The University Of Sheffield.

Advanced Methods for Fusion of Remote Sensing Images and Oil Spill Segmentation

A thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy

By:

Fodio S. Longman

Department of Automatic Control and Systems Engineering

March 2020

Dedicated to My Mother Fatima, my wife Aisha and my sons Muhammad and Umar....

Acknowledgements

First and foremost, I would like to thank God Almighty for giving me the strength, knowledge, and ability to undertake this research study. I would like to express my deepest appreciation and sincere gratitude to my supervisor Professor Lyudmila Mihaylova, for the continous support of my PhD study and research, for her patience, enthusiasm and immense knowledge. Her guidance and motivation helped me in all the time of the research and writing of this thesis. I couldn't have imagined a better supervisor for my PhD.

Again, I would like to thank Professor Daniel Coca for his encouragement. My sincere thanks also to my friends, for their support, encouragement, and time especially Jinny Robson, Uche Diala, Miguel Arteita, Dr. Bashir Mohammed, Hon. Sagir Attahiru Bafarawa, Aminu Jega, Barrister Mujahid Indabawa, Dr. Umar Sani Abdullahi and Kabiru Sambo

I am deeply indebted to my family, my parents for their prayers, support and encouragement throughout my study. My heartfelt appreciation goes to my wife Aisha, my sons Muhammad and Umar for their love, patience and understanding during my study.

I would also like to offer my special thanks to my role model Alh. (Dr) Attahiru Dalhatu Bafarawa (Garkuwan Sokoto) who treats me as a son and ensured i got a scholarship to pursue a PhD. I pray that the Almighty Allah grant him the highest place in paradise. Many thanks to Dr. Abdullahi S.G. Adiya for his support and many others that have contributed in one way or the other towards the successful completion of my PhD.

Lastly, I would like to gratefully acknowledge the funding support from the Petroleum Trust Development Fund (PTDF) in Nigeria and my employer, the National Space Research and Development Agency (NASRDA) for granting me study leave to pursue a PhD.

Abstract

Remote sensing systems on board satellites (spaceborne) or aircraft (airborne) have continued to play significant role in disaster management and mitigation, including for oil spill detection due to their ability to obtain wide area coverage images and other data from a distance. A single remotely sensed image can cover hundreds of kilometres of the earth surface enabling wider monitoring and change detection observation. When oil spill occur, remote sensing systems equipped with different sensors covering the spectral bands of the electromagnetic spectrum are deployed to obtain images for damage assessment, scientific analysis or to ascertain spill location, amount of oil spilled and the type of oil for efficient planning, management and illegal dumping of ballasts identification for legal actions.

In the design of such remote sensing systems, there are usually considerate trade-offs that are inevitable due technological limitations of such systems, resulting in spatial and spectral amendments. Panchromatic sensors for example obtain images at high spatial resolutions but lower spectral resolutions, while hyperspectral images obtain high spectral images but in lower spatial resolutions. Additionally, optical systems depend on external energy sources to obtain the images while others can acquire data irrespective of weather conditions. By combining data originating from different sources, scientists, analyst and planners can have images of higher quality than the individual images and can take advantage of the complimentary information embedded in diverse data acquired.

This thesis presents a new framework for oil spill detection by combining data originating from different imaging sensors of remote sensing systems. Firstly, the new framework for oil spill segmentation utilises the fusion of images to improve image quality and to take advantage of complimentary information available in the different resolutions of SAR images. The framework adopts the wavelet image fusion technique where the individual images are converted from spatial to frequency domain and decomposed to approximations and detail coefficients, allowing image properties to be transferred using a maximum fusion rule. While this method improves spatial resolution of images and retains colour information, it is observed that the scale of decomposition needs to be sensibly selected since smaller scales creates mosaic effects and large scale values causes loss of colour contents making it unsuitable for images with different spectral channels. To solve the problem of multi-modality in images, a Gaussian Process (GP) regression approach is utilised using a custom covariance that learns the geometry and intensity of pixels and also handles the change of support problem inher-

ent in multi-resolution images. Established performance metrics in the literature are used to evaluate the quality of the fused images when compared with a reference data. Additionally, a qualitative and quantitative review of pansharpening methods for hyperspectral images is carried out specifically for the oil spill detection application. The pansharpened results are compared in terms of unmixing performance with a reference hyperspectral image. This review can help researchers interested in this field of study to determine what methods are best for pansharpening and unmixing and to answer the question of whether pansharpening improves unmixing result. This can be extended for other applications that include weather forecasting, spectral analysis etc. Lastly, the a new covariance kernel is developed to solve image fusion problems in multi-band images by treating differently each spatial and spectral channels as input to the Gaussian process allowing different spatial and spectral pixels of the images to be learned and combined. The developed approach allows the transfer of information between different image modalities enabling local structure of high spatial resolution images that forms the base of the estimated image to be recovered. The developed fusion approaches achieves compelling enhancement when compared with the state of the art. Furthermore, segmentation is done on the fused and reference images with the developed fused image picking up more objects from the image than other methods. This can be attributed to the ability of the approach to sharpen the resolution of the spectral channels that supports pixel coordinates from high spatial image that improves edges of the image.

Contents

List of Figures

List of Tables

List of Acronyms

ANN : Artificial Neural Networks

ANM: Arborescence Network Matching

ASAR: Advanced Synthetic Aperture Radar

BLE: Bayesian Linear Estimator

CC: Correlation Coefficient

CS: Component Substitution

CS: Component Substitution

ddCRP: Distance Dependent Chinese Restaurant Process

DoG : Difference of Gaussian

DWT : Discrete Wavelet Transform

ERGAS : Erreur Relative Globale Adimensionnelle de Synthese

ESA : European Space Agency

ERS : European Remote Sensing

FCLS : Fully Constrained Least Square Method

GFPCA : Guided Filter Principal Component Analysis

GLCM : Gray Level Co-occurence Matrix

GP : Gaussian Process

GS : Gram Schmidt

HH : Horizontal transmit, Horizontal receive

HSI : Hyperspectral Image

HV : Horizontal transmit, Vertical receive

IHS : Intensity Hue Saturation

IWT : Inverse Wavelet Transform

KNN : K Nearest Neighbour

MAP : Maximum a posteriori

MCMC : Markov Chain Monte Carlo

MODIS : Moderate Resolution Imaging Spectroradiometer

MRA : Multi Resolution Analysis

MRF : Markov Random Field

PAN : Panchromatic Image

PC : Principal Component

PCA : Principal Component Analysis

PPNM : Polynomial Post Nonlinear Mixture

RANSAC : Random Sample Consensus

RMSE : Root Mean Squared Error

ROC : Receiving Operating Characteristics

ROI : Region of Interest

SA : Simple Averaging

SE : Squared Exponential

SAM : Sample Angle Mapper

SAR : Synthetic Aperture Radar

SFIM : Smoothening Filter-Based Intensity Modulation

SIFT : Scale Invariant Feature Transform

SVME : Support Vector Machine Ensemble

SVM: Support Vector Machine

VCA : Vertex Component Analysis

VH : Vertical transmit, Horizontal receive

VV : Vertical transmit, Vertical receive

VTV : Vector Total Variation

List of Symbols

$\mathbf{P}$	Panchromatic Image
$\widetilde{\mathbf{H}}$	Interpolated hyperspectal Image
$\mathbf{H}$	Hyperspectral Image
r	Spatial resolution ratio between Images
$\mathbf{F}^k$	kth row in $\widehat{\mathbf{F}}$
$()^T$	Transposition Operation
$\widehat{\mathbf{F}}$	Fused or Pansharpened Image
$\mathbf{X}$	Matrix
$\mathbf{x}$	Vector
$\mathbf{I}(x, y)$	Image
$\mathbf{x}_*$	Testing Data
$\mathbf{x}^\top$	Transpose
$\mathbf{X}^\top$	Transpose of $\mathbf{X}$
$*$	Convolutional Operator
$\mathcal{N}(\mu, \Sigma)$	Normal Distribution with mean μ and covariance Σ
$P(.)$	The probability Operator
ϵ	Sample noise
$\mathbb{R}^*$	Real numbers of Dimension $*$
$\lvert . \rvert$	Determinant
$\| \, . \, \|_F$	Frobenious Norm
$\sum_{i=1}^{N}$	Sum operation with index 1-N
$\boldsymbol{\theta}$	Vector of Hyperparameters
$P(. \mid .)$	Conditional Probability Operator
$\propto$	Proportional to
$exp(.)$	Exponential Function

λ	spectral wavelength
$\oslash$	Element-Element Division
ψ	Wavelets
$\langle . \rangle$	Inner Product
$\boldsymbol{\omega}$	Abundance Matrix
$\mathcal{P}$	Average of overlapping Patches

Chapter 1

Introduction

1.1 Overview

Oil and its products are vital to our daily life activities. Products from oil such as petrol and gas are used both industrially and domestically. This requires transporting oil between continents and countries across the world. During this transportation, oil spill can happen [1, 2]. Oil spills are caused by a number of factors for instance tanker accidents during oil transportation or transfers and dumping of ballasts by vessels during operations [3]. It is observed, that most operational oil spills tally with major shipping routes or show up at offshore oil installations or platforms [4]. Globally, it is estimated that annually, 48% of spills leading to sea pollution are caused by fuels, 29% by crude and only 5% by tanker accidents [5]. Additionally, oil spill statistics between the years 1988-2000 shows that a total of 2,475 oil spills occurred, releasing an approximate 800,000 litres of oil in Toronto and its regions alone [3]. Recently, an oil tanker collision in Eastern China spilled around 138,000 tonnes of crude oil into the marine environment leading to a significant degradation and polluting the environment [6]. In Table 1.1, a summary of some major oil spills around the world are shown. The data is available on https://ourworldindata.org/oil-spills and was last accessed on the 11th november, 2019 at 4:00pm GMT.

Marine oil spills (e.g. The BP oil spill of 2010 in the Gulf of Mexico) severely pollute the marine environment causing immense damage both socially and economically [8]. The sustenance of persons living along the coastal areas and tourism are most affected particularly those that rely on fishing as means of livelihood. Sea birds are affected with behavioural changes that results in the loss of eggs and their eventual deaths [3]. When oil spill happens,

Table 1.1: List of some oil spills around the world [7]

Ship/Vessel	Location	Dates	Min Tonnes	Max Tonnes
MV Solomon	Solomon Islands	05/02/2019	75	80
Sanchi Oil Tanker Collision	East China Sea	06/11/2018	138,000	138,000
Doon Iowa Derailment	Lyon County Iowa	22/06/2018	520	520
Agia Zoni II	Saronic Gulf, Greece	10/09/2017	2,500	2,500
Con Edison	New York	07/05/2017	97	101
Ennore Oil Spill	India Chennai	28/01/2017	251	251
Black Sea	Tuapse, Russia	24/12/2014	unknown	unknown
MV Marathassa	English Bay (Vancouver)	22/03/2014	546	546
Bullenbai	Isla Refinery	01/11/2013	43	163
North Dakota Pipeline Spill	Tioga, North Dakota	25/09/2013	2,810	2,810
Mayflower	Arkansas, Magnolia	30/03/2013	680	950
Nigeria	Bonga Field	21/12/2011	5,500	5,500
Deep Water Horizon	Gulf of Mexico	20/04/2010	492,000	627,000

it quickly spreads on the surface forming a thin layer often referred to as "Oil Slick" that causes quick environmental deterioration [2]. Additionally, there is a growing concern by governments, the public and media to ascertain the location and extent of the spill to enable adequate preventive measures to be taken and to mitigate the impact of the damage caused [1]. Early detection of oil spill therefore, can go a long way in alleviating the effects caused by oil spill disaster, avoidance of the widespread and continuous pollution, and to help in contingency planning and clean-up process.

A conventional approach to oil spill monitoring is the use of trained personnel to visually inspect and report suspected spill location, and in many instances sample the type of oil spilled. This is however a non-effective approach because visual identification of oil spills can be difficult due to the obvious fact that oil slicks on water surfaces can be mistaken for other substances e.g. weeds or fish sperm [9], prompting the need for more effective approach in detecting and identifying oil spills. In the last decade, the monitoring of oil spills have taken a new turn with the use of remote sensing data. Remote sensing systems on-board a satellite (spaceborne) or aircraft (airborne) are used to obtain images of the earth from a distance which can help more efficiently in monitoring and detection of oil spills. When an oil spill occur, these systems equipped with various sensors in different bands of the electromagnetic spectrum are deployed to acquire data (mostly images) for detection and monitoring purposes e.g in hyperspectral and multispectral bands or Synthetic Aperture Radar (SAR) [5, 8, 10, 11], with each sensor having both individual advantages and drawbacks as a result of spatial and spectral specifications that describes the sensor. Hyperspectral and Multispectral sensors (e.g. AVIRIS, Hyperion) for example are high in spectral resolutions with bands varying between

4-few hundreds. The advantage of this specification is that materials present in the scene of interest can be identified through spectroscopic analysis hence facilitating the identification of not only spill, but the type, amount and quantity of spilled oil can be estimated with spectral unmixing [12]. Unfortunately, these sensors being passive and depending on external source of energy (the sun) to acquire energy are affected by weather conditions and cloudiness. On the other hand, SAR systems such the Canadian RADARSAT 1&2 and the European Space Agency's Envisat have been adopted widely for oil spill detection and monitoring as they are not affected by local weather conditions or cloudiness [13]. The ability of SAR instrument to detect oil is however affected by sea conditions. For instance, low sea state conditions (1-3 on the Douglas scale) will not produce enough sea clutter in the sorrounding area to contrast with oil, while very high sea conditions (7-9 on Douglas scale) will scatter radar sufficiently, blocking detection inside the wave troughs [1]. SAR is an active microwave sensor that exploits Bragg Scattering effect in the microwave radiation that comes from sensor mechanisms and incides on the surface of the sea [14]. The C-band of the microwave spectrum allows the interaction of short gravity waves that are generated by winds that blow over the sea surface and enabling the formation of dark patches on SAR images. Under appropriate wind conditions, oil spill will appear as dark formations on acquired images using SAR [15]. Unfortunately, other sea phenomena known as "lookalikes" also produce dark patches making the detection with SAR images a complex process by creating false alarms.

The use of remote sensing images for oil spill detection is evident in many scientific researches [2, 3, 8, 11, 14, 16–19]. In the past years and many approaches and methods have been proposed towards automated and semi-automated solutions. In particular, SAR images are the most widely used for this purpose. The proposed methods mostly share a common process flow and only differs by approaches taken within each process step. A standard block structure for oil spill detection is presented in Fig 1.1, with the most important process being segmentation.

Recents developments in the design and deployment of remote sensing systems have enabled a wide range of sensors to be produced and deployed for many applications in earth observation. Furthermore, different sensors can be mounted on the same aircraft to acquire image data in different modalities, providing the research community with a wide range of data in different bands of the EMS, multi-frequency (C, L, X, S etc) in the case of SAR sensors, multi-angle, multi-polarisations and multi resolutions. The advantage of multi modal data for oil spill detection cannot be over emphasised. For instance with multi-modal

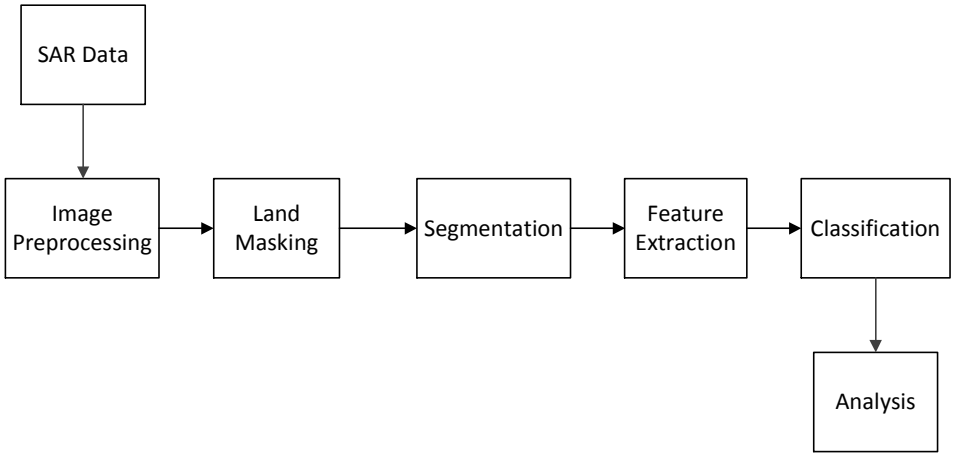


Figure 1.1: Common block structure of most proposed oil spill detection methods [20].

data configuration, new approaches for oil spill detection can be explored by combining complimentary data sources with an aim of obtaining more detailed information of the scene of interest through Image fusion.

Image fusion however posses several challenges that include firstly aligning the different images as a result of differences in modality through image registration. Another problem is the issue of sensor mechanisms that deals with the mode of acquisition of images and lastly the selection of suitable image fusion method that will efficiently bring together the complementary and supplementary information together since image fusion can be performed at different levels including pixel and decision levels. In the case of SAR images there is also the issue of speckle noise that create a poor visualisation. Hence de-speckling is a necessary pre-processing step.

The focus of this study therefore, is to investigate a new detection framework that include the fusion of different modalities of image data to the oil spill segmentation framework and determine if the quality of the fused images is better than the individual images before fusion. Additionally, answer the question of whether the fusion of images in the optical region can improve the unmixing of the different materials on sea surfaces including oil spills and other materials. Additionally, a comparison of band sharpening methods for optical data is done and extended to address the research question of whether band sharpening can improve spectral unmixing by comparing with both linear and non-linear spectral unmixing methods of the state-of-the art.

1.2 Motivation

Remote sensing have continued to play an important role in earth observation especially for disaster monitoring where it can be deployed in all the phases of disaster management including for response, mitigation and recovery. Recent developments have enabled more airborne and spaceborne systems to be modelled and launched, acquiring images in the different regions of the Electro Magnetic Spectrum (EMS), providing a wide range of data for scientific research, assessment and analysis. In oil spill detection, SAR data is regarded as the fundamental tool for oil spill detection due to its all weather data acquisition capability, high spatial resolution and large coverage. Despite this, the detection of oil spill in SAR images depends on wind conditions. On the other hand, hyperspectral imagers can acquire images in contiguous spectral bands of up to few hundreds enabling more materials within a scene of interest to be identified as well as other sensors have all been utilised for spill detection using remote sensing.

The number of sensors deployed when oil spill occur is proportionate to the amount of data acquired. Additionally attempts are being made to increase the performance of different imaging sensors for example the provision of polarimetric data in SAR sensors, or the deployment of different sensors on an airborne or spaceborne system to acquire multi-modal data simultaneously e.g. the hyperspectral Imager Suite (HISUI) [21] system that can acquire both hyperspectral and multispectral data providing information at spectroscopic level that is beneficial to a wide range of remote sensing and many other image processing applications [22].

This thesis aims to utilise the different images obtained by these sensors with different modalities through image fusion to improve the quality of images for improving segmentation process and by extension the detection of oil spill in remote sensing images. It is considered also, that sensors deployed for this application can obtain multiple images of a same scene from different angles and at different times. If these images are combined that can be of benefit for change detection, object recognition and other computer vision analysis and applications. Further more, the trade-offs that usually happen when designing remote sensing systems can be remedied by taking the complimentary advantages of individual images and combining them to get a more accurate information and improve image perception and quality.

1.3 Thesis Overview

This thesis is structured in the following chapters:

- **Chapter 1**: Introduces the research topic and the aim of the study, followed by the motivation of this research, outline of the thesis, its summary of contributions and the list of published papers from the research.

- **Chapter 2**: Reports the background of the study that includes sensors utilised for oil spill remote sensing, reviews of fundamental image fusion approaches including wavelets and regression approaches using Gaussian processes, followed by an overview of state-of-the-art image segmentation techniques utilised for oil spill detection.

- **Chapter 3**: Presents a quantitative and qualitative review of band pansharpening methods in the literature and further investigates whether band pansharpening improves spectral unmixing of optical images in the context of oil spill unmixing. Linear and non-linear spectral unmixing methods are compared for this purpose on each of the band pansharpening approaches. A real life Hyperspectral image of the Gulf of Mexico oil spill as acquired by SpecTIR hyperspectral imager is utilised as a reference and semi-synthetic hyperspectral and panchromatic images are simulated from the reference data for the study.

- **Chapter 4**: Proposes an initial framework for oil spill segmentation using fused Synthetic Aperture Radar (SAR) images from different modalities. Radarsat-2 & ESA's Envisat C-band SAR images of the Gulf of Mexico oil spill are utilised for this study. The proposed framework takes as input multi-modal SAR data from the different sensors and firstly co-register them in space and time using scale invariant feature transform (SIFT) registration approach. The purpose is to align the images before fusion. Next, a Discrete wavelet Transform (DWT) in the HAAR approach is utilised to perform image fusion at pixel level where the images are converted from spatial to frequency domain using a max fusion rule. The framework is concluded by segmenting the dark areas of the fused image using a Polynomial Fitting curve on the image histogram and utilising threshold technique. This is extended with Edge detection using Canny approach to further detect the border edges of the dark spots in the fused image.

- **Chapter 5**: A Gaussian Process Regression algorithm is utilised for the fusion of multi-modal SAR images to improve oil spill segmentation. A non-stationary covariance co-

variance kernel is employed to handle the change of support problems exists in multi-resolution images. The approach considers different image modalities and images in different spatial and spectral resolutions for fusion. The developed prior covariance kernel which is a product of sparse covariance kernel and a rational quadratic kernel is utilised to model high spatial resolution pixel coordinates and their corresponding intensity values, forming a base covariance function from which a new modality image is constructed. To segment the dark areas, K-means clustering algorithm is utilised since different modalities are involved. The approach converts the image into *lab* color spaces and initialise the value of clusters from this and allowing more objects to be identified in the segmentation step. The GP approach is extended to handle multi-band images where the images to be fused each contains more than a single band. This is achievable by learning seperately the bands as individual band bases of the fused image to be recovered.

- **Chapter 6**: Summarises the thesis results, reports conclusion and propose directions for future work.

The thesis overview describes the chapters presented in this work. The contributing chapters are structured to account for different scenarios of fusion based on data format. Chapter 3 compares hyperspectral pansharpening methods proposed in the literature where the aim is to improve the spatial resolution of hyperspectral images by sharpening them with a high spatial panchromatic image. In chapter 4, the fusion framework is introduced to the oil spill segmentation procedure using the discrete wavelet algorithm algorithm and utilising SAR data in C-band (single band from each sensor) from different sensors. The Images were acquired at different times from different angles, hence the chapter also utilises an image registration algorithm to achieve a common frame between the images before the fusion process. The decomposition is done up to level 4 vand results are compared. In chapter 5, a new fusion approach is presented that utilises machine learning methods and in particular the GP to model the spatial and spectral characteristics of the images. In the approach, the polarimetric RADAR-Sat 2 is utilised and fused with a single polarised channel of the Envisat. This approach is further extended to handle multi-band images where the individual images both have spectral channels making it robust in handling diverse data for fusion.

1.4 Main Contributions

The thesis explores new frameworks for fusing remote sensing images from different modalities. In particular, oil spill segmentation is explored with fused images to answer the research question of whether the fusion or pansharpening of remote sensing images can improve the segmentation process of dark formation on SAR images and spectral unmixing in hyperspectral images. The contributions of this research are listed as follows

- A qualitative and quantitative review of pansharpening methods is carried out using an oil spill reference Hyperspectral image of the Gulf of Mexico oil spill and a further comparative study is done for both linear and non-linear spectral unmixing approaches to compare the performance of the pansharpening methods in material unmixing.

- A new framework is developed for oil spill segmentation in fused SAR images. The framework takes as input complimentary SAR images of a same scene acquired by different platforms at different time intervals. Firstly the images are registered in time and space by extracting from the images SIFT features and matching corresponding features so that per-pixel alignment is achieved. DWT pixel level image fusion is the utilised to fuse the images using a select maximum fusion rule. The segmentation of dark areas on the fused image is carried out using step is done by combining a polynomial fitting curve segmentation approach and edge detection.

- A new image fusion approach is developed using Gaussain Process regression. The proposed approach is based on the design of a non-stationary covariance kernel which is a product of a sparse covariance to handle the change of support problem and a rational quadratic kernel that models the geometry of pixels and their corresponding intensities forming the spatial resolution prior of the estimated image to be recovered.

- An Adaptive multi-output Gaussian Process to fuse multi-band images is introduced. The model utilises spatial information of the high spatial resolution bands using a sparse covariance kernel to learn the geometry of image pixels forming a base prior of spatial bands for the estimated image to be recovered. A new learning strategy is adopted where individual bands are learnt separately.

1.5 Publications

The following publications with relevance to this study are listed below

Peer Reviewed Conferences

- [*P*1] Fodio S Longman, Lyudmila Mihaylova and Daniel Coca, "Oil Spill Segmentation in Fused Synthetic Aperture Radar Images" , in Proceedings of 4th International Conference on Control Engineering & Information Technology, 16-18 Dec 2016, Hammamet, Tunisia. IEEE. ISBN 978-1-5090-1055-4.

- [*P*2] Fodio S Longman, Lyudmila Mihaylova and Le Yang, "A Gaussian Process Approach for Fusion of Remote Sensing Images for Oil Spill Segmentation", in Proceedings of the 21st International Conference on Information Fusion, 10-13 July 2018, Cambridge, UK. IEEE. ISBN 978-0-9964527-6-2.

- [*P*3] Fodio S Longman, Lyudmila Mihaylova, Le Yang and Topouzelis Konstantinos, "Multi-band Image Fusion using Gaussian Process Regression with Sparse Rational Quadratic Kernel", in Proceedings of the 22nd International Conference on Information Fusion, 02-05 July 2019, Ottawa, Canada. IEEE.

Journal Paper in Preparation

- [*P*4] Fodio S Longman, Asma'u Ahmed, Lyudmila Mihaylova and Daniel Coca "A Comparative Analysis of Hyperspectral Pansharpening Methods for Spectral Un-mixing of Oil Spill" to be submitted to the International Journal on Multi-Sensor, Multi-Source Information, ELsevier (2019).

Chapter 2

Background and Related Work

2.1 Oil Spill Remote Sensing: Background

Remote sensing systems have continued to play an increasingly important role in disaster mitigation and response. These systems, offer the advantage of being able to observe events in remote and mostly inaccessible areas through the acquisition of images and other data. A single remotely sensed image can cover hundred of kilometres of land on the earth's surface. Thanks to such advantages, remote sensing image data can be used extensively in the monitoring of different disasters on earth including for oil spill detection and monitoring. Remote sensing systems are deployed to monitor usually on a 24-hour basis the spread, location and damages caused. Between 1970 and 2016, it is reported that about 600,000 tonnes of oil was spilled into the marine environment by tanker accidents alone [7]. Remote sensing systems are designed to operate in frequencies within the electromagnetic spectrum range from visible, to microwave, as shown in Fig 2.1.

Conventionally, remote sensing systems have provided a secondary role in oil spill detection [10]. However, recent development in sensor design, deployment and instrumentation have enabled a promising role for remote sensing in oil spill detection, with the ability to offer a 24-hr monitoring of the open sea and provide adequate mapping of spill location and its extent [1]. Additionally, remote sensing sensors can provide the means to quantify the amount and type of oil spilled hence providing ample information for disaster response and polluted environmental recovery. These systems, * usually deployed for oil spill monitoring can be classified into two namely; active and passive sensors or according to platform whether they are mounted on an aircraft (airborne) or on a space satellite (spaceborne). Sensors deployed

*remote sensing systems and sensors are used interchangeably in this chapter

for this purpose utilises the sea surface properties that include colour, reflectance, temperature, sea roughness and wind conditions to operate [23]. A detailed review of these systems is discussed in [1, 3, 10, 24, 25].

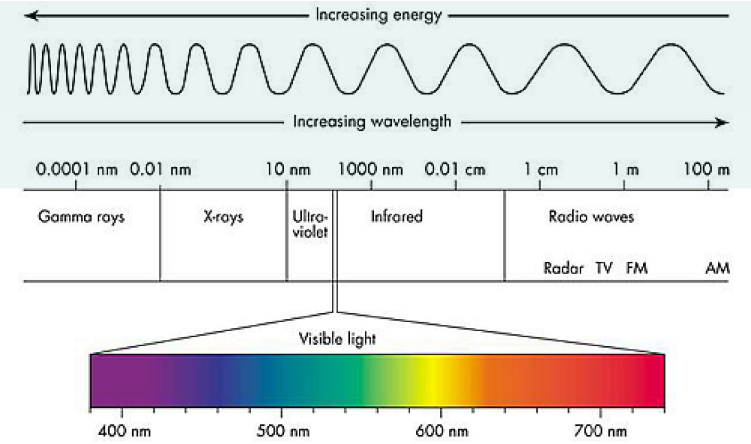
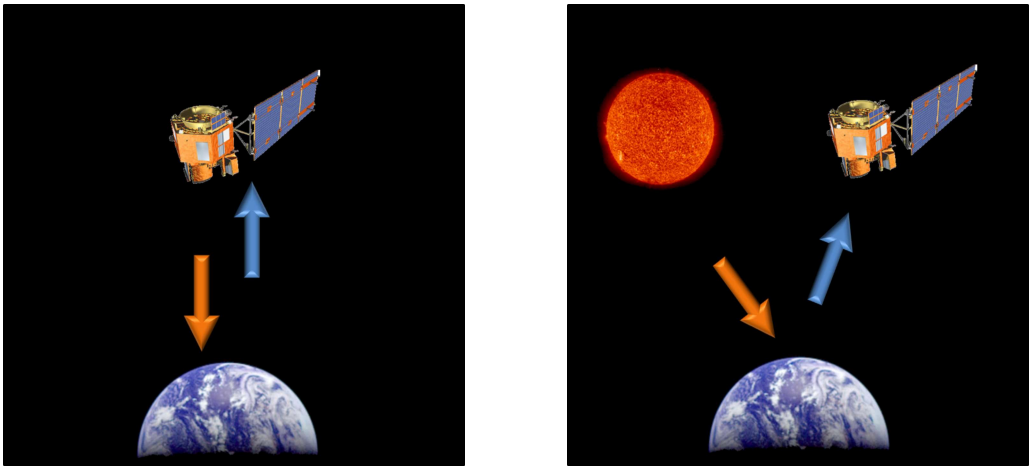


Figure 2.1: Wavelength range of the electromagnetic spectrum (EMS) [1].

Active sensors illustrated in Fig 2.2 (a), are mostly radar instruments that measure transmitted signals when reflected by the sensor. They are widely used for oil spill remote sensing since their electromagnetic waves can be deployed for this purpose. Oil on water surface reflects on these sensors and the transmitted signal can then be measured. The advantage of using these sensors include their ability to obtain remote sensing images irrespective of weather conditions or time of the day. Examples of active sensors utilised for oil spill detection and surveillance are RadarSat-2 and EnviSat SAR systems [5, 8]. On the other hand, Passive sensors illustrated in Fig 2.2 (b), are microwave instruments that measure natural emissions from the sea surface or the oil on the surface. They rely on an external energy source (sun) and requires daylight and cloudless skies to obtain useful images for spill detection. Fig 2.1 shows the wavelength range for data acquisition of the sensors. In this chapter, a review of the most utilised sensors for oil spill detection and analysis are presented according to their strengths and weaknesses.

2.1.1 Visible Light Sensors

Regardless of their many drawbacks, visible sensors operating within the visible range of the electromagnetic (400-700nm) [1] are still very much used to detect and monitor oil spills. They include thermal and visible scanning systems and aerial photography using cameras [27]. Usually mounted on aircrafts (airbone sensors) to conduct aerial surveillance and obtain



(a) Active Sensor (b) Passive Sensor

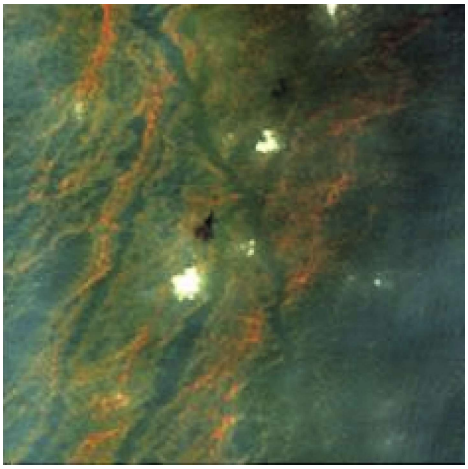
Figure 2.2: Illustrative view of active and passive sensors, respectively [26].

image data. Within the visible region, oil on water surface has higher reflection than water for this reason it is difficult to distinguish between other sea phenomenon or the background [3]. However, advancements in technology have enabled the development of Hyperspectral sensors that acquire data in ten to hundreds of contiguous bands that enable more detailed information of the scene of interest to be obtained. Hyperspectral sensors such as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [28] with high spectral resolutions can be utilised to distinguish between light reflectance on water surface and oil with an added advantage for change detection and material concentration analysis through spectral unmixing.

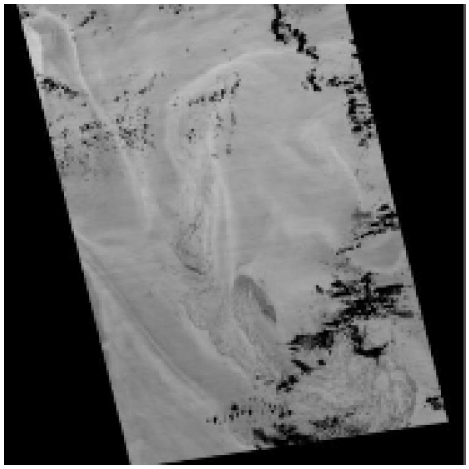
2.1.2 Infrared Sensors

In the long wave region $(8 - 14\mu\text{m})$ of the electromagnetic spectrum, oil absorbs solar radiation and re-emits a portion of it as thermal energy [10]. Oil having higher emissivity infrared than water and being optically thick will emit infrared radiation when heated. Thus, the use of infrared sensors for detecting oil spill therefore is to measure the emissions from spilled oil on water surfaces. Infrared images acquired by these sensors for spill detection shows that thick oil on water surfaces appears hot and its intermediate thickness may become cool, making thin oil slicks not to be visible within this range [1]. Study [24] on the differences between the transition of hot and cold sea surface shows that the minimal distinguishable thickness range in between $10 - 70\mu\text{m}$. A drawback to the use of infrared sensors for oil spill detection is their inability to distinguish water in oil emulsions due to the high thermal conductivity

and the fact that emulsions usually contains about 70% of water with the thermal properties being very closely similar to water. Additionally, previous study [29] on infrared sensors revealed that the spectral formation is absent in images obtained within the long wave region. Spectral information can help in oil spill detection since the spectral information can help to distinguish between the different materials contained in an image of the scene of interest especially to distinguish between the different thickness range of oil in relation to other sea conditions. Images acquired with infrared have also shown that, interference with other sea conditions may be visible, which can complicate the detection process. However, infrared cameras on a good note are relatively cheap and readily available for remote sensing including oil spill detection but cannot measure the thickness of oil spilled. Fig 2.3 (a) illustrates a visible sensor image of the DWH oil spill describing the simplicity of acquiring images with these sensors when the weather is cloudless. However, Fig 2.3 (b) shows the infrared image of the Gulf spill also acquired in 2010 by the infrared sensor ASTER. The dark spots in the middle and lower right part of the image are islands and shoreline of the Gulf while oil spill are not very visible as sheens. This could be due to cloud cover as such sensors are affected by weather conditions.



(a) Visible Sensor Image of the DWH oil spill



(b) Infrared Sensor Image of DWH oil spill

Figure 2.3: Image acquired by visible and infrared sensors, respectively [1].

2.2 Microwave Sensors

Sensors in the microwave region including Radio Detection and Ranging (RADAR) are commonly used sensors for oil spill detection due to their all weather capability. Passive mi-

crowave radiometers can measure emitted energy at millimetre-centimetre wavelengths in the range of $1 - 1000$GHz. When deployed for oil spill detection, they are capable of detecting oil films on water surfaces by radiated signals from space that measures the reflectance of oil. Passive radiometers can estimate the emissivity of oil at 0.8 compared to that of water at 0.4, [30] when utilised for oil spill detection process. The use of passive microwave sensors in spill detection is largely due to its all weather data acquisition capability and relatively moderate spatial resolution images they produce. Radiometers, which are also in the microwave region can be used for oil thickness measurements and estimation by tuning to different frequencies. Within this region, SAR is the most used sensor deployed for oil spill detection.

2.2.1 Radio Detection and Ranging

RADAR is an active microwave sensor operating in the microwave region. Its use include the determination of range, angle or velocity of an object in a scene of interest. Effectively, radar can also be used for detection on the sea surface e.g. a ship disposing its ballasts which is a form of oil spill and environmental pollution. Evidence acquired this way can be used for prosecution of defaulting companies and organisation. Typically, a RADAR system will consist of a transmitter that produces electromagnetic waves in the microwave domain, a transmitting and receiving antenna to transmit and decode signals and the processor that determines objects in the scene of interest. RADAR is widely used for oil spill detection using Synthetic Aperture Radar (SAR) and Side-Looking Aperure Radar (SLAR). SAR is usually deployed on a satellite which means its mostly spaceborne while SLAR is usually mounted on an aircraft (airborne sensing).

2.2.1.1 Synthetic Aperture Radar

SAR is an active microwave sensor that is utilised to obtain 2D images [18]. SAR systems are widely deployed for spill detection due to their all weather conditions and ability for large coverage [5]. These systems, are usually characterised by their frequency band (C, L, X, K, P, etc.), polarisation(VH, VV, HH, HV) that defines the geometry, and the incidence angle that describes the angular relationship between the beam of the RADAR and the object on the ground that is often referred to as the target. A summary of SAR bands and their operational frequencies is given in Table 2.1.

SAR sensors obtain images when capillary waves on the ocean surface reflect on the RADAR energy to produce a bright image often referred to as the "sea clutter" [32]. When

Frequency Bands	Wavelengths	Frequency (GHz)
Ka	0.8-1.1	40-26.5
K	1.1-1.7	26.5-18
Ku	1.7-2.4	18-12.5
X	2.4-3.8	12.5-8
C	3.8-7.5	8-4
S	7.5-15	4-2
L	15-30	2-1
P	30-300	1-0.3

Table 2.1: Frequency bands utilised for SAR systems [31]

oil is spilled on the ocean's surface, the capillary waves becomes dampened and the presence of oil can be seen on SAR images as a dark formation. Unfortunately other sea phenomena including sea weeds, whale and fish sperm, wind calms etc. also appear as dark formation on SAR images creating look-alikes and false alarms. For this reasons, SAR images need to be pre-processed to effectively detect oil spills. Previous studies [5], have included wind condition of the sea as an ancillary data to reduce false alarm rates on dark formations. An important factor in detecting oil spills in SAR images is the wind condition. Low sea state conditions (1-3 on Douglas Scale [33]) will not produce enough sea clutter in the surrounding sea to contrast with the oil and very high sea conditions will scatter the RADAR signals adequately hence blocking detection inside the wave troughs [1, 8]. Despite this drawback, SAR sensors still remain the most active tool for oil spill detection on sea surfaces [34], with the ability to obtain images irrespective of weather condition, cloudiness or fog [32].

In [5, 8, 15–18, 35, 36] SAR images have been utilised for oil spill detection using different approaches and image processing techniques with comparative analysis indicating better resolution in SAR images than images produced by other sensors especially those in the optical region of EMS. Additionally, when analysing the different bands of SAR for oil spill detection, X-band SAR had been proven to provide better data than L and C-band [25]. Other forms of SAR data such as Polarimetry and Interferometry have also been explored to improve spill detection using SAR. In [4], the use of interferometric SAR data to reduce the level of uncertainty in dark formations being oil spills or otherwise. Interferometry allows the acquisition of SAR images under similar geometry using the phase difference in the transmitted signal [37] at two or more positions and computing an additional information of the scene of interest [38]. However, the use of SAR interferometry is only possible in two different configurations; the bistatic mode that involves two SAR sensors either spaceborne or airborne and multistatic

involving more than two SAR sensors.

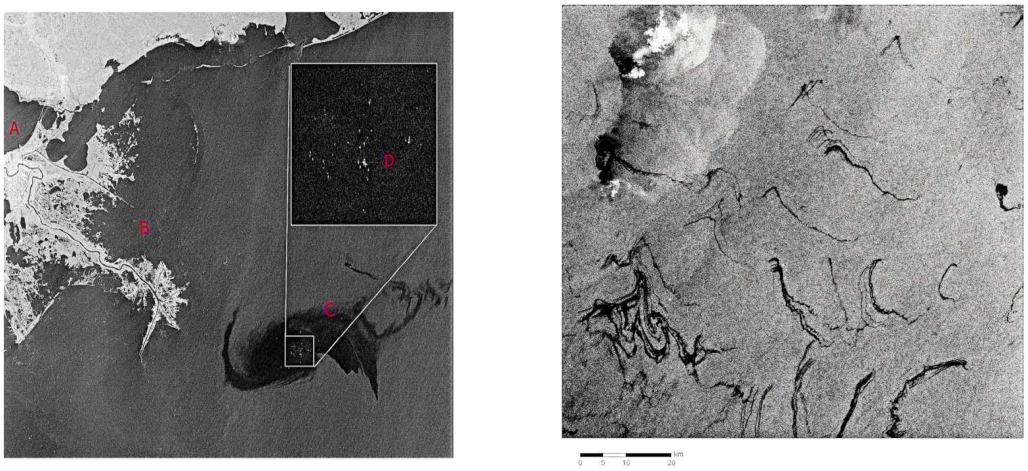
Advanced systems design in SAR technology have enabled the deployment of complex SAR systems with the ability to generate polarised data using SAR polarimetry [39]. Typically SAR systems can operate in different polarisations e.g. VH polarisation means that the transmitted signal has vertical polarisation and the backscattered signal is received by the SAR antenna horizontally [17]. Previous studies [34, 40, 41] have reported that although VV polarisation is the most used polarisation for oil spill detection, the HH polarisation have been proven to provide better results and performance with less oil-look-alikes, and it is well suited for monitoring and ship detection. Ideally, the HH polarisation can be utilised for general ocean monitoring. A fully polarimetric SAR system (e.g. The CosmoSkyMed with 4 constellations, PALSAR etc.) [42] will acquire all compositions of a co-polarized signal (VV & HH) or cross polarized (VH & HV) SAR signals using a 2×2 scattering matrix:

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} , \tag{2.1}$$

In (2.1) the matrix $\mathbf{S}$ describes a fully polarimetric scattering matrix [43], where the subscripts defines transmitted and received signals, respectively. The use of polarimetric SAR data have proven to provide useful information when the full polarimetric channels (HH, HV, VV, VH) are utilised compared to using a single channel (VV for e.g.). Furthermore a general consensus among researchers [41, 43, 44] in this field is that the use of fully polarimetric SAR data will provide adequate information that can improve detection of oil spills. In [41] the use of polarimetric data is utilised for oil slick observation on sea surfaces. The approach extracts polarimetric features from SAR images following established electromagnetic models that enables an efficient oil slick observation in an unsurpervised approach. In addition, extensive information with regards the damping of oil on sea surface can be estimated to enable efficiency in environmental remediation [41].

On the other hand, Side-Looking Synthetic Aperture Radar (SLAR) is another configuration of SAR utilised for oil spill monitoring services allowing for large area coverage irrespective of weather conditions. SLAR is relatively cheap when compared to SAR. Firstly because it is mounted on an aircraft and secondly it utilises a horizontal antenna to obtain images along the flight path while normally a SAR utilises the forward motion of the aircraft to enhance resolution [24]. A drawback to the use of SLAR for oil spill monitoring is that this system has lower range and resolution when compared to SAR and produces artifacts on the images

obtained. However, while SAR requires a substantial electronic image refinement due to its inherent speckles to enhance its resolution, this may not usually be the case with SLAR since the resolution of images obtained is dependent on range. In [45], the use of SLAR is utilised for oil spill monitoring by exploring two approaches. Firstly, regions with noise in the image as a result of aircraft movement are detected and marked to improve detection. Secondly a segmentation step backed by a saliency map is employed to detect the regions in the image that are likely to be oil spills. It can be deduced from the approach that speckles created as a result of backscattering in SAR or artifacts created as a result of flight movement can affect the performance of using these systems for oil spill monitoring and detection. The segmentation step introduced in [45] can be improved using machine learning approaches as proposed in [46] where the use of deep selectional autoencoders is utilised by evalauting different layers of the network and employing a codification technique that maintains a selection of image pixels using a groundtruth of selected pixels over a range of [0 1]. The approach is robust to artifact noise and independent to weather conditions. In Fig 2.4. a SAR image of the Gulf of Mexico oil spill is presented. The zoomed square on Fig 2.4 (a) shows the rescue shapes in whites dots around the location of the spill. The black spots on the image represents the spread of oil around the gulf. In Fig 2.4 (b), an infrared image of the same location is presented. However, the effect of cloud cover makes it difficult to distinguish the black spots as some are likely to be look alikes.



(a) Visible Sensor Image of the DWH oil spill

(b) Infrared Sensor Image of DWH oil spill

Figure 2.4: (a) Deep Water Horizon (DWH) oil spill captured by Radarsat and (b) Leaking oil platforms in Caspian sea acquired by Radarsat-2 [24]

2.2.1.2 Other Sensors Used for Oil Spill Remote Sensing

Other sensors such as ultraviolet (UV) and laser fluorosensors (LF) have also been utilised for the detection of oil spills. In the wavelength range of 0-38nm, ultraviolet sensors are useful to determine the relative thickness of oil spills due to the high reflectivity of thin oil layers [47]. However, the drawback to the use of this sensor is that only oil thickness below 10 microns can be detected. Additionally, sea materials such as sea weeds, wind sheen or sun glint cannot be distinguished from oil spills, hence raising false alarm rate and look-alike. Its primary use to spill detection can be limited to being an auxiliary data that are usually overlayed on infrared images to create oil spill thickness maps [3].

Laser Fluorosensors on the other hand, have the potential to discriminate between the different types of oil spilled. Oil naturally, contains compounds that absorbs ultraviolet light that enabling electrical excitation usually removed through the emission of fluorescence in the visible region of the EMS that strongly indicates the presence of oil. The ability of this sensor to distinguish between the types of oil is because different oil types have unique flourescent intensities and properties [24]. When the aim of detection is to distinguish between different the thickness of spilled oil, LF sensors are the best with this having the capability to distinguish between heavy,medium and light oil types[48]. LF sensors utilises the "gating" technique; that extends the seperation of oil from other sea phenomena enabling a further detection of the water columns. In Fig 2.5. the fluorescence spectra of respective light crude oils after laser activation at 308nm is shown with detailed information on this discussed in [25].

