6-28-2010

Algorithms for spectral and spatio-spectral feature selection and classification for tunable sensors : theory and application

Biliana Paskaleva

Algorithms for Spectral and Spatio-Spectral Feature Selection and Classification for Tunable Sensors: Theory and Application

by

Biliana Stefanova Paskaleva

B.S., Electrical Engineering, Technical University of Varna, 1992
M.S., Electrical Engineering, University of New Mexico, 2004

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Engineering

The University of New Mexico

Albuquerque, New Mexico

May, 2010

Dedication

To my parents.

Acknowledgments

This dissertation would not have been possible without the constant support and encouragement of my advisor, Professor Majeed M. Hayat, whose exceptional dedication and professionalism set the standards for me. Thank you for giving me this tremendous opportunity and for believing in me!

A debt of gratitude is owed to my committee members Professor T. Caudell, Professor S. Krishna, Professor S. Prasad, and Professor B. Santhanam, and the professors at UNM who, throughout the years, taught me the fundamentals needed to complete this work. I also want to acknowledge the faculty and staff at CHTM, and especially Professor S. Krishna, for giving me the opportunity to work with cutting edge technology. My fellow graduate students provided an exciting and stimulating atmosphere and their camaraderie gave me additional strengths to complete this work. Special thanks to Woo-Yong (Eric) Jang, Jorge E. Pezoa and David E. Ramirez for all your help and support!

Dr. Steven Bender from the Los Alamos National Laboratory shared his deep knowledge of multispectral and hyperspectral remote sensing and provided many invaluable insights and interesting ideas for this work. Many sincere thanks Steve!

I am also grateful to Dr. T. Williams and Dr. M. Wood from the University of Hawaii, for providing the AHI hyperspectral data, and Dr. L. Crossey and Dr. V. Atudorei from the Department of Earth and Planetary Sciences at UNM for providing the rock samples.

For the past three years I was fortunate enough to work with an extraordinary team at Sandia National Laboratories who taught me about the value of team work and showed tremendous patience and understanding while I was juggling school, work and family. Thank you Julie Bouchard, Joseph Castro, Robert Fogler, Charles Hembree, Rob Hoekstra, Scott Hutchinson, Kenneth Kambour, Ellen Lemen, Mary Moya, and John Shadid.

There were many times when this moment seemed distant and unreachable, but

my family and especially my husband Pavel never stopped believing in me. Pavel taught me that persistence and hard work always pay off and I am grateful for having him by my side.

I would like to acknowledge the support of the National Science Foundation under awards IIS-0434102 and ECS-401154. Additional support was provided by the National Consortium for MASINT Research through a Partnership Project led by the Los Alamos National Laboratory.

Algorithms for Spectral and Spatio-Spectral Feature Selection and Classification for Tunable Sensors: Theory and Application

by

Biliana Stefanova Paskaleva

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Engineering

The University of New Mexico

Albuquerque, New Mexico

May, 2010

Algorithms for Spectral and Spatio-Spectral Feature Selection and Classification for Tunable Sensors: Theory and Application

by

Biliana Stefanova Paskaleva

B.S., Electrical Engineering, Technical University of Varna, 1992
M.S., Electrical Engineering, University of New Mexico, 2004

Ph.D, Electrical Engineering, University of New Mexico, 2010

Abstract

The main thrust of this dissertation is to develop, prototype and test new customized algorithms for infrared spectral imaging and remote sensing of terrestrial features and objects, with particular focus on a general class of sensors with noisy and overlapping spectral bands. While the principal driver of this dissertation is the bias-tunable quantum-dot-in-a-well (DWELL) sensor, developed at the Center for High Technology Materials (CHTM), the scope of the proposed algorithms is broad and extends to traditional sensors with fixed bands. The algorithms formulated in this work also represent a step towards enabling future compressive sensing capabilities based on bias tunable detectors such as the DWELL imager.

The key contributions of this dissertation span three distinct types of algorithms for multispectral (MS) and hyperspectral (HS) imagery. The first algorithm is the *Canonical Correlation Feature Selection* (CCFS), which utilizes the spectral content of the data to form a weighted linear superposition of the bias-tunable DWELL bands in order to achieve algorithmic spectral matching in the presence of noise for the purpose of feature selection and classification. Mathematically, the CCFS can be interpreted as a customized projection algorithm, which minimizes a generalized distance between the spectrum of an object and the linear space spanned by the filters bands, while maximizing the signal-to-noise ratio.

The second algorithm extends the CCFS to a *spatio-spectral feature selection framework*. In this framework the spatial content of the HS imagery is used to enhance the canonical spectral features exposed by the CCFS algorithm by applying a sequence of spatial masks to the canonical features. The resulting stacked hyperpixels combine spatial and spectral features that have been obtained independently from each other. Consequently, the extension of the CCFS rests on the assumption of spatial-spectral separability of the hyperspectral image, i.e., that the spatial content of a hyperspectral image is independent of the spectral bands, but can be used to enhance the latter.

The third algorithm is a *joint spatio-spectral algorithm* for MS and HS image segmentation, which integrates the spectral information into the spatial feature extraction process in order to achieve *simultaneous* spatio-spectral feature selection. Unlike the second algorithm, this algorithm does not assume spatial-spectral separability of the hyperspectral image. Instead, it utilizes the concept of *spectral ratio contrast* to define edge signatures using bands that maximize the spectral contrast between any two materials based on their spatio-spectral signatures. The edge signatures are then fused with a spatial mask to obtain a three-dimensional spatio-spectral mask in which the third (spectral) dimension of each pixel can be independently chosen.

This dissertation also provides a thorough and systematic validation and testing of the three algorithms using laboratory data and real infrared imagery from the DWELL Focal Plane Array (FPA) and the Airborne Hyperspectral Imager (AHI). Our results include supervised classification and spectral unmixing and abundance estimation using AHI hyperspectral imagery, rock classification study using laboratory and DWELL FPA imagery, and HS and MS edge detection using the AHI and the DWELL FPA imagery.

Contents

List of Figures

List of Tables

Glossary

AHI	Airborne Hyperspectral Imager
CCFS	Canonical Correlation Feature Selection
CCSS-FS	Canonical Correlation Spatio-Spectral Feature Selection
CHTM	Center for High Technology Materials
DWELL	Quantum-dot-in-a-well
FPA	Focal plane array
HS	Hyperspectral
LWIR	Long-wave infrared
MCG	Multi-Color Gradient
MS	Multispectral
MWIR	Mid-wave infrared
QDIP	Quantum-dot infrared photodetector
SNR	Signal-to-noise ratio
SRC	Spectral Ratio Contrast

SWIR	Short-wave infrared
UNM	University of New Mexico

Chapter 1

Introduction

1.1 Motivation

Spectral imaging for remote sensing of terrestrial features and objects has emerged as a supplement to high-spatial-resolution, large-aperture satellite imaging systems. Subsequently, the significant interest and advance in the infrared (IR) sensing technologies prompted the development of sophisticated short-wave (SW), mid-wave (MW) and long-wave (LW) IR multispectral (MS) and hyperspectral (HS) sensing systems. By using tens or even hundreds of spectral bands operating in the 0.4–18 μm range, these systems offer highly resolved spectral imaging. One example is the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), which captures high spatial resolution data in 14 bands, from the visible to the thermal IR portion of the spectrum. Another example is the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [5], which contains 224 contiguous spectral channels (bands) with wavelengths from 0.4 to 2.5 μm, and is focused on identifying, measuring, and monitoring the constituents of the Earth's surface and atmosphere. The Airborne Hyperspectral Imager (AHI) [6] and the Spatially Enhanced Broadband

Array Spectrograph System (SEBASS) [7] are two other examples of HS sensors that operate in the LWIR portion of the spectrum and contain, respectively, 256 and 128 narrowband channels.

A typical infrared MS or HS system relies on either deploying multiple detectors, each sensing at a specific range of wavelengths, or a single broadband detector combined with a bank of IR optical filters, each tuned to a specific wavelength. In either case, the sensor represents a highly complex opto-mechanical instrument that requires precision alignment and calibration. Once the calibration is complete and the sensor is deployed, the sensor functionality cannot be easily modified. As a result, the sensor cannot be easily adapted to take advantage of a specific sensing situation and one is typically forced to acquire all available imagery data before its relevance can be determined. This leads to the acquisition of maximum and often massive amounts of data that has to be stored for subsequent processing in applications such as classification, abundance estimation, image segmentation and analysis, etc. Besides the large storage demands, the analysis of this MS and HS imagery requires powerful hardware systems and efficient processing algorithms.

The quantum-dot-in-a-well (DWELL) IR photodetector [8, 9, 10], developed and fabricated in the Center for Hight Technology Materials (CHTM) at the University of New Mexico (UNM), is a new emerging technology that has the potential to provide an unprecedented flexibility in the sensing process [8, 11] through continuous spectral tunability. The asymmetric DWELL bandstructure leads to bias-dependent spectral response, which is attributable to the quantum-confined Stark effect. As a result, the photocurrent produced by the application of each bias voltage can be thought of as an output of a distinct band. This means that in the context of MS and HS sensing, a single DWELL detector can be utilized as a MS IR sensor; the photocurrents measured at different operational biases can be viewed as outputs of different spectrally broad and overlapping bands [12]. The DWELL detectors are

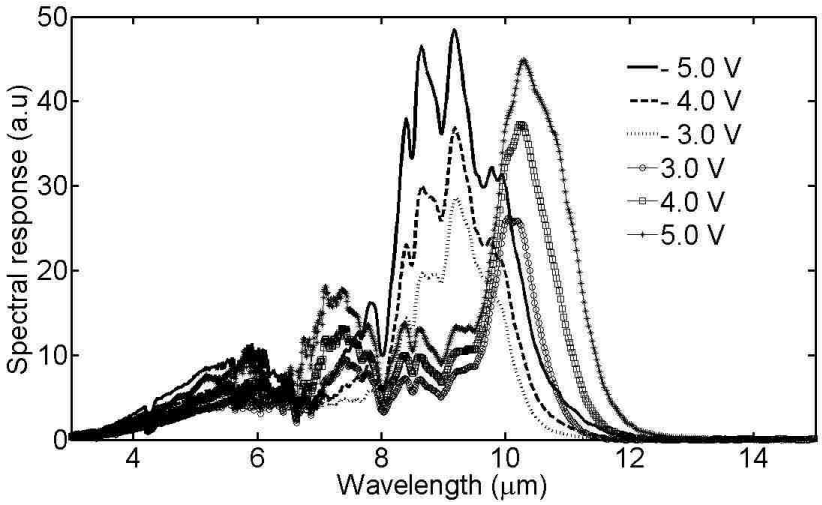


Figure 1.1: Spectral response of a DWELL detector as a function of the applied bias voltage.

based on a mature GaAs growing technologies and have desirable features such as intrinsic sensitivity to normal IR incidence light and low dark currents [8, 11].

Such on-demand electronically controlled tunability, that can be optimized for specific applications, has the potential to greatly simplify the opto-mechanical complexity of the MS and HS sensing systems. Furthermore, this capability is an excellent fit to *compressive sensing* once the sensor is combined with algorithms and reconfigurable readout integrated circuits (ROICs). In particular, compressive sensing can greatly reduce the massive amounts of data that must be acquired by traditional MS sensors by utilizing only the bands that are relevant to the scene.

The flexibility afforded by on-demand electronically controlled tunability is not without a price, however. For instance, as Fig. 1.1 shows, the DWELL's spectral response is relatively broad ($\approx 1-2\mu$m). As a result, the spectral bands corresponding to different bias voltages overlap significantly, making the photocurrents generated by the DWELL bands highly correlated. Another complication is bias-dependance of

the noise (dark current) present in the photocurrents. As a result, efficient utilization of the DWELL requires the development of specialized post-processing algorithms and methods that take into consideration these attributes. Robustness with respect to bias- or band-dependent noise is one of the traits required of these algorithms. Correlated data is another issue that has to be addressed at the post-processing stage. The algorithms should be also able to exploit the on-demand controlled tunability of the DWELL to improve the overall efficacy of the sensing process. However, at present there are very few algorithms dealing with these issues and the literature devoted to this subject is rather sparse. Some recent results have been reported in [12, 13, 14, 15] and the references cited therein, but besides these sources not much work has been done to address data processing for sensors with noisy and overlapping bands. This dissertation aims to contribute towards the development of this much needed capability.

1.2 Contributions of this dissertation

The main thrust of this dissertation is to develop, prototype and test new customized algorithms for a general class of sensors with noisy and overlapping bands. While our principal driver is the DWELL sensor, the scope of the algorithms proposed in this dissertation is significantly broader and extends to traditional sensors with fixed bands. These algorithms also represent a step towards future compressive sensing capability based on bias tunable detectors such as DWELL.

Algorithms for MS and HS image analysis can be broadly divided into three categories based on how they utilize the acquired imagery data. The first category includes methods which exploit only the spectral content of the MS and HS data. The most well-known among these methods are the Principal Component Analysis (PCA), Independent Component Analysis (ICA), Projection Pursuit (PP) [16, 17],

Matching Pursuits [18, 19], and the Maximal Noise Fraction (MNF) transform [20], to name just few. The second category comprises of methods that utilize both the spectral and the spatial information, but in a disjoint manner; see for example [21, 22, 23, 24] for representative work. Finally, in the third category we find methods that are characterized by their joint use of spectral and spatial information in a way that exploits the intrinsic correlation between the spectral and the spatial content of HS and MS imagery. The multi-color gradient (MCG) edge detector [25, 1, 26], the morphological color gradient (MoCG) [27], and the total variation methods for restoration of vector valued images [28, 29, 30] are examples of methods that belong to this category.

The key contributions of this dissertation are the development and the verification of three algorithms that target sensors with overlapping bands, such as the DWELL. These algorithms fall into each one of the three categories described above. Below we present a concise summary of the main contributions.

The first contribution of this work is the *Canonical Correlation Feature Selection* (CCFS) algorithm [12, 14]. The CCFS utilizes the idea of forming a weighted linear superposition of the bias-dependent photocurrents; in this way it resembles the DWELL-based algorithmic spectrometer (DAS) formulated and examined in [31, 32, 33, 15], which achieves continuous DWELL spectral tuning for optimal spectral reconstruction. With the CCFS we pursue a different objective. Instead of determining sets of weights to be used for spectral reconstruction, the training step of the CCFS algorithm determines sets of weights that are optimally suited for a specific classification problem. More precisely, for a given spectrum, representing a class of objects of interest, and in the presence of noise, the CCFS seeks a set of weights that are used to form an optimal superposition of the DWELL's bias-dependent bands. Such *superposition band* can be thought of as the most "informative generalized direction" in the DWELL spectral space for the given spectrum in the presence

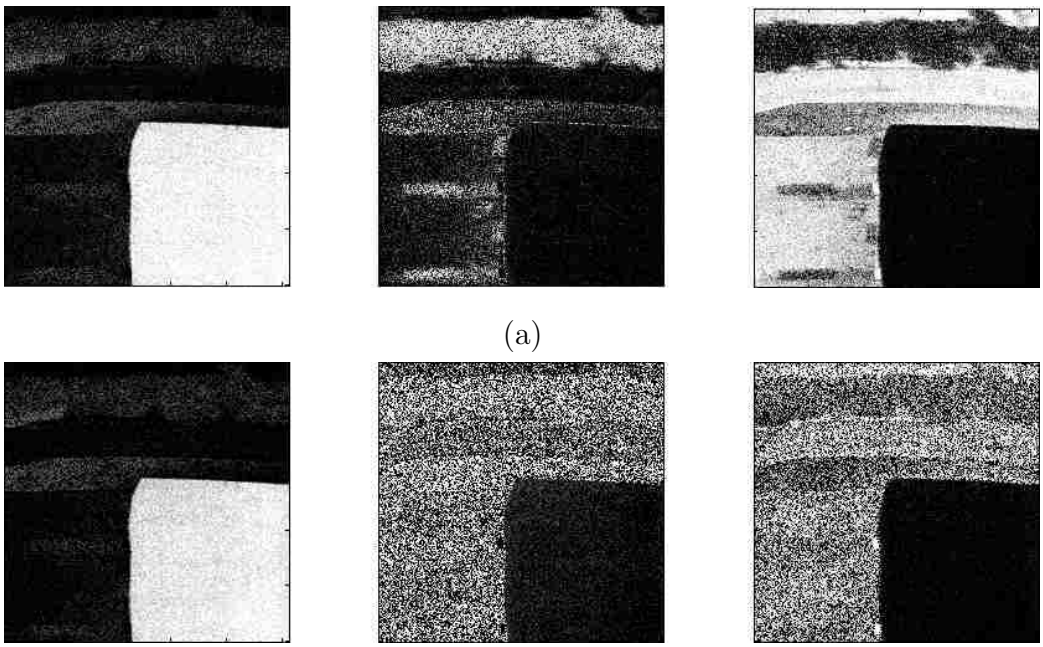


Figure 1.2: Left to right: abundance estimation maps for building, vegetation and road endmembers, and for an SNR level of 20dB: (a) using three superposition features selected by the CCFS algorithm from a subset of 50 bands in the range 7.7 μm to 8.6 μm; (b) using three features selected by the noise-adjusted PP from a subset of 50 bands in the range 7.7 μm to 8.6 μm.

of noise. Thus, the CCFS algorithm accomplishes algorithmic spectral matching in the presence of noise for the purpose of classification.

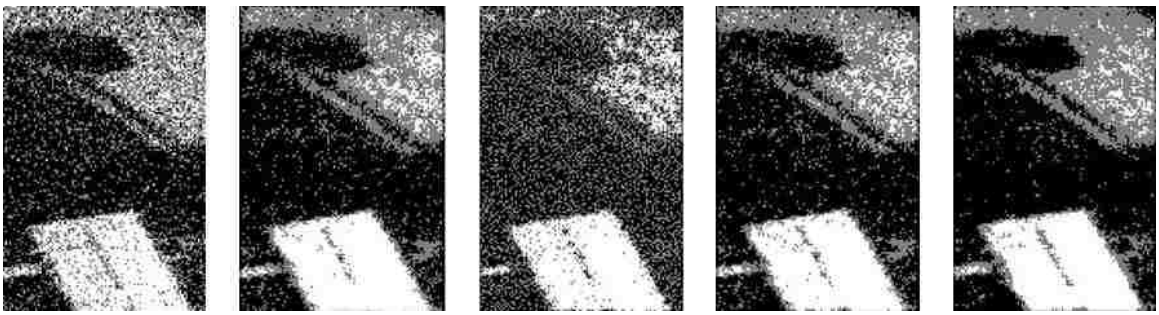
Figure 1.2 (a–b) compares the performance of the CCFS algorithm with the noise-adjusted PP algorithm in the presence of noise. The figure shows three groups of fractional abundance maps, one for building, vegetation and road endmembers, from left to right respectively, and for a signal-to-noise ratio (SNR) level of 20dB. The CCFS (a) and the noise-adjusted PP (b) are applied to 50 consecutive AHI bands in the range 7.7 to 8.6 μm. The maps show improved performance of the

CCFS compared to the noise-adjusted PP. Specifically, PP was not able to clearly discriminate between the endmembers of vegetation and road in this SNR case.

Rigorous derivation and a precise formulation of the CCFS algorithm is presented in Chapter 2 where we also provide interpretation of the optimal feature-selection criterion in terms of SNR.

Our second contribution is extension of CCFS to a *spatio-spectral* feature selection framework. The objective of this work is to explore integration of spatial and spectral information in a way that enhances the canonical spectral features exposed by the CCFS, using the spatial content of the hyperspectral imagery. To this end we assume *separability* of the hyperspectral image, i.e, that the spatial content of a hyperspectral image is independent of the spectral bands, but can be used to enhance the latter. The extension of the CCFS algorithm utilizes the concept of *spectrally enhanced spatial features* [21, 22], which are obtained by considering the pixels simultaneously in the spectral domain, using their spectral content, and in the spatial domain by applying a spatial feature extraction approach. The resulting Canonical Correlation Spatio-Spectral Feature Selection (CCSS-FS) agorithm consists of two distinct stages: a spatially independent spectral feature selection, based on the CCFS, followed by spatially enhanced classification.

Figure 1.3 compares thematic maps of AHI imagery created from the classification results based on different subsets of spectrally enhanced spatial features. For an SNR of 10dB, these results show a noticeable improvement in the classification accuracy when three canonical superposition features, selected with respect to three classes of interest (road, ground and buildings), are extended with 'Mean', 'Median' and 'Gaussian' spatial features, as shown in Fig. 1.3 (i) and (v). The investigation of different combinations of spectrally enhanced spatial features has demonstrated that in this problem, the combination presented in Fig. 1.3 (v) gives consistently, for different SNR values, the best classification results. Chapter 4 provides a detailed



(i) (ii) (iii) (iv) (v)

Figure 1.3: Performance of the CCSS-FS algorithm for different combinations of spectral and spatial features and an SNR of 10dB. Left to right: thematic maps for (i) three CCFS features; (ii) three CCFS features extended with three 'Mean' spatial features; (iii) three CCFS features extended with three 'Laplacian' spatial features; (iv) case (ii) extended with three 'Gaussian' spatial features; and (v) case (iv) extended with three 'Median' spatial features.

description of the CCSS-FS algorithm and results from a validation study using AHI HS imagery.

Our third contribution is a *joint spatio-spectral* algorithm for MS and HS image segmentation. Image segmentation is one of the most difficult and important tasks in image processing [2, p.567] required for computerized analysis of digital images. For MS and HS images this task is further complicated by the presence of features such as iso-luminant edges, [2, p.336] and [34]. Such attributes manifest the fact the spectral and the spatial contents of an MS or HS image cannot in general be treated independently from each other. In other words, an MS or HS image is in general non-separable in the spectral and the spatial domain. Sufficient conditions for separability of HS and MS images are investigated in Chapter 4.

Consequently, the objective here is to incorporate the spectral information into the spatial feature extraction process, i.e., to perform a *simultaneous* spatio-spectral

feature selection. The solution to this problem is not as intuitive as in the CCSS-FS algorithm, where we assume separability of spatial and spectral domains, and is much more challenging to achieve. We focus on the development of edge detection algorithms for MS and HS images based on a three-dimensional spatio-spectral mask, in which the third (spectral) dimension of each pixel can be independently chosen. In other words, we propose a mask that does not operate in a single image plane but instead fuses information from multiple planes. The action of the three-dimensional mask is compared schematically with the MCG and with a standard gray-scale gradient operator in Fig. 1.4.

The spectral plane for each element in the three-dimensional mask is selected in a way that enhances the targeted spatial feature. Our approach uses the concept of *spectral ratio contrast* to define an edge signature for an edge between two materials. An edge signature is a combination of spectral ratios calculated using bands that enhance the spectral contrast between the two materials.

The joint spatio-spectral algorithm, termed spectral ratio contrast (SRC) edge detector, also has two stages. The first stage is a training step which identifies the bands that maximize the spectral contrast between two given materials. The second stage is the feature extraction using the three-dimensional mask with bands defined at the training step.

Figure 1.5 compares the edge maps derived using the SRC algorithm with the edge maps obtained by the MCG approach and by the application of two well-known gray-scale edge detectors to individual DWELL FPA bands. Figure 1.5 (ii) clearly shows the ability of the SRC to capture the "weak", almost iso-luminant edge between the granite and limestone classes in this scene, which is missed by the other approaches.

