 60 cm Pan), and Wordview2 (1.84 m MS; 46 cm Pan). In the following, the proposed pan-sharpening algorithm is presented [120].

Fig. 4.16 shows block diagram of the proposed method, which consists of a number of essential stages:

- 1) The input images are co-registered, and the MS data is resampled into the same spatial reference and grid as the Pan band.
- 2) The Pan and the MS images are histogram matched. This is because that the final pan-sharpened image be less sensitive to the different illumination condition in the Pan image.
- 3) In this step, an algorithm based on the shiftable contourlet transform is used to generate two initial results of the high-resolution MS images. The objective is to produce two modified high-resolution MS images, in which one has high spatial similarity to the Pan image and the other one has high radiometric quality. Therefore, we have used two different fusion rules to integrate high frequency contourlet coefficients of the Pan and MS images to generate two initial results of high-resolution MS image or the pan-sharpened image.
- 4) Finally, we can find an optimal pan-sharpened image applying the multi-objective particle swarm optimization and using the two initial pan-sharpened results. Specifically, the pan-sharpened image is obtained via a weighted combination of the two initial results, in which the weights are locally estimated via a multi-objective particle swarm optimization algorithm to generate a pan-sharpened image with high spatial and spectral qualities.

In the following subsections, we have provided more detailed explication of the image fusion process.

#### **4.3.1 Generation of two initial pan-sharpened results**

Injection the spatial details of the Pan image into the MS image, without reducing spectral quality or decreasing color distortion of MS images is the main challenge in the pan-sharpening methods. An ideal fusion method must be able to

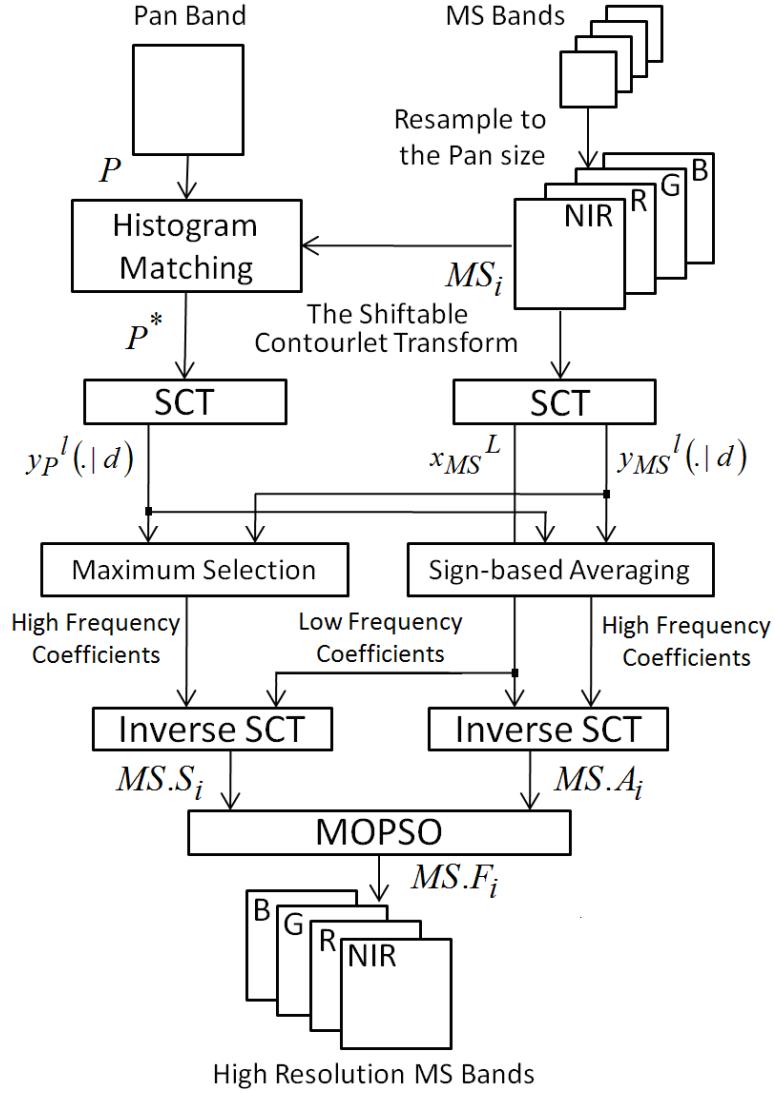


Figure 4.16: Flowchart of the proposed pan-sharpening algorithm.

simultaneously reach both spatial and spectral qualities, and not one at the cost of the other. In other words, the ideal fusion method should be able to preserve original spectral and spatial information of the MS images while increasing the spatial resolution [121]. For reaching to this goal, we have proposed a two steps algorithm, in which two initial results of the high-resolution MS images are generated in the first step. Then, in the second step, the final pan-sharpened image is obtained through a weighted combination of the two initial results using an optimization algorithm.

The Pan and MS images are locally different because of different resolutions, and therefore we cannot directly combine them in the spatial domain. For this reason, we generate two initial results, which are more appropriate for the weighted combination. We have used a multi-resolution-based algorithm to generate the two initial results. In addition, the SCT is used for the multi-resolution analysis. We have experimentally shown that the SCT is better transformation compared to the wavelet and contourlet transforms at the end of this Section.

The first initial result is obtained using a simple maximum selection fusion rule in the shiftable contourlet transform domain. The objective is to generate a high-resolution MS image with high spatial similarity to the Pan image. First, the Pan and MS images are decomposed using the SCT. Since, the high frequency contourlet coefficients reflect the image edges and detailed information; we have used the following formula to inject detailed information or high frequency coefficients of the Pan image into the MS image:

$$y_1^l(\cdot|d) = \begin{cases} y_P^l(\cdot|d) & |y_P^l(\cdot|d)| > |y_{MS}^l(\cdot|d)| \\ y_{MS}^l(\cdot|d) & \text{otherwise} \end{cases} \quad (4.51)$$

where the  $y_1^l(\cdot|d)$ ,  $y_P^l(\cdot|d)$  and  $y_{MS}^l(\cdot|d)$  are the high frequency shiftable contourlet coefficients of the first high-resolution MS, the Pan and low-resolution MS images, respectively.  $l$  is the decomposition level,  $d$  is the direction of contourlet coefficients in the transform domain, and  $(\cdot)$  is the shorthand notation of the spatial position.

In addition, the low frequency shiftable contourlet coefficients of the first high-resolution MS image are selected from the low-resolution MS image. This is because the low frequency coefficients contain the background information of an image, and therefore choosing background information from the low-resolution MS image in the first high-resolution MS image could preserve the color information.

The inverse SCT of the combined contourlet coefficients generates the first high-resolution MS image (denoted by  $PS1$ ):

$$PS1 = \Psi^{-1}(y_1, x_{MS}^L) \quad (4.52)$$

where  $\Psi$  is the inverse SCT,  $y_1$  is the high frequency contourlet coefficients using (4.51), and  $x_{MS}^L$  is the low frequency contourlet coefficients of the low-resolution MS image in the last decomposition level  $L$ .

The second initial result is obtained using a new sign-based averaging fusion rule in the shiftable contourlet transform domain. The objective is to generate a high-resolution MS image with high radiometric quality, while the detailed information from the Pan image is also available to some extent. The high frequency shiftable contourlet coefficients of the second high-resolution MS image are obtained using following formula:

$$y_2^l(\cdot | d) = \begin{cases} \frac{y_P^l(\cdot | d) + y_{MS}^l(\cdot | d)}{2} & si(y_P^l(\cdot | d)) = si(y_{MS}^l(\cdot | d)) \\ si(y_P^l(\cdot | d)) \times \frac{|y_P^l(\cdot | d) - y_{MS}^l(\cdot | d)|}{2} & otherwise \end{cases} \quad (4.53)$$

where the  $y_2^l(\cdot | d)$ ,  $y_P^l(\cdot | d)$  and  $y_{MS}^l(\cdot | d)$  are the high frequency shiftable contourlet coefficients of the second high-resolution MS, the Pan and low-resolution MS images, respectively.  $l$  is the decomposition level,  $d$  is the direction of contourlet coefficients in the transform domain,  $(\cdot)$  is the shorthand notation of the spatial position, and  $si(\cdot)$  is the sign function.

The simple averaging method produces a lot of near zero coefficients, because the coefficients may be in different signs. The near zero coefficients produce artifacts in the fused image after inverse transform. In order to avoid this problem, we propose a new method for averaging high frequency coefficients in (4.53), which takes sign of the coefficients into account. Similar to the first method, the low frequency shiftable contourlet coefficients of the low-resolution MS image are

used for low frequency coefficients of the second high-resolution MS image. The inverse SCT of the combined contourlet coefficients generates the second high-resolution MS image (denoted by  $PS2$ ):

$$PS2 = \Psi^{-1}(y_2, x_{MS}^L) \quad (4.54)$$

where  $\Psi$  is the inverse SCT,  $y_2$  is the high frequency contourlet coefficients using (4.53), and  $x_{MS}^L$  is the low frequency contourlet coefficients of the low-resolution MS image in the last decomposition level  $L$ .

The two initial high-frequency MS results, which are obtained for four samples of Pan and MS images, are shown in Fig. 4.17. As it can be seen in Fig. 4.17, the first high-frequency MS image has high spatial quality because of transferring detailed information or high frequency contourlet coefficients from both Pan and low-resolution MS images. In addition, the second high-resolution MS image, while has detailed information from the Pan image is so close to the low-resolution MS image and has high radiometric quality. Spatial and radiometric qualities can be also measured by some objective functions. Indeed, there are many different ways to analyze the spatial and radiometric qualities. Here we have used the relative dimensionless global error in synthesis (ERGAS) and the relative average spectral error (RASE) as the radiometric and spectral qualities. The ERGAS calculates the amount of spectral and radiometric distortions in the image [122]. The formula for the ERGAS is given by:

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{n=1}^N \left( \frac{RMSE(n)}{\mu(n)} \right)^2} \quad (4.55)$$

where  $h/l$  is the ratio between pixel sizes of the Pan and MS images,  $\mu(n)$  is the mean of the  $n$ th band, and  $N$  is the number of bands. In addition, the root mean squared error (RMSE) is:

$$RMSE = \sqrt{\frac{1}{L \times K} \sum_{i=1}^L \sum_{j=1}^K (F(i, j) - MS(i, j))^2} \quad (4.56)$$

where  $F$  is the high-resolution fused image,  $MS$  is the original multi-spectral image, and  $L \times K$  is the image size.

The RASE characterizes the average performance in the spectral bands [63]:

$$RASE = \frac{100}{M} \sqrt{\frac{1}{N} \sum_{n=1}^N RMSE^2(n)} \quad (4.57)$$

where  $M$  is the mean radiance of the  $N$  spectral bands of the original MS image.

The lower the values of the RASE and ERGAS indexes, the higher the radiometric and spectral qualities. To judge the spatial quality of the pan-sharpened image, we have used the method proposed by Zhou et al. [123]. First the Pan and the high-frequency MS images are filtered using the Laplacian filter:

$$\text{Laplacian filter} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Then, we compare the resulting filtered images by considering the correlation coefficients between each band and the Pan image. The closer the average correlation coefficient (CC) is to one indicating better spatial quality. We have obtained the three objectives (ERGAS, RASE, and CC to the Pan) for the two initial pan-sharpened results of the high-resolution MS images shown in Fig. 4.17. As it can be seen in Table 4.8, it is so clear that the first high-resolution MS image has high spatial similarity to the Pan image and the second one has high radiometric quality. Indeed, the results show that the two initial results have the information we need in the next step. Specifically, in the next step, useful information within the two initial results are transferred into the final pan-sharpened image using a weighted combination scheme, in which the weights are optimally estimated using the multi-objective particle swarm optimization.

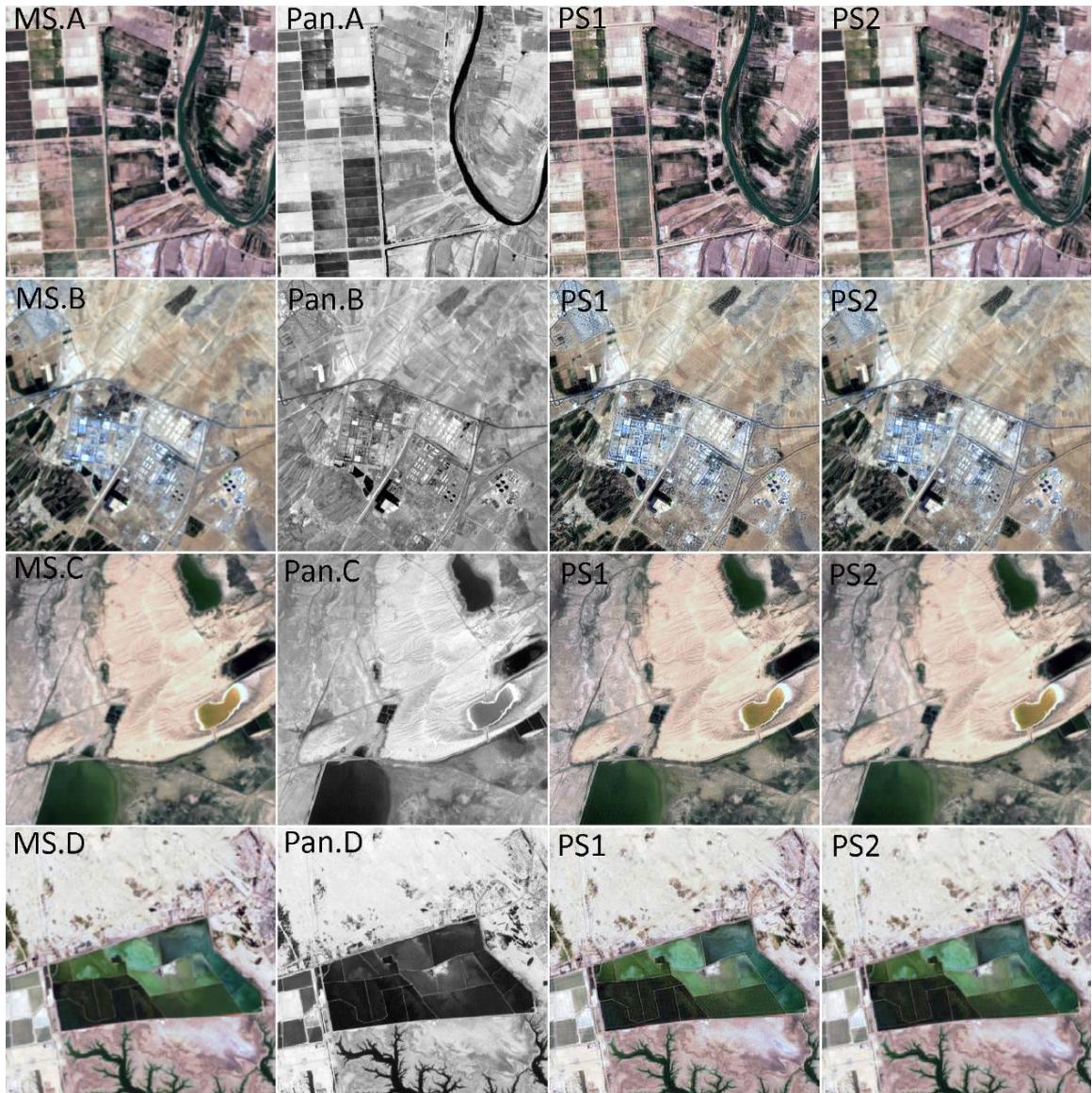


Figure 4.17: True-color composites of the two initial pan-sharpened images for different multi-spectral and panchromatic images from the Landsat-7 dataset.

### 4.3.2 Multi-objective optimization

Having the two initial results of the high-resolution MS images (*PS1* and *PS2*); the final pan-sharpened (*FPS*) image is obtained using a weighted combination of the *PS1* and *PS2* images, in which the weights are optimally estimated via a multi-objective particle swarm optimization (MOPSO) algorithm to generate a pan-sharpened image with high spatial and radiometric qualities:

Table 4.8: Different objective measures for the two initial pan-sharpened results from different datasets shown in Fig. 4.17.

Image		Index	Radiometric Quality		Spatial Quality
			ERGAS	RASE	CC to the Pan
A	PS1		2.78	11.15	<b>0.966</b>
	PS2		<b>1.14</b>	<b>4.53</b>	0.901
B	PS1		2.77	11.14	<b>0.965</b>
	PS2		<b>1.24</b>	<b>4.99</b>	0.896
C	PS1		1.43	5.75	<b>0.961</b>
	PS2		<b>0.97</b>	<b>3.74</b>	0.912
D	PS1		1.52	6.08	<b>0.971</b>
	PS2		<b>0.704</b>	<b>2.81</b>	0.906

$$FPS_i = W1_i \times PS1_i + W2_i \times PS2_i \quad (4.58)$$

where  $FPS$ ,  $PS1$  and  $PS2$  are the final pan-sharpened image and the two initial results, respectively.  $i$  is the number of bands of the MS image,  $W1$  and  $W2$  are the values in the range of  $[0,1]$ , and  $W1 + W2 = 1$ . In addition, the weights are locally estimated using MOPSO algorithm. The  $W1$  and  $W2$  parameters are not the same dimensions as the  $PS1$  and the  $PS2$  images, and the ' $\times$ ' operation does not represent Hadamard products between pairs of images. It should be mentioned that the parameters ( $W1$  and  $W2$ ) are separately obtained in each sliding window ( $M \times N$ ) and each band of the two initial results. For an MS image with size  $L \times K \times 4$ , the numbers of parameters are  $L \times K \times 4 / M \times N$ .

As we have mentioned, the final pan-sharpened image should have high spatial and spectral qualities. In addition, the spatial and radiometric qualities are measured by different indexes. Therefore, we have chosen two objectives for the optimization problem. In fact, the optimization problem is a bi-objective optimization, which is a subset of multi-objective optimization.

The first objective is the correlation coefficient (CC) to the Pan image, which measures the spatial quality of the pan-sharpened image. We have measured the CC to the Pan in each band and each sliding window ( $M \times N$ ) of the pan-sharpened ( $PS$ ) image in (4.58) compared to the original Pan image using the method proposed by Zhou et al. [123]. First the Pan and the  $PS$  images are filtered using the Laplacian filter. Then, we compare the resulting filtered images by considering the correlation coefficients between each band of  $PS$  and the Pan images.

$$F_1 = CC(FPS, Pan) = \frac{\sum_{m=1}^M \sum_{n=1}^N (PS_{n,m} - \bar{PS})(Pan_{n,m} - \bar{Pan})}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N (PS_{n,m} - \bar{PS})^2 (Pan_{n,m} - \bar{Pan})^2}} \quad (4.59)$$

where  $FPS$  and  $Pan$  are the final pan-sharpened and original panchromatic images, and  $M \times N$  is the size of sliding window.