2.3 Methods for Image Fusion

Image Fusion enables the combination of image information acquired by multiple sensors, or from a single sensor acquiring images at different times or from different angles, with an aim of acquiring a more complete information in a single image than it is in the individual images [49]. In remote sensing images for example, a variety of remotely sensed data like hyperspectral, multispectral and SAR imagery are captured by different earth observation sensors on a daily basis and further processed for many applications such as weather prediction, mapping, or for environmental monitoring like oil spill detection etc. For most of the applications mentioned, the analysis of only one source of imagery may not be sufficient and hence the need for complimentary information from other sensors to better understand the observed

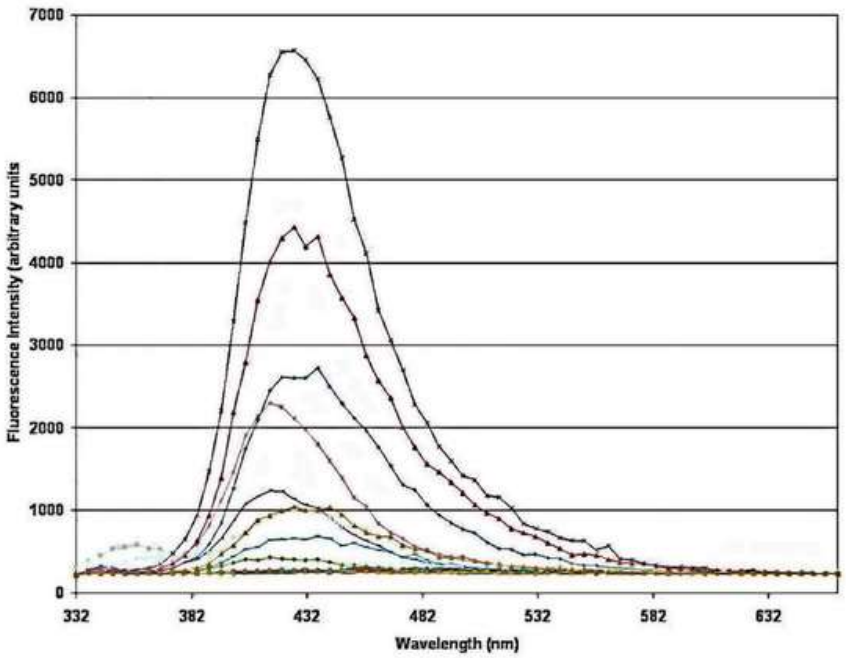


Figure 2.5: Flourescence spectra of some light crude oil obtained in the laboratory after activation of 308nm [24]

scene, the earth surface and to be able to make informed decisions that are reliable. Image fusion provides us with such opportunity to combine the complimentary and supplementary information contained in images acquired by different sensors, at different times and from different angles.

Image Fusion can be implemented at three levels. The pixel level image fusion is the lowest level of fusing images. In this approach, the fused image is constructed by obtaining the pixel values of different images using some algorithms, e.g. the Simple Averaging (SA) under strict registration conditions [50]. This type of fusion has the advantage of minimum loss of information during the fusion process, yet it is computationally expensive since the process involves all pixels which will result in very slow processing speed. The feature level fusion takes into consideration features extracted from different images to be fused such as edges, texture, and shape etc, using same or different preprocessing step, the extracted features are combined to form an ideal feature set which is used to form the fusion and for further analysis like classification. The decision level fusion is a high level fusion [51]. In this approach, the images to be fused are processed seperately and the information refined and combined, while the differences in information is resolved using some decision rules. Two types of decision fusion have been identified in literature. The first is to classify the image using

the different classifiers for a single source image, or from two different sources for example an optical image and a SAR image, classified seperately and then combined to produce a refined classification map. Various methods have been used to achive decision fusion such as the statistical methods, logical resoning methods, fuzzy set theory and neural network. In ideal situations, the decision fusion has the advantage of fault tolerance and it is also good in real time applications. However the preprocessing is expensive, it also has the smallest data quantity and has the highest ability of anti-interference compared to the other levels of fusion. Also, with decison fusion the probability and reality of fused results are high and the performance of multi sensor system is improved as illustrated in [50, 52].

In addition to the levels of image fusion described above, individual methods and approaches for fusing different image modalities have been reported in the literature. Pansharpening methods for example, where the aim is to combine panchromatic images of high spatial resolution with a multispectral image of high spectral resolution with an aim obtaining an image with high spatial resolution of the former and high spectral resolution of the latter. In this thesis, a detected chapter reviews the pansharpening methods in relation to their performances on spectral unmixing [27, 53] of oil spills. In this chapter the focus will be on the fusion methods utilised for image fusion in this thesis. Firstly, an image is defined as

$$\mathbf{I}(x, y), \tag{2.2}$$

a two dimensional function with the variables x and y being the spatial coordinates and the amplitude I is the intensity of the image at the coordinates (x, y). For a high spectral image, the image is slightly defined differently by adding the spectral parameter λ. Thus a high spectral resolution image is given as

$$\mathbf{I}(x, y, \lambda), \tag{2.3}$$

where the λ defines the spectral extent of the image. the variables (x, y) are as defined above and the intensity I, and the total number of measurement in I is given as $.x \times y \times \lambda$.

2.3.1 Simple Average

Considering the fact that image regions in focus are of higher pixel intensity than other regions of the image, the simple average method obtains an output image putting all the regions of the image in focus, so that the resultant fused image is obtained by taking the

average intensity value of the corresponding pixels in the input images. If $\mathbf{I}_1$ and $\mathbf{I}_2$ are two input images to be fused, the resultant fused image $\widehat{\mathbf{F}}$ is given by

$$\widehat{\mathbf{F}} = \frac{(\mathbf{I}_1 + \mathbf{I}_2)}{2} \ , \tag{2.4}$$

where $\widehat{\mathbf{F}}$ is the resultant fused image and $\mathbf{I}_1$ and $\mathbf{I}_2$ are the input images, respectively. In addition to the simple average fusion method, there is the select maximum method; here the algorithm selects the image focus regions from each of the input images by choosing the highest value for each of the input pixels which results in a highly focused output. The fused image is obtained by selecting the maximum intensity of the corresponding pixels from the input images such that:

$$\widehat{\mathbf{F}} = \sum_{i=0}^{x} \sum_{j=0}^{y} max(\mathbf{I}_1 \mathbf{I}_2) \ . \tag{2.5}$$

Similarly, the select minimum method takes the minimum intensity value of the corresponding pixels of the input images to obtain the fused image such that:

$$\widehat{\mathbf{F}} = \sum_{i=0}^{x} \sum_{j=0}^{y} min(\mathbf{I}_1 \mathbf{I}_2) \ . \tag{2.6}$$

Combined with the multiplicative method, these methods described above are regarded to as the arithmetic and combinations based image fusion methods [50] and are often referred to as the simplest form of image fusion methods.

2.3.2 Wavelet Transform Image Fusion Methods

The wavelets transform methods allow images to be processed at multiple resolutions, thus making it an efficient method of image fusion. Additionally, the wavelets methods provides insight into the image's spatial and frequency characteristics unlike the fourier transform methods, where only the frequency attribute of the image are revealed. As a background information, lets consider an image $I(x, y)$ whose size is $m \times n$, its forward discrete transform $T(u, v)$ can be expressed in terms of general relation as

$$T(u, v, ...) = \sum_{x,y} I(x, y) g_{u,v,...}(x, y) \ , \tag{2.7}$$

where x and y are the spatial coordinates and $u, v, ..$ are transform domain variables. Additionally, given $T(u, v, ...)$, $I(x, y)$ can be obtained by a generalised inverse discrete transform

such that:

$$\mathbf{I}(x,y) = \sum_{u,v,\ldots} T(u,v,\ldots) h_{u,v,\ldots}(x,y) \ , \tag{2.8}$$

where $g_{u,v,\ldots}$ in (2.7) and $h_{u,v,\ldots}$ in (2.8) above, are the forward and inverse transformation kernels, that determine the nature and computational complexity of the transform pair.

In image fusion, the wavelets advantage of multi resolution is explored. Images are decomposed into different coefficients, and the coefficients of different images are then combined to form new coefficients [50]. This is the basic idea of the wavelets methods. This method have been used in image fusion applications including medical image fusion and remote sensing image fusion as it improves spatial resolution of the fused image while also preserving the colour appearance for interpretation and further analysis [54]. A tutorial on wavelet based image fusion is done in [55]. Additionally, a review of this method is done in [56].

In [57], an image fusion approach based on wavelet transform is proposed. The algorithm first computes the wavelet transforms of the images containing the low-high bands, and considering that the larger absolute transform coefficients corresponds to sharper brightness changes of the image which are mostly the salient features [58]. The algorithm combines feature selection using an area based selection, since most useful features of the image are greater than a pixel, thus a pixel-pixel maximum selection is not efficient. In oil spill detection, the wavelets methods have been proposed by [59] to fuse different bands of SAR, precisely the S and X bands. The idea is to use the fusion technique to see if that improves the discrimination of different type of oil on water. The approach is as with every wavelets transform the decomposition of the image signal into different sub-components at different resolutions. However, this algorithm considered the Daubechies wavelets in conjuction with a choose maximum selection rule. Additionally, the algorithm uses segmented images which means the dark spots are segmented after fusion of the images; unfortunately, the results did not show improvement in the descrimination of the different types of oil as a result of fusing different bands of SAR using the wavelets system, the approach suggested a more quantitative analysis to measure the level of improvement. In another study, [60] proposed a method based on inversion of wavelet coefficient for oil spill detection in radar images. This approach is similar to the approach in [59] but differs in the sense that in the inverse approach, the wavelet transformation is applied to inverted images of SAR in addition to the frame differencing technique. As with all multi resolution methods, the wavelet methods have the advantages of

temporal coherence, spectral consistency and robustness to aliasing under proper conditions [53]. on the other hand, its disadvantage is that implementation is complex due to the need for spatial filter designs and it is also computationally expensive when compared to other methods. Additionally, while this method enhances the spatial resolution by fusing images, the decomposition scale needs be selected sensibly since it can affect the fusion output. Large scaling causes the colour contents to be lost which makes the approach un-suitable for high spectral images and small scaling will create mosaics on the fused image [61].

2.3.3 Bayesian Fusion Methods

The Bayesian fusion methods are probabilistic approaches for fusing image information from same or different image sensors with same or different imaging characteristics. It is ideal for image fusion since it only needs a single measure to describe uncertainty unlike the other theories like the Dempster-Shafer theory [62]. The fusion of multi modal images eg. MS or PAN can coveniently be formulated within the Bayesian framework as described in [53]. As a background, lets consider a simple Bayesian approach to an image data fusion problem. Let D_i be the set of observed images, and $i = 1, ..., D_n$, $\mathbf{X}$ is the unknown object of the scene represented in the image. Therefore, to obtain the fused image $\widehat{\mathbf{F}}$, the following steps are evaluated

1. The model to relate the observed data D_i to the unknown scene X which is referred to as the forward model

2. Account for the uncertainty associated with the model in 1 above as well as the instability of the image measurement sensor, this is known as the likelihood of the parameter X when D_i is observed, denoted as $P(D_i \mid X)$. However assigning the likelihood $P(D_i \mid X)$, requires a deterministic relation between D_i and X, that accounts for the physical process of data acquisition and the probability modelling that accounts for the uncertainty referred to as noise (ϵ). From each individual likelihood function of $P(D_i \mid X)$, the function $P(D_1, ..., D_n \mid X)$ can be defined if the data is assumed to have been acquired independently and there is no correlation between the different observing sensors.

3. Express the prior knowledge of X by assigning it a prior probability function $P(X)$. This is an important step especially when the likelihood function does not contain much information and its not unimodal. Once appropriate models of the the likelihood function

$P(D_1, ..., D_n \mid X)$ and the prior probability $P(X)$ are assigned, the next step is to combine them through the Baye's rule in order to obtain the posterior. Thus, the posterior can be formulated as $P(X \mid D_1, ..., D_n)$ is given as:

$$P(X \mid D_1, ..., D_n) = \frac{P(D_1, ..., D_n \mid X)P(X)}{P(D_1, ..., D_n)}\ , \tag{2.9}$$

where $P(D_1, ..., D_n)$ is a normalising factor, $P(X \mid D_1, ..., D_n)$ is the posterior function, $P(D_1, ..., D_n \mid X)$ is the likelihood function and $P(X)$ is the prior probability function of X. Once this is done, it is possible to infer any knowledge about X. A review of Bayesian approaches to image fusion can be found in [53].

In [21], a Bayesian fusion technique for remotely sensed multi band images was proposed. In this approach, the observed images are related to the high spectral and high spatial resolution image to be recovered through physical degradations such as the spatial and spectral blurring or subsampling as a result of the image sensor characteristics. An appropriate prior distribution exploiting the geometrical considerations is introduced since the problem is formulated within a Bayesian estimation framework. A Markov Chain Monte Carlo (MCMC) [63] algorithm is further designed to generate samples so as to compute the Bayesian estimator for the scene of interest. This approach is different from other Bayesian approaches proposed, for instance [64] proposed a Bayesian fusion method to fuse multi modal aerial images. The idea is to first propagate information from high resolution images into other low resolution modalities, while allowing the images to have different spectral channels. This is similar to the approach of wavelets were image coefficients are created from different coefficients as illustrated in [50]. The relationship between the images is non-deterministic and non-linear; for this reason a Gaussian process framework is used to define a stochastic prior over the estimated images. Lastly, its covariance function is computed to replicate the local structure of the high resolution image so as to allow the model collect a high resolution estimate from a low resolution channel to obtain the fused image $\widehat{\mathbf{F}}$.

Similarly, in [65], a Bayesian Linear Estimator (BLE) is proposed to fuse MS and PAN images. The BLE is applied to the observation models between images with different spatial and spectral characteristics. The method only estimates the mean vector and the covariance matrix of the high resolution MS image, without assuming a joint distribution between the PAN image and the low resolution MS image. The method also shows efficiency in enhancing the spatial resolutions of several principal components of the MS image unlike in the PCA

method where only the first principal component is enhanced. Additionally, Bayesian methods different framework can be formulated which depends on the strength or limitation of the prior which gives this method an advantage over the other fusion methods.

The advantages of the Bayesian fusion methods include their ability to integrate a prior knowledge of a scene of interest or object of interest into the fusion framework. Additionally, Bayesian inference can handle the fusion of multiple sensors with respect to the system state. This is achievable when the optimal stochastic belief of the system state is incorporated as the prior. Furthermore, Bayesian fusion being data driven; the use of prior information based on the data can be seen as an effective approach in cases where the observed features of an image (in case of fusing images) for the fused data are large enough to provide objective estimates of the target.

Where the observations (images obtained for e.g.) are time dependent variant, the prior will then change with any new observation (new image obtained) such that the Bayesian framework is designed recursively, making the prior time dependent and on the previous observation (image) obtained. In a time-independent observation, the vector $\mathbf{z}^{t-1} = \{z_{t-1}, \ldots, z_0\}$ defines the obtained observations and its prior $P(X \mid \mathbf{z}^{t-1})$ at time $t-1$. If the system is time independent, then (2.10) is valid.

$$P(X) = P(X \mid \mathbf{z}^{t-1}) = P(X_t \mid \mathbf{z}^{t-1}) = P(X_{t-1} \mid \mathbf{z}^{t-1}) \, . \tag{2.10}$$

Bayesian inference with time dependency, the difference in this case with the time independent approach is that (2.10) is not satisfied, as illustrated in (2.11). However, the prior can still be inferred by utilising $P(X_{t-1} \mid \mathbf{z}^{t-1})$ posterior in (2.10) with a new time dependent model $P(X_t \mid X_{t-1})$. Hence, the satisfaction of difference between the two system models can be given as

$$P(X_t \mid \mathbf{z}^{t-1}) \neq P(X_{t-1} \mid \mathbf{z}^{t-1}) \, , \tag{2.11}$$

$$P(X_t \mid \mathbf{z}^{t-1}) = \sum_{xt-1} P(X_t \mid x_{t-1}) \cdot P(x_{t-1} \mid \mathbf{z}^{t-1}) \, , \tag{2.12}$$

Summarily, a Bayesian inference in a time dependent model is given as:

$$P(X_t \mid \mathbf{z}^t) = \frac{1}{P(z_t)} \cdot P(z_t \mid X_t) \cdot P(X_t \mid \mathbf{z}^{t-1}) \, , \tag{2.13}$$

$$P(X_t \mid \mathbf{z}^t) = \frac{1}{P(z_t)} \cdot P(z_t \mid X_t) \cdot \sum_{x_t - 1} P(X_t \mid x_{t-1}) \cdot P(x_{t-1} \mid \mathbf{z}^{t-1}) \; . \tag{2.14}$$

2.3.4 Gaussian Process Regression

The use of prior information in a Bayesian model framework enables us to make more informed decision about the model and allows the selection of an appropriate function for the model. Additionally, the availability of more data can further enhance the chances of a more reliable prediction of the estimate and in the case of image fusion, the more reliable the prior information we have about the observations (images to be fused) the better the prediction of the estimate, the fused image. Bayesian non-parametric models are well suited for cases of pattern recognition, data learning and extrapolation with large datasets [66]. Non-parametic Bayesian models include Gaussian Processes (GP) that have been explored for smoothening and filtering [67].

In this thesis, the aim is to combine images from different modalities obtained by different imaging sensors to reconstruct an image with higher information than the individual images. To achieve this, there is need to automatically extract the rich information contained in the individual images. A GP enables us to explore this information through learning that utilises covariance kernels [*]. The GP in an image fusion approach can be utilised as a distribution over functions with some prior support for a given function to generate the fused image ($\widehat{\mathbf{F}}$) [68]. The distribution of the prior function and its properties (smoothness of the function and its periodicity) can be determined when using an appropriate covariance function [69]. The ability of GP to perform inference through the use of kernel learning makes it flexible and interactive. However, the use of this model is limited by computational complexity that requires $\mathcal{O}(N^3)$ for processing and an additional $\mathcal{O}(N^2)$ for storage for any dataset with N number of observations, this has limited the performance of GP for large datasets. Furthermore, since the processing requires learning that is dependent on the covariance kernel, there is need for the design of new covariance kernels to suit different models and problems. The design of custom kernels is not an easy task since it encodes the assumption of the underlying function [64]. In this background review, a brief on parametric modelling is given with the objective that deviates from weight space regression into the function space and the non-parametric modelling,both serving as ingredient of the GP regression model.

Consider a set of observations for training in a regression model, $\mathbf{y} = (y(x_1), \ldots, y(x_n))^T$,

[*]covariance, covariance function and kernels are used interchangeably in this thesis

next the set of observations are evaluated over some known inputs $X = (x_1, \ldots, x_n)^T$. In essence the model needs to predicts a new output $y(x_*)$ at a new input location x_*. As an example, the training dataset can be a series of images obtained at different times and we may want to predict a new image at another time. Additionally the images (inputs) can be multi-dimensional i.e. $x \in \mathbb{R}^M$. A conventional regression approach in this case is to assume the frame of the function utilised to generate the data, and further learn the parameters that characterise this function so that errors from the prediction step can be minimised. Based on the training dataset, two approaches can be used for this function, the first is to assume a linear approach if the training set has a linear inclination and represent the function as $f(x, \mathbf{w}) = \mathbf{w}^T x$ where $\mathbf{w}$ describes the vector of weights on which the function is parametrised. The second approach can be utilised if the function tends to be a complex representation of the dataset. In this case, a linear model with an added non-linear vector of features can be utilised for regression. i.e. $f(x, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\eta}(x)$, where $\boldsymbol{\eta} = (\eta_1(x), \eta_2(x), \ldots, \eta_k(x))^T$. The error of the prediction between the model and the actual system can be calculated by taking the squared differences of the predicted and the actual as described in [68].

The second approach is to assume a noisy training dataset such that the function can be described as

$$y(x) = f(x, \mathbf{w}) + \epsilon(x), \tag{2.15}$$

where ϵ is a random variable that defines the noise parameter and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian distribution with a zero mean and standard deviation σ. When the dataset is assumed to be noisy, the weights $\mathbf{w}$ can be evaluated using the maximum likelihood $P(\mathbf{y} \mid X, \mathbf{w})$, such that

$$P(y \mid X, \mathbf{w}) = \mathcal{N}(y(x); f(x, \mathbf{w}), \sigma^2) \tag{2.16}$$

$$P(\mathbf{y} \mid X, \mathbf{w}) = \prod_{i=1}^{n} \mathcal{N}(y(x_i); f(x_i, \mathbf{w}), \sigma^2), \tag{2.17}$$

where n defines the number of observations or datapoints in $\mathbf{y}$. It is important to mention that both approaches to the regression problem are susceptible to over-fitting. However, solutions to the over-fitting problem can be to introduce an additional complexity penalty term to the maximum likelihood as illustrated in [66]. Hence, a Gaussian Process model can

be introduced thus:

A GP therefore can be described as a stochastic process, that defines the collection of random variables [70]. The GP wholly defined by its mean $m(\mathbf{x})$ and its covariance $k(\mathbf{x}, \mathbf{x}')$, also wriiten as $\mathbf{K}$. In another defination, [71] defines it as a "collection of random variables, any finite number of which have a joint Gaussian distribution". The mean and covariance functions can further be described as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{2.18}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))^T]. \tag{2.19}$$

The GP can then be described as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \ . \tag{2.20}$$

Random variables in this case refer to the values of the function $f(\mathbf{x})$. Hence, the definition of GP as a collection of random variables refers to the value of the function $f(\mathbf{x})$ at the location of $\mathbf{x}$ [69]. The mean function $m(\mathbf{x})$ can assume a zero value since the GP can be tuned to model the mean swiftly [72], while the covariance kernel is determined by the hyperparameters of the model. The definition also means the consistency condition otherwise referred to as the marginalisation attribute i.e. If the GP state $(y_1, y_2) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, then it must satisfy $y_1 \sim \mathcal{N}(\mu_i, \Sigma_{ij})$, where Σ_{ij} defines the related submatrix of Σ, and $\Sigma_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, where k is a positive semi-definite kernel function [72].

Considering that the covariance $k(x, x')$ operates on finite pair of input, for a pair of inputs x and x' in any random space $X \in \hat{X}$, i.e. The collection of function values $f(x)$ evaluated at the input points will have a joint Gaussian distribution such that

$$[f(x_1), \cdots, f(x_N)] \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}), \tag{2.21}$$

where

$$\boldsymbol{\mu}_i = \mathbb{E}[f(x_i)] \tag{2.22}$$

$$\mathbf{K}_{ij} = k(x_i, x_j) \ . \tag{2.23}$$

Additionally, the GP can be utilised for a supervised learning problem [64]. For example, consider a training dataset $\boldsymbol{\upsilon}$, with N set of input, output variables $\{\mathbf{X}, \mathbf{y}\}$, where the variable $\mathbf{X} = \{x_i \in \mathbb{R}^D\}_{i=1}^N$ define the locations of inputs in the function space and $\mathbf{y} = \{y_i \in \mathbb{R}\}_{i=1}^N$ [73]. The objective of the GP therefore, is to predict the new outputs $\mathbf{y}^*$ observed over a set of query points $\{x^*\} = \{\mathbf{x}_i^* \in \mathbb{R}^D\}_i^M$, where M defines the set of query inputs. According to [66, 73], the observed output y is suppose to be the outcome of a structural inclination of the functions $\mu_0, f(x)$ and ϵ. Thus, the output y_i is given as:

$$y_i = \mu_0 + f(x) + \epsilon, \tag{2.24}$$

where the noise of the individual output is described as an un-constrained Gaussian distribution. The noise term is to factor in process noise or noise resulting from the observation error. Furthermore, the observed and queried outputs $\mathbf{y}$ and $\mathbf{y}_*$ are related by assuming that the functions of the model $\mathbf{f}$ and $\mathbf{f}_*$ are evaluated on a common GP. The covariance function, $k(x_i, x_j)$ (subsequently discussed in more detailed) is then utilised to model the pairwise connection between two outputs evaluated as a function $f(x_i)$ and $f(x_j)$ of their location in the input data (input space) $(x_i, x_j \in \mathbb{R}^D)$. The covariance allows for a joint prior distribution of the outputs $\mathbf{y}$ over the function $\mathbf{f}_*$ and is given as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0\,, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{x}^*) \\ K(\mathbf{x}^*, \mathbf{X}) & K(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}\right). \tag{2.25}$$

To conduct inference with GP, the prior is logically conditioned on the values of the observed outputs y so as to obtain an objective predictive distribution of the model function f_* in closed for such that:

$$p(\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*) \sim \mathcal{N}(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star), \tag{2.26}$$

where

$$\boldsymbol{\mu}_\star = k(\mathbf{X}^*, \mathbf{X}) K_{\mathrm{X}}^{-1} \mathbf{y} \tag{2.27}$$

$$\boldsymbol{\Sigma}_\star = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X}) K_{\mathrm{X}}^{-1} k(\mathbf{X}, \mathbf{X}^*)\,. \tag{2.28}$$

The predictive distribution equations above can be regarded as a Bayesian formulation to conduct inference and can further be utilised for machine learning since the description of

the Gaussian prior in (2.22) can appropriately be used to evaluate the marginal likelihood of the observed output $\mathbf{y}$ [69] as will be shown in the hyperparameter learning sub-section.

2.3.4.1 Covariance Function

The covariance function or kernel [66] is the main ingredient of the GP predictor [71], since it encodes our assumption of the function we intend to model. Hence, the covariance plays a fundamental role in any GP regression task and help to define the inter-dependence between function values [74]. Typically, the kernel is specified by its hyperparameters $\boldsymbol{\theta}$, that characterize the model. A classical assumption of near points $\mathbf{x}$ and $\mathbf{x}'$ is that they are likely to produce a similar output $\mathbf{y}$. Subsequently, training points close to a test point will under normal circumstance provide sufficient information of the prediction at that point. In a GP setting, the kernel or covariance function should be robust to describe these attribute of similarity of points or their closeness. In evaluating the covariance function of a GP, a standard requirement is that the covariance function must always give a positive semi-definite covariance matrix for any order of inputs, hyperparameters and query points. Additionally, the kernel $\mathbf{K}$ must be shown to be positive definite so that for any point $x_i \in R^d$ and any point $a_i \in Rc$, $i = 1 \cdots, N$ it satisfies the following

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j k(x_i, x_j) \geq 0. \tag{2.29}$$

The above condition can be satisfied conventionally by constructing the kernel from the collection of well defined covariance functions available in the literature. For example the Squared Exponential kernel, or the mater'n class of covariances [8]. It is also shown in [75], that new valid covariance functions can be developed through addition, product or integral of valid covariances [64] as described in the following equations, assuming $k_1(\mathbf{x}, \mathbf{x}')$ and $k_2(\mathbf{x}, \mathbf{x}')$ are valid covariance functions, then

$$k(\mathbf{x}, \mathbf{x}') = q(k(\mathbf{x}, \mathbf{x}'))\ , \tag{2.30}$$

$$k(\mathbf{x}, \mathbf{x}') = \exp(k(\mathbf{x}, \mathbf{x}'))\ , \tag{2.31}$$

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}')\ , \tag{2.32}$$

$$k(\mathbf{x}, \mathbf{x}') = \iota(x)k_1(\mathbf{x}, \mathbf{x}')\iota(x')\ , \tag{2.33}$$

$$k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x_a}, \mathbf{x'_a}) + k(\mathbf{x_b}, \mathbf{x'_b})\ , \tag{2.34}$$

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x_a}, \mathbf{x'_a})k_2(\mathbf{x_b}, \mathbf{x'_b})\ , \tag{2.35}$$

where $\iota(x)$ in (2.33) defines any function, q in (2.30) is a polynomial with non-negative coefficients [66], k_1 and k_2 are valid covariances in their respective spaces, accordingly. When selecting the covariance function for a model, a careful consideration should be given to other factors that relate to the observation (data), e.g. the stationarity and isotropy qualities.

The assumption of input points $\mathbf{x} - \mathbf{x}'$ in a stationary setting is that of displacement of the points rather than the actual position of inputs, which means the inputs are invariant to translations. This is suitable in situations where the structure of the observation is static . Where this is not the case, more appropriate and complex models need to be designed that are non-stationary. An example of a stationary covariance kernel is the popularly used Squared Exponential (SE) which is an isotropic kernel also referred to as distance kernel [66]. A covariance is termed isotropic when the inputs are described by the function $\parallel \mathbf{x} - \mathbf{x}' \parallel$. The SE covariance is given as

$$k_{se}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\Big(-\frac{1}{2\ell^2}\ (\mathbf{x}, \mathbf{x}')^2 \Big). \tag{2.36}$$

Here the hyperparameters of the model are ℓ and σ_f^2 where ℓ controls the sensitivity of the function and it is the function's characteristic horizontal length scale while σ_f^2 controls the variation of the function vertically [72]. In isotropic form the SE kernel can be given as

$$k_{se}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\Big[-\frac{1}{2}\Big(\frac{|\mathbf{x}, \mathbf{x}'|}{\lambda} \Big)^2 \Big], \tag{2.37}$$

where the hyperparameters are given as $\boldsymbol{\theta} = \{\lambda, \sigma_f^2, \sigma_n^2\}$, and describes the variance of the signal, length scale and noise, respectively. The optimisation of the hyperparameters is discussed in the next sub-section. On the other hand, the anisotropic SE kernel is in the form

$$k_{se}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}\left(|\mathbf{x}, \mathbf{x}'|^\top \mathbf{\Omega}^{-1} (\mathbf{x}, \mathbf{x}') \right) \right], \tag{2.38}$$

where $\mathbf{\Omega}$ defines a positive definite matrix of length scales and replaces the length scale in the isotropic form. This, can scale upto to $\frac{D(D+1)}{2}$, depending on the parametrisation [73] while other hyperparameters are retained as those in the isotropic kernel.

2.3.4.2 Hyperparameter Learning

Learning the hyperparameters of a GP model is usually done by optimising the marginal likelihood [76]. In the Bayesian framework for a GP formulation, the prior distribution $\mathbf{y} \sim \mathcal{N}(0, [k((\mathbf{x}, \mathbf{x}') + \sigma_n^2 \mathbf{I}])$ can be utilised for an optimal solution of the hyperparameters $\boldsymbol{\theta}$ from a given dataset $\{\mathbf{X}, \mathbf{y}\}$ of input and output pairs. The approach that optimises the marginal likelihood is the most used in a GP regression problem [66, 69]. Other approaches could be sampling like the Genetic Algorithm (GA). However, sampling methods are not computationally efficient and it is difficult to select what sampling method to use for hyperparameter learning. When utilising the marginal likelihood approach for optimisation, it is important to note and consider that the prior function is responsive only to the structure of the kernel $k(\mathbf{x}, \mathbf{x}')$, where the log of this prior is given by

$$log(P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta})) = -\frac{1}{2}\mathbf{y}^\top [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} - \frac{1}{2} log|\mathbf{K} + \sigma_n^2 \mathbf{I}| - \frac{n}{2} log\pi \,, \tag{2.39}$$

where $\mathbf{K} = k(\mathbf{X}, \mathbf{X}' \mid \boldsymbol{\theta})$ and $|.|$ specifies the determinant of a matrix. The constituent of (2.39) play important roles in the optimisation process. First term of the equation finds data fit, the second term is the model complexity term and the third term is a constant that ensures the marginal likelihood is robust to over-fitting.

GP models have been used for image processing applications including for image fusion [64], image dehazing [77], image denoising [78] and for improving the resolutions of images [74]. The distinguishing factor in most of these approaches is in the design of new covariance functions suitable in addressing the individual problems and in improving the computational complexity inherent in the learning process especially when dealing with large datasets. While established covariances in the literature still remain relevant for most regression problems, it is not always the case for all approaches. In image fusion for example, the problem is to combine images acquired through different processes and modalities. The use of established covariance kernels may not adequately model the differences in sensor char-

acteristics and so on. Hence there is need for the development of new covariance functions that will address the inter-modality issue. Another lingering problem with images is that of change of support, situation where a pixel is of different spectral channel with the neighbouring pixel and when dealing with multi-band images. While GP has proven worthy in modelling inter-dependency, a lot of work needs to be done in terms of learning the parameters of such models. This thesis, aims to solve the problem of fusing multi-modal images with GP.

2.4 Image Segmentation

Segmentation is the process of partitioning the image into regions or multiple segments so as to change the representation of the image into something that is more meaningful and easier to analyse [2, 79]. In oil spill detection using SAR images, segmentation is a key step in the detection process chain. It allows the dark formations on the SAR image to be partitioned from the rest of the image. The dark areas as stated earlier are potential candidates for oil spill. This is illustrated in Figure 2.6. Segmentation methods generally can be grouped into local and global segmentation techniques.

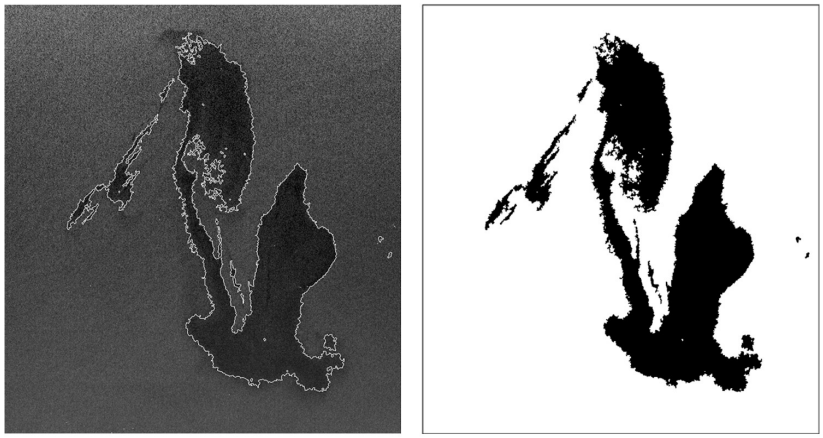


Figure 2.6: Segmentation of Dark Formation on a SAR Image [80]

2.4.1 Thresholding

Thresholding technique is regarded as the simplest as it is the most commonly used method of image segmentation [81]. The idea of this technique is the selection of a threshold value. Here, a value t is selected either manually by trying a range of values of t and seeing which

value gives the best result of the segmentation process or the selection of t can be done automatically. For example a pixel located at a lattice position (i, j) of an image with a grayscale value fij is allocated to category 1 if

$$fij \leq t. \tag{2.40}$$

or 0 if the expression is otherwise. Here 1 can denote object and 0 background. Oil spills are characterised by low backscattering levels hence suggesting the use of thresholding for dark spot segmentation [5]. An adaptive threshold technique has been applied in [15, 16] to segment oil spill in RADARSAT and Envisat SAR images. The idea is to create a pyramid of the image by averaging the pixels of the original image, and the next level of pyramid is created with half the pixel size of the original image and so on. An adaptive thresholding is then applied to each level of the pyramid and combined with a clustering step so as to have a better separation of the spill from the background.

In [2], global thresholding was applied to the SAR image to segment the dark areas. The otsu method is used to compute the threshold value by converting the intensity image into a binary image, and then taking the compliment of the image such that ones are now zeros and vice versa. The pixels that correspond to the digital number values less than the threshold value are taken as pixels of the dark regions.

In [82], hysteresis thresholding is used for detecting oil spill in European Remote Sensing (ERS) SAR images. Here, a search is done within the eight neighbourhood directions which is then followed by merging of the responses. In this method, linear features are reported accentuated [5].

Thresholding based segmentation technique requires that the image histogram has a number of peaks and that each of these peaks corresponds to a region of the image. It has the advantage of not needing to have a prior information of the image thus has less computational complexity. However, it does not work well for an image without any obvious peaks which may be the case in some SAR image data. Additionally, thresholding methods do not consider spatial details and hence cannot guarantee that the segmented regions are contigous [81] and as a result false alarm rate is high .

2.4.2 Clustering

Clustering method as with other segmentation methods aims to group image data into meaningful regions. It can be achieved using supervised or un-supervised approach. The clustering segmentation method uses techniques such as the Fuzzy C-Means and K-means clustering. K-means clustering however seems to be very popular due to its simplicity and fast execution speed.

K-means clustering aims to partition n number of observations into k clusters, with each observation n_i belonging to a cluster to the nearest mean, which serves as a pattern of the cluster. K-means clustering can be in the following methods iterative, numerical, non-numerical or unsupervised [81]. On the other hand, Fuzzy C-Means clustering also referred to as soft clustering has the ability to represent the relationship between input pattern data and clusters more naturally [81]. It uses the fuzzy partition matrix to describe the degree of uncertainty of each cluster thus influencing fuzziness [83].

In [84], clustering is proposed to segment oil spill areas in hyper spectral images, the key idea is the emphasis on the spectral profile of oil, and the presence of non oil spill regions are suppressed to create image slices. The image slices maximises the possibility of accurate identification since it is derived from the perspective of the spectral profile of oil. The algorithm considered the many bands contained in hyperspectral images, hence reduced the dimensionality from n bands to a single band. Once the slices are generated, clustering is applied to label oil spill regions. The idea of image slicing is to minimize classification errors considering both spatial and spectral details.

Guangmin et al. [83], proposed the use Fuzzy C-Means clustering technique for segmenting oil spill in remote sensing images captured by UAV system. First they selected the Y' (luma) and Cb (blue difference) components of the YCbCr color space to enable the construction of oil color model. The fuzzy clustering is then applied to classify the color vector. From the color model, cluster centres are chosen from the model range and the clustering result forms the segmentation. The drawback to this method is its sensitivity to noise.

In [85], a multi scale segmentation method of oil spill in SAR images which is based on JSEG() and spectral clustering is proposed. This approach efficiently overcame the effect of anti-noise which is a problem when segmenting SAR images using the traditional spectral clustering based on gray features. Multi-scale J-images are used to extract the multi-features and the Laplace matrix is clustered by the K-means method. Finally, a decision-level fusion

strategy is used to fuse the segmentation results from different scales. The problem with this method is its limitations on direction sensitivities and the difficulties in selecting the best feature combination.

2.4.3 Other Segmentation Methods Proposed

Other segmentation methods have also been proposed in literature. In [14], a Bayesian adaptive oil spill segmentation of SAR images via graph cuts is proposed. The approach used real ERS and Envisat SAR images of verified oil spill. The method modelled the density of the observed amplitudes given the oil spill region as a finite mixture of Gamma distributions which renders robustness to the backscattering fluctuations within each region. The prior, an M-level Markov Random Field (MRF) is defined on a 2D grid to enforce continuity in a statistical sense, and the maximum *a* posteriori (MAP) segmentation is computed by means of graph-cut techniques.

Miguel et al [80] proposed a segmentation method of oil spill in SAR images using an adaptive stochastic minimisation which measured the extent of marine oil spill. The experiment was done using Envisat/ASAR images of the Deep Water Horizon oil spill of the Gulf of Mexico. The method explored a binary segmentation scheme based on the MRF theory to obtain a more integrated field. A modified simulated annealing schedule is then used to perform a joint conditional estimation of model parameters. To get a finer detection, the pixel neighbourhood system of the *a* priori model is continuously updated at each step of the maximisation algorithm.

The use of super pixel method to segment oil spill in SAR images is proposed in [86]. The approach is to first partition the SAR image into grids. Secondly otsu segmentation is then applied for adaptive thresholding during the image binarysation. In the end the suspicious pixels are clustered and the false positives are eliminated by another approach that combines the space distance, intensity deviation and size threshold together. The method performed well on real SAR image of RADARSAT-1.

Although, a variety of segmentation methods have been proposed in the detection of oil spill using SAR images, the key idea of all is to detect all suspicious oil spills and to also preserve their shapes [5]. The next step in the oil spill detection framework is to extract features from the image. This will not be discussed in this report. Once features are extracted the next step is to distinguish the oil spills from look alikes using a classifier.

2.5 Summary

This chapter presents a background of oil spill remote sensing. Firstly, remote sensing systems in passive and active forms utilised for oil spill detection are discussed. The advantages and disadvantages of systems and sensors utilised for the application are presented with more emphasis on the microwave sensors and in particular SAR which is the most active sensor utilised for oil spill detection due to its ability to acquire images irrespective of weather condition or cloudiness. Additionally, the backscattering of radar and its effect on capillary waves allows the dampening on oil sea surfaces causing the formation of dark patches on the SAR images. A general assumption among scientist in this area is that oil spill appear as dark formations on SAR images. However, other sea phenomena such as weeds and sea sperm appear as dark formations on SAR images, hence complicating spill detection with these sensors. On the other hand, passive sensors are affected by cloudiness and rely on external energy (sun) to acquire data, but their ability to capture images in contagious bands enables the identification of other materials on the sea surface and also enables the identification of the type of oil spilled and even the quantity through spectral unmixing. This thesis aims to combine image data originating from different sensors and platforms. In the second part of the chapter, the theoretical background of image fusion methods used in this thesis are introduced. Considering that Wavelet based methods create mosaic in fused images, this thesis proposes a Gaussian Process regression approach to fuse multi-modal images using a non-stationary covariance that models pixel location and their intensities, and handles the change of support problem inherent in multi-modal images. Additionally, Gaussian Process regression models are discussed in more detail and introduced. The last part of this chapter reviews segmentation methods developed for oil spill detection in the literature. Segmentation of images plays a vital role in oil spill detection especially when utilising SAR images to distinguish the dark patches from the image background.

Chapter 3

Pansharpening Methods for Spectral Unmixing Assessment of Oil Spill

3.1 Introduction

The continued development of remote sensing systems and the design of new sensors for earth observation have enabled extensive use of remote sensing images for many applications including for environmental monitoring, oil spill monitoring and detection, change detection, disaster management and security operations to mention a few. In the design of these sensors, often considerate trade-off becomes inevitable due to the technological drawbacks of such sensors hence necessitating spatial and spectral resolution adjustments. Major causes of limitations in design and development of remote sensing satellites especially sensors designed for the optical range dwell on two factors namely: 1). The radiation energy source that facilitates acquisition of images when it strikes the sensor and 2). The size of the image data acquired by the sensor.

Remote sensing systems such as Panchromatic (PAN) imaging sensors (e.g. SPOT 6/7) acquire images with high spatial resolution but with lower spectral channels due to their inherent expansive bandwidth while on the hand, Multispectral (MS) sensors (e.g. Worldview 2/3) and Hyperspectral (HS) imagers (e.g. Aviris, Hyperion) generate image data with low spatial resolution but with higher spectral channels ranging from less than 10 bands to more than a hundred due to their inherent narrower bandwidth. To improve the quality of images and benefit from the complimentary advantages of the multi-modal data acquired by different sensors, methods that combine image data from differing imaging systems have

been proposed in the literature in a procedure popularly referred to in the computer vision society as "Pansharpening". Methods for this approach aim to combine the high spatial PAN images with the high spectral HS or MS to produce an image with both characteristics and of higher quality than the individual images. In this chapter, a quantitative and qualitative study of pansharpening * methods is carried out using the Gulf of Mexico oil spill data acquired on 12/06/2010 by SpecTIR; a hyperspectral imager with 2.2m spatial resolution and 360 spectral bands between the range of 390-2450nm to investigate the performance of the different pansharpening methods developed in the literature. A further study into spectral un-mixing is done on the different images formed by each method to additionally, investigate the performance of pansharpened images against HS images for spectral un-mixing of oil spill.

3.2 Methods for Pansharpening

Pansharpening methods [53, 87] have been studied widely in the computer vision field and used for many image processing applications including for thematic mapping, change detection, image analysis and in understanding a scene of interest (scene interpretation) [88]. These methods, can be classified into four (4) according to [53, 87, 88]. In the following up sections, a discussion and critical comparison of these methods is done using the reference hyperspectral data obtained by SpecTIR and simulated semi-synthetic PAN and HS images accordingly.

- Component Substitution (CS)
- Multi-Resolution Analysis (MRA)
- Hybrid Methods
- Bayesian Methods

3.2.1 Component Substitution Analaysis

Implemented in popular remote sensing softwares (e.g. ENVI, ERDAS Imagine etc.), the CS based methods are the most widely used pansharpening approaches due to their simplicity. They rely on a component substitution usually obtained by spectral conversion of the high

*Pansharpening and fusion will be used throughout this chapter interchangeably

spectral resolution data e.g the MS or HS data with the high spatial PAN data. Conventionally, the procedure uses spectral bands of the MS data and projects them to another space by transforming the spectral information with a component of the spatial HS data substituted to the PAN data. Lastly, the inverse is taken to represent the fused image. CS family comprises of methods such as the Principal Component Analysis (PCA), Intensity-Hue-Transform (IHS) and Gram-Schmidt (GS). Additionally, the extended versions of these methods include. A typical approach to the CS method is described in Fig 3.1 where HS PAN′ represent the transformed HS data, IP is the Inverse Projection of the HS data and the components refer to the substituted components of the spectral MS data. The CS fusion technique can be expressed mathematically as

$$\widehat{\mathbf{F}}^k = \widetilde{\mathbf{H}}^k + G_k(\mathbf{P} - \mathbf{S}_L), \tag{3.1}$$

where k= 1, . . . , N and defines the spectral dimension of the fused and the high spectral data, $\widehat{\mathbf{F}}^k$ represents the kth band of the fused * or estimated image, $\mathbf{G} = [G_1, \ldots, G_N]^T$ is the vector of injection gains, with $\mathbf{S}_L$ being the component to be substituted and it is defined as

$$\mathbf{S}_L = \sum_{i=1}^{N} w_i \widetilde{\mathbf{H}}_i, \tag{3.2}$$

The weights vector $\mathbf{w} = [w_1, \ldots, w_i, \ldots, w_N]^T$ is the measure of the spectral overlap that exist between the high spectral image data (MS or HS) and the high spatial image data (HS).