Spectral ratios and cross-spectral ratios have been previously used for quantitative vegetation monitoring. Examples include the Normalized Difference Vegetation

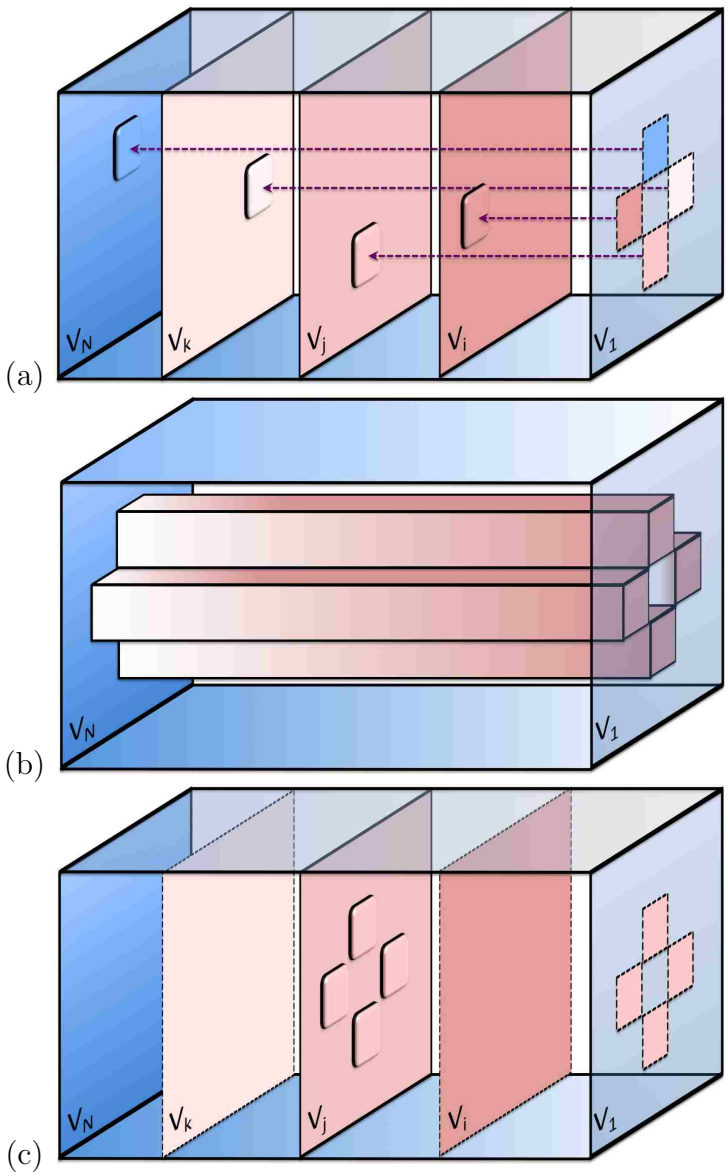
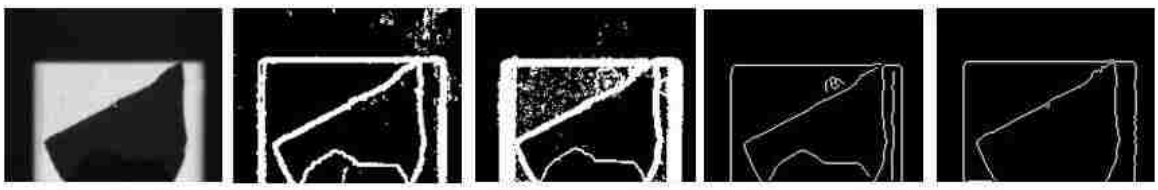


Figure 1.4: Joint spatio-spectral feature extraction using cross-band ratios (a) vs. multi-color gradient [1] approach (b) and standard [2] intensity-based gradient (c) .

Index (NDVI) [35], the Soil-Adjusted Vegetation Index (SAVI) [36] and the Atmospherically Resistant Vegetation Index (ARVI) [37]. Other applications of spectral and cross-spectral ratios include regional seismic discrimination [38, 39, 40, 41] and deblurring of noisy multichannel images [30], among others. However, to the best



(i) (ii) (iii) (iv) (v)

Figure 1.5: Comparison between edge maps obtained by the SRC and three benchmark edge detectors. Left to right: (i) DWELL FPA image comprising of phyllite, granite and limestone at bias 0.7 V; (ii) the SRC edge map; (iii) the MCG edge map; (iv) the Canny [3] edge map at 0.5 V; (v) the Sobel [4] edge map at 0.5 V.

of our knowledge, the application of spectral ratio contrast to edge detection in this dissertation is new and previously unexplored research direction.

Another contribution of this dissertation is the extensive validation of the three algorithms using laboratory data and real IR imagery from the DWELL Focal Plane Array (FPA) and AHI. In particular, we carried out the following studies. First, we performed supervised classification, spectral unmixing and abundance estimation of hyperspectral imagery obtained from the AHI using the CCFS. CCFS was also applied to rock classification study using laboratory data and DWELL FPA imagery.

Second, we carried out a performance validation of CCSS-FS for target detection and classification applications using remotely sensed imagery collected by the AHI. We also studied the sensitivity of the CCSS-FS algorithm with respect to the initial set of sensor bands and with respect to the number and types of spatial features utilized during the classification stage.

Finally, we performed edge detection for HS and MS imagery using the AHI and the DWELL FPA data. We compared the joint spatio-spectral SRC algorithm with the multi-color gradient edge detection approach [26] and several gray-scale edge

detector such as Canny [3] and Sobel [4], applied to individual image bands.

1.3 Overview of the dissertation

For the convenience of the reader, whenever appropriate, the chapters provide concise review of the relevant background information and related work, as well as brief summaries of our conclusions and findings.

We have organized this dissertation as follows. In the focus of Chapter 2 is the rigorous mathematical development of the CCFS algorithm. We formulate the mathematical model for sensors with noisy and spectrally overlapping bands in Section 2.2 and then use this theory to develop the CCFS algorithm in Section 2.3.

Performance of the CCFS algorithm and comparison with noise-adjusted Projection Pursuit approach [42] is provided in Chapter 3. There we test the CCFS algorithm on three different applications. In Section 3.1 we study separability and classification analysis of rock species using the CCFS and laboratory spectral data obtained with a single-pixel quantum-dot infrared photodetector (QDIP). The second application of the CCFS is for supervised classification, spectral unmixing, and abundance estimation of hyperspectral imagery obtained from AHI. Results from this study are presented in Section 3.2. Our third application also deals with rock-type classification; however, in this case we apply the CCFS to real DWELL FPA imagery acquired at the CHTM. These results are presented in Section 3.3 and demonstrate, for a first time, the MS capability of the DWELL FPA by considering three different scene configurations.

In Chapter 4 we extend the CCFS approach to a spatio-spectral feature-selection and classification framework for hyperspectral imagery. Details of the extension are presented in Section 4.2, while Section 4.4 contains validation results using AHI data

in the context of supervised classification. Specifically, we study the sensitivity of the spatio-spectral feature-selection approach with respect to the initial set of sensor bands and the improvement of the classification accuracy with respect to the number and type of spatial features, utilized during the classification stage. In Section 4.3 of this chapter we briefly discuss separability of the spatial-spectral information in MS and HS imagery.

The joint spatio-spectral SRC algorithm for MS and HS edge detection is formulated and studied in Chapter 5. The details of of our approach, including definition of edge signatures, complexity estimates, and implementation are discussed in Section 5.2. In Section 5.3 we present validation study of the SRC algorithm and comparison with the benchmark MCG edge detector using the DWELL FPA and AHI imagery data. For the convenience of the readers, the benchmark MCG approach [1] is reviewed in Appendix A.

1.4 Publications generated by this dissertation

A list of our publications, related to this dissertation, is as follows:

1. B. S. Paskaleva, W-Y. Jang, M. M. Hayat, Y. D. Sharma, S. C. Bender, and S. Krishna, "Multispectral Classification with Bias-tunable Quantum Dots-in-Well Focal Plane Arrays," IEEE *Trans. Geoscience and Remote Sensing.* Special issue on hyperspectral imaging and signal processing. Submitted.

2. W-Y. Jang, B. S. Paskaleva, M. M. Hayat, S. C. Bender, and S. Krishna, "Multispectral Classification with Bias-tunable Quantum Dots in a Well Focal Plane Array," IEEE *Photonics Society (formerly LEOS) 2009 Annual Meeting: Photodetectors and Imaging*, Antalya, Turkey, Oct. 5–9, 2009.

3. B. S. Paskaleva, M. M. Hayat, and S. Krishna, "A new approach for spatio-spectral feature selection for sensors with spectrally overlapping bands," 2008 IEEE *International Geoscience and Remote Sensing Symposium* (IGARSS 2008), July 6–11, 2008, Boston, MA.

4. B. S. Paskaleva, M. M. Hayat, W-Y. Jang, S. Bender, and S. Krishna, "Application of the canonical-correlation feature-selection algorithm to tunable quantum-dot detectors: material discrimination and gas detection," *2008 International Symposium on Spectral Sensing Research (2008 ISSSR)*, Hoboken, NJ, June 23–27, 2008.

5. W-Y. Jang, B. S. Paskaleva, M. M. Hayat, and S. Krishna, "Spectrally adaptive nanoscale quantum-dot sensors," Wiley Handbook of Science and Technology for Homeland Security, in press.

6. B. S. Paskaleva, M. M. Hayat, Zh. Wang, J. S. Tyo, and S. Krishna, "Canonical correlation feature selection for sensors with overlapping bands: theory and application," IEEE *Trans. Geoscience and Remote Sensing*, vol. 46, pp. 3346–3358, 2008.

7. B. S. Paskaleva, U. Sakoglu, Zh. Wang, M. M. Hayat, J. S. Tyo, and S. Krishna, "Algorithmic tunability of quantum-dot infrared detectors," IEEE *LEOS Newsletter*, vol. 20, No. 5, pp. 33–36, October 2006.

8. B. S. Paskaleva, M. M. Hayat, M. M. Martinez, Zh. Wang, and J. S. Tyo, "Feature selection for spectral sensors with overlapping noisy spectral bands," SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII, *Proceedings of the SPIE*, Vol. 6233, Orlando, Fl, June 2006.

9. B. S. Paskaleva and M. M. Hayat, "Optimized algorithm for spectral band selection for rock-type classification," SPIE Defense and Security Symposium

for Multispectral, Hyperspectral, and Ultraspectral Imagery XI, *Proceedings of the SPIE*, vol. 5806, pp. 131–138, Orlando, Fl, March 2005.

10. B. S. Paskaleva, M. M. Hayat, M. M. Moya, and R. J. Fogler, "Multispectral rock-type separation and calssification," Earth Observing Systems IX. Edited by Barnes, William L.; Butler, James J., *Proceedings of the SPIE*, vol. 5543, pp. 152–163, Denver, Co, 2004.

Chapter 2

Canonical Correlation Feature Selection

The main focus of this chapter is a rigorous development of a novel canonical correlation feature selection (CCFS) algorithm that is particularly well-suited for spectral sensors with overlapping and noisy bands. Our approach combines a generalized canonical correlation analysis framework and a minimum mean-square-error criterion for the selection of feature subspaces. This criterion induces ranking of the best linear combinations of the noisy overlapping bands and, in doing so, guarantees a minimal generalized distance between the centers of classes and their respective reconstructions in the space spanned by sensor bands.

2.1 Background and overview of relevant work

In the past two decades, infrared spectral imaging in the wavelength range of 4–18 μm has found many applications in night vision, battlefield imaging, missile tracking and recognition, mine detection and remote sensing, to name just few. Examples of

spectral imagers operating in the 3–5 μm and 8–12 μm atmospheric windows include the Airborne Hypersectral Imager (AHI) and the Spatially Enhanced Broadband Array Spectrograph System (SEBASS), which contain, respectively, 256 and 128 narrow-band channels. However, the price of offering such sophisticated spectral imaging is enormous due to the complexity of the optical systems that render the detailed spectral information. Efforts have been made to develop two-color and even multi-color focal-plane arrays (FPA) for long-wave (LW) applications [43, 44]; these sensors can be electronically tuned to two or more regions of the spectrum. Clearly, such tunable sensors offer greater optical simplicity as the spectral response is controlled electronically rather than optically. However, most existing multi-color sensors are limited in that the spectral sensitivity can only be electronically switched, but not continuously tuned.

More recently, a new technology has emerged for continuously tunable midwave-infrared (MWIR) and longwave-infrared (LWIR) sensing that utilizes intersubband transition in nanoscale self-assembled systems; these devices are termed quantum-dot infrared photodetectors (QDIPs). QDIP sensors promise a less expensive alternative to the traditional hyperspectral (HS) and multispectral (MS) sensors while offering more tuning flexibility and continuity compared to multi-color sensors [44]. QDIPs are based on a mature GaAs-based processing, they are sensitive to normally incident radiation and have lower dark currents compared to their quantum-well counterparts [8, 11]. Unfortunately, QDIPs have low quantum efficiencies, and much effort is currently underway to enhance that efficiency through increasing the number of QD layers as well as using new supporting structures such as photonic crystals [45, 46].

Additionally, QDIPs with a dot-in-a-well (DWELL) configuration exhibit a bias-dependent spectral response, which is attributable to the quantum Stark effect, whereby the detector's responsivity can be altered in shape by varying the applied

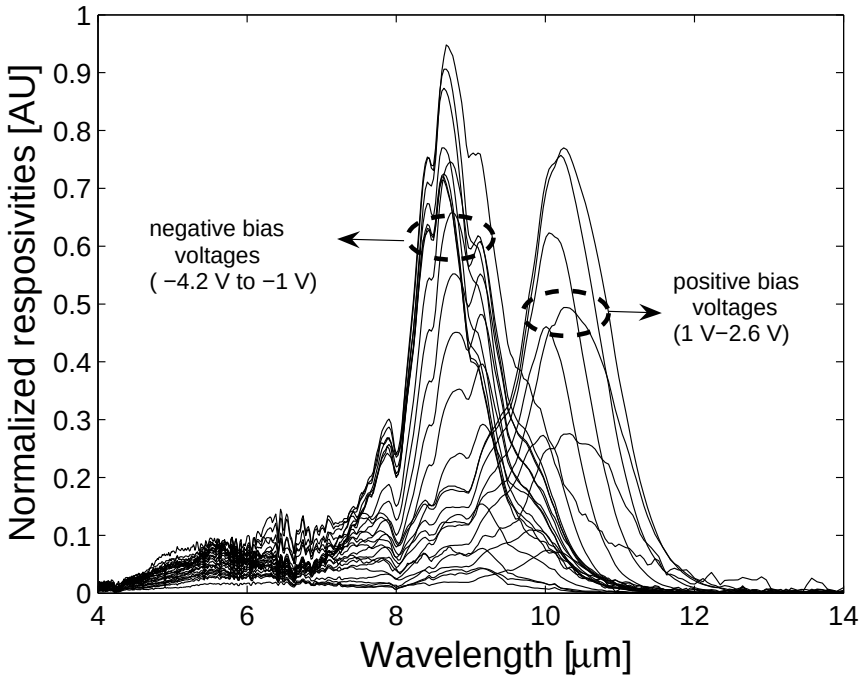


Figure 2.1: Normalized spectral responses of QDIP 1780. The left cluster of spectral responsivities corresponds to the range of negative bias voltages between -4.2 V and -1 V. The right cluster of spectral responsivities corresponds to the range of positive bias voltages between 1 V and 2.6 V.

bias. Figure 2.1 shows the bias-dependant spectral responses of the QDIP device[1] used in this work, measured with a broadband source and a Fourier transform infrared spectrophotometer at a temperature of 30 K. Bias voltages in the range -4.2 V to -1 V and 1 V to 2.6 V, in steps of 0.2 V, were applied to this device. Therefore, a single QDIP can be exploited as a MS infrared sensor; photocurrents of a single QDIP, driven by different operational biases, can be viewed as outputs of different spectrally broad and overlapping bands.

[1]This QDIP was fabricated by Professor Krishna's group at the Center for High Technology Materials at the University of New Mexico.

The inherent and often significant spectral overlap in the bands of a QDIP sensor produces a high level of redundancy in the output photocurrents of these bands. This redundancy, which is not dissimilar to the redundancy present in the outputs of the cones of the human eye, necessitates the development of lower dimensional, uncorrelated representations of the sensed data.

Comprehensive exploration of the bias-dependent tunability of the QDIP and DWELL detectors has led to the development of customized algorithms, which enable more efficient MS sensing. One example is the DWELL-based algorithmic spectrometer (DAS) reported in [31, 32, 33, 15]. DAS achieves continuous DWELL spectral tuning for optimal target spectra reconstruction. The main idea is to probe a target spectra by a DWELL photodetector sequentially in time while changing the applied bias. As a result, a set of bias-dependent photocurrents, each corresponding to a spectral band, is generated. Then, for each tuning wavelength, a projection algorithm is used to produce a set of optimal superposition weights, one per bias. The resulting superposition photocurrent will approximate the value of the spectrum of the target for the specified tuning wavelength. This stage is iterated for each tuning wavelength within the spectral regions of interest.

The presence of noise in the photocurrents, i.e., dark current and Johnson noise, further complicates extraction of reliable spectral information from the highly overlapping and broad spectral bands of QDIP devices. Johnson noise results from the random motion of electrons in resistive elements and occurs regardless of any applied voltage [47]. On the other hand, current resulting from the generation and recombination process within the photoconductor will cause fluctuation in the carrier concentration and, hence, fluctuation in the conductivity of the semiconductor [47]. Generation and recombination noise, or so-called shot noise, becomes important in small band-gap semiconductors, in which the Johnson noise can also be high. Finally, at very low frequencies (e.g., less than 1 KHz) flicker noise, also known as $1/f$ noise,

also becomes an issue; it arises from surface and interface defects and traps in the bulk of the semiconductor. However, for integration times of 1 ms or smaller, this noise is not important. Noise in QDIP detectors is dominated by the Johnson noise at temperatures less than 40 K, and by the shot noise at higher temperatures (e.g., 77 K or above).

It is well known that in the presence of noise the existing feature-reduction techniques may not always yield reliable information compression. It was shown by Green *et al.* [20] that in the Principal Component Analysis (PCA) approach, the variance of the MS/HS data does not always reflect the actual signal-to-noise ratio (SNR), due to unequal noise variances in different spectral bands. Therefore, it is possible that a band with a low variance may have a higher SNR than a band with a high variance. As a result, modified approaches such as the Maximal Noise Fraction (MNF) transform were developed [20] based on maximizing the SNR; this method first whitens the noise covariance and then performs PCA. Other techniques include "higher-order methods" such as the Projection Pursuit (PP) and the Independent Component Analysis (ICA) [16, 48, 49]; these methods search for "interesting" projection directions generating features that maximally deviate from "Gaussianity," or directions that maximize a certain projection index. Following the idea of the MNF transform [20], Lennon and Mercier in [42] proposed to adjust both PP and ICA to the noise in such a way that the SNRs of the noise-adjusted components are significantly increased compared to the SNRs of the components determined by the original algorithms.

In this Chapter we develop the mathematical theory for spectrally adaptive feature-selection approach for a general class of sensors with overlapping and noisy spectral bands. Using this theory we develop the canonical correlation feature selection (CCFS) algorithm. Our work draws upon the geometrical sensing model developed by Wang *et al.* [50, 51], in which the sensing process is viewed as a pro-

jection of the *scene space*, defined as the space of all spectra of interest, onto a space spanned by the sensor bands, termed the *sensor space*.

Similar to the DAS algorithm [33], the CCFS is also based upon the principle of forming a weighted linear superposition of the bias-dependent photocurrents; however, its objective is quite different. Instead of determining sets of weights to be used for target spectral reconstruction, the CCFS algorithm determines sets of weights that are optimally suited for specific detection/classification problems. In particular, for a given spectrum, representing a class of targets of interest, the CCFS seeks a set of weights used to form a superposition of the QDIP's bias-dependent bands. Such superposition band can be thought of as the most "informative generalized direction" for the given target spectrum in the presence of noise. In other words, the CCFS algorithm can accomplish *algorithmic spectral matching for the purpose of target classification*. Moreover, this process of selecting a superposition band is repeated in a hierarchical fashion to yield a canonical set of superposition bands that will generate, in turn, the best set of features for classes of objects.

It is instructive to compare the CCFS with the matching pursuit (MP) algorithm reported in [18, 19]. The main goal of the MP algorithm is to represent efficiently any function $f(t)$ in terms of a family of functions D (called a dictionary), where the family is usually extremely redundant. Depending on the local properties of $f(t)$, the idea is to select adaptively a subset of elements of D that best represents $f(t)$, i.e., MP is a procedure that expands functions over a set of waveforms, selected appropriately among large and redundant dictionary. Therefore, the MP algorithm can be viewed as a version of the greedy algorithm. In contrast, with the CCFS we also aim to find the best representation out of a redundant set, however, the search is not restricted to the subset of the functions in the dictionary (filter space). Roughly speaking, we can think of MP as a band selection procedure (selection of a set of bands from the original set), whereas the CCFS is primarily a feature selection

(extraction) process.

This Chapter is organized as follows. The mathematical theory for feature selection for sensors with noisy and spectrally overlapping bands is developed in Section 2.2. We use this theory in Section 2.3 to formulate the CCFS algorithm and summarize our conclusions in Section 2.4.

2.2 Mathematical model for spectral sensing

2.2.1 Preliminaries

We start by reviewing the relevant concepts in spectral sensing. The spectral characteristics of bands are represented by a finite set of real-valued square-integrable spectral filters, or simply bands, $\{\hat{f}_i(\lambda)\}_{i=1}^k$, where the variable λ represents wavelength. The spectral response of the ith band is given by $\hat{f}_i(\lambda) = R_0 f_i(\lambda)$, where the unit of $\hat{f}_i(\lambda)$ is response per watt of power incident on the detector. The scalar R_0 can be thought of as the peak responsivity, and will assume the units required by $\hat{f}_i(\lambda)$, while the functions $\{f_i(\lambda)\}_{i=1}^k$ will be treated as dimensionless functions. Similarly, the emitted spectra of materials of interest can be described by another set of square-integrable functions of wavelength, $\{\hat{p}_i(\lambda)\}_{i=1}^m$. The emitted spectra of the ith type material can be represented by $\hat{p}_i(\lambda) = P_0 p_i(\lambda)$, where P_0 is another constant that carries the units of the emitted radiance [W/cm^2/sr/μm]. As a result, the *spectral pattern* $p_i(\lambda)$ can be assumed dimensionless. We define the universal linear space containing all the spectral patterns of interest and all spectral responses as the *spectral space*, Φ. For example, Φ can be the Hilbert space $L_2([0,\infty))$ of all real-valued square-integrable functions. The subspaces spanned by the spectral bands $\{f_i(\lambda)\}_{i=1}^k$ and the spectral patterns $\{p_i(\lambda)\}_{i=1}^m$ are respectively termed the *sensor space*, $\mathcal{F}$, and the *pattern space*, $\mathcal{P}$.