The second objective is the peak signal to noise ratio (PSNR), which measures the radiometric quality of the pan-sharpened image. As we have mentioned in the previous subsection, the ERGAS and RASE are the standard indexes to measure the radiometric and spectral qualities, but we cannot directly use them here. This is because the ERGAS and RASE are global indexes which are obtained using  $N$  bands of pan-sharpened images compared to the original MS images, but here we measure radiometric quality in each band of the pan-sharpened image compared to the corresponding band of the low frequency MS image. Therefore, we have used PSNR as the radiometric quality. We have measured the PSNR in each sliding window ( $M \times N$ ) of the pan-sharpened ( $PS$ ) image in (4.58) compared to the low-resolution MS image using the following:

$$F_2 = PSNR(FPS, MS) = 10 \log \frac{255^2}{RMSE^2} \quad (4.60)$$

in which RMSE (root mean square error) is defined as:

$$RMSE^2 = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N [FPS(n, m) - MS(n, m)]^2 \quad (4.61)$$

where  $FPS$  and  $MS$  are the final pan-sharpened and low-resolution MS images, and  $M \times N$  is the size of sliding window.

In other words, the optimization is a maximization problem, and the goal is to estimate the optimal parameters ( $W_1$  and  $W_2$ ) in each sliding window of the two initial results, which simultaneously maximize the two objectives expressed in (4.59) and (4.60). For this reason, we have used the multi-objective particle swarm optimization, which will be explained in the next subsection.

### 4.3.3 Multi-objective particle swarm optimization

Kennedy and Eberhart [110] firstly proposed the PSO algorithm for optimization. PSO is a population-based search algorithm based on the simulation of the social behavior of birds within a flock. In PSO, particles are flown through hyper-dimensional search space. Changes to the position of the particles within the search space are based on the social-psychological tendency of individuals to emulate the success of other individuals [124].

The position of each particle is changed according to its own experience and that of its neighbors. Let  $x_i(t)$  denotes the position of particle  $p_i$ , at iteration  $t$ . The position of  $p_i$  is then changed by adding a velocity  $v_i(t)$  to the current position, i.e.:

$$x_i(t) = x_i(t-1) + v_i(t) \quad (4.62)$$

The velocity vector reflects the socially exchanged information, and generally is defined in the following way:

$$v_i(t) = W \times v_i(t-1) + C_1 r_1 \times (x_{pbest_i}(t) - x_i(t)) + C_2 r_2 \times (x_{leader}(t) - x_i(t)) \quad (4.63)$$

where  $W$  is the inertia weight employed to control the impact of the previous history of velocities on the current velocity of a given particle, and  $x_{pbest}$  has been

the personally best position of a given particle, so far. That is, the position of the particle that has provided the greatest success.  $x_{leader}$  is the position of the best particle of the entire swarm.  $C_1$  and  $C_2$  are learning factors and represent the attraction that a particle has toward either its own success or that of its neighbors, and  $r_1, r_2 \in [0,1]$  are random values.

In summary, the way in which the general single objective PSO algorithm works is as follows: First, the swarm is initialized. This initialization includes both positions and velocities. The corresponding  $pbest$  of each particle is initialized and the *leader* is located. Then, for a maximum number of iterations, each particle flies through the search space updating its position using (4.62) and (4.63), and its  $pbest$  and, finally, the *leader* are updated, too.

In order to apply the PSO strategy for solving multi-objective optimization problems, it is obvious that the original scheme has to be modified. The solution set of a problem with multiple objectives does not consist of a single solution. Instead, in the multi-objective optimization, we aim to find a set of different solutions (the so-called Pareto optimal set). There are many approaches to implement multi-objective problems using PSO. Here, we use Pareto approach, which consists of finding as many non-dominant solutions as possible and returning a set of non-dominant solution to the user [124].

In this method, the  $x_{pbest}$  and  $x_{leader}$  are updated at each iteration to minimize the distance of the Pareto front produced by the proposed algorithm with respect to the true (global) Pareto front (assuming we know its location). In the proposed algorithm, the  $x_{pbest}$  is updated at each iteration using the following:

$$x_{pbest_i}(t) = \begin{cases} x_i(t) & \text{if } F_1(x_i(t)) > F_1(x_i(t-1)) \text{ and } F_2(x_i(t)) > F_2(x_i(t-1)) \\ x_{pbest_i}(t-1) & \text{otherwise} \end{cases} \quad (4.64)$$

where  $x_{pbest}$  is the personal best position of a given particle,  $i$  is the particle number,  $t$  is the iteration, and  $F_1$  and  $F_2$  are the two objectives defined by (4.59) and (4.60).

In addition, a single *leader* is selected between the Pareto front solutions, which are stored in the external archive. In order to obtain Pareto front solutions, first the dominant solutions are selected between all the particles using the following:

$$x_d(i) = \begin{cases} x_i(t) & \text{if } F_1(x_i(t)) > F_1(x_i(t-1)) \text{ and } F_2(x_i(t)) > F_2(x_i(t-1)) \\ \phi & \text{otherwise} \end{cases} \quad (4.65)$$

where  $x_i(t)$  denotes the position of particle,  $i$  is the particle number,  $t$  is the iteration, and  $F_1$  and  $F_2$  are the two objectives defined by (4.59) and (4.60).

Having the dominant solutions, the Pareto front solutions are defined as follows:

$$PF(l)_{k=1,2\dots N, l\neq k} = \begin{cases} x_d(l) & \text{if } F_1(x_d(l)) > F_1(x_d(k)) \text{ and } F_2(x_d(l)) > F_2(x_d(k)) \\ \phi & \text{otherwise} \end{cases} \quad (4.66)$$

where  $x_d(l)$  are the dominant solutions between all particles, and  $F_1$  and  $F_2$  are the two objectives defined by (4.59) and (4.60).  $l=1,2\dots N$ , and  $N$  is the number of dominant solutions.

As an example, Fig. 4.18 (a) shows the particles at first and second iterations, and Fig. 4.18 (b) shows the dominant and Pareto front solutions. After finding Pareto set, they are stored to an external archive, and then the external archive is updated with respect to the previous stored Pareto front set using the following:

$$PF^*(l)_{k=1,2\dots N, l\neq k} = \begin{cases} PF(l) & \text{if } F_1(PF(l)) > F_1(PF(k)) \text{ and } F_2(PF(l)) > F_2(PF(k)) \\ \phi & \text{otherwise} \end{cases} \quad (4.67)$$

where  $PF$  are the Pareto front solutions defined by (4.66), and  $F_1$  and  $F_2$  are the two objectives defined by (4.59) and (4.60).  $l=1,2\dots N$ , and  $N$  is the number of Pareto front solutions in the external archive.

Now, a single leader is chosen between the updated Pareto front set, which has minimum Euclidian distance to the true Pareto front. We have chosen the best

values of the two objectives as the true Pareto front, which are obtained for the final pan-sharpened image produced by the proposed algorithm using (4.58) in each iteration of multi-objective optimization. As we have explained about the optimization procedure, a number of particles are selected to estimate each parameter. Indeed the true Pareto front is related to a particle, which results in best values of the two objectives for the final pan-sharpened image:

$$TP_1 = \max_{j=1,2,\dots,NP} CC(FPS_j, Pan) \quad (4.68)$$

$$TP_2 = \max_{j=1,2,\dots,NP} PSNR(FPS_j, MS) \quad (4.69)$$

where  $FPS$  is the final pan-sharpened image using (4.58), Pan and MS are the original panchromatic and multi-spectral images, CC is the correlation coefficient defined by (4.59), and PSNR is defined by (4.60).  $j$  is the particle number,  $NP$  is the number of particles, and  $TP_1$  and  $TP_2$  are the true Pareto front.

It should be mentioned that the true Pareto front are updated in each iteration. After determining the true Pareto front by (4.68) and (4.69), the *leader* is selected between the updated Pareto front set, which has minimum Euclidian distance to the true Pareto front:

$$d^2(i) = [F_1(PF^*(i)) - TP_1]^2 + \alpha^2 \times [F_2(PF^*(i)) - TP_2]^2 \quad (4.70)$$

Then,

$$\text{If } d(i) = \min_{j=1,2,\dots,K} d(j), \text{ Then Leader} = PF^*(i) \quad (4.71)$$

where  $PF^*$  are the updated Pareto front solutions,  $F_1$  and  $F_2$  are the two objective defined by (4.59) and (4.60),  $TP_1$  and  $TP_2$  are the true Pareto front,  $K$  is the number of updated Pareto front solutions in the external archives, and  $\alpha = TP_1/TP_2$ . For better enlightenment, an example of the true Pareto front, the Euclidian distances

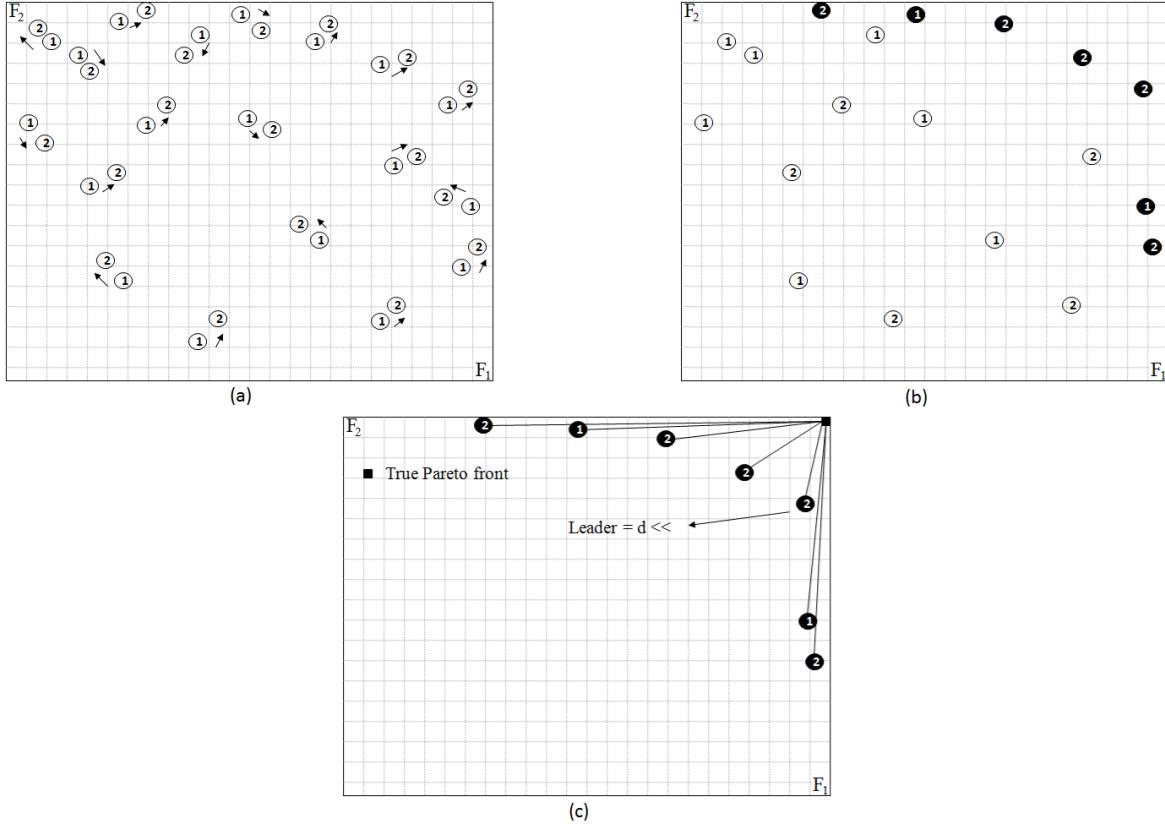


Figure 4.18: (a) The particles at first and second iterations, (b) the dominant (white circles) and Pareto front solutions (black circles), and (c) the true Pareto front, the Euclidian distances between Pareto front set and the true Pareto front, and the *leader*.

between Pareto front set and the true Pareto front, and the *leader* are shown in Fig. 4.18 (c).

#### 4.3.4 Parameters selection for the proposed pan-sharpening algorithm

In this sub-section, the processing parameters for the proposed pan-sharpening algorithm are determined. The parameters consist of the decomposition level of the SCT to generate two initial pan-sharpened images, the block size of the sliding window for local parameter estimation in (4.58), and the PSO parameters selected for the multi-objective optimization.

As we have mentioned, the model parameters ( $W_1$  and  $W_2$ ) in (4.58) are locally estimated over a sliding window. Therefore, the block size for local parameter estimation must be defined. Fig. 4.19 (a) and (b) shows the spatial and radiometric

qualities of the final pan-sharpened image versus the block size for local parameter estimation. In addition, Fig. 4.19 (c) shows the computation time of the proposed algorithm using different block sizes for multi-objective optimization. In this experiment, we have used the average results obtained for 20 MS ( $512 \times 512$ ) and corresponding Pan ( $1024 \times 1024$ ) images from our Landsat-7 ETM+ datasets. In addition, we have used two decomposition levels for the SCT and the PSO parameters are manually set, which are listed in Table 4.9. It should be mentioned that the number of iterations and particles is set based on convergence of the optimization algorithm. More particles could speed up the convergence of the PSO algorithm; however, it increases its computation time. In addition, using too much particles and iterations is ineffective on the convergence of PSO. The inertia weight ( $w$ ) is set to control the impact of the previous history of velocities on the current velocity of a given particle. High values of  $w$  gives a global search and low values of  $w$  gives a local search. Indeed, we want to consider both the local and global search properties of the inertia weights in the PSO algorithm, and therefore give a medium value to it. In addition, the learning factors and the inertia weight are linearly decreased with respect to the iteration. In other words, we have used dynamic learning factors and inertia weight. This is because the larger inertia weights and the learning factors at the beginning help to find good seeds and the later small inertia weights facilitate a fine search.

Table 4.9: The PSO parameters selected for the multi-objective optimization.

Operation	Parameter
Particle Number	15
Inertia Weight	$W = 0.65$
Learning Factors	$C_1 = 1, C_2 = 1$
Maximum Iteration	40

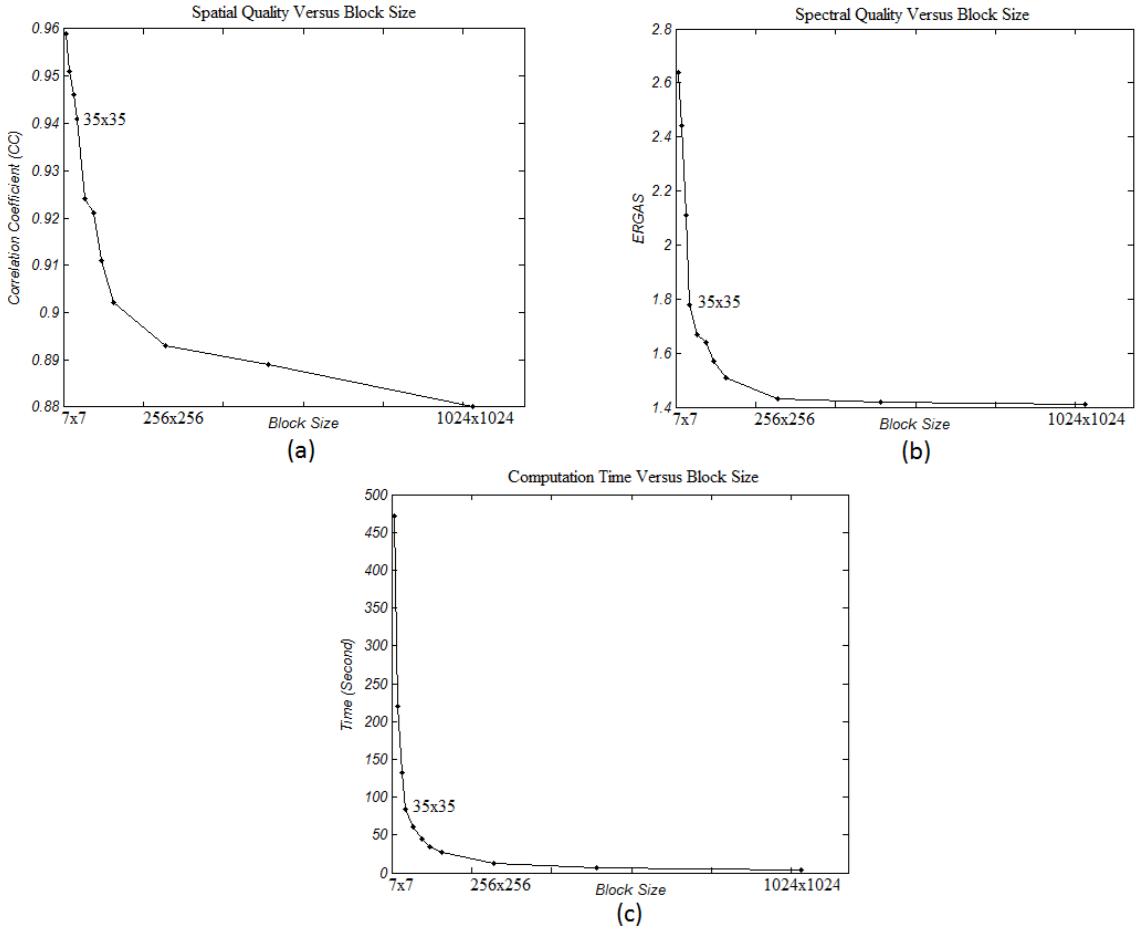


Figure 4.19: (a) Overall CC to the Pan index, (b) ERGAS index, and (c) computation time of the proposed algorithm versus the size of the sliding window for local parameter estimation.

As it can be seen from Fig. 4.19 (a) and (b), the smaller block size results in a pan-sharpened image with higher spatial similarity (CC to the Pan), and larger block size results in a pan-sharpened image with higher radiometric and spectral qualities (ERGAS, and RASE). In addition, the smaller block size increases the number of local parameters and the computation time. Therefore, there is tradeoff between spatial and radiometric qualities of the final pan-sharpened image and the computation time of the proposed algorithm. Based on the experiments, we have selected  $35 \times 35$  block size for local parameter estimation, which results in acceptable spatial and radiometric qualities of the final pan-sharpened image and the computation time.

Another parameter is the decomposition levels of the SCT to generate the two initial pan-sharpened images in the proposed algorithm. Table 4.10 shows the decomposition levels used in the algorithm of the SCT versus the spatial and radiometric qualities of the final pan-sharpened image for different satellite imagery. In this experiment, we have used of  $35 \times 35$  block size for local parameter estimation. Table 4.10 shows that the decomposition level which maximizes both of the spatial and radiometric qualities of the final pan-sharpened image for the Landsat-7 dataset is two levels. Our explanation for this selection is that the gap for the spatial quality (CC to the Pan) between first and second decomposition levels is higher than the gap between second and third decomposition levels. While the gap for the radiometric quality (ERGAS and RASE) between second and third decomposition levels is very higher than the gap between first and second decomposition levels. Using the similar judgment, three decomposition levels are selected for the Quickbird and Wordview2 satellites datasets.