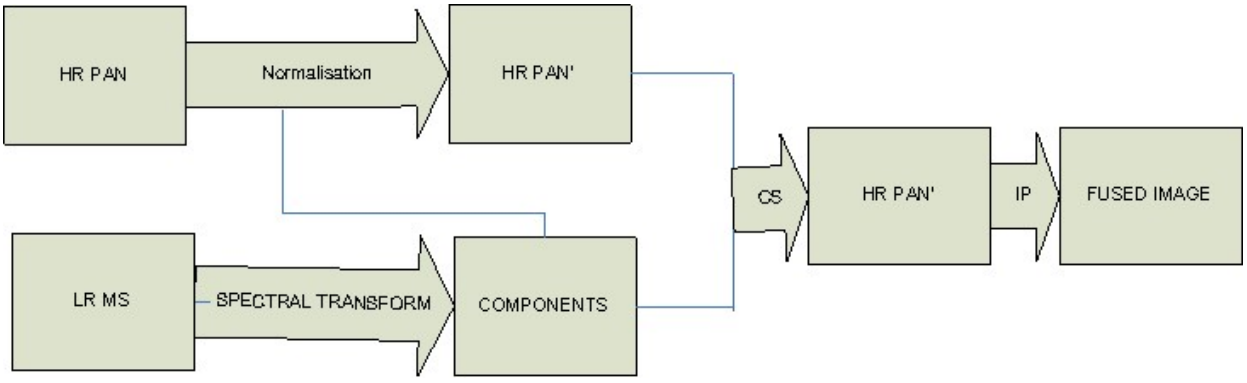


Figure 3.1: A framework of the traditional approach for CS pansharpening methods [53]

*Pansharpened and fused images will be used interchangeably throughout the chapter

3.2.1.1 Principal Component Analysis

PCA employs a spectral transformation approach by linearly transforming the original data through rotation which results in the formation of principal components also referred to as PC's [89]. The idea behind this fusion approach is in its ability to concentrate the spatial information that is shared by all channels in its first PC and the spectral information specific to each spectral band in the high spectral data is considered in the other PC's. This is same as expressing the spectral information as $N-1$, where N represent the number of spectral channels in the MS or HS image data. The fusion process can be formulated using (3.2) and (3.3), respectively. The coefficient vectors of weights $\mathbf{w}$ and injected gains $\mathbf{G}$ are formed from the PCA fusion process originating from the high spectral image data.

3.2.1.2 Gram-Schmidt Algorithm

GS method like the PCA, is a transformation based pansharpening approach. Initially proposed by [90], the GS method is a patented work by Kodak and implemented in popular image processing software ENVI as a spectral pansharpening tool [87]. The algorithm begins with a pre-processing step that involves interpolating the high spectral bands of the MS or HS data to the corresponding scale of the high spatial PAN data, this way, images are lexically ordered as vectors with dimensions corresponding to the number of image pixels at the scale of the high spatial PAN data [53, 87, 88]. Secondly, a component is substituted from the HS or MS data and replaced with the PAN image data through an orthogonal decomposition of the component and then decomposed to recover the fused or pansharpened result [91], this can be expressed as

$$g_k = \frac{\mathrm{cov}(\widetilde{\mathbf{H}^k}, \mathbf{S}_L)}{\mathrm{var}(\mathbf{S}_L)} \,, \tag{3.3}$$

where the number of bands ranges from $k = 1, \ldots, N$ in the high spectral data. $\mathrm{cov}(\widetilde{\mathbf{H}^k}_i, \mathbf{S}_L))$ defines the covariance between spectral channels and component substituted while $\mathrm{var}(\mathbf{S}_L)$ is the variance of substituted components. Low resolution PAN image can be obtained by averaging the HS bands using the weights from (3.3) i.e.

$$w_i = 1/N, \quad i = 1 \ldots N \,. \tag{3.4}$$

An extended form of the GS method known as the Adaptive GSA is described in [92] where

the $\mathbf{S}_L$ is generated by linearisation of (3.3) with $\mathbf{w}$ estimated through the minimisation of MSE between the downsampled high spatial PAN data and the estimated components [53]

3.2.1.3 Intensity-Hue-Transform

Intensity Hue Transform (IHS) exploits the human visual system and achieves this by processing image data into Intensity (I), Hue (H) and Saturation (S). The image data as a preprocessing step is first converted into the IHS information color space. A drawback to this method is in its limitation of accommodating only the visible bands of the electromagnetic spectrum i.e Red (R), Green (G) and Blue (B) components. When dealing with MS or HS images, this is a major drawback since these images contain more than 3 spectral channels. An improvement of this limitation is proposed in [93] which considers more than 3 spectral channels and generalises the number bands by formulating the problem for any random sets of spectral weights that are non-negative

$$\widehat{\mathbf{F}}_k = \widetilde{\mathbf{H}}_k + \left(\sum_{i=1}^{N} w_i \right) (\mathbf{P} - \mathbf{S}_L), \quad k = 1, \ldots, N \,, \tag{3.5}$$

where $\widehat{\mathbf{F}}_k$ defines the fused or pansharpened image, $\mathbf{S}_L$ is as defined in (4.3) and the coefficients of weights $\{w_i\}_k$ equals $1/N$ [87], and $m\lambda$ defines the number of spectral bands in MS data.

3.2.2 Multi Resolution Analysis

Multi-Resolution Analysis (MRA) based methods [88], rely on the spatial filtering of the high spatial image data (PAN) in generating the spatial details to be infused into the spectral channels of the high spectral image data (HS). Proposed initially as a single level decomposition technique e.g. the High Pass Filters (HPF) and Discrete Wavelet Transform (DWT). The role of the PAN image in the fusion process is basically in finding the difference between high spatial PAN image ($\mathbf{P}$) and its filtered low-pass form ($\mathbf{P}_L$). This is mathematically expressed as

$$\widehat{\mathbf{F}}_k = \widetilde{\mathbf{H}}_k + g_k(\mathbf{P} - \mathbf{P}_L), \quad k = 1, \ldots, N. \tag{3.6}$$

Generally, the low-pass filtered form $\mathbf{P}_L$ is generated by a continual decomposition of the PAN image, resulting in low spatial resolution versions of the high spatial PAN image. The

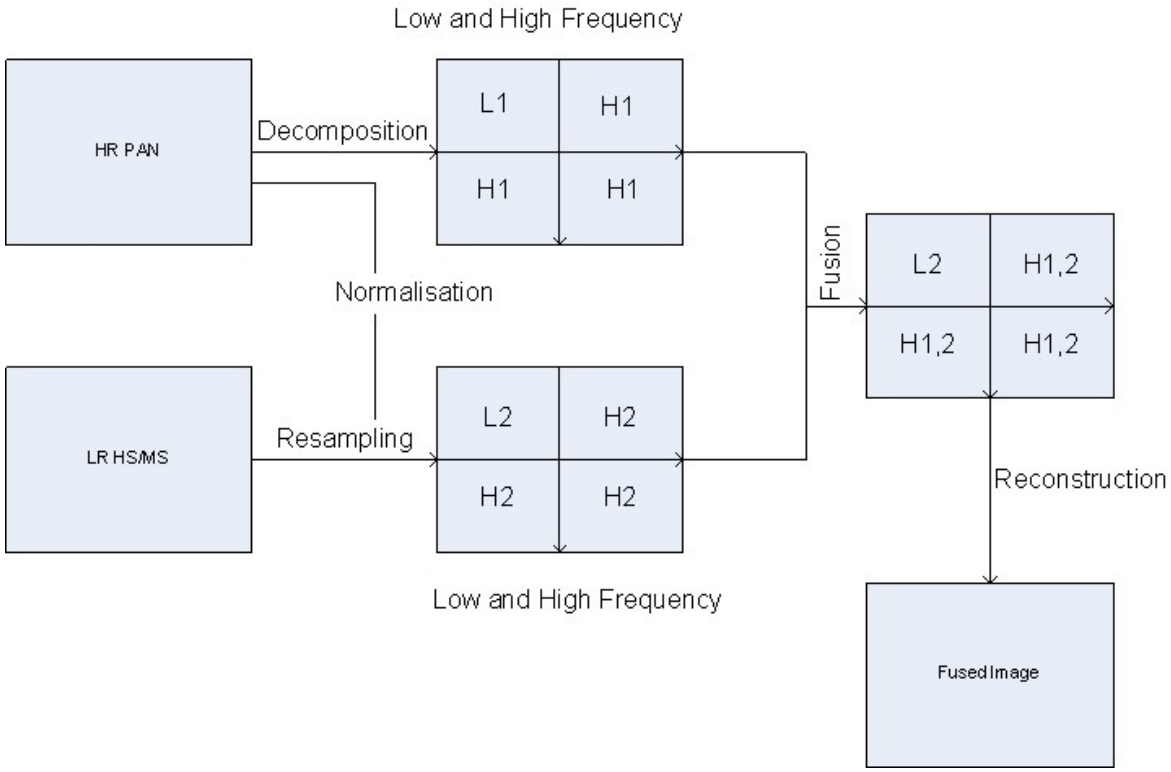


Figure 3.2: A conventional MRA-based methods approach for pansharpening images

value of k ranges from 1 to N, where N represent the number of spectral bands in the high spectral MS or HS data. A description of the basic idea behind the MRA pansharpening methods is shown in Fig 3.2. The advantages of MRA based techniques include spectral and temporal consistency, while the disadvantage is in its computational complexity when compared with the CS based methods, this is so because the fusion process in MRA is further complicated by the additional design or implementation of the spatial filters [53]. A careful attention should be given to how the LPF versions of $\mathbf{P}_L$ are obtained from $\mathbf{P}$ since the quality of the pansharpened or fused image will largely depend on this operation. Additionally, the decomposition of the original high spectral data (MS or HS) ranges from a simple single level decomposition which is usually achieved with the use of LPF or to more complex approaches of the MRA. Secondly, the gains needed to be injected into the fusion process need to be defined in either of two form as shown in [53]. The first form which is also defined as the additive injection utilises a suitably fitted matrix whose elements added together gives $\mathbb{1}$ as shown in (3.7)

$$\mathbf{G}_k = \mathbb{1}, \quad k = 1, \ldots, N \, . \tag{3.7}$$

The second choice of coefficients is the so called High Pass Modulation (HPM) where the injected details are taken as weights of the ratio R between $\widetilde{\mathbf{H}}$ and $\mathbf{P}_L$ i.e. the interpolated high spectral image and the LPF version of the high spatial PAN. This is expressed as

$$\mathbf{G}_k = \widetilde{\mathbf{H}}_k \oslash \mathbf{P}_L \qquad k = 1, \ldots, N \,, \tag{3.8}$$

where the $\oslash$ defines an element-element division of pixels. The purpose of this, is to balance the contrast between spectral and spatial channels of the multi-modal images to be fused. In [53], however, it is shown that the injection gain G_k depends on multiplying the combination of the spatial and spectral channels which is quite similar to the above expression in (4.8). This multiplication can be expressed as

$$\widehat{\mathbf{F}}_k = \widetilde{\mathbf{H}}_k + \frac{\widetilde{\mathbf{H}}_k}{\mathbf{P}_L}(\mathbf{P} - \mathbf{P}_L) = \widetilde{\mathbf{H}}_k \cdot \frac{\mathbf{P}}{\mathbf{P}_L} \,, \tag{3.9}$$

Here, $\widetilde{\mathbf{H}}_k \cdot \frac{\mathbf{P}}{\mathbf{P}_L}$, defines the multiplicative between two modalities. Spectral malformation are minimised using the MRA based methods if the filter used in generating the $\mathbf{P}_L$ details is distinctive, hence the generated details can then be injected with the gains $\mathbf{G}_k$ to the bands of the spectral image (HS or MS) data. When dealing with HS images for pansharpening, [53] noted that the ratio between the spatial resolution of the images may not always result in a whole number which makes it difficult to extend MRA methods designed for MS images. Additionally, it is rare to acquire PAN and HS on the same platform. In the following up sub sections, methods in the MRA are discussed.

3.2.2.1 Smoothening Filter-based Intensity Modulation

Smoothening Filter-based Intensity Modulation (SFIM), was introduced in [94], the SFIM method is a fusion and pansharpening technique that aims to improve spatial resolution details without distorting the spectral properties. This method utilises an averaging technique of the LPF that generates the spatial details from $\mathbf{P}$. Normally, the implementation of the MRA as described by (3.1) utilises the application of a Linear Time Invariant (LTI) , [95] LPF H_{LP} to $\mathbf{P}$ in generating $\mathbf{P}_L$, and (3.1) can be re-written as

$$\widehat{\mathbf{F}}_k = \widetilde{\mathbf{H}}_k + g_k(\mathbf{P} - \mathbf{P}_L * H_{LP}), \quad k = 1, \ldots, N \,, \tag{3.10}$$

where $*$ defines the convolution operation between the differences in $\mathbf{P}$ and $\mathbf{P}_L$ and the value

of k ranges from 1 to the number of bands N. Summarily, the SFIM places the H_{LP} employing simple averaging and utilises the HPM to inject the spatial details.

3.2.2.2 Laplacian Pyaramid

The formation of the low pass filtered forms of $\mathbf{P}$ to generate $\mathbf{P}_L$ can be implemented using a decomposition approach often referred to as the "Pyramid Decomposition". First proposed in [96], this approach utilises a Gaussian filter as the LPF at each step of the decomposition level to generate a Gaussian Pyramid. The Difference of Gaussian (DoG) between the decomposition levels results in what is termed the "Laplacian Pyramid" [53]. Additionally, the Gaussian filter can conveniently be adjusted to complement the sensor's Modulation Transfer Function (MTF) [97] which allows the extraction of fine details from the PAN image that are not captured by the HS or MS sensor [87]. The ability of the filter in complementing the MTF rests with the only parameter of tuning in this case the standard deviation (σ) of the Gaussian filter that can be matched with the sensor's designed parameters including its amplitude response at Nyquist frequency, a value provided by the sensor's manufacturer [98]. In this chapter, a comparison of the different approaches under the Laplacian Pyaramid technique are evaluated. Firstly, the general approach here referred to as MTF-GLP and the MTF utilising HPM here referred to as the MTF-GLP-HPM.

3.2.3 Hybrid Methods

The Hybrid techniques exploit the ideas of other classes of pansharpening methods (CS, MRA) either by combining them or using an instance from the methods. The Guided-Filter-PCA (GFPCA) is one method that employs the hybrid approach.

3.2.3.1 Guided-Filter-PCA

The Guided Filter-PCA (GFPCA) attempts to find stability in the trade-off between the spatial and spectral resolutions of images to be fused which have proven to be a challenging task in image fusion applications [53]. The Guided Filter, relies in its ability to capture the structures of images to enhance spatial resolution since the spatial properties are preserved in the filtering process. Its efficiency in HS image classification have been shown in [99] where the structural properties are augmented with the PC's of the HS image for a classification task. The idea in this method is that the structural properties of the guidance image are preserved

within the image's PC's and transferred to the filtered image i.e. the structural properties of the guidance image are not lost in the filtering process, another advantage of this approach is that it can substitute the CS method since spectral distortions inherent in CS methods can be avoided with the guided filter. The fusion framework using the GFPCA is illustrated in Fig 3.3.

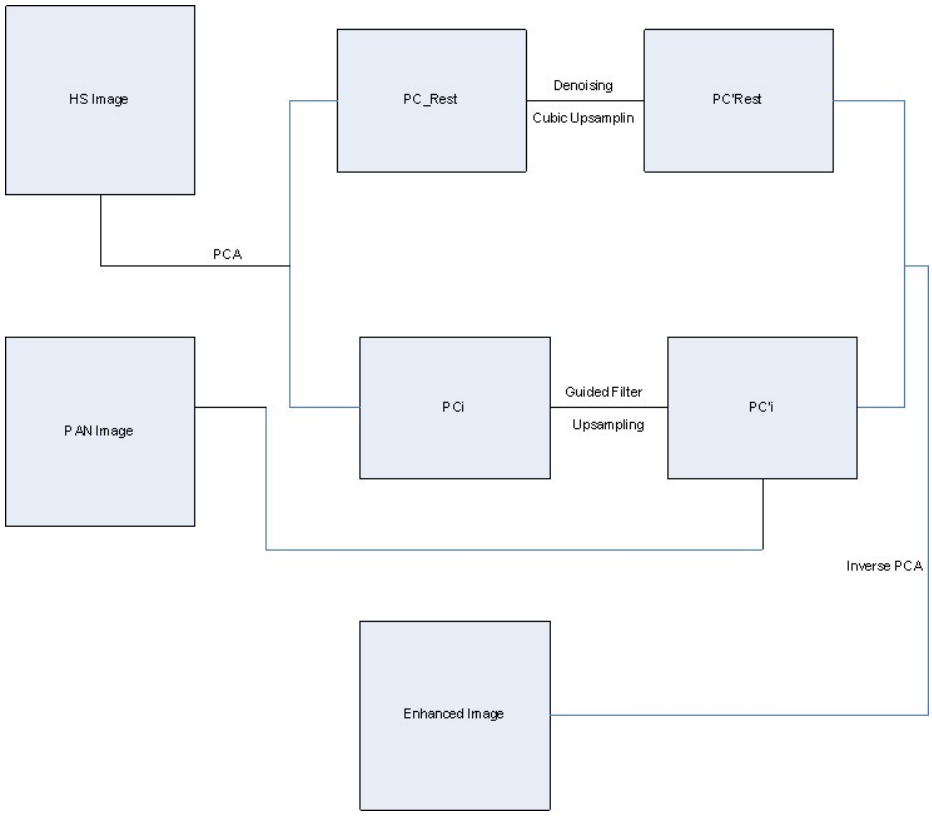


Figure 3.3: A Conventional Hybrid GFPCA Fusion for Pansharpening

Firstly, spectral bands of the HS images are decorrelated so that image information is seperated from noise and allowing the fusion process to be accelerated while improving computation. This is possible since the first PC's contain mostly information (spectral) and can be represented as $p \ll N$, and the remaining PC's $N - p$ contain noise where N is the number of spectral bands in the image. As described in Fig 3.3, the filter magnifies the first PCA channels of the HS image, while retaining the spatial property of the PAN image and the remaining PC's are upsampled using a cubic interpolation [100]. Representing the ith PC channel as PC_i, attained from the high spectral image data $\mathbf{H}$, and its spatial resolution marked up to the scale of the high spatial image in this case $\mathbf{P}$ using bicubic interpolation earlier described, the resultant of the filtering process PC_i is the affine transformation of $\mathbf{P}$ within a local window

ω_i with size $(2d+1) \times (2d+1)$ given as:

3.2.4 Bayesian Methods

Bayesian methods for image fusion and pansharpening follows a perceptive approach that allows the fusion process to be formulated within the posterior distribution of a Baye's model [101]. The idea behind this approach is in its ability to formulate a suitable prior for the scene of interest. The model have been used for image fusion applications especially for pansharpening of MS and PAN images. In [21], a multi-band image fusion utilising the Bayes's approach is proposed. Firstly, the physical degradations that define the blurring or subsampling as a result of sensor characteristics are considered, and then an appropriate prior that exploit the geometrical considerations is introduced within the Bayesian estimation framework. Lastly, a Markov Chain Monte Carlo (MCMC) is utilised to generate samples that compute the estimator for the scene of interest. This approach is different from many Bayesian approaches for instance [64] propagates the information of the higher resolution images into the lower resolution images enabling the images to have different spectral channels. This latter approach is quite similar to the wavelet transform method that allows images to be decomposed as coefficients [50] at different resolutions. Wavelet transform can be implemented using many approaches e.g. Daubechies and Haar [102]. The comparison in this chapter for the Bayesian fusion methods, follows the observation models introduced in [53] and given here as

$$\mathbf{H} = \widehat{\mathbf{F}}\mathbf{V}\mathbf{S} + \boldsymbol{\epsilon}_H \,, \tag{3.11}$$

$$\mathbf{M} = \mathbf{R}\widehat{\mathbf{F}} + \boldsymbol{\epsilon}_M \,, \tag{3.12}$$

where $\widehat{\mathbf{F}}$, $\mathbf{H}$ and $\mathbf{M}$ define the fused, hyperspectaral and multispectral images. Generally, the fusion approach in Bayesian tends to find the uncertainty associated with a model and encode some properties of the sensor for example it instability to measurement for inference [21]. Other variables within the observation in (3.12) and (3.13) can be defined as

- $\mathbf{V} \in \mathbb{R}^{n \times n}$ defines a cyclic convolution operator that correlate with the spectral response of the HS sensor at the same scale of resolution with either the $\mathbf{M}_k$ or $\mathbf{P}_k$ image [103].
- $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a downsampling matrix with factor d earlier defined in (4.1)

- $\mathbf{R} \in \mathbb{R}^{n_k \times m_k}$ defines the spectral response of **M** or **P** sensors.

- ϵ_H and ϵ_M defines the individual noise terms in **H** and **M**, respectively. This, assumes a zero mean and a normal distribution with covariances Λ_H and Λ_M.

For this comparison the high spatial image to be fused is a PAN image with a single spectral channel i.e. **P**, and $N = 1$. For this, the high resolution image $\widehat{\mathbf{F}}$ to be recovered from the fusion process is assumed to live in a lower dimensional space and utilises the geometrical considerations of **H**, the high spectral resolution image. With this assumption, the expression of distributions of (3.12) and (3.13) becomes

$$\mathbf{H}|\mathbf{U} \sim \mathcal{MN}_{mN,\, m}(\mathbf{TUVS},\ \Lambda_H\, \mathbf{I}_m)\ , \tag{3.13}$$

$$\mathbf{M}|\mathbf{U} \sim \mathcal{MN}_{nN,\, n}(\mathbf{RTU}, \Lambda_M, \mathbf{I}_n)\ , \tag{3.14}$$

Here, $\mathcal{MN}$ defines the normal distribution of the matrices and the fused image $\widehat{\mathbf{F}}$ is here expressed as $\widehat{\mathbf{F}} = \mathbf{TU}$ and $\mathbf{T} \in \mathbb{R}^{mN \times \tilde{m}N}$ and N defines the spectral channels in the images to be fused. The columns of **T** forms the basis for the dimensional subspace whose size is $\tilde{m}_N \ll m_N$ and the matrix $\mathbf{U} \in \mathbb{R}^{\tilde{m}_N \times n}$ consists of the coefficients of the fused image $\widehat{\mathbf{F}}$ and regarding to **T**.

With the above expressions, and considering that the noise terms $\mathbf{ffl}_H$ and $\mathbf{ffl}_M$ are independent, the posterior according to Baye's theorem can be defined as

$$P(\mathbf{U}|\mathbf{H}, \mathbf{M})\ \propto\ P(\mathbf{H}|\mathbf{U})\ P(\mathbf{M}|\mathbf{U})\ P(\mathbf{U}) \tag{3.15}$$

or otherwise

$$-log\ P(\mathbf{U}|\mathbf{H}, \mathbf{M}) \doteq \underbrace{\frac{1}{2}\|\Lambda_H^{-\frac{1}{2}}(\mathbf{H} - \mathbf{TUVS})\|_F^2}_{\text{High spectral data term}} \tag{3.16}$$

$$+ \underbrace{\frac{1}{2}\|\Lambda_M^{-\frac{1}{2}}(\mathbf{M} - \mathbf{RTU})\|_F^2}_{\text{High spatial data term}} + \underbrace{\lambda\phi(\mathbf{U})}_{\text{Normalisation}} \tag{3.17}$$

The methods under the Bayesian approach to image fusion are tuned to define the normal-

isation term $\lambda\phi(\mathbf{U})$. In the next sub section, a brief of the two approaches will be discussed namely:

3.2.4.1 Naive Bayesian Approach

The Bayesian Naive approach to image fusion assigns the Gaussian prior to the columns of $\mathbf{U}$ assuming the entries of the column are mutually independent [53], such that the expression becomes

$$P(\boldsymbol{u}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \mathcal{N}(\boldsymbol{\mu}_i\boldsymbol{\Sigma}_i) \ , \tag{3.18}$$

Here $\boldsymbol{\mu}_i$ represent the fixed image described by $\widetilde{\mathbf{H}}$ and $\boldsymbol{u}_i = (i = 1, \ldots, n)$. The hyperparameters of the model to be estimated are defined as the matrix $\boldsymbol{\Sigma}_i$ here assumed to be identical entries with $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$, hence reducing the number of parameters to be optimised as shown in [104]. Furthermore, this parameter is further estimated alongside other parameters of interest to deduce the matrix $\mathbf{U}$ from the posterior distribution $p(\mathbf{U}|\mathbf{H}, \mathbf{M})$. Next the the samples to be used for the estimation are generated. It is worthy to mention here, that aside the MCMC earlier introduced, other sampling methods such as the Gibbs sampling [103] and more recently by solving a Sylvester equation [105] in relation to the optimization problem and improving significantly the computation of reaching convergence. can also be utilised. As mentioned earlier, this chapter follows the review presented in [53] on the Bayesian methods where the MCMC approach is utilised to generate samples required to approximate the Bayesian estimators for the posterior.

3.2.4.2 Sparse Bayesian Approach

The introduction of sparsity in image fusion is mostly to improve the computational complexity inherent in combining image details especially when dealing with multi-band images where the number of bands can range into hundreds. The difference in this approach when compared with the Naive is that rather than integrating the Gaussian prior into the fusion process, a sparse representation is adopted to regularise the fusion process. A dictionary of linearly combined components of sparse image representation are utilised and learned to improve image depiction that enables a sparse image dependent regularisation of the fusion process. In [21, 53] the regularisation term is expressed as

$$\phi(\mathbf{U}) \propto -logP(\mathbf{U}) \doteq \frac{1}{2}\sum_{k=1}^{N} \|\mathbf{U}_k - \mathcal{P}(\bar{\mathbf{D}}_k\bar{\mathbf{A}}_k)\|_F^2 \,, \tag{3.19}$$

Here $\propto$ defines an equality aside the additive constant since additive constants will not change the position of the optimal value and the functions in consideration need to be optimised [53]. $\bar{\mathbf{D}}$ defines the dictionary from which the image elements are represented, $\bar{\mathbf{A}}$ defines the code for the number of bands in the image and $\mathcal{P}$ is a linear operator that defines the averages overlapping patches.

3.2.4.3 Hysure Method

This approach utilises a convex regularisation solution within the Bayesian framework and adapt a Vector Total Variation (VTV) [11] form to $\phi(\mathbf{U})$, taking into consideration both spatial and spectral properties of the images to be fused. Additionally, a new convex approach is designed to determine the spatial and spectral response of the considered sensors [106]. This appraoch can be considered as a blind super-resolution method [21, 53, 107] utilising two convex solutions for the fusion process. The VTV regulariser can be expressed as

$$\phi\mathbf{U} = \sum_{i=1}^{C_a} \sqrt{\sum_{k=1}^{N} \{[\mathbf{U}\mathbf{D}_h]_{ij}^2 + [(\mathbf{U}\mathbf{D}_v)_{ij}]^2\}}, \tag{3.20}$$

where $\mathbf{A}_{ij}$ indicate the element of the ith row and jth column of $\mathbf{A}$, and the products by matrices $\mathbf{D}_h$ and $\mathbf{D}_v$ calculate the vertical and horizontal disjunct differences of an image with periodic border conditions , respectively. Lastly, the pansharpened high spectral image $\widehat{\mathbf{F}}$ is the result of an optimisation solution given as

$$\min_{\mathbf{U}} \quad \frac{1}{2}\|(\mathbf{H} - \mathbf{TUVS})\|_F^2 + \frac{N}{2}\|\mathbf{M} - \mathbf{RTU})\|_F^2 + N_\phi\, \phi(\mathbf{U}), \tag{3.21}$$

and the spatial and spectral responses of the sensors can be evaluated by solving the optimisation problem with quadratic regularisers and their respective regularisation parameters as shown in [105].

3.2.4.4 Coupled Non-negative Matrix Factorisation

The Coupled Non-negative Matrix Factorisation (CNMF) method explores a matrix factorization based on a linear spectral un-mixing model [108]. The adoption of matrix factorisation to fusion relies on two underlying factors as illustrated in [53] that 1). The dictionary $\bar{\mathbf{D}}$ of the

signal subspace is learnt from the high spectral image $\mathbf{H}$ resulting in the factorisation $\widehat{\mathbf{F}}$=$\bar{\mathbf{D}}\mathbf{U}$ and secondly 2). In cases where the noise term from (4.13) is trivial, i.e. $\epsilon_M \simeq 0$. Initially introduced in [109], this method can be adopted for HS pansharpening in cases where the high spatial image has only one spectral band e.g. $\mathbf{P}$ with a single band because the approach attempts to un-mix the individual images for fusion to obtain individual endmembers and the high spatial resolution image abundance maps. Where an endmember is earlier defined in the previous chapter. The utilisation of spectral unmixing is basically the foundation of the CNMF approach.

The spectral unmixing of $\mathbf{H}$ with the assumption that the spectrum of each pixel of the image is a linear combination of the various endmembers, and hence express the high spectral image as $\mathbf{H} \in \mathbb{R}^{m\lambda \times n}$ and the linear spectral unmixing can be formulated as

$$\mathbf{H} = \mathbf{F}\boldsymbol{\omega} \ , \tag{3.22}$$

where $\mathbf{F} \in \mathbb{R}^{m\lambda \times p}$ is the matrix containing the spectral description of the endmembers, and $\boldsymbol{\omega} \in \mathbb{R}^{p \times n}$ defines the abundance matrix that holds the corresponding abundances of the various endmembers at pixel level while p defines the number of endmembers. In the fusion process, the matrix $\mathbf{F}$ is initiated using a linear spectral unmixing model like the Vertex Component Analysis (VCA) [110] and $\boldsymbol{\omega}$ is estimated using the high spatial $\mathbf{P}$ with a single band spectrum.

3.3 Quality Assessment of Pansharpened Images

The fusion or pansharpening of varying image data may not guarantee a perfect image which is why it is important to evaluate the quality of fused or pansharpened images [8]. Unfortunately this is not a very easy task especially when dealing with real-life high spectral and high spatial images. Firstly, due to amount of data embedded in high spectral images and secondly because there is usually no reference image to compare with. A conventional approach to evaluating the quality of fused and pansharpened images is the Wald's protocol [111]. However, [87] pointed out that in regards to pansharpening methods, the Wald's protocol is not efficient in situations where the reference image is not available to make comparative analysis. In view of this, two approaches have been presented for the evaluation process

1. Lower the image resolution by artificially inducing this and comparing with the original

high spectral image

2. Evaluate using statistical indexes that find the relationship between original images and the fused or pansharpened image.

The first approach though providing a concise evaluation may be affected by the process of lowering the image resolution particularly when spatial filters are employed in the process, because the image can be affected visually with smoothening effect of the filter and this may affect the results of comparisons. In the second situation, the relationship between the individual images fused and the pansharpened image are compared using established indexes [112]. However, the work of [53] based a general standard to assess the quality of fused images using the Wald's protocol introduced in [113] using two (2) properties that ideally the fused image should possess 1.The fused image should be consistent when compared with the reference image such that if the fusion process is to be reversed, then the original high spectral image ($\mathbf{H}$) should be retrieved if the degradation procedure the fused image is right 2. The second property is that of synthesis that requires the pansharpened or fused image be closely alike with an image of high resolution acquired by a different sensor. To satisfy the second property, the correlation between complimentary bands in the individual images to be fused need to be adequately conserved alongside image features that define the components and objects of the image. It is important also to note that the issue of evaluating the quality of pansharpened and fused images is still a very active area of research in the computer vision society as shown in [112, 114, 115].

In this chapter, the Wald's protocol for assessing the quality of fused and pansharpened images is adopted in a procedure depicted in Fig.3.4. Firstly, the observed high spatial PAN image $\mathbf{P}$ and the high spectral Hyperspectral image $\mathbf{H}$ are generated from the Reference Hyperspectral Image of the real-life data. The pansharpened results discussed in the previous section are then evaluated using the synthetically simulated data and the quality of the fusion results are evaluated against the reference image using fusion quality assessment indices. In the next section, a brief description of the reference data is given

3.4 Data Description

The data used in this comparative analysis is a SPECTIR hyperspectral image [116] obtained on june 12, 2010 at about 4pm during the heat of the Gulf of Mexico [8] oil spill

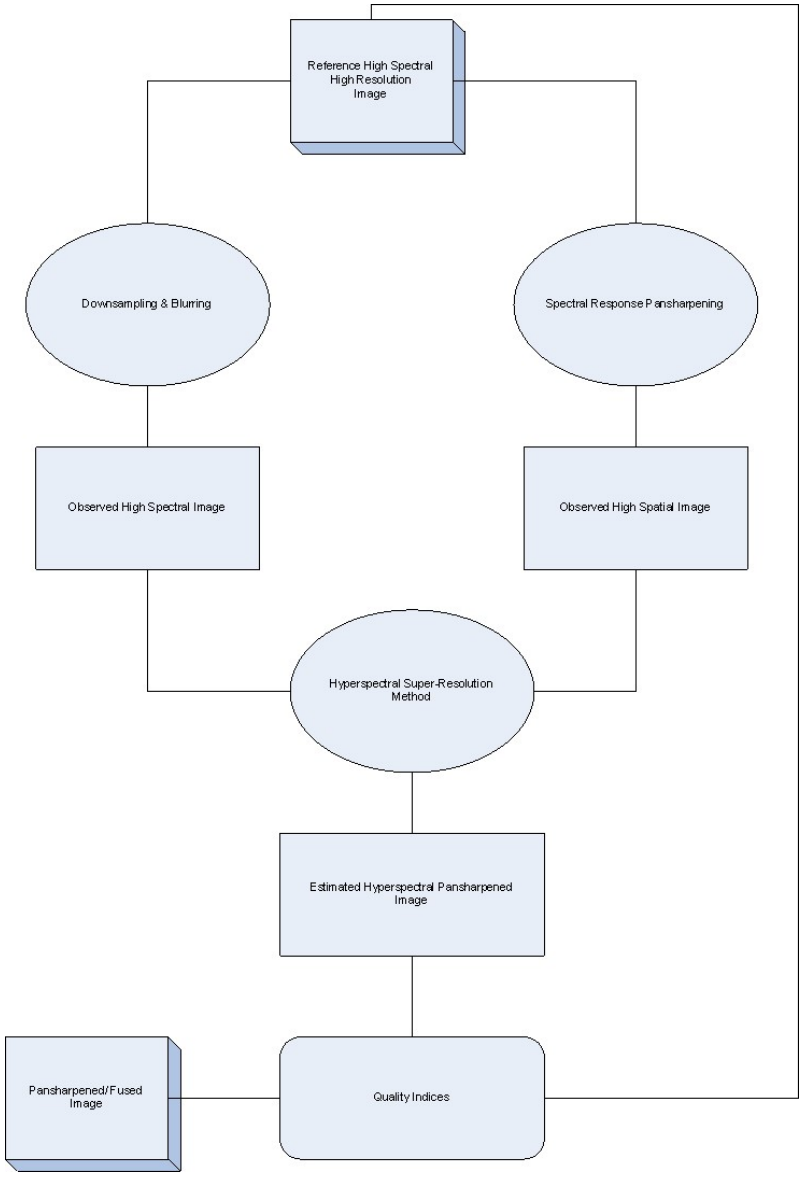


Figure 3.4: Flow chart of the pansharpening procedure with the SPECTIR image as described in [111]

crises in the United States. The Gulf of Mexico oil spill occurred on the 20th April, 2010 as a result of the explosion of British Petroleum (BP)'s tanker discharging an estimated 4.9million barrels of crude into the Gulf. The image data can be downloaded online at *www.spectir.com/download.html* as a sample data provided for scientific research and evaluation.The SPECTIR hyperspectral sensor provides a 2.2m spatial resolution image with 360 spectral bands covering 390-2450 nm wavelengths range. A subset of the image that is cloud free has been carved out for this analysis. Firstly a pre-processing step of the data is done

to remove water logged bands and bands with high signal to noise ratio (SNR) leaving the remaining spectra for 224 bands for further analysis. Secondly an Atmospheric correction is done with FLAASH software embedded in the ENVI 4.0. Although, an in-situ is not provided with the image, there is however helpful and applicable information on the Gulf of Mexico as discussed in [10].

3.4.1 Description of the Synthetic Dataset

Table 3.1 shows the description of the simulated synthetic data for the PAN image $\mathbf{P}$ and the simulated Hyperspectral image $\widetilde{\mathbf{H}}$. Recall, that in the pre-processing step, noise sensitive and water vapour bands have been removed leaving the actual reference image $\mathbf{H}$ with a total of 224 bands as described in section 3.4. The simulated data presented in Table 3.1. are generated according to the Wald's protocol as illustrated in Fig 3.4. The SPECTIR hyperspectral image has been used as the reference image and the PAN and Hyperspectral images are generated accordingly.

Data Type	Dimension	Spatial Resolution	Bands
Reference Data	560×280	2.2m	224
Panchromatic	560×280	4m	1
Hyperspectral	140×70	20m	224

Table 3.1: Characteristic of the synthetic dataset

3.5 Quality Assessment Measures

In this section we describe the quality measures used to evaluate the performance of the individual pansharpening and fusion methods for high spatial PAN ($\mathbf{P}$) and the high spectral hyperspectral image ($\mathbf{H}$), respectively. In evaluating the quality of the resultant high spatial and high spectral image, it is important to specify what quality need to be measured. In literature, quality assessment measures are classified into three namely; spatial quality measures, spectral quality measures and global image quality assessment measures. In this chapter, the estimated image is evaluated using atleast a measure from each of the classes.

3.5.1 Correlation Coefficient

The correlation coefficient (CC) is a spatial quality assessment measure used to evaluate the spatial similarity between an observed image and the estimated image in other words the

individual images to be fused and the resultant fused or pansharpened image. Additionally, the CC also compares geometric distortion of the images [53] and is given by

$$CC(\widehat{\mathbf{F}}, \mathbf{H}) = \frac{1}{N} \sum_{i=1}^{N} CCS(\widehat{\mathbf{F}}^i, \mathbf{H}^i) \ , \tag{3.23}$$

Here CCS defines the cross correlation for a single band image given as

$$CCS(\mathbf{B}, \mathbf{L}) = \frac{\sum_{j=1}^{n} (\mathbf{B}_j - \mu_B)(\mathbf{L}_j - \mu_L)}{\sqrt{\sum_{j=1}^{n} (\mathbf{B}_j - \mu_B)^2 \sum_{j=1}^{n} (\mathbf{L}_j - \mu_L)^2}} \ , \tag{3.24}$$

where the matrices $\mathbf{B}, \mathbf{L} \in \mathbb{R}^{1 \times n}$ defines two common one band images and $\widehat{\mathbf{F}}^i$ is the ith element of the pansharpened image. The value of $\mu_I = (1/n) \sum_{j=1}^{n} \widehat{\mathbf{F}}_j$ defines the mean of $\widehat{\mathbf{F}}$. Ideally, the value of this quality measure is 1 [8].

3.5.2 Spectral Angle Mapper

Spectral Angle Mapper (SAM) is a spectral quality measure, that measures the angle between correlative pixels in the reference and estimated images, respectively using a predefined space and considering each band to be a coordinate axis [87]. Given the pixel vectors $\mathbf{b}, \mathbf{l} \in \mathbb{R}^N$ the value of SAM can be calculated either in radians or degrees using

$$SAM(\mathbf{b}, \mathbf{l}) = arccos\left(\frac{\langle \mathbf{b}, \mathbf{l} \rangle}{||\mathbf{b}|| \ ||\mathbf{l}||} \right), \tag{3.25}$$

here, $\langle \mathbf{b}, \mathbf{l} \rangle = \mathbf{b}^T \mathbf{l}$ defines the scalar product and $|| * ||$ is the vector of the $\ell 2$ norm. SAM also measures the degree of preservation of the spectral shapes and its optimal value is ideally 0 [53]. In this analysis, the SAM value is taken per pixel and the final value obtained is the result of averaging the values obtained on each pixel of the images.

3.5.3 Root Mean Square Error

Remote Mean Square Error (RMSE) is used as a pansharpening image quality assessment measure to determine the difference between the reference and estimated images. It is the measure of the ℓ 2 error between the matrices $\widehat{\mathbf{F}}$ and $\mathbf{H}$. This is given as

$$RMSE(\widehat{\mathbf{F}}, \mathbf{H}) = \frac{||\widehat{\mathbf{F}} - \mathbf{H}||_F}{\sqrt{n * N}} \ , \tag{3.26}$$

where $||\mathbf{H}||_F = \sqrt{\text{trace}(\mathbf{H}^T\mathbf{H})}$ defines the Frobenious norm of $\mathbf{H}$. An ideal value of this measure is 0.

3.5.4 Erreur Relative Globale Adimensionnelle de Synthese

With this measure, the global quality of the pansharpened or fused image can be determined. Meaning both spatial and spectral properties are taken into consideration with this measure. The ideal value of this measure is 0 [50, 53, 112]. ERGAS is calculated using

$$ERGAS(\widehat{\mathbf{F}}, \mathbf{H}) = 100r\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{RMSE_k}{\mu_k}\right)}, \tag{3.27}$$

where the value of r defines the ratio between the linear resolutions of the source images $(\mathbf{P}, \widetilde{\mathbf{H}})$ and is given by:

$$r = \frac{\mathbf{P}_v}{\widetilde{\mathbf{H}}_v}, \tag{3.28}$$

where $\mathbf{P}_v$ and $\widetilde{\mathbf{H}}_v$ represent the linear spatial resolutions of the PAN and the Hyperspectral images, respectively. $RMSE_k$ is given as

$$RMSE_k = \frac{||\widehat{\mathbf{F}}^k - \mathbf{H}^k||_F}{\sqrt{n}}. \tag{3.29}$$

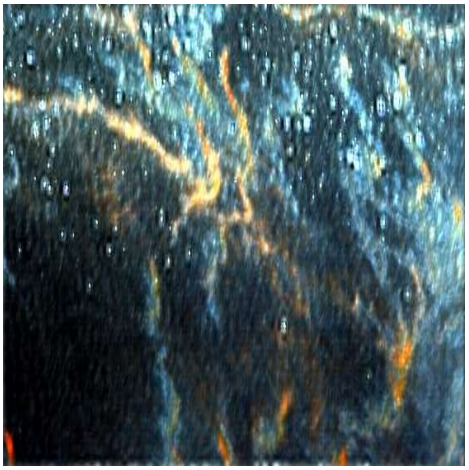
The value of μ_k defines the sample mean of the kth band of the reference image $\mathbf{H}$.

3.6 Pansharpening Results

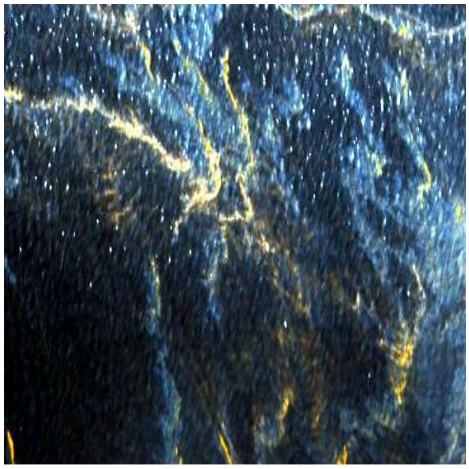
In this section, the individual results of the various methods discussed are visually presented. The reference image is shown to make visual comparisons. Though it may be difficult to compare performance of the methods visually, the results of comparing the methods using the quality measures discussed are also presented and the computational complexity result of each method, respectively.

3.7 Spectral Unmixing (SU)

In this section, the unmixing comparison is done. The aim here is to compare the performance of the estimated images recovered from individual pansharpening methods and the reference

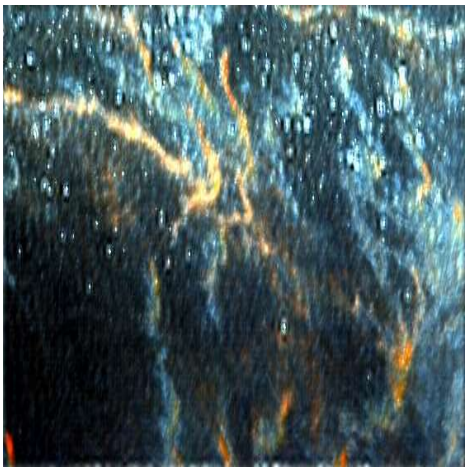


(a) Reference Hyperspectral Image

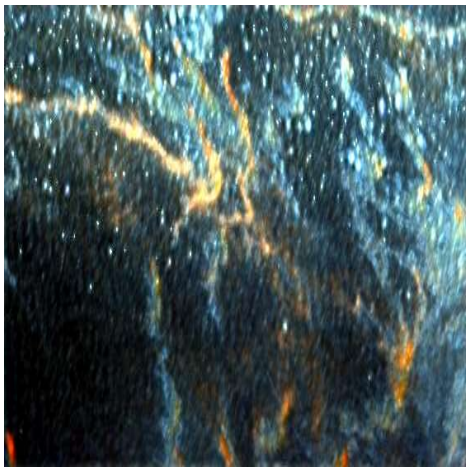


(b) Interpolated Hyperspectral Image

Figure 3.5: Reference Image and the Interpolated Hyperspectral Image



(a) SFIM Method Pansharpened Result

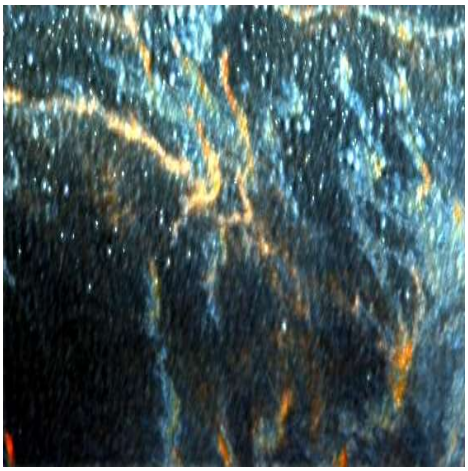


(b) MTF_GLP Method Pansharpened Result

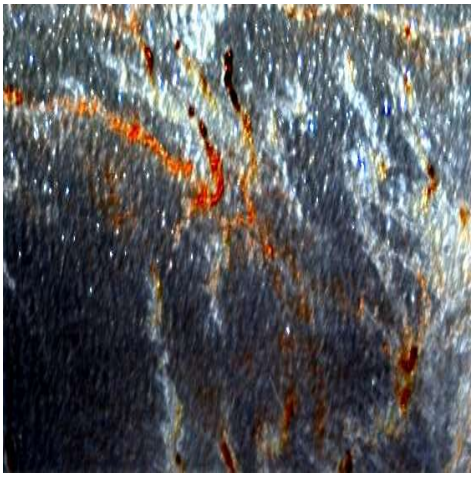
Figure 3.6: Pansharpened results for SFIM & MTF-GLP methods, Respectively.