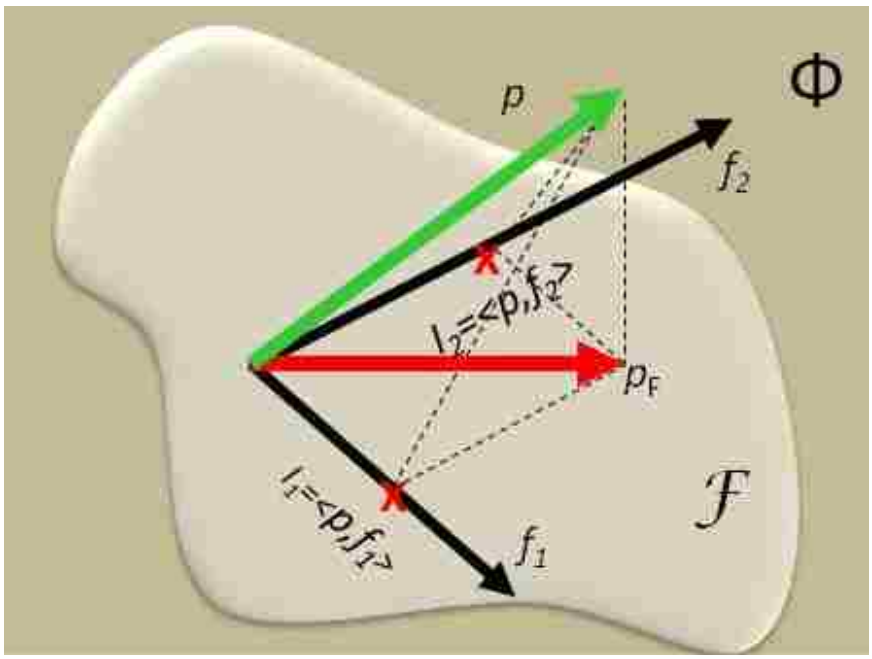


Figure 2.2: Spectral sensing model for sensors with non-overlapping bands.

Ideally, the process of sensing a pattern with a spectral sensor can be represented mathematically as an inner product between the pattern and each one of the sensor bands,

$$\langle p, f_i \rangle \triangleq \int_{-\infty}^{\infty} p(\lambda) f_i(\lambda) d\lambda, \tag{2.1}$$

producing a set of photocurrents, one for each band as shown in Fig. 2.2. In actuality, however, the photocurrents are perturbed by noise, yielding the noisy photocurrent I_i for the ith band sensing the pattern p,

$$I_i = \int_{\lambda_{\min}}^{\lambda_{\max}} p(\lambda) f_i(\lambda) d\lambda + N_i, \tag{2.2}$$

where N_i represents additive pattern-independent noise associated with the ith band, and the interval $[\lambda_{\min}, \lambda_{\max}]$ represents the common spectral support. Conceivably, different bands yield different noise levels, for example, due to different bias voltages in the case of a QDIP. The concept is illustrated in Fig. 2.3 again for the simple case

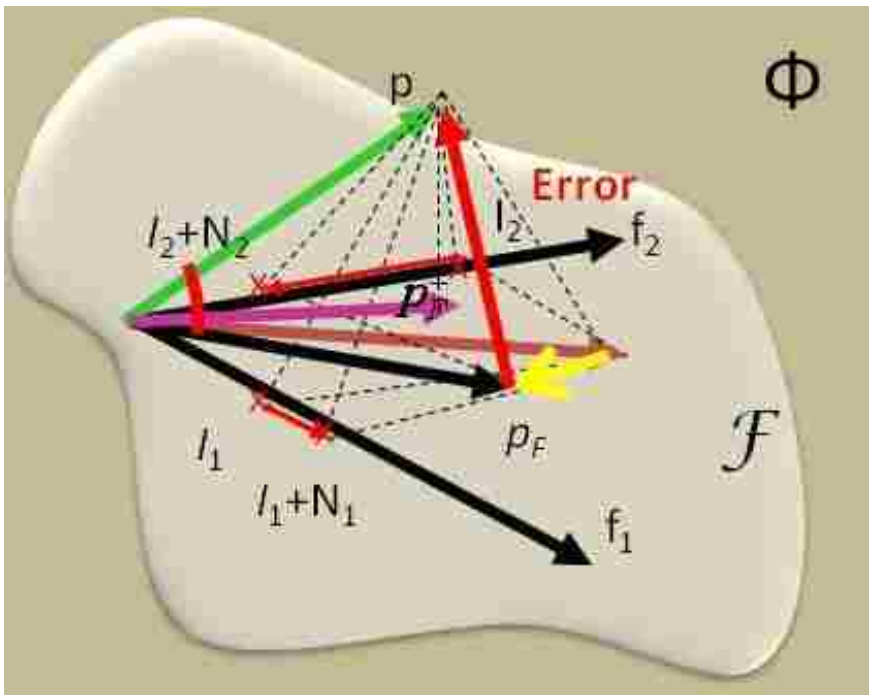


Figure 2.3: Spectral sensing model for sensors with broad and overlapping noisy bands.

of sensor with two bands and scene containing just one class of materials. For a given spectral pattern, the output corresponding to a single spectral band constitutes the *feature* of that pattern with respect to the band. A *spectral signature* is then defined as a k-dimensional vector in $\mathbb{R}^k$, whose coordinates are the measured photocurrents (features) associated with each spectral band.

2.2.2 Problem-specific feature selection

We now develop the key building blocks for our canonical feature selection algorithm. Specifically, we will seek to optimally replace the k-dimensional spectral signature in $\mathbb{R}^k$ with a single spectral feature. This transformed feature, $\tilde{I}$, for the pattern p,

is defined as weighted linear combination of all features, i.e., $\tilde{I} = \sum_{i=1}^{k} a_i I_i$, where the weights a_i are to be optimized for each pattern p. We term such a feature $\tilde{I}$ a *superposition current*. Equation (2.2) can then be expressed in the following form:

$$I_i = \sum_{i=1}^{k} a_i(\langle p, f_i \rangle + N_i) = \langle p, \sum_{i=1}^{k} a_i f_i \rangle + \sum_{i=1}^{k} a_i N_i. \tag{2.3}$$

From (2.3) we can deduce a useful analogy for the superposition current. Comparing this equation with (2.2), we see that the superposition current can be viewed as the output of an imaginary band, $f = \sum_{i=1}^{k} a_i f_i$. We will term the band f a *superposition band* since it is a weighted superposition of the sensor's bands and it is also associated with the superposition current.

Hitherto, the problem of determining the best superposition current, $\tilde{I}$, for a given spectral pattern can be thought of as the problem of determining the optimal superposition band f in $\mathcal{F}$ that offers the best approximation of p. Note that for a given superposition band f in $\mathcal{F}$, the approximation (or representation) of p rendered by this band is

$$p_f \triangleq (\langle p, \sum_{i=1}^{k} a_i f_i \rangle + \sum_{i=1}^{k} a_i N_i) f, \tag{2.4}$$

which is a vector in $\mathcal{F}$ that is along the direction of f but whose length is random due to noise.

Accordingly, one suitable criterion for the selection of a superposition band is to minimize the distance between the spectral pattern and its representation according to the superposition band. More precisely, we would select a set of coefficients $a_1, \ldots, a_k$ so that the L_2 norm of the error vector, $\|p - p_f\|$, is minimized. Noting that $f = \sum_{i=1}^{k} a_i f_i$, we have

$$p_f = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j (\langle p, f_i \rangle + N_i) f_j.$$

Hence, for a given pattern p, we propose an optimal superposition band, represented by the vector $\mathbf{a}^*$, as

$$\mathbf{a}^* \triangleq \underset{\mathbf{a}\in\mathbb{R}^k, \|f\|=1}{\operatorname{argmin}} \mathsf{E}\Big[\Big\|p - \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j(\langle p, f_i\rangle + N_i) f_j\Big\|^2\Big], \tag{2.5}$$

where $\mathbf{a} = (a_1, \ldots, a_k)^T$ is a weight vector associated with the superposition band f.

To provide a better insight into the criterion in (2.5), and particularly the constraint $\|f\| = 1$, let us assume for the moment that the noise is absent. In this case, one can show that the minimization of the noiseless versions of the criterion (2.5) is equivalent to computing the projection $p_{\mathcal{F}}$ of p onto $\mathcal{F}$. More precisely, let $p_{\mathcal{F}}$ be the orthogonal projection of p onto the subspace $\mathcal{F}$. By the minimum-distance property of the projection $p_{\mathcal{F}}$ [52, Theorem 4.11]

$$\inf_{g\in\mathcal{F}} \|p - g\| = \|p - p_{\mathcal{F}}\|.$$

The lemma bellow shows that $p_{\mathcal{F}}$ can be obtained, up to a sign difference, by projecting p onto unit-norm vectors in $\mathcal{F}$ and then selecting the vector that yields the minimum error between the projection along that unit vector and p.

Lemma 1 *Define* $f_p \triangleq \pm\dfrac{p_{\mathcal{F}}}{\|p_{\mathcal{F}}\|}$, *then*

$$\inf_{f\in\mathcal{F}} \|p - \langle p, f\rangle f\| = \min_{f\in\mathcal{F}, \|f\|=1} \|p - \langle p, f\rangle f\| = \|p - \langle p, f_p\rangle f_p\| = \|p - p_{\mathcal{F}}\|.$$

Proof. By using the fact that $(p - p_{\mathcal{F}})$ is orthogonal to $p_{\mathcal{F}}$, [52, Theorem 4.11], we obtain

$$\langle p, f_p\rangle f_p = \frac{\langle (p - p_{\mathcal{F}}) + p_{\mathcal{F}}, p_{\mathcal{F}}\rangle}{\langle p_{\mathcal{F}}, p_{\mathcal{F}}\rangle} p_{\mathcal{F}} = p_{\mathcal{F}}. \tag{2.6}$$

Therefore,

$$\|p - \langle p, f_p\rangle f_p\| = \|p - p_{\mathcal{F}}\|. \tag{2.7}$$

Hence, because $\inf_{g\in\mathcal{F}} \|p-g\| = \|p-p_{\mathcal{F}}\|$, (2.7) along with the fact that $\|f_p\| = 1$ together imply

$$\inf_{f\in\mathcal{F},\|f\|=1} \|p - \langle p, f\rangle f\| = \|p - \langle p, f_p\rangle f_p\|. \tag{2.8}$$

Thus, we have proved that the infimum in (2.8) is achieved at $f = f_p$, or

$$\inf_{f\in\mathcal{F},\|f\|=1} \|p - \langle p, f\rangle f\| = \min_{f\in\mathcal{F},\|f\|=1} \|p - \langle p, f\rangle f\| = \|p - \langle p, f_p\rangle f_p\|. \qquad \square$$

With the above interpretation of $p_{\mathcal{F}}$, and by realizing that the inner product associated with a superposition band represented by the weight vector $\mathbf{a}$ is corrupted by the additive noise $\sum_{i=1}^{k} a_i N_i$, as seen from (2.3), we arrive at the optimization criterion stated in (2.5). This justifies our selection of (2.5) as a criterion in the noiseless case and motivates its use as a meaningful criterion in the general case when the photocurrents are corrupted by additive noise.

The following lemma characterizes the minimization in (2.5).

Lemma 2 *Let $f = \sum_{i=1}^{k} a_i f_i$, $\mathbf{a} = (a_1, \ldots, a_k)^T$, and consider p_f given by (2.4). Without loss of generality, assume that $\|p\| = 1$, and further assume that the noise components in (2.4), $N_1, \ldots, N_k$, are zero-mean and independent random variables with variances σ_i^2, $i = 1, \ldots, k$. Then,*

$$\operatorname*{argmin}_{\mathbf{a}\in\mathbb{R}^k,\|f\|=1} \mathsf{E}\left[\left\|p_f - p\right\|^2\right] = \operatorname*{argmax}_{\mathbf{a}\in\mathbb{R}^k,\|f\|=1} \Big\{\langle p, f\rangle^2 - \sum_{i=1}^{k} a_i^2\sigma_i^2\Big\}. \tag{2.9}$$

Proof. Note that

$$\mathsf{E}\Big[\left\|p - p_f\right\|^2\Big] = \|p\|^2 - 2\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \langle p, f_i\rangle\langle p, f_j\rangle$$

$$+\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \langle p, f_i\rangle\langle p, f_j\rangle \|f\|^2$$

$$+\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \mathsf{E}[N_i N_j]\|f\|^2 - 2\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \mathsf{E}[N_i]\langle p, f_j\rangle$$

$$+2\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \mathsf{E}[N_i]\langle p, f_j\rangle \|f\|^2. \tag{2.10}$$

Using the stated assumptions on noise statistics and the norm of p, we obtain

$$\begin{aligned}
&\underset{\mathbf{a}\in\mathbb{R}^k,\|f\|=1}{\operatorname{argmin}} \mathsf{E}\Big[\big\|p - p_f\big\|^2\Big] \\
&= \underset{\mathbf{a}\in\mathbb{R}^k,\|f\|=1}{\operatorname{argmin}} \Big\{1 - \langle p, f\rangle^2 + \sum_{i=1}^{k} a_i^2\sigma_i^2\Big\} \\
&= \underset{\mathbf{a}\in\mathbb{R}^k,\|f\|=1}{\operatorname{argmax}} \Big\{\langle p, f\rangle^2 - \sum_{i=1}^{k} a_i^2\sigma_i^2\Big\}.
\end{aligned} \tag{2.11}$$

□

Lemma 2 provides useful information about the structure of the mean-square-error (MSE) in (2.9). If we define the SNR associated with the superposition band, f, represented by $\mathbf{a}$, as

$$\mathrm{SNR}_{\mathbf{a}} = \frac{\langle p, f\rangle^2}{\sum_{i=1}^{k} a_i^2\sigma_i^2} \tag{2.12}$$

the criterion (2.11) can be written in terms of $\mathrm{SNR}_{\mathbf{a}}$ as follows:

$$\underset{\mathbf{a}\in\mathbb{R}^k,\|f\|=1}{\operatorname{argmin}} \mathsf{E}\Big[\big\|p - p_f\big\|^2\Big] = \underset{\mathbf{a}\in\mathbb{R}^k,\|f\|=1}{\operatorname{argmax}} \Big\{\big(\mathrm{SNR}_{\mathbf{a}} - 1\big)\sum_{i=1}^{k} a_i^2\sigma_i^2\Big\}.$$

The quantity $\langle f, p\rangle^2$ in (2.12) reflects how much energy from the scene is preserved during the spectral sensing process and relates this energy to the mutual position, i.e., angle, between the pattern p and any sensor band f_i that contributes to the

superposition band. More precisely, defining the interior angle, θ_{p,f_i}, between the spectral pattern p and any sensor band f_i as

$$\theta_{p,f_i} = \cos^{-1}\left(\frac{\langle p, f_i\rangle}{\|p\|\|f_i\|}\right),$$

if a given pattern p is "almost collinear" to any of the sensor bands $\{f_i\}_{i=1}^{k}$, then θ_{p,f_i} will be nearly zero and the quantity $\langle p, f_i\rangle$ will attain its maximum value. In such cases, the contribution of that spectral band to the direction of the superposition band needs to be maximized in order to maximize the SNR for the superposition band. If $\mathcal{P} \subset \mathcal{F}$, then the angle between p and any f_i will be zero, meaning that the pattern space will be completely captured by the sensor space. On the other hand, if the angle between a given pattern $p \in \mathcal{P}$ and a spectral band $f_i \in \mathcal{F}$ is close to $\pi/2$, then this indicates lack of correlation between the spectral pattern and the spectral band. In such a case, the pattern cannot be sensed reliably by that particular band and the contribution of that band in the superposition band needs to be minimized.

In the presence of noise, due to the superpositions process, the noise variance corresponding to the superposition band will accumulate, resulting in lower SNR and therefore higher approximation error. As a result, the optimal superposition band in a noisy environment may not coincide with the direction of projection of the pattern onto the sensor space, and the amount of the deviation will depend upon the SNR for the individual bands.

In the next section, we use and extend the principle of optimal superposition band presented in this section to formulate the CCFS. This algorithm allows us to search for a set of weight vectors that yield the "best" collection of "sensing directions" minimizing the MSE in sensing a class of patterns.

2.3 Canonical correlation feature-selection algorithm

We begin by reviewing germane aspects of canonical correlation (CC) analysis [53, 54, 55] of two Euclidean subspaces. In essence, based on a computed sequence of principal angles, θ_k, between any two finite-dimensional Euclidean spaces $\mathcal{U}$ and $\mathcal{V}$, CC analysis yields the so-called canonical correlations, $\rho_k = \cos(\theta_k)$, between the two spaces. The first canonical correlation coefficient, ρ_1, is computed as $\rho_1 = \max_{i,j} \mathbf{u}_i^T \mathbf{v}_j$, where the vectors $\mathbf{u}_i$ $(i = 1, \ldots, m)$ and $\mathbf{v}_j$ $(i = 1, \ldots, n)$ are unit length vectors that span $\mathcal{U}$ and $\mathcal{V}$, respectively. The two vectors for which the maximum is attained are then removed, and ρ_2 is computed from the reduced sets of the bases. This process is repeated until one of the remaining subspaces becomes null.

The CC analysis approach, however, is not applicable to cases for which the inner products between vectors are accompanied by additive noise, as in the case of the photocurrents seen in (2.2). In this case, a stochastic version of "principal angle" must be introduced and used. This new criterion was precisely introduced in Lemma 2. Thus, in our approach we will follow the general principle of CC analysis while embracing the minimization stated in (2.9) as a criterion for maximal correlation.

In our formulation of the CCFS algorithm we will restrict the attention to finite dimensional spaces. Let us assume that all the spectral patterns and the sensor's bands belong to an n-dimensional subspace of the Hilbert space Φ. Thus, without loss of generality, we can think of the Hilbert space Φ as $\mathbb{R}^n$, and the functions $p \in \mathcal{P}$ and $f \in \mathcal{F}$ as Euclidean vectors $\mathbf{p}$ and $\mathbf{f}$ in $\mathbb{R}^n$, where $\mathbf{p}$ and $\mathbf{f}$ are the coordinate vectors of f and p, respectively. Furthermore, the inner product $\langle p, f \rangle$ can be represented by the dot product $\mathbf{p}^T\mathbf{f}$.

Assume further that $\mathcal{F}$ is the span of k ($k \leq n$) linearly independent spectral bands, represented by the columns of a matrix $\mathbf{F} = [\mathbf{f}_1|\ldots|\mathbf{f}_k]$. We term $\mathbf{F}$ the *filter matrix*. Let $\mathcal{P}$ denote the span of a set of m linearly independent patterns $\{\mathbf{p}_i\}_{i=1}^m$ representing the means of each one of m classes of interest. The matrix $\mathbf{P} = [\mathbf{p}_1|\ldots|\mathbf{p}_m]$ is termed the *pattern matrix*. We will assume further that $m < k$.

The CCFS algorithm begins the search for the first canonical band by determining m weight vectors $\mathbf{a}_i$, $i = 1, \ldots, m$, one for each class of interest. In particular, for the mean of the lth class, we determine a vector of weights $\mathbf{a}_l = (a_{l,1}, \ldots, a_{l,k})^T$ as

$$\begin{aligned}\mathbf{a}_l &= \underset{\mathbf{a}\in\mathbb{R}^k, \|\mathbf{Fa}\|=1}{\operatorname{argmin}} \mathsf{E}\left[\left\|\mathbf{p}_l - (\mathbf{p}_l^T\mathbf{Fa} + \mathbf{n}^T\mathbf{a})\mathbf{Fa}\right\|^2\right] \\ &= \underset{\mathbf{a}\in\mathbb{R}^k, \|\mathbf{Fa}\|=1}{\operatorname{argmax}} \left\{\left(\frac{\left(\mathbf{p}_l^T\mathbf{Fa}\right)^2}{\mathbf{a}^T\Sigma_{\mathbf{N}}\mathbf{a}} - 1\right)\mathbf{a}^T\Sigma_{\mathbf{N}}\mathbf{a}\right\},\end{aligned} \tag{2.13}$$

where the last equality follows from the identity (2.9) in Lemma 2, specialized to the present setting. Specifically, in (2.13) each component $a_{l,j}$ weights the corresponding sensor band $\mathbf{f}_j$, $j = 1, \ldots, k$. Note that (2.13) is the equivalent matrix representation of (2.5), where

$$\mathbf{n} = \left(N_1, \ldots, N_k\right)^T$$

is a random vector whose components N_i are independent, zero-mean random variables with variance σ_i^2. We reiterate our earlier assertion in Section 2.2 that for each pattern $\mathbf{p}_i$ minimizing (2.13) is equivalent to selecting a direction $\sum_{j=1}^k a_{i,j}\mathbf{f}_j$ in $\mathcal{F}$ that satisfies (2.9) and exhibits minimal combined noise variance and angle between the pattern and the direction.

The minimization process outlined in (2.13) is repeated m times determined by the number of classes of interest, where each class is represented by its mean $\mathbf{p}_i$, $i = 1, \ldots, m$. This process yields a set of m superposition bands, or sensing directions, $\mathbf{f}^1 = \mathbf{Fa}_1, \cdots, \mathbf{f}^m = \mathbf{Fa}_m$, each one optimized with respect to the mean of each class. If the feature-selection algorithm stops here and the so determined set of m

superposition bands is used, it can be the case that these bands span a very small subspace of the sensor space, because almost collinear patterns will determine almost collinear directions.

The algorithm continues by selecting from this optimized set of superposition bands the one that is the most "collinear" with its corresponding mean, i.e., the superposition band that gives the minimum MSE for a particular class:

$$\begin{aligned}\tilde{\mathbf{f}}^1 &= \underset{\mathbf{f}^i;i=1,\ldots,m}{\operatorname{argmin}} \mathsf{E}\left[\left\|\mathbf{p}_i - \left(\mathbf{p}_i^T\mathbf{f}^i + \mathbf{n}^T\mathbf{a}_i\right)\mathbf{f}^i\right\|^2\right] \\ &= \underset{\mathbf{f}^i;i=1,\ldots,m}{\operatorname{argmax}} \left\{\left(\frac{\left(\mathbf{p}_i^T\mathbf{f}^i\right)^2}{\mathbf{a}_i^T\Sigma_{\mathbf{N}}\mathbf{a}_i} - 1\right)\mathbf{a}_i^T\Sigma_{\mathbf{N}}\mathbf{a}_i\right\}, \end{aligned} \tag{2.14}$$

where the last equality follows from Lemma 2. We term the superposition band $\tilde{\mathbf{f}}^1$ the *first canonical band*.

To ensure complete cover of the scene space within the filter space, the search for the second canonical band $\tilde{\mathbf{f}}^2$ is conducted in the orthogonal complement of $\tilde{\mathbf{f}}^1$ and it is with respect to the means of the remaining classes. More precisely, if $\tilde{\mathbf{f}}^1 = \mathbf{f}^{\ell_1}$, for some $\ell_1 \in \{1,\ldots,m\}$, then the ℓ_1th class is excluded from the search for $\tilde{\mathbf{f}}^{\mathbf{2}}$.

In general, if $\tilde{\mathbf{f}}^j$ is the jth optimal superposition band, then $\tilde{\mathbf{f}}^{j+1}$ is selected by searching in the orthogonal complement of $\tilde{\mathbf{f}}^1,\ldots,\tilde{\mathbf{f}}^j$ and over all classes less the $\ell_1,\ldots,\ell_j$th classes, where ℓ_i is defined through $\tilde{\mathbf{f}}^i = \mathbf{f}^{\ell_i}$. We continue in this fashion until we obtain a set of m canonical bands $\tilde{\mathbf{f}}^1,\ldots,\tilde{\mathbf{f}}^m$. The process is illustrated in Fig. 2.4 for a simple case of a sensor with two bands and a scene with two classes. Note, that the canonical order of the superposition bands does not depend on the presentation order of the classes of interest, because at the end of each optimization cycle the algorithm always selects the pair that yields the smallest estimation error. Each one of these canonical bands can be applied to the data to yield the so-called CC features.