Table 4.10: Average performance results obtained from different MS and Pan images in the datasets using different decomposition levels of the SCT in the proposed algorithm.

Decomposition Levels	CC to the Pan			ERGAS			RASE		
	L-7	Q	W-2	L-7	Q	W2	L-7	Q	W-2
1	0.901	0.897	0.889	1.26	1.19	1.05	5.09	5.20	4.83
2	<u>0.947</u>	0.928	0.921	<u>1.78</u>	1.57	1.23	<u>7.28</u>	6.87	6.29
3	0.951	<u>0.951</u>	<u>0.949</u>	2.76	<u>2.34</u>	<u>2.85</u>	11.09	<u>9.11</u>	<u>9.72</u>
4	0.957	0.957	0.951	3.24	3.97	4.13	13.04	13.79	14.11
5	0.962	0.961	0.958	3.64	4.32	4.56	14.64	16.14	16.56
6	0.968	0.969	0.963	4.12	4.98	5.04	15.53	17.11	17.34

### 4.3.5 Comparisons between different multi-resolution transformations in the proposed algorithm

Another important aim of this paper is to show that the SCT is a better transformation approach than the CT, and the DT-DWT. As we have discussed in the Chapter 3 that DWT suffers from shift variance and poor directionality, which are essential to the quality of wavelet-based image fusion [77]. DT-DWT is a modified version of the DWT and was proposed to overcome shift variance and directionality limitations of the DWT while maintaining the perfect reconstruction property with limited redundancy [80]. In addition, the CT is an alternative multi-resolution method provides an efficient directional representation and is efficient in capturing intrinsic geometrical structures of the natural image along the smooth contours [83]. Similar to the DWT, the CT is not shift-invariant and results in artifacts along the edges to some extent. We have mentioned in Chapter 3 that the SCT is the new shiftable and modified version of the CT. Indeed, the relation of the SCT to the CT is similar to that of the DT-DWT to the DWT. Therefore, the SCT is a better transformation than the CT because of the shift-invariant property, and better than the DT-DWT because of the directional selectivity.

First, we have objectively shown that the SCT is better transformation compared to the CT and DT-DWT transforms for transferring the spatial details of the Pan image into the high-resolution MS image, without reducing radiometric quality. Table 4.11 shows the objective results obtained from 20 MS and Pan images in the database using different multi-resolution transformations (DT-DWT, CT, and SCT) in the proposed algorithm.

In this experiment, we have used the obtained results for 20 MS ( $512 \times 512$ ) and corresponding Pan ( $1024 \times 1024$ ) images from the Landsat-7 ETM+ datasets. We have used two decomposition levels for the DT-DWT, 2 decomposition levels and 64 directions in each level of decomposition for the SCT and CT. For the DT-DWT-based method, the available “AntonB” mother wavelet is used. For the CT-based method, the available “9-7” is used for pyramidal filter and the “pkva” is used

for the directional filter banks, and for the SCT-based method the “nalias” is used for the Laplacian low pass filter, “pkva” is used for the diamond or fan filter, and “9-7” is used for the wavelet filter.

For visual comparison, we have generated a pair of Pan and MS images, in which the Pan image is shifted up to two pixels in both vertical and horizontal directions with respect to the MS image. Then we have used the DT-DWT and CT instead of SCT in the proposed pan-sharpening algorithm to generate the two initial pan-sharpened images using (4.52) and (4.54). Fig. 4.20 shows the visual comparison of different methods. It can be seen in Fig. 4.20, using the SCT in the proposed algorithm results in a pan-sharpened image with higher spatial and radiometric qualities compared to the other transformations. The result of using CT in the proposed algorithm causes artifacts around the edges. Our explanation for this is that the SCT is a shiftable transform, while the CT is not shift invariant, and therefore a shift dependent fusion method leads to unstable and flickering results. The human visual system is primarily sensitive to moving light stimuli, so moving artifacts, when introduced by the fusion process, are highly distracting to the human observer [77]. In addition, the detailed information of the SCT+MOPSO approach is better compared to the DT-DWT+MOPSO, which indicates that Contourlet are better than wavelet for spatial detail injection, because the DT-DWT has only six directional sub-bands, while we have used 64 directions in the SCT method in each level of decomposition.

#### 4.4 Summary

In the Chapter just completed, the proposed methods for image fusion were described in different applications. In the first application, new wavelet-based method for fusion of spatially registered multi-focus images is proposed. Generally, we have formulated the image fusion process as a two-class classification problem: in and out of focus classes. First, a twelve dimensional feature vector using DT-DWT sub-bands of the source images are extracted, and then a trained two-class fisher classifier projects it to the class labels. The classifier output is used as a

Table 4.11: Performance results obtained from 20 MS and Panchromatic images using different multi-resolution transformations in the proposed algorithm.

Index		Multi-Resolution Transformation		
		SCT	CT	DT-DWT
CC to the Pan	Min	0.929	0.912	0.906
	<b>Ave.</b>	<b>0.947</b>	0.933	0.926
	Max	0.958	0.948	0.943
	Std.	0.0157	0.0190	0.0187
ERGAS	Min	1.33	1.56	1.65
	<b>Ave.</b>	<b>1.78</b>	<u>2.43</u>	<u>2.85</u>
	Max	2.01	2.91	3.23
	Std.	0.389	0.754	1.062
RASE	Min	6.01	6.78	7.08
	<b>Ave.</b>	<b>7.28</b>	8.45	9.23
	Max	8.53	10.21	10.76
	Std.	1.26	1.71	1.91

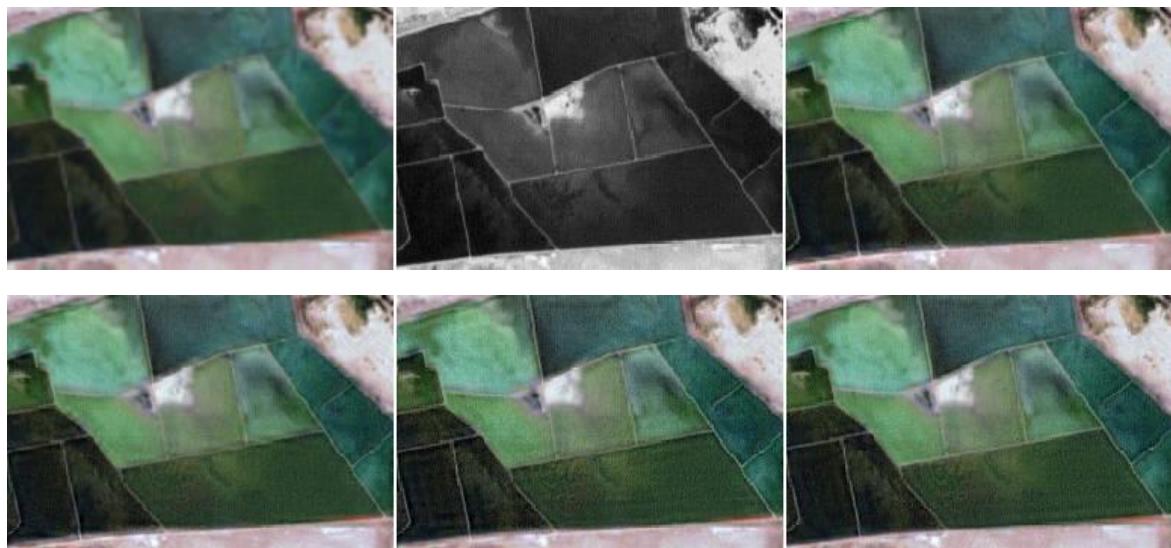


Figure 4.20: Visual comparison between different multi-resolution transformations in the proposed algorithm. Top row, left to right: Low resolution MS image, Pan image with two pixels shift, and result of DT-DWT+MOPSO. Bottom row: Results of CT+MOPSO, SCT+MOPSO, and SCT+MOPSO without shifting.

decision map for fusing high frequency wavelet coefficients of multi-focus source images in different directions and decomposition levels of the DT-DWT. In addition, there is an uncertainty for selecting high frequency wavelet coefficients in smooth regions of source images, which causes some misclassified pixels in the classification output or the decision map. In order to solve this uncertainty and integrate as much information as possible from the source images into the fused image, we propose an algorithm based on fuzzy logic, which combines outputs of two different fusion rules based on a dissimilarity measure from the source images: Selection based on the decision map and weighted averaging. An estimation of the decision map is also used for fusing low frequency wavelet coefficients of the source images instead of simple averaging. After fusing low and high frequency wavelet coefficients of the source images, the final fused image is obtained using the inverse DT-DWT.

In the second research topic, a new wavelet-based algorithm for the fusion of spatially registered infrared and visible images is presented. We specifically propose new fusion rules for fusion of low and high frequency wavelet coefficients of the source images in the second step of the wavelet-based image fusion algorithm. First, the source images are decomposed using DT-DWT. Then, a fuzzy-based approach is used to fuse high frequency wavelet coefficients of the IR and visible images. Particularly, Fuzzy logic is used to integrate the outputs of three different fusion rules (weighted Averaging, selection using PDM, and selection using RDM), based on a dissimilarity measure of the source images. The objective is to utilize the advantages of previous pixel- and region-based methods in a single scheme. The PDM is obtained based on local activity measurement in the DT-DWT domain of the source images. A new segmentation-based algorithm is also proposed to generate the RDM using the PDM. In addition, a new optimization-based approach using population-based optimization is proposed for the low frequency fusion rule instead of simple averaging. After fusing low and high frequency wavelet coefficients of the source images, the final fused image is obtained using the inverse DT-DWT.

Finally, in the last research subject in this dissertation, a novel approach based on MOPSO is presented for panchromatic sharpening of multispectral image. This new method could transfer spatial details of the Pan image into the high-resolution version of the MS image, while color information from the low-resolution MS image is well preserved. The Pan and MS images are locally different because of different resolutions, and therefore we cannot directly combine them in the spatial domain. For this reason, we generate two initial results, which are more appropriate for weighted combination. First, the Pan and the MS images are histogram matched. Then we use SCT for decomposing the histogram matched Pan and MS images. The SCT is a new shiftable and modified version of the Contourlet transform. In this step, an algorithm based on the SCT is used to generate two initial results of the high-resolution MS images. Our objective is to produce two modified high-resolution MS images, in which one has high spatial similarity to the Pan image and the other one has high radiometric quality in each band. Therefore, we have used two different fusion rules to integrate high frequency contourlet coefficients of the Pan and MS images to generate two initial results of high-resolution MS image or the pan-sharpened (PS) image. Finally, we can find an optimal PS image applying the MOPSO algorithm and using the two initial PS results. Specifically, the PS image is obtained via a weighted combination of the two initial results, in which the weights are locally estimated via a multi-objective particle swarm optimization algorithm to generate a PS image with high spatial and radiometric qualities.

## Chapter 5. Experimental results

In this Chapter, we will compare the proposed image fusion algorithms with other well-known methods. Generally, there are two ways for evaluation of image fusion results including subjective and objective indices as shown in Fig. 5.1. Usually in cases where there is a ground-truth image to compare performance (such as combining images with different focuses), the objective evaluation method will be a method with more reliability. Although in most cases, there is not a proper ground-truth image for comparing performance. In these cases, quantitative or objective assessment is performed without a ground-truth image. In addition, it is necessary to do the subjective assessment alongside the objective evaluation.

An example of a system using a subjective audience to optimize an image fusion system is shown in Fig. 5.1 a). Image fusion of the outputs of two imaging sensors surveying a scene in the environment produces a fused representation of the scene. This representation is displayed to an audience which evaluates it, either by comparing it to input images and other fused images, or by performing tasks using the information obtained from the display. The responses from the subjects are usually not in a readily usable format and have to be processed to obtain a meaningful performance evaluation. Only then can the result be passed back to the image fusion system in order to improve its parameters.

All the prerequisites of subjective tests defined above, are costly and highly impractical. In addition, their complexity means that subjective tests can not be run

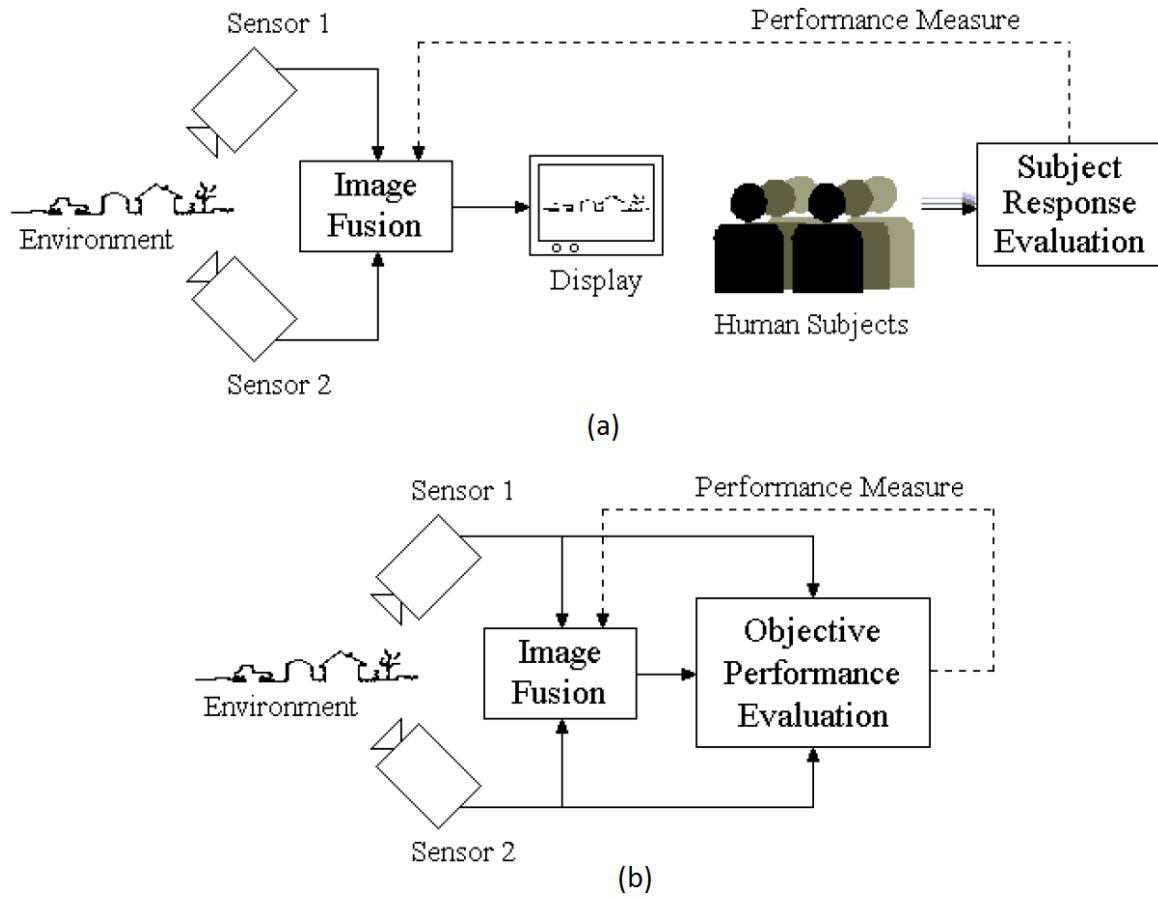


Figure 5.1: (a) Subjective and (b) objective, and fusion parameter optimization

in real time and generally require a lengthy planning and organization period. As a result, the need arose for automated, objective, measures of pixel-level image fusion performance. The basic advantages of such objective metrics over subjective methods is illustrated in Fig. 5.1 b), as compared to the above example of subjective fusion parameter optimization. The objective performance assessment system needs no display equipment and all the images are automatically processed, either in hardware or software. This means that there is a significant reduction in time taken to convey the evaluations back to the fusion process. In fact, using such an arrangement implemented in software also enables conventional optimization algorithms to be successfully applied to fusion. Furthermore, there is no need for any effort on the part of the subjects (decreasing the cost) as the system can be made to operate with the bare minimum of supervision.

In this Chapter, the qualitative and quantitative evaluation will be used to fit various applications of image fusion. The experimental results are discussed in three parts. In subsection 5.1, the results of the multi-focus image fusion are provided. Then, the results of the visual and infrared image fusion will be presented in subsection 5.2. Finally, in subsection 5.3, the results of the remote sensing image fusion will be discussed.

## 5.1 The results of multi-focus image fusion

In this Section, we compare our proposed multi-focus image fusion algorithm with some of the best state-of-the-art fusion methods. The proposed image fusion method was tested against several methods consisting of the simple averaging, principle component analysis (PCA)<sup>1</sup> [125], spatial frequency (SF)<sup>2</sup> [21], and the Dual-Tree Discrete Wavelet Transform<sup>3</sup> with three different fusion rules: Maximum Selection (MS), Weighted average (WA) [45], and Window-Based Verification (WBV) [39]. For the DT-DWT based methods, the available “AntonB” mother wavelet is used. In addition, the result of using the handmade decision map in the DT-DWT based image fusion is used as the best result (Best). In other words, the handmade decision map is used for selecting wavelet coefficients between the source images, instead of using the classification output as the decision map in the proposed algorithm.

The evaluation indices of image fusion include subjective and objective indices. Objective indices can be divided into three categories [126]. One category reflects the image features, such as entropy and gradient. The second reflects the relation of the fused image to the source images, such as mutual information and Petrovic index [107]. Finally, the third reflects the relation of the fused image to the standard image, such as correlation coefficient and peak signal to noise ratio (PSNR).

---

<sup>1</sup> Image Fusion Toolbox for MATLAB developed by Oliver Rockinger: <http://www.metapix.de/toolbox.htm>

<sup>2</sup> The Image Fusion Toolkit for Matlab developed by Eduardo Canga: <http://www.imagefusion.org/>

<sup>3</sup> Available at: [taco.poly.edu/WaveletSoftware/dt2D.html](http://taco.poly.edu/WaveletSoftware/dt2D.html)

Two metrics are considered in this paper. The first metric is the Xydeas and Petrovic metric, proposed in [107], which considers the amount of edge information transferred from the input images to the fused images using a Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images. The second metric is the PSNR index, which is defined as:

$$PSNR = 10 \log \frac{255^2}{RMSE^2} \quad (5.1)$$

in which the root mean square error (RMSE) is defined as the following formula:

$$RMSE^2 = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [R(x, y) - F(x, y)]^2 \quad (5.2)$$

An image with higher PSNR index will have better fused image. It should be mentioned that for image fusion experiments, a ground-truth image was prepared by cutting and pasting method [39], which is performed by transferring in focus regions of source images into one image. Image fusion results will be discussed in the two following subsections.

### 5.1.1 Fusion results of natural images

In the first experiment, nine pairs of natural multi-focus source images, as shown in Figs. 4.3 and 5.2, are used to evaluate the proposed method. The multi-focus images consisting of “Disk”, “Lab”, “Pepsi” and “Clock” are publicly available at the Image fusion web site [16], “Flower” and “Book” are publicly available at [127], and additional multi-focus images consisting of “Doll”, “Rose” and “Jug” are provided by us.