Methods	**Computational Time**
SFIM	6.02
MTF-GLP	6.13
MTF-GLP-HPM	6.01
GS	5.46
GSA	6.27
PCA	9.29
GFPCA	4.55
CNMF	18.56
BayesNaive	2.68
BayesSparse	180.01
Hysure	124.96

Table 3.2: Computational Times of Methods

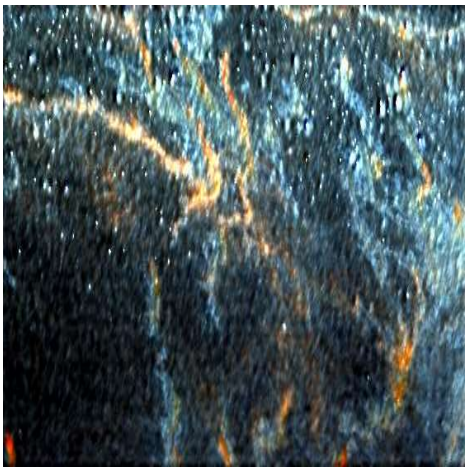


(a) MTF-GLP-HPM Pansharpened Result

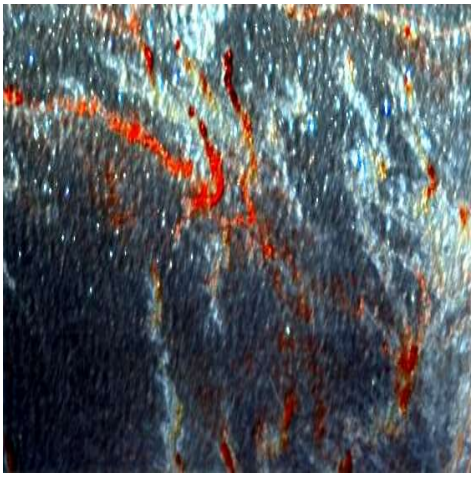


(b) GS Method Pansharpened Result

Figure 3.7: Pansharpened Result for MTF-GLP-HPM & GS Methods, Respectively.



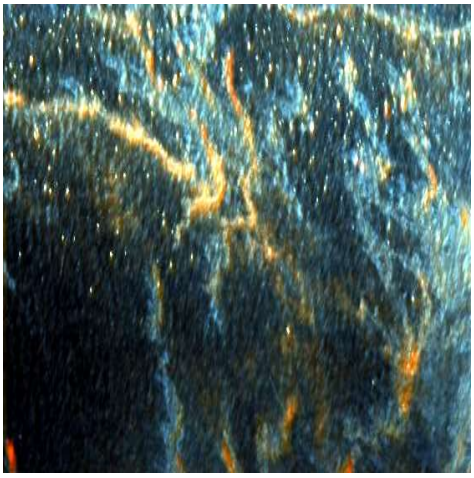
(a) GSA Pansharpened Result



(b) PCA Method Pansharpened Result

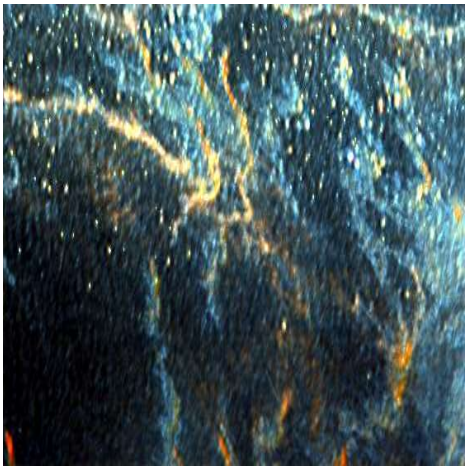
Figure 3.8: Pansharpened Result for GSA & PCA Methods, Respectively.

(a) GFPCA Pansharpened Result

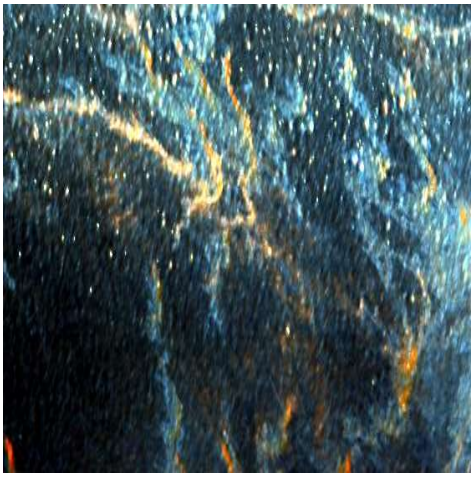


(b) CNMF Method Pansharpened Result

Figure 3.9: Pansharpened Result for GFPCA & CNMF Methods, Respectively.



(a) Bayes Naive Pansharpened Result

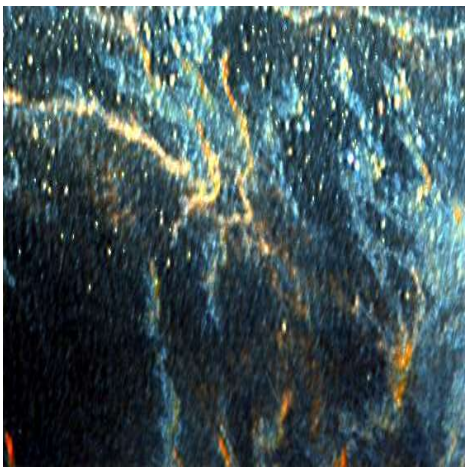


(b) Bayes Sparse Method Pansharpened Result

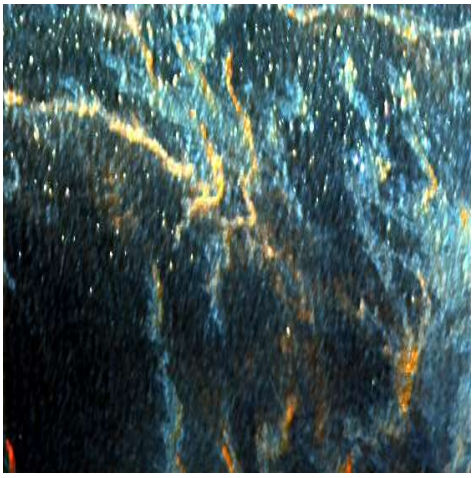
Figure 3.10: Pansharpened Result for Hysure Methods, Respectively.

hyperspectral image in unmixing the endmembers present in the scene of interest. Recall, that the pansharpened image $\widehat{\mathbf{F}}$, the interpolated hyperspectral image $\widetilde{\mathbf{H}}$ and the Reference Hyperspectral image ($\mathbf{H}$) have 224 spectral channels each, respectively. The first task of spectral unmixing is to define the spectrum mixing model that represents how the endmembers * converge to form the spectrum as measured by the sensor [119]. In cases where the number of endmembers is known, the SU can be done using a supervised approach or un-surpervised when the endmembers are unknown. In either case, defining the mixture model is an important step of the unmixing process [12]. Firstly, the first assumption is adopted knowing

*an endmember is defined as a mineral that is at the extreme end of a mineral series in terms of purity

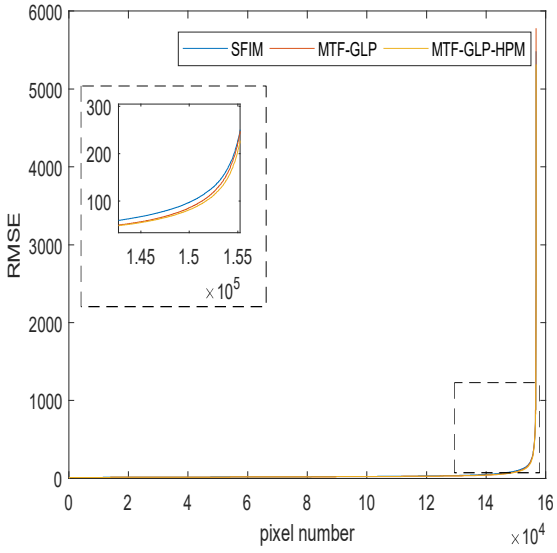


(a) Bayes Naive Pansharpened Result

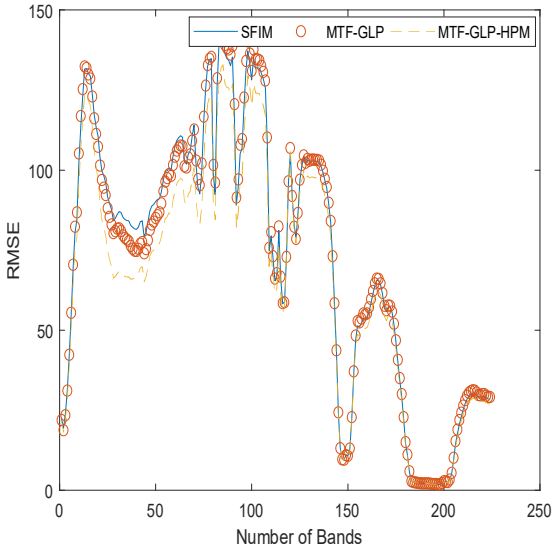


(b) Hysure Method Pansharpened Result

Figure 3.11: Pansharpened Result for Bayesian Methods, Respectively.



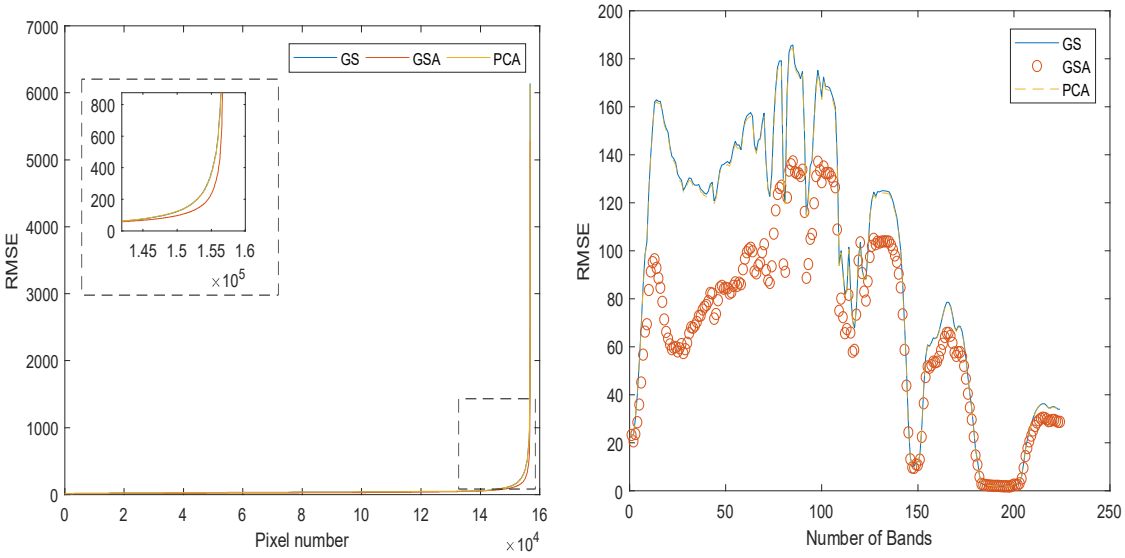
(a) Per Pixel RMSE of MRA based Methods



(b) Per Band RMSE of MRA based Methods

Figure 3.12: Per Pixel & Per Band RMSE Comparison of MRA based Methods

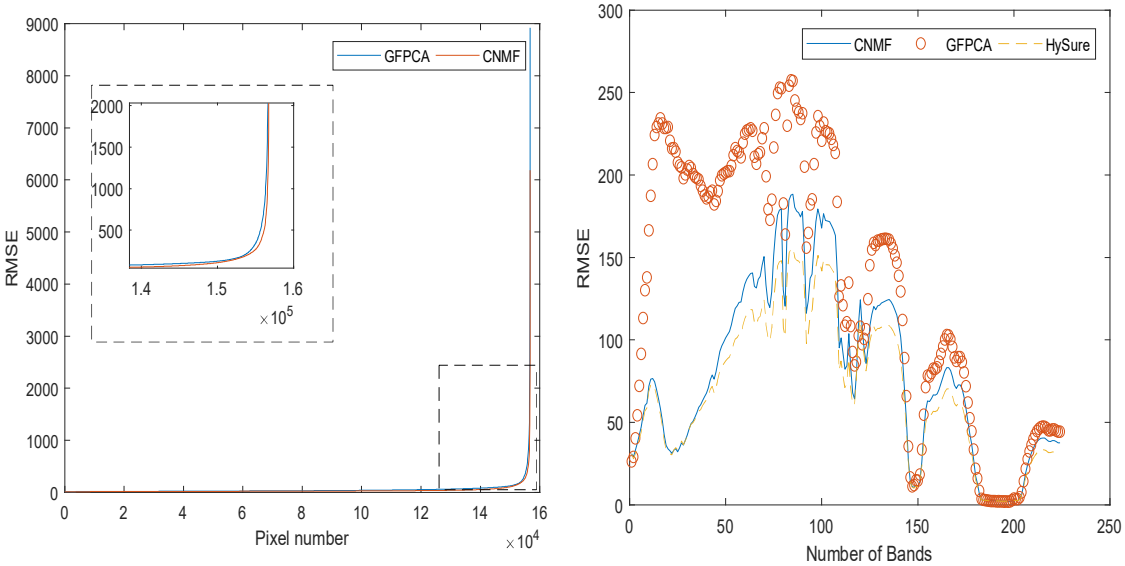
the endmembers of the scene interest and adopting the linear SU technique. Secondly the comparison is done again using the non-linear SU appraoch for the scene of interest. Since the data contains some cloud, this is added to the number of endmembers present. For this case, the endmembers are oil,water and cloud. Recall that the measured spectrum is given as $\boldsymbol{y}_p \in \mathbb{R}^N$ where N defines the number of spectral band and the Linear Mixing Model (LMM) can be defined as



(a) Per Pixel RMSE of CS based Methods

(b) Per Band RMSE of CS based Methods

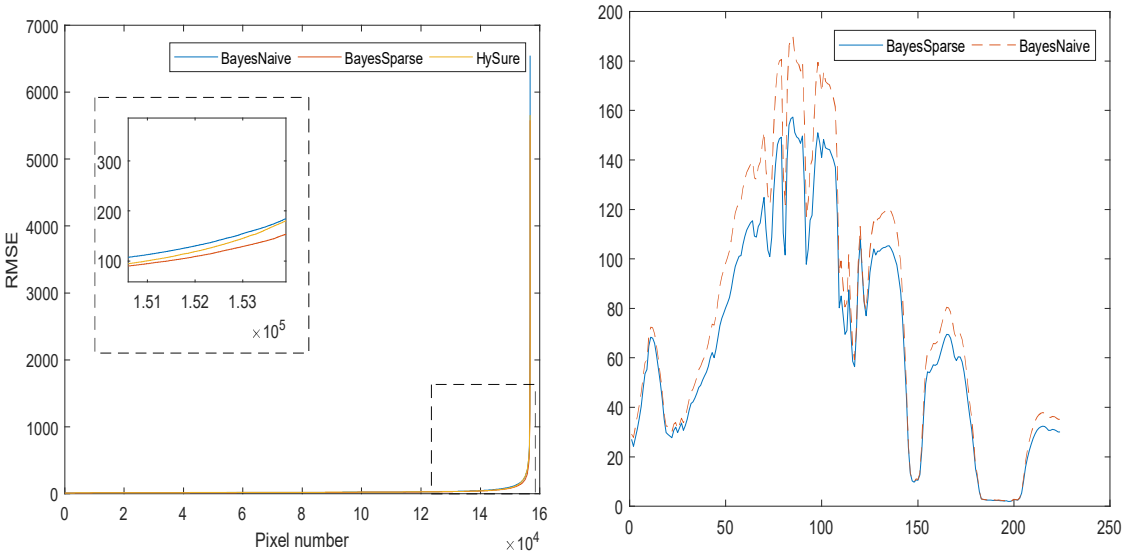
Figure 3.13: Per Pixel & Per Band RMSE Comparison of CS based Methods



(a) Per Pixel RMSE of Hybrid based Methods

(b) Per Band RMSE of Hybrid based Methods

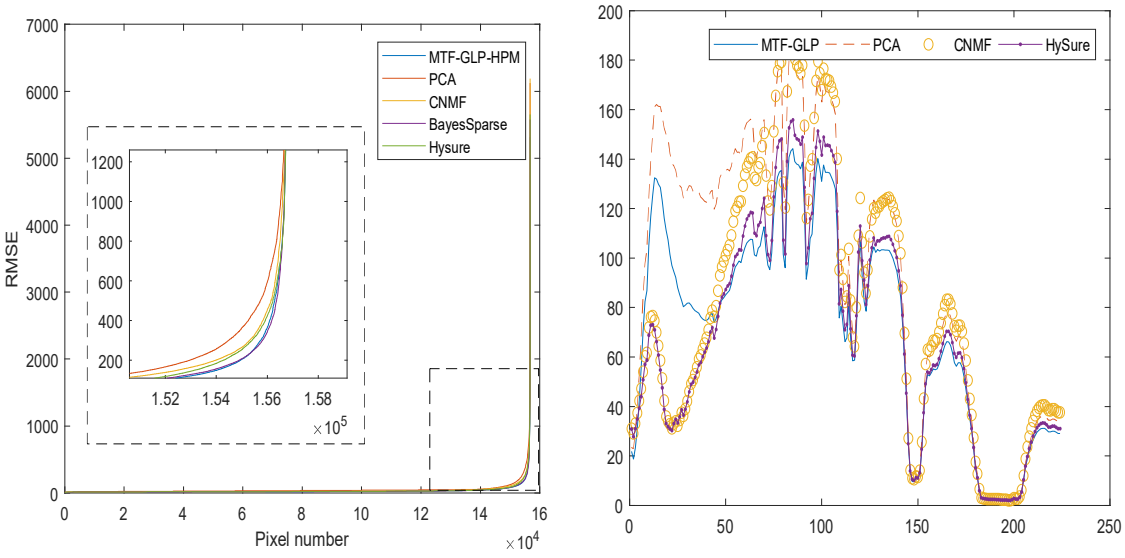
Figure 3.14: Per Pixel & Per Band RMSE Comparison of Hybrid based Methods



(a) Per Pixel RMSE of Bayesian based Methods

(b) Per Band RMSE of Hybrid based Methods

Figure 3.15: Per Pixel & Per Band RMSE Comparison of Hybrid based Methods



(a) Per Pixel RMSE of Hybrid based Methods

(b) Per Band RMSE of Hybrid based Methods

Figure 3.16: Per Pixel & Per Band RMSE Comparison of Hybrid based Methods

$$\boldsymbol{y}_p = \sum_{e=1}^{E} a_{e,p}\mathbf{m}_e + \boldsymbol{\epsilon}_p \tag{3.30}$$

where E defines the number of endmembers in the scene of interest, $\mathbf{m}_e$ is the spectral signa-

ture of the *eth* endmember, $a_{e,p}$ defines the abundance of *eth* endmember in the *pth* pixel of the image and $\boldsymbol{\epsilon}_p$ is the additive noise or model error.

In oil spill detection and analysis using optical sensors (Hyperspectral, Multispectral etc.), the process of unmixing plays a great role since the final aim is to estimate the amount of oil spilled which can be known with measuring the spectra for oil. In [116], spectral unmixing is utilised to evaluate the characterisation of spilled oil and to quantify the thickness of oil spilled as illustrated in [117] or to determine the abundance for spill coverage calculation described in [118].

To compare the unmixing of pansharpened images against the hyperspectral data, two linear and two non-linear SU methods are utilised.The research question here is to find out if pansharpening of images have an added advantage when unmixing endmembers of the scene of interest when compared to non-pansharpened hyperspectral data. In subsequent sub-sections the linear and non-linear SU methods are briefly discussed. With more details of this discussed in [12, 27, 119].

3.7.1 Linear Spectral Unmixing

Linear approach to unmixing is widely adopted due to its simplicity and easy implementation. The assumption here is that the mixing scale is clear i.e the mixing is visible and the incident angle between the source of energy in this case the sun and the acquiring sensor interact with one material [12]. An important consideration of this approach is the prior knowledge of the number of endmembers present in the scene of interest and their corresponding spectral reflectances [27]. Examples of such linear approaches include the Vertex Component Analysis (VCA) and the Fully Constrained Least Squares (FCLS) method. The two methods will be used for comparison in this chapter.

3.7.1.1 Vertex Component Analysis (VCA)

VCA approach to linear spectral unmixing was first proposed in [110]. In this approach the mixing model is assumed for each observed spectral vector and is given as

$$\mathbf{L} = \mathrm{x} + \mathrm{n} = \mathbf{M} \underbrace{\Gamma \boldsymbol{\alpha}}_{\wp} + \mathrm{n}\,, \tag{3.31}$$

where the vector $\mathbf{L}$ defines the number of spectral bands in the hyperspectral image, $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2 \ldots, \mathbf{m}_p]$ describes the mixing model matrix and $\mathbf{m}_i$ represent the ith endmember

signature and p defines the total number of endmembers present in the scene of interest, and $\wp \equiv \Gamma\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}$ signifying the scale that models the illumination change due to the surface terrane and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha p]^T$ is the abundance estimation vector of all endmembers in the scene of interest, the notation $(.)^T$ defines the vector transpose as described earlier in this chapter. Finally, n defines the additive noise in the data in this case. The VCA approach is based on the geometry of convex sets [27] and relies on the fact that the endmembers occupy vertices of a simplex [12] in addition to assuming spectrally pure pixels in the dataset [120]. The data is iteratively (high spectral image e.g. **H**) projected to the direction that is orthogonal to the subspace covered by the endmembers already fixed (the value is initiated). A new endmember signature is then determined that corresponds to the maximum of the projection. This procedure is performed iteratively until all endmembers within the scene of interest are exhausted [12]

3.7.1.2 Fully Constrained Least Squares Approach

In linear SU, an important consideration is the prior knowledge of the abundances. In actuality, this is difficult and therefore needs the abundance vector $\boldsymbol{\alpha}$ to be estimated using a posteriori information acquired from the data. The FCLS method first introduced in [121] make use of the simplex to create a set of attainable solutions for the unmixing process. It achieves this by assigning a null (0) value to negative values in the abundance estimation while the remaining fraction of material signatures are normalised to 1 [27]. In other words, all abundance values in the negative are discarded and other values normalised and retained. The estimate of the abundance vector α using the least squares from (3.32) is given by

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{L}\ , \tag{3.32}$$

where the variables of (3.33) are as defined in (3.32), respectively. The algorithm imposes constraints to the unmixing procedure namely; an Abundance Non-negative Constraint (ANC) that discards negative values and the Abundance Sum-one Constraint (ASC) that sum to one all other values by normalisation as described earlier. It is also important to point out that these constraints are not carried out simultaneously but rather sequentially as described in [121].

3.7.2 Non-Linear Spectral Unmixing

The validity of linear spectral unmixing techniques are in most cases debatable due to the requirement for a priori knowledge of the number of endmembers and their corresponding abundances which is in reality difficult to estimate [27, 119, 121]. For this reason, non-linear approaches have been considered to unmix high spectral images. Non-linear SU takes into consideration the multi scattering effect and the interaction between the source of energy (sunlight) and the endmembers in the scene of interest. In addition, the abundances of the endmembers can be regarded as corresponding spaces occupied by the material within a pixel of the high spectral image; in which case additional constraints can be considered for the abundance coefficients expressed in (3.31) [119]. In summary, the non-linear methods are applicable when the conditions that holds for a linear approach are not fulfilled for example when the mixing of endmembers is not at macroscopic level [122]. In this comparison, two popular non-linear spectral unmixing techniques are utilised.

3.7.2.1 Generalarised Bilinear Mixing Model

The bilinear approaches for non-linear spectral unmixing explore the elevation differences in the lay-out of materials (endmembers) contained in the scene of interest. Considering that the mixture model is determined by the direction of photons interacting with at least a material in the linear mixing. However, in situations where the interaction between the photons is with more than one material, the spectral unmixing procedure becomes more complex prompting non-linear unmixing solutions. The GBMM non-linear model is utilised to account for the effect of multiple scattering of photons by considering [123] and their interaction with materials. A generic approach to this is the initiation of a second term to handle the multiple photon interaction problem hence providing a solution to the mixed spectrum contained in each pixel of the acquired image of the scene of interest. This solution is given as

$$\boldsymbol{y}_i = \sum_{e=1}^{E} a_e \mathbf{m}_e + \sum_{i=1}^{E-1} \sum_{j=i+1}^{E} y_{i,j}\, a_i a_j \mathbf{m}_i \odot \mathbf{m}_j + \epsilon \ . \tag{3.33}$$

In (3.34) above, the first term relates to the linear mixing model, while the double sums account for the non-linear effect. $\mathbf{m}_i \odot \mathbf{m}_j$ defines the term-term product of the spectral signatures, $\boldsymbol{y}$ describes the real parameter vector in the range (0,1) that estimate the interaction between the different spectral components and the ϵ is the additive noise from the acquiring

sensor. GBMM also, utilises the positivity and sum-one constraints earlier described for non-linear SU approaches.

3.7.2.2 Polynomial Post Non-linear Mixture (PPNM)

In this non-linear approach to SU, the spectrum of a pixel is taken as a non-linear transformation of the non-linear function g_p of the linear mixture of spectral signatures affected by some noise ϵ_p [124]. Additionally, the reflectance of the acquired image is assumed to be a non-linear function of pure spectral components (endmembers) where the model combines linear and quadratic functions of a_e. A motivation to this non-linear approach is the "Weierstrass Approximation Theorem". Introduced in [125], this theorem states that at desired precision [27], a continued function defined over an interval can be approximated uniformly using a polynomial.The PPNM model is given as

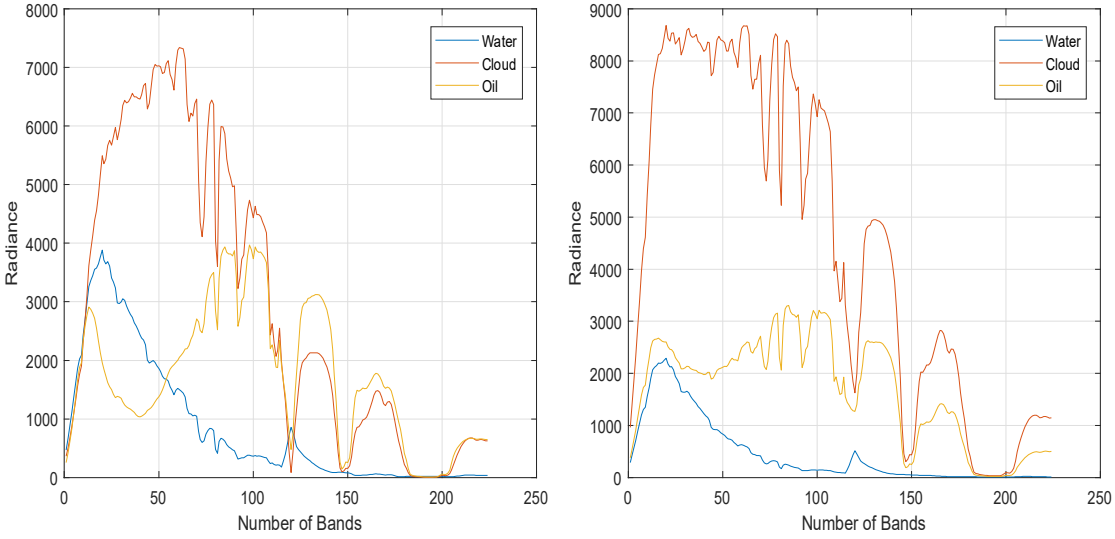
$$y_p = g_p\left(\sum_{e=1}^{E} a_e m_e\right) + \epsilon_p \ . \tag{3.34}$$

3.8 Spectral Unmixing Results

In this section, the results of the spectral unmixing are presented. Furthermore a comparison of these results is done on selected methods of the pansharpening process. A method is selected each from the discussed families (MRA, CS, Bayesian etc.) of pansharpening approaches and compared with the reference **H** hyperspectral image that is being utilised as a ground truth data. The whole spatial (560×280) extent of the images are utilised for the comparison to allow for more information to be utilised and for the procedure to be as close to real life scenarios or situations. In most cases, only windows of interest points are selected which is not always the situation in reality. Additionally, the Signal to Noise Ratio are varied through 10,30 and 50dB for comparison purpose.

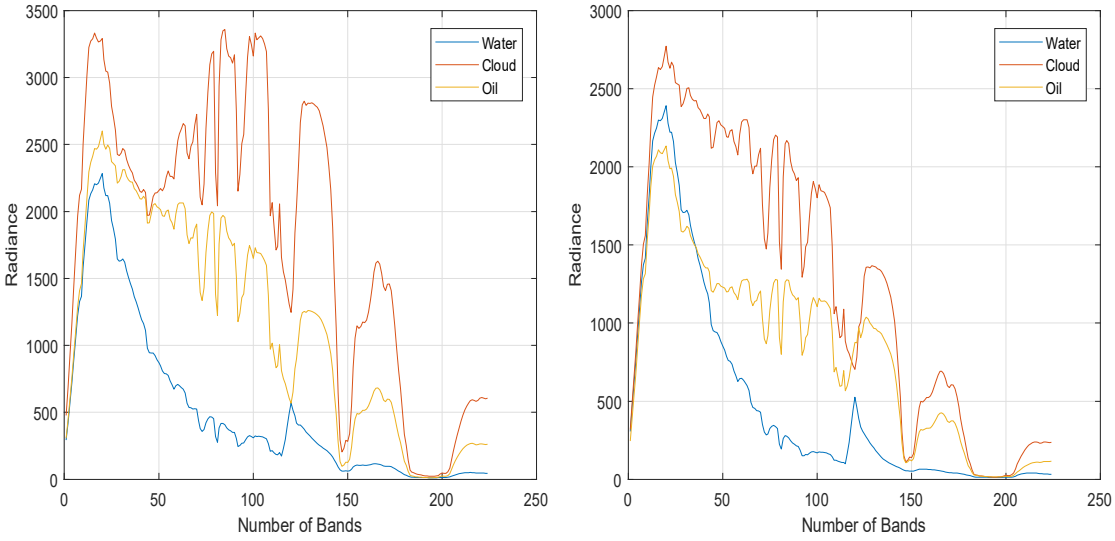
3.8.1 Linear Spectral Unmixing Results

The results of applying the linear spectral unmixing techniques including VCA and FCLS are first presented and compared.



(a) Abundance Estimation of REF Image (b) Abundance Estimation of SFIM Method

Figure 3.17: Per Band Radiance Abundance Estimation of REF & SFIM Methods



(a) Abundance Estimation of CNMF Method (b) Abundance Estimation of GFPCA Method

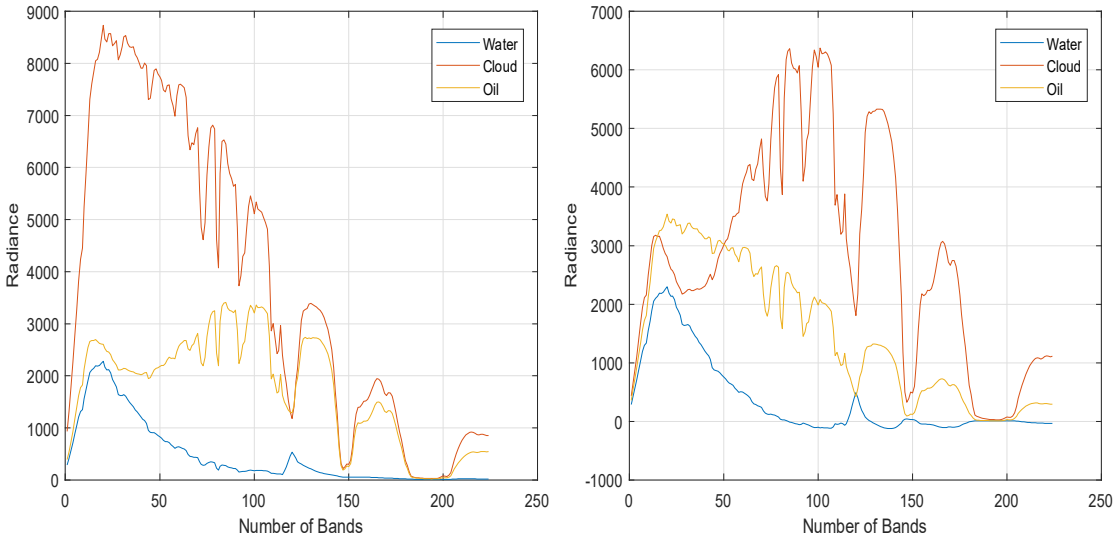
Figure 3.18: Per Band Radiance Abundance Estimation of CNMF & GFPCA Methods

Methods	VCA	FCLS	GBM	PPNM
REF	0.3740	11.6818	594.6941	134.0163
SFIM	0.2927	12.7537	568.7361	118.2755
CNMF	0.4681	27.9684	765.9219	99.8286
MTF-GLP-HPM	0.2940	25.6423	747.2242	105.3257
Bayesian Sparse	0.4487	12.0487	552.8717	124.9561
GFPCA	0.3717	11.1359	561.3685	141.2291

Table 3.3: Computational Complexity in secs of Unmixing process at 50dB SNR

Methods	VCA	FCLS	GBM	PPNM
REF	0.2770	27.6924	572.2157	84.3539
SFIM	0.3332	10.7886	542.2355	118.8568
CNMF	0.3313	27.9440	720.1771	97.4011
MTF-GLP-HPM	0.2770	10.5654	560.3979	123.0741
Bayesian Sparse	0.3136	11.3573	545.6049	129.0148
GFPCA	0.2918	20.5452	643.1309	125.4968

Table 3.4: Computational Complexity in secs of Unmixing process at 20dB SNR



(a) Abundance Estimation of MTF-HPM Method (b) Abundance Estimation of Bayesian Method

Figure 3.19: Per Band Radiance Abundance Estimation of MTF-HPM & Bayesian Methods

3.9 Summary

This chapter presents a quantitative and qualitative review of pansharpening methods using a reference hyperspectral image of the Gulf of Mexico oil spill acquired by the SPECTIR high spectral resolution sensor. A controlled simulation is carried out to generate high spatial res-

Methods	VCA	FCLS	GBM	PPNM
REF	0.8402	28.0592	601.1827	104.9304
SFIM	0.5451	40.2813	97.4032	77.4820
CNMF	2.3449	27.5251	604.4095	82.8860
MTF-GLP-HPM	0.6795	40.7462	85.4772	79.5284
Bayesian Sparse	24.5764	15.9938	549.0199	130.9064
GFPCA	0.9642	21.7123	599.0622	135.3783

Table 3.5: Computational Complexity in secs of Unmixing process at 10dB SNR

olution PAN image and a high spectral resolution HS image according to the Wald's protocol as illustrated by Fig 3.4. The simulated images are fused together using the different pansharpening methods described in section 3.2 and compared with the reference image using developed image fusion performance metrics according to Wald's protocol. A further analysis of the selected pansharpening methods is carried out for the oil spill application using linear and non-linear spectral unmixing methods described in section 3.7. Additionally, the computational times of the unmixing process are compared at different SNR values of 50dB, 30dB and 10dB for linear and non-linear unmixing approaches, respectively. Tables 3.2 and 3.3, shows the quantitative with regards to the performance evaluation measures discussed in section 3.5. In Fig. 3.5 - 3.11, the quick look representations of the results obtained by the pansharpening methods are displayed. The quick look images are generated by selecting the visible range bands comprising the R=705nm, G=625nm and B=435nm to form the RGB bands. Visual inspection of the images show that GS, PCA and GFPCA do not perform well for pansharpening of hyperspectral data with PAN data due to their inherent ability in rendering the spatial features of the image hence generating significant distortions to the spectral information.

However, GSA and Naive Bayesian approaches have shown promising results and took lower time computational time for the pansharpening process as shown in Table 3.3. Other approaches are found to perform better but take longer computational times to complete the pansharpening process. These methods include the Hysure and Sparse Bayesian approaches. This can be attributed to the ability of these approaches to utilise the spectral and spatial degradation applied to the estimated (fused) image. However, their performances may be affected by conditions that rely on parametric knowledge of the blurring approach and the corresponding spectral response of sensors. It is however shown in [53] that this problem can be solved by estimating the parameters along with the fused image. Or from the input images as a pre-processing step. In summary, the pansharpened methods evaluated in this chapter

shows that the multi-resolution analysis (MRA) methods are appropriate for fusing wide ranging images that include spaceborne high spectral images. On the other hand, Component Substitution (CS) methods shows defective fusion results with lower computation times. CNMF methods shows fair fusion results and lower computational time. Lastly, Bayesian Sparse method shows better performance result with higher computation processing time.

Spectral unmixing results can be translated in similar faith with the pansharpening results in this regard. The abundance estimation per band in radiance of three endmembers namely (water, cloud and oil) of the scene of interest are evaluated as shown in Fig. 3.17, 3.18, 3.19, respectively. Computational times are compared between the reference hyperspectral image and selected pansharpening methods from the groups (CS MRA, Bayesian, CNMF). As in pansharpening, the GSA method does not perform well when compared to the hyperspectral reference image. Sparse Bayesian method shows better performace at the cost of high computational time requirement for the unmixing process. Pansharpening can be seen from the results to be affecting the unmixing procedure and suggesting that fused images can actually improve the unmixing of oil spill in images.

Chapter 4

A Framework for Oil Spill Segmentation in Fused Synthetic Aperture Radar Images

4.1 Introduction

Existing oil spill detection and monitoring methods have relied on images from single remote sensing systems at a time to carry out oil spill analysis. However, when oil spill occur, different imaging sensors and systems of the electromagnetic spectrum such as SAR, HSI, PAN and MSI imaging sensors are deployed to generate image data in diverse forms, with each sensor having a unique advantage over the other. For instance SAR systems in the microwave region of the electromagnetic spectrum can obtain image data regardless of weather conditions or cloudiness, making it a classic tool for oil spill monitoring. In the optical region, Hyperspectral sensors can acquire images in contiguous bands that allow separation and easier identification of materials within a scene of interest. Multiple sensors make it possible to acquire diverse images of the same scene of interest, providing the scientific community with a wide range of data in different modalities to explore and make more meaningful analysis.

SAR systems are the most commonly used remote sensing systems utilised for oil spill detection. In addition to their all weather image acquisition potential, they also have a greater range coverage and a higher resolution compared to other sensors [5]. SAR is an active microwave sensor that acquires two dimensional (2D) images [36], its performance in detecting oil spill largely depends on sea conditions and the ability of oil films to decrease the backscat-

tering of the sea surface, resulting in the formation of dark patches. A general assumption among researchers in this area is that oil spill appear as dark formations on SAR images due to the dampen effect on capillary waves [18], [4]. However, not all dark formations on SAR images are oil spills, necessitating the need for robust detection technique and verification. In Fig 4.1, an example of two challenging dark formations for oil spill detection are presented

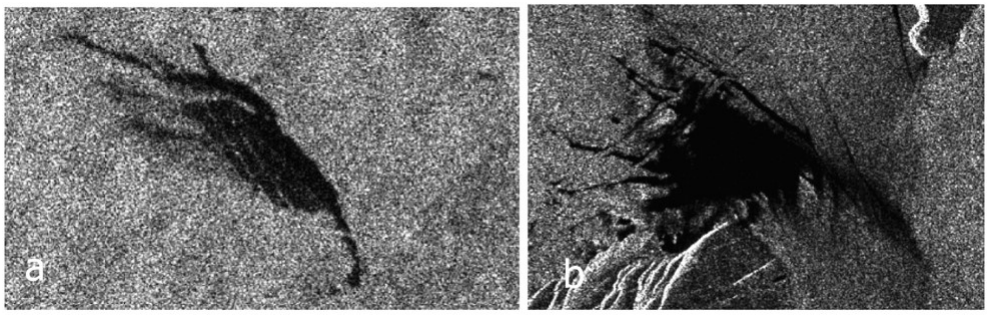


Figure 4.1: Example of challenging dark formations: (a) verified oil spill acquired 6/09/2005, Ancona Italy, (b) Verified look-alike acquired 25/08/2005, Otranto, Italy. [18]

Technological advancement have enabled more spaceborne SAR systems (e.g. the Sentinels) to be designed and launched, providing a wide range of data in multi modal configuration, including multi frequency SAR systems in C-band (e.g. EnviSat, RadarSat), X-band (Terra Sar-X) and L-band, multi frequency $(10^\circ - 70^\circ)$ and multi polarisations (dual, quad, etc.) in multiple resolutions features [126]. With the availability of multi-modal data, images utilised for oil spill monitoring can be improved in quality by combining data originating from different sources through image or sensor fusion, with the aim of obtaining information of greater quality than individual sensor data used conventionally in different studies [5, 13, 18, 35, 36, 127–129] to monitor and detect oil spill.

Combining SAR images or fusing SAR images, however, poses several challenges due to the multi-modal nature of the images to be fused, the differences in sensor characteristics and mode of obtaining the images. The first and most important task is that the individual images need to be co-registered in space and time. A suitable fusion algorithm needs to be choosen carefully to effectively and efficiently bring together the complimentary information from the individual modalities [8]. It is also important to define the level at which the data will be fused, since data fusion can be performed at different information levels including pixel, feature and decision levels [130].

This chapter proposes a framework that takes as input two SAR images from different modalities, pre-process the images, co-register the images and perform image fusion at pixel

level using Wavelets Transform (WT) and segment the dark areas using curve fitting and edge detection. This is the first time in our knowledge that such system framework is introduced for this application. Previous studies [12, 126, 131] have assumed the images are co-registered and aligned. In Fig 4.2 The framework of the system is shown

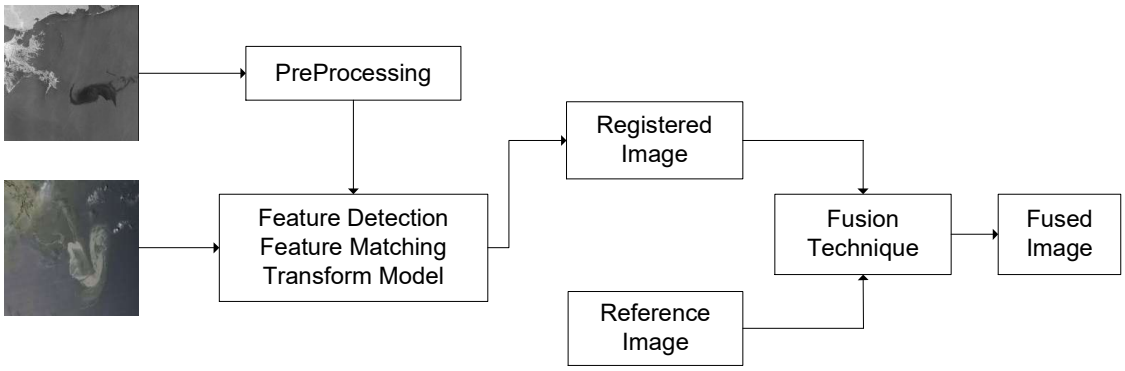


Figure 4.2: Developed SAR Image Fusion Framework

The proposed framework of SAR image fusion shown above comprises a pre-processing step followed by an automatic feature based image registration and then fusion at pixel level using wavelets

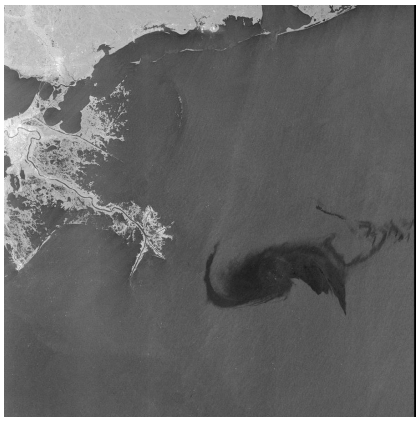
4.2 Dataset

In this section, the dataset used for the experiments in this chapter is described. The dataset utilised are real SAR images of the oil spill that occurred in the Gulf of Mexico as a result of a tanker blow-out on the 20th of may, 2010. During this spill, SAR systems including EnviSat and RadarSat acquired images for research and disaster management assessment. Images acquired by these sensors are utilised for the fusion framework illustrated in Fig 4.2, and subsequently segmentation of dark areas (Assumed oil spill locations).

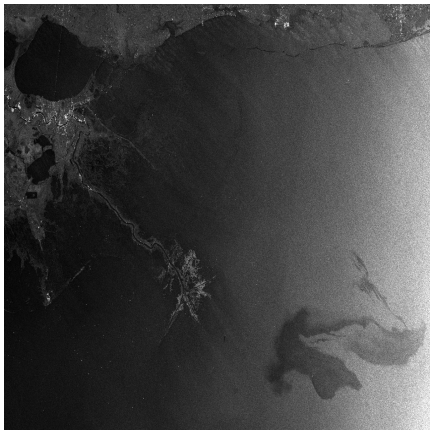
Table 4.1: Characteristics of the Dataset

Satellite	Instrument	Resolution	Band	Dimension	Date Acquired
Radarsat-2	SAR	$10m$	C	865×905	29/04/10
Envisat	ASAR	$150m \times 150m$	C	930×1271	26/04/10

The RADARSAT-2 SAR image is a ScanSAR wide single beam mode acquired on April



(a) ScanSAR RadarSat Image



(b) ASAR EnviSat Image

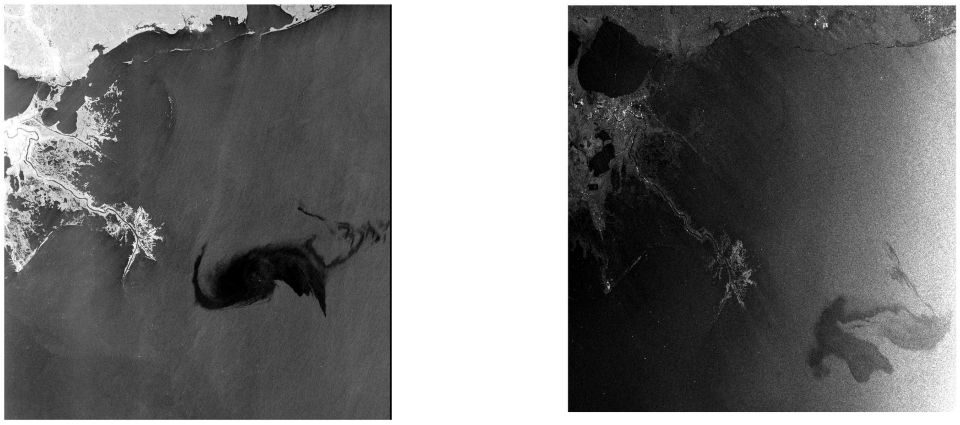
Figure 4.3: SAR Images acquired by RadarSat and EnviSat Satellites

26th,2010. The ScanSAR mode provides images with very wide swaths in a single pass of the satellite[127]. The original size of the SAR image is 865×905 with a spatial resolution of $100m \times 100m$. The satellite operates in C band. The Envisat image from the ASAR instrument on board ESA's Envisat Satellite is of size 930×1271 with a spatial resolution of $150m \times 150m$ in the wide scan mode and it also operates in the C-band. Both images represent the Gulf of Mexico oil spill scene and were acquired on the 26th and 29th of April 2010, respectively. For the purpose of this experiment, both images were resized to 512×512. The resized images are shown in Fig. 4.3.