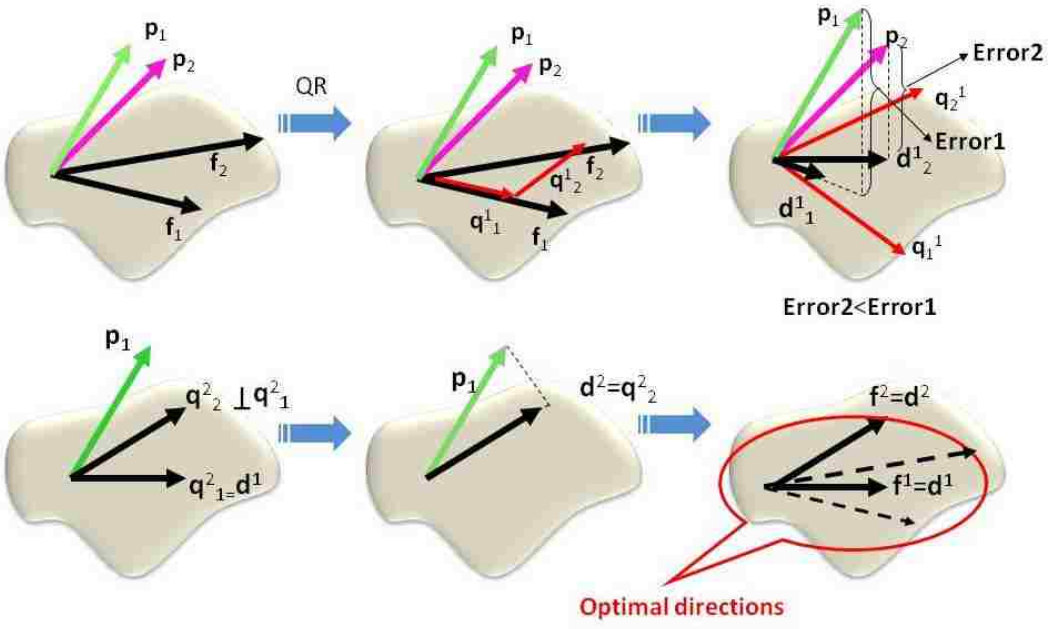


Figure 2.4: Canonical correlation feature selection algorithm.

2.3.1 Implementation of the CCFS

Because the spectral bands $\mathbf{f}_i$, $i = 1, \ldots, k$, are highly correlated, they provide a numerically ill-conditioned basis set for $\mathcal{F}$. Instead of solving (2.13) directly, we may replace this problem by an equivalent problem for which the minimization is carried out with respect to an orthonormal basis set for $\mathcal{F}$. This replacement will also speed up the numerical implementation of the optimization.

More precisely, let $\mathbf{F} = \mathbf{QR}$ be the reduced QR factorization of the matrix $\mathbf{F}$. Then the minimization problem

$$\underset{\mathbf{a}\in\mathbb{R}^k, \|\mathbf{QRa}\|=1}{\operatorname{argmin}} \mathsf{E}\left[\left\|\mathbf{p}_i - (\mathbf{p}_i^T\mathbf{QRa} + \mathbf{n}^T\mathbf{a})\mathbf{QRa}\right\|^2\right] \tag{2.15}$$

is equivalent to the problem in (2.13). Moreover, the optimization criterion in (2.15)

can be recast in the equivalent form

$$\begin{aligned} &\underset{\mathbf{b}\in\mathbb{R}^k,\|\mathbf{Qb}\|=1}{\operatorname{argmin}} \mathsf{E}\Big[\big\|\mathbf{p}_i - \mathbf{p}_i^T\mathbf{Q}\mathbf{b}\mathbf{Q}\mathbf{b} - \mathbf{n}^T\mathbf{R}^{-1}\mathbf{b}\mathbf{Q}\mathbf{b}\big\|^2\Big] \\ &\quad = \underset{\mathbf{b}\in\mathbb{R}^k,\|\mathbf{Qb}\|=1}{\operatorname{argmin}} \Big[1-(\mathbf{p}_i^T\mathbf{Q}\mathbf{b})^2 + (\mathbf{R}^{-1}\mathbf{b})^T\Sigma_{\mathbf{N}}\mathbf{R}^{-1}\mathbf{b}\Big], \end{aligned} \tag{2.16}$$

where $\mathbf{b} = \mathbf{Ra}$ is the set of weights for the ith class mean derived with respect to the orthonormal basis set $\{\mathbf{q}_i\}_{i=1}^k$ for $\mathcal{F}$, and $\mathbf{q}_i$ is the ith column of $\mathbf{Q}$. We recall that the representation with respect to the original sensor space is $\mathbf{f} = \sum_{i=1}^k a_i\mathbf{f}_i = \mathbf{Fa}$, where $\mathbf{a} = \mathbf{R}^{-1}\mathbf{b}$. Then,

$$\mathbf{q} = \sum_{i=1}^k b_i\mathbf{q}_i = \mathbf{Qb} = \mathbf{FR}^{-1}\mathbf{b} = \mathbf{Fa}\,.$$

Therefore, minimizing (2.16) with respect to $\mathbf{b}$ is equivalent to minimizing (2.15) with respect to the coefficients $\mathbf{a}$.

Let $\mathbf{n}^1 = \mathbf{nR}^{-1}$ where $\mathbf{n}$ is a random vector. The mean of the transformed random vector can be calculated according to the formula

$$\mathsf{E}[\mathbf{n}^1] = \mathsf{E}[\mathbf{nR}^{-1}] = \mathbf{R}^{-1}\mathsf{E}[\mathbf{n}] = 0\,. \tag{2.17}$$

Likewise, using that

$$\mathsf{E}\Big[\mathbf{n}^T\mathbf{n}\Big] = \Sigma_{\mathbf{N}} = \operatorname{diag}(\sigma_i^2)$$

we easily find the following formula for the covariance of $\mathbf{n}^1$:

$$\begin{aligned} \mathsf{E}\Big[(\mathbf{n}^1)^2\Big] &= \mathsf{E}\Big[(\mathbf{nR}^{-1})^2\Big] \\ &= \mathsf{E}\Big[(\mathbf{R}^{-1})^T\mathbf{n}^T\mathbf{n}\mathbf{R}^{-1}\Big] = (\mathbf{R}^{-1})^T\mathsf{E}\Big[\mathbf{n}^T\mathbf{n}\Big]\mathbf{R}^{-1} = (\mathbf{R}^{-1})^T\Sigma_{\mathbf{n}}\mathbf{R}^{-1}\,. \end{aligned}$$

In the sequel we use the notation

$$\Sigma_N^1 \triangleq (\mathbf{R}^{-1})^T\Sigma_{\mathbf{n}}\mathbf{R}^{-1} = \mathsf{E}\Big[(\mathbf{n}^1)^2\Big]\,. \tag{2.18}$$

Form the above discussion it follows that we have replaced the minimization with respect to the ill-conditioned sensor basis in (2.13) by minimization with respect to an orthonormal basis. In the new optimization problem the filter matrix $\mathbf{F}$ is replaced by $\mathbf{Q}$ and the noise components $\mathbf{n}$ are replaced by their linear transformations $\mathbf{n}^1$, still with zero mean. However, the transformed noise components are not independently distributed anymore and their covariance matrix is given by (2.18).

Let us now discus in more details the jth step of the CCFS. We will use the symbols $\mathbf{a}^j$, $\mathbf{b}^j$, $\mathbf{Q}^j$, $\mathbf{R}^j$ and $\mathbf{f}^j$ to denote the quantities computed at the jth step of the CCFS. The optimal direction at step j is given by

$$\tilde{\mathbf{f}}^j = \tilde{\mathbf{Q}}^j \tilde{\mathbf{b}}^j \,. \tag{2.19}$$

To find the orthogonal complement of $\tilde{\mathbf{f}}^j$ at step $j+1$, we compute the QR factorization of $\mathbf{F}^{j+1}$, where

$$\mathbf{F}^{j+1} = [\tilde{\mathbf{f}}^j | \mathbf{q}_1^j | \dots | \mathbf{q}_{k-j}^j] \subset \text{span}\{\mathbf{r}_i\}_{i=1}^k \,. \tag{2.20}$$

Then, $\mathbf{F}^{j+1} = \mathbf{Q}^{j+1}\mathbf{R}^{j+1}$, where $\mathbf{Q}^{j+1} = [\mathbf{q}_1^{j+1} | \mathbf{q}_2^{j+1} | \dots | \mathbf{q}_{k-j+1}^{j+1}]$ and $\mathbf{q}_1^{j+1} = \tilde{\mathbf{f}}^j$. Therefore, the orthogonal complement of $\tilde{\mathbf{f}}^j$ is given by

$$(\mathbf{q}_1^{j+1})^\perp = (\tilde{\mathbf{f}}^j)^\perp = \tilde{\mathbf{Q}}^{j+1} = [\mathbf{q}_2^{j+1} | \dots | \mathbf{q}_{k-j+1}^{j+1}] \,. \tag{2.21}$$

It follows that at the $j+1$-st step of the CCFS, we have to solve the minimization problem

$$\min_{\mathbf{b}^{j+1}} \mathsf{E}\Big[\| \mathbf{p} - (\langle \mathbf{p}, \tilde{\mathbf{Q}}^{j+1}\mathbf{b} \rangle + \mathbf{N}^{j+1}) \tilde{\mathbf{Q}}^{j+1}\mathbf{b} \|^2 \Big] \,, \tag{2.22}$$

where $\mathbf{n}^{j+1}$ is the noise vector containing the noise components with respect to each basis vector $\{\mathbf{q}_i^{j+1}\}_{i=1}^{k-j+1}$. At the $j+1$-st step, this vector is linearly transformed according to

$$\mathbf{n}^{j+1} = \mathbf{n}^1 \mathbf{X}^{j+1} \,, \tag{2.23}$$

where $\mathbf{X}^{j+1} = \mathbf{R}^{1(-1)}\mathbf{Q}^{1T}\tilde{\mathbf{Q}}^{j+1} = \mathbf{T}\tilde{\mathbf{Q}}^{j+1}$, and $\mathbf{T} \triangleq \mathbf{R}^{1(-1)}\mathbf{Q}^{1T}$. The transformed covariance is given by

$$\mathsf{E}\Big[(\mathbf{N}^{j+1})^2\Big] = \mathsf{E}\Big[(\tilde{\mathbf{Q}}^{j+1})^T\mathbf{Q}^1(\mathbf{R}^1)^{-T}\mathbf{N}^T\mathbf{N}(\mathbf{R}^1)^{-1}\ (\mathbf{Q}^1)^T\tilde{\mathbf{Q}}^{j+1}\Big] \tag{2.24}$$

$$= (\tilde{\mathbf{Q}}^{j+1})^T\mathbf{T}\Sigma_N\mathbf{T}^T\tilde{\mathbf{Q}}^{j+1}\,. \tag{2.25}$$

We can conclude that at each optimization step, the noise mean remains the same. However, after the first optimization step, the covariance matrix is not diagonal anymore, and the correlation among the noise components depends on the current and the initial orthonormal bases for the sensor space.

Because the optimal directions have physical meaning only with respect to the sensor bands, at each step the algorithm recalculates the weights (coefficients) of the linear combinations with respect to these bands. We have already shown that at the first step $\tilde{\mathbf{f}}^1 = \mathbf{Q}^1\mathbf{b}^1 = \mathbf{F}^1\mathbf{a}^1$ where $\mathbf{F}^1 \triangleq \mathbf{F}$, and $\mathbf{a}^1 = (\mathbf{R}^1)^{-1}\mathbf{b}^1$. At each step, we seek the coefficients of the representation in the form

$$\tilde{\mathbf{f}}^j = \mathbf{F}^1\mathbf{a}^j\,. \tag{2.26}$$

Substituting (2.19) into (2.26) and using the decomposition $\mathbf{F}^1 = \mathbf{Q}^1\mathbf{R}^1$, equation (2.26) can be rewritten as

$$\tilde{\mathbf{Q}}^j\tilde{\mathbf{b}}^j = \mathbf{Q}^1\mathbf{R}^1\mathbf{a}^j\,. \tag{2.27}$$

Therefore the coefficients with respect to the original filter space at each step are:

$$\mathbf{a}^j = (\mathbf{R}^1)^{-1}\mathbf{Q}^{1T}\tilde{\mathbf{Q}}^j\tilde{\mathbf{b}}^j\,. \tag{2.28}$$

2.3.2 Summary of the CCFS

In this section we outline the key steps of the CCFS. The algorithm is implemented in Matlab$^{\text{TM}}$ using the Optimization toolbox.

1. Given a set $\{\mathbf{p}_1, \ldots, \mathbf{p}_m\}$ of m classes of interest, estimate the mean of each class.

2. Compute the QR factorization of the basis $\mathbf{F}$ for the filter space.

3. Solve the optimization problem in (2.11) for each class mean to find the optimal superposition bands $\{\mathbf{f}^1, \ldots, \mathbf{f}^m\}$.

4. From the set $\{\mathbf{f}^1, \ldots, \mathbf{f}^m\}$, select the superposition band $\tilde{\mathbf{f}}^1$, which gives the minimum MSE according to (2.14).

5. Using QR factorization find the orthogonal complement $(\tilde{\mathbf{f}}^1)^\perp$ of the superposition band determined in Step 4.

6. Retain $(\tilde{\mathbf{f}}^1)^\perp$ while discarding $\tilde{\mathbf{f}}^1$ and the associated class from the pattern space.

7. Repeat Steps 3-6 until the pattern space $\mathcal{P}$ is empty.

2.4 Conclusions

We have developed a problem-specific feature-selection algorithm, termed CCFS, that is appropriate for the general class of sensors whose bands are both noisy and spectrally overlapping. Our approach is based upon statistical projection-like concepts in Hilbert spaces in conjunction with canonical correlation analysis. For a given class of patterns, the CCFS algorithm seeks for a set of weights that are used to determine the optimal superposition band or sensing direction. The obtained sensing direction is optimal in a sense that it provides the best MMSE estimate of the mean of a class in the sensor space. In particular, the superposition band yields the best sensing direction, taking into account both information content and noise. The superposition-band selection procedure is repeated sequentially as many times as the number of the classes of interest, producing a canonical set of superposition bands.

At each stage, the algorithm excludes from the search for the optimal direction the class that has been selected in the prior stage; moreover, every superposition band is selected from a subspace of the sensor space that is in the orthogonal complement of the previous sensing direction.

Chapter 3

Application of the CCFS to Classification and Remote Sensing Problems

In this Chapter we demonstrate the efficacy and the scope of the CCFS algorithm. To this end, three different applications are considered. The first application is separability and classification analysis of rock species using laboratory spectral data and a quantum-dot infrared photodetector (QDIP). The QDIP belongs to the class of bias tunable DWELL photodetectors. In this study the CCFS algorithm is applied to the spectral responses of the QDIP and laboratory measured spectral rock data for the purpose of separability and classification analysis of seven classes of rocks [56, 57]. The second application deals with supervised classification and spectral unmixing, and abundance estimation of hyperspectral (HS) imagery obtained from the Airborne Hyperspectral Imager (AHI) sensor.

Our third application also deals with rock-type classification; however, in this case we apply the CCFS to real DWELL FPA imagery acquired at the CHTM.

Our results demonstrate, for a first time, the MS capability of the DWELL FPA by considering three different scene configurations. We conduct thorough analysis of the separability and classification errors between granite, limestone and an optical filter, as a function of the applied bias. Finally, the CCFS algorithm is applied to the classification problem of discriminating between granite and hornfels. We compare the CCFS classification results with classification results obtained using different combinations of the original DWELL bands.

Results from the first and the third applications demonstrate that proper post-processing can facilitate the emergence of bias tunable sensors, such as QDIP and DWELLs, as a promising technology for midwave- and longwave-infrared remote sensing and spectral imaging. Furthermore, our second study shows that the scope of the CCFS also extends to more traditional sensors with fixed bands, such as AHI.

3.1 Rock-type classification using laboratory data

In the last few decades, the LWIR wavelengths have been used successfully to distinguish a number of primary silicates (feldspars, quartz, opaline silica) that are spectrally bland or have features that are non-unique at shorter wavelengths [58]. Thus, the thermal-infrared (TIR) region of the spectrum is excellent for examining pure samples as well as mineralogically complex geologic materials (i.e., rocks) and is gaining popularity as a remote sensing wavelength range for geologic applications [59], [60]. Our previous investigation of the rock-type classification problem, using Multispectral Thermal Imager (MTI) that operates in the shortwave (SW), MWIR and LWIR portions of the spectrum has shown that the MTI sensor in conjunction with supervised Bayesian classifier offers high discrimination accuracy among the different rock types; hence, MTI performance can serve as a good benchmark in this case study [56]. (MTI was designed to be a satellite-based system for terrestrial

observation with an emphasis on obtaining qualitative information of the surface temperature [61]. Currently, MTI operates with set of 15 bands, covering the broad range from 0.45 μm to 10.7 μm.)

3.1.1 Definition of training and testing sets

Generally, rocks can be divided into three main geological groups: *igneous*, *metamorphic* and *sedimentary*, which correspond to the different geological processes involved in the rock's formation. Geologists have further divided these three main rock categories into seven generic classes, which we have adopted in this study. To create the training and testing data sets, we select a number of spectra of common rock samples in different grain sizes from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) hyperspectral database [62]. Table 3.1 describes the rock classes and the endmembers included in the training set.

The limited number of endmembers, shown in Table 3.1, however, prevents direct application of a Bayesian classifier. This fact forces us to increase the size of the training set by perturbing the endmebers in each rock-class with different mixing materials. To create the perturbations we use a simple, two-component linear mixing model, where each mixture is considered as a linear combination of a representative endmember and a mixing endmember, weighted by the correspondent abundance function β. For the abundance function, we use five randomly chosen values of β between 1% and 10% for the mixing endmembers, and $(100-\beta)$% for the representative endmembers. Using the above mixing model, we create spectral mixtures of the representative endmembers with minerals, vegetation, soil and water [56]. We also create mixtures between fine- and coarse-size rocks, and between coarse- and fine-size rocks, according to their geological properties that make such mixtures realistic. All mixing endmembers used to enlarge the training set are presented in Table 3.2.

Group	**Endmembers**
Hornfelsic	hornfels (fine, coarse)
Granoblastic	pink quartzite, marble (fine, coarse) and gray quartzite (coarse)
Schistose	gray slate, chlorite schistose (fine, coarse) and chlorite
Semischistose	felstic gneiss (fine, coarse)
Igneous	andesite, basalt, diorite, gabbro, granite, rhyolite (fine, coarse), tan rhyolite and tuff (cup 8, 9)
Clastic Sedimentary	shale, siltstone, fossiliferous limestone and red sandstone (fine, coarse)
Chemical Sedimentary	limestone (fine, coarse) and dolomite

Table 3.1: Rock-type groups and their representative endmembers.

Minerals	andradite, anorthite, dolomite, quartz and topaz
Rocks	basaltic andesite, diorite gneiss, limestone and siltstone (fine and coarse)
Water, Soil	distilled water, see water, dark brown loam, fine sandy loam and brown to dark brown sand
Vegetation	conifer and grass (green), spruce cellulose, citrus pectin, white peppermint, CA buckwheat, brown sycamore and brown leaf (dry)

Table 3.2: Mixing endmembers used to create random perturbations of the representative endmembers listed in Table 3.1.

Minerals	andradite, antigorite, erionite, fluorite, quartz and spodumene
Rocks	basalt, pink marble and black shale (fine and coarse)
Water, Soil	see foam, grayish brown loam, dark grayish brown silty loam, reddish fine sandy loam and dark reddish fine sandy loam
Vegetation	deciduous (green), cotton cellulose, citrus pectin, sycamore-loer (yellow), CA brown buckwheat and sycamore (dry)

Table 3.3: Mixing endmembers used to create random perturbations of the representative endmembers listed in Table 3.1 in order to create Test Set-1 and Test Set-2.

Figure 3.1 shows spectral signatures of the endmembers for the class hornfelsic, fine and coarse size in thick black lines, as well as their mixtures with rocks, minerals, soils and vegetation.

For our study we created two testing sets where the mixing endmembers used to create these sets are shown in Table 3.3. In Set-1, the representative endmembers in Table 3.1 were perturbed with rocks listed in Table 3.3. For the abundance function, we use five randomly chosen values within the range 1% to 10%. Set-2 is an enlargement of Set-1 with the addition of mixtures of the representative endmembers, as shown in Table 3.1, with soils, minerals and vegetation listed in Table 3.3.

The addition of all the mixtures helps to increase the rank of the covariance matrix to 13 in the case of QDIP and 11 in the case of MTI, which still fails short of full rank for 26-dimensional data in the case of QDIP and 13-dimensional data in the case of MTI. To mitigate this problem, we select a subset of 13 arbitrary QDIP bands. The performance of this subset is averaged over different arbitrarily selected subsets of 13 bands. In the case of MTI, we were able to identify high correlation for bands C and L with their adjacent spectral bands, so they were removed without

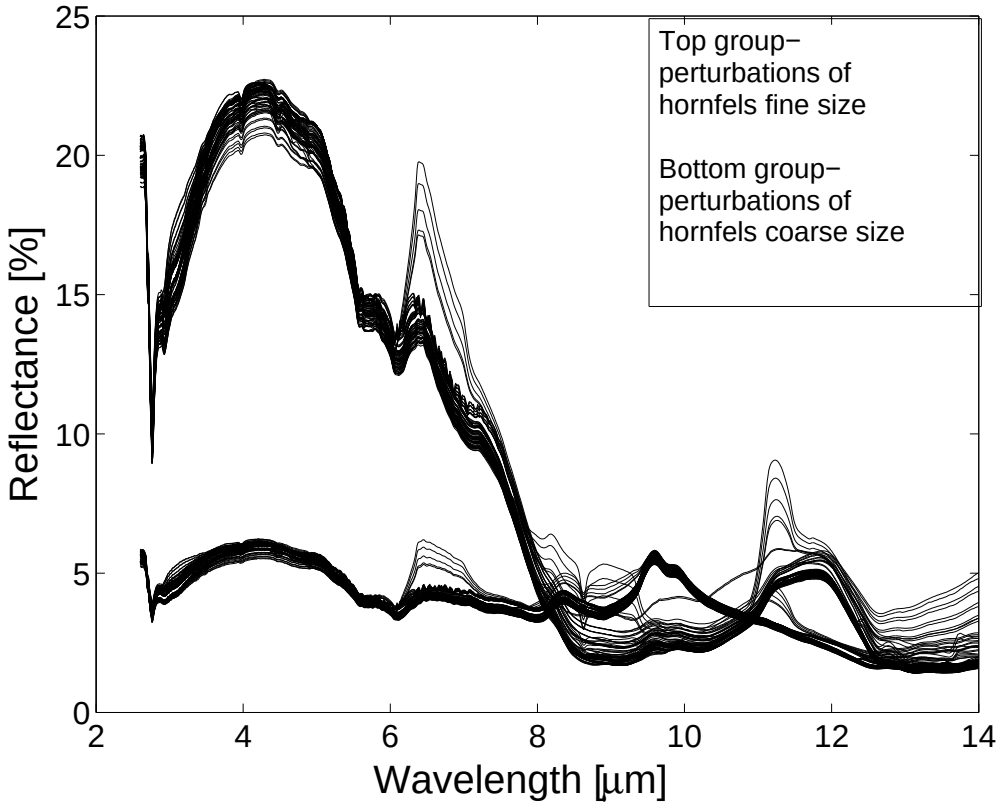


Figure 3.1: Reflectivity of the hornfels showing fine (top group) and coarse size (bottom group), as well as their perturbations.

loosing relevant information.