Fig. 5.3 shows objective results for different multi-focus image fusion algorithms. It should be mentioned that the average results between nine pair test images are presented. It can be seen in Fig. 5.3 that our proposed image fusion



Figure 5.2: Multi-focus source images, from top left, clockwise: “Pepsi”, “Clock”, “Jug”, and “Lab” images.

algorithm has better PSNR index compared to the other methods (more than +4.5 dB), and the obtained results for the proposed method are very close to the best results (less than 2.2 dB). The best result is related to manually picking in focus wavelet coefficients from the transformed source images using hand-made decision map to generate best possible fused image in the wavelet domain. In addition, evaluation of different methods using Petrovic metric has shown that our proposed algorithm has better performance for transferring edges information from source images to the corresponding fused image. When looking more closely at the results, we observe the following:

- Synthesizing the composite image by averaging the corresponding pixels of the source images is a simple method for image fusion. In addition to simplicity, this method usually causes many undesirable effects on the fused images such as low contrast. The averaging method produces worst results in terms of PSNR and Petrovic indexes compared with other methods.
- The PCA method produces fused image by weighted averaging, in which the weights for each input image are obtained from the eigenvector corresponding

to the largest eigenvalue [125]. The PCA approach also generates weak objective results compared with other methods.

- Shutao et al. in [21] developed a method based on spatial frequency, which is computationally simple and can be used in real-time applications. The spatial frequency measures the overall activity level in an image. The spatial frequency method suffers from blocking effect, and this problem leads to weak objective results. The consequence of blocking effect on the Petrovic index is more than SNR metric. Because the Petrovic index measures similarity between edges information from source images and the fused image, and the blocking effect, change the edges information of fused image compared to the source images.
- As we have discussed in Chapter 4, previous wavelet-based methods (MS, WA, and WBV) cannot produce reliable decision maps, and therefore they result in lower performances compared to the proposed method. Another important reason is that these methods use averaging method for low frequency fusion rule. As we have shown in subsection 4.1.4, the proposed method for low frequency fusion rule outperforms averaging method. The WBV method produces best results among the previous wavelet-based method. Because it uses a window-based verification to modify the decision maps, which is obtained using maximum selection method for high frequency fusion rule.
- The proposed method, which is based on fuzzy logic and classification, generates best results compared to other methods. Generation of very accurate decision maps for high and low frequency fusion rules is the main reason for outshining the proposed method.

In addition to objective assessment, two samples have been selected here to subjectively evaluate different results. Fig. 5.4 shows the fused results using different methods for multi-focus “Disk” images. As it can be observed in Fig. 5.4, the PCA method produces blur image with low contrast condition. The result of spatial frequency (SF) method has blocking effects, which can be seen around the clock. The previous wavelet-based methods including MS, WA, and WBV have

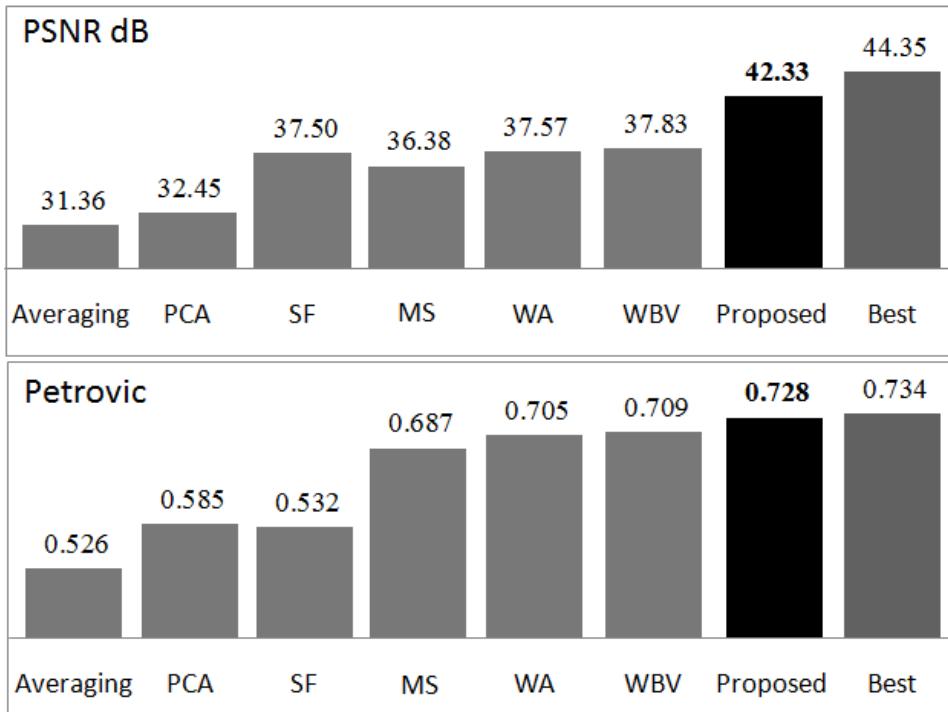


Figure 5.3: PSNR and Petrovic indexes for different image fusion methods.

caused blurring in the fused images, which can be seen on the left of the clock and top of the book. The result of our proposed method avoids the mentioned problems (blurring, blocking effect), also produces clearer fused image with high contrast condition. Fig. 5.5 shows another subjective result for multi-focus “Lab” images. As it can be seen in Fig. 5.5, specifically in the bottom row (magnified head regions of the fused results), the PCA method produces blur image with low contrast condition. The result of SF method has blocking effects (around the head). The previous wavelet-based methods including MS, WA, and WBV have caused artifacts in the fused images (around the head). These artifacts are because of using inaccurate decision maps (produced by these methods) for high frequency fusion rule when multi-focus images have miss registration problem. The result of our proposed method avoids the mentioned problems (blurring, blocking effect, artifacts), also produces clearer fused image with high contrast condition.

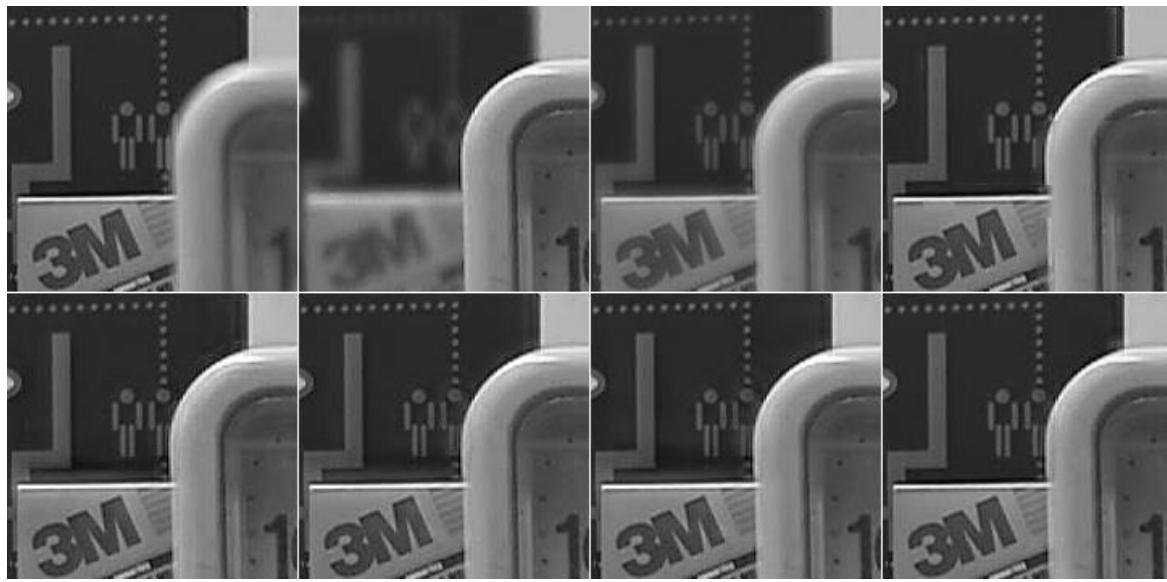


Figure 5.4: Subjective fusion results. Top row, left to right: a part of multi-focus “Disk” images, the fused result using PCA, SF methods. Bottom row, left to right: the fused result using MS, WA, WBV, and the proposed methods.

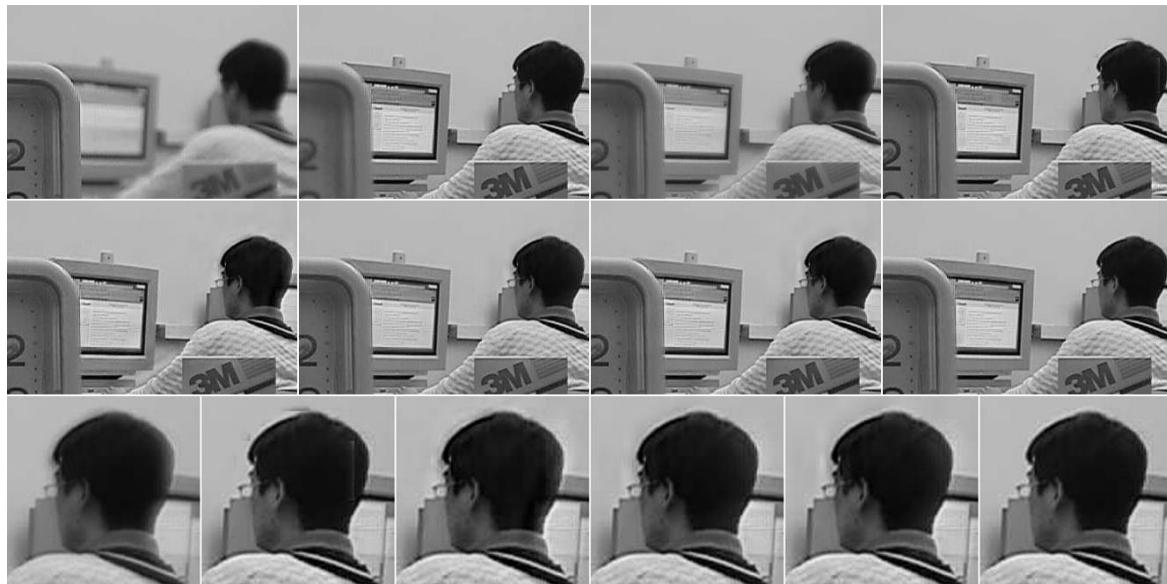


Figure 5.5: Subjective fusion results. Top row, left to right: a part of multi-focus “Lab” images, the fused results using PCA, SF methods. Middle row, left to right: the fused results using MS, WA, WBV, and the proposed methods. Bottom row, left to right: magnified head regions of the fused results produced with PCA, SF, MS, WA, WBV, and the proposed method.

### 5.1.2 Fusion results of artificial images

To confirm the effectiveness of the proposed algorithm, ten standard images shown in Fig. 5.6 are used as ground truth. From each image, Gaussian blurring creates two out-of-focus images. Then, the blurred images with different focus points are taken as the source images, and the original image is taken as the reference image for image fusion process.

Fig. 5.7 shows objective results for different multi-focus image fusion algorithms. It can be seen in Fig. 5.7 that our proposed image fusion algorithm has better PSNR index compared to other methods (more than +3.17 dB), and the results obtained for the proposed method are very close to the best results (less than 2.05 dB). In addition, evaluation of different methods using Petrovic metric has shown that our proposed algorithm has better performance for transferring information of edges from source images to the corresponding fused image.

## 5.2 The results of IR and visible image fusion

The proposed image fusion method was tested against several state-of-the-art image fusion methods including the simple averaging, the Laplacian Transform<sup>1</sup> (LT) [34], the Dual-Tree Discrete Wavelet Transform with three different fusion rules including Maximum Selection (MS), Weighted average (WA) [45], Window based verification (WBV) [47], and Lewis's region-based algorithm [49]. In the LT method five decomposition levels is used for image decomposition and maximum absolute selection rule is used as the high frequency fusion rule. For the DT-DWT based methods, the available "AntonB" mother wavelet is used for the filter banks and five decomposition levels is used for image decomposition. In addition, the simple averaging is used for the low frequency fusion rule in the LT and DT-DWT based methods.

---

<sup>1</sup> Image Fusion Toolbox for MATLAB developed by Oliver Rockinger: <http://www.metapix.de/toolbox.htm>



Figure 5.6: Ten ground truth images.

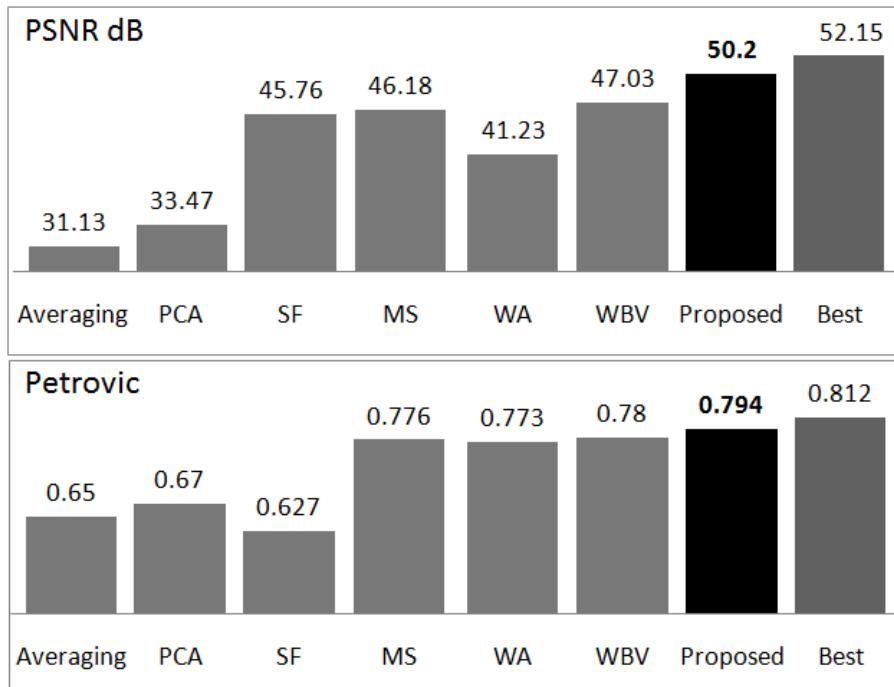


Figure 5.7: PSNR and Petrovic indexes for different image fusion methods.

The images used in experiments are surveillance images from TNO Human Factors, publicly available at the Image Fusion web site [16]. Image sequence “UN Camp” consists of 32 pair images and the image sequences “Trees” and “Dune” contain 19 pair images. In addition, we have applied the contrast stretching to enhance the representation of the source image. Other image processing techniques like histogram equalization can also be used.

Two metrics are considered in this paper, which do not require ground truth images for evaluation. The first metric is the Xydeas and Petrovic index, which is proposed in [107], which considers the amount of edge information transferred from the input images to the fused image using a Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images. The second metric is the Entropy index, which measures the information content in an image. An image with high information content will have high entropy. The entropy index is defined using (4.47).

It should be mentioned that the original size of the source images is  $270 \times 360$ . Because of using the dual-tree discrete wavelet transform in the proposed algorithm, the images size should be a factor of two. Therefore, we have generated  $512 \times 512$  images by repeating border elements of original images symmetrically. The possible decomposition levels for a  $512 \times 512$  image is  $N = \log_2 512 = 8$ . Different decomposition levels result in different performance of the proposed algorithm. In order to obtain best decomposition level, we have studied the effect of different decomposition levels on the performance of fused images. Table 5.1 shows the objective results of final fused images using different decomposition levels in the proposed algorithm. It can be seen in Table 5.1 that the effect of decomposition levels on the proposed algorithm is more noticeable for the Petrovic index compared to entropy index. The Petrovic index considers the amount of edge information transferred from the input images to the fused image [107]. The experimental results demonstrate that for  $N-1$  decomposition levels, the best results are obtained for different datasets.

Figs. 5.8, 5.9, and 5.10 show the results of different image fusion algorithms for the “UN camp”, “Dune” and “Trees” multi-sensor images, respectively. Subjective or visual comparison between different methods indicates superiority of the proposed method against other image fusion algorithms. In Fig. 5.8, it is clear that the fence details from the visible image and the person details from the IR image are far better transferred into the fused image in the proposed method than in the other algorithms. The fused image obtained by the averaging method is dark and

image contrast is low. The different contrast conditions of the IR and visible images causes this problem. While the fused images obtained by the previous pixel-based methods in the multi-resolution domain i.e. LT, MS, WA, and WBV algorithms have low contrast conditions, the fused images also have distortions (darkness) around the person and the roof. In addition, these methods are sensitive to the noise and image misregistration.

In the fused image obtained by Lewis's region-based method, which is also implemented in the dual-tree discrete wavelet transform domain, some details such as the contours of trees and the bright points are not transferred into the fused image. In Fig. 5.9, the roads' details from the visible image and the person's details from the IR image, and also in Fig. 5.10 the trees' details from the visible image and the person's details from the IR image, are better transferred into the fused image in the proposed method. Generally, it can be seen in Figs 5.8, 5.9, and 5.10 that the person is brighter in the proposed method, and the contrast of the fused images is far better compared to other methods.

Table 5.2 shows the average performance results from different image fusion methods and different datasets. It should be mentioned that the obtained results are different from the results listed in previous publications, because here we have applied the contrast stretching to enhance the representation of the source images.

The metric's values confirm the subjective assessment, that our proposed image fusion algorithm generally integrates more details from the visible and IR images into the fused images. It can be seen in Table 5.2 that the entropy index for the proposed method is better compared to other fusion methods, which indicates that the fused image obtained by the proposed method has more visual information and better contrast compared to the fused images using other fusion algorithms. In addition, evaluation of different methods using Petrovic metric has shown that the proposed algorithm has better performance for transferring edges information from the source images to the corresponding fused image.