4.3 Image Pre-Processing

SAR images contain speckle noise and often have poor visualisation. The pre-processing step helps to reduce this noise by filtering the images and improving the visualisation by enhancement to obtain the best possible image perception. A Gaussian Filter, that allows the preservation of edges, texture and fine details of the image is used to reduce the effect of speckle noise. Firstly, the image features are sharpened and the contrast increased. The moving kernel of the Gaussian filter is then utilised to reduce speckle effect using its hump shape [129]. Additionally, the filter preserves structural and textural features and can improve the overall quality of the images to enhance better analysis. The Gaussian filter is mathematically defined as:

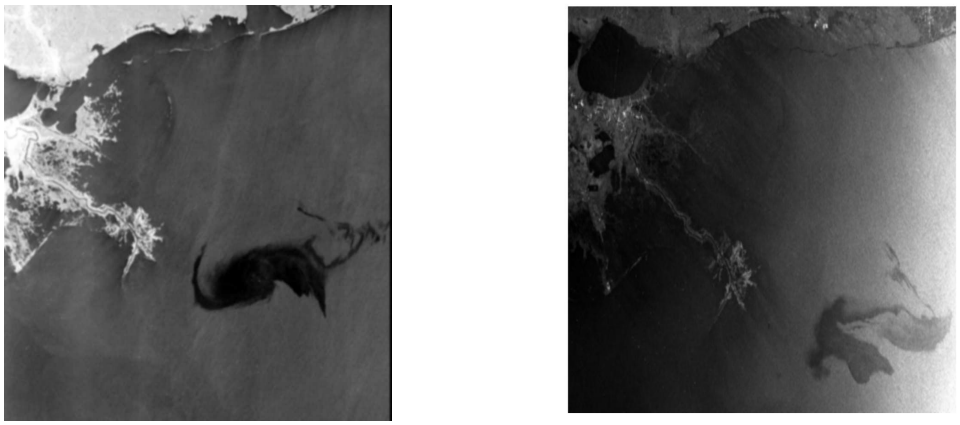
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \ , \tag{4.1}$$



(a) Adjusted ScanSAR (b) Adjusted ASAR

Figure 4.4: Adjusted SAR Images with Enhancement

where σ is the standard deviation of the distribution, x and y are the directions of the moving kernel of the filter. The Gaussian filter is applied to the SAR image after the contrast enhancement step, as a pre-processing step to improve the image quality and to preserve the texture of the dark areas (assumed oil spill positions).



(a) Denoised ScanSAR (b) Denoised ASAR

Figure 4.5: Denoised SAR Images Using Gaussian Filter

4.4 Image Registration

The mode of acquisition of images from different sensors differs due to the multi-modal nature and calibration of sensors. To combine images of a same scene acquired by different

sensors, a vital step of aligning the images need to be considered.

Image Registration is a crucial step in most image analysis tasks where the final information is gained from the combination of various data sources like in image fusion [132]. The aim of image registration is to map the similarity between images acquired of a same scene, at different times and from different angles using different imaging sensors or cameras [133]. In remote sensing, image registration is utilised to perform tasks that include environmental monitoring, change detection, weather forecasting and to improve resolution of images. Additionally, the registration of images can establish the correspondence between images and determine the geometric transformation that aligns one image to the other [9]. The registration of images can be done in either of two forms area based or feature based. The results of enhancing the images and filtering of speckles is presented in Fig. 4.4 and Fig 4.5, respectively.

The process of registering two or more images is done in four (4) steps basically, and majority of registration methods proposed in the literature adopts these steps;

- Detection of Image Features: Detecting features in an image can best be described as locating the interest points of that image and the objects of the image can be described using the features detected. Image features could include edges, lines, regions or contours and can be grouped into Salient or distinctive features. Detecting features of an image is an important step of the registration process. In detecting features, image control points (centers of gravity, lines and distinctive points) are identfied.

- Feature Matching: Once features are detected, the next step is to match corresponding features across the images to be registered. In other words the correspondence between the images is established. Feature matching entails spatially aligning the images acquired at different times of a same scene [134].

- Transformation Model Estimation: In achieving the matching of features, the mapping function uses a set of parameters that need to be estimated to make the matching process more robust and accurate. These parameters are computed by means of the established feature correspondence during matching.

- Resampling and Transformation: The final step in registering images is to resample the images and geometrically transform them using the mapping function described in the previous step. Resampling the images involves the computation of intensities within the non-integer coordinates through the utilisation of suitable interpolation.

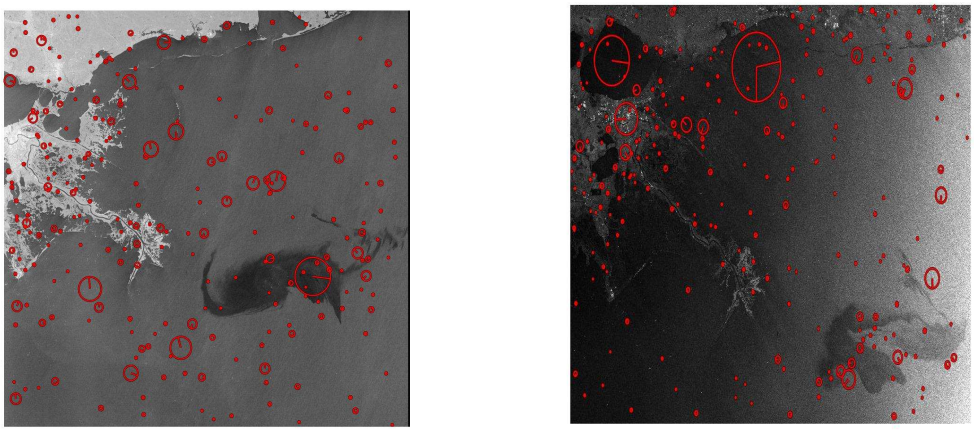
For efficient registration of images, a careful selection of appropriate methods for each step described above needs to be considered. In detecting features for example, the question of what features should be detected or extracted is very important. Generally, image features can be grouped into two namely; Local and Global features that describe the contents of the image [134]. Global features describe the image wholly using a singular property of the image for example its color or texture that involves all pixels of the image. On the other hand, Local features put into consideration key points otherwise known as interest points. In local features, the number of keypoints n detected by the detection algorithm is directly proportional to the number of vectors n describing each property of the image detected e.g. shape, color, orientation, etc. In the literature, several algorithms such as Scale Invariant Transform (SIFT), Speed Up Robust Feature (SURF) etc. have been developed to extract, detect and match features of images [135]. In this framework, an automatic feature based image registration is considered. Firstly, the SIFT algorithm is utilised to detect features from SAR images obtained by different sensors of a same scene.

4.5 Extraction and Matching of Features

The SIFT algorithm was first introduced in [136], to extract distinctive invariant features from images that can be used to perform reliable matching of features from images acquired from different angles of an object or scene in an image. The SIFT model employs a feature based approach to detect invariant features from images and match corresponding ones. When compared to other algorithms like SURF [137], the features detected using SIFT have advantages that include being robust, features detected with SIFT are invariant to scaling, rotation distortion to affinity and noise. To extract SIFT features, four computational steps need to be satisfied, these steps qualify a feature as a SIFT. In the next section, the four (4) computational steps in actualising SIFT feature detection are described. Fig.4.6. shows extracted SIFT features on the SAR images to be fused.

4.5.1 Scale Space and Extrema Detection

In the first step of computation for SIFT, the algorithm searches through all scales and the locations in the image through the use of a difference-of-Gaussian function [136] to locate those interest points of the image that are invariant to both scale or orientation of objects in the image. Firstly, locations and scales that can frequently be allocated over different angles or



(a) SIFT Features ScanSAR

(b) SIFT Features ASAR

Figure 4.6: Extraction of SIFT Features

viewpoints of an object in the image are identified by searching for features that are relatively invariant to change of scales in the image using the scale space function introduced in [138]. A widely adopted assumption is that the only possible scale-space kernel is the Gaussian kernel [139].

Hence, the scale-space extrema can be defined as a function, $L(x, y, \sigma)$, a product of convolution of the variable-scale Gaussian here defined as $G(x, y, \sigma)$, with the image $I(x, y)$. The scale-space extrema can be defined mathematically as

$$L(x, y, \sigma) = G(x, y, \sigma) * \mathbf{I}(x, y), \tag{4.2}$$

where the variables x and y are the spatial plane coordinates of the image and $*$ is the convolutional operator in x and y, respectively. The variable scale Gaussian can also be defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{4.3}$$

To detect stable interest points, [140] proposed a difference-of-Gaussian function which is computed by finding the difference between two scales within a neighbourhood that are seperated by a constant multiplicative factor usually set to $\sqrt{2}$ [137]. This is mathematically expressed as

$$
\begin{aligned}
D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\
&= L(x,y,k\sigma) - L(x,y,\sigma).
\end{aligned}
\tag{4.4}
$$

The advantage of using this function $D(x,y,\sigma)$ includes its computational efficiency when we consider that the images will still undergo a further scale-space feature description, thus allowing the function $D(x,y,\sigma)$ to be computed by a simple subtraction of the processed images defined here as L. In Fig 4.7, the difference-of-Gaussian is described.
In [138], it is shown that $D(x,y,\sigma)$ gives a close estimation to scale-normalised Laplacian of Gaussian $\sigma^2\nabla^2 G$. Additionally, to locate actual invariant scale in the image to be registered, [138] pointed the need for normalisation of the Laplacian with the factor of σ^2. Therefore, to identify the local maxima and minima of the function $D(x,y,\sigma)$, each interest point is compared to its neighbourhood which is usually eight (8) in the image and nine neighbours in the scale and below as shown in Fig 4.7.

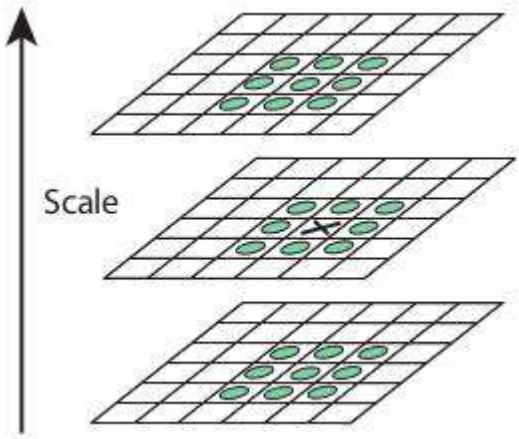


Figure 4.7: Difference-of-Gaussian Pyramid [136]

4.5.2 Localisation of Keypoints

The next computational step in actualising SIFT is keypoint localization. This follows from identification of interest points where pixels are compared with their neighbourhood. Keypoint localisation entails assigning the interest point to a nearby data to allow for location, scaling and the rotation of its principal curvatures [136]. This step allows the elimination of noise-sensitive points and also non-edge points, with the aim of improving stability of the

feature matching step and improve the feature's immunity to noise [8]. The work of [141] further improves the step by determining the interpolated location of the maximum using a three dimensional (3D) quadratic function fitted to the local sample points used in [140], this approach has shown significant improvement to the stability of detected features and matching.

4.5.3 Orientation Assignment

Assigning stable orientation to detected keypoints enables the keypoint descriptor to be described more accurately in relation to the orientation of the features. The advantage of this step is that steady features and stability to image rotation is achieved. Consider the Gaussian smoothed image L selected from the scale of a keypoint, and compared with the nearest scale, hence allowing all computations to be performed in an manner that is scale invariant, for every sample of the smoothed image $L(x, y)$ at the selected scale, the magnitude of the gradient defined as $m(x, y)$ and its orientation $\theta(x, y)$ are computed with the pixel differences here defined as

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}\,, \tag{4.5}$$

$$\theta(x,y) = tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))\,, \tag{4.6}$$

where L represent the smoothed image, the variables x and y are the spatial coordinates of L. In (4.5) and (4.6), the absolute value and direction of a pixel's gradient are defined. Additionally, the scaling of L defines also the scale of each keypoint within the image, respectively. Orientation histograms [142] are formed from sample points over the region of the keypoint. The histogram contains 36 bins that cover the circle range of orientation corresponding to 360 degrees. In the histogram, its peak correlate to stable directions of the neighbourhood gradients [136].

4.5.4 Keypoint Descriptor

The aim of keypoint description is to ensure that local image regions are invariant to changes that include noise, angle of view change and illumination. A possible approach is to sample local image intensities over the keypoints using a suitable scaling that enables invariance and cohesion to be achieved. Firstly, magnitude of the gradients and orientation at each sample

point in the neighbourhood are computed using (4.6) and (4.7) and shown on the left hand side of Fig 4.4. and their weight taken using a Gaussain window as illustrated by the blue circle overlaid. Next, the samples are added to form histogram of the orientations as stated above; summarising the contents of an 8×8 region shown on the right image of Fig 4.4, where the length of each arrow shown represents the sum of the magnitude of gradients close to its direction within the region. Orientation histograms formed are integrated to a vector of fixed length that is normalised making it invariant to changes described above and achieving invariance, hence forming the SIFT descriptor [137]. Finally, a 2×2 array of descriptor is computed from the 8×8 set of image samples. The approach adopted in this framework is a set of 4×4 descriptors computed from a 16×16 sample array.

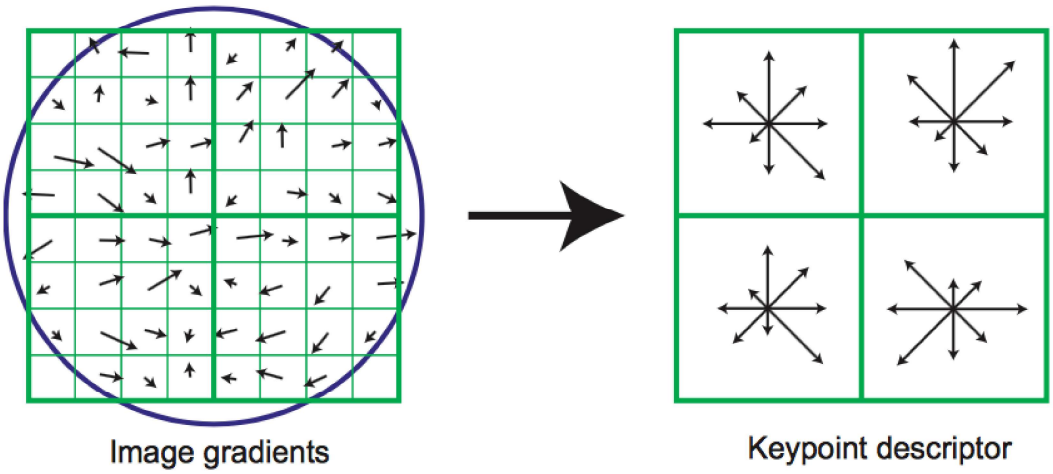
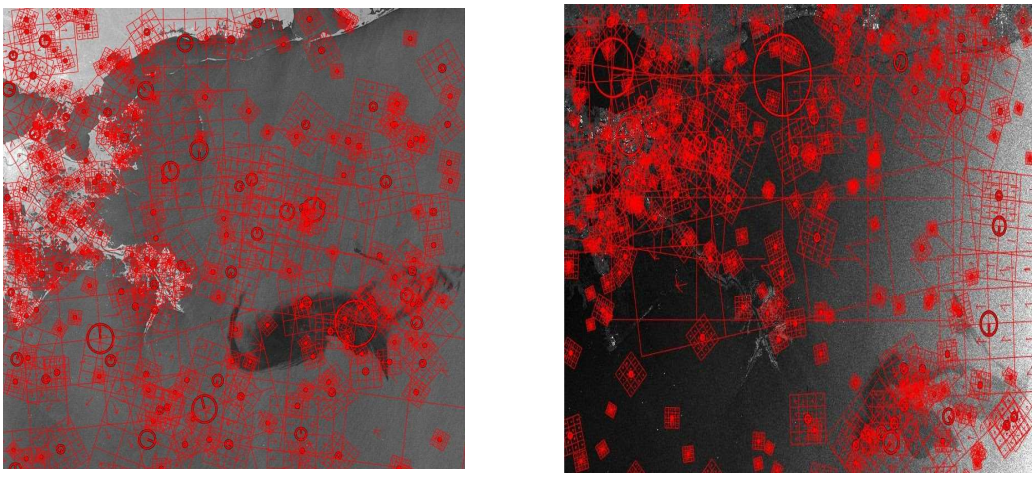


Figure 4.8: Keypoint Descriptor [136]

4.5.5 Feature Keypoints Matching

In the matching step, corresponding features are located and matched. Firstly, best candidate match for each keypoint are found using the nearest neighbourhood identification. The database of keypoints created in the extraction step is utilised for this purpose. Features without a corresponding match are discarded by comparing neighborhood distances between the closest and the second closest pixel to ensure that correct matches have the nearest neighborhood matches. While this could be a trivial step, because features in one image considered may not have corresponding matches in the other images. For this reason, methods such as the Random Sample Consensus (RANSAC) are used. For the feature matching step, the



(a) Keypoint Descriptors on ScanSAR

(b) Keypoint Descriptors ASAR

Figure 4.9: SIFT Features and Descriptors

RANSAC algorithm is adopted in this chapter with the aim of achieving efficient matching of extracted features from the SAR images to be fused. It has the ability to reject all error matches so that only tentative matches are retained for registration.

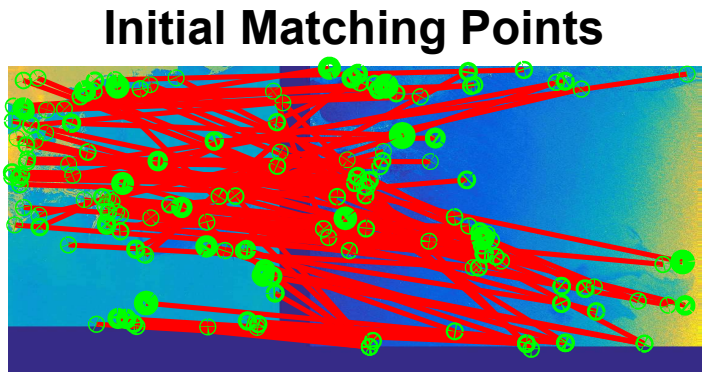


Figure 4.10: Initial Matching Points between Images

4.5.6 Transformation Estimation

RANSAC is first introduced in [143], it is a robust transformation estimation algorithm [142], that can handle the mapping of features in the presence of many outliers. The method has been used to map image features in automatic image registration tasks [144]. The algorithm works by finding the homography of each image pair to be registered, enabling tentative

matching and identification of inliers while discarding the outliers [8]. RANSAC generally utilises the resampling technique by generating candidate solutions using the minimum number of data points required to estimate the model parameters [145]. SAR images are likely to contain high proportion of outliers due to the effect of speckle noise, although, the effect of this has been reduced with the Gaussian filter during pre-processing stage, a robust fitting of features for efficient registration is neccessary.

The distinguishing factor between RANSAC and most predominant resampling methods is in its ability to use much information from the image data to obtain the initial solution before it proceeds to eliminate the outliers. To achieve this, a dataset $\mathbf{M}$ is created consisting of all matched points and N number of data points. An affine transformation model $\mathbf{A}$ is then established between all the feature points. A subset S of four (4) matched feature points are randomly selected to compute the model ν using the selected data points. Finally, the image registration model between the two SAR images can be represented by affine transformation as

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = S \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} , \tag{4.7}$$

where the variables (x_1, y_1) and (x_2, y_2) represent the control points coordinates in the SAR images to be registered. t_x and t_y are the translational values in the geometric coordinates x and y of the images, S is the scaling factor, while θ is the angle of rotation, respectively. When making iterations to eliminate outliers, a common practice is to choose a higher number to ensure that the probability of that one of the random sample sets is an inlier. The probablity ρ is usually set to 0.99 [145], so that the probability τ of a sample point being an inlier is given as

$$1 - \rho = (1 - \tau^m)^N , \tag{4.8}$$

and

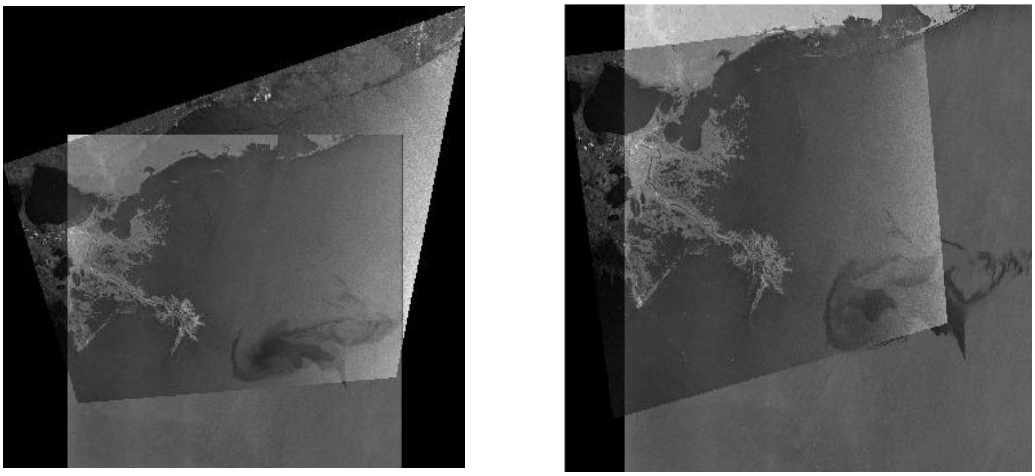
$$N = \frac{log(1-\rho)}{log(1-(1-v)^m)} , \tag{4.9}$$

where ρ represent the probability of a selected set from a random sample is not an outlier. N

is the number of iterations for a minimum number of points, here represented as m and τ is the probability of a sample point being an inlier. The algorithm can further be summarised in five (5) steps

Algorithm 4.1 Transformation Estimation Algorithm

1: Determine the parameters of the model by selecting randomly the minimum number of datapoints
2: Calculate the parameters of the model using (3.7)
3: Determine the number of points from database of features using an established tolerance γ,
4: **If** number of inliers ϱ compared to η (total number of points in the set) is > ϕ (threshold),
5: **DO**
$1 - \rho = (1 - \tau^m)^N$
Otherwise
Repeat 1-4
end



(a) Registered Image

(b) Registered Image Varying Points

Figure 4.11: Image Registration Results

4.6 Wavelet image Fusion Approach

Once the images are registered in time and space, the images can then be combined using image fusion techniques. Image fusion enables the enhancement of images by combining images acquired by different sensors with different sensor characteristics [146], with the aim of obtaining an image with combined characteristics and features of the individual images combined. Image fusion can be performed at different levels including pixel, feature and decision

levels, depending on the application in consideration. In detecting oil spill for example, [147] established that fusion techniques are effective for detecting oil spill with SAR images. Additionally, [126, 131] explored image fusion techniques to improve oil spill detection by fusing different bands of SAR images using multi frequency fusion methods by fusing SAR images with HSI images in a multi modal image fusion technique at pixel levels, with both studies reporting significant improvement in classification of oil spills and look alikes. In fusing SAR images, consideration should be given to the characteristics of the different imaging sensors since SAR images are formed from the backscattering process as a result of the interaction between the microwave sensor and ground features, resulting in images that are relatively rich in high frequencies within the frequency spectrum. For this reason, [49] suggested that fusion techniques based frequencies have the advantage to combine information per the nature of the SAR images. In this chapter, the wavelet based image fusion approach is considered due to its robustness and its ability to provide a representation of directional information with the creation of sub-bands of the image ie. High-low, low-high, and high-high sub bands [57].

The advantage of wavelet based image fusion is in its ability to decompose images into approximations, hence preserving information contained in images. Additionally, its multi scale approach makes it suitable for images with different resolutions. Furthermore, approximations formed can be combined to form new ones such that all information are captured accordingly. In the next section, a brief background of wavelet is given to introduce the reader with the fundamentals of this method.

4.6.1 Wavelets Transform

In this section a quick background to wavelet transformation used for viewing or processing images at multiple resolutions is given. Recall from image registration, that multi resolution images can be formed from a pyramid of images where at each level the differences between an image at that level and the predicted image from the preceding level are stored. The image, can be reconstructed by adding up the differences from each level of the pyramid. With wavelets, multi resolution images can be generally represented and analysed by decomposition.

Given an image $I(x, y)$ with size $m \times n$, its forward discrete transform given as $F(a, b, \dots)$

can be described typically in the form

$$F(u, v, \dots) = \sum_{x,y} \mathbf{I}(x, y) k_{a,b,\dots}(x, y) \ , \tag{4.10}$$

where the variables x and y define the spatial coordinates of the image M is the size in rows, N size in columns and $u, v, \dots$ are the transformation domain variables. Given the transformation and its domain $T(u, v, \dots)$, the image $I(x, y)$ can be reconstructed using a generalised inverse discrete transform (IDT) described in [148].

$$\mathbf{I}(x, y) = \sum_{a,b,\dots} F(a, b, \dots) q_{a,b,\dots}(x, y) \ , \tag{4.11}$$

here $q_{(a,b,\dots)}$ and $h_{(a,b,\dots)}$ represents the forward and inverse transformation kernels, respectively. The kernels defines the behaviour, computational complexity and effectiveness of the transformation. The coefficients of $F(a, b, \dots)$ can be viewed as the expansion coefficients of a sequence of expanding I given as

$$q_{a,b}(x, y) = k^*_{a,b}(x, y) = \frac{1}{\sqrt{MN}} e^{j2\pi(ax/M + by/N)} \ , \tag{4.12}$$

where $j = \sqrt{-1}$, and $*$ is the complex conjugate operator, with $u = 0, 1, \dots, M-1$ and $v = 0, 1, \dots, N-1$. Additionally, a and b (transform domain variables) represent the vertical and horizontal frequencies, respectively. Furthermore, the kernels can be seperated, and they are both orthogonal and normalised (orthonormal) such that

$$q_{a,b}(x, y) = q_a(x) q_b(y), \tag{4.13}$$

$$q_a(x) = \frac{1}{\sqrt{M}} e^{j2\pi ax/M} \qquad \text{and} \qquad q_b(y) = \frac{1}{\sqrt{N}} e^{j2\pi ay/N} \ . \tag{4.14}$$

This is also orthonormal because:

$$\left\langle q_r, q_s \right\rangle = \partial_{rs} = \begin{cases} 1 & r = s \\ 0 & \text{otherwise} \end{cases} . \tag{4.15}$$

In (4.15) $\langle\rangle$ defines the inner product and describes the orthonormal behaviour of the kernel. The separability behaviour of the kernels enables a faster computation of the 2D transform allowing the row-column or column row passes of a 1-D transform to be used. Again, Due to the orthonormality, the forward and inverse kernels are the complex conjugates of themselves.

The difference therefore between the Discrete Wavelet Transform (DWT) and the Discrete Fourier Transform (DFT) is that DFT is expressed by the two general equations that are based on its single pair transformation kernels while DWT refers to the class of transformations that are different not only in the transformation kernels but also in the underlying identity of the functions [148], i.e whether those functions form orthonormal or biorthogonal basis, in addition to the number of different resolutions created by decomposition or computation of the image signals. DWT therefore consists of diverse and distinctive transformations that cannot be described by a single function or equation. However, each DWT can be defined by its transform kernel pair or the set of parameters that actually defines the pair, and the transformations are related through the basis of expanding their functions that result in small waves with varying frequencies. These small waves, can further be described by their properties. The first of which is:

Property 1: Separability, Scalability and Translatability

$$\psi^H(x,y) = \psi(x)\varphi(y) , \tag{4.16}$$

$$\psi^V(x,y) = \varphi(x)\psi(y) , \tag{4.17}$$

$$\psi^D(x,y) = \psi(x)\psi(y) , \tag{4.18}$$

where $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$ represent the horizontal, vertical and diagonal wavelets, respectively and a seperable 2D scaling function given as:

$$\varphi(x,y) = \varphi(x)\varphi(y) . \tag{4.19}$$

Additionally, each of the properties listed above (2D functions) is a product of 1D real,

square-integratable scaling and wavelet functions defined as:

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)\ , \tag{4.20}$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)\ , \tag{4.21}$$

where the translation k determines the location of the 1D functions along the $x$axis and the scale j determines the thickness of the functions along x, while $2^{j/2}$ determines the amplitude of the waves, $\psi(x) = \psi_{0,0}(x)$ and $\varphi(x) = \varphi_{0,0}(x)$ are the related scalings and integer translates of ψ defined as the mother wavelength.

Other properties relate to the multi-resolution compatibility that relies on the scaling function φ satisfying the following conditions.

- $\varphi_{j,k}$ is orthogonal to its integer translates [148].
- $I(x) = 0$ is the only function that can be represented at every scale
- Functions resulting from the expansion of $\varphi_{j,k}$ at lower resolutions are contained in the functions resulting from higher φ. where φ represent the scaling of higher and lower signals.
- Any of such functions can be defined by a random precision as $j \to \infty$.

4.6.2 SAR Image Fusion with Wavelets

$\mathbf{I}_1(x, y)$ and $\mathbf{I}_2(x, y)$ represent two co-registered intensity SAR images acquired by different sensors over the same locations and at different times T_1 and T_2, respectively. The aim is to fuse these images using the Discrete Wavelet Transform (DWT) earlier introduced and to obtain a fused image $\mathbf{I}_F(x, y)$ with greater quality than the individual images [149]. The proposed framework of fusing the imagesusing wavelets decomposition is illustrated in Fig. 4.12.

The DWT based fusion technique improves the spatial resolution of the fused image while preserving the colour appearance for further interpretation and analysis [54]. Preserving the colour information is critical in oil spill segmentation since oil spill is assumed to appear as dark formations on SAR images. Additionally, SAR images often contain some peculiarities [150]. The first step is to compute in each image its decomposition into coefficients of

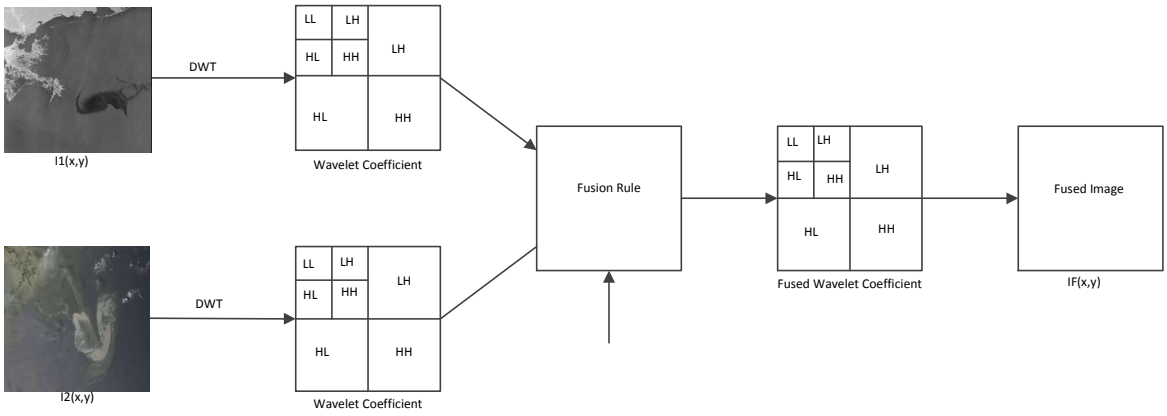


Figure 4.12: Proposed SAR wavelet based Image Fusion

multi-resolution images by converting the image from spatial to the frequency domain on downsampled images and applying wavelet filters L for lowpass filter and H for high pass filters to the rows x and columns y of the images $\mathbf{I}_1$ and $\mathbf{I}_2$, respectively. Since SAR is an active microwave sensor that acquires images in 2D [36], the coefficients obtained from scaling and wavelet functions will be

$$\varphi_{LL}(x,y) = \varphi(x)\varphi(y), \quad \psi_{LH}(x,y) = \varphi(x)\psi(y)\ , \tag{4.22}$$

$$\psi_{HL}(x,y) = \psi(x)\varphi(y), \quad \psi_{HH}(x,y) = \psi(x)\psi(y)\ . \tag{4.23}$$

Equations (4.24) and (4.25) represent the introduces scaling and wavelet functions as the tensor products of 1D aggregates of the introduced scaling and wavelet functions earlier defined in (4.20) and (4.21), respectively.

The filtering and down-sampling of each image's rows and columns will result in coefficient matrices given as $I_L(x,y)$ and $I_H(x,y)$. On the vertical end, sub band images are created in the first decomposition level represented by $I_{LL}(x,y)$, $I_{LH}(x,y)$, $I_{HL}(x,y)$ and $I_{HH}(x,y)$ as shown in Fig 4.5. The sub-images created represent the approximation coefficients, in I_{LL}, the horizontal detail coefficients in I_{LH}, the vertical detail coefficients in I_{HL} and the diagonal coefficients in $I_{HH}(x,y)$ [151]. The next level decomposition is done only on the $I_{LL}(x,y)$ to obtain further coefficient matrices and so on until the desired decomposition is achieved.

4.6.3 Merging Decomposed Wavelets

The decomposition of individual images at the first phase of processing result in the four sub-bands as described earlier. In Fig 4.13, the decomposition is illustrated where the smaller square boxes represent the spatial pixels that are utilised for the fusion process, forming the base of the fused image $I_F(x,y)$ construction. The hierarchical pyramid nature of DWT decomposition allows the frequency bands to decrease further as the decomposition level is increased. In other words, the sizes of the frequency bands is directly proportional to the decomposition level. As an example, a DWT process with 4 decomposition levels will have $A = 3(4) + 1$ frequency bands and the transformation with V levels of decomposition will always have LL^V since for each decomposition there is only a single LL low-low frequency while the rest contain a high frequency signal as shown in Fig 4.13. It is also important to mention that the fusion of images occurs only on same resolution level even though the decomposition of images can be at varying decompositions.

The fusion approach in this chapter begins with the simplest form of DWT the Haar transform and then compare results with the Daubechies approach. To achieve the fusion of SAR images with DWT, there is need for the fusion rule that defines the fusion of the decomposed coefficients at a given resolution. Recall that the fusion is done only on corresponding bands with the same resolution level [55]. While different fusion rule exist, which can depend on the number of images to be fused. For this application, two SAR images are considered for the DWT fusion process.

Given the source images $\mathbf{I}_1(x,y)$ and $\mathbf{I}_2(x,y)$ to be fused, and the fused image to be recovered as $\mathbf{I}_F(x,y)$, the multi scale decomposition described earlier be denoted as D_I and the Activity level C_I, if $Q = (m,n,k,l)$ where Q is the index corresponding to the multi scale decomposition, the variables m and n are the spatial location of a frequency band, k is the level of decomposition and l the frequency band of the multi scale description. Hence, the multi scale decomposition D_I and the activity level C_I can further be described as multi scale decomposition value and activity level of the equivalent coefficient, respectively. To measure the activity level C_1 at Q, where Q defines the location of a coefficient, the coefficient based activity level is utilised in this chapter; in which each coefficient is considered individually and is defined by

$$C_I = |D_I(Q)| \quad \text{or} \qquad C_I(Q) = (D_I(Q))^2 \,, \tag{4.24}$$

where C_I is defined by the absolute value of the corresponding coefficient from the multi

scale decomposition [55]. In the next step of the fusion process, the coefficients are grouped or combined using a fusion rule. The select maximum approach. It is important to mention here, that the coefficients with high frequencies are considered since the LL band contains only positive values while other high frequency bands are varying around 0. Hence in this approach, transform with higher values that relate to the salient features including edges and boundaries are selected from the activity value computation described earlier, while other transform are discarded. From the activity measure, consider

$$\mathbf{I_F}(Q) = \mathbf{I}_i(Q) \ , \tag{4.25}$$

where $\mathbf{I}_i$ represent the source images to be fused depending on which has the higher transform value or satisfies the select maximum rule. The select maximum can then be expressed as

$$\mathbf{I}_i(Q) = max(\mathbf{I}_1(Q), \mathbf{I}_2(Q)) \ , \tag{4.26}$$

In the next step step of the fusion process, the inverse of the DWT is taken to give the fused image. This is mathematically defined as:

$$\mathbf{I_F}(x,y) = IDWT[\Phi DWT(\mathbf{I}_1(x,y), DWT(\mathbf{I}_2(x,y))] \ , \tag{4.27}$$

where Φ defines the fusion rule in this case the select maximum earlier defined in (3.26).
The fusion result is shown in Fig. 4.14 and Fig. 4.15, respectively. The decomposition levels are varied from 1-4 and the result are compared using image fusion performance metrics acording to the Wald's protocol.

4.7 Image Fusion Performance Metrics

In this section, we evaluate the performance of the fusion technique and test the fused image against the individual images using established quality assessment measures described in the literature. According to [53], several of these measures have been defined to determine the quality of images from fusion techniques and the similarity that exists between fused images and individual images fused. A conventional approach for such quality test is the Wald's protocol where the model is based on the properties of consistency and synthesis of the images [8]. Given the individual images $\mathbf{I}_1(x,y)$, $\mathbf{I}_2(x,y)$ and the generated fused image

(a) Haar Transform 1-Level Decomposition

(b) Haar Transform 2-Level Decomposition

Figure 4.13: Fusion Results at 1 & 2 Decomposition Levels with Haar Transform

(a) Haar Transform 3-Level Decomposition

(b) Haar Transform 4-Level Decomposition

Figure 4.14: Fusion Results at 3 & 4 Decomposition Levels with Haar Transform

$\mathbf{I_F}(x, y)$, the quality of fused image is tested using firstly correlation coefficient (CC) between pixels to determine the geometric deformation in the fused image when compared with any of the individual images. The correlation coefficient is here defined as

$$CC(\mathbf{I}_i, \mathbf{I}_F) = \frac{\sum_{mn}(\mathbf{I}_i - \bar{\mathbf{I}}_i)(\mathbf{I}_F - \bar{\mathbf{I}}_F)}{(\sum_{mn}(\mathbf{I}_i - \bar{\mathbf{I}}_i))^2(\sum_{mn}(\mathbf{I}_i - \bar{\mathbf{I}}_i))^2}\ , \tag{4.28}$$

where I_i represent a pixel of the reference image (either of the individual images) with size $m \times n$ and $\bar{I}_i$ is its sample mean. Similarly, I_F is for the fused image with its sample mean given as $\bar{I}_F$. Secondly, we also test the quality using Spectral Angle Mapper (SAM) that

computes the spectral angle difference between pixels of the fused image and the source or reference images expressed here as

$$SAM(\mathbf{I}_i, \mathbf{I}_F) = \frac{1}{n}\sum_{j=1}^{n} SAM(i_{ij}, f_j) \tag{4.29}$$

$$SAM(\mathbf{I}_{\{i\}}, F_{\{i\}}) = arccos\left(\frac{\langle I_{\{i\}}, F_{\{i\}}\rangle}{\|\mathbf{I}_{\{i\}}\| \, \|F_{\{i\}}\|}\right) \tag{4.30}$$

where i_j define the jth columns of the individual images and the fused image, $\langle I_{\{i\}}, F_{\{i\}}\rangle$, the inner product and $||.||$ the norm, respectively. Finally, the RMSE is also evaluated and here defined as:

$$RMSE = \left(\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[F(i,j) - I(i,j)]^2)}{m \times n}\right), \tag{4.31}$$

where $F(i,j)$ defines the pixels of the fused image, $I(i,j)$ pixels of the reference or source image and $m \times n$ the size of the images.

Fusion Method	CC	RMSE	SAM
DWT Level1	0.2306	0.3609	11.3992
DWT Level2	0.2416	0.3600	11.3822
DWT Level3	0.2384	0.3603	11.3950
DWT Level4	0.2387	0.3603	11.4005

Table 4.2: Quality Measures of the Fusion Result

In Table 4.2, the results of assessing the quality of the fused images are presented. The performance assessment is to measure the degree of information retained in the fused image when compared with the individual images fused in terms of spatial and geometric distortions. CC as defined earlier, compares the geometric deformation of the fused when compared to the individual images fused with results nearer to zero expressing good performance. From the figures in Table 4.2, the results of CC shows the first level of decomposition with least performance since the ideal value of CC is 1, signifying that the correlation between images gets better as the images are decomposed further, hence the level of decomposition needs to be chosen sensibly. On the other hand, RMSE values shows the second level of decomposition performing slightly better than the third and fourth levels. The ideal value of this measure is zero.

4.8 Segmentation of Dark Areas

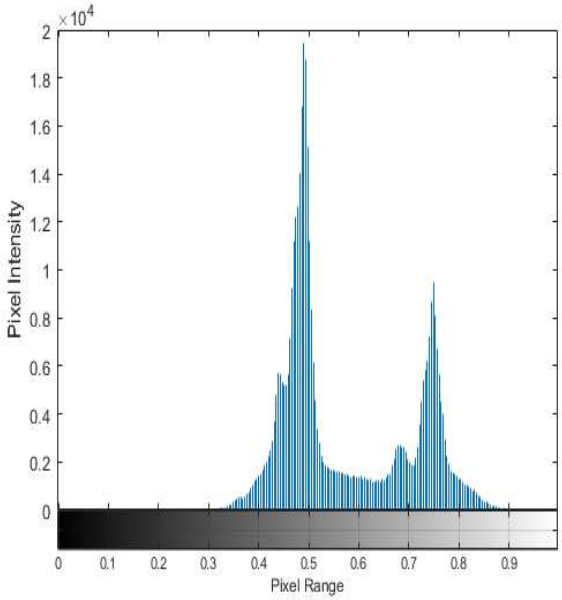
The next step of the framework involves segmenting the dark areas (assumed oil spills). Segmentation allows the partition of an image into meaningful components. The aim of this step is to subdivide the fused image into its constituents object and regions so that objects within the image can easily be identified. In this case, the interest is to efficiently segment the dark areas of the fused image. Segmentation methods such as adaptive threshold have been used to segment dark areas in SAR images for oil spill detection. In this method, a threshold is predefined on the image so that all pixels below the threshold value can be grouped into the fore, and above the threshold into the background in what is typically known in the computer vision society as background subtraction [152]. In [2], however, a global thresholding approach is considered. Firstly, a complimentary of the image is taken and the thresholding is done using the Otsu's method, which means rather than selecting a fixed value, all possible threshold values are evaluated and an optimal value is calculated in this process thereby minimizing the inter class variance of the thresholding resulting from the normalization of pixels in the ranges of [0 1]. The issue with this approach is that tiny spaces are created within the segmented image which can be remedied by image dilation procedure and smoothening. The segmentation in this framework uses the polynomial fitting curve and extends it with edge detectors to enable the detection of all edges around the dark areas in the fused image. This approach, allows the construction of an appropriate curve fitting so that the image is now represented as a smooth function [153].

Given the fused image $\mathbf{I}_F(x, y)$, the segmentation process starts by getting the histogram of the image which is a graphical representation of the pixel distribution and it represents the data to be segmented. The image histogram basically reflects the light value otherwise known as the tonal for each pixel in an image and is defined as

$$X(s_k) = n_k \, , \tag{4.32}$$

where s_k defines the kth intensity level in the pixel value range [0 1], n_k is the total number of pixels in $\mathbf{I}_F(x, y)$ with its intensity defined by s_k. Additionally, the histogram can normally be normalised by taking the average of all elements in $X(s_k)$ with the total number of pixels in $\mathbf{I}_F(x, y)$ here denoted by n.

(a) Haar Transform 4-Level Decomposition



(b) Corresponding Image Histogram

Figure 4.15: Fused Image and its Corresponding Histogram

$$p(s_k) = \frac{X(s_k)}{n} \tag{4.33}$$

$$= \frac{n_k}{n} \, . \tag{4.34}$$

The histogram of the fused Image at 4th level decomposition is shown in Fig 4.15 alongside the fused image. The light distribution in $\mathbf{I}_F(x, y)$ can simply be viewed from the image histogram where high peaks indicate areas in the image with higher tonal points and intensity values. The normalisation of the histogram can help in adjusting the image for further processing.