We employ a supervised Bayesian classifier with the assumptions for normal class populations and equal priors [63]. The second assumption is reasonable as the training set was defined by geologists in accordance with the geological properties of the rocks; thus, the number of samples in the training set for a certain group does not represent the frequency of occurrence of the rocks in the nature. Instead, number of samples per class reflects the rock diversity within a given class.

3.1.2 Separability and classification results

To set a benchmark for the performance of the CCFS algorithm, we begin by presenting the separability and classification results in the ideal case when noise is absent, and without using the CCFS algorithm.

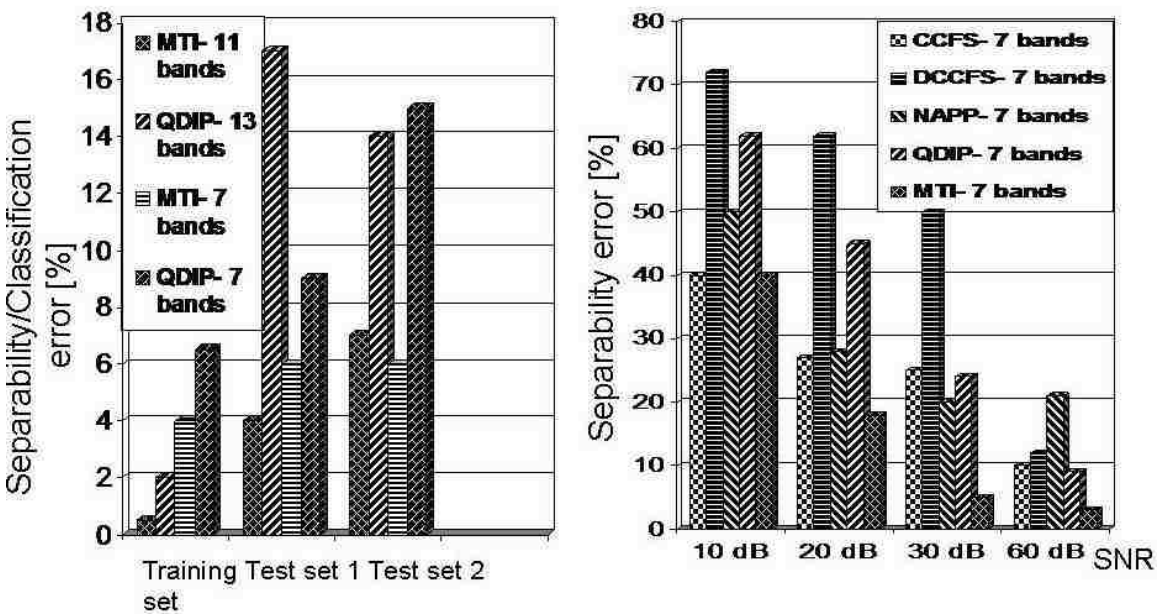


Figure 3.2: Left: comparison in rock-type separation and classification in the absence of noise. Right: Comparison in rock-type separation for CCFS, DCCFS, noise-adjusted PP, 7 QDIP bands and 7 MTI bands in presence of noise with average SNR values of 10, 20, 30 and 60dB.

We first compare separability and classification performance for QDIP and MTI sensors. Four sets of separability and classification results are summarized in Fig. 3.2, left and right respectivly. The first set corresponds to using 13 arbitrary bands out of the 26 QDIP bands. The second set of results corresponds to using 11 out of the 15 MTI bands (bands A-E, G, I, O, J, K and M) [61]. The third set is based on

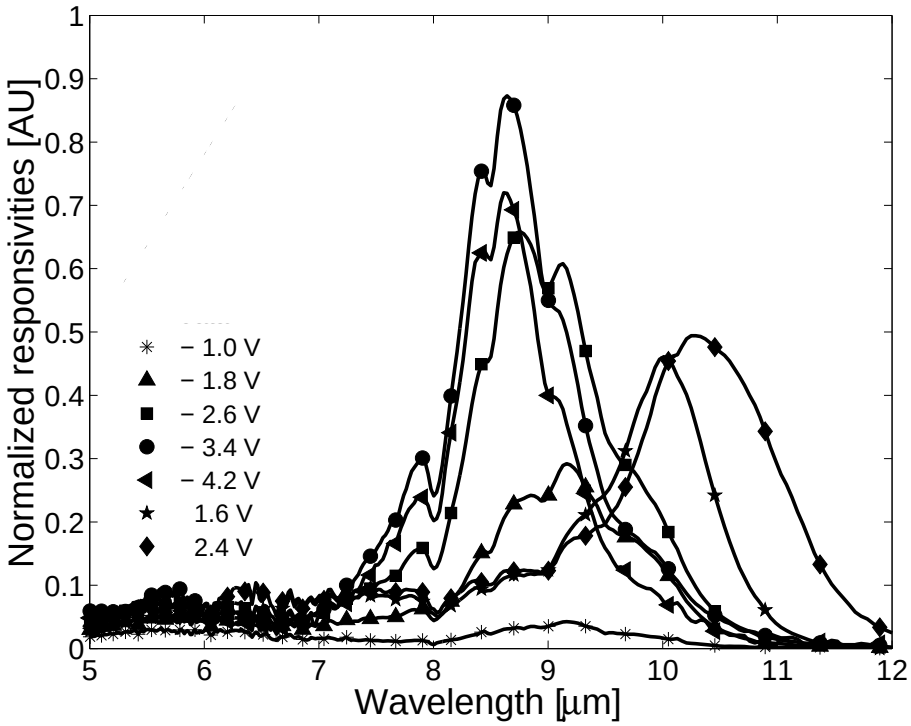


Figure 3.3: Seven QDIP bands used in the rock-type classification.

a subset of 7 arbitrarily selected QDIP bands, shown in Fig. 3.3. The final, fourth set of results is based on 7 MTI bands (bands G, I, O, J, K, M and N) selected to approximate the spectral range of the QDIP bands. The results presented in Fig. 3.2 (left) suggest that the MTI and QDIP bands yield comparable performance in the absence of noise.

3.1.3 Effect of noise

In this section we consider the presence of noise and compare the separability and classification results for the CCFS algorithm with four different cases, each using 7 bands and for four different SNR values. The results are averaged over 100 indepen-

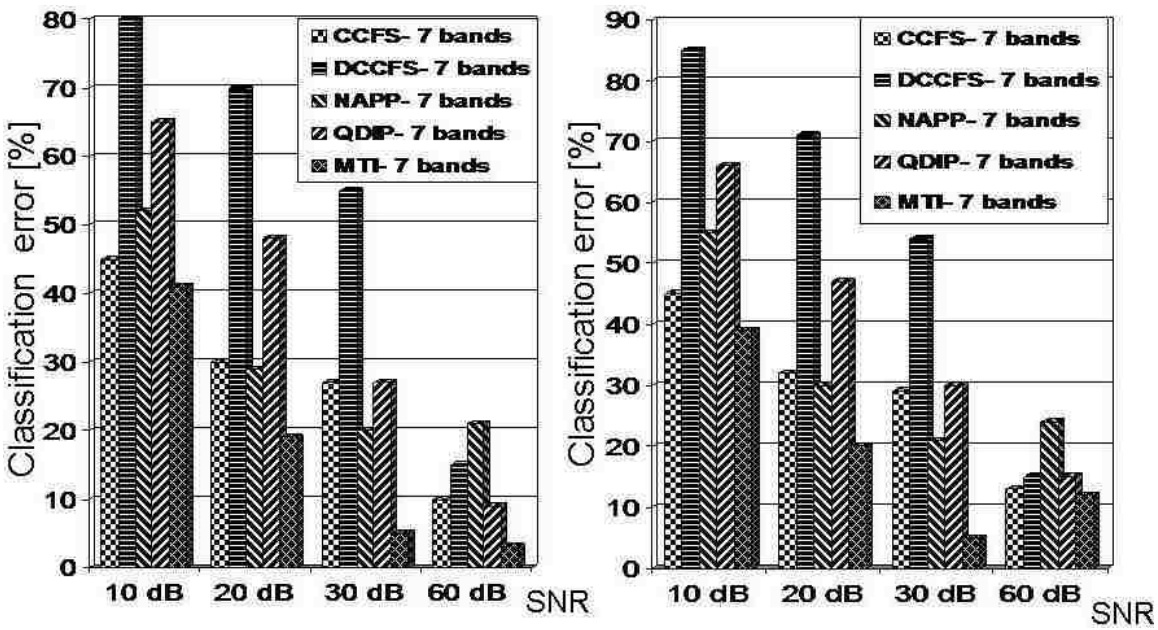


Figure 3.4: Comparison in rock-type classification for CCFS, DCCFS, noise-adjusted PP, QDIP bands and MTI bands in presence of noise with average SNR values of 10, 20, 30 and 60dB. Left: Test Set-1. Right: Test Set-2.

dent noise realizations for each SNR value. We define the SNR as

$$\text{SNR} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{(\mathbf{f}_i^T \mathbf{p}_j)^2}{\sigma_i^2}, \tag{3.1}$$

where M and N respectively denote the number of sensor's bands and the number of the spectral patterns included in the training set. To achieve a particular SNR value, we scale the noise mean and variance.

Throughout this section we assume that the bias-dependent noise components N_i, where i is the band (or bias) index, are zero-mean normally distributed random variables. This follows from the fact that amplitude distributions for both thermal and shot noise converge to normal distributions by the central limit theorem. For the large number of electrons generating the thermal noise, the amplitude distribution of the thermal noise converges to zero-mean normal distribution. On the other hand,

the actual numbers of generation-recombination events underlying the shot noise will exhibit a Poisson distribution [47]. However, this number will become approximately normally distributed for a large average number of generation-recombination events [64]. Therefore, the amplitude distribution of the total noise will be also normal with mean equal to the mean of the shot noise and a variance equal to the sum of the variances of the two types of noise. Since the mean of the shot noise is deterministic and known (being equal to the DC value of the measured dark current), it can be subtracted from the noise without having any ramifications on the analysis or the algorithm development.

For this case study, the number of selected superposition bands is determined by the number of classes of interest - 7. The first case is termed *deterministic* CCFS (DCCFS) and it employs the CCFS algorithm but without accounting for the photocurrent noise during the selection process. In the second case, termed *noise-adjusted PP* [42, 65], we use seven features extracted using the noise-adjusted PP algorithm. Finally, the last two cases correspond to the classifiers used in Fig. 3.2 (left) applied to noisy data; these cases are termed *QDIP-7 bands* and *MTI-7 bands*. Figures 3.2 (right) and 3.4, left and right, compare the separability and classification performances respectively (for the training and testing sets) for the five cases described above.

The first observation is that embedding the noise statistics in the canonical feature-selection leads to a significant improvement in the classification. As we can see from the results presented in Fig. 3.2 (right) and Fig. 3.4, for the first three SNR cases (average SNR of 10dB, 20dB and 30dB), the CCFS algorithm performs almost twice as good as the DCCFS algorithm. In the limiting case of a very high SNR, the performance of the CCFS and DCCFS algorithms becomes almost identical, as expected, and the classification error drops to 10-15%.

We next compare the CCFS algorithm with the arbitrary selection of 7 QDIP

bands. For the average SNR of 10dB, as shown in Fig. 3.2 (right) the separability error from the latter case is 63%, compared to 41% in the CCFS case. This result underscores the higher sensitivity of QDIP bands to significant noise levels compared to the canonical superposition bands. Notably, by using the CCFS algorithm we were able to achieve a significant improvement in the classification performance (approximately 20%). As expected, when the average SNR increases, the performances of the two cases become comparable.

The separability and classification results also indicate that the CCFS approach offers classification capabilities comparable to those offered by the MTI bands when high levels of noise are present. When the SNR increases to 30dB, as shown in Fig. 3.4, the classification results corresponding to the MTI bands almost reach the noiseless case classification error, as shown in Fig. 3.2 (left); however, this trend is much slower in the case of CCFS. The results suggest that the bands designed via the CCFS approach are still more susceptible to noise, compared to the MTI bands. Such a conclusion should not be surprising in view of the fact that the MTI sensor contains well-separated spectral bands with almost non-overlapping finite supports and distinct spectral characteristics. As a result, even for high noise levels, the photocurrents obtained with MTI bands are often well separated.

3.1.4 Comparison with the projection-pursuit approach

We now turn attention to a comparison between the CCFS algorithm and the noise-adjusted version of the PP feature-selection algorithm [16, 48, 66]. In this study we adopt the state-of-the art fast ICA for the implementation of the PP algorithm and its noise-adjusted version [49, 65], which are used as a benchmark.

For low average SNR of 10dB, the separability and classification accuracy achieved with the CCFS algorithm is approximately 10% better than the one obtained with the

noise-adjusted PP. As the SNR increases, the performance of the two algorithms becomes very similar, yielding almost identical separability and classification accuracy in the cases of the average SNR of 20dB and 30dB, as shown in Fig. 3.2 (right) and Fig. 3.4. However, Fig. 3.2 (right) and Fig. 3.4 show that when the SNR reaches extremely high values, the CCFS algorithm once again outperforms the noise-adjusted PP approach, yielding a 10% classification error compared to the 20% error by the noise-adjusted PP for the training set and testing Set-1.

3.1.5 Conclusions

The CCFS algorithm was applied to a QDIP LWIR sensor as a realistic representative of the class of sensors with highly overlapping and noisy bands. As demonstrated by the separability and classification results, in the presence of noise, the CCFS algorithm can effectively reduce the sensor-space dimensionality while maintaining good separability and classification results. The CCFS algorithm outperforms the noise-adjusted PP technique in the cases of low and high SNR. The CCFS algorithm promises robustness to the photocurrent noise by yielding sensing directions with maximal information content and minimized cumulative noise associated with each direction.

3.2 Spectral unmixing and abundance estimation using AHI hyperspectral imagery

Our second case study concerns application of the CCFS to AHI imagery for two classes of problems. In the first problem we investigate the application of the CCFS for supervised Bayesian classification of three spectral classes from the image shown in Fig. 3.5. The second problem uses the CCFS in the context of spectral unmixing

and abundance estimation for three endmembers from the image shown in Fig. 3.6.

AHI is an LWIR pushbroom hyperspectral imager with a 256-by-256 element Rockwell TCM2250 HgCdTe FPA mechanically cooled to 56K [6]. The AHI sensor contains 256 spectral bands in the range 7–11.5 μm, with 0.1 μm spectral resolution for each spectral band. Further details on the AHI system and related data acquisition and calibration issues can be found in [6].

The AHI scene used in the first problem is shown in Fig. 3.5 for $\lambda = 10.0967$ μm, and consists of classes of objects such as roads (R), vegetation (V) and building (B) roofs. The size of this AHI image is 4451 by 256 pixels with 256 spectral bands. To perform supervised classification, we select by visual examination three representative areas for each one of the three classes of interest and use spectral signatures corresponding to these areas as training sets for the classifier. We create the test sets by selecting three areas that represent different spatial locations of the same image but correspond visually to the same classes. The training and testing sets contain 1250 pixels each, 450 pixels per class. The three sections of the scene, shown in Fig. 3.5 for $\lambda = 10.0967$ μm, represent the three classes of interest; these approximate regions are used to extract the training (left) and testing (right) sets.

After the training and testing spectral sets are determined, we apply Bayesian classification, in conjunction with the CCFS, to both sets and calculate separability and classification errors for different SNR cases. AHI spectral bands are approximated uniformly by triangular pulses with peaks at the central frequencies and base widths of 0.1 μm. Similarly to the rock-type classification problem described earlier, we consider four average SNRs in the range 10 to 60dB. For each SNR we determine sets of three superposition bands and then apply them to the spectral content of each pixel in the training and testing regions as shown in Fig. 3.5.

We also examined the application of the CCFS to the spectral unmixing and

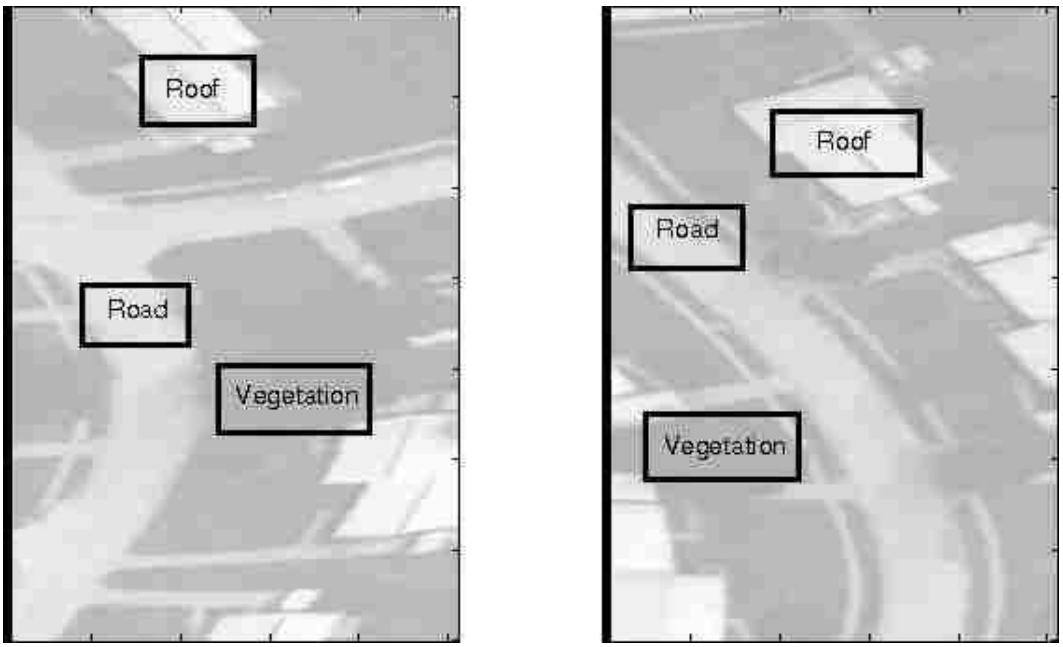


Figure 3.5: Training (left) and testing (right) areas selected from AHI test-flight image of an urban area at 10.0967 μm. The rectangular boxes indicate the approximate areas used to select the training and testing sets for the endmembers.

abundance estimation problem of AHI imagery. The scene used in this study is a different AHI test-flight image, sections of which are shown in Fig. 3.6. The scene represents a snapshot of an urban area at $\lambda = 7.8267$ μm and contains different classes of objects such as buildings, roads, vegetation, parking lots and cars.

Spectral unmixing consists of three main stages: feature-extraction, endmembers determination followed by unmixing and fractional abundance estimation. Unmixing methods can generally be classified by the endmember determination process as automatic and interactive; the automatic methods estimate the number of the endmembers, their spectral signatures and abundance patterns using only the mixed data, the mixing model with no *a priori* information about the ground materials and any human intervention [67, 68, 69]. In interactive unmixing, an analyst or expert

Figure 3.6: Segments of AHI test-flight image of an urban area at 7.8267 μm.

chooses the "pure pixels" from the image or the endmember spectra from spectral library and then estimates the fractional abundance patterns of the component materials in the image. For this study we use the interactive method, while following the three stages described above.

First, by means of visual inspection, three main endmember categories, i.e., building-roof (B), roads (R) and vegetation-ground (V) are identified in the scene areas captured in the image in Fig. 3.6. The representative spectral signatures are determined by calculating the mean of each region corresponding to the designated

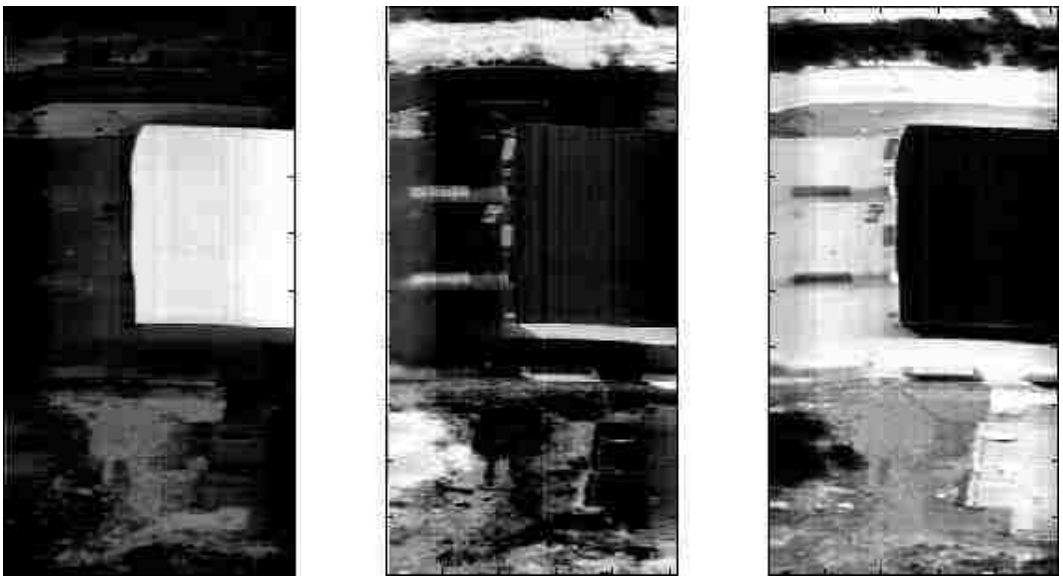


Figure 3.7: Left to right: abundance estimation maps for endmebers building, vegetation and road, respectively, using three uniformly spaced AHI spectral bands in the range 7.7 to 8.6 μm.

endmember category. Endmembers determination is followed by spectral feature-extraction where the CCFS is applied to determine the three most informative directions in the AHI spectral space with respect to the three endmembers and in the presence of noise. The extraction of the three superposition features, one for each endmember, follows the same approach as done in the supervised classification problem described earlier.

3.2.1 Abundance estimation

The last step is to estimate the abundance fraction of each endmember in every pixel from the tested area. Assuming a linear mixing model, the fractions of the

endmembers can be determined by solving the problem of minimizing

$$e = \| \mathbf{x} - \mathbf{Sb} \|^2,$$

where $\mathbf{S}$ is the 3×3 matrix resulting from the application of the CCFS algorithm to the data. The 3 columns of $\mathbf{S}$ correspond to the 3 endmembers and the 3 rows are the superposition features. The vector $\mathbf{x}$ represents the mixed spectrum and $\mathbf{b}$ is the 3×1 fractional abundance vector. Considering the physical meaning of the mixing model, the elements of the abundance vector $\mathbf{b}$ can be subject to two constrains:

$$\sum_{i=1}^{3} b_i = 1 \quad \text{and} \quad b_i \geq 0, \quad i = 1, 2, 3.$$

3.2.2 Results and discussion

To set a benchmark for the performance of the CCFS algorithm in the supervised classification and abundance estimation problems, we first discuss the results in the absence of noise. Bayesian classification results for the three classes of interests (roof, road and vegetation,) for five randomly selected subsets of the AHI spectral bands, show perfect separability and classification. As for the problem of spectral unmixing and abundance estimation, Fig. 3.7 presents the abundance maps of the three endmembers (B, V and R) when using three uniformly separated AHI spectral bands in the range 7.7 to 8.6 μm. The size of the subimage used here is 500 by 256 pixels. It is seen from Fig. 3.7 that each map is able to correctly estimate the fraction of abundance of the corresponding endmember.