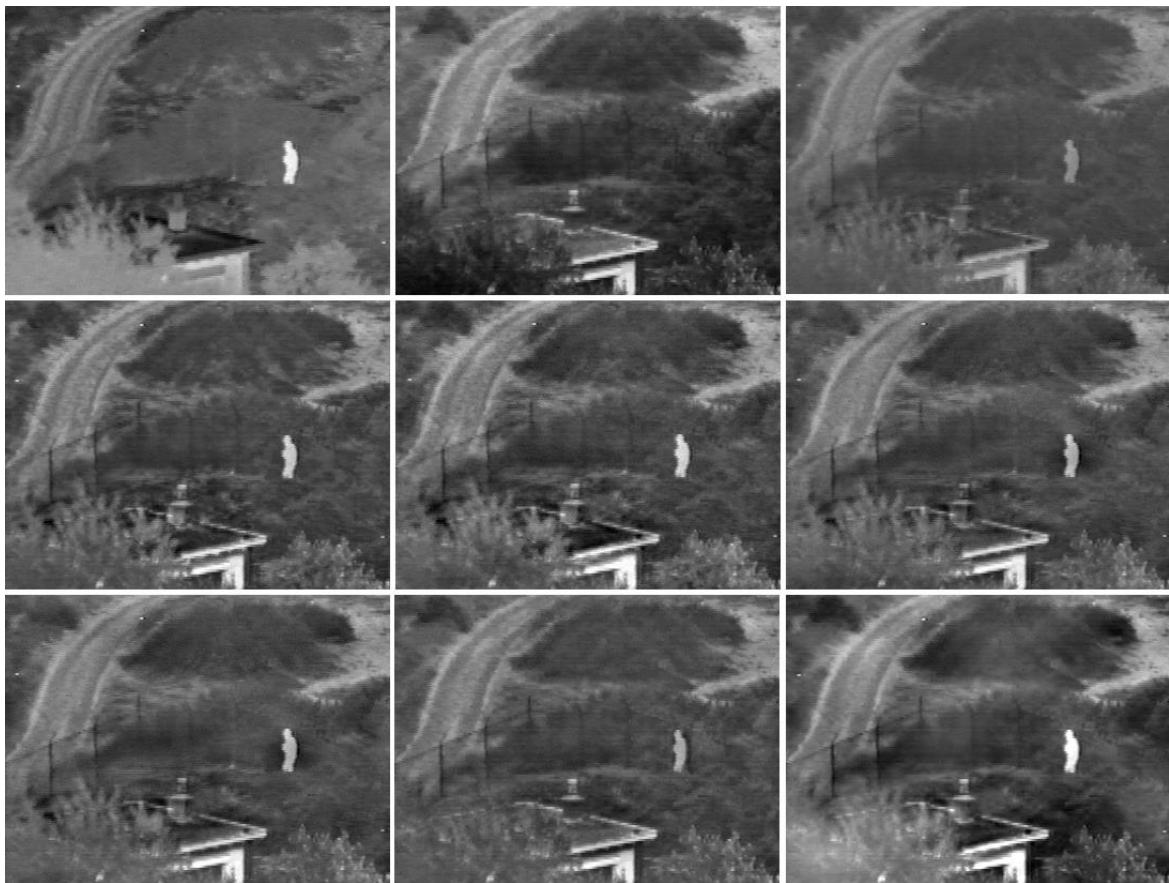


Figure 5.8: Subjective fusion results of “UN Camp” images: Original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [34], MS, WA [45] methods (middle-row), and fused images using WBV [47], Lewis [49], the proposed methods (bottom-row).

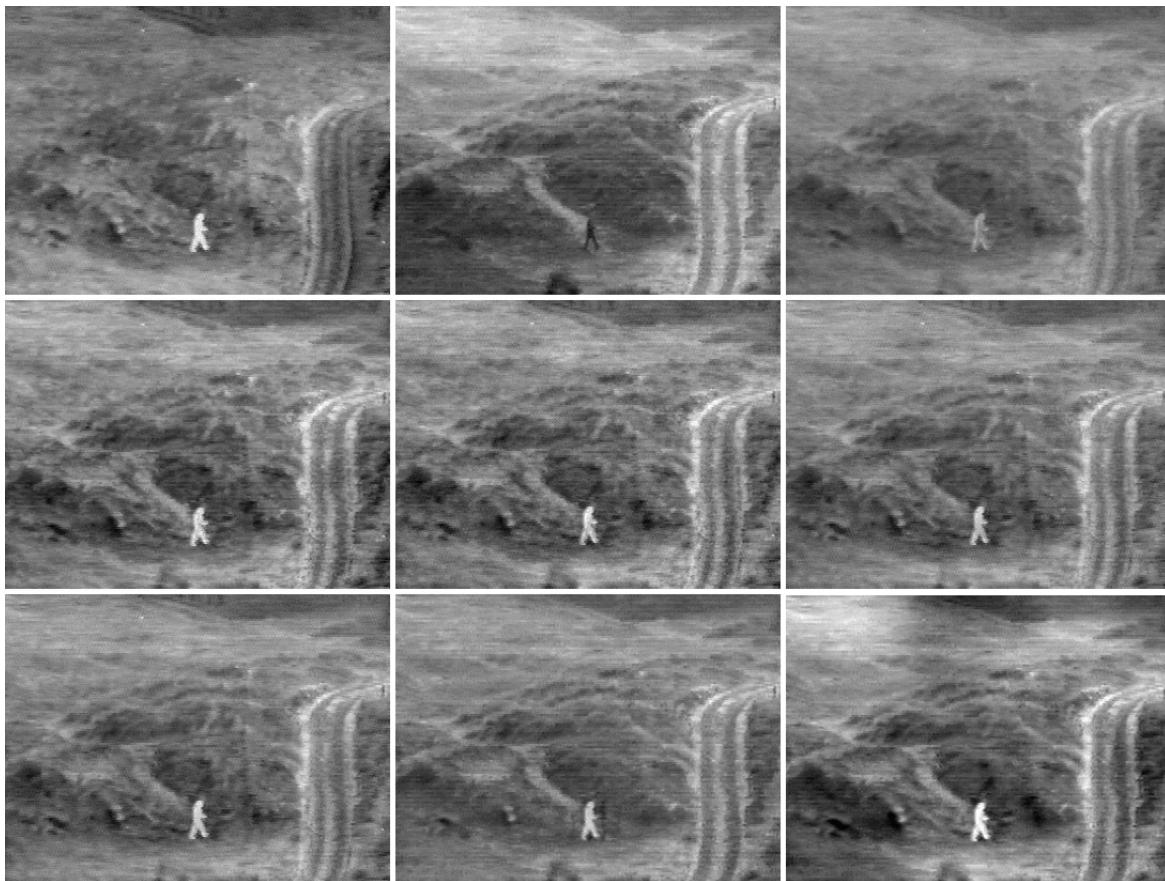


Figure 5.9: Subjective fusion results of “Dune” images: Original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [34], MS, WA [45] methods (middle-row), and fused images using WBV [47], Lewis [49], the proposed methods (bottom-row).

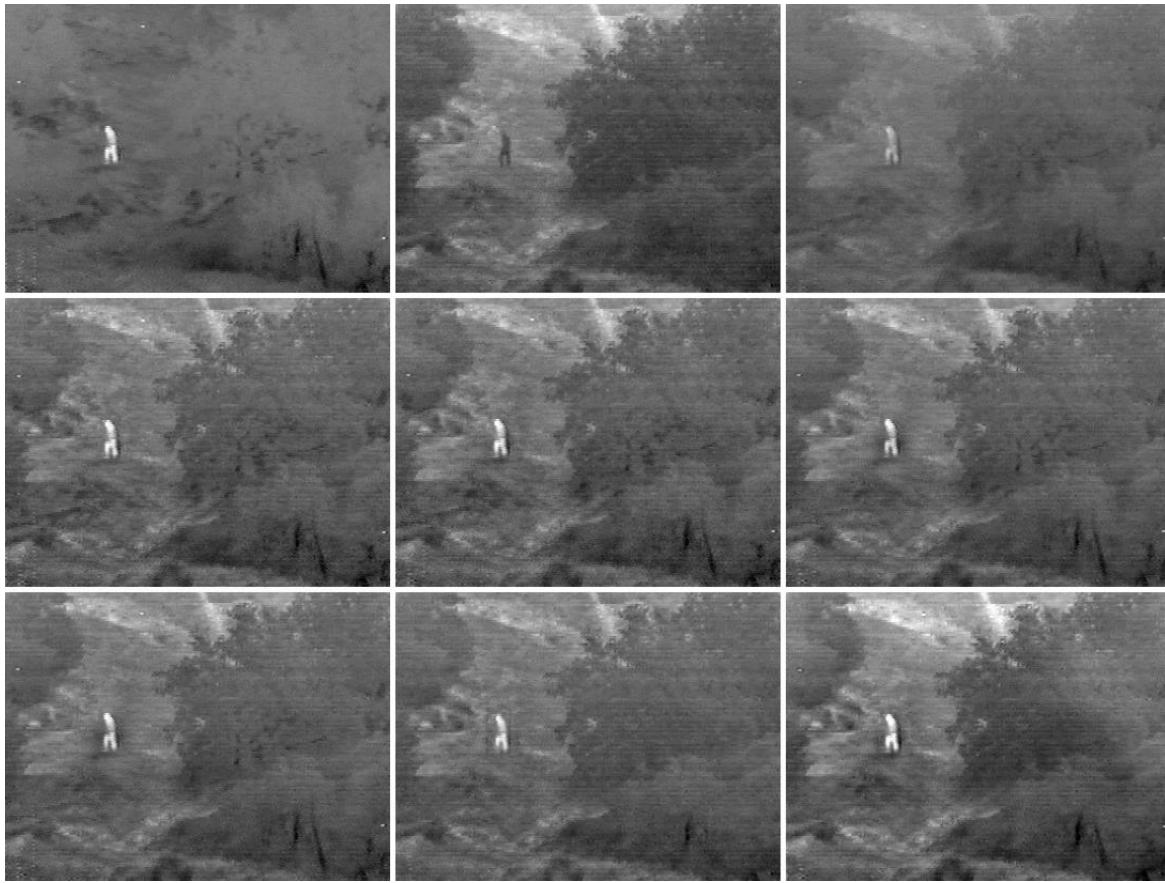


Figure 5.10: Subjective fusion results of “Trees” images: Original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [34], MS, WA [45] methods (middle-row), and fused images using WBV [47], Lewis [49], the proposed methods (bottom-row).

Table 5.1: The objective results of final fused images using different decomposition levels in the proposed algorithm.

Metric	N	1	2	3	4	5	6	7	8
	Dataset								
Petrovic	UN Camp	0.162	0.200	0.297	0.421	0.495	0.532	0.552	0.481
	Dune	0.224	0.267	0.364	0.452	0.504	0.534	0.537	0.443
	Trees	0.248	0.316	0.436	0.539	0.592	0.614	0.630	0.520
Entropy	UN Camp	6.869	6.940	6.980	7.031	7.164	7.246	7.305	7.223
	Dune	7.065	7.122	7.169	7.253	7.363	7.433	7.453	7.671
	Trees	6.577	6.667	6.706	6.786	6.905	7.011	7.051	7.145

Table 5.2: Average performance results of different image fusion methods.

Metric	Method	UN Camp	Dune	Trees
<b>Petrovic</b>	Averaging	0.332	0.338	0.351
	LT [34]	0.471	0.451	0.541
	MS	0.439	0.409	0.472
	WA [45]	0.421	0.396	0.494
	WBV [47]	0.424	0.390	0.492
	Lewis [49]	0.503	0.485	0.553
	Proposed	<b>0.552</b>	<b>0.530</b>	<b>0.630</b>
<b>Entropy</b>	Averaging	6.29	6.75	6.47
	LT [34]	6.71	7.03	6.61
	MS	6.79	7.01	6.63
	WA [45]	6.72	6.92	6.58
	WBV [47]	6.76	6.97	6.57
	Lewis [49]	6.66	6.88	6.54
	Proposed	<b>7.41</b>	<b>7.56</b>	<b>7.11</b>

### 5.3 The results of remote-sensing image fusion

There are many different ways to analyze the results of pan-sharpened images and compare different methods. When comparing different methods, we are interested in spatial, radiometric, and spectral qualities. In judging spatial quality, it is much easier to check the sharpness of the edges. However, when judging spectral and radiometric qualities, it is much more difficult to match the colors of the result to the original multispectral ones by visual inspection. Many metrics analyze the spectral quality. Here, we have used four metrics to judge spectral and radiometric qualities: The relative dimensionless global error in synthesis (ERGAS) proposed by Wald [122], the Relative Average Spectral Error (RASE) [63], Spectral Angle Mapper (SAM) [128], and Q<sub>4</sub> index [129]. The Q<sub>4</sub> quality index is a generalization

of the Universal Image Quality Index ( $Q_{avg}$ ) [130]. The RASE, ERGAS, and  $Q_4$  indexes measure both the radiometric and spectral qualities, and the SAM metric only measures the spectral quality.

To judge the spatial quality of the pan-sharpened image, we have used the correlation coefficient (CC) between the Pan and pan-sharpened images [123]. Here, we compare the spectral and radiometric qualities (ERGAS, RASE, SAM, and  $Q_4$  indexes) of different pan-sharpened results with respect to their spatial quality (CC to the Pan index). In other words, an algorithm, which generates better tradeoff between the spatial, radiometric and spectral qualities, is chosen as the best method, and higher spatial or radiometric or spectral quality is individually ineffective.

It should be mentioned that to assess the objective or numerical results (ERGAS, RASE, SAM, and  $Q_4$  indexes) of the pan-sharpened images, spatially degraded Pan and MS images from the original datasets are used, and the obtained results are compared to the original data. In addition, three subsets of the original (not degraded) datasets were extracted to compare the visual effect of different pan-sharpening algorithms (It is more usual to use real data for visual comparison). Therefore, there is not a reference high-resolution MS image in the visual or subjective comparisons. The results will be discussed in the following three subsections. First, the experimental datasets are presented that are used for evaluation of different algorithms. Then, we demonstrate the results of standard pan-sharpening algorithms. Finally, the results of multi-resolution-based methods will be illustrated.

### 5.3.1 Experimental datasets

The presented method is evaluated by performing pan-sharpening on datasets acquired by Landsat-7 ETM+, Quickbird, and Wordview-2 satellites. All images are radiometrically calibrated and orthorectified to uniform ground resolution. The MS and Pan images are co-registered for each dataset. All imageries have different

characteristics in terms of spectral bands and range, and spatial resolutions, etc. These characteristics are summarized in Table 5.3. Table 5.4 represents a summary of the datasets as related to their size and location. The true-color composites sample imageries used for the experiments are shown in Fig. 5.11.

- 1) The Landsat-7 ETM+ dataset is shown in Fig. 5.11 (a), which covers different urban, forested, Mediterranean, agricultural areas. Because only three bands are within the wavelength span of the Pan image, only the green, red, and near-infrared bands are considered for pan-sharpening.
- 2) The second dataset is very high resolution image data collected by Quickbird satellite. Two QuickBird imageries are shown in Fig. 5.11 (b) and (c). The first dataset shown in Fig. 5.11 (b) is an imagery representing mountainous regions consisting of forest and having a lot of uniform texture details, and urban features such as manmade buildings. The second dataset shown in Fig. 5.11 (c) covers different Mediterranean and agricultural areas. The four MS bands of the QuickBird span the visible and NIR wavelengths are within the wavelength span of the Pan image, and we have considered all of them for pan-sharpening.
- 3) The third very high resolution dataset has been considered, which has been acquired by the Wordview2 satellite is shown in Fig. 5.11 (d). This imagery representing urban features such as manmade buildings. We have considered the four MS bands consisting of the Blue, Green, Yellow, and Red bands for pan-sharpening. It is possible to pan-sharpen the Red-Edge band, because it is also within the wavelength span of the Pan image.

In addition, for simplicity to work with the images and pan-sharpening algorithms we have generated 20 MS ( $512 \times 512$ ) and corresponding Pan ( $1024 \times 1024$ ) images from the Landsat-7 ETM+ dataset, 10 MS ( $256 \times 256$ ) and corresponding Pan ( $1024 \times 1024$ ) images from the Quickbird datasets, and 10 MS ( $256 \times 256$ ) and corresponding Pan ( $1024 \times 1024$ ) images from the Wordview2 dataset. The Pan and MS images are co-registered for each dataset and the MS

images were spatially resampled by a factor of two for the Landsat-7 ETM+ dataset, and by a factor of four for Quickbird and Wordview2 datasets, to produce new 14.25 m, 0.6 m, and 0.46 m resolution MS images, respectively.

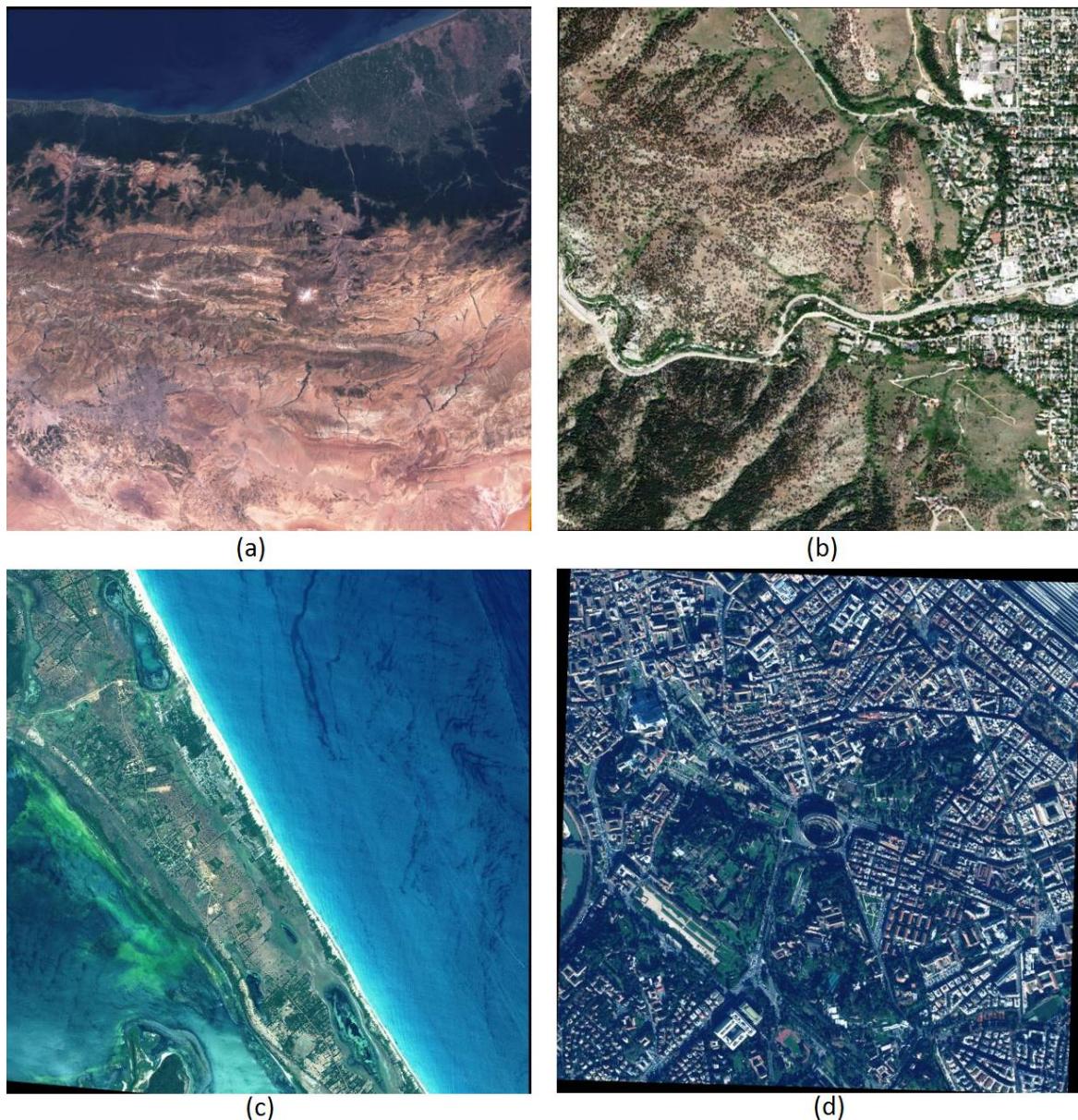


Figure 5.11: True-color composites of multi-spectral data used in the experiment. (a) Lansat-7 ETM+ scene from Iran, (b) Quickbird scene from Boulder, CO, USA, (c) Quickbird scene from Kokilai-Lagoon, Srilanka, and (d), Wordview2 scene from Rome, Italy.