4.8.1 Polynomial Curve Fitting

Polynomial fitting curve can be described as a function that constructs a curve of best fit to a series of data points, and can be implemented at certain degrees or levels. A first degree polynomial can be defined as

$$y = ax + b \, . \tag{4.35}$$

The function, defines a line with its slope a and coordinates x.The first degree polynomial hence, defines the precise fit through any two points. A second order polynomial is given by

$$y = ax^2 + bx + c \tag{4.36}$$

The above function can exactly fit a curve through any three points denoted as a, b and c. If the order is increased, the third order degree polynomial is given by:

$$y = ax^3 + bx^2 + cx + d \ . \tag{4.37}$$

Prior to fitting the curve, firstly the histogram that defines the input data for the polynomial fitting curve is preprocessed which enables the object in the image to be detected according to the light intensity distribution as described earlier. The polynomial defined as $P(x)$ of a predefined degree n is used to fit the data $P(x_i)$ to the output y_i in a least square sense, here defined as

$$P(x) = P_1 x^n + P_2 x^{n-1} + \cdots + P_n x + P_{n+1} \ . \tag{4.38}$$

The coefficients $\widehat{x}$ of the polynomial $P(x)$ is defined as

$$\widehat{x} = \frac{x - \bar{x}}{\sigma_x} \ , \tag{4.39}$$

where $\bar{x}$ is the mean and σ_x defines variance of the coefficients. In Fig. 4.16 the polynomial fitting curve is shown on the image histogram, where the red line represent the fitting of best fit on the histogram data. To extract boundary information of the dark area (assumed oil spill), the segmentation step is further extended with edge detection using Canny edge detector firstly introduced in [154]. A brief description of the edge detection approach follows Edge detectors such as Sobel, Prewitt and Canny [155] play important roles in image segmentation by detecting intensity discontinuities that exist in image pixels by utilising the first and second order derivatives. A popular choice for first order derivate in image processing tasks is the gradient given as

$$\zeta \mathbf{G} = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{bmatrix} \ , \tag{4.40}$$

and the magnitude of $\zeta \mathbf{G}$ is defined as

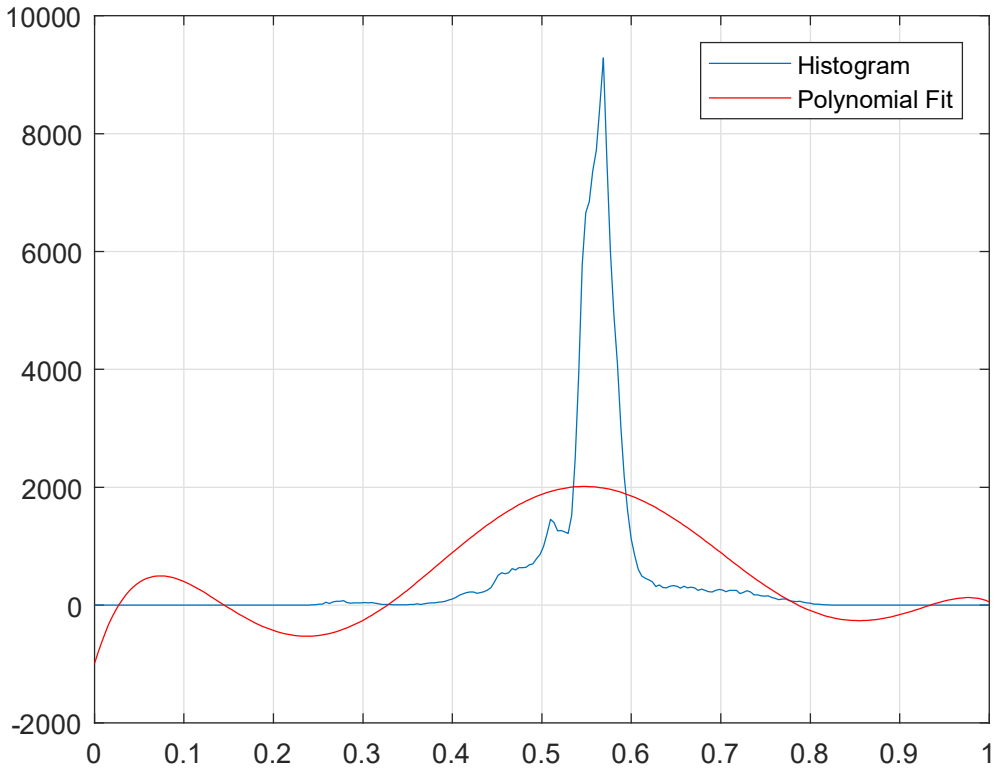


Figure 4.16: Polynomial curve fitting on the fused image histogram

$$\zeta G = mag(\zeta \mathbf{G}) = [g_x^2 + g_y^2]^{\frac{1}{2}} \tag{4.41}$$

$$= [(\delta f/\delta x) + (\delta f/\delta y)^2]^{\frac{1}{2}} \ . \tag{4.42}$$

The computation is improved by approximating the magnitude and eliminating the square root or by taking the absolute value of the quantities resulting in

$$\zeta G \approx g_x^2 + g_y^2 \tag{4.43}$$

or

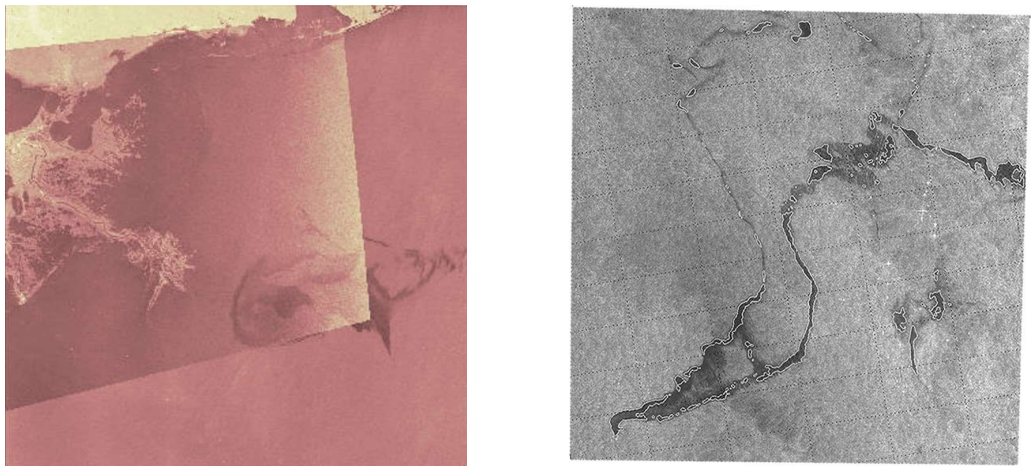
$$\zeta G \approx |g_x| + |g_y| \ . \tag{4.44}$$

The approximations still retain the derivate behaviour by being zero where the intensity

values are constant and variable in areas of the image where intensity values are changing. It is important to mention that the gradient vector is able to point towards the direction of maximum change in $\mathbf{I}_F$ (image to be segmented) in relation to pixel coordinates (x, y), and the angle of this change is expressed as

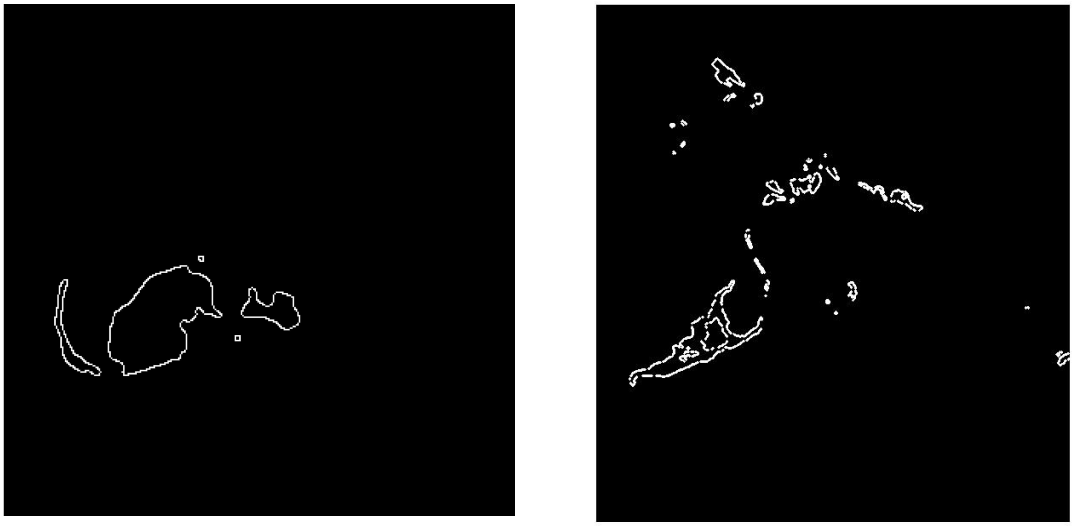
$$\theta(x, y) = tan^{-1} \begin{bmatrix} g_x \\ g_y \end{bmatrix} . \tag{4.45}$$

With the Canny edge detector, the image is filtered using (4.1) earlier introduced. The gradient, and angle of change within the pixels of the fused image are then computed over a 3×3 neighborhood of rows and columns to determine the edge points. The center pixel of this neighborhood is the average of the pixels within the window to provide smoothening [148] on all edge points, where an edge point defines a point in the direction of $\zeta\mathbf{G}$ whose strength is the local maximum. In determining the edge points, ridges are formed in the image and a further track of the top is done so that all pixels not on the top are set to zero, thus enabling a thin line to be formed in the segmented image in a procedure popularly referred to as non-maximal suppression. Next the suppressed pixels go through a hysteresis thresholding step that involves assigning two threshold values TH_1 and TH_2. Pixels with greater threshold values than the assigned value are labelled as strong candidate pixels while pixel values that fall between the two threshold values are labelled weak candidates. Lastly, strong and weak candidate pixels are incorporated by linking the edges using an 8-neighbor approach. The result of segmenting the images in Fig. 4.17 is shown in Fig. 4.18.



(a) Haar Transform 3-Level Decomposition (b) Haar Transform 4-Level Decomposition

Figure 4.17: Fused Image and a test non-Fused Image of C-Band SAR



(a) Segmentation with Polynomial Fitting and Canny Edge Detection on Fused Image

(b) Segmentation with Edges

Figure 4.18: Segmentation Results with Polynomial Fitting Curve and Edge Detection

4.9 Summary

In this chapter, an initial framework for oil spill segmentation with fused images is presented. The framework takes as input two SAR images acquired by different sensors and platforms with different spatial and spectral resolutions and fuse them using DWT to improve the quality of the individual images. The framework begins with pre-processing the individual SAR images by denoising image speckles inherent in these images using a Gaussian filter, and a further image enhancement to improve the overall image perception as seen in Fig. 4.4 and Fig. 4.5, respectively. The pre-processing step is followed by image registration step that aligns the images so that the similarities between them are mapped using SIFT which involves detecting invariant features that are immune to rotation, affine transformation and noise. The matching step is done with RANSAC which is robust and efficiently handles the mapping of features in the presence of outliers. The percentage of inlier matches which is greater than 65% from tentative ones is promising as shown in Fig 4.10. This can be attributed to the speckles reduction procedure during pre-processing. The decomposition of images using vertical and horizontal filters and creating sub-bands with DWT allows vital information for the fusion process to be captured. Results of different decomposition levels are evaluated using the performance metrics. It is noticed that, the higher the level of decomposition the better the fusion results with this approach. Again, it is noticed that image fusion with decomposition

creates the mosaic effects when the scale of decomposition is small as evidently shown with the performance measures in Table 4.2 and when the scale is large, the colour contents in the fused images could be lost hence making it incompatible for images with different spectral bands [61]. The fused image is then segmented using polynomial fitting curve and extended with edge detection to distinctively partition lines and shape of the dark patches(assumed oil spill locations) in the image. Polynomial Fitting curve has the advantage of smoothness, and reduces quantization errors caused by manual segmentation. Additionally, extending this polynomial fit step with canny edge detection ensures that nearly exact shape of the contour is retained which improves accuracy of the segmentation process as shown in Fig.4.19.

Chapter 5

A Gaussian Process Regression Approach for Fusion of Remote Sensing Images for Oil Spill Segmentation

5.1 Introduction

Oil spills are caused by accidental discharge or illegal dumping of oil ballasts by oil vessels and drilling platforms into the environment, causing enormous damage both socially and environmentally [10]. Marine oil spill for example, pollutes sea water, destroy wildlife, coastal beaches and affects the overall quality of life of marine inhabitants, raising concerns on oil transportation across the sea and a growing interest in developing efficient methods for oil spill detection [89].

The Macando blow-out that occurred on the 20th April 2010, is of national significance in the United States. It was an accidental oil spill caused by the explosion and sinking of the Deep Water Horizon offshore platform making the sea floor oil gusher to flow for 87 days and releasing more than 200 million gallons of oil into the Gulf of Mexico, resulting in loss of lives and damage to the marine ecosystem. The company responsible, British Petroleum (BP) was made to pay the largest environmental fine in history, a total of about 18.7 billion dollars [156]. Early detection of oil spill will help towards efficient disaster management. To detect and monitor oil spill, remote sensing systems with sensors on-board a satellite or aircraft are used

to acquire images of the earth from distance. Sensors in different bands of the electromagnetic spectrum have been applied, e.g. in hyperspectral and multi spectral bands, or Synthetic Aperture Radar (SAR) for oil spill detection in marine environment [1, 5, 8, 10, 157]. However, SAR images are the most widely used for this purpose as they are not affected by local weather conditions and cloudiness [13].

SAR is an active microwave sensor that acquires two dimensional (2D) images [36]. The performance of detecting oil spill in a SAR image, largely depends on sea conditions and the ability of oil films to decrease the backscattering of the sea surface, resulting in dark formations. A general assumption is that oil spill appears as dark areas on SAR images due to the dampen effects on capillary waves [4, 18]. However, not all dark formations are oil spills, necessitating the need for a robust detection technique and verification. Fig 5.1 illustrates an example of two challenging dark formations for detection. While Fig. 5.1(a) is a verified spill, Fig 5.1(b) is a verified look-alike showing the difficulty in visually ascertain the dark formations formed by the backscattering effect of SAR.

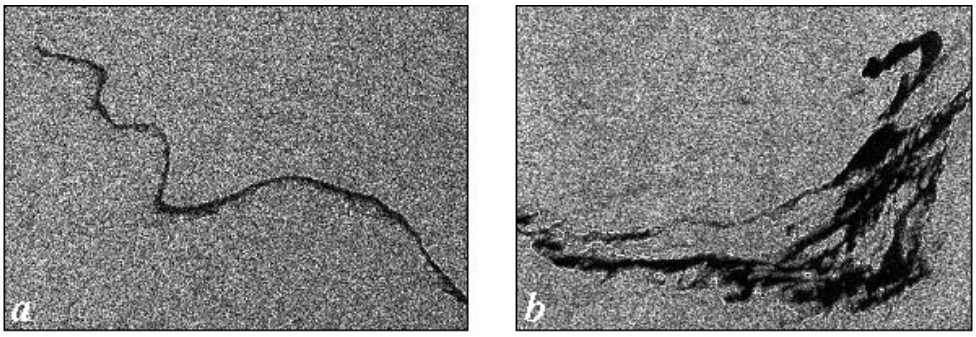


Figure 5.1: Example of challenging dark formations: (a) verified oil spill (b) Verified look-alike. [158]

Recent developments have enabled more spaceborne SAR systems (e.g. the Sentinels) to be designed and launched, providing the scientific community with wide range of data in multi-modal configuration, including multi-frequency (C, L, X etc.), multi-angle ($10^\circ - 70^\circ$), multi-polarisation (dual, quad, etc) and multi-resolution features. With the availability of multi-modal SAR data, new methods to detect oil spill can be explored by combining data originating from different sources, with the aim of obtaining information of greater quality than individual sensor data used in previous studies for oil spill detection [5, 13, 18, 35, 36, 127–129] .

Fusion of SAR images, however, imposes several challenges due to multi-modalities, differences in sensor characteristics and image acquisition modes. The individual images need

to be registered in space and time. A suitable fusion algorithm needs to be chosen that will efficiently bring together the complimentary information from the individual modalities [8]. Additionally, it is important to define the level of fusion since data fusion can be performed at different information levels including pixel, feature and decision level [130].

Previous studies have reported significant improvement in oil spill classification, segmentation and discrimination with fused SAR images compared to using the individual images alone. In [131], fusion of SAR and hyperspectral images (HSI) is performed at pixel level, although the fusion method used is not described in the paper, the approach focused on fusing images from different sensors (SAR and HSI) and not on multi-frequency SAR data. The works of [8, 126] explored fusion of multi-frequency (S & X-bands) and multi-resolution and multi-modal (C-band) SAR images by adopting the wavelet transform approach.

Wavelet transform as shown in the previous chapter of this thesis improves the spatial resolution of the fused image while preserving the colour appearance for interpretation [54], this is important since oil spill appears as dark formation on SAR images. With wavelets, images are converted from the spatial domain to the frequency domain and then decomposed into approximation and detail coefficients while preserving information, allowing image properties to be transferred using a fusion rule. However, if the decomposition scale is small, mosaic effects occurs on the fused image. On the other hand, if the scale is large, the colour contents of the fused images are lost, making it unsuitable for images with different spectral channels [61].

This chapter proposes to solve oil spill segmentation problems by fusing multi-resolution SAR images using a Gaussian process regression approach. The approach is based on the design of a non-stationary covariance kernel to handle the change of support problem that exists in multi resolution images, The approach extends the work from [64] over different image modalities. A prior covariance function, the product of an intrinsically sparse covariance kernel and a rational quadratic kernel is utilised to model the high resolution pixel coordinates and their intensity values, forming a base covariance from which the new modality image is constructed. It considers that Gaussian process models have been used in object recognition in situations where the images are in different resolutions with the training data [159]. Additionally, GP priors are adaptable for inter-modality data encoding with multiple output behaviour [64]. The aim is to construct an image with high spatial and high spectral resolution.

The rest of the chapter is organised as follows. In the following up sections, the proposed

framework and a brief description of the registration is provided, with more detailed explanation of the process already presented in the previous chapter of this thesis. Subsequently, GP is introduced and the proposed kernel design and fusion process are described. In addition to this, the performance validation measures of the approach are evaluated and finally K-means segmentation is utilised to distinctively seperate the dark regions (assumed oil spill) and other coastal areas of the scene of interest as captured in the estimated (fused) image.

5.2 Proposed Framework

The proposed framework is shown in Fig 5.2. It comprises a pre-processing step that performs image filtering to reduce speckles inherent in SAR images, and image enhancement, to improve image visualisation and to obtain the best possible image perception, respectively [8]. The next process is an automatic image registration that aligns the images so that a common spatial frame is realised. Lastly, the fusion and performance evaluation stages complete the system framework.

5.2.1 Gaussian Processes

Recall from chapter 2 of this thesis that a Gaussian Processes (GP), is determined by a mean function and a covariance function also known as the covariance kernel. The mean $m(\mathbf{x})$ and the covariance $k(\mathbf{x}, \mathbf{x}')$ of a space function $f(\mathbf{x})$ are given as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{5.1}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{5.2}$$

and the GP can be described as:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{5.3}$$

GP is a stochastic process, defined as a collection of random variables [69]. For convenience, the mean function is often assumed a zero value since GP can be adjusted to model the mean swiftly [72], while the covariance kernel is determined by some hyperparameters. A detailed explanation on kernels and hyperparameter adaptation is discussed in [69]. To achieve the mapping of inputs to an output space, GP imposes a Gaussian prior distribution over the

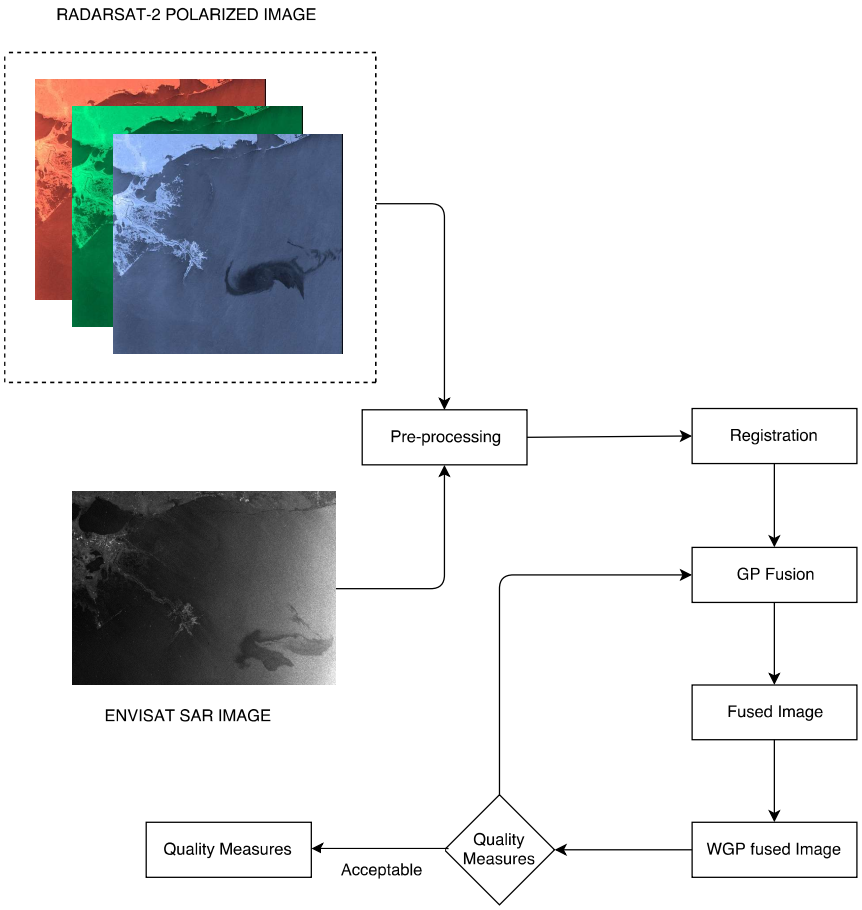


Figure 5.2: The Proposed GP fusion framework

space functions $f(\mathbf{x})$, to map inputs $\mathbf{x}_i \in \mathbb{R}^D$ to the output space $\mathbf{y}_i \in \mathbb{R}$, where the output $\mathbf{y}$ is a noisy observation represented as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\epsilon}, \tag{5.4}$$

where

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \tag{5.5}$$

is a Gaussian distribution with a zero mean and standard deviation σ.

To make predictions, GP learns the hyperparameters from the given training dataset here represented as Q. The training is done on N input-output pairs such that: $Q = \{(y_1, x_1), \ldots (y_N, x_N)\}$, where the function values are normally distributed with the modelled mean and covariance defined as

$$[f(x_1)^T f(x_2)^T \ldots f(x_N)^T]^T = \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}')) \ , \tag{5.6}$$

here, $m(\mathbf{x})$ and $K(\mathbf{x}, \mathbf{x}')$ defines the GP mean and covariance, respectively.

Representing the number of observations N as $\{\mathbf{X}, \mathbf{y}\}$, where $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^{\mathbb{D}}\}_{i=1}^{N}$ and $\mathbf{y} = \{y_i \in \mathbb{R}\}_{i=1}^{N}$ and test points M given as $\{\mathbf{X}^*\} = \{\mathbf{x}_i^* \in \mathbb{R}^{\mathbb{D}}\}_{i=1}^{M}$, the joint density of the observations N and the test points M is given as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \mathcal{N}\left(\boldsymbol{\mu} \begin{pmatrix} \mathbf{X} \\ \mathbf{X}^* \end{pmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{X}^*) \\ K(\mathbf{X}^*, \mathbf{X}) & K(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix} \right), \tag{5.7}$$

Here, $\mu(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is a positive semi-definite covariance function. From (5.7) above, the predictive distribution of the mean and covariance functions can be defined as

$$p(\mathbf{y}^* | \mathbf{x}, \mathbf{y}, \mathbf{x}^*) \sim \mathcal{N}(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \, , \tag{5.8}$$

where

$$\boldsymbol{\mu}_\star = k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} \mathbf{y} \, , \tag{5.9}$$

$$\boldsymbol{\Sigma}_\star = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} k(\mathbf{X}, \mathbf{X}^*) \, . \tag{5.10}$$

Here, $\boldsymbol{K}_X$ is defined as:

$$\boldsymbol{K}_X = k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 \mathbf{I} \, , \tag{5.11}$$

and $\sigma_n^2 \mathbf{I}$ is the sensor measurement noise, $\mathbf{I}$ is an N-dimensional identity matrix. Subsequently, we will revert to our earlier notations of the mean $(\boldsymbol{\mu}_\star)$ as $m(\mathbf{x})$ and the covariance $(\boldsymbol{\Sigma}_\star)$ as $k(\mathbf{x}, \mathbf{x}')$. Considering that the $m(\mathbf{x})$ of the process is not always confined to a zero value, the mean function can be modelled to conveniently express the prior information allowing the predictive mean to be written as

$$\mathbf{y}^* = \mathbf{m}(\mathbf{X}_*) + K(\mathbf{X}_*, \mathbf{X}) {K_X}^{-1} (\mathbf{y} - \mathbf{m}(\mathbf{X})), \tag{5.12}$$

where $\boldsymbol{K}_X = K + \sigma^2 \mathbf{I}$ and $\boldsymbol{K}$ = $k(\mathbf{X}, \mathbf{X})$.

5.2.2 Covariance Function

The covariance function or kernel plays the central role in a GP. It encodes the inference of the underlying process by defining the correlation between function values [74]. In this chapter,

we aim to construct an image with a new modality, by combining SAR images with different image resolutions. To achieve this, firstly we require the covariance function to handle the change of support problem that exists in multi-modal images by conducting inference over image pixels with different resolutions. This is achieved by extending the observation point kernel to adapt to a multiple task kernel function over spaces, and utilising an integral kernel derivation described in [160]. The assumption here is there exist a 2D GP function $\mathbf{f}$ within the spectral channels of the low resolution image, such that the designed model of a pixel is the result of observing the output function $\mathbf{f}_*$ over areas of the high resolution image $k(\mathbf{H_A}, \mathbf{H'_A})$ rather than points $k(\mathbf{x}, \mathbf{x}')$ which is the norm in a standard GP. A simple average relationship is then established between observed pixels and $\mathbf{f}_*$. A detailed derivation of defining covariance over areas is described in [64].

Secondly, a prior of the new modality image structure is defined. The covariance function prior design is based on the spatial characteristic of the high resolution image, forming the base for the new modality image construction. For the fusion problem, Let $\mathbf{H_A}$ denote the locations of the high-resolution image pixels and $\mathbf{L_A}$ the locations of the low-resolution image pixels, the covariance between two high-resolutions image pixels is defined as

$$k(\mathbf{H_A}, \mathbf{H'_A}) = \frac{1}{|\mathbf{H_A}||\mathbf{H'_A}|} \iint_{\mathbf{x}\in\mathbf{H_A}} \iint_{\mathbf{x}'\in\mathbf{H'_A}} k(\mathbf{x}, \mathbf{x}')dxdx' \,, \tag{5.13}$$

where $k(\mathbf{H_A}, \mathbf{H'_A})$ defines the covariance between two high resolution image areas and $|\mathbf{H_A}|$ is the surface area of $\mathbf{H_A}$

To design the prior, we consider that image data are normally non-smooth, and exhibit discontinuities with spatial non-stationarity. For this reason, the spatial information of the high-resolution image is used as the input space of the covariance function, added with the observed pixels of the high-resolution image to achieve contextual non-stationarity and to address image discontinuity problem. In the first step of the prior design, we exploit an intrinsically sparse covariance function proposed by [161] to obtain a sparse covariance. The sparse kernel is smooth but not infinitely differentiable making it suitable for application exhibiting discontinuities. A detailed description and derivation of this kernel is given in [161]. Let V_S be an intrinsically sparse kernel defined as

$$V_S(\mathbf{H_A}, \mathbf{H'_A}; \sigma_0, l) =$$

$$\begin{cases} \sigma_0 \left[\frac{2+\cos(2\pi \frac{d}{l})}{3} (1 - \frac{d}{l}) + \frac{1}{2\pi} \sin(2\pi \frac{d}{l}) \right] & \text{if } d < l \\ 0 & \text{if } d \geq l \ , \end{cases} \tag{5.14}$$

where d is the distance between the midpoint coordinates of the high-resolution pixel areas here defined as

$$d = |mid(\mathbf{H_A}) - mid(\mathbf{H'_A})| \ . \tag{5.15}$$

The variables σ_0 and l are the parameters of the kernel. where σ_0 determines the average distance of the function from the mean, l is the characteristic length scale that determines the length of change of the function.

In the second step, we link the observed high-resolution image pixels using a Rational Quadratic Kernel (RQK). The RQK is equivalent to adding together several Squared Exponential (SE) with different length-scales, enabling smoothness transfer and efficiency in handling the change of support problem. Let V_P be the RQK defined as

$$\begin{aligned} &V_P(\mathbf{I_H}(\mathbf{H_A}), \mathbf{I_H}(\mathbf{H'_A}); \alpha, \theta_p) \\ &= \sigma^2 \left(1 + \frac{(\mathbf{I_H}(\mathbf{H_A}) - \mathbf{I_H}(\mathbf{H'_A}))}{2\alpha\theta_p^2} \right)^{-\alpha} , \end{aligned} \tag{5.16}$$

Here, $\mathbf{I_H}(\mathbf{H_A})$ and $\mathbf{I_H}(\mathbf{H'_A})$ represent the pixel intensity values in $\mathbf{H_A}$ and $\mathbf{H'_A}$, respectively. The variables α and θ_p are the parameters of the function. The function V_S provides a smooth, sparse and neighbouring covariance kernel, while V_P link image pixels within the covariance based on related information contained in the high-resolution image. Additionally, the RQK function V_P enables image smoothness transfer and handles the change of support problem. Hence, the image prior covariance function is then defined as a product of two independent processes (a) A positive semi-definite spatial covariance V_S and (b) A positive semi-definite pixel intensity covariance function V_P

$$k(\mathbf{H_A}, \mathbf{H'_A}) = \sigma^2 V_S((\mathbf{H_A}, \mathbf{H'_A}; \boldsymbol{\theta}) V_P(\mathbf{I_H}(\mathbf{H_A}), \mathbf{I_H}(\mathbf{H'_A}); \boldsymbol{\theta}). \tag{5.17}$$

Using the high-resolution areas $(\mathbf{H_A}, \mathbf{H'_A})$ and the augmented intensity values $(\mathbf{I_H}(\mathbf{H_A}), \mathbf{I_H}(\mathbf{H'_A})))$ of the high-resolution image pixels, the GP model is then trained to learn the hyperparameters represented as $\boldsymbol{\theta}$ of the image prior $k(\mathbf{H_A}, \mathbf{H'_A})$ as described in the next section.

5.2.2.1 Hyperparameter Adaptation

Hyperparameters refer to the parameters of the prior $k(\mathbf{H_A}, \mathbf{H'_A})$ that includes parameters of the mean, covariance and noise term $\sigma_n^2\mathbf{I}$. Here the parameters of the model are defined as $\theta = \{\theta_s, \theta_p, \sigma_f, \alpha, l\}$, where σ_f is an amplitude hyperparameter while θ_p controls sensitivity, α determines the relative weighing for large-scale and smaller-scale variations. When $\alpha \rightarrow \infty$ the behaviour of the kernel is identical to the SE kernel. The parameters of the prior need to be selected appropriately as they determine the quality of the output image. To optimize the hyperparameters, the Bayes approach is considered because it allows the use of continuous optimization methods enabling faster computation [72]. To achieve this, the marginal likelihood is maximised such that

$$P(\mathbf{y}|X) = \int P(\mathbf{y}|\mathbf{f}, X)P(\mathbf{f}|X)d\mathbf{f}\,. \tag{5.18}$$

From (5.12), the likelihood $\mathbf{y}|\mathbf{f} \sim \mathcal{N}(\boldsymbol{f}, \sigma_n^2\mathbf{I})$, with the GP prior over the latent function f from (11), gives the log of the marginal likelihood

$$log(P(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})) = -\frac{1}{2}\mathbf{y}^T K_X^{-}1\mathbf{y} - \frac{1}{2}log|K_X| - \frac{n}{2}log2\pi. \tag{5.19}$$

In (5.23), the first term of the log of marginal likelihood finds data fit, the second is a model complexity term while the third is a constant, making it robust to over-fitting.

The optimized parameters of the covariance function from (5.23) are used to calculate $k(\mathbf{H_A}, \mathbf{H'_A})$ which forms the base of the new modality image we aim to construct. Using the prior covariance, we find the covariance $k(\mathbf{H_A}, \mathbf{L'_A})$ that couple the high-resolution pixels with the observed low-resolution pixels. First, the observed $\mathbf{L_A}$ pixels are approximated by the $\mathbf{H_A}$ pixels, an integration over $\mathbf{L_A}$ with respect to point $\mathbf{x}$, the sum of the piecewise integration over the pixels of $\mathbf{H_A}$ [64], defined as

$$k(\mathbf{H_A}, \mathbf{L'_A}) = \frac{1}{T_H}\sum_{\mathbf{H'_A} \in \mathbf{L'_A}} k(\mathbf{H_A}, \mathbf{H'_A}), \tag{5.20}$$

where T_H is the number of high-resolution $\mathbf{H_A}$ areas that are contained in $\mathbf{L_A}$ area. Accordingly, the corresponding covariance between the low-resolution pixels is defined as

$$k(\mathbf{L_A}, \mathbf{L'_A}) = \frac{1}{T_H T'_H}\sum_{\mathbf{H_A} \in \mathbf{L_A}}\sum_{\mathbf{H'_A} \in \mathbf{L'_A}} k(\mathbf{H_A}, \mathbf{H'_A}). \tag{5.21}$$

5.2.3 Fused Image Reconstruction

To fuse the two image modalities, the training data of the model comes from the low-resolution image comprising the $\mathbf{L_A}$ spatial areas and the pixel intensity values $\mathbf{I_L}(\mathbf{L_A})$ that were observed. The GP model is then queried over the high resolution $\mathbf{H_A}$ spatial areas where the $\mathbf{I_H}(\mathbf{H_A})$ intensity values have earlier been defined in (5.20). The fused image is constructed by querying the predictive mean of the GP model whilst performing a normalisation as described below. Additionally, a constant mean value $\mu = 0.5$, is assumed over the image, this is justified because image pixels are observed to be continuous within the range of 0-1. Hence, the predictive mean in (5.16) becomes

$$\widehat{\mathbf{F}}_i = \mu + k(\mathbf{H_A}, \mathbf{L_A}) \left[k(\mathbf{L_A}, \mathbf{L'_A}) + \sigma_n^2 \boldsymbol{I} \right]^{-1} (\mathbf{I_L}(\mathbf{L_i}) - \mu), \tag{5.22}$$

where i represent the ith spectral band of the low-resolution image, and $\boldsymbol{I}$ is an identity matrix equivalent to the number of pixels in the low-resolution image. This implies that we query the GP model by the number of spectral channels present in the low-resolution image. Finally, the new modality image $\widehat{\mathbf{F}}$ is the concatenated sum of all $\widehat{\mathbf{F}}_i$'s.

5.2.4 Fusion Quality Metrics

It is important to evaluate the quality of the output fused image from the model using well-established image quality assessment measures. Methods such as Image Correlation Coefficient (CC) and Mean Squared Error (MSE) are widely used to determine the quality of images and measure the similarity between fused image and a reference image. However, image quality measures can be classified into three categories depending on the aim of the fusion. This can be to measure spatial, spectral or global quality of the image. In [162] for example, a Wald's Protocol is used to test the quality of pan sharpened images focusing on consistency and synthesis. In this chapter, the following global quality performance measures are used to test the quality of the fused image

1. Root Mean Squared Error (RMSE): RMSE evaluates the difference between the fusion model output and the reference image [50], providing a complete image quality pointer with results closer to zero indicating a high performance of the algorithm. RMSE is given as

$$RMSE(\tilde{\mathbf{R}}_*, \widehat{\mathbf{F}}_*) = \left(\frac{\sum_{i=1}^{M} \sum_{j=1}^{N} [\tilde{\mathbf{R}}_*(i,j) - \widehat{\mathbf{F}}_*(i,j)]^2)}{m \times n} \right), \tag{5.23}$$

where $\mathbf{R}_*(i,j)$ and $\widehat{\mathbf{F}}_*(i,j)$ are the pixel values of the fused and reference image, respectively and $m \times n$ are the rows and columns of the image that define the image size.

2. Image Correlation Coefficient (CC): CC is a fusion quality measure that characterises the geometric distortion between the reference image and the estimated (fused) image. The higher the correlation between the images the better the estimation of the spectral values. CC is defined as

$$CC(\mathbf{R}_*, \widehat{\mathbf{F}}_*) = \frac{\sum_{mn}(\mathbf{R}_* mn - \bar{\mathbf{R}}_*)(\widehat{\mathbf{F}}_* mn - \bar{\mathbf{A}}_*)}{(\sum_{mn}(\mathbf{R}_* - \bar{\mathbf{R}}_*))^2(\sum_{mn}(\mathbf{R}_* - \bar{\mathbf{R}}_*))^2}\ , \tag{5.24}$$

where $\mathbf{R}_*$ is a pixel of the reference image with size $(m \times n)$, $\bar{\mathbf{R}}_*$ is the mean of the reference image. Similarly, $\widehat{\mathbf{F}}_*$ with size $(m \times n)$ represent a pixel of the fused image and $\bar{\widehat{\mathbf{F}}}_*$ is its mean.

3. Erreur Relative Globale Adimensionnelle de Synthese (ERGAS): This measure offers a global indication of quality of the estimated image [89], based on normalised average error of each band of the image [112]. The ideal value of ERGAS is 0 [130]. Increase in the value of ERGAS could mean a distortion in the estimated image, on the hand increase in the value of ERGAS indicates that the estimated image $\widehat{\mathbf{F}}_*(x,y)$ is similar to the reference image $\tilde{\mathbf{R}}(x,y)$. ERGAS is defined as

$$ERGAS(\tilde{\mathbf{R}}, \widehat{\mathbf{F}}_*) = 100r\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{RMSE_j}{\mu_j}\right)^2} \tag{5.25}$$

where r is the ratio between the resolution of the images fused, μ_j is the sample mean of the ith band of $\mathbf{F}_*$.

5.3 Segmentation

Segmentation is the subdivision of the image into separated regions [35], grouping similar pixels to homogeneous image segments so that increase in heterogeneity over the image is very much reduced allowing image pixels to be classified correctly in a decision oriented application. In oil spill detection, segmentation is a pre-requisite for classifying oil spill and look-alike. In this stage, dark areas that are assumed oil spill based on appearance are segmented out from the image and features are extracted that form the base for classification. In this chapter, the segmentation step is done using the K-means clustering algorithm.

5.3.1 K-means Segmentation

K-means clustering allows partitioning of data into a k number group of the data [163], classifying the given data (image) into k number of disjoint clusters. To achieve this, the algorithm is divided into two steps. In the first step, it calculates the k centroid in the image using Euclidean distance, and in the second step it groups each image pixel to a cluster nearest to a k_i centroid from the respective pixel. Summarily, K-means is an iterative method that minimizes all distances from each pixel to its cluster k_i centroid over all clusters k. Using the output image ($\widehat{\mathbf{F}}_*$) from (5.26), with size $m \times n \times \lambda$, where λ is the number of bands in $\widehat{\mathbf{F}}_*$. The aim is to segment the image into k number of clusters, let $a(m_i, n_j)$ be an input pixel of $\widehat{\mathbf{F}}_*$ to be assigned to a cluster, and c_k be the centroid of the clusters; first the number of clusters k is initiated, secondly, for each pixel of the image, the Euclidean distance d is calculated, between the centre c_k of the centroid and the pixel using

$$d = \| \, a_*(m_i, n_j) - c_k \, \| \quad . \tag{5.26}$$

Next, all pixels are assigned to the nearest k_i using d. A new position of k_i is recalculated using

$$c_k = \frac{1}{k} \sum_{n \in c_k} \sum_{m \in c_k} a_*(m_i, n_j) \; . \tag{5.27}$$

Finally, the cluster of pixels are reshaped to form the segmented image.

5.3.2 Dataset

The dataset used in this chapter are as presented in Table 1. Firstly, multi-modal and multi temporal SAR images of the Gulf of Mexico oil spill as acquired by the Canadian RADARSAT-2 ScanSAR instrument are utilised. This instrument is fully polarimetric (HV,VV,VH) in wide beam mode with a nominal swath of 500km. Secondly, the European Space Agency (ESA)'s Envisat system with a single band (VV) polarisation also in ScanSAR and wide swath mode is also utilised. A summary of the data description is given in Table 5.1.

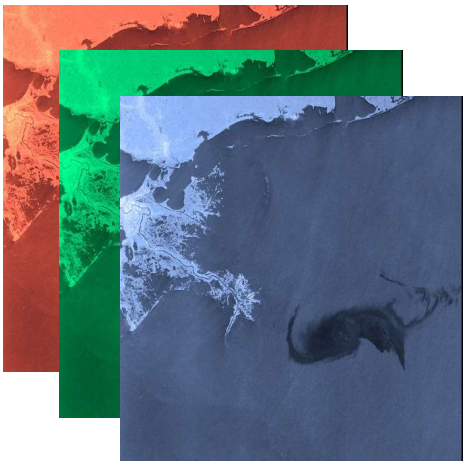
Table 5.1: Characteristics of the Dataset:

Satellite	Instrument	Resolution	Band	Dimension	Date Acquired
Radarsat-2	SAR	$100m$	C	865×905	$29/04/10$
Envisat	ASAR	$150m$	C	930×1271	$26/04/10$

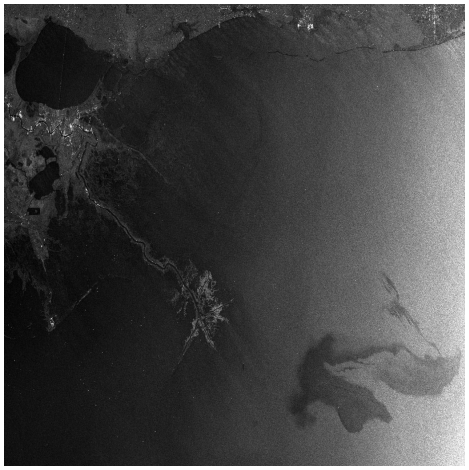
5.3.3 Discussion

The first task is to pre-process the SAR images to be fused, to reduce speckle noise and enhance the image using a Gaussian filter as discussed in Section 5.2. The registration stage is next, following the steps described in Subsection 5.2.1 to align the images and to find correspondence between them. A mosaic of the two images is presented in [8] to show the progression of dark area (supposed spill location) over the multi-temporal period. An average fusion result is shown in Fig 5.4(a); a product of adding the two images together and taking the average. The GP fusion algorithm described in Subsection 5.2.2. is applied to fuse the multi resolution images. Subsequently, we compare the results of the proposed algorithm with [64] using global image quality measures described in Subsection 5.2.5 to test the quality of the output image and the performance of the proposed method presented in Table 5.2. In Figs. 5.4b and 5.5b we present the results of the fusion process of the proposed method and the method of [64]. It is noticed that the proposed approach achieves compelling enhancement visually; this is attributed to the intrinsically sparse covariance function that provides a much smoother prediction of the function. In Fig. 5.7, the RMSE per image pixel of the output image $\widehat{\mathbf{F}}_*$ and the output image from the method of [64] is also compared. Again, the proposed method achieves a better performance in this measure. The evaluation time to ouput the fused image from evaluating the covariance kernel function with optimised hyperparameter values is also compared. The proposed method achieved this in 0.2sec compared to [64] which took 0.63sec. Lastly, K-means technique described in Section III is applied to segment the dark formations (assumed oil spill) in $\widehat{\mathbf{F}}_*$ by converting the image into CIELAB color space also known as *lab* color space which gives the initial value of K, as described in (5.30) and (5.31), respectively. The segmentation result is shown in Figs. 5.6(a) and 5.6(b), objects in the image are clustered to different regions in Fig. 5.6c.

Image segmentation results presented in Fig 5.6 shows that the fusion algorithm proposed can improve the detection of objects within the scene of interest and objects can be segmented more accurately since the resolutions both spatial and spectral are more enhanced by injection of spatial details. In Fig 5.6 (a) The segmentation results shows the island and shorelines of the Gulf, while 5.6(b) segments the dark areas (assumed oil spill locations). In Fig 5.6 (c) the image is completely segmented to show the different objects recognised by the clustering algorithm. The approach presented differs in the sense that it can uniquely cluster similar objects accurately. This is important in segregating the actual locations and geometry of



(a) Polarimetric SAR Image



(b) Grayscale SAR Image

Figure 5.3: RADARSAT-2 ScanSAR and Envisat ASAR images of Gulf of Mexico Oil Spill acquired 29/04/2010 and 26/04/2010, respectively.

(a) Simple Average Result

(b) Reid et al

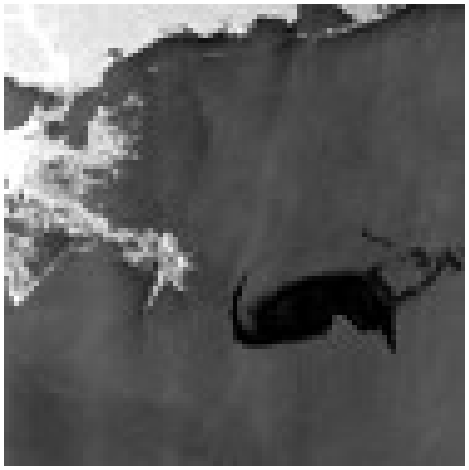
Figure 5.4: Fusion Results (a) Simple Averaging Method (b) Reid et al Method

affected areas.

Table 5.2: Quality Measures of the Fusion Result

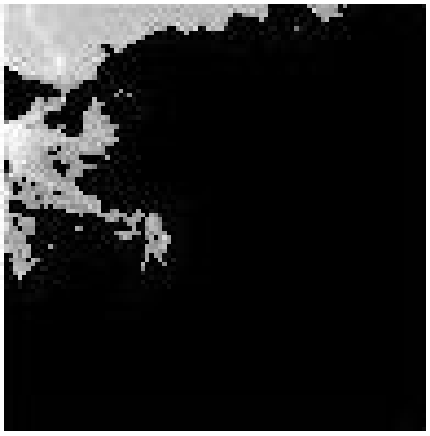
Fusion Method	CC	RMSE	ERGAS
DWT Fusion	0.9997	0.4937	23.3466
Reid et al	0.9997	0.7571	5.5657
Proposed	0.6414	0.4059	5.8099

(a) DWT Fusion at level 4 Decomposition



(b) Proposed GP Approach

Figure 5.5: Fusion Results for (a) DWT and (b) Proposed GP Approach



(a) Cluster1

(b) Cluster2

(c) Cluster3

(d) Segmentation results in different clusters of the fused Image

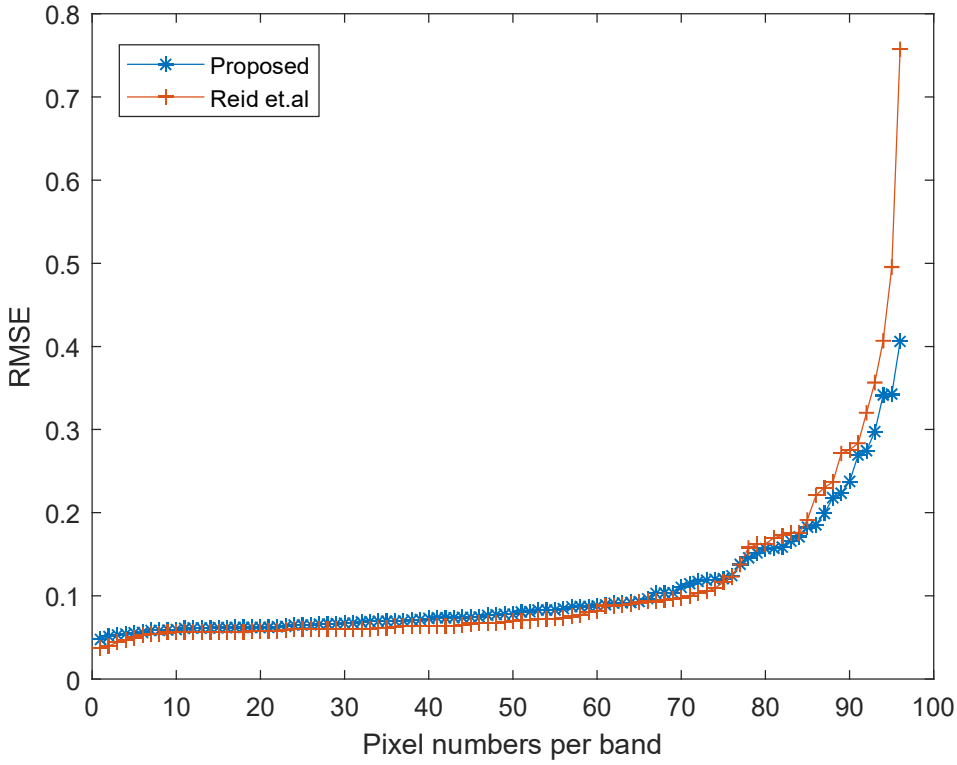


Figure 5.7: Comparison of RMSE per pixel of Methods

5.4 Multi Band Image Fusion

The approach for fusion presented in this chapter is extended for multi-band datasets. In this case the images to be fused have more than one spectral and spatial bands, respectively. Previously, in the preceeding chapters, the methods have focused on fusing high spatial resolution and high spectral resolution images. In which case, the high spatial resolution images have single band and the high spectral resolution images having more than one band. The flexibility of the GP is utilised for this approach. Recall that GP being a powerful data driven approach has led to efficient solutions of image processing problems including for modelling of low level image features, image de-noising and exploring the structural redundancy for super-resolution image reconstruction. GPs have also been applied to high resolution object reconstruction for lower resolution images by considering local structures in natural images defined by their pixel neighbourhood [74]. Additionally, GP regression provides a flexible framework for fusing multiple datasets from heterogeneous sensors [164]. In [64], a multi-task GP approach is developed for multi-modal image fusion. Other image fusion approaches

such as Wavelet Transform have been used in [126], [8], [54] to improve spatial resolution and maintain the spectral property of the images. This is achieved by converting the image from its spatial domain to the frequency domain, and decomposing the image into approximated and fine details without information loss as described in chapter 4 of this thesis. The fusion approach allows image properties to be inferred, thanks to the fusion rule including Haar, Debuchies and other wavelets [19].