Next, we consider the effect of noise and compare the performance of the CCFS algorithm, in supervised classification and spectral unmixing, to that obtained using the noise-adjusted PP. As in the rock-type classification example, 4 different SNR values are considered in the range 10 to 60dB. The search for the three optimal directions for both the CCFS and the noise-adjusted PP is performed over two different

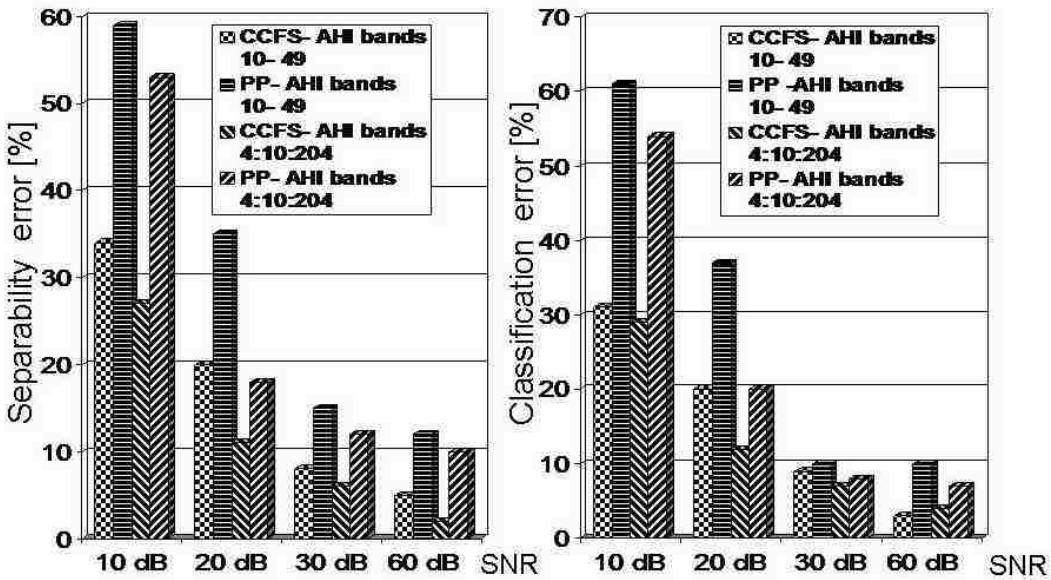


Figure 3.8: Separability (left) and classification (right) results for two subsets of AHI bands and when the CCFS and the noise-adjusted PP are used.

subsets of the AHI bands. The first subset consists of 40 consecutive AHI bands in the range 7.7 μm to 8.6 μm, and the second set consists of 21 uniformly spaced bands in the range 7.7 μm to 11.2 μm.

The separability and classification results for supervised classification of road, roof and vegetation classes, averaged over 50 noise realizations, are presented in Fig. 3.8 for both the CCFS and the noise-adjusted PP algorithms. The performance of the CCFS in this application is consistent with that corresponding to the rock-type classification problem, and it demonstrates good classification in modest SNR scenarios of 10–30dB. Feature-selection from 21 uniformly spaced AHI bands (for both CCFS and noise-adjusted PP) gives better separability and classification than feature-selection from 40 consecutive AHI bands. This results can be explained by the fact that the 40 consecutive AHI bands exhibit higher spectral correlation compared to the 21 uniformly separated bands, and thus, they are potentially more sensitive

to the presence of noise. The noise-adjusted PP shows comparable performance to the CCFS algorithm; however, as in the rock-type classification problem, the CCFS gives improved separability and classification compared to the noise-adjusted PP for the lowest (10dB) and highest (60dB) SNR cases.

We note that for the applications in our study we have observed a very high sensitivity of the performance of the fast ICA implementation of the PP to the initial guess for the projection matrix. In some cases, the classification and separability errors were low; however, in other cases they were much higher than the averaged errors presented in the tables. One possible explanation is that the initialization of the projection matrix by random numbers may not necessarily yield a good initial guess for the hyperspectral data involved.

Figure 3.9 (a-c) shows three groups of fractional abundance maps, for SNR values of 20, 30 and 60dB, respectively, and when the CCFS is applied to 50 consecutive AHI bands in the range 7.7 to 8.6 μm. The corresponding results for the noise-adjusted PP algorithm are shown in Figs. 3.10 (a-b). The size of the subimage used for this problem is 250 by 256 pixels and it represents a subsection of the image shown in Fig. 3.7.

It is seen that the CCFS algorithm shows once again good performance. The CCFS and the noise-adjusted PP perform similarly for the SNR value of 10dB (results not shown). Figures 3.9 (a) and 3.10 (a) compare the abundance maps created using the three CCFS features and the three noise-adjusted PP features, respectively, for the SNR value of 20dB. The maps show improved performance of the CCFS compared to the noise-adjusted PP, which is not able to clearly discriminate between the endmembers of V and R in this SNR case. As expected, the results improve for both CCFS and noise-adjusted PP as the SNR is increased, as shown in Figs. 3.9 (b) and 3.10 (b).

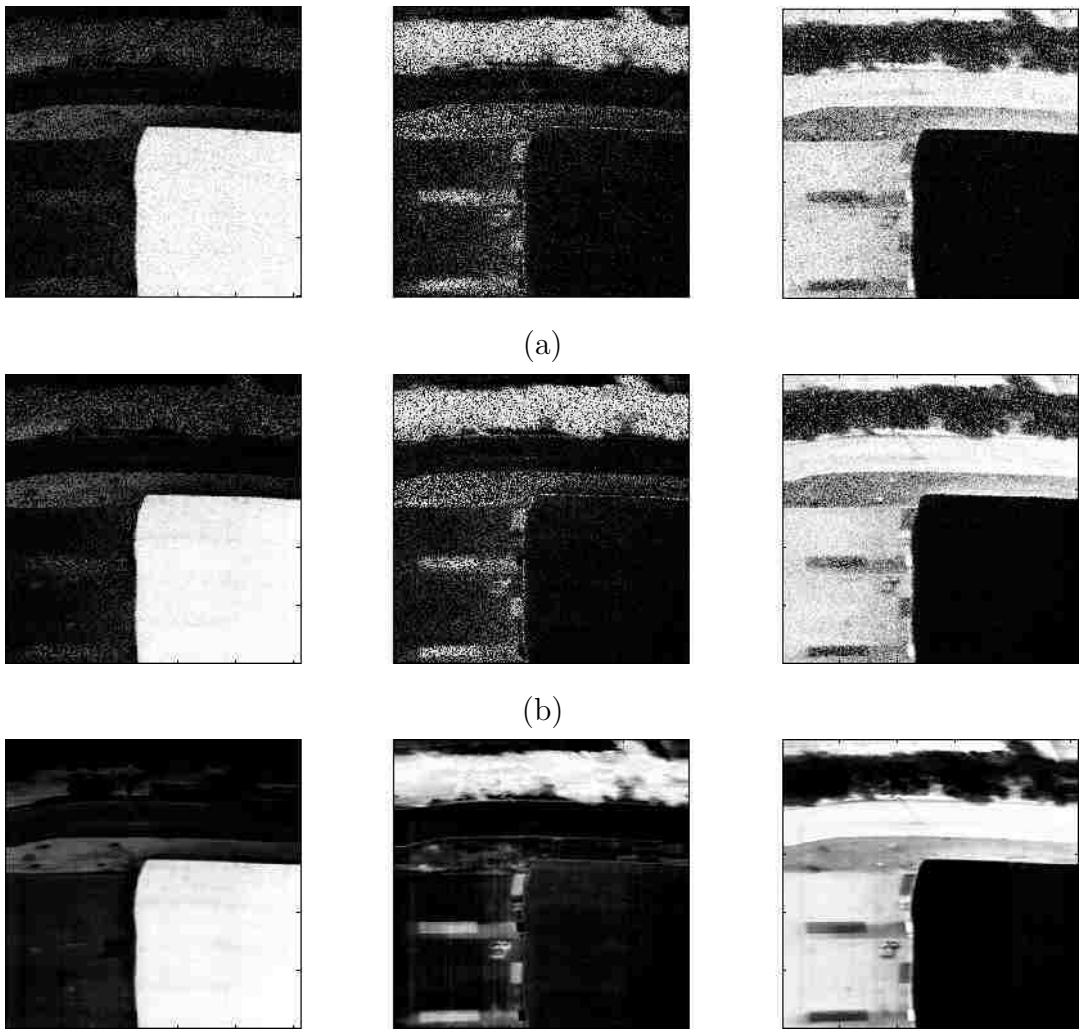


Figure 3.9: Abundance estimation maps for B, V and R endmembers (left to right) using three superposition features selected by the CCFS algorithm from a subset of 50 bands in the range 7.7 μm to 8.6 μm, and for SNR levels of (a) 20dB; (b) 30dB; (c) 60dB.

For the high SNR case of 60dB, we compare the performance of the CCFS described by the abundance maps in Fig. 3.9 (c) to the AHI image in Fig. 3.6 and to the abundance maps shown in Fig. 3.7, representing the noiseless case when three

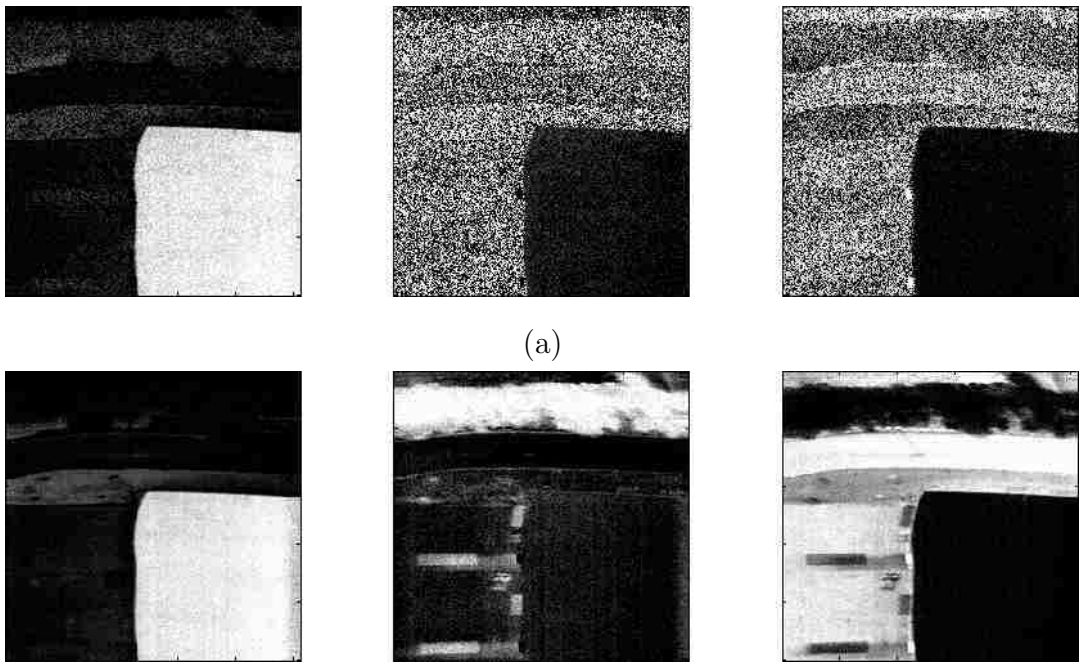


Figure 3.10: Abundance estimation maps for building, vegetation and road endmembers (left to right) using three superposition features selected by the noise-adjusted PP from a subset of 50 bands in the range 7.7 μm to 8.6 μm, and for SNR level of (a) 20dB and (b) 30dB.

AHI bands are used. The results show that at high SNR values, the performance of the CCFS algorithm approaches the noiseless limit.

3.2.3 Conclusions

The following conclusions can be drawn from the examples presented in this section. Our results indicate that the CCFS algorithm offers a noticeable improvement over the noise-adjusted PP algorithm in the cases of low and high SNR. Of course, these improvements come at a price of using numerical optimization procedures to com-

pute the CCFS weights, which is the most expensive step in the CCFS algorithm. However, the cost of the optimization step can be significantly reduced by a judicious choice of the initial guess for the CCFS weights. Our implementation takes advantage of the fact, that in the absence of noise, the optimization algorithm essentially computes the standard orthogonal projection; we therefore choose the coefficients of this projection as an initial guess for the optimization algorithm. In our calculations we have observed that this choice of the initial guess results in substantial reduction in the number of optimization steps needed for convergence.

3.3 Rock-type classification using DWELL FPA imagery

This section presents our third case study. In this study we demonstrate, for a first time, a MS classification capability of a DWELL FPA. To this end, we image a given scene repeatedly using a sequence of bias voltages in the tuning range of the FPA. Then we apply several classification and feature-selection techniques to the totality of readouts, over multiple biases, for each pixel to identify the "class" of the material captured by that pixel. The MS capability is demonstrated for two classification problems: separation among different combinations of three IR filters and discrimination between rocks. Classification is performed using Euclidean- and Mahalanobis-distance classifiers in conjunction with the CCFS algorithm.

The DWELL FPA used in our study is a 320 by 256 detector array, developed and fabricated at CHTM. For details about the fabrication process we refer the readers to [70, 71]. Figure 3.11 shows a representative cartoon of the DWELL FPA.

The DWELL FPA responses have been characterized in [70] by using CamIRa

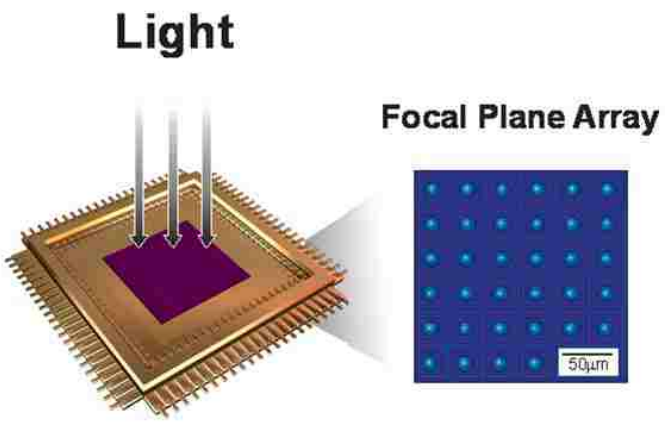


Figure 3.11: An illustration of IR FPA with uniform pixels.

demonstration system[1]. The tuning range of the DWELL FPA is between 0.3 to 1.2 V with an optimal operating temperature[2] of 30 K. The image acquisition laboratory setup at the CHTM used for this study is shown in Fig. 3.12.

3.3.1 Bias tunability of the DWELL FPA

Three different scenes, shown in Fig. 3.13 (a-c), are used to demonstrate the DWELL FPA bias tunability. The first scene, shown in Fig. 3.13 (a), consists of two optical filters with passbands at 4-5 μm and 8.5 μm, termed MW_2 and LW_3, respectively, and a background consisting of a blackbody source. The filters are manufactured by Northumbria Optical Coatings Ltd. The blackbody is manufactured by MIKRON

[1]Manufactured by *SE-IR Corporation*, 87A Santa Felicia Drive, Goleta, CA 93117, USA.

[2]Recently, an optimized DWELL FPA was reported in [70] demonstrating an increase in the operating temperature up to 80 K. The higher operating temperature has been achieved by a strain reduction and an increased number of stacks in the active region improving the responsivity and the absorption quantum efficiency.

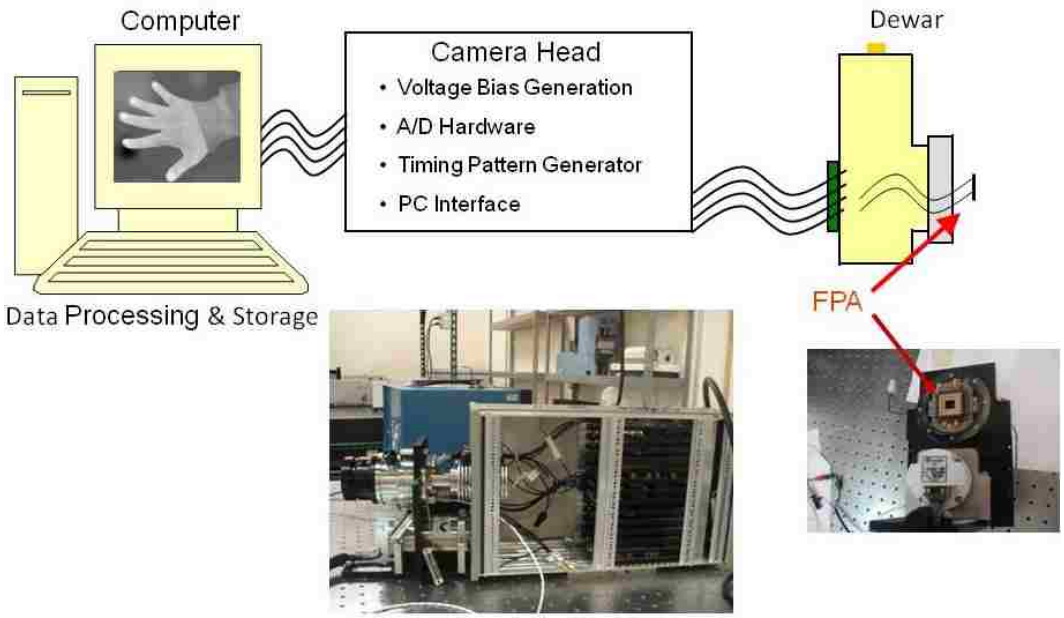


Figure 3.12: An illustration of the laboratory setup at the CHTM for image acquisition with the DWELL FPA. Diagram is curtesy of Woo-Yong Jang.

company (model M315) and provids a temperature between ambient 5^oC and 350^oC, a control to within 0.2^oC, and an emissivity of +0.99.

The scene in Fig. 3.13 (b) includes granite (G), limestone (L) and the MW_2 filter. The scene in Fig. 3.13 (c) includes granite and hornfels (H). All images shown in Fig. 3.13 (a-c) are taken at 0.6 V. The left column of images in Fig. 3.13 (a-c) corresponds to raw data. Normalized images at 0.6 V are shown on the right column in Fig. 3.13, (a-c). The DWELL FPA data is normalized at each pixel by the approximate area of the multi-bias pixel response prior to the spectral ratio calculations in order to eliminate the intensity effect. More details about the normalization are given in the sequel.

Figure 3.14 and Fig. 3.15, left and right, show plots of the spectral ratios for pairs of sensed materials as a function of the applied bias. Figure 3.14 shows the spectral

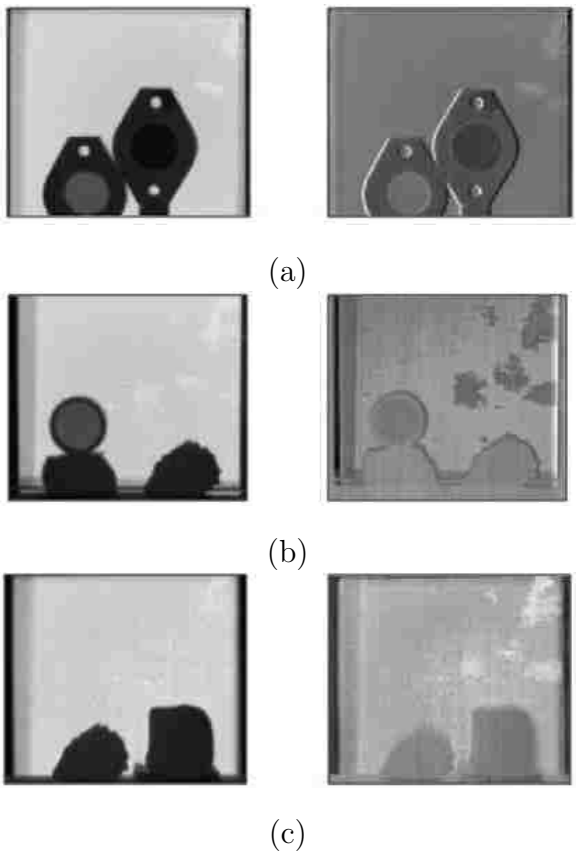


Figure 3.13: DWELL FPA images, at 0.6 V. Left column shows raw imagery and right column shows normalized imagery; (a) filters MW_2 (left) and LW_3 (right); (b) MW_2 (top), limestone (left), granite (right); (c) granite (left), hornfels (right).

ratios calculated for pairs of materials from the scene in Fig. 3.13 (a). The spectral ratios vary between 0.4 to almost 1.4 when the applied bias changes in the range from 0.3 V to 1.2 V with a step of 0.1 V. The fact that the ratio values change from one bias to another indicates the DWELL FPA can sense different spectral contents of the targets observed in a scene simply by changing the applied bias.

The widest spectral ratio range is observed between the filter LW_3 and the background and the narrowest range is observed between the filter LW_3 and the metal holder. These results are not surprising since the filter LW_3 shown in Fig. 3.16 (a)

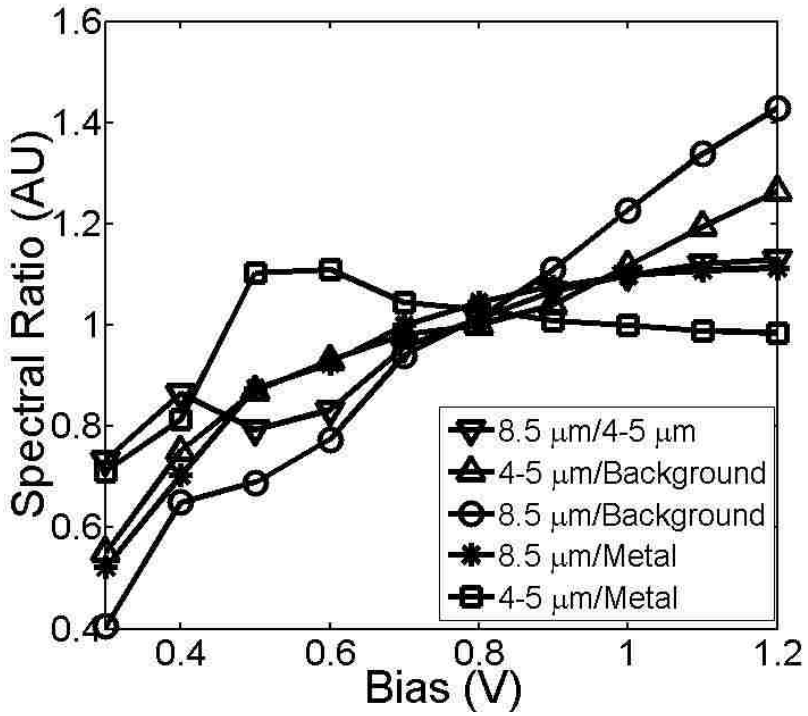


Figure 3.14: Ratios of pixel values for various pairs of the objects MW_2, LW_3, metal holder and the background, as a function of the DWELL FPA bias.

has a passband that is far away from the peak response of the DWELL FPA, and thus, it transmits a very small portion of the background. On the other hand, the metal holder is a solid object that does not transmit light and its spectral response is expected to be quite similar to that of the LW_3 filter response.