Table 5.3: Satellite datasets characteristics.

Features	Imagery Type		
	Landsat-7 ETM+	Quickbird	Wordview-2
MS Image-Spectral Bands and Range	Blue: 450-515 nm Green: 525-605 nm Red: 630-690 nm Near-IR: 760-900 nm MID-IR 1: 1550-1750 nm MID-IR 2: 2080-2350 nm	Blue: 450-520 nm Green: 520-600 nm Red: 630-690 nm Near-IR: 760-900 nm	Coastal: 400-450 nm Blue: 450-510 nm Green: 510-580 nm Yellow: 585-625 nm Red: 630-690 nm Red Edge: 705-745 nm Near-IR1: 770-895 nm Near-IR2: 860-1040 nm
Pan image-Spectral Range	520-920 nm	450-900 nm	450-800 nm
Spatial Resolution-MS Image	28.5 m	2.4 m	1.84
Spatial Resolution-Pan Image	14.25 m	0.60 m	0.46 m
Resolution Ratio of MS and Pan Images	2	4	4

Table 5.4: Datasets' Summary

Features	Imagery Type			
	Landsat-7 ETM+	Quickbird		Wordview-2
Location	Tehran, Iran	Boulder, CO, USA	Kokilai-Lagoon Srilanka	Rome, Italy
MS Image Size	7348×6208	828×815	3198×2667	1151×1150
Pan Image Size	14696×12416	3312×3260	12792×10668	4604×4600

### 5.3.2 Comparisons between different state-of-the-arts Pan-sharpening algorithms

In this subsection, we compare the proposed method, which is based on shiftable contourlet transform and multi-objective particle swarm optimization with some of

the best state-of-the-art techniques consisting of the IHS, Brovey [61], PCA<sup>1</sup>, P+XS<sup>1</sup> [65], Ehlers [66], and Bayesian Data Fusion (BDF)<sup>2</sup> [131]. The parameters of each method have been set according to the values given by their respective authors in the corresponding referred papers.

Table 5.5 summarizes the obtained results. As it can be seen in Table 5.5, the proposed method provides better spectral and radiometric qualities (ERGAS, RASE, SAM, and Q<sub>4</sub> indexes) in all cases and better spatial quality (CC to the Pan index) in almost all cases. When looking closer at the results, we observe the following:

- The proposed method outperforms the IHS technique in spatial, radiometric and spectral qualities. The CC index of the IHS method is acceptable, while the corresponding spectral and radiometric qualities are very weak.
- Our method gives better results than the Brovey algorithm in the radiometric quality. However, the Brovey method gives better result in CC to the Pan index, but the corresponding ERGAS, RASE, and Q<sub>4</sub> values for it are very weak, which indicates that the pan-sharpened image has low radiometric quality. It should be mentioned that the angle from the Spectral Angle Mapper (SAM) between the original and pan-sharpened MS images is zero. This is because the operation of multiplication only changes the norm of a spectral vector [132]. Therefore, it does not affect the SAM index between the original and pan-sharpened MS images. A value of SAM equal to zero denotes absence of spectral distortion, but radiometric distortion is possible and each band of MS image is greatly changed due to different values being involved in the multiplication from pixel to pixel (the two pixel vectors are parallel but have different lengths).

---

<sup>1</sup> Available at: [www.math.ucla.edu/~wittman/pansharpening/index.html](http://www.math.ucla.edu/~wittman/pansharpening/index.html).

<sup>2</sup> ORFEO Accompaniment Program, ORFEO Tool Box, 2007. Available: [http://smsc.cnes.fr/PLEIADES/A\\_prog\\_accomp.htm](http://smsc.cnes.fr/PLEIADES/A_prog_accomp.htm)

- Our method provides better results than the PCA algorithm. While the proposed method has a little difference in the CC index, results in very high gap in the spectral and radiometric quality indexes for different datasets.
- The proposed method outperforms the P+XS and Ehlers techniques in spatial, radiometric and spectral qualities for different datasets.
- Our method gives better results than the BDF algorithm in the spectral and radiometric quality. The BDF method provides very good spatial quality, while the corresponding spectral and radiometric quality indexes are weak.

For subjective evaluation of methods, there are two important criteria: spectral and spatial qualities. In other words, the final pan-sharpened image at the same time as provides high spatial quality (or spatial details from the panchromatic image), should have high spectral and radiometric qualities (or color information from the original multi-spectral image).

Table 5.5: Average Performance of different state-of-the-arts Pan-sharpening algorithms against the proposed method for different datasets.

Index	Dataset	Pan-sharpening algorithms							RV
		IHS	Brovey	PCA	P+XS	Ehlers	BDF	MOPSO	
CC to the Pan	Landsat-7	0.937	<b>0.967</b>	0.961	0.880	0.931	0.963	0.947	1
	Quickbird	0.915	0.971	<b>0.974</b>	0.792	0.884	0.971	0.951	
	Wordview2	0.923	<b>0.967</b>	0.916	0.743	0.898	0.963	0.949	
ERGAS	Landsat-7	5.55	6.16	6.38	2.51	2.96	3.23	<b>1.78</b>	0
	Quickbird	10.98	9.86	9.71	2.88	3.85	4.84	<b>2.34</b>	
	Wordview2	6.94	9.96	9.87	3.43	3.89	5.48	<b>2.85</b>	
RASE	Landsat-7	22.10	27.20	25.56	10.30	11.87	13.23	<b>7.28</b>	0
	Quickbird	43.00	40.62	36.92	11.23	15.02	19.15	<b>9.11</b>	
	Wordview2	23.17	32.18	29.54	11.14	12.92	19.38	<b>9.72</b>	
SAM	Landsat-7	1.69	<b>0.0</b>	2.35	1.70	1.41	1.99	1.01	0
	Quickbird	3.55	<b>0.0</b>	4.63	2.36	2.21	3.15	2.14	
	Wordview2	2.65	<b>0.0</b>	6.34	2.52	2.37	3.11	1.69	
Q <sub>4</sub>	Landsat-7	0.928	0.891	0.848	0.904	0.927	0.881	<b>0.925</b>	1
	Quickbird	0.893	0.843	0.819	0.921	0.939	0.895	<b>0.943</b>	
	Wordview2	0.941	0.899	0.893	0.949	0.964	0.928	<b>0.971</b>	

Figs. 5.12 and 5.13 give subjective results for two samples of Pan and MS images from the Wordview2 and Quickbird datasets, respectively. The two samples are selected from the real (not degraded) datasets (Table 5.4), and therefore there are not reference images for comparison (The resolution ratio between the MS and Pan images is 4). It can be seen in Figs. 5.12 and 5.13 that the results of the IHS and Brovey methods are sharp enough, but the colors are changed compared to the low-resolution MS images. The result of the PCA method is not clear and the colors are changed. For the results of the P+XS method, the colors are preserved, but the spatial details from the Pan image is not appropriately transferred to the pan-sharpened image. The Ehlers method suffers from miss-registration problem. This is because that in the Ehlers method, a low frequency version of intensity image from the MS bands is added to the high frequency version of Pan image, and if the Pan and MS images are not accurately registered, causes an unclear image with thick edges. In addition, the Ehlers method uses Fourier Transform for spatial details extraction from the Pan image, while Fourier Transform is not an efficient representation for 2-D signals. The BDF method results in very good spatial quality in the pan-sharpened image and details information from the Pan image is appropriately preserved, while the spectral and radiometric information from the low-resolution MS image has changed to some extent. The result of the proposed algorithm visually demonstrates very good spatial quality and the colors are well preserved from the low-resolution MS image. Indeed, the visual inspection of true-color composites confirms the numerical results of Table 5.5, in which the proposed method gives better trade-off between spatial and spectral qualities.

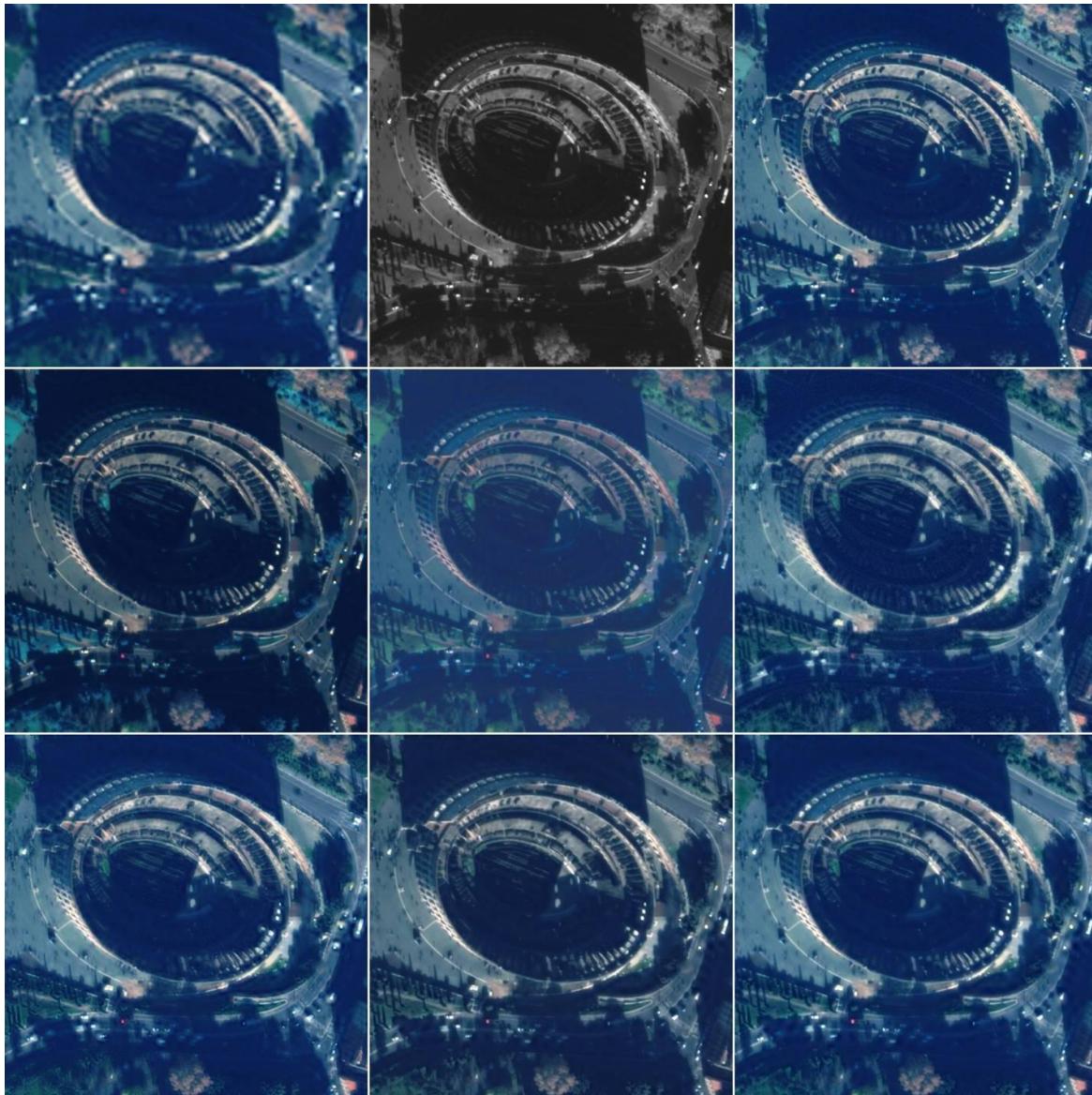


Figure 5.12: Subjective results of different pan-sharpening methods. True-color composites of Low Resolution Multi-spectral image form the Wordview2 dataset, corresponding panchromatic image, and the result of IHS method (top-row, left to right). The results of Brovey, PCA, and P+XS methods (middle-row). The results of Ehlers, BDF, and the proposed methods, (bottom-row).

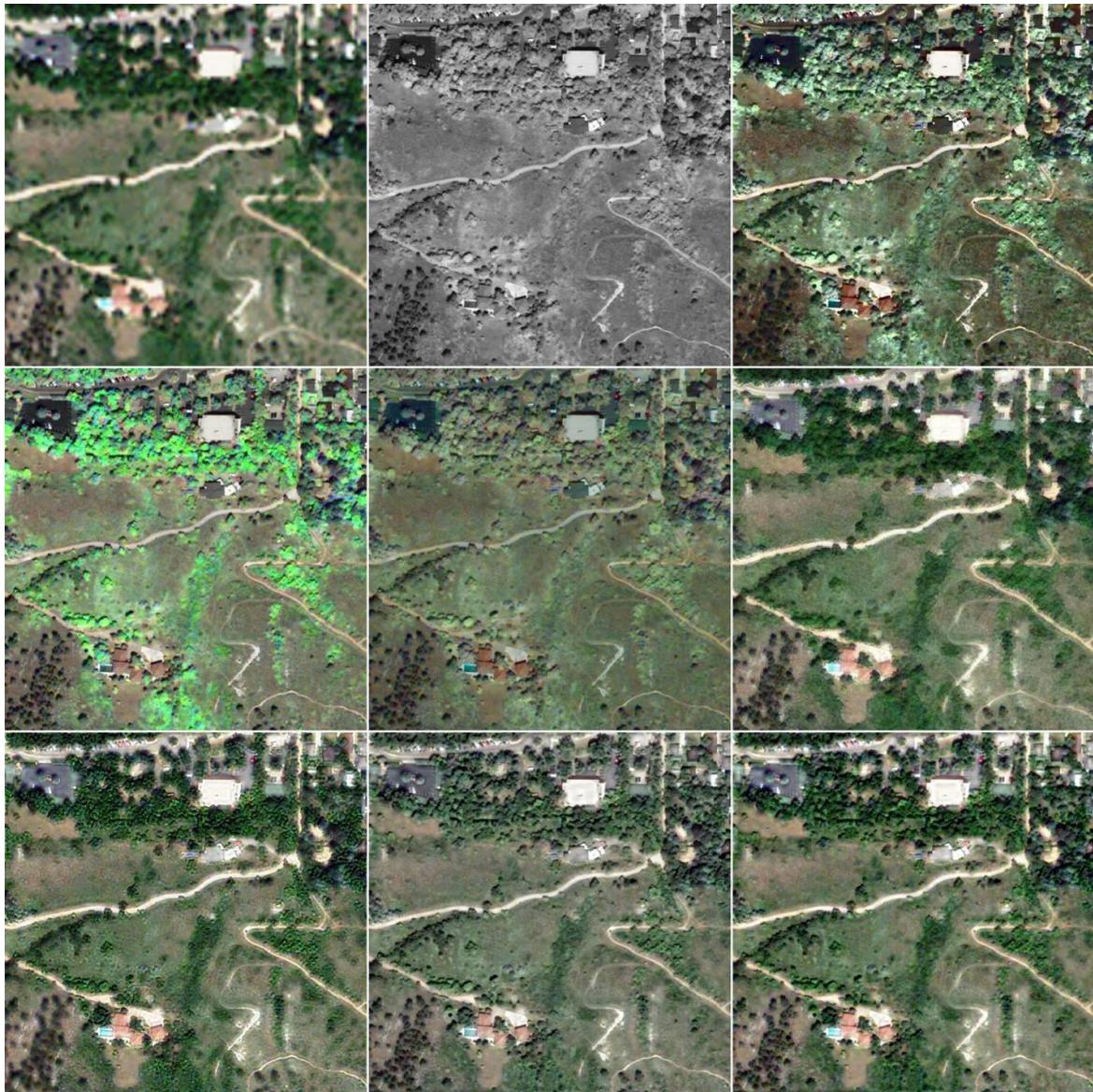


Figure 5.13: Subjective results of different pan-sharpening methods. True-color composites of Low Resolution Multi-spectral image form the Quickbird dataset, corresponding panchromatic image, and the result of IHS method (top-row, left to right). The results of Brovey, PCA, and P+XS methods (middle-row). The results of Ehlers, BDF, and the proposed methods, (bottom-row).

### 5.3.3 Comparisons between different multi-resolution-based Pan-sharpening algorithms

In this experiment, we have compared the proposed algorithm, which is based on SCT and MOPSO algorithm with the multi-resolution-based pan-sharpening methods. In the multi-resolution-based methods, first the MS and Pan images are

decomposed. Then, the high-frequency coefficients of the final pan-sharpened image are obtained using averaging (A) and substitutive (S) fusion rules. The final pan-sharpened image is obtained using inverse multi-resolution transform. Here we have used DT-DWT<sup>1</sup> and CT<sup>2</sup> transforms for the multi-resolution analysis: DT-DWT+A, DT-DWT+S [64], CT+A, and CT+S. In addition, we have compared the proposed method with two hybrid multi-resolution algorithms. In the first hybrid method, IHS transform is used to separate spectral information of the low-resolution MS image, and then multi-resolution-based algorithm is used to inject detailed information from the histogram-matched Pan image to the intensity image. Finally the inverse IHS transform generates pan-sharpened image: IHS+DT-DWT+S [133], and IHS+CT+S [88]. In the second hybrid method, contourlet and principal component analysis (PCA) methods have been used for spatial and spectral transformations, respectively. Then, the first principal component of high variance is used for replacing or injecting it with high spatial details from the high-resolution histogram-matched Pan image. Injecting of the spatial details is based on multi-resolution-based algorithm, in which high frequency contourlet coefficients of the Pan image are replaced with the high frequency contourlet coefficients in the first principal component. Finally, the inverse contourlet and PCA transformations generates final pan-sharpened image [59].

Table 5.6 summarizes the obtained results for different datasets. As it can be observed in Table 5.6, the proposed method provides appropriate spectral and radiometric qualities (ERGAS, RASE, SAM, and Q<sub>4</sub> indexes) with respect to the spatial quality (CC to the Pan index) compared to other methods. When looking closer at the results, we observe the following:

- The CC to the Pan index for the methods, which use substitutive fusion rule (S) are better than other algorithms, while the spectral and radiometric quality indexes for them are worse. This indicates that the pan-sharpened images using

---

<sup>1</sup> Available at: [taco.poly.edu/WaveletSoftware/dt2D.html](http://taco.poly.edu/WaveletSoftware/dt2D.html).