Here, the approach combines high resolution image data acquired by a UAV exploring marine environments with multispectral data acquired by the Sentinel-2 satellite for the same scene. This work extends the approach proposed in [19] for fusing images with multiple bands. Firstly, a multi-output regression framework is introduced by applying an independent GP [165] to each output band in the high resolution image. Secondly, common spectral channels in the visible range are selected from the multi-modal images (3-bands in each dataset). The covariance function of the model is the product of an intrinsically sparse kernel and a Rational Quadratic Kernel (RQK) utilised to model the coordinates and intensities of pixels in the high resolution image which forms the spatial base of the estimated image to be recovered. This work is motivated by the ability of GP's to sufficiently model the relationship between different modalities with varied outputs, with an aim at extending its inherent point data operation over areas [64]. The contribution of this work is the adaptation of a multi-output variable GP to model the spatial information from selected bands of the high spatial resolution image and to use a mapping cross covariance to combine spectral information. The proposed kernel serve as a base stochastic prior for each band of the estimated image.

5.4.1 The Proposed Gaussian Processes Fusion Approach

Recall: The GP can be described by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ also known as the kernel. It is a stochastic process that defines a collection of random variables [70] . The covariance function is defined by some hyperparameters that characterise its behaviour while the mean function may conveniently assume a zero value since GP can be modified to model a non-zero mean [72]. More insight into different kernels and modelling the mean function can be found in [69]. Consider a function $f(\mathbf{x})$, whose mean and kernel are defined as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{5.28}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))], \tag{5.29}$$

where $\mathbb{E}$ is the expectation operator and the GP is then described as a non-linear function expressed as:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{5.30}$$

A GP model can achieve the mapping of inputs $\mathbf{x}_i \in \mathbb{R}^{\mathcal{D}}$ to an output space $\mathbf{y}_i \in \mathbb{R}$ by imposing a Gaussian prior over the latent function where the output vector $\mathbf{y}$ is a noisy observation represented as

$$\mathbf{y} = f(\mathbf{x}) + \epsilon \tag{5.31}$$

and

$$\epsilon \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}) \tag{5.32}$$

is a Gaussian distribution with a zero mean and standard deviation σ.

The parameters of its kernel are then learnt using N input-output pairs from a given training dataset $\mathcal{D}$, such that $\mathcal{D} = \{(\mathbf{x_1}, \mathbf{y_1}), \ldots (\mathbf{x_N}, \mathbf{y_N})\}$. The function values are normally distributed with the mean $m(\mathbf{x})$ and the covariance $k(\mathbf{x}, \mathbf{x}')$ given by:

$$[f(\mathbf{x_1})^T f(\mathbf{x_2})^T \ldots f(\mathbf{x_N})^T]^T = \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))\ , \tag{5.33}$$

here, $m(\mathbf{x})$ and $k((\mathbf{x}, \mathbf{x}')$ are as defined in (1) and (2), respectively.
The entries of the covariance matrix $\mathbf{K}_{ij}$ are calculated by evaluating each element of $k(\mathbf{x}, \mathbf{x}')$ using the user defined kernel e.g the Squared Exponential (SE) with its hyperparameters [69]. This can be expressed as

$$k(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} k(\mathbf{x_1}, \mathbf{x_2}) & k(\mathbf{x_1}, \mathbf{x_2}) & \ldots & k(\mathbf{x_1}, \mathbf{x_N}) \\ k(\mathbf{x_2}, \mathbf{x_1}) & k(\mathbf{x_2}, \mathbf{x_2}) & \ldots & k(\mathbf{x_2}, \mathbf{x_N}) \\ \vdots & \vdots & \vdots & \vdots \\ k(\mathbf{x_N}, \mathbf{x_1}) & k(\mathbf{x_N}, \mathbf{x_2}) & \ldots & k(\mathbf{x_N}, \mathbf{x_N}) \end{bmatrix} \tag{5.34}$$

As part of the GP training, the hyperparameters of the kernel are optimized. This is achieved by maximising over the hyperparameters and the marginal likelihood. This is illustrated in

the next section. Consequently, the joint distribution of the training outputs $\mathbf{y}$ and the test outputs $\mathbf{y}^*$ with the mean function is defined by

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu} \begin{bmatrix} \mathbf{X} \\ \mathbf{X}^* \end{bmatrix}, \begin{bmatrix} K_{\mathbf{X}\mathbf{X}} + \sigma_n^2 \mathbf{I} & K_{\mathbf{X}\mathbf{X}^*} \\ K_{\mathbf{X}^*\mathbf{X}} & K_{\mathbf{X}^*\mathbf{X}^*} \end{bmatrix} \right) \tag{5.35}$$

In (8) above, $\mathbf{X}$ and $\mathbf{X}^*$ define the design matrices for the training and test data, respectively. When $\mathbf{y}^*$ is conditioned on the observations $\mathbf{y}$ the predictive distribution becomes

$$p(\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*) \sim \mathcal{N}(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \tag{5.36}$$

where

$$\boldsymbol{\mu}_\star = k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} \mathbf{y} \tag{5.37}$$

$$\boldsymbol{\Sigma}_\star = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} k(\mathbf{X}, \mathbf{X}^*) \tag{5.38}$$

Here, $\boldsymbol{K}_X$ is defined as

$$\boldsymbol{K}_X = k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 \mathbf{I} \tag{5.39}$$

and $\sigma_n^2 \mathbf{I}$ is the standard deviation of the measurement noise, $\mathbf{I}$ is an N-dimensional identity matrix. With the learnt hyperparameter values of the kernel, the GP can then predict the output $\mathbf{y}^*$ using the predictive distribution of input and outputs points.

5.4.2 Multi-Output Variable Gaussian Processes

Normally, a GP model assumes single output value. However, multiple output values are possible in practice. A common approach is to model each output value as an independent GP model [165]. Additionally, new kernels can be constructed over multi-dimensional inputs by adding or multiplying between kernels defined on each individual input. An additive function can simply be expressed as $f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ and can easily be encoded into GP models. It allows a flexible way to model functions having more than one input. Consider the functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ drawn independently from a GP prior, for two different datasets $\mathcal{D}_1$ and $\mathcal{D}_2$,

$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \ , \tag{5.40}$$

$$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}')) \ , \tag{5.41}$$

$$\vdots$$

$$f_n(\mathbf{x}) \sim \mathcal{GP}(0, k_n(\mathbf{x}, \mathbf{x}')) \ , \tag{5.42}$$

where the input ouput data pairs is given by:

$$\mathcal{D}_1 = \{(\mathbf{x_1}, \mathbf{y_1}), \ldots (\mathbf{x_N}, \mathbf{y_N}\}) \ , \tag{5.43}$$

$$\mathcal{D}_2 = \{(\mathbf{x_2}, \mathbf{y_2}), \ldots (\mathbf{x_N}, \mathbf{y_N}\}) \ , \tag{5.44}$$

$$\vdots$$

$$\mathcal{D}_n = \{(\mathbf{x_n}, \mathbf{y_n}), \ldots (\mathbf{x_N}, \mathbf{y_N}\}) \ . \tag{5.45}$$

The distribution of the sum is given by

$$f_1(\mathbf{x}) + f_2 + \ldots f_n(\mathbf{x}) \sim \mathcal{GP}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \cdots \boldsymbol{\mu}_n, \quad \mathbf{K}_1 + \mathbf{K}_2 + \cdots \mathbf{K}_n) \ , \tag{5.46}$$

where $\mathbf{K} = k(\mathbf{x}, \mathbf{x}')$. The above expression can be used to sum any number of components. Consequently, the modelling of functions of multiple dimensions will result to additive structure across the dimensions such that

$$f(\mathbf{x}, \mathbf{x}') \sim \mathcal{GP}(0, \mathbf{K}_1 \mathbf{K}_2 \ldots \mathbf{K}_n) \ , \tag{5.47}$$

where the individual covariances are defined by $\mathbf{K}_1$, $\mathbf{K}_2, \cdots \mathbf{K}_n$ and assuming a 0 mean, respectively.

5.4.3 Proposed Approach

The multi-band image fusion presented here is motivated by the works in [19],[64] and extends them to multi-band images from heterogeneous imaging sensors. The aim is to combine two image modalities with different spatial and spectral characteristics. The approach considers the change of support problem inherent in image fusion where pixels are of different

resolution with their neighbours. The proposed approach is given in Fig 5.8.

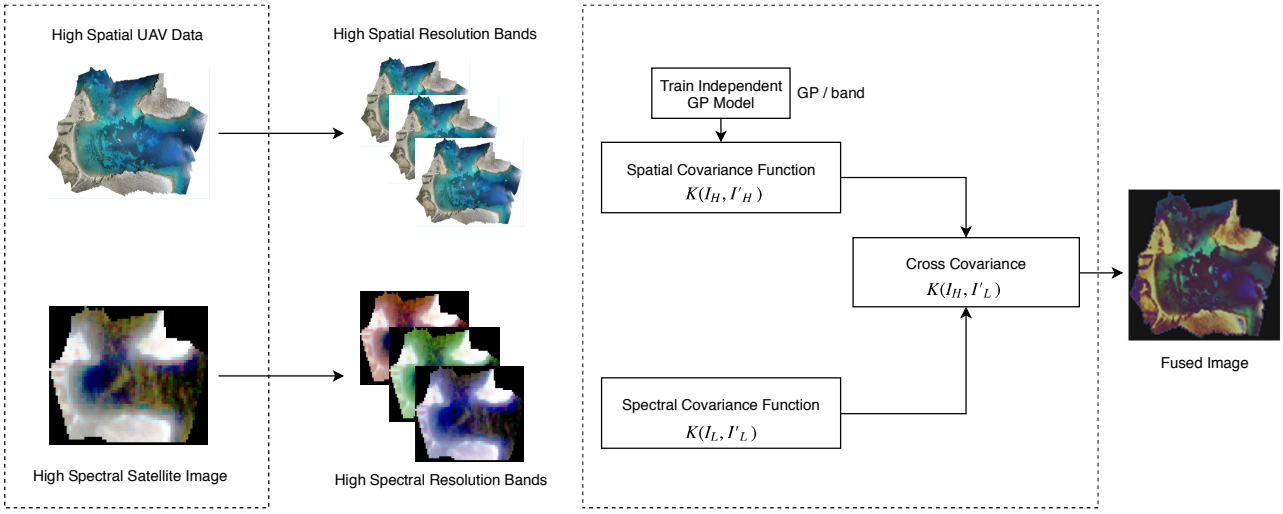


Figure 5.8: Multi-Band GP Regression Fusion Framework

The images to be fused are here represented as $\mathcal{I}\mathbf{H}_{m\times n\times\lambda}$ and $\mathcal{I}\mathbf{L}_{m\times n\times\lambda}$, representing the high spatial and high spectral resolution images acquired by two heterogeneous sensors for the same scene $\mathcal{X}$. Here, $._{m\times n}$ defines the spatial and $._{\lambda}$ the spectral extent of the images. The total number of measurements ($\mathcal{T}_z$) in each image is $._{m\times n\times\lambda}$ dimensions for $\mathcal{I}\mathbf{H}$ and $\mathcal{I}\mathbf{L}$, respectively. The images to be fused represent debased versions of the image to be recovered. For simplicity, we select common spectral channels in the visible range composed of Red, Green and Blue (RGB) in the images to be fused which means $\lambda = 3$. The problems we aim to solve are 1) To reconstruct a high-spatial and high spectral image from the complimentary images $\mathcal{I}\mathbf{H}$ and $\mathcal{I}\mathbf{L}$; 2) To build a covariance kernel that can handle pixel inference where the change of support problem exist; 3) To calculate the cross covariance that models the multi-modal and multi-band images. For this, we introduce a multi-output Gaussian Processes to model each spatial band of the high spatial resolution image. This forms the spatial prior of the image to be recovered.

5.4.3.1 Covariance Function

The covariance function or kernel plays a fundamental role in a GP model. It has the ability to encode our assumption of the function we want to model, and defines the correlation between function values [74]. The kernel is characterised by its hyperparameters that define its behaviour. A commonly used covariance function is the Squared Exponential (SE) with hyperparameters σ_f^2 and l. Conventionally, a covariance kernel is defined on $\mathbf{x} - \mathbf{x}'$ (point-

point bases) and explores the relationship between these points. However, when dealing with images the covariance function can be extended over areas by relating pixel observation to the function $\mathbf{f}(\mathbf{x})$, where a pixel observation is defined by $P_k(_{x_i,y_i,z})$, here $_{x_i,y_i}$ defines the geometric location of the *ith* pixel and $_z$ its intensity value at that location in the *ith* band of the multi-band image. A detailed derivation of defining covariance over areas of the image is given is discussed in [64].

$$\mathbf{P}(\mathbf{H_A}) = \frac{1}{|A_H|} \iint_{\mathbf{x}\in\mathbf{H_A}} f(x)dx \ , \tag{5.48}$$

where, $\mathbf{H_A}$ is the geometry areas per band of the high spatial resolution pixels, and $\mathbf{H_L}$ is the low spatial and high spectral pixels, the aim is firstly, to design the prior that defines the structure of the high spatial and high spectral image to be recovered. For this, we model the high spatial resolution pixels of the high spatial image by defining areas, where an area consists of the intensity of the pixel and the geometry of it at that location. We assume a simple average relationship between pixels and the GP function. The covariance of two high spatial resolution pixels in the *ith* band is then defined as

$$k_B(\mathbf{H_A}, \mathbf{H'_A}) = \frac{1}{|\mathbf{H_A}||\mathbf{H'_A}|} \iint_{\mathbf{x}\in\mathbf{H_A}} \iint_{\mathbf{x'}\in\mathbf{H'_A}} k(\mathbf{x}, \mathbf{x'})dxdx' \ , \tag{5.49}$$

where $k_B(\mathbf{H_A}, \mathbf{H'_A})$ defines the covariance between two high spatial resolution pixels in the *ith* band of the image, and B ranges from 1-number of bands in the high spatial resolution image, $|\mathbf{H'_A}|$ is the surface area of $\mathbf{H_A}$. In designing the prior covariance function, the following factors are considered

- Image data are normally non-smooth
- Image data exhibit discontinuity
- Spatial non-stationarity of images

The spatial information in $\mathcal{I}H$ is used as the input space of the prior covariance function added with the observed pixels to achieve contextual non-stationarity that addresses the discontinuity problem. Firstly, we explore an intrinsically sparse covariance introduced in [161] to enable sparsity and reduce computational complexity inherent in the full storage of the covariance matrix $\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}$. The proposed kernel is suitable for applications exhibiting discontinuities. It is smooth but not infinitely differentiable. The the intrinsically sparse covariance $\mathbf{M_S}$ is defined by

$$\mathbf{MS}_B(\mathbf{H_A}, \mathbf{H'_A}; \sigma_0, l)_i =$$

$$\begin{cases} \sigma_0 \left[\frac{2+\cos(2\pi \frac{d}{l})}{3} (1 - \frac{d}{l}) + \frac{1}{2\pi} \sin(2\pi \frac{d}{l}) \right] & \text{if } d < l \\ 0 & \text{if } d \geq l \ . \end{cases} \tag{5.50}$$

The proposed kernel is a function of the midpoint coordinates of areas (pixel coordinates and the intensity value at that location) in the *ith* band and reduces to zero for displacements larger than d. The variables σ_0 and l are the hyperparameters of the kernel. Here $\sigma_0 > 0$ is a constant coefficient that determines the average distance of $\mathbf{f}(\mathbf{x})$ from $\boldsymbol{\mu}_*$, $l > 0$ is the characteristic length scale that determines the length of change of $\mathbf{f}(\mathbf{x})$, and d is defined by

$$\mathbf{d}_i = |\mathbf{H_A} - \mathbf{H'_A}| \ . \tag{5.51}$$

The second step of the prior covariance design utilises the Rational Quadratic Kernel (RQK) [69] to model the pixel intensities of each band in the high spatial resolution image, allowing the pixels to be linked with the observed paired pixel geometries. The RQK is equal to putting together several SE kernels with different characteristic lengths that enables smoothness transfer. It improves efficiency in handling the change of support problem. Let $\mathbf{M_P}$ be the RQK kernel defined by

$$\begin{aligned} &\mathbf{MP}_B(\mathbf{I}(\mathbf{H_A}), \mathbf{I}(\mathbf{H'_A}); \alpha, \theta_p)_i \\ &= \sigma^2 \left(1 + \frac{(\mathbf{I}(\mathbf{H_A}) - \mathbf{I}(\mathbf{H'_A}))}{2\alpha\theta_p^2} \right)^{-\alpha} , \end{aligned} \tag{5.52}$$

Here, $\mathbf{IH_A}, \mathbf{IH'_A}$ represent the pixel intensity values in the $\mathbf{H_A}$ area of the *ith* band of the high spatial resolution image. The role of $\mathbf{M_S}$ is to ensure sparsity that improves computation while $\mathbf{M_P}$ links pixels within the covariance using the multi-band spatial information from the high spatial resolution multi-band image. The hyperparameters of the RQK are θ_p and α, respectively. Lastly, the prior covariance can be defined as the product of two independent kernels that models the spatial and pixel intensity of the high spatial resolution image forming the spatial base of the estimated image. This is mathematically given by

$$k_B(\mathbf{H_A}, \mathbf{H'_A}) = \sigma^2 \mathbf{MS}((\mathbf{H_A}, \mathbf{H'_A}; \boldsymbol{\theta_1}) \mathbf{MP}(\mathbf{I}(\mathbf{H_A}), \mathbf{I}(\mathbf{H'_A}); \boldsymbol{\theta_2}) \ . \tag{5.53}$$

The GP model is then trained using this kernel to learn the hyperparameters of the model

here represented as θ_1 and θ_2 where k_B represents the number of spatial bands in the high spatial resolution image. This means an independent GP is trained for each spatial band of the high spatial image. The product prior function augments the geometry of pixels and the intensities. In the next section we describe the learning of the hyperparameters of the model.

5.4.4 Hyperparameter Optimization

The hyperparameters of the GP model refers to the parameters of the prior covariance function evaluated on each band of the multi- band high spatial resolution image. The parameters of the model include $\boldsymbol{\mu}_*$, $\sigma_n^2\mathbf{I}$. The hyperparameter vector associated with the covariance function is defined as $\boldsymbol{\theta} = \{\sigma_0, \theta_p, \alpha, l\}$. In the proposed covariance function θ_p controls sensitivity and α determines the relative weighting for scale variations. When $\alpha \to \infty$, the behaviour of the RQK is similar to the SE kernel. It is important to initialise the parameters sensibly as they determine the quality of the estimated image to be recovered. A non-optimal solution is likely to produce fusion result that is blurry or even with high frequency artifacts [74].
To optimize $\boldsymbol{\theta}$, the marginal likelihood is maximised, which is given by:

$$P(\mathbf{y}|\mathbf{X}) = \int P(\mathbf{y}|\mathbf{f}, \mathbf{X})P(\mathbf{f}|\mathbf{X})d\mathbf{f}\ . \tag{5.54}$$

From (5.12) and (5.16), the likelihood $\mathbf{y}|\mathbf{f} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2\mathbf{I})$ and the model prior over the latent function $\mathbf{f}$ gives the logarithm of the marginal likelihood

$$log(p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})) = -\frac{1}{2}\mathbf{y}^T K_X^{-1}\mathbf{y} - \frac{1}{2}log|K_X| - \frac{n}{2}log2\pi. \tag{5.55}$$

In (5.55), the first term finds data fit, the second term is the model complexity term and the third term is a constant that ensures the marginal likelihood is robust to over-fitting.

5.4.5 Cross Covariance

The optimized $\boldsymbol{\theta}$ values from the optimization step are used to calculate the base covariance $k(\mathbf{H_A}, \mathbf{H'_A})$ that forms the spatial base of the estimated image to be recovered. While applying independent GP to multi-output regression problems is seen as a sub-optimal approach because the cross correlation is not put into consideration, the proposed approach solves this problem by mapping image areas where an area consists of both the pixel geometry and intensity values. $\mathbf{H_A}$ areas are discretely integrated by assuming they correspond to the $\mathbf{H_L}$ areas as shown in [64].

The optimized base covariance is then used to calculate the cross covariance $k(\mathbf{H_A}, \mathbf{H'_L})$ per band in $\mathbf{H_L}$. The role of this kernel is to couple per band the high spatial resolution pixels with the corresponding high spectral pixels. The cross covariance is given by

$$k_B(\mathbf{H_A}, \mathbf{L'_A}) = \frac{1}{\mathcal{T}_\mathcal{H}} \sum_{\mathbf{H'_A} \in \mathbf{L'_A}} k(\mathbf{H_A}, \mathbf{H'_A}) \ , \tag{5.56}$$

where $\mathcal{T}_\mathcal{H}$ defines the total number of $\mathbf{H_A}$ areas correponding to $\mathbf{H_L}$ areas. The function is evaluated over the corresponding bands in $\mathbf{H_L}$. Similarly, the cross covariance between areas in $\mathbf{H_L}$ can be calculated using

$$k_B(\mathbf{L_A}, \mathbf{L'_A}) = \frac{1}{\mathcal{T}_\mathcal{H} \mathcal{T}'_\mathcal{H}} \sum_{\mathbf{H_A} \in \mathbf{L_A}} \sum_{\mathbf{H'_A} \in \mathbf{L'_A}} k(\mathbf{H_A}, \mathbf{H'_A}) \ . \tag{5.57}$$

5.4.6 Image Fusion and Reconstruction

To fuse the multi-band image modalities and reconstruct the estimated image, the training data of the model is extracted from the high spatial resolution image $\mathbf{H_A}$ consisting of the spatial geometries of pixels and the augmented intensity values within the bands (3 high spatial bands) observed. The GP model is then queried over the high spectral image $\mathbf{H_L}$ pixel areas and their intensities. The reconstruction is done by querying the predictive mean of the model (see (5.12)-(5.14)), again we evaluate this function by querying corresponding bands in the modalities. A constant mean value of 0.5 is assumed because image values are continues within the range of 0-1. The predictive mean in (10) becomes

$$\mathcal{I}_B^* = \bar{} + k_B(\mathbf{H_A}, \mathbf{H_L}) \left[k_B(\mathbf{L_A}, \mathbf{L'_A}) + \sigma_n^2 I \right]^{-1} (\mathbf{I_L}(\mathbf{L_i}) - \mu) \ . \tag{5.58}$$

5.4.7 Non-reference Image Fusion Performance Metric

It is important to evaluate the performance of the fusion model to validate the result. Conventionally, this is done by comparing the fusion results with a reference image using fusion performance metrics e.g. image Correlation Coefficient (CC) [8], Reconstruction Error (RE) or the Universal Image Quality Index (UIQI) [21]. In situations where a reference image is not available, non-reference image fusion metrics have been developed. We propose to use the Fast- Feature Mutual Information (Fast-FMI) introduced in [166] as a measure to validate the fusion model. Fast-FMI calculates the mutual information between corresponding regions

in the fused and source images, respectively. Firstly, the mutual information is normalized using

$$\frac{\ell(\mathcal{I}_i) + \ell(\mathcal{I}^*)}{2}\,, \tag{5.59}$$

Here, $\ell(\mathcal{I}_i)$ and $\ell(\mathcal{I}^*)$ defines the entropies of corresponding windows in the source images and the fused image, respectively. Secondly, the mutual information between the source images and the fused image is defined by

$$\mathbf{I}(i;\mathcal{I}^*) = \frac{2}{n}\sum_{i=1}^{n}\frac{\mathbf{I}_i(\mathcal{I}_i;\mathcal{I}^*)}{\ell(\mathcal{I}_i) + \ell(\mathcal{I}^*)}\,. \tag{5.60}$$

Finally, the non-reference fusion metric is given by

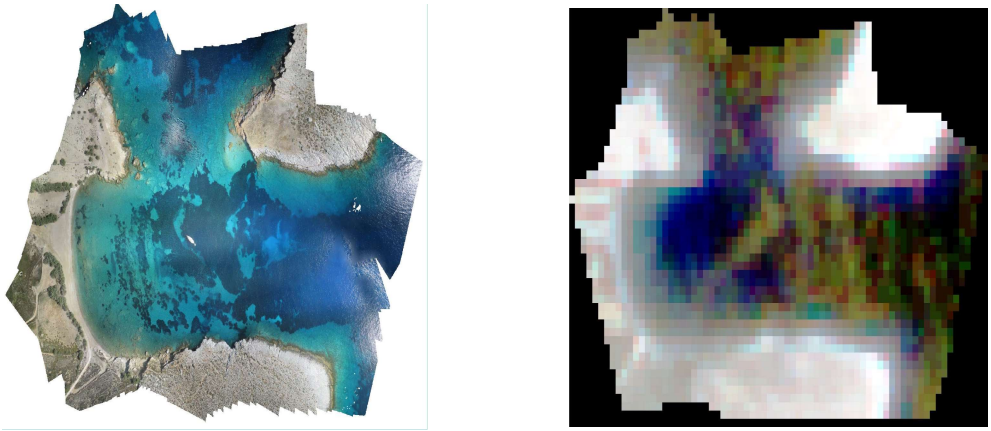
$$FMI_{\mathcal{I}^*}^{\mathcal{I}_i} = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\mathbf{I}_i(\mathcal{I}_H;\mathcal{I}^*)}{\ell(\mathcal{I}_H) + \ell(\mathcal{I}^*)} + \frac{\mathbf{I}_i(\mathcal{I}_L;\mathcal{I}^*)}{\ell(\mathcal{I}_L) + \ell(\mathcal{I}^*)}\right). \tag{5.61}$$

5.4.8 Dataset

In this section, the dataset utilised for the image fusion is described. Multi-band data from a UAV and satellite data from the sentinel-2 satellite are utilised. While the UAV data is high in spatial resolution, it lacks spectral information. On the other hand, the satellite data has low spatial resolution and high spectral resolution. The aim of this model therefore, is to bring the complimentary information from the two heterogenous sensors to reconstruct a high spatial and high spectral image for the scene of interest. Sentinel-2 is a multispectral earth observation satellite and forms part of the European Union (EU) Copernicus Programme for environmental monitoring. It has 13 bands in the visible, near infrared and short wave infrared part of the spectrum, with 10m spatial resolution in the visible range comprising the Red, Green, Blue and Near Infrared bands.

5.5 Multi-band Image Fusion Approach Results

In this section the multi band image fusion results are presented. The result of the proposed approach is compared with the SA method that sums up the images and take the average as the fusion result and also the DWT transform method presented in the previous chapter. It also evaluates the performance of the proposed Multi-Output variable GP model. The images are registered with pixel correspondence established between them. The purpose



(a) (b)

Figure 5.9: Quick Look of UAV and Sentinel-2 Multi Band Images

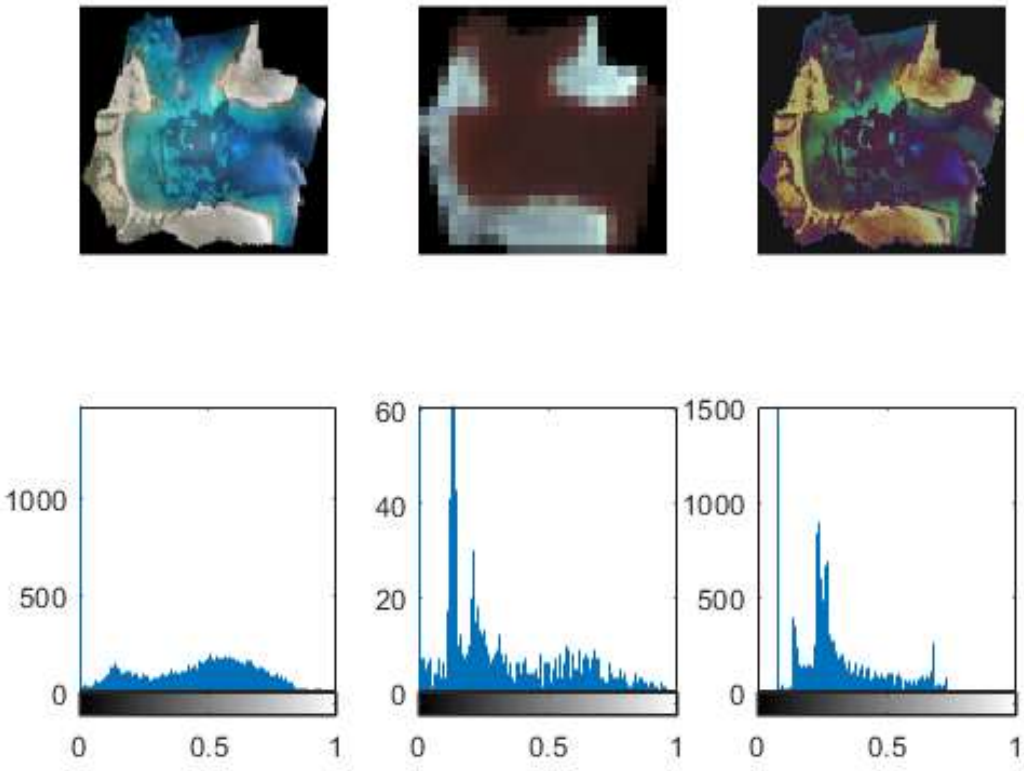
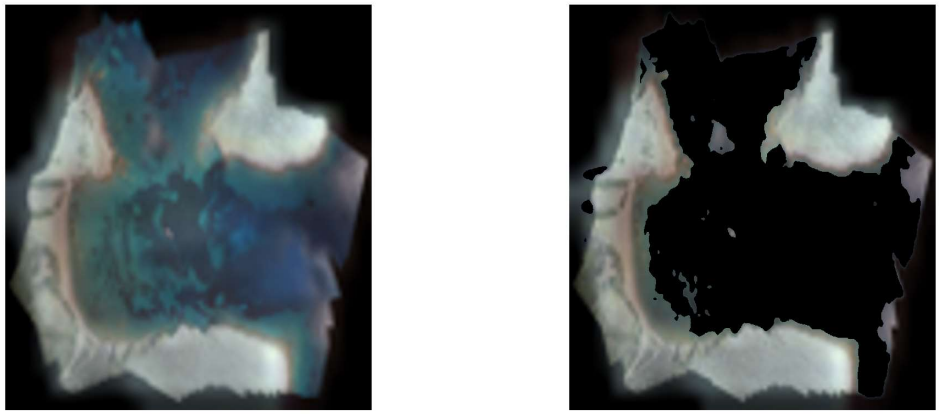
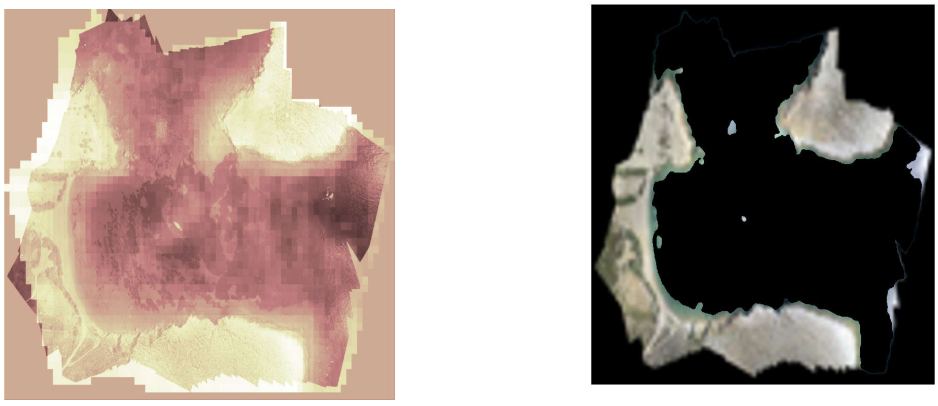


Figure 5.10: (Top) left-right: UAV Image, Satellite Image, Proposed Result (below) Corresponding Histograms



(a) (b)

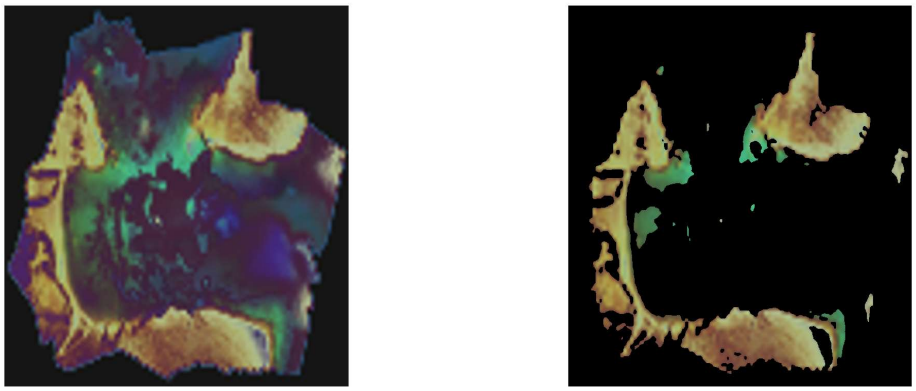
Figure 5.11: Fused Image and Segmented ROI Image using the Simple Average Method



(a) (b)

Figure 5.12: Fused Image and Segmented ROI Image Using the Method

of image registration as highlighted in the previous chapter is to determine the geometric transformation that aligns one image to the other [133]. The aim, is to recover a high spatial and high spectral resolution image by fusing the two complimentary images. Firstly, the images are resized to 100×100 and normalized so that image pixels are in the range of 0-1 and to reduce the computational complexity inherent in regression models. Secondly, common spectral bands in the visible range (Red, Green and Blue) are selected from the images. Pixel geometry per spatial band in $\mathcal{I}_{\mathcal{H}}$ and the corresponding intensity values have been extracted forming $\mathbf{H_A}$ and $\mathbf{I}(\mathbf{H_A})$, respectively. The variables extracted form the input of the prior function are shown in (5.52) - (5.57). The model is then trained using the proposed base kernel and the marginal likelihood is maximised to optimize the hyperparameters of the



(a) (b)

Figure 5.13: Fused Image and Segmented ROI Image Using Proposed Multi-Band Fusion Approach

model. A cross covariance is then calculated using the optimized base kernel that couples the high spatial pixels and the corresponding high spectral bands per band using a mapping of 4:1 corresponding areas between $\mathbf{H_A}$ and $\mathbf{H_L}$, respectively. In Fig. 5.10, the high spatial and high spectral resolution images with the fused image result are shown using the considered model and their corresponding image histogram. The image histogram shows the graphical representation of pixel intensity on the x-axis and the corresponding number of pixels on the y-axis. The results shows that there are more pixels in the fused image when compared to the individual images, implying more information (pixels) in the fused image.

In Fig. 5.11-5.13, result of the proposed multi-band fusion approach is compared with the results from applying SA and IWT methods. Simple averaging, sums the source images and finds the average while the wavelet approach decomposes the images into approximations and fuses them using a rule. Both methods do not consider spatial location of pixels and their corresponding intensities while performing fusion. However, the proposed approach on the other hand, uses both characteristics of the image as the input space to the model. This allows transferral of information between the image modalities that enables local structure of images to be learned. In the results, simple averaging shows ringing leading to artifacts and blurred image, while the inverse wavelet transform exhibits mosaic effects which could be due to small decomposition scale. The proposed method however, shows visually appealing image with improved edges that alleviates artifacts as can be seenn in the top right corner of Fig. 5.10.

Additionally, the performance of the proposed method is further evaluated by segmenting

the fused image using K-means clustering algorithm [19] and comparing the segmented results with the other methods. Again, the proposed method shows compelling results, picking up more objects from the image and the ROI than the other methods as shown in Fig. 5.13(b). This compelling performance can be attributed to the ability of the proposed approach to sharpen the resolution of the spectral channels that supports the pixel coordinates from the high spatial image, hence improving the edges.

Finally, the Fast-FMI non-reference image fusion performance metric described in (5.59)-(5.61) is utilised to validate the performance of the proposed approach. The proposed method is compared with other established image fusion methods and subjected to the performance test. Visual comparisons in Fig. 5.10 (top) shows the proposed method having visually appealing performance than the other methods. A window of 3×3 is used which corresponds to evaluating $(3 \times 3)^2$ regions and finding the MI between them. Fast-FMI performance metric reduces the computational complexity from $\mathcal{O}(n)^2$ to $\mathcal{O}(n)$ when compared to similar non-reference performance measures.

Table 5.3: Fast-FMI Results

Method	Fast-FMI
Simple Averaging	0.4321
IWT	0.4532
Proposed	0.3954

Table 5.3 above, shows the results obtained by running the FAST-FMI algorithm on the fused image. The results are compared with the proposed multi-band GP fusion approach, IWT, and SA methods. Again, as described earlier the FAST-FMI algorithm validates the performance of the fusion approach where a reference data is not available. The individual images fused are compared with the fused image and the mutual information retained by the fused image is used to measure the performance of the fusion algorithm. The proposed approach performs better than the other methods. This is attributed to the new multi band covariance kernel and its ability to learn the individual bands seperately, hence extracting useful information for the image reconstruction.

5.6 Summary

In this chapter, a GP regression approach is utilised to fuse multi-modal images from different sensors and platforms with different spatial and spectral resolutions. The approach

utilises the spatial location of pixels and their corresponding intensity values to model the spatial envelope of the fused image to be recovered. The new covariance function, which is a product of an intrinsically sparse covariance that defines the behaviour of the model is sparse in nature enabling faster computation of input output points and improving model performance. The Rational Quadratic Kernel (RQK) models the pixel coordinates and intensity of the high spatial resolution image. Additionally, the new covariance serves as the base stochastic prior for the estimated image to be recovered. Furthermore, the approach enables the transfer of information between the different modalities while allowing local image structures of the high spatial image to be transferred unto high spectral image. The hyperparameters of the model as described in this chapter are learnt by maximising the marginal likelihood over the input points(geometrical locations and intensity values) of the high spatial resolution image. The learnt parameters are used the evaluate the cross covariance that model the high spatial areas and the spectral areas using a mapping of 4:1 that corresponds to the ratio of difference in resolution of the individual images (20:5) in this case. The sparsity of the covariance kernel introduced is modified to accommodate multi-band data where the images are in different modalities and both having more than a band each. For simplicity, bands in the visible range (RGB) of the EMS are selected, and a GP is trained for each band in the spectral space. The performance of the proposed fusion model is analysed using both referenced and non-referenced quality assessment measures with the proposed method exhibiting better performances than the compared methods. In addition to image fusion performance metrics, a further segmentation step of the fused images is done using K-means segmentation algorithm where the ROI is segmented for each of the compared results. Again, the proposed approach shows compelling results and picking up more objects in the image than the other methods. This is attributed to the ability of the proposed method to handle the change of support problem inherent in images of multi-modal nature using appropriate covariance envelope to model pixel inferences [64]. The change of support problem is associated with situations where a pixels resolution is different from its nearest neighbourhood.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Technological limitations as a result of spatial and spectral trade-off in sensor design, allow the acquisition of diverse data from different remote sensing imaging sensors onboard divergent platforms (satellite or airborne). In this work, data acquired by such sensors are improved to take advantage of the complimentary information from the different systems through image or sensor data fusion, which enables the combination of data originating from different sources. Through data fusion, and the approaches presented, researchers and scientists can explore new methods, develop robust approaches and implement applicable solutions towards disaster mitigation, detection and analysis of oil spills. This thesis has presented advanced methods for fusion of remote sensing images and oil spill segmentation. A new framework is presented that includes a robust fusion approach using machine learning method to combine images acquired by different sources. The methods proposed are:

- A new framework is developed for oil spill segmentation in fused SAR images. The framework takes as input complimentary SAR images of a same scene acquired by different platforms at different time intervals. Firstly the images are registered in time and space by extracting from the images SIFT features and matching corresponding features so that per-pixel alignment is achieved. DWT pixel level image fusion is the utilised to fuse the images using a select maximum fusion rule. The segmentation of dark areas on the fused image is carried out using step is done by combining a polynomial fitting curve segmentation approach and edge detection.
- A new image fusion approach is developed using GP regression. The proposed ap-

proach is based on the design of a non-stationary covariance kernel which is a product of a sparse covariance to handle the change of support problem and a rational quadratic kernel that models the geometry of pixels and their corresponding intensities forming the spatial resolution prior of the estimated image to be recovered.

- A new Adaptive multi-output GP to fuse multi-band images is introduced. The model utilises spatial information of the high spatial resolution bands using a sparse covariance kernel to learn the geometry of image pixels forming a base prior of spatial bands for the estimated image to be recovered. A new learning strategy is adopted where individual bands of the multi-band images are learnt separately to improve the overall fusion process.

The rest of this chapter is structured as follows: section 6.2 summarises the chapters presented in the thesis with some important conclusions derived from the study. Finally, section 7.2 of this chapter outlines directions for future work that are based on the study and conclusions that are arrived at from the proposed approaches.

6.2 Summary

The work presented in this thesis, develops new framework that utilises image fusion for oil spill segmentation. Firstly, remote sensing systems utilised for oil spill detection and monitoring are reviewed to introduce the reader to the diverse remote sensing systems utilised for the application (oil spill). Furthermore, the advantages and disadvantages of these systems are discussed. State of the art methods for image fusion and segmentation are introduced. In chapter three, a qualitative and quantitative review of hyperspectral pansharpening methods is carried out. Additionally, the pansharpened results are further examined for spectral unmixing of materials in the scene of interest using a reference hyperspectral image of the Gulf of Mexico oil spill acquired by SPECTIR hyperspectral sensor. The result shows significant improvement in quality of images from the different methods especially for the sparse Bayesian method where prior spatial and spectral parameters of the estimated image are considered in the fusion process, although at a high cost of computation. Additionally, the introduction of sparsity in the fusion model allows the regularisation of the fusion process. In unmixing the endmembers, non-linear approaches especially the PPNM method look promising with the pansharpened images when compared to the hyperspectral image.

In chapter four, a framework for oil spill segmentation in fused SAR is presented. Images of the Gulf of Mexico oil spill obtained by Radarsat-2 and EnviSat SAR systems both in the C-band of the microwave region of the EMS are utilised. The images are in different spatial resolutions and were acquired at different times from different angles. The aim is to combine image data of a same scene acquired at different times, from different angles and different imaging sensors and to compare the spread of oil over time. The framework begins with a pre-processing step that reduces the speckles inherent in SAR images as a result of backscattering effect and to enhance the image using a Gaussian filter. Next the images are aligned so that a common frame is achieved. The registration step is done by utilising the SIFT algorithm that extracts invariant features from both images and performs a matching step of corresponding features. The registration is completed by resampling and transformation of the images using the RANSAC algorithm. The fusion stage is actualised with the Wavelet transform that transforms the co-registered images from spatial to frequency domain and decomposed into details and approximation coefficients that are then combined using a fusion rule and the fused image is recovered by taking the inverse of the wavelets. The framework is concluded with a segmentation step using polynomial fitting curve on the histogram of the fused image and adopting a global threshold to distinguish foreground and background extended with the Canny edge detector that enables the detection of boundary lines of dark formations (assumed oil spill) of the fused image. The performance of the fusion approach is done using the Wald's protocol over four levels of decomposition. It is observed from this chapter, that image fusion with wavelets creates mosaic effects when the scale of decomposition is not selected appropriately. For this reason, a more efficient model need to be developed so that decomposition scale can be automatically selected according to the data.

In chapter five, a robust fusion approach is developed with Gaussian Process (GP) regression to solve the problems of mosaic in the resulting fused images as was the case with wavelets and to solve the problem of inter-modality issues with different images. Additionally, to consider also, inter-band dependency problem and the change of support problem. To address these problems, a non-stationary covariance kernel is utilise to handle the change of support problem and adopting a prior covariance function that models image pixels and their corresponding intensity values as input to the GP model. The high spatial resolution image is usedto model the covariance since it forms the spatial base of the estimated image

to be recovered. The approach also considers that GP priors are adaptable for inter-modality data compression with multiple output behaviours. The introduction of sparsity in the model allows for the accommodation of large data and many pixels and lowers the computational complexity that is inherent in modelling point-point inputs in a GP. The covariance function is further extended to handle multi-band images where the input images both contain spectral and spatial channels. The channels are modelled as outputs of independent GP model and new covariances formed by adding the individual covariances of each individual input allowing for a flexible way to model the inter-band dependencies between images. The resultant fused image shows compelling enhancement result when compared to other methods. It is noticeable that the approach is responsive to image registration errors as with other fusion approaches. The results are further examined by segmenting them using K-means colour segmentation. The value of the centriods is selected to be three to tally with the spectral channels of the fused images. Again, the approach developed shows better segmentation results with no artifacts when compared with other methods. This can be attributed with the ability of the proposed method to sharpen the resolution of the image and enhance the edges.