Figure 3.15, left, shows the pairwise spectral ratio plots between the granite, limestone, and the background. Figure 3.15, right, shows the pairwise spectral ratio plots between the granite, hornfels, and the background. As observed from the plots, the ratios between granite and limestone, and between granite and hornfels do not exhibit wide range as, for example, the granite-background ratio or the limestone-background ratio. Note in Fig. 3.16, right, that all three rock types have similar

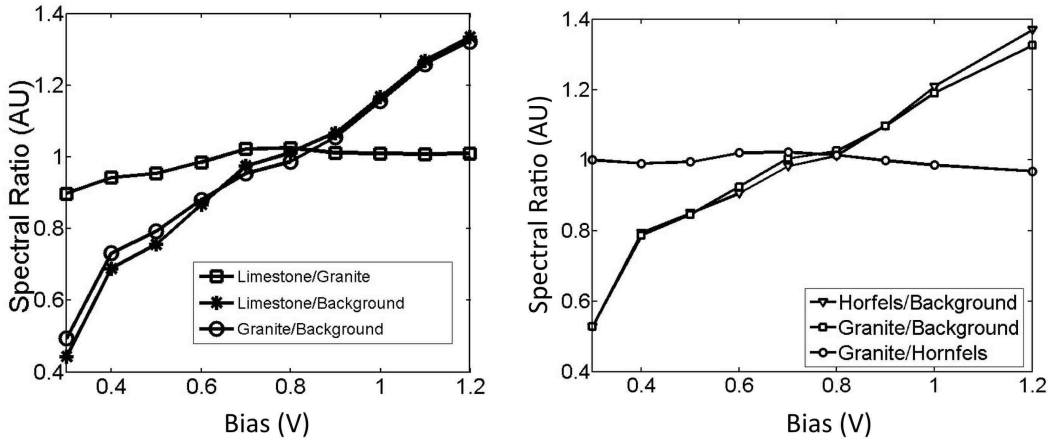


Figure 3.15: Left: ratio of pixel values for the pairs granite-limestone, granite-background and limestone-background, as a function of the applied DWELL FPA bias. Right: ratio of pixel values for the pairs granite-hornfels, granite-background and hornfels-background, as a function of the DWELL FPA bias.

spectra in the 4-8 μm range. The spectral ratios between the granite-limestone and granite-hornfels however, show some variations with respect to the applied bias. The classification results presented bellow also demonstrate that the spectral contrast captured by the bias-tunable DWELL FPA is sufficient to discriminate between the rocks.

3.3.2 Classification problems

The first classification problem considered in this section is that of separating between MW_2 and LW_3 spectral filters and the metal holder. For this problem we consider the scenes shown in Fig. 3.13 (a). The second classification problem is to discriminate between pairs of rocks drawn from the set of the three distinct rock types: granite, hornfels and limestone. The scene configurations for this problem are shown in

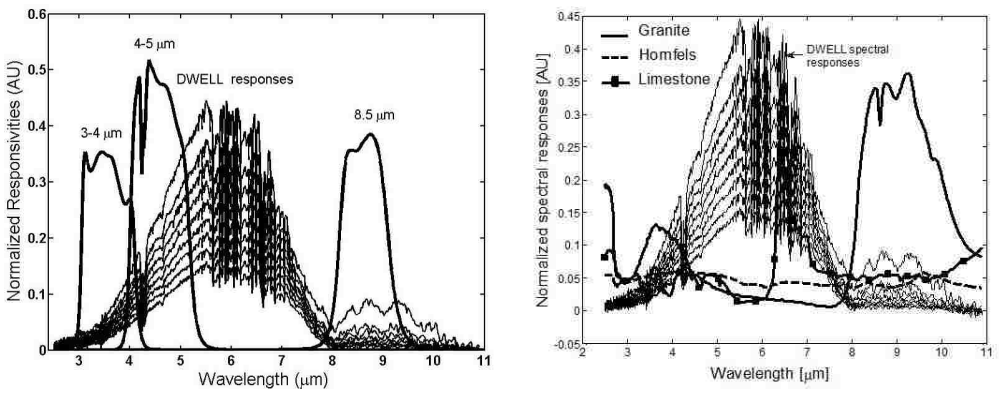


Figure 3.16: Left: normalized spectral responses of a single-pixel DWELL at a bias range of 0.3 V to 1.2 V, and the normalized spectra of the three filters: MW_1, MW_2 and LW_3. Right: normalized spectral responses of a single-pixel DWELL at a bias range of 0.3 V to 1.2 V, and the normalized spectra of the three rocks: granite, hornfels and limestone.

Filter classification	**Identified classes**
Scene (a)	MW_2 and LW_3 filters, metal and background
Rock classification	**Identified classes**
Scene (b)	MW_2 filter, limestone, granite and background
Scene (c)	granite, hornfels and background

Table 3.4: Summary of identified classes for the filter and rock classification problems.

Fig. 3.13 (b-c) . The classes identified for both classification problems are summarized in Table 3.4.

Two types of normalization techniques are applied to the raw digital numbers (DNs) that are retrieved directly as an output of the DWELL FPA. First, as an integrated part of the image acquisition process, at each bias voltage, pixel's DN

values are radiometrically corrected by a two-point nonuniformity correction (NUC) algorithm to compensate for the spatially nonuniform response of the detectors within the FPA [72]. The two-point nonuniformity correction is performed using blackbody temperatures at 22° C and 150° C.

Next, for every radiometrically corrected pixel and its replicas at each bias voltage, the pixel's value is normalized as follows:

$$\underline{I}_j = \frac{I_j}{\Delta v \sum_{i=1}^{n} I_i}. \tag{3.2}$$

This is equivalent to normalization by the area enclosed under the multi-bias response of each pixel in the DWELL FPA. The normalized multi-bias response of a pixel can then be written as:

$$\underline{\mathbf{I}} = (\underline{I}_1, \ldots, \underline{I}_n). \tag{3.3}$$

This normalization minimizes the role of broadband emissivity in the discrimination process and emphasizes the spectral contrast. The normalized images at 0.6 V for both classification problems are shown on the right columns in Fig. 3.13, (a-c).

For both classification problems we perform a supervised classification comprising of training and testing steps. To determine representative multi-bias signatures for each class listed in Table 3.4 we follow the same approach as used in [12]. Specifically, for each class we compute statistical means and covariance matrices using spatially uniform regions that are visually associated with that class. Subsequently, Euclidean- and Mahalanobis-distance classifiers are trained by the classes' mean multi-bias signatures and the covariance matrices.

At the testing step, the trained classifiers are used to classify the objects in Table 3.4 from a set of testing scenes. These scenes capture the same images as the training scenes but were acquired at different times. As a result, the testing

	Number of pixels (training set)	**Number of pixels (testing set)**
Scene (a)	MW_2: 154, LW_3: 108, metal: 126	MW_2: 330, LW_3 : 320, metal: 260
Scene (b)	G: 340, L: 360, MW_2: 360	G: 420, L: 450, MW_2: 300
Scene (c)	G: 224, H: 308	G: 525, H: 870

Table 3.5: The number of pixels in the training and testing data sets for the filter and rock classification problems.

scenes carry inherent variability in the data due to the difference in the measurement conditions from day-to-day and the presence of ambient and system noise. The testing images are normalized in the same fashion as the training images. The size of training and testing data sets for the filter and rock classification problems are listed in Table 3.5.

3.3.3 MS classification results

The thematic maps for the filter and rock classification problems using Euclidean-distance classifier are presented in Figures 3.17–3.20. These maps show the distribution of the derived classes over the spatial area captured by the DWELL FPA. Each map defines a partitioning of the area into sets, each including the points with identical class labels. In order to investigate the effect of the bias selection on the classification accuracy, the classification is performed for multiple combinations of biases.

The thematic maps for the filter classification problem, specified in Table 3.4, are shown in Fig. 3.17. The calculated average classification errors per class are summarized in Table 3.6. The results in Fig. 3.17 are obtained using four different

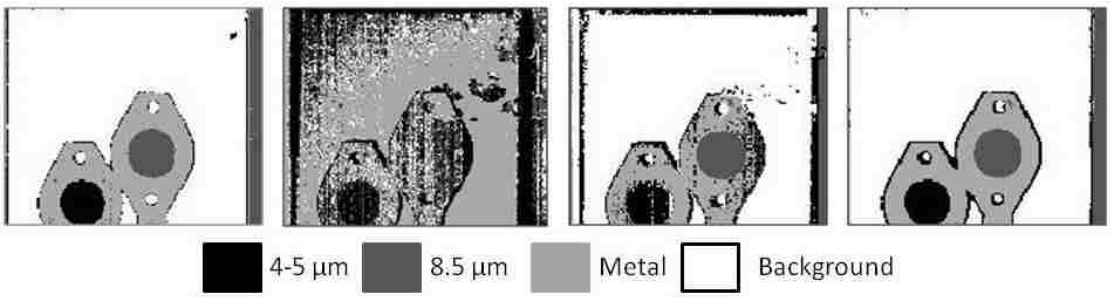


Figure 3.17: Thematic maps for the filter classification problem: left to right: (i) one bias at 0.3 V; (ii) one bias at 0.7 V; (iii) two biases at 0.6 V and 0.7 V; (iv) all ten biases in the range of 0.3 V to 1.2 V, step 0.1 V.

sets of bias voltages, shown from left to right as follows: (i) one bias at 0.3 V; (ii) one bias at 0.7 V; (iii) two biases at 0.6 V and 0.7 V; (iv) all ten biases in the range of 0.3 V to 1.2 V.

For the first bias voltage set the Euclidean-distance classifier consistently shows good classification. In contrast, for the second bias voltage set the Euclidean-distance

Bias (V)	**MW_2** Error [%]	**LW_3** Error [%]	**Metal** Error [%]
0.3	0	0	5
0.7	18	58	46
0.6, 0.7	3	9	7
0.3 – 1.2	0	0	5

Table 3.6: Classification errors for the filter classification problem using the Euclidean-distance classifier. The errors are calculated for the number of pixels defined in Table 3.5, testing sets.

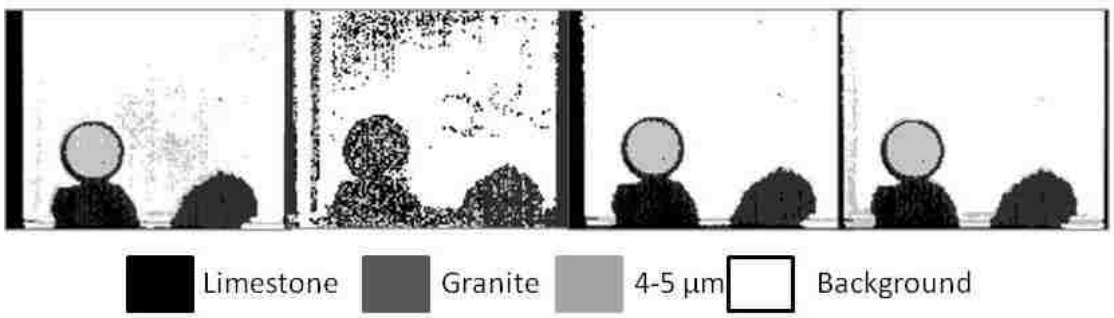


Figure 3.18: Thematic maps for the MW_2-G-L classification problem: (i) one bias at 0.4 V; (ii) one bias at 0.7 V; (iii) two biases at 0.3 and 0.4 V; (iv) all ten biases in the range of 0.3 V to 1.2 V, step 0.1 V.

classifier cannot discriminate successfully between the filters, metal holders and background, as shown by thematic map (ii) in Fig. 3.17. This result indicates that the bias voltage at 0.7 V is not a good choice for this scene. However, adding a second bias voltage at 0.6 V to the second set (resulting in our third bias voltage set) improves the classification, as shown by thematic map (iii) in Fig. 3.17. Finally,

Bias (V)	**MW_2** Error [%]	**Limestone** Error [%]	**Granite** Error [%]
0.4	2.076	29.81	1.91
0.7	62.62	47.26	17.94
0.3, 0.4	0.34	12.77	3.82
0.3 – 1.2	0.34	17.84	1.43

Table 3.7: Classification errors for the MW_2-G-L classification problem using the Euclidean-distance classifier. The errors are calculated for the number of pixels defined in Table 3.5, testing sets.

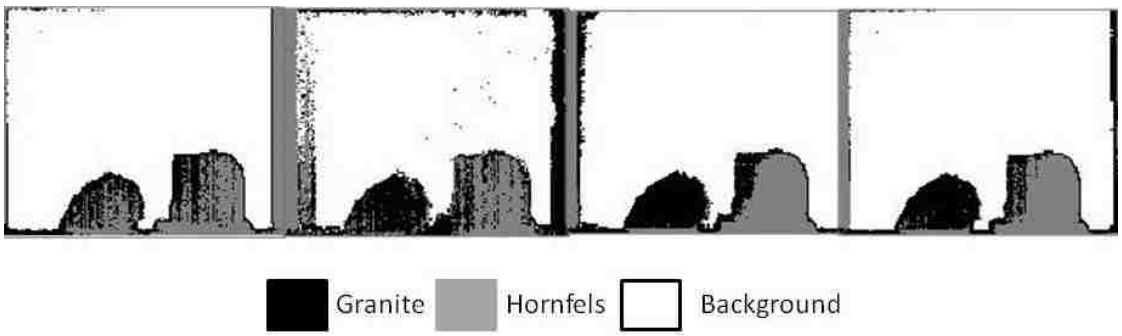


Figure 3.19: Thematic maps for the granite-hornfels classification problem: (i) one bias at 0.3 V; (ii) two biases at 0.6 V and 0.7 V; (iii) one bias at 1.2 V; (iv) all ten biases in the range of 0.3 V to 1.2 V, step 0.1 V.

thematic map (iv) in Fig. 3.17 indicates almost perfect classification results for the fourth set of bias voltages, i.e., when all ten biases are used.

Thematic maps and classification errors for the rock classification problems are shown in Figures 3.18–3.19, and Tables 3.7–3.8, respectively. For the G-L-MW_2

Bias (V)	**Granite** Error [%]	**Hornfels** Error [%]
0.3	55	46
1.1	0	20
0.6, 0.7	5	27
0.3 – 1.2	1	17
CCFS-2 features	1	16

Table 3.8: Classification errors for the granite-hornfels classification problem using the Euclidean-distance classifier. The errors are calculated for the number of pixels defined in Table 3.5, testing sets.

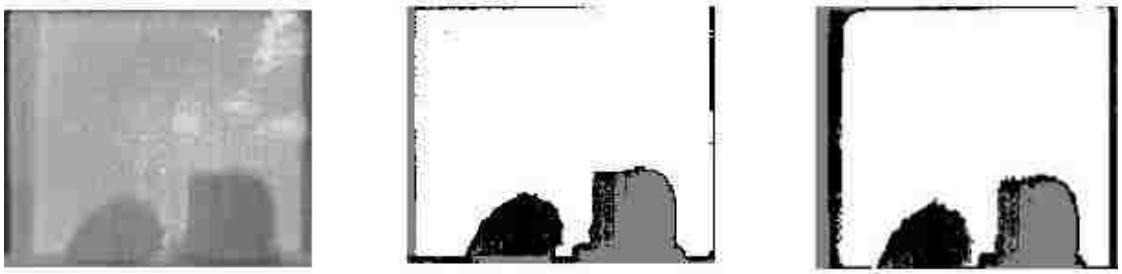


Figure 3.20: The leftmost plot shows the normalized image of granite (left) and hornfels (right) at 0.6 V. Middle: thematic maps for granite-hornfels classification problem when all biases in the range of 0.3 to 1.2 V, with step 0.1 V, are used. Right: thematic maps for granite-hornfels classification problem when two superposition bands obtained by the CCFS are used.

classification problem we use four different sets of bias voltages defined as follows: (i) one bias at 0.4 V; (ii) one bias at 0.7 V; (iii) two biases at 0.3 and 0.4 V; and (iv) all ten biases in the range of 0.3 V to 1.2 V. The first and the second thematic maps in Fig. 3.18 show that the first bias voltage set gives more accurate results than the second one, i.e., bias at 0.4 V is more effective for this scene content than the bias at 0.7 V. Using the third bias-voltage set, which combines two biases at 0.3 V and 0.4 V, improves the classification accuracy compared to the first two cases (the third thematic map in Fig. 3.18.) Moreover, from the fourth thematic map in Fig. 3.18 we see that the third bias set gives results comparable to those using the fourth bias set, i.e., when all ten DWELL FPA bands are used.

The thematic maps between granite and hornfels are presented in Fig. 3.19. The four different bias voltage sets used for this classification problem are as follows: (i) one bias at 0.3 V; (ii) two biases at 0.6 V and 0.7 V; (iii) one bias at 1.2 V; and (iv) all ten biases in the range of 0.3 V to 1.2 V. In contrast to the filter classification problem, where the bias at 0.3 V led to almost perfect classification results, we see

that in the present setting the same bias voltage results in significant classification errors between the two rocks. Nonetheless, the third thematic map in Fig. 3.20 shows that accurate classification for the G-H classification problem by a single bias voltage is still possible, but for a different value at 1.2 V. Interestingly enough, when using the second set of bias voltages (0.6 V and 0.7 V) we obtain classification results whose accuracy falls between the two previous cases: it is better than that with 0.3 V but worse than the accuracy achieved with 1.2 V. As expected, the fourth bias voltage set, i.e., all ten DWELL FPA bands yields almost perfect classification results.

To summarize, our results for the filters, G-L-MW_2 and G-H classification problems demonstrate that accurate classification can be achieved by either considering a broader range of spectral information, namely by using all bias voltages, or by using specific biases, or combination thereof. However, as our results show, the optimal sub-selection of the bias range depends on the specific classification problem. Our next results show that this ambiguity can be reduced by using the CCFS algorithm to select optimal subsets of bands for the granite-hornfels (G-H) classification problem.

We perform classification for this problem using two superposition CCFS bands in conjunction with the Euclidean-distance classifier. The first superposition band is optimized with respect to granite and the second is optimized with respect to hornfels. The corresponding thematic map is shown in Fig. 3.20, right. Comparison between the two thematic maps in Fig. 3.20, middle and right, shows that the two bands selected by the CCFS perform almost the same as when all ten bands are used. Moreover, the classification errors presented in Table 3.8 indicate that in general, the two superposition bands give better accuracy than that obtained from two randomly selected bands, for example the combination of 0.6 and 0.7 V.

In the next section we examine how the between-class separability and the classification accuracy depend on the selection of the bias voltages. For this study we use the G-L-MW_2 classification problem.

3.3.4 Separability analysis and bias selection

The idea of using a measure of between-class separability to select spectral bands or features has been widely used in machine learning and computer vision. Let $\boldsymbol{\mu}_{\mathrm{G}} = (\mu_{\mathrm{G}}(v_1), \ldots, \mu_{\mathrm{G}}(v_m))$ and $\boldsymbol{\mu}_{\mathrm{L}} = (\mu_{\mathrm{L}}(v_1), \ldots, \mu_{\mathrm{L}}(v_m))$ denote the means of class granite and limestone, respectively, for given biases $v_1, \ldots, v_m$.

We define the *normalized separability* between the two rock types at bias voltage v_i as follows:

$$S_{v_i} = \frac{|\mu_{\mathrm{G}}(v_i) - \mu_{\mathrm{L}}(v_i)|}{\|\boldsymbol{\mu}_{\mathrm{G}} - \boldsymbol{\mu}_{\mathrm{L}}\|_2}, \tag{3.4}$$

where $|\mu_{\mathrm{G}}(v_i) - \mu_{\mathrm{L}}(v_i)|$ is the distance between the means of the classes granite and limestone, respectively, when only bias voltage v_i is applied, and $\|\boldsymbol{\mu}_{\mathrm{G}} - \boldsymbol{\mu}_{\mathrm{L}}\|_2$ denotes the Euclidean-distance between the (vector) mean of classes granite and limestone when all biases are used. The normalized separability metric provides information about the contribution of the individual biases to the overall separability achieved when all bias voltages are used.

For bias voltages in the range of 0.3 to 0.5 V the normalized separability between the granite and limestone is in the range of $40-50\%$, as shown in Fig. 3.21 (a). This means that bands at 0.3, 0.4 or 0.5 V contribute almost half of the total separability between the two rocks. However, at 0.6 V the normalized separability drops to approximately 16%. In the range of 0.9 V to 1.1 V the individual band's contributions are all bellow 20%. Figure 3.21 (b) shows the average classification error between granite, limestone and the MW_2 from the scene in Fig. 3.13 (b), as a function of the applied bias. The average classification error is calculated by averaging the number of misclassified pixels for the classes of interest over the number of tested pixels per class and over the number of classes.

Comparison between the results presented in Fig. 3.21 (a) and (b) demonstrates that in general bias voltages that exhibit higher contribution to the overall separa-

bility lead to lower classification error. For example, in the range of 0.3 to 0.5 V, for all bands that are characterized by high granite-limestone separability, the averaged classification error is between 5 to 12%. The bias at 0.7 V, characterized by lower contribution to the overall separability, leads to highest classification error of 42%. For the range of 0.9 V to 1.2 V, where the bands exhibit relatively low contribution to the overall separability, the classification error varies between 10 to 15%.

Figure 3.22 (a) shows the progression in the normalized separability between granite and limestone as bias voltages are added one by one in an increasing order. In reference to the normalized separability calculated as described by (3.4), let

$$V = \{v_1, \ldots, v_n\}$$

denote the set of all bias voltages and

$$\alpha = i_1, i_2, \ldots, i_k\,, \quad 1 < k \leq 10$$

be a multi-index where $1 \leq i_m \leq n$. We define the subset V_α of V as follows:

$$V_\alpha = \{v_{i_1}, \ldots, v_{i_k}\}\,.$$

The progression of the normalized separability as a function of the number of bias voltages can now be re-cast in the following form:

$$S_{V_\alpha} = \frac{\|\boldsymbol{\mu}_G(V_\alpha) - \boldsymbol{\mu}_L(V_\alpha)\|_2}{\|\boldsymbol{\mu}_G - \boldsymbol{\mu}_L\|_2}. \tag{3.5}$$

We observe that the addition of the bias at 0.4 V to the bias at 0.3 V increases the contribution to the total separability (when all biases are used) from 50% to 70%. Furthermore, the addition of the bias at 0.5 V increases the contribution up to 80%. However, note that sequential addition of the biases in the range of 0.6 V to 1.2 V increases the contribution to the total separability only by 20%. This observation is consistent with the results shown in Fig. 3.21 (a). Similarly, Fig. 3.22 (b) shows

the progression of the average classification error for granite, limestone and MW_2 for two classifiers (based upon the Euclidean and Mahalanobis distances) as a function of the number of applied biases.

Two cases are considered. In the first case, the bands are added in sequential order from low bias to high bias, one at a time. As expected, the highest error (20%) is achieved when bias 0.3 and bias 0.4 V are used and the lowest error is achieved when all biases are used. Note that for all bands used, the Mahalanobis-distance classifier gives lower error than the Euclidean-distance classifier. In the second case, the biases are added sequentially in descending order, one at a time. As in the first case, the highest error is achieved when two bias voltages are used (1.2 V and 1.1 V, respectively) and the lowest error is achieved again when all biases are used.

Notably, the error magnitude depends on the order in which the biases are added. Clearly, two DWELL biases at 1.2 V and 1.1 V lead to more than twice the increase in the classification error ($\sim$ 50%) compared to biases at 0.3 V and 0.4 V (18%). The trend is similar up to 5-6 biases used for classification. These results again are consistent with the results presented in Fig. 3.21 (a) and (b) showing that bands corresponding to the bias voltage in the range of 0.3 to 0.5 V give better performance in terms of separability and accuracy of the classification between the three objects: granite, limestone and the filter.