<sup>2</sup> Available at: [www.mathworks.com/matlabcentral/fileexchange/8837-contourlet-toolbox](http://www.mathworks.com/matlabcentral/fileexchange/8837-contourlet-toolbox).

these methods are sharp enough, but the spectral information from the low-resolution MS image is not appropriately transferred to the pan-sharpened image.

- The spectral and radiometric quality indexes for methods, which use averaging fusion rule (A), are better than other algorithms, while the corresponding CC to the Pan index for them is very low. This indicates that the colors from the low-resolution MS image are well preserved, but the image is not clear or sharp.
- We have discussed in previous Section that the shiftable Contourlet transform is a better transformation than the CT because of the shift-invariant property, and better than the DT-DWT because of the directional selectivity. The objective results demonstrate that the SCT is a better transformation compared to the DT-DWT and CT for both substitutive and averaging fusion rules.
- The results of using IHS transform as a spectral transformation in the multi-resolution-based scheme shown better spectral and radiometric qualities compared to the multi-resolution-based methods, which use substitutive fusion rule (S).
- The PCA+CT method, which is used spatial and spectral transformations results in better trade-off between the spatial and spectral qualities compared to the multi-resolution-based methods, which use averaging fusion rule (A). However, the spatial quality for different datasets is very low, which is the consequence of using spectral transformation (PCA) in the pan-sharpening scheme.
- The proposed pan-sharpening method is shown a better trade-off between the spatial (CC to the Pan index), radiometric and spectral qualities (ERGAS, RASE, SAM, and Q<sub>4</sub> indexes) compared to other techniques. In other words, the results show that the proposed method could transfer spatial details of the Pan image into the final pan-sharpened image, without reducing spectral and radiometric qualities of the MS image.

In addition, Fig. 5.14 gives subjective results for a sample of Pan and MS images from the Wordview2 datasets. The sample are selected from the real (not degraded)

datasets (Table 5.4), and therefore there is not a reference image for comparison (The resolution ratio between the MS and Pan image is 4). It can be seen in Fig. 5.14, the colors in the results of the CT+A, and DT-DWT+A methods are preserved, while the images are not sharp. The results of the CT+S, and DT-DWT+S methods are sharp enough, while the colors specially the blue, orange, and red areas from the low-resolution MS image are changed. The result of the IHS+DT-DWT+S method locally introduces small artifacts due to the adding inaccurate spatial details of low-resolution MS image to the Pan image. The IHS+CT+S method is also results in similar pan-sharpened image. The result of PCA+CT method is not very accurate in spectral or color preservation, and the spatial information of the Pan image is not appropriately transferred into the pan-sharpened image. The result of proposed algorithm visually demonstrates very

Table 5.6: Average Performance of different multi-resolution-based Pan-sharpening algorithms against the proposed method for different datasets.

Index	Dataset	Pan-sharpening algorithms										R V
		DTDWT +A	DTDWT +S	CT+A	CT +S	IHS+ DTDWT+S	IHS+ CT+S	PCA+ CT	SCT +S	SCT +A	MO -PSO	
CC to the Pan	L-7	0.906	0.971	0.912	0.976	0.961	0.964	0.921	0.983	0.915	0.947	1
	Q	0.907	0.964	0.905	0.963	0.960	0.957	0.915	0.967	0.909	0.951	
	W-2	0.889	0.967	0.884	0.964	0.963	0.961	0.912	0.972	0.891	0.949	
ERGAS	L-7	1.34	3.68	1.21	3.37	3.27	3.21	1.85	3.08	1.16	1.78	0
	Q	1.83	3.59	1.96	3.84	3.45	3.67	2.70	3.42	1.81	2.34	
	W-2	2.19	4.26	2.27	4.41	4.14	4.23	3.04	4.18	2.13	2.85	
RASE	L-7	5.40	13.76	5.21	13.21	12.32	12.03	7.43	12.11	4.97	7.28	0
	Q	7.12	13.95	7.60	14.89	13.51	14.37	10.27	13.67	7.01	9.11	
	W-2	7.23	14.21	7.51	14.63	13.62	14.09	10.13	14.17	7.18	9.72	
SAM	L-7	0.728	1.43	0.794	1.57	1.47	1.46	1.37	1.41	0.715	1.01	0
	Q	1.47	3.10	1.62	3.38	2.77	2.86	2.57	3.01	1.42	2.14	
	W-2	1.23	2.27	1.26	2.42	2.18	2.21	1.79	2.13	1.11	1.69	
Q <sub>4</sub>	L-7	0.928	0.838	0.933	0.826	0.924	0.921	0.928	0.859	0.941	0.925	1
	Q	0.951	0.881	0.942	0.864	0.931	0.927	0.939	0.887	0.954	0.943	
	W-2	0.969	0.956	0.971	0.942	0.952	0.944	0.965	0.961	0.983	0.971	

accurate spectral preservation and superior performances in terms of radiometric and geometric accuracy. Visual assessment of true-color composites confirms the numerical results of Table 5.6, in which the proposed method provides better trade-off between spatial and spectral qualities.

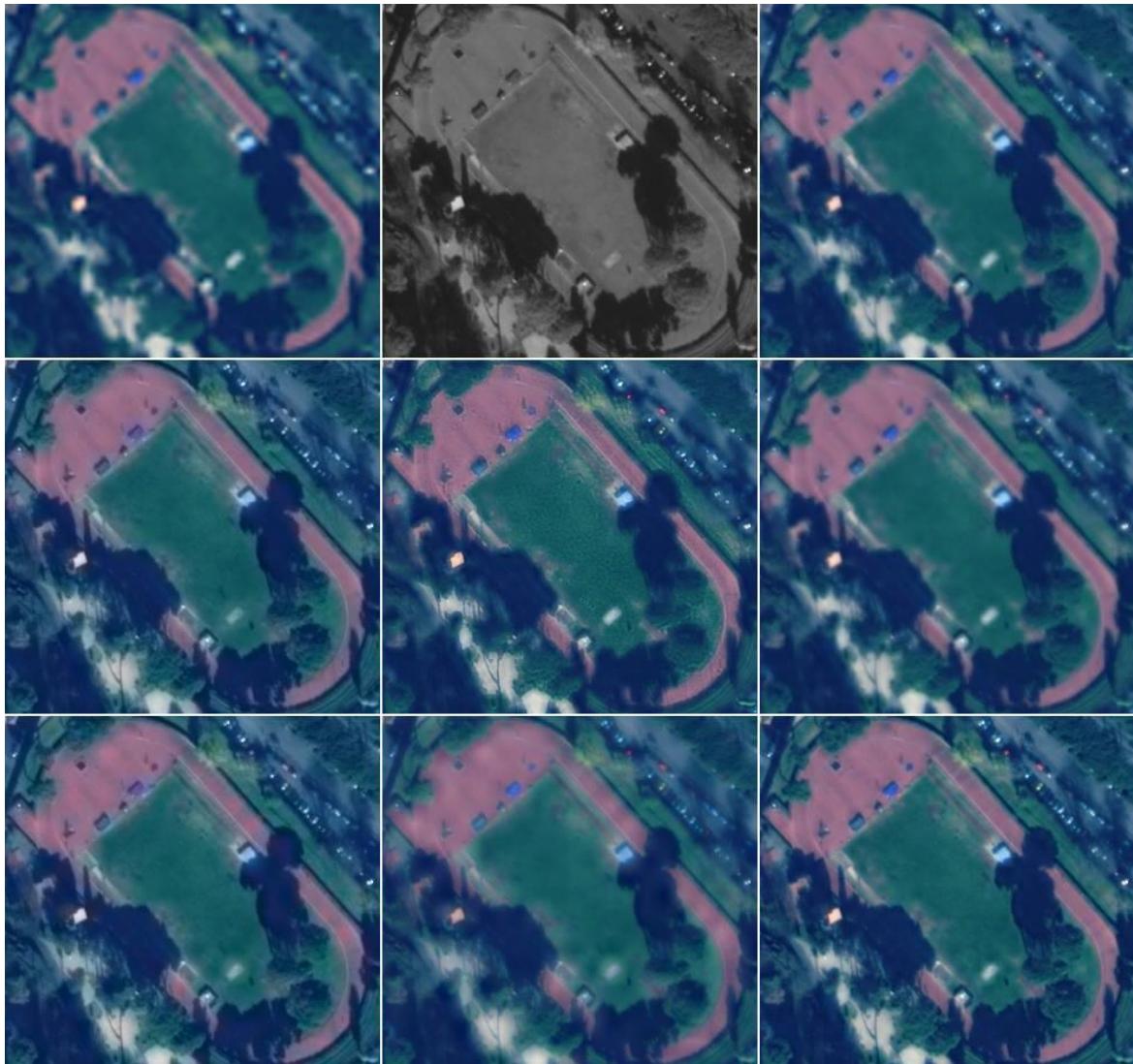


Figure 5.14: Subjective results of different multi-resolution-based pan-sharpening methods. True-color composites of Low Resolution Multi-spectral image from the Wordview2 dataset, corresponding panchromatic image, and the result of the DT-DWT+A method (top-row, left to right). The results of the DT-DWT+S, IHS+DT-DWT+S, and CT+A methods (middle-row). The results of CT+S, PCA+CT, and the proposed methods (bottom-row).

## Chapter 6. Conclusions and future works

The subject of this dissertation as stated in the introduction Chapter is pixel-level image fusion is. It represents fusion of visual information of the same scene, from any number of registered image signals, obtained using different sensors. The goal of pixel-level image fusion can broadly be defined as: to represent the visual information present in any number of input images, in a single fused image without the introduction of distortion or loss of information.

In this study we have been able to introduce novel image fusion methods in three different research areas. In first research subject, we have presented a new wavelet-based multi-focus image fusion method using Fisher classifier and Fuzzy logic. Proposing new fusion rules for merging high and low frequency wavelet coefficients, which is the second step in the wavelet-based image fusion, is the main novelty of this method. Generating a reliable decision map using the information of high frequency wavelet coefficients in the six directional sub-bands of the DT-DWT is the main idea. In addition, we have used fuzzy logic based on our observation to solve existent uncertainty in the smooth regions of source images and in the border of in- and out-of-focus regions. This new method used DT-DWT for finer frequency decomposition and shift invariant property as compared to discrete wavelet transform. The experimental results demonstrated that the proposed method outperforms the standard fusion methods in the fusion of multi-focus images.

In the second research topic, we have presented a new wavelet based multi-sensor image fusion method using fuzzy logic and particle swarm optimization.

Proposing new fusion rules for merging high and low frequency wavelet coefficients, which is the second step in the wavelet-based image fusion, is the main novelty of this method. We specifically have presented a fuzzy-based approach for fusing high frequency wavelet coefficients of the IR and visible images, which combines the outputs of three different fusion rules based on a dissimilarity measure of the source images. This new method applies the advantages of the previous pixel and region-based methods using fuzzy logic. Proposing an optimization method for fusing low frequency wavelet coefficients of the source image is another novelty of this method, which applies the population-based optimization in the low frequency fusion scheme to maximize the entropy of the final fused image. In addition, the proposed method used the DT-DWT for finer frequency decomposition and shift invariant property compared to discrete wavelet transform. The experimental results demonstrated that the proposed method outperforms the standard fusion methods in the fusion of IR and visible images.

Finally, a novel pan-sharpening algorithm based on the shiftable Contourlet transform and multi-objective particle swarm optimization is proposed in last research area in this dissertation. The shiftable Contourlet transform is known to provide shiftable transform compared to the Contourlet transform, and better directional representation compared to the wavelet transform. In addition, the multi-objective particle swarm optimization is used to generate high-resolution MS images with high spatial similarity to the Pan image and high radiometric quality in each band. Therefore, as expected the proposed method provides better pan-sharpened results based on the global metrics (CC to the Pan, ERGAS, RASE indexes), which measure the spatial and radiometric qualities. The experimental results also show that the produced pan-sharpened images have good spectral quality based on the Q<sub>4</sub> and SAM metrics.

The work of this dissertation can be applied and extended to a variety of research topics. A few examples are enumerated below:

- ❖ In the proposed algorithm for IR and visible image fusion, we have used an image segmentation algorithm, which is sensitive to the noise and for noisy image results in oversegmentation with incorrect regions. In addition, we have not considered the noise parameters for the high frequency fusion rule, and therefore the results will be sensitive to the noise. As a result, we have to use an image-denoising algorithm as a pre-processing step to work with the proposed algorithm in noisy situation. One aspect that we would like to explore in the future is to design a robust image fusion algorithm for fusing noisy multi-sensor images, in which the two processes of image fusion and denoising are considered all together.
- ❖ One aspect that we would like to explore in the future is to analyze the evaluation indices system to acquire a meaningful measurement, and using better combination of the evaluation indices in the multi-objective optimization procedure. We are also considering improving the optimization performances using other evolutionary algorithms such as Genetic algorithm further. Using segmentation map of the source images instead of rectangular window for local parameter estimation in the proposed method can improve the pan-sharpened results. Future work should be done on these promising issues.

## References

- [1] L. Klein, Sensor and Data Fusion Concepts and Applications, SPIE Press, 1999
- [2] J. Aggarwal, Multisensor Fusion for Computer Vision, Springer Verlag, 1993
- [3] D. Hall, Mathematical Techniques for multisensor Data Fusion, Artech House Inc. 1992
- [4] R. Brooks, S. Iyengar, Multi-sensor Fusion, Prentice-Hall, 1998
- [5] Z. Zhang, Investigations of Image Fusion, PhD Thesis, Lehigh University, USA, 1999
- [6] V. S Petrovic, Multisensor Pixel-level Image Fusion, PhD Thesis, Lehigh University, USA, 2001
- [7] G. Peilla, Adaptive Wavelets and their Applications to Image Fusion and Compression, PhD Thesis, Amsterdam University, Holland, 2003.
- [8] R. S. Blum, and Z. Liu, Multi-Sensor Image Fusion and Its Applications, Taylor & Francis Group, New York, 2006.
- [9] L. Brown, "A survey of image registration techniques," ACM Computing Surveys, 24(4), 1992, pp. 325-376.
- [10] H. Li, B. Munjanath, and S. Mitra, "A contour-based approach to multi-sensor image registration," IEEE Transactions Image Processing, 4(3), 1995, pp. 320-324.
- [11] A. Toet, L. van Ruyven, and J. Velaton, "Merging thermal and visual images by a contrast pyramid," Optical Engineering, 1989, 28(7), pp. 789-792.
- [12] D. Rajan, and S. Chaudhuri, "Generalized Interpolation and Its Application in Super- Resolution Imaging," Image and Vision Computing, 19 (13) , pp. 957-969, 1 November 2001
- [13] C. Chiang, and T. E. Boult, "Efficient Super-Resolution via Image Warping," Image and Vision Computing, 18 (10), pp. 761-771, July 2000.

- [14] Zh. Zhang, and R. S. Blum, "A Categorization of Multiscale-Decomposition-Based Image Fusion Schemes with a Performance Study for a Digital Camera Application," PROCEEDINGS OF THE IEEE, VOL. 87, NO. 8, AUGUST 1999, pp.1315-1326.
- [15] Y. Song, M. Li, Q. Li, and L. Sun, "A New Wavelet Based Multi-focus Image Fusion Scheme and Its Application on Optical Microscopy," Proceedings of the 2006 IEEE International Conference on Robotics and Biomimetics December 17 - 20, 2006, Kunming, China, pp. 401-405.
- [16] Available at: <http://imagefusion.org>
- [17] J. A. Richards, "Thematic mapping from multitemporal image data using the principal component transformation," Remote Sensing of Environment 16, 1984, pp. 36-46.
- [18] L. Bo, and L. Hai-lian, "Pixel level image fusion scheme based on accumulated gradient and PCA transform," Journal of Communication and Computer, ISSN1548-7709, USA, Feb. 2009, 6 (2), pp. 49-54.
- [19] S. d. Bethune, F. Muller, and M. Binard, "Adaptive intensity matching filters: a new tool for multi-resolution data fusion," In Proceedings of the Sensor and Propagation Panel's 7th Symposium on Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar, 1997, 28, pp. 1-15.
- [20] E. Lallier, "Real-time pixel-level image fusion through adaptive weight averaging," Technical report, Royal Military College of Canada, Kingston, Ontario, 1999.
- [21] S. Li, James, T. Kwok, and Yaonan Wang, "Combination of Images with Diverse Focuses Using the Spatial Frequency," Information Fusion, Volume 2, Issue 3, pp. 169-176, September 2001.
- [22] S. Li, James T. Kwok and Y. Wang, "Multifocus Image Fusion Using Artificial Neural Networks," Pattern Recognition Letters, 23 (80), pp. 985-997, June 2002

- [23] S. Li, and B. Yang, “Multifocus image fusion using region segmentation and spatial frequency,” *Image and Vision Computing*, 26 (2008) pp. 971–979.
- [24] R. Maruthi, and K. Sankarasubramanian, “Multi Focus Image Fusion Based on the Information Level in the Regions of the Images,” *Journal of Theoretical and Applied Information Technology*, 2007, pp.80-87.
- [25] R. K. Sharma, and M. Pavel, “Adaptive and statistical image fusion,” *Society for Information Display Digest of Technical Papers* 27, 1996, pp. 969-972.
- [26] B. Ma, S. Lakshmanan, and A. O. Hero, “Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion,” *IEEE Trans. Intell. Trans. Syst.*, 1(3), 2000, pp. 135-147.
- [27] R. S. Blum, and J. Yang, “Image fusion using the expectation-maximization algorithm and a Gaussian mixture model,” In *Advanced Video-Based Surveillance Systems*, Boston, 2003.
- [28] V. Maik, D. Cho, J. Shin, and J. Paik, “Regularized Restoration Using Image Fusion for Digital Auto-Focusing,” *IEEE transactions on circuits and systems for video technology*, 17 (10), OCTOBER 2007, pp. 1360-1369.
- [29] R. K. Sharma, T. K. Leen, and M. Pavel, “Bayesian sensor image fusion using local linear generative models,” *Opt. Eng.*, 40 (7), 2001, pp.1364 – 1376.
- [30] A. M. Djafari, ‘Bayesian approach with hierarchical Markov modelling for data fusion in image reconstruction applications’, in *Proceedings of the Fifth International Conference on Information Fusion (FUSION)*, Vol. 1, 2002, pp. 440–447.
- [31] W. A. Wright, and F. Bristol, Quick Markov random field image fusion, *Proc. SPIE*, 3374, 302 – 308, 1998.
- [32] J. Yang, and R. S. Blum, “A statistical signal processing approach to image fusion using hidden Markov models,” in *Multi-Sensor Image Fusion and Its Applications*, Marcel Dekker/CRC, 2005
- [33] A. Rosenfeld, and M. Thurston, “Edge and curve detection for visual scene analysis,” *IEEE Trans. Comput.*, C-20, 1971, pp. 562–569.