6.3 Direction for Future Work

- In chapter three of this thesis, pansharpening methods are reviewed and the results are compared for spectral unmixing of endmembers in the scene of interest to measure the spectral abundance of the materials. Spectral unmixing allows the identification of the spectral signatures of materials and enables the quantification of their spatial distribution over an image. It is also shown that the unmixing procedure can be achieved using linear or non-linear methods depending on the mixing process or model. However, it is difficult to accurately guess the mixing model and to select the appropriate method that is best for unmixing the endmembers based on a given data as there is no defined or developed criteria for this. Therefore, there is need to develop un-surpervised methods that can automatically select an appropriate unmixing method given the data. An attempt to solve this problem is presented in [27] where a hybrid method is developed using artificial neural networks trained over pixel windows of the image to switch between linear and non-linear spectral unmixing methods. Other approaches is to utilise deep learning to better interpret the data in an un-supervised way. This is needed to fully appreciate the effect of pansharpening methods for different applications including

for oil spill monitoring in optical images.

- In chapter four, an oil spill framework with fused SAR images is introduced. The framework includes pre-processing, image registration and then fusion using wavelets. An important requirement for fusing multi-resolution and multi-modal images is an effective and robust registration technique. The registration is to align the images to be fused so that correspondence between image pixels is established and the geometric transformation that aligns one image to another [9]. In the fusion results presented in chapter four, there is a visible mosaic issue in the fused image. This is partly due to selection of decomposition scale and mis-alignment of pixels in the registration process taht results in registration errors. It is observed also, that a single pixel misalignment will produce poor performance metrics when compared with a reference image as illustrated in the SAM results of the wavelet methods. The registration step needs to be improved by taking accurate measurements of the spatial coordinates of pixels so that errors in pixel locations can be reduced drastically. This is possible in satellite systems where sensor calibration information is given.

- In chapter 5, a Gaussian Process (GP) regression approach for fusion of multi-modal, multi-resolution SAR images is proposed. A sparse covariance kernel is utilised to model the geometry and intensity of pixels thereby reducing the pixel mis-alignment from the registration step. However, the process of learning complex datasets as remote sensing images through explicable covariance kernels under normal circumstance of a GP requires $\mathcal{O}(n^3)$ for computation and $\mathcal{O}(n^2)$ for storage. This problem limits the performance of GP when dealing with large datasets for example multi-band images. In the future, there is the need to put into consideration while designing custom covariances that can accomodate large datasets running into hundreds of thousands of pixels. This can be achieved by using inducing points methods introduced in [167] where the GP is decomposed variationally, depending only on a set inducing variables that factorises the GP model and enable variation inference. Or the use of a more recent approach by [68] that utilises a structured kernel interpolation (SKI) framework that generalises the inducing points for scalable GP's. In this approach kernel approximations can be utilised to aid faster computations through kernel interpolation and the effect of inducing point and the number of inducing points appropriate can be established.

- Lastly, There is need to compare the different fusion performances for the classification

of oil spill by selecting appropriate image features to serve as the input of the classifier to complete the detection process as described in many oil spill detection frameworks. A combination of multi-modal images e.g. hyperspectral images with SAR images and SAR images in different bands e.g. L-band SAR with C-band SAR is also desirable, since the fusion methods developed in this research only concentrated on multi-resolution images.

Bibliography

[1] M. Fingas and C. Brown. Review of oil spill remote sensing. *Marine Pollution Bulletin*, 83(1):9–23, 2014.

[2] A. Sheta, M. Alkasassbeh, M. Braik, and H.A. Ayyash. Detection of oil spills in sar images using threshold segmentation algorithms. *International Journal of Computer Applications*, 57(7), 2012.

[3] M. Jha, J. Levy, and Y. Gao. Advances in remote sensing for oil spill disaster management: state-of-the-art sensors technology for oil spill surveillance. *Sensors*, 8(1):236–255, 2008.

[4] A.D. Lawal, G. Radice, M. Ceriotti, and A.U. Makarfi. Investigating SAR algorithm for spaceborne interferometric oil spill detection. *International Journal of Engineering and Technical Research*, 4(3):123–127, 2016.

[5] C. Brekke and A.H. Solberg. Oil spill detection by satellite remote sensing. *Remote sensing of environment*, 95(1):1–13, 2005.

[6] S. Khanna, M.J. Santos, S.L. Ustin, K. Shapiro, P.J. Haverkamp, and M. Lay. Comparing the potential of multispectral and hyperspectral data for monitoring oil spill impact. *Sensors*, 18(2):558, 2018.

[7] M. Roser. Oil spills. *Our World in Data*, 2019. https://ourworldindata.org/oil-spills.

[8] F.S. Longman, L. Mihaylova, and D. Coca. Oil spill segmentation in fused synthetic aperture radar images. In *Proceedings of the 4th International Conference on Control Engineering & Information Technology (CEIT)*, pages 1–6. IEEE, 2016.

[9] W. Jia, J. Zhang, and J. Yang. SAR image and optical image registration based on contour and similarity measures. *Proc. GSEM*, 2009.

[10] I. Leifer, W.J. Lehr, D. Simecek-Beatty, E. Bradley, R. Clark, P. Dennison, Y. Hu, S. Matheson, C.E. Jones, B. Holt, et al. State of the art satellite and airborne marine oil spill remote sensing: Application to the bp deepwater horizon oil spill. *Remote Sensing of Environment*, 124:185–209, 2012.

[11] X. Bresson and T.F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse problems and imaging*, 2(4):455–484, 2008.

[12] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

[13] M. Cococcioni, L. Corucci, and B. Lazzerini. Issues and preliminary results in oil spill detection using optical remotely sensed images. In *Proceedings of OCEANS 2009-EUROPE*. 2009.

[14] S. Pelizzari and J.M. Bioucas-Dias. Bayesian segmentation of oceanic sar images: Application to oil spill detection. *arXiv preprint arXiv:1007.4969*, 2010.

[15] A.S. Solberg, G. Storvik, R. Solberg, and E. Volden. Automatic detection of oil spills in ers sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4):1916–1924, 1999.

[16] A.H. Solberg, C. Brekke, and P.O. Husoy. Oil spill detection in Radarsat and Envisat SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):746–755, 2007.

[17] A.H.S. Solberg. Remote sensing of ocean oil-spill pollution. *Proceedings of the IEEE*, 100 (10):2931–2945, 2012.

[18] K.N. Topouzelis. Oil spill detection by sar images: dark formation detection, feature extraction and classification algorithms. *Sensors*, 8(10):6642–6659, 2008.

[19] F.S. Longman, L. Mihaylova, and L. Yang. A Gaussian process regresion approach for fusion of remote sensing images for oil spill segmentation. In *proceedings of 2018 21st International Conference on Information Fusion (FUSION)*, pages 62–69. IEEE, 2018.

[20] S.A. Pelizzari. *Oil Spill Detection Using SAR Images*. PhD thesis, Ph. D thesis, Lisboa: UTL. Instituto Superior Técnico, 2011.

[21] Q. Wei, N. Dobigeon, and J.Y. Tourneret. Bayesian fusion of multi-band images. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1117–1127, 2015.

[22] R. Arablouei. Fusing multiple multiband images. *Journal of Imaging*, 4(10):118, 2018.

[23] K. Pilžis and V. Vaišis. Aplinkos apsaugos inžinerija/environmental protection engineering.

[24] M. Fingas and C. Brown. A review of oil spill remote sensing. *Sensors*, 18(1):91, 2018.

[25] M. Fingas. *Oil spill science and technology*. Gulf professional publishing, 2016.

[26] S. Hennig. Exploring the benefits of active vs. passive spaceborne systems. *Earth Imag. J.*, 10(6), 2013.

[27] A.M. Ahmed, O. Duran, Y. Zweiri, and M. Smith. Hybrid spectral unmixing: Using artificial neural networks for linear/non-linear switching. *Remote Sensing*, 9(8):775, 2017.

[28] S. Sun, C. Hu, L. Feng, G.A. Swayze, J. Holmes, G. Graettinger, I. MacDonald, O. Garcia, and I. Leifer. Oil slick morphology derived from aviris measurements of the deepwater horizon oil spill: Implications for spatial resolution requirements of remote sensors. *Marine pollution bulletin*, 103(1-2):276–285, 2016.

[29] J.W. Salisbury and D.M. D'Aria. Infrared (8–14 μm) remote sensing of soil particle size. *Remote Sensing of Environment*, 42(2):157–165, 1992.

[30] F.T. Ulaby, R.K. Moore, and A.K. Fung. Microwave remote sensing: Active and passive. volume 3-from theory to applications. 1986.

[31] Y.K. Chan and V.C. Koo. An introduction to synthetic aperture radar (sar). *Progress In Electromagnetics Research*, 2:27–60, 2008.

[32] M. Fingas and C.E. Brown. Oil spill remote sensing: a review. In *Oil spill science and technology*, pages 111–169. Elsevier, 2011.

[33] B.C. Douglas. Global sea rise: a redetermination. *Surveys in Geophysics*, 18(2-3):279–292, 1997.

[34] M. Migliaccio. Synthetic Aperture Radar for oil spill monitoring: a brief review. *in Proceedings of Anais XVII Brazilian Symposium on Remote Sensing*, 2015.

[35] K. Topouzelis, V. Karathanassi, P. Pavlakis, and D. Rokos. Oil spill detection: SAR multi-scale segmentation & object features evaluation. In *In Proceedings of the 9th International symposium on remote sensing (SPIE)*, pages 77–87, 2002.

[36] K. Topouzelis, V. Karathanassi, P. Pavlakis, and D. Rokos. Detection and discrimination between oil spills and look-alike phenomena through neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(4):264–270, 2007.

[37] F. Chen, R. Lasaponara, and N. Masini. An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring. *Journal of Cultural Heritage*, 23:5–11, 2017.

[38] M. Cherniakov. *Bistatic radar: emerging technology*. John Wiley & Sons, 2008.

[39] W. Wang, Y. Ji, and X. Lin. A novel fusion-based ship detection method from pol-sar images. *Sensors*, 15(10):25072–25089, 2015.

[40] M. Migliaccio, A. Gambardella, and M. Tranfaglia. Sar polarimetry to observe oil spills. *IEEE Transactions on Geoscience and Remote Sensing*, 45(2):506–511, 2007.

[41] M. Migliaccio, F. Nunziata, and A. Buono. Sar polarimetry for sea oil slick observation. *International Journal of Remote Sensing*, 36(12):3243–3273, 2015.

[42] Y. Yamaguchi. Disaster monitoring by fully polarimetric sar data acquired with alos-palsar. *Proceedings of the IEEE*, 100(10):2851–2860, 2012.

[43] H. Zheng, Y. Zhang, Y. Wang, X. Zhang, and J. Meng. The polarimetric features of oil spills in full polarimetric synthetic aperture radar images. *Acta Oceanologica Sinica*, 36 (5):105–114, 2017.

[44] H. Li, W. Perrie, Y. He, J. Wu, and X. Luo. Analysis of the polarimetric sar scattering properties of oil-covered waters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8):3751–3759, 2014.

[45] P. Gil and B. Alacid. Oil spill detection in terma-side-looking airborne radar images using image features and region segmentation. *Sensors*, 18(1):151, 2018.

[46] A.J. Gallego, P. Gil, A. Pertusa, and R. Fisher. Segmentation of oil spills on side-looking airborne radar imagery with autoencoders. *Sensors*, 18(3):797, 2018.

[47] K. Grüner, R. Reuter, and H. Smid. A new sensor system for airborne measurements of maritime pollution and of hydrographic parameters. *GeoJournal*, 24(1):103–117, 1991.

[48] M.F. Fingas and C.E. Brown. Review of oil spill remote sensing. *Spill Science & Technology Bulletin*, 4(4):199–208, 1997.

[49] R. Chandrakanth, J. Saibaba, G. Varadan, P.A.D. Raj, et al. Fusion of high resolution satellite SAR and optical images. In *Proceedings of 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM)*, pages 1–6. IEEE, 2011.

[50] V.R. Pandit and R. Bhiwani. Image fusion in remote sensing applications: A review. *International Journal of Computer Applications*, 120(10), 2015.

[51] R. Heideklang and P. Shokouhi. Decision-level fusion of spatially scattered multi-modal data for nondestructive inspection of surface defects. *Sensors*, 16(1):105, 2016.

[52] Z. Wang, Y. Tie, and Y. Liu. Design and implementation of image fusion system. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, volume 10, pages V10–140. IEEE, 2010.

[53] L. Loncan, L.B. De Almeida, J.M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G.A. Licciardi, M. Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geoscience and remote sensing magazine*, 3(3):27–46, 2015.

[54] M. Berbar, S. Gaber, and N. Ismail. Image fusion using multi-decomposition levels of discrete wavelet transform. In *Proceedings of International Conference on Visual Information Engineering, 2003. VIE 2003.*, pages 294–297. IET, 2003.

[55] G. Pajares and J.M. De La Cruz. A wavelet-based image fusion tutorial. *Pattern recognition*, 37(9):1855–1872, 2004.

[56] K. Amolins, Y. Zhang, and P. Dare. Wavelet based image fusion techniques—an introduction, review and comparison. *ISPRS Journal of photogrammetry and Remote Sensing*, 62(4):249–263, 2007.

[57] H. Li, B. Manjunath, and S.K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3):235–245, 1995.

[58] Q. Tian, N. Sebe, M.S. Lew, E. Loupias, and T.S. Huang. Image retrieval using wavelet-based salient points. *Journal of Electronic Imaging*, 10(4):835–850, 2001.

[59] R. Guida, S.W. Ng, and P. Iervolino. S-and X-band SAR data fusion. In *2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pages 578–581. IEEE, 2015.

[60] I. Kuzmanić, J. Šoda, Z. Kulenović, and M. Vujović. Inversion of wavelet coefficients in oil spills detection in radar images for environment risk reduction in adriatic sea. In *2009 International Symposium ELMAR*, pages 33–36. IEEE, 2009.

[61] X. Yang, J.H. Pei, and W. Yang. Disadvantage of the methods based on wavelet transform in high-resolution and multispectral fusion image. *Journal of Infrared and Millimeter Waves*, 21(1):77–80, 2002.

[62] J. Sander and J. Beyerer. Bayesian fusion: modeling and application. In *2013 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6. IEEE, 2013.

[63] G. Winkler. *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*, volume 27. Springer Science & Business Media, 2012.

[64] A. Reid, F. Ramos, and S. Sukkarieh. Bayesian fusion for multi-modal aerial images. In *Robotics: Science and Systems*, 2013.

[65] Z. Ge, B. Wang, and L. Zhang. Remote sensing image fusion based on bayesian linear estimation. *Science in China Series F: Information Sciences*, 50(2):227–240, 2007.

[66] A.G. Wilson. *Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes*. PhD thesis, University of Cambridge, 2014.

[67] M.P. Deisenroth, R.D. Turner, M.F. Huber, U.D. Hanebeck, and C.E. Rasmussen. Robust filtering and smoothing with gaussian processes. *IEEE Transactions on Automatic Control*, 57(7):1865–1871, 2011.

[68] A. Wilson and H. Nickisch. Kernel interpolation for scalable structured gaussian processes (kiss-gp). In *International Conference on Machine Learning*, pages 1775–1784, 2015.

[69] C.E. Rasmussen and C.K. Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.

[70] W. Aftab, A. De Freitas, M. Arvaneh, and L. Mihaylova. A Gaussian process approach for extended object tracking with random shapes and for dealing with intractable likelihoods. In *Digital Signal Processing (DSP), 2017 22nd International Conference on*, pages 1–5. IEEE, 2017.

[71] C.E. Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, 2003.

[72] K.P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[73] A.S. Reid. *Gaussian Process Models for Analysis of Remotely Sensed Geo-Spatial Data*. University of Sydney, 2011.

[74] H. He and W.C. Siu. Single image super-resolution using gaussian process regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*, pages 449–456. IEEE, 2011.

[75] C.M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[76] G. Malkomes, C. Schaff, and R. Garnett. Bayesian optimization for automated model selection. In *Advances in Neural Information Processing Systems*, pages 2900–2908, 2016.

[77] X. Fan, R. Gao, and Y. Wang. Example-based haze removal with two-layer gaussian process regressions. 2014.

[78] E. Gilboa, J.P. Cunningham, A. Nehorai, and V. Gruev. Image interpolation and denoising for division of focal plane sensors using gaussian processes. *Optics express*, 22(12): 15277–15291, 2014.

[79] S.U. Mageswari, M. Sridevi, and C. Mala. An experimental study and analysis of different image segmentation techniques. *Procedia engineering*, 64:36–45, 2013.

[80] M. Moctezuma and F.F. Parmiggiani. Adaptive stochastic minimization for measuring marine oil spill extent in synthetic aperture radar images. *Journal of Applied Remote Sensing*, 8(1):083553, 2014.

[81] M.H. Narkhede. A review on graph based segmentation. *Journal of innovative research in electrical, electronics, instrumentation and control engineering*, 1(7):3, 2013.

[82] T.F. Kanaa, E. Tonye, G. Mercier, V.d.P. Onana, J.M. Ngono, P. Frison, J. Rudant, and R. Garello. Detection of oil slick signatures in sar images by fusion of hysteresis thresholding responses. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*, volume 4, pages 2750–2752. IEEE, 2003.

[83] G. Sun, H. Ma, D. Zhao, F. Zhang, L. Jia, and J. Sun. Oil spill image segmentation based on fuzzy c-means algorithm. In *2015 International Conference on Computer Science and Intelligent Communication*. Atlantis Press, 2015.

[84] K.A.S. Al Abri, N. Poojary, and J. Menezes. A novel segmentation technique for clustering oil spill data from hyperspectral images. 2012.

[85] C. Wang, L.Z. Xu, X. Wang, and F.C. Huang. A multi-scale segmentation method of oil spills in SAR images based on jseg and spectral clustering. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 7(1):425–432, 2014.

[86] Z. Chen, C. Wang, X. Teng, L. Cao, and J. Li. Oil spill detection based on a superpixel segmentation method for SAR image. In *Proceedings of 2014 IEEE Geoscience and Remote Sensing Symposium*, pages 1725–1728. IEEE, 2014.

[87] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G.A. Licciardi, R. Restaino, and L. Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, 2014.

[88] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu. Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: practical discussion and challenges. *Information Fusion*, 46:102–113, 2019.

[89] L. Loncan, S. Fabre, L.B. Almeida, J.M. Bioucas-Dias, W. Liao, X. Briottet, G.A. Licciardi, J. Chanussot, M. Simões, N. Dobigeon, et al. Hyperspectral pansharpening: a review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):27–46, 2015.

[90] C.A. Laben and B.V. Brower. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, January 4 2000. US Patent 6,011,875.

[91] T. Maurer. How to pan-sharpen images using the gram-schmidt pan-sharpen method-a recipe. *International archives of the photogrammetry, remote sensing and spatial information sciences*, 1:W1, 2013.

[92] H. Li, L. Jing, and Y. Tang. Assessment of pansharpening methods applied to worldview-2 imagery fusion. *Sensors*, 17(1):89, 2017.

[93] T.M. Tu, S.C. Su, H.C. Shyu, and P.S. Huang. A new look at ihs-like image fusion methods. *Information fusion*, 2(3):177–186, 2001.

[94] J. Liu. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18): 3461–3472, 2000.

[95] G. Palubinskas. Fast, simple, and good pan-sharpening method. *Journal of Applied Remote Sensing*, 7(1):073526, 2013.

[96] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.

[97] M.M. Khan, L. Alparone, and J. Chanussot. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE transactions on geoscience and remote sensing*, 47(11):3880–3891, 2009.

[98] J. Choi, H. Park, and D. Seo. Pansharpening using guided filtering to improve the spatial clarity of vhr satellite imagery. *Remote Sensing*, 11(6):633, 2019.

[99] X. Kang, S. Li, and J.A. Benediktsson. Spectral–spatial hyperspectral image classification with edge-preserving filtering. *IEEE transactions on geoscience and remote sensing*, 52(5): 2666–2677, 2013.

[100] G. Moser and J. Zerubia. *Mathematical Models for Remote Sensing Image Processing*. Springer, 2018.

[101] L. Jiang, H. Zhang, and Z. Cai. A novel bayes model: Hidden naive bayes. *IEEE Transactions on knowledge and data engineering*, 21(10):1361–1371, 2008.

[102] M.I. Mahmoud, M.I. Dessouky, S. Deyab, and F.H. Elfouly. Comparison between haar and daubechies wavelet transformions on fpga technology. *World Academy of Science, Engineering and Technology*, 26:68–72, 2007.

[103] Q. Wei, N. Dobigeon, and J.Y. Tourneret. Bayesian fusion of hyperspectral and multispectral images. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3176–3180. IEEE, 2014.

[104] X.R. Wang, A.J. Brown, and B. Upcroft. Applying incremental em to bayesian classifiers in the learning of hyperspectral remote sensing data. In *2005 7th International Conference on Information Fusion*, volume 1, pages 8–pp. IEEE, 2005.

[105] Q. Wei, N. Dobigeon, and J.Y. Tourneret. Fast fusion of multi-band images based on solving a sylvester equation. *IEEE Transactions on Image Processing*, 24(11):4109–4121, 2015.

[106] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez. Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(10):2376–2385, 2005.

[107] M. Simões, J. Bioucas-Dias, L.B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015.

[108] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2011.

[109] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled non-negative matrix factorization (cnmf) for hyperspectral and multispectral data fusion: Application to pasture classification. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 1779–1782. IEEE, 2011.

[110] J.M. Nascimento and J.M. Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.

[111] L. Wald. Quality of high resolution synthesised images: Is there a simple criterion? 2000.

[112] P. Jagalingam and A.V. Hegde. A review of quality metrics for fused image. *Aquatic Procedia*, 4:133–142, 2015.

[113] L. Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.

[114] C. Kwan, B. Budavari, A.C. Bovik, and G. Marchisio. Blind quality assessment of fused worldview-3 images by using the combinations of pansharpening and hypersharpening paradigms. *IEEE Geoscience and Remote Sensing Letters*, 14(10):1835–1839, 2017.

[115] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli. Quality assessment of pansharpening methods and products. *IEEE Geoscience and Remote Sensing Society Newsletter*, 1(161):10–18, 2011.

[116] V. Karathanassi. Spectral unmixing evaluation for oil spill characterization. *Int. J. Remote Sens. Appl*, 4:1–6, 2014.

[117] Y. Lu, Q. Tian, X. Wang, G. Zheng, and X. Li. Determining oil slick thickness using hyperspectral remote sensing in the bohai sea of china. *International Journal of Digital Earth*, 6(1):76–93, 2013.

[118] M. Fingas and C. Brown. Review of oil spill remote sensing. *Marine Pollution Bulletin*, 83(1):9–23, 2014.

[119] N. Dobigeon, J.Y. Tourneret, C. Richard, J.C.M. Bermudez, S. McLaughlin, and A.O. Hero. Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal Processing Magazine*, 31(1):82–94, 2013.

[120] A. Barberis, G. Danese, F. Leporati, A. Plaza, and E. Torti. Real-time implementation of the vertex component analysis algorithm on gpus. *IEEE Geoscience and Remote Sensing Letters*, 10(2):251–255, 2012.

[121] D.C. Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3):529–545, 2001.

[122] N. Dobigeon, Y. Altmann, N. Brun, and S. Moussaoui. Linear and nonlinear unmixing in hyperspectral imaging. In *Data Handling in Science and Technology*, volume 30, pages 185–224. Elsevier, 2016.

[123] C. Li, Y. Ma, J. Huang, X. Mei, C. Liu, and J. Ma. Gbm-based unmixing of hyperspectral data using bound projected optimal gradient method. *IEEE Geoscience and Remote Sensing Letters*, 13(7):952–956, 2016.

[124] Y. Altmann, A. Halimi, N. Dobigeon, and J.Y. Tourneret. Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1009–1012. IEEE, 2011.

[125] M.H. Stone. The generalized weierstrass approximation theorem. *Mathematics Magazine*, 21(5):237–254, 1948.

[126] R. Guida, S.W. Ng, and P. Iervolino. S- and X-band SAR data fusion. In *2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pages 578–581, Sept 2015.

[127] M. Marghany. Automatic detection of oil spills in the Gulf of Mexico from RADARSAT-2 SAR satellite data. *Environmental Earth Sciences*, 74(7):5935–5947, 2015.

[128] S. Singha, T.J. Bellerby, and O. Trieschmann. Detection and classification of oil spill and look-alike spots from sar imagery using an artificial neural network. In *Proceedings of 2012 IEEE International conference on Geoscience and Remote Sensing Symposium (IGARSS)*,, pages 5630–5633. IEEE, 2012.

[129] R. Chandrakanth, J. Saibaba, G. Varadan, P.A.D. Raj, et al. Fusion of high resolution satellite SAR and optical images. In *Proceedings of 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM)*, pages 1–6. IEEE, 2011.

[130] V.R. Pandit and R. Bhiwani. Image fusion in remote sensing applications: A review. *International Journal of Computer Applications*, 120(10), 2015.

[131] L. Dabbiru, S. Samiappan, R.A.A. Nobrega, J.A. Aanstoos, N.H. Younan, and R.J. Moorhead. Fusion of synthetic aperture radar and hyperspectral imagery to detect impacts of oil spill in Gulf of Mexico. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1901–1904, July 2015.

[132] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

[133] M. Subramanyam and Mahesh. Automatic feature based image registration using SIFT algorithm. In *Proceedings of 2012 Third International Conference on Computing Communication & Networking Technologies (ICCCNT)*, pages 1–5. IEEE, 2012.

[134] M. Hassaballah, A.A. Abdelmgeid, and H.A. Alshazly. Image features detection, description and matching. In *Image Feature Detectors and Descriptors*, pages 11–45. Springer, 2016.

[135] D. Mistry and A. Banerjee. Comparison of feature detection and matching approaches: Sift and surf. *GRD Journal*, 2, 2017.

[136] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[137] R. Chen, M. Hawes, L. Mihaylova, J. Xiao, and W. Liu. Vehicle logo recognition by Spatial-SIFT combined with logistic regression. *In Proceedings of the International Conference on Information Fusion 2016*, 2016.

[138] T. Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270, 1994.

[139] J. Babaud, A.P. Witkin, M. Baudin, and R.O. Duda. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1): 26–33, 1986.

[140] D.G. Lowe et al. Object recognition from local scale-invariant features. In *iccv*, volume 99, pages 1150–1157, 1999.

[141] M. Brown and D.G. Lowe. Invariant features from interest point groups. In *BMVC*, volume 4, 2002.

[142] M. Subramanyam and Mahesh. Automatic feature based image registration using SIFT algorithm. In *Proceedings of 2012 Third International Conference on Computing Communication & Networking Technologies (ICCCNT)*, pages 1–5. IEEE, 2012.

[143] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[144] L.R. Dung, C.M. Huang, and Y.Y. Wu. Implementation of RANSAC algorithm for feature-based image registration. *Journal of Computer and Communications*, 1(06):46, 2013.

[145] K.G. Derpanis. Overview of the ransac algorithm. *Image Rochester NY*, 4(1):2–3, 2010.

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

[147] L. Lopezl, M. Moctezuma, and F. Parmiggianil. Contextual approach for oil spill detection in sar images using image fusion and markov random fields. In *2006 49th IEEE International Midwest Symposium on Circuits and Systems*, volume 2, pages 137–139. IEEE, 2006.

[148] R. Gonzalez. Digital image processing using matlab-gonzalez woods & eddins. pdf. education. 2004.

[149] J. Ma, M. Gong, and Z. Zhou. Wavelet fusion on ratio images for change detection in sar images. *IEEE Geoscience and Remote Sensing Letters*, 9(6):1122–1126, 2012.

[150] Z. Tang, J. Wang, and S. Huang. Wavelet transform application for image fusion. In *Wavelet Applications VII*, volume 4056, pages 462–470. International Society for Optics and Photonics, 2000.

[151] V. Naidu and J. Raol. Pixel-level image fusion using wavelets and principal component analysis. *Defence Science Journal*, 58(3):338, 2008.

[152] A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21, 2014.

[153] S. Biswas, D. Ghoshal, and R. Hazra. A new algorithm of image segmentation using curve fitting based higher order polynomial smoothing. *Optik*, 127(20):8916–8925, 2016.

[154] C.G. Harris, M. Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[155] Ş. Öztürk and B. Akdemir. Comparison of edge detection algorithms for texture analysis on glass production. *Procedia-Social and Behavioral Sciences*, 195:2675–2682, 2015.

[156] J.W. Griggs. Bp gulf of mexico oil spill. *Energy LJ*, 32:57, 2011.

[157] D.H.S. Group. Final report on the investigation of the Macondo well blowout. *Center for Catastrophic Risk Management, University of California at Berkeley*, 2011.

[158] P. Trivero and W. Biamino. Observing marine pollution with synthetic aperture radar. *Geoscience and remote sensing new achievements*, pages 397–418, 2010.

[159] C.M. Christoudias, R. Urtasun, M. Salzmann, and T. Darrell. Learning to recognize objects from unseen modalities. In *European Conference on Computer Vision*, pages 677–691. Springer, 2010.

[160] A. Melkumyan and F. Ramos. Multi-kernel gaussian processes. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 1408, 2011.

[161] A. Melkumyan and F. Ramos. A sparse covariance function for exact Gaussian process inference in large datasets. In *IJCAI*, volume 9, pages 1936–1942, 2009.

[162] M.F. Yakhdani and A. Azizi. *Quality assessment of image fusion techniques for multisensor high resolution satellite images (case study: IRS-P5 and IRS-P6 satellite images)*. na, 2010.

[163] N. Dhanachandra, K. Manglem, and Y.J. Chanu. Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54: 764–771, 2015.

[164] S. Vasudevan. Data fusion with Gaussian processes. *Robotics and Autonomous Systems*, 60(12):1528–1544, 2012.

[165] L. Yang, K. Wang, and L.S. Mihaylova. Online sparse multi-output gaussian process regression and learning. *IEEE Transactions on Signal and Information Processing over Networks*, 2018.

[166] M. Haghighat and M.A. Razian. Fast-FMI: non-reference image fusion metric. In *Application of Information and Communication Technologies (AICT), 2014 IEEE 8th International Conference on*, pages 1–3. IEEE, 2014.

[167] J. Hensman, N. Fusi, and N.D. Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

Appendices

Appendix A

Abundance Estimates of Endmembers using the Unmixing Methods

The abundance estimates from evaluating the linear and non-linear spectral unmixing methods evaluated on the Reference Hyperspectral image and the Bayesian Sparse method are presented here for illustrative purpose. The abundance estimate is for three endmembers in the scene of interest, namely water, oil and cloud. The presented abundances are evalauted at an SNR of 50dB.

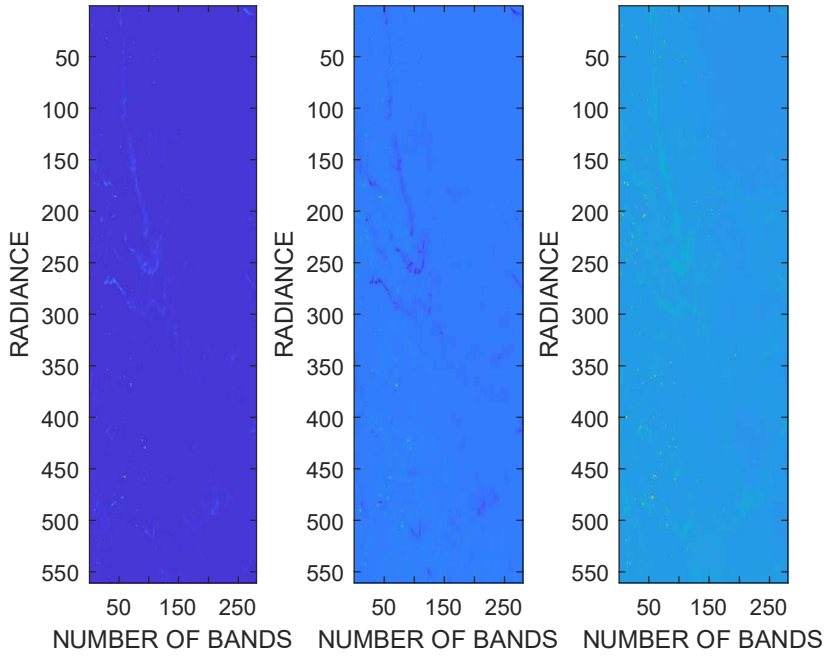


Figure A.1: Abundance Estimate of Endmembers using VCA Method

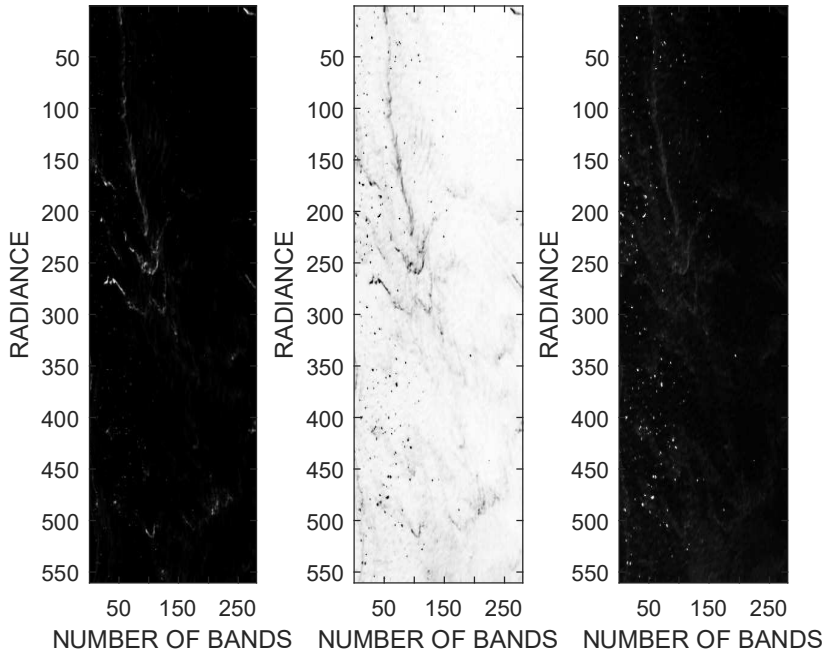


Figure A.2: Abundance Estimate of Endmembers using FCLS Method

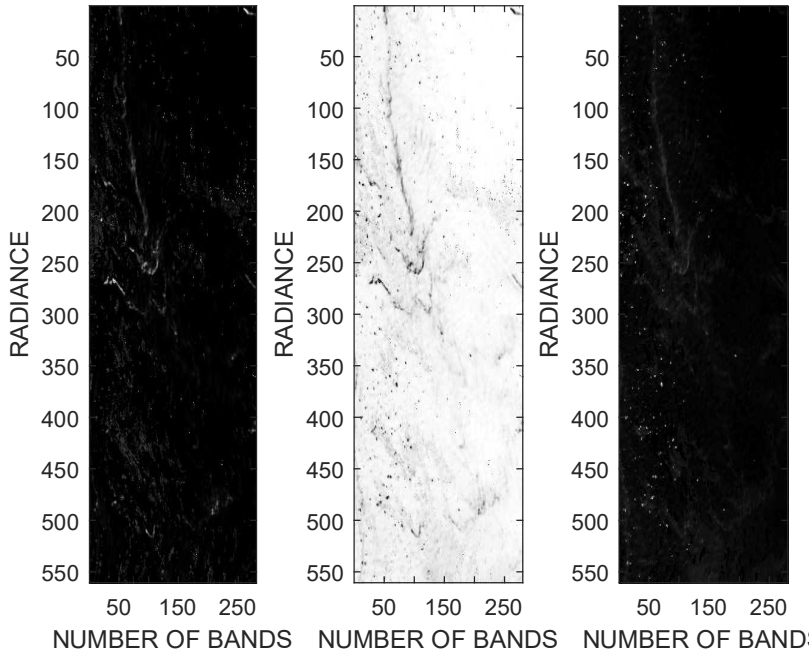


Figure A.3: Abundance Estimate of Endmembers using Bilinear Method

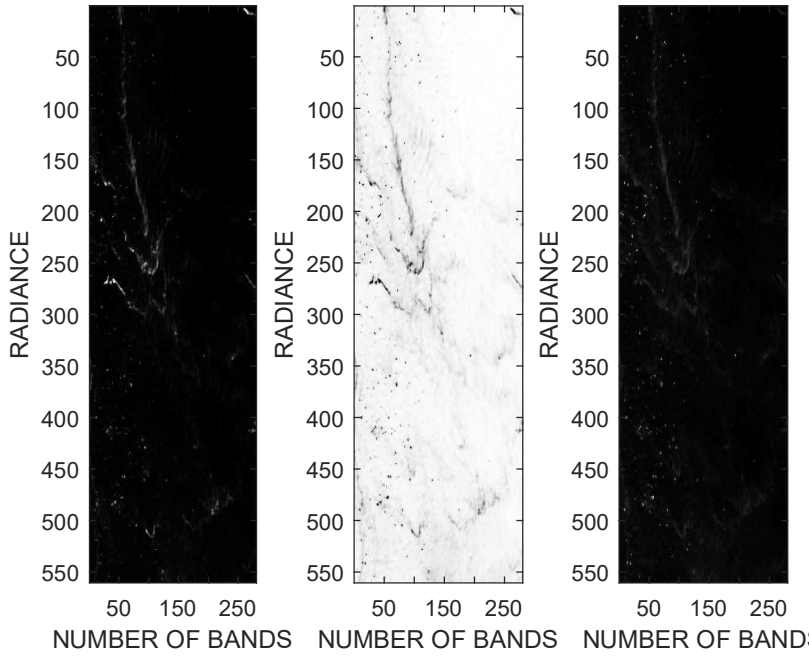


Figure A.4: Abundance Estimate of Endmembers using PPNM Method

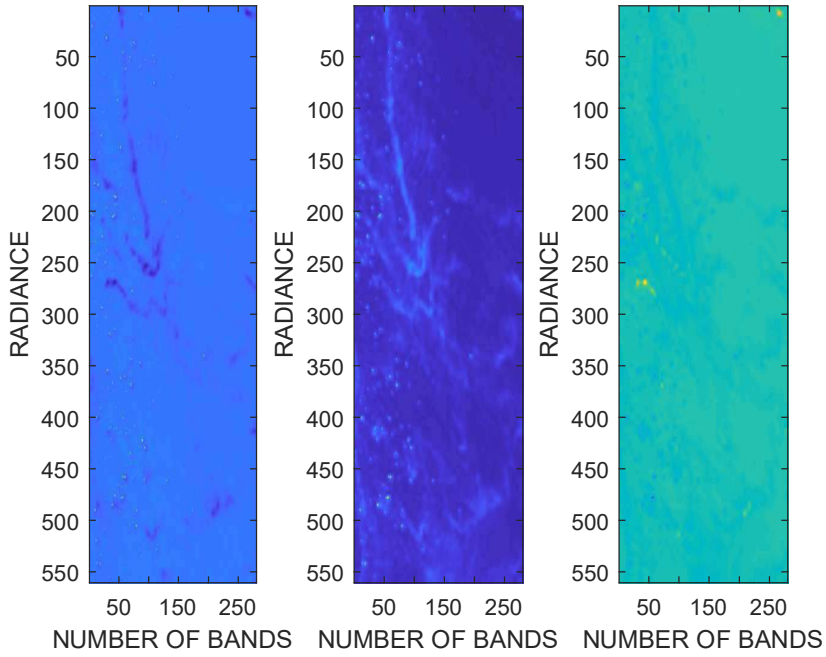


Figure A.5: Abundance Estimate of Endmembers using VCA Method

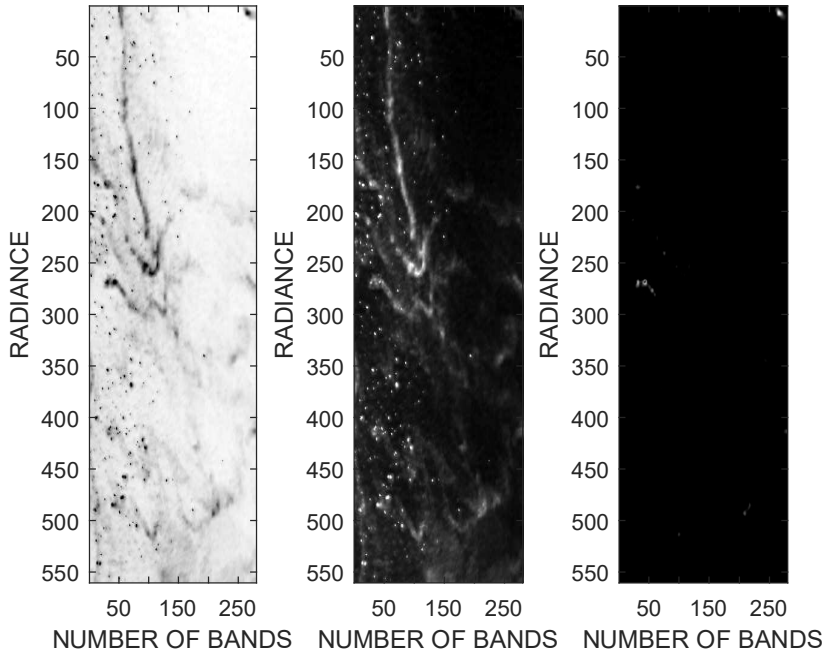


Figure A.6: Abundance Estimate of Endmembers using FCLS Method

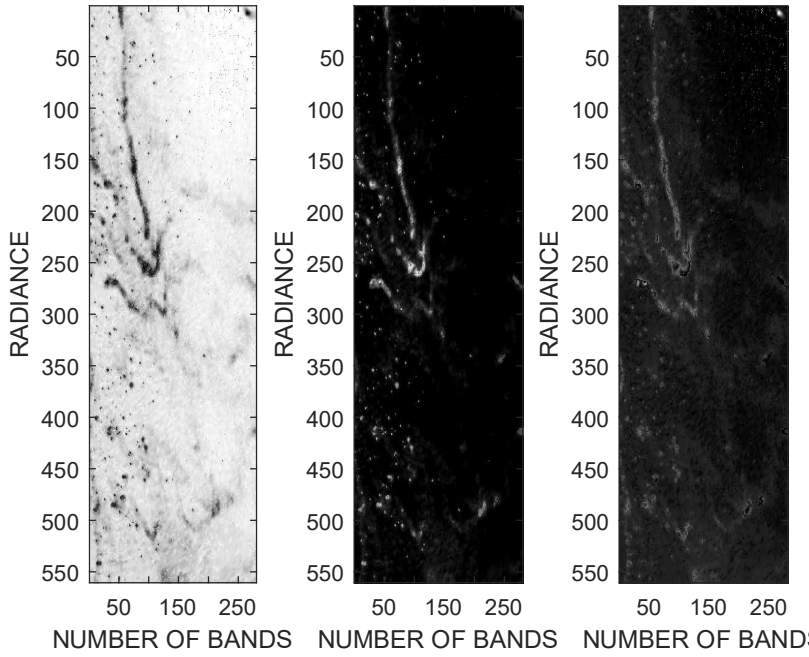


Figure A.7: Abundance Estimate of Endmembers using Bilinear Method

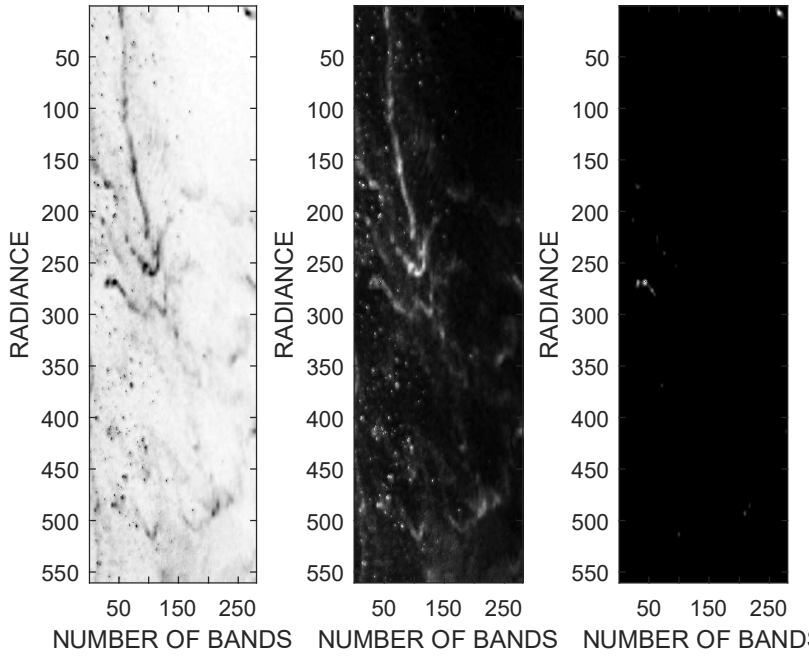


Figure A.8: Abundance Estimate of Endmembers using PPNM Method

Appendix B

Defining Covariance over areas rather than points in GP

Firstly an image pixel is modelled as an output of observing the output function $f(x)$ over the high resolution area $\mathbf{H_A}$. Next, a simple average relationship is assumed to relate an image pixel to the function such that

$$\mathbf{P}(\mathbf{H_A}) = \frac{1}{|H_A|} \iint_{\mathbf{x} \in H_A} f(x) dx \ , \tag{B.1}$$

where $|H_A|$ is the surface area of $\mathbf{H_A}$ and the function $f(x)$ is modelled as a GP as shown in (5.34) where $k(\mathbf{x}, \mathbf{x}')$ describe the covariance function between points $\mathbf{x}$ and $\mathbf{x}'$. substituting $\mathbf{x}'$ from the observation model and replacing it with an area $|A_H|'$, the observation model can be written as:

$$k(x, H_A) = \frac{1}{|H'_A|} \iint_{x' \in H'_A} k(x, x') dx \tag{B.2}$$

This is possible because $k(x, x')$ is flexible to be added or looped into multiple forms [72] [69]. With this, (B.2) can be repeated for all points over areas to get:

$$k(\mathbf{H_A}, \mathbf{H'_A}) = \frac{1}{|H_A||H'_A|} \iint_{x \in H_A} \iint_{x' \in H'_A} k(x, x') dx dx' \tag{B.3}$$

This is valid since covariance functions of a GP can summed or multiplied into compounded forms as shown in chapter 2 of this thesis.