- [34] P. J. Burt, and E. H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Commun.*, 31(4), 1983, pp. 532–540.
- [35] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publisher, Dordrecht, 1994.
- [36] D. Mueller, A. Maeder, and P. O’Shea1, *The Generalized Image Fusion Toolkit (GIFT)*, 2006.
- [37] P. J. Burt, “A gradient pyramid basis for pattern selective image fusion,” In: Proc. of the Society for Information Display Conf, 1992, pp. 467–470.
- [38] A. Toet, “A morphological pyramidal image decomposition,” *Pattern Recognition Letter*, 9 (3), 1989, pp. 255–261.
- [39] H. Li, B. S. Manjunath, and S.K. Mitra, “Multi-sensor image fusion using the wavelet transform,” *Graphical Models and Image Processing*, Vol. 57, No. 3, 1995, pp. 235-245.
- [40] G. Pajares, J.M. Cruz, “A wavelet-based image fusion tutorial,” *Pattern Recognition* 37 (9), 2004, pp.1855–1872.
- [41] O. Rockinger, “Image sequence fusion using a shift invariant wavelet transform,” In: *IEEE Internat. Conf. on Image Process*, 1997, pp. 288–291.
- [42] J. Nu’fez, X. Otazu, O. Fors, V. Pala, and R. Arbiol, “Multiresolution based image fusion with additive wavelet decomposition,” *IEEE Trans. Geosci. Remote Sens.* 37 (3), 1999, pp. 1204–1211.
- [43] Y. Chibani, A. Houacine, “Redundant versus orthogonal wavelet decomposition for multisensor image fusion,” *Pattern Recognition* 36 (4), 2003, pp. 1785–1794.
- [44] P. Hill, N. Canagarajah, D. Bull, “Image fusion using complex wavelets,” In: Proc. of the British Machine Vision Conf, 2002, pp. 487–496.
- [45] P. J. Burt, and R. J. Kolczynski, “Enhanced image capture through fusion,” *Proceedings of the 4th International Conference on Computer Vision*, 1993, pp. 173–182.
- [46] A. Toet, and E. M. Franken, Fusion of visible and thermal imagery improves situational awareness, *Displays*, 18, pp. 85–95, 1997.

- [47] H. Li, B. S. Manjunath, and S. K. Mitra, Multisensor image fusion using the wavelet transform, *Graphical Models and Image Processing*, 57 (3), pp. 235–245, 1995
- [48] I. Koren, A. Laine, and F. Taylor, Image fusion using steerable dyadic wavelet transforms, in *Proceedings of the IEEE International Conference on Image Processing*, Washington D.C., (1995) 232–235.
- [49] J. Lewis, R. O'Callaghan, S. Nikolov, D. Bull, and N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, *Information Fusion*, 8 (2), pp. 119–130, 2007
- [50] G. Piella, H. Heijmans, Multiresolution image fusion guided by a multimodal segmentation, in: *Proceedings of Advanced Concepts of Intelligent Systems*, Ghent, Belgium, pp. 1-8, 2002
- [51] N. Cvejic, D. Bull, and N. Canagarajah, Region-Based Multimodal Image Fusion Using ICA Bases, *IEEE SENSORS JOURNAL*, 7 (5), pp.743-75, 2007
- [52] T. Wan, N. Canagarajah, and A. Achim, Segmentation-Driven Image Fusion Based on Alpha-Stable Modeling of Wavelet Coefficients, *IEEE TRANSACTIONS ON MULTIMEDIA*, 11 (4), pp. 624-633, 2009
- [53] L Wald, “Some terms of reference in data fusion,” *IEEE Trans. Geosci. Remote Sens*, 37 (3), pp. 1190–1193, 1999.
- [54] J. Wahlen, “Comparison of standard and image-filter fusion techniques,” in *Data Mining III*, Southampton, U.K.: WIT Press, 2002
- [55] Y Zang, and R Wang, “Multi-resolution and multi-spectral image fusion for urban object extraction,” in *Proc. 20th ISPRS Congr*, pp. 960–966, 2004
- [56] R Colditz, T Wehrmann, M Bachmann, K Steinnocher, M Schmidt, G Strunz, and S Dech, “Influence of image fusion approaches on classification accuracy: A case study,” *Int. J. Remote Sens.*, 27 (15), pp. 3311–3335, 2006
- [57] T M Tu, W-Ch Cheng, Ch-P Chang, P S. Huang, and J-C Chang, “Best Tradeoff for High-Resolution Image Fusion to Preserve Spatial Details and Minimize Color Distortion,” *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 4 (2), APRIL 2007, pp. 302-306.

- [58] M. Choi, “A New Intensity-Hue-Saturation Fusion Approach to Image Fusion with a Tradeoff Parameter,” IEEE Transaction on Geoscience and Remote Sensing, 44, pp. 1672–1682, 2006.
- [59] Sh. P. Vijay, N. H. Younan, and R. L. King. “An Efficient Pan-Sharpening Method via a Combined Adaptive PCA Approach and Contourlets,” IEEE Transaction on Geoscience and Remote Sensing 46, pp. 1323-1335, 2008.
- [60] N. Mitianoudis and T. Stathaki, “Adaptive image fusion using ICA bases,” in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, May 2006.
- [61] C. Pohl, and J. L. Genderen, “Multisensor image fusion in remote sensing: Concepts, methods and applications,” Int. J. Remote Sens., 19 (5), pp. 823–854, Mar. 1998.
- [62] K. Amolins, Y. Zhang, and P. Dare, “Wavelet based image fusion techniques — An introduction, review and comparison,” ISPRS Journal of Photogrammetry & Remote Sensing 62, pp. 249–263, 2007.
- [63] T. Ranchin, L. Wald, “Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation,” Photogrammetric Engineering & Remote Sensing 66, pp. 49–61. 2000
- [64] S. Ioannidou, and V. Karathanassi, “Investigation of the Dual-Tree Complex and Shift-Invariant Discrete Wavelet Transforms on Quickbird Image Fusion,” IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, 4 (1), pp. 166-170, 2007
- [65] C. Ballester, V. Caselles, L. Igual, and J. Verdera, “A Variational Model for P+XS Image Fusion,” Int. J. of Comput. Vision, 69 (1), pp. 43–58, August 2006.

- [66] M. Ehlers, “Spectral Characteristics Preserving Image Fusion Based on Fourier Domain Filtering,” In: Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IV, Proceedings of SPIE, pp. 1–13, 2004
- [67] Sh. Li, J. T. Kwok, and Y. Wang, “Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images,” Information Fusion 3 (2002) 17–23, pp. 17-23
- [68] M. G. lez-Audicana, J. L. Saleta, O. G. Catalan, R. Garcí'a, “Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition,” IEEE Transactions on Geoscience and Remote Sensing 42 (6), pp. 1291–1299, 2004.
- [69] J. Portilla, V. Strela, M. Wainwright, E. Simoncelli, “Image denoising using Gaussian scale mixtures in the wavelet domain,” IEEE Transactions on Image Processing, 2003, pp.1338-1351.
- [70] G. Fan, X. Xia, “Image denoising using local contextual hidden markov model in the wavelet domain,” IEEE Signal Processing Letters, 2001, pp. 125-128.
- [71] M. Wakin, “Image compression using multiscale geometric edge models,” Master’s thesis, Rice Univ, Houston, TX, 2002.
- [72] S. G. Mallat, A Wavelet Tour of Signal Processing. Academic Press, San Diego, 1998.
- [73] Y. Meyer, Wavelets and Operators. Cambridge University Press, 1992.
- [74] G. Strang, and T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley MA02181 USA, 1997.
- [75] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image Coding Using the Wavelet Transform, IEEE Trans on Image Processing, 2 (2) , pp. 205-220, April 1992

- [76] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, “The Dual-Tree Complex Wavelet Transform,” IEEE Signal Processing Magazine, 2005, pp. 124–152.
- [77] O. Rockinger, “Image sequence fusion using a shift-invariant wavelet transform,” in Proc. ICIP, 1997, vol. 3, pp. 288–291.
- [78] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transform,” IEEE Trans. Inf. Theory, 38 (2), Mar. 1992, pp. 587–607.
- [79] N. G. Kingsbury, “A Dual-Tree Complex Wavelet Transform with Improved Orthogonality and Symmetry Properties,” Proc. IEEE Conf. on Image Processing, 2000.
- [80] N. G. Kingsbury, “Complex Wavelets for Shift Invariant Analysis and Filtering of Signals,” Journal of Applied and Computational Harmonic Analysis, 10(3), pp. 234-253, 2001.
- [81] I. W. Selesnick, “The Design of Approximate Hilbert Transform Pairs of Wavelet Bases,” IEEE Trans. on Signal Processing, 50(5), pp.1144-1152, 2002.
- [82] M. N. Do, and M. Vetterli, “Pyramidal directional filter banks and curvelets,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Thessaloniki, Greece, Oct. 2001
- [83] M. N. Do, and M. Vetterli, “The contourlet transform: An efficient directional multiresolution image representation,” IEEE Trans. Image Process, 14, pp. 2107–2116, Dec. 2005
- [84] J. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” Vis. Res., 20 (10), pp. 847–856, 1980.
- [85] A. B. Watson, “The cortex transform: Rapid computation of simulated neutral images,” Comput. Vis., Graph. Image Process, 39 (3), pp. 311–327, 1987.
- [86] A. L. Da Cunha, and M. N. Do, “Filter design for directional multiresolution decomposition,” in Proc. SPIE—Wavelet XI, 5914, pp. 256–265, Sep. 17, 2005

- [87] A. L. Da Cunha, J. P. Zhou, M. N. Do, "The nonsubsampled contourlet transform: theory, design and applications," *IEEE Transactions on Image Processing*, 15(10), pp. 3089-3101, 2006.
- [88] Y. X. Hui and J. L. Cheng, "Fusion Algorithm for Remote Sensing Images Based on Nonsubsampled Contourlet Transform," *ACTA AUTOMATICA SINICA*, 34 (3), pp. 274-281, 2008.
- [89] T. T. Nguyen, and S. Oraintara, "The shiftable complex directional pyramid—Part 1: Theoretical aspects," *IEEE Trans. Signal Process.*, 56 (10), Oct. 2008.
- [90] R. H. Bamberger, M. J. T. Smith, "A filter bank for the directional decomposition of images: theory and design," *IEEE Transaction on Signal Processing* 40 (4), pp. 882-893. 1992
- [91] S. I. Park, M. J. Smith, R. M. Mersereau, "Improved structure of maximally decimated directional filter banks for spatial image analysis," *IEEE Transaction on Image Processing* 13 (11), pp. 1424-1431. 2004
- [92] J. Saeedi, K. Faez, "A classification and fuzzy-based approach for digital multi-focus image fusion," *Pattern Analysis and Application Journal* 16 (3), pp. 365-379, July 2013.
- [93] J. Saeedi, K. Faez, "Fisher Classifier and Fuzzy Logic Based Multi-Focus Image Fusion," *IEEE International Conference on Intelligent Computing and Systems Intelligent*, pp. 420-425, 2009.
- [94] J. Saeedi, K. Faez, and S. Mozaffari, "Multi-focus Image Fusion Based on Fuzzy and Wavelet Transform," *CIARP 2009, LNCS 5856*, Springer-Verlag Berlin Heidelberg, pp. 970–977, 2009.
- [95] S. Wei, and W. Ke, "A Multi-Focus Image Fusion Algorithm with DT-CWT," *International Conference on Computational Intelligence and Security*, 2007, pp. 147-151.
- [96] Y. Zheng, H. Li, and D. Doermann, "Machine printed text and handwriting identification in noisy document images," *IEEE Trans. Pattern Anal. Mach. Intell.*, 26 (3), 2004, pp. 337–353.

- [97] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model," *IEEE Trans. On Image Processing*, 16 (8), 2007, pp 2117-2128.
- [98] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second Ed. New York: Academic Press, 1990.
- [99] K. Kannan, and S. A. Perumal, "Optimal Decomposition Level of Discrete Wavelet Transform for Pixel based Fusion of Multi-focused Images," *International Conference on Computational Intelligence and Multimedia Applications*, 2007, pp. 314-318.
- [100] V. S. Petrovic, and C. S. Xydeas, "Gradient-Based Multiresolution Image Fusion," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 13 (2), 2004, pp. 228-237.
- [101] J. Saeedi, K. Faez, "Infrared and visible image fusion using fuzzy logic and population-based optimization," *Elsevier Applied Soft Computing Journal* 12 (3), pp. 1041–1054, March 2012.
- [102] J. Saeedi, K. Faez, "The New Segmentation and Fuzzy Logic based Multi-Sensor Image Fusion," *24<sup>th</sup>IEEE International Conference Image and Vision Computing New Zealand*, pp. 328-333, 2009.
- [103] N.G. Kingsbury, The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters, in Proc. 8th IEEE DSP Workshop, Utah, Aug. pp. 9–12. 1998
- [104] L. Vincent, and P. Soille, Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Machine Intel.*, 13, pp. 583–593. 1991
- [105] P. Scheunders, and J. Sijbers, Multiscale watershed segmentation of multivalued images, *IEEE Proceedings*, 2002.
- [106] R. C. Gonzalez, R. E. Woods, *digital image processing*, Tom Robbins, Upper Saddle River, New Jersey, 2002.

- [107]V. Petrovic, C. Xydeas, On the effects of sensor noise in pixel-level image fusion performance, in: Proceedings of the Third International Conference on Image Fusion, 2, pp. 14–19, 2000
- [108]S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by Simulated Annealing, *Science* 220, pp. 671–680, 1983
- [109]S. Rao, Optimization: theory and applications, Wiley, New York, 1984
- [110]J. Kennedy and R. C. Eberhart, A discrete binary version of the particle swarm algorithm, In Proceedings of the 1997 IEEE Conference on Systems, Man, and Cybernetics, Piscataway, New Jersey, pp. 4104–4109, 1997
- [111]E. Goldberg, Genetic Algorithms in Search, Optimization & Machine Learning, Addison-Wesley, 1989
- [112]Y. Shi, and R.C. Eberhart, Parameter selection in particle swarm optimization, In Proceedings of the Seventh Annual Conference on Evolutionary Programming, 591-600, 1998
- [113]Y. Shi, and R.C. Eberhart, Empirical study of particle swarm optimization, In Proceedings of the IEEE Congress on Evolutionary Computation (CEC), pp. 1945-1950, 1999
- [114]J. Kennedy, The behavior of particles. In: Porto VW, N. Saravanan, D. Waagen, and AE. Eiben (eds) Evolutionary Programming VII, Springer, pp. 581–590. 1998
- [115]A. Carlisle, and G. Dozier, An Off-The-Shelf PSO, Proceedings of the Particle Swarm Optimization Workshop, pp. 1–6. 2001
- [116]A. Garzelli, F. Nencini, “Pan-sharpening of very high resolution multispectral images using genetic algorithms,” *International Journal of Remote Sensing* 27 (15), pp. 3273–3292. 2006
- [117] A. Garzelli, F. Nencini, L. Capobianco, “Optimal MMSE Pan Sharpening of Very High Resolution Multispectral Images,” *IEEE Transaction on Geoscience and Remote Sensing* 46 (1), pp. 228-236. 2008

- [118] X. Li, "A non-dominated sorting particle swarm optimizer for multiobjective optimization," In: Proc. of the Genetic and Evolutionary Computation Conference, Chicago, USA, 3 August, pp. 37-48. 2003
- [119] C.A. Coello, G. T. Pulido, M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," IEEE Transaction on Evolutionary Computing 8 (3), pp. 256-279. 2004
- [120] J. Saeedi, K. Faez, "A new pan-sharpening method using multiobjective particle swarm optimization and the shiftable contourlet transform," Elsevier Photogrammetry & Remote Sensing Journal 66, pp. 365-381, 2011.
- [121] C. Thomas, T. Ranchin, L. Wald, J. Chanussot, "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 (5), pp. 1301-1312. 2008
- [122] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion?," In: Proc. of the third conference Fusion of Earth data: merging point measurements, raster maps and remotely sensed images, Sophia Antipolis, France, 26-28 January, 6 p., 2000
- [123] J. Zhou, D. L. Civco, J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," International Journal of Remote Sensing 19 (4), pp. 743-757. 1998
- [124] M. R. Sierra, C. A. Coello, "Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art," International Journal of Computational Intelligence Research 2 (3), pp. 287-308. 2006
- [125] O. Rockinger, and T. Fechner, "Pixel-level image fusion: the case of image sequences," Proc. SPIE, 3374, 1998, pp. 378 – 388.
- [126] Y. Niu and L. Shen, A Novel Approach to Image Fusion Based on Multi-Objective Optimization, Proceedings of the 6th World Congress on Intelligent Control, 2009, pp. 9911-9915.
- [127] <http://www.imgfsr.com>.

- [128] R. H. Yuhas, A. F. H. Goetz, J. W. Boardman, Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: Proc. Summaries of the Third Annual JPL Airborne Geoscience Workshop, 1 Jun, pp. 147–149. 1992
- [129] L. Alparone, S. Baronti, A. Garzelli, F. Nencini, “A global quality measurement of Pan-sharpened multispectral imagery,” IEEE Geoscience and Remote Sensing Letters 1 (4), pp. 313–317, 2004
- [130] Z. Wang, A. C. Bovik, A Universal Image Quality Index. IEEE Signal Processing Letters 9 (3), 81–84. 2002
- [131] D. Fasbender, J. Radoux, P. Bogaert, Bayesian Data Fusion for Adaptable Image Pansharpening. IEEE Transaction on Geoscience and Remote Sensing 46 (6), pp. 1847-1857, 2008
- [132] Q. Du, N. H. Younan, R. King, V. P. Shah, On the Performance Evaluation of Pan-Sharpening Techniques. IEEE Geoscience and Remote Sensing Letters 4 (4), pp. 518-522, 2007.
- [133] Y. Zhang, G. Hong, An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. Information Fusion 6 (3), pp. 225-234, 2005.



## Image Fusion in the Multi-Scale Transforms Domain

The research presented in this thesis is concerned with the problem of multi-sensor pixel-level image fusion. Generally, the image fusion's task is used for three different applications consisting of fusing the multi-focus images, fusing the infrared and visible images, and fusing the multi-spectral and panchromatic images.

We formulate the image fusion process as a two-class problem: in focus and out of focus classes, in which the decision map for selecting important coefficients between input images is obtained using two-class fisher classifier. In the proposed method for fusing infrared and visible images, first, the input images are decomposed using dual-tree discrete wavelet transform and then, we use a dissimilarity measure of source images to combine three different fusion rules for selecting high frequency wavelet coefficients between source images. Finally, a new method for fusion of remote sensing images is proposed. In this method, the aim is to improve spatial and spectral quality of the fused image, simultaneously. We use the shiftable contourlet transform and multi-objective particle swarm optimization for this purpose.