Table 3.9 presents the results of an exhaustive search for optimal bias and for the optimal combinations of biases, in the range of two to ten, as a function of minimizing the average error between granite, limestone and the MW_2 filter for the Mahalanobis-distance classifier. The overall trend in the results presented in Table 3.9 demonstrates that, as the number of biases included in the optimal combination increases, the classification error decreases. For example, the optimal combination of two bias voltages gives a classification error of approximately 6%, while using all biases leads to an error of less than 1%. Note however, that optimal combination of five and

Mahalanobis distance Error [%]	**Biases (V)**
5.83	0.3, 1.2
1.16	0.8, 0.9, 1.2
0.36	0.6, 0.8, 0.9 1.2
0	0.3, 0.6, 0.8, 0.9, 1.2
0	0.3, 0.6, 0.7, 0.8, 0.9, 1.2
0.08	0.3, 0.6, 0.7, 0.8, 0.9,1.1, 1.2
0.12	0.3, 0.4, 0.5, 0.7, 0.8, 0.9, 1.1, 1.2
0.35	0.3, 0.5, 0.6, 0.7, 0.8, 0.9,1.0, 1.1, 1.2
0.34	0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2

Table 3.9: Combinations of biases that minimize the error in the Mahalanobis-distance classifier for the MW_2-G-L classification problem.

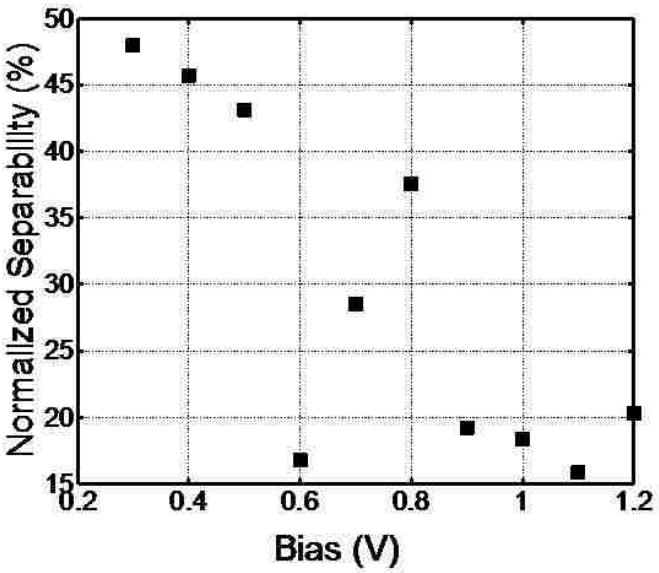
six biases gives almost the same classification error as the case when all biases are used, thus optimized subsets of five or six biases are sufficient to achieve perfect classification for this problem.

3.3.5 Conclusions

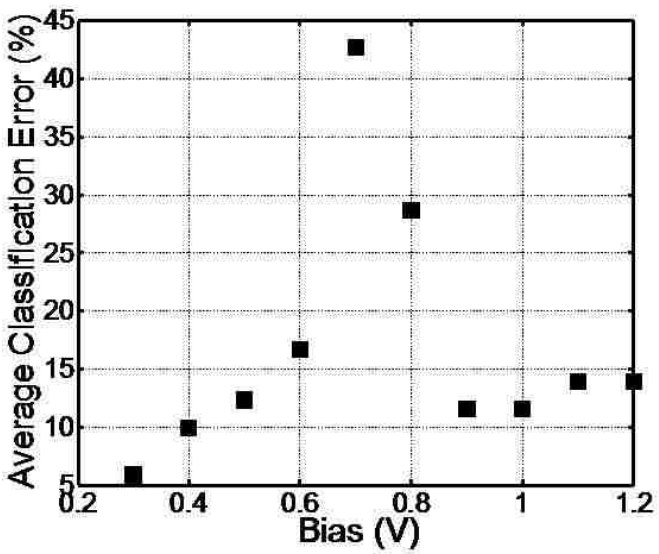
In this section we have demonstrated for the first time the MS-based classification of the DWELL FPA by exploiting the DWELL's bias tunability along with traditional and customized algorithms. The DWELL FPA performance has been validated using two classification problems: (1) separation between three IR spectral filters and (2) discrimination among two pairs of rocks and a filter. The second classification problem is more challenging than the first one as the rocks exhibit lower overall

spectral contrast within the tuning range of the DWELL FPA.

Our verification studies with the DWELL FPA imagery allow us to draw several conclusions. First, the studies show that, as a result of its bias tunability, the DWELL FPA can successfully capture spectral contrast between different materials, which, in turn, enables their accurate classification. Second, the results from the separability and classification analysis for optimal bias selection in both problems demonstrate that accurate classification can be achieved by either considering a broader range of spectral information, i.e., by using all bias voltages, or by using specific biases, or combination thereof. Our results also indicate that the sub-selection of the bias range depends on the classification problem. As expected, the selection of biases varies from case to case. Finally, a customized feature-selection algorithms that specifically addresses the abundant spectral overlap and noise in the DWELL bands, such as the CCFS, can additionally enhance the MS capability of the DWELL FPA by selecting only few optimized superposition bands that yield the same classification results as when using all DWELL FPA bands.

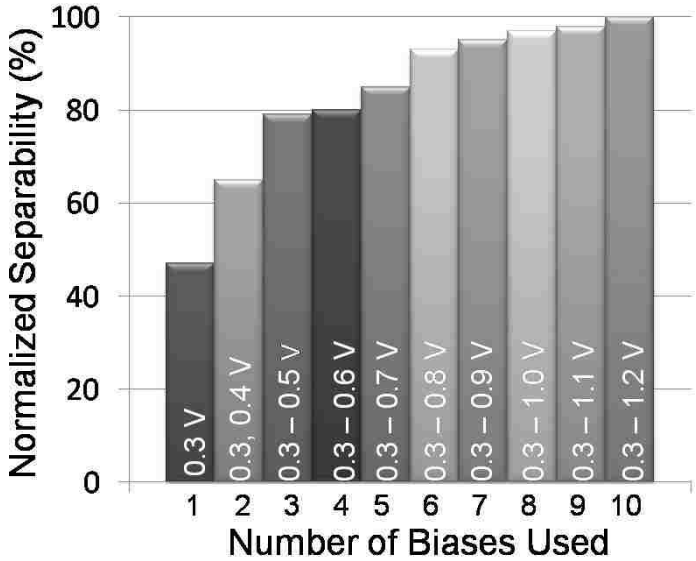


(a)

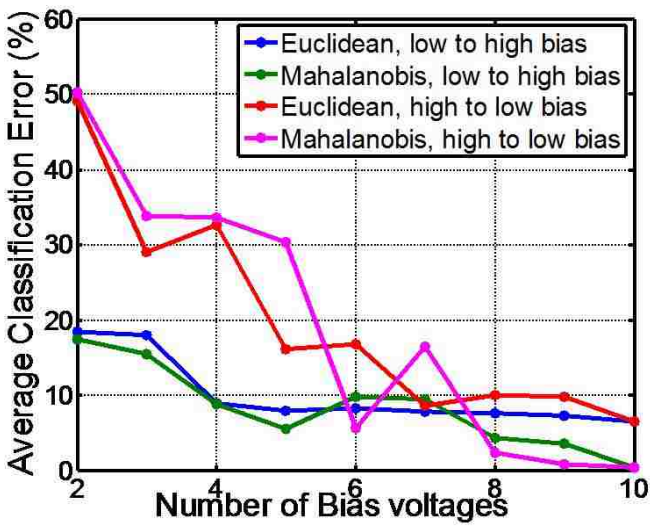


(b)

Figure 3.21: (a) Separability between granite and limestone for each individual bias used; (b) Classification error between granite, limestone and filter as a function of each individual bias used.



(a)



(b)

Figure 3.22: (a) Normalized separability between granite and limestone when the biases (bands) are added sequentially in an increasing order; (b) Average classification error between granite, limestone and filter when the biases (bands) are added sequentially in both an increasing and a decreasing order.

Chapter 4

Spatio-Spectral Feature Selection

This chapter extends the canonical-correlation feature-selection (CCFS) approach, developed in Chapter 2, to a collective spatio-spectral feature-selection and classification framework for hyperspectral (HS) and multispectral (MS) imagers. The main idea is to integrate the spatial and the spectral information in a way that enhances the canonical spectral features exposed by the CCFS, using the spatial content of the hyperspectral imagery. To this end we assume *separability* of the hyperspectral image, i.e, that the spatial content of the image is independent of the spectral bands, but can be used to enhance the latter.

The extension of the CCFS algorithm in this Chapter utilizes the concept of *spectrally enhanced spatial features*, which are obtained by using both the spectral and the spatial contents of an HS/MS image. The resulting Canonical Correlation Spatio-Spectral Feature Selection (CCSS-FS) method consists of two distinct stages: a spatially independent spectral feature selection, based on the CCFS, followed by spatially enhanced classification.

The performance of the CCSS-FS framework is tested on classification applications using remotely sensed imagery collected by the Airborne Hyperspectral Imager

(AHI) [6].

We have organized this chapter as follows. In Section 4.1 we provide the necessary background and review the relevant work. The extension of the CCFS algorithm is developed in Section 4.2, and in Section 4.4 we present the verification studies of the extended algorithm. Discussion of separability of the spatial and spectral information content in MS/HS images is given in Section 4.3. We close the Chapter with a summary of our conclusions in Section 4.5.

4.1 Background and overview of relevant work

HS and MS images can be viewed as a three-dimensional array of real numbers, or a hypercube, whose elements are called pixels. Throughout this dissertation we employ the following formal notation for HS and MS images:

$$\mathbf{u} \in \mathbb{R}^{I \times J \times K}, \quad \mathbf{u} = \{u_k(i,j) \in \mathbb{R} \,|\, 1 \leq i \leq I,\ 1 \leq j \leq J,\ 1 \leq k \leq K\}\,. \quad (4.1)$$

The (i, j) plane represents the spatial image or the scene, consisting of pixels with finite resolution, and the third dimension k is the spectral domain, represented by tens or even hundreds of spectral bands, constituting spectral replicas of each pixel.

One of the important characteristics of the data structure represented by (4.1) is the complex nature of the relations between pixels and their neighbors along the i, j and the k directions within the image hypercube. First, due to possible manufacturing limitations, a pixel's spectral replicas are typically correlated, because adjacent bands may have overlapping spectral ranges. Second, pixel replicas from different bands are spatially correlated since they are images of the same object but with respect to different waveranges. Third, due to the scene topology, there is high spatial correlation between the adjacent pixels in each spectral image. Such complex interplay between the spatial and spectral domains calls for approaches that allow

us to take this interplay under consideration during the feature selection and the classification stages.

The integration of the spatial and spectral information for improved classification of HS and MS images is a subject of an active ongoing research. Several different ideas have already been proposed and tested in various detection and classification problems. The most popular approach is to include both spectral and textural information in the classifier via composite kernels. This is usually achieved by means of the so-called "stacked" approach, in which feature vectors are build from concatenation of spectral and spatial features. Benediktsson *at al.* in [73] have proposed the use of extended morphological profiles (EMPs). These profiles rely on classical mathematical morphology operations such as erosion and dilation [74], which are used to process a binary or grayscale image with a set of known shape, called structuring element. These basic operations are used to construct opening and closing operators; with these operators, if the structure of the image cannot be contained in the structuring element then it is totaly removed, otherwise it is preserved. A morphological profile [73] is defined as a composition of the opening and closing profiles, both of them viewed as n dimensional vectors, where n is the number of openings or closings, respectively.

Extension of this concept to HS and MS imagery requires spectral feature selection stage in order to determine few spectral components. In [75] the full spectral information was used to construct the EMPs; however, Benediktsson *et al.* [73] proposed to use only the first few principal components (PCs), and then extend each one of them with the morphological information. The extended profile is a single stacked vector used as an input feature for the classifier.

In [76] Gamba *et al.* propose to combine spatial characterization using Markov Random Fields (MRF) with spectral neuro-fuzzy classification. To this end, Gamba *et al.* start with a spectral classification step performed by a neuro-fuzzy classifier.

The result of this step is then used in an MRF-based spatial analysis stage, which is performed jointly with a maximum likelihood spectral-based reclassification stage. In doing so, the pattern recognition capabilities of the neuro-fuzzy classifier, which has demonstrated excellent performance at a single-pixel level, are combined with the spatio-spectral capabilities of the MRF framework.

The segmentation hierarchies approach [77] is another idea that can be utilized for spatio-spectral analysis of MS and HS imagery. Loosely speaking, a segmentation hierarchy is a set of several segmentations of the same image at different levels of detail. The segmentation at a coarser level can be produced by simply merging the regions at finer levels of detail. In MS and HS imagery, spatially adjacent regions iteratively merge through a specified merge selection process based on spectral criteria. We refer the readers to [77] for further details.

In the light of the above discussion, the objectives of this Chapter are as follows. First, we seek to extend the CCFS algorithm to a collective spatio-spectral feature-selection and classification framework for HS and MS imagers, using the concept of "stacked" or concatenated feature vectors, which represent *spectrally enhanced spatial features*. This approach parallels that of [73] with two key distinctions. Instead of using principal component analysis to select the spectral content, we apply the CCFS algorithm, and instead of using morphological operations to define the spectrally enhanced spatial features, we employ spatial masks such as 'Sobel' [4], 'Median', and 'Variance' [2], to name just few.

Our second objective is to validate the performance of the CCSS-FS framework in classification problems using real HS imagery. To this end we use remotely sensed imagery collected by the AHI. Sensitivity of the spatio-spectral feature-selection approach with respect to the initial set of sensor bands and with respect to the number and types of spatial features utilized during the classification stage is also investigated in this chapter.

4.2 Extension of the CCFS to a spatio-spectral feature selection.

To develop the extension of the CCFS algorithm we utilize the idea of *spectrally enhanced spatial features*. This approach can be viewed as an extended kernel method that deals with the non-stationary nature of the spectral signatures [73]. As already mentioned, the spectrally enhanced spatial features are obtained by considering both the spectral and the spatial information content of each pixel, and can be viewed as "stacked" or concatenated feature vectors.

Typical spatial feature extraction techniques utilize functions, called masks or kernels, that operate on pixel values in a predefined neighborhood $M \times N$ of a pixel with spatial coordinates (i, j) [2, 74]. The kernel's coefficients determine the type of the extracted spatial feature. However, for HS and MS images a spatial feature contains as many spectral components K, as the number of the spectral bands used to obtain the image. Consequently, such spatial features may be thought of as a result of an application of an $M \times N \times K$ dimensional mask to each hyperspectral pixel. We call such spatial feature a *spectrally enhanced spatial feature*. Note that if the kernel represents a simple delta function, the spectrally enhanced spatial feature will be the spectral signature itself. Therefore, a spectral signature can be viewed as the simplest case of a spectrally enhanced spatial feature.

To define the extension of the CCFS algorithm along these lines, we consider a two stage spatio-spectral feature-extraction technique. At the first stage, and for a given classification/detection problem, the optimal superposition bands are determined by applying the CCFS algorithm from Chapter 2 to each class of interest. At the second stage, after the most informative superposition bands have been determined and applied to the spectral imagery, various spatial feature-extraction techniques based on suitable $M \times N$ spatial masks, can be applied based on certain *a priori* information

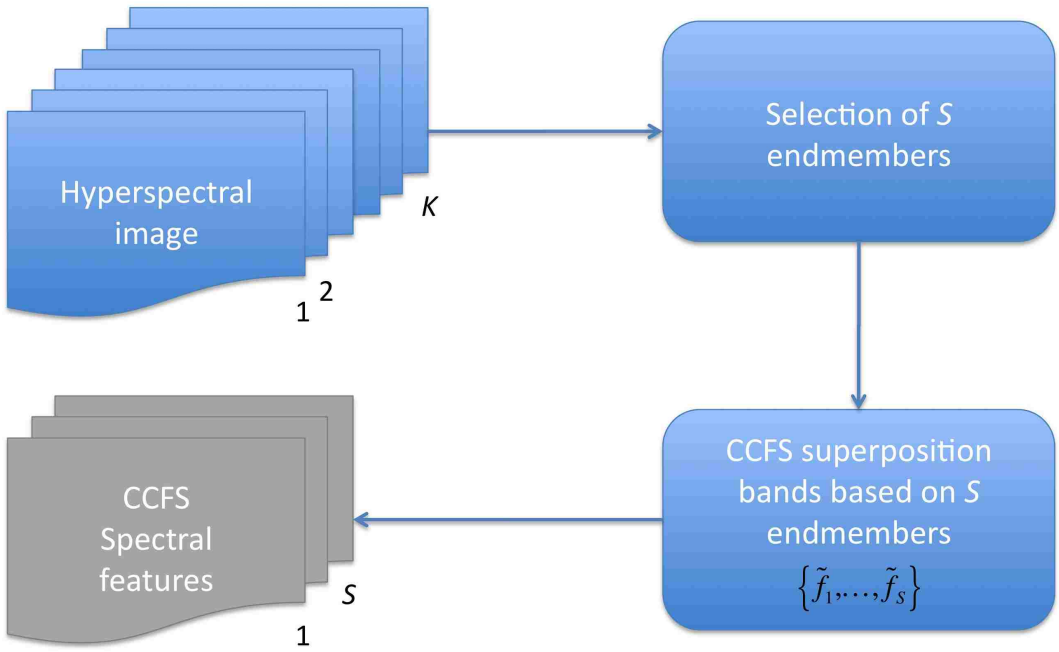


Figure 4.1: The first stage of the CCSS-FS algorithm.

about the scene. A set of spatial features can possibly contain information about edges, points and lines, as well as texture features, and morphological profile features, among others [21]. Moreover, each one of these spatial features is now a spectrally enhanced superposition feature, that contains as many spectral components as the number of the selected superposition bands from the CCFS stage. Thus, an extended pixel profile is created that can be used as an input to the classifier.

A formal description of the CCSS-FS framework is as follows. Assume that $\mathbf{u} \in \mathbb{R}^{I\times J\times K}$ is a MS/HS image, as defined in (4.1). At the first stage of the CCSS-FS we apply the CCFS algorithm from Chapter 2 to $\mathbf{u}$ in order to obtain a set $\{\tilde{f}_1,\ldots,\tilde{f}_S\}$ of superposition bands, where S is the number of classes of interest, or endmembers. The S superposition bands represent the most informative spectral directions for the given set of endmembers, where typically $S \ll K$.

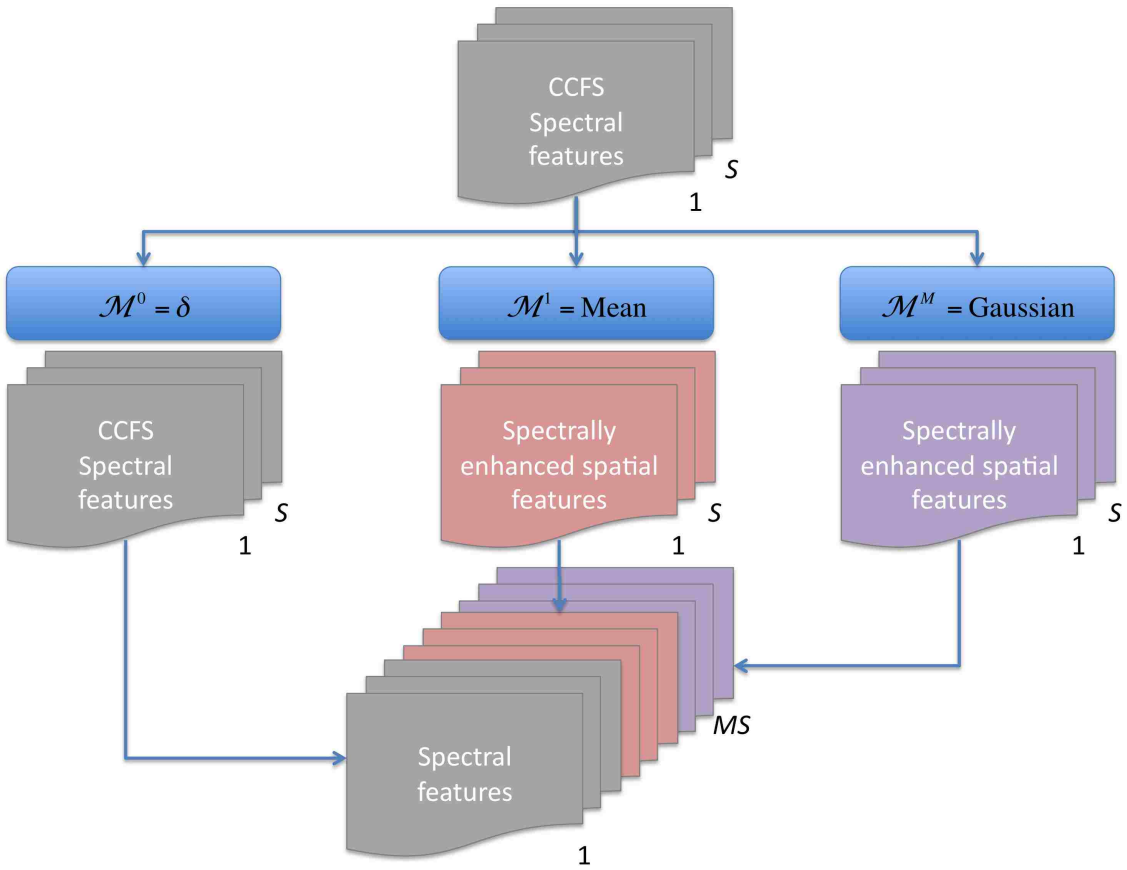


Figure 4.2: The second stage of the CCSS-FS algorithm.

Using the superposition bands, we define a new *reduced* image hypercube

$$\mathbf{u}^0 \in \mathbb{R}^{I \times J \times N}, \mathbf{u}^0 = (\mathbf{u}_1^0, \ldots, \mathbf{u}_S^0)$$

where each image plane $\mathbf{u}_s^0$ is obtained by the application of superposition band $\tilde{f}_s$ to the original image hypercube $\mathbf{u}$. This completes the first stage of the CCSS-FS algorithm, which is shown in Fig. 4.1.

At the second stage we select a set $\{\mathcal{M}^1, \ldots, \mathcal{M}^M\}$ of spatial masks based, for example, on certain *a priori* information about the scene captured in $\mathbf{u}$. For simplicity, we restrict attention to masks with square $N \times N$ regions of operations, whose

response on a gray-scale image u is given by

$$\mathcal{M}^m(i,j) = \sum_{-N \leq n_1, n_2 \leq N} \omega^m(n_1, n_2) u(i + n_1, j + n_2)\,. \tag{4.2}$$

We recall that in (4.2) $u(i,j)$ is the pixel's intensity value at the center $\{i,j\}$ of the mask, $\omega^m(n_1,n_2)$ are the weights at offsets $\{n_1,n_2\}$ relative to the center of the mask, and N is the diameter of $\mathcal{M}^m$.

To define the spectrally enhanced spatial features we first apply each one of the masks $\mathcal{M}^m$ to the reduced image hypercube $\mathbf{u}^0$ to obtain the sequence of its replicas

$$\mathbf{u}^m = \mathcal{M}^m(\mathbf{u}^0)\,, 1 \leq m \leq M\,.$$

To complete the second stage of the CCSS-FS the replicas and the original reduced hypercube are concatenated to obtain another image hypercube

$$\widetilde{\mathbf{u}} = (\mathbf{u}_1^0, \ldots, \mathbf{u}_S^0, \mathbf{u}_1^1, \ldots, \mathbf{u}_S^1, \ldots, \mathbf{u}_1^M, \ldots, \mathbf{u}_S^M)\,, \quad \widetilde{\mathbf{u}} \in \mathbb{R}^{I \times J \times MS}\,. \tag{4.3}$$

The image hypercube in (4.3) redefines the original image hypercube $\mathbf{u}$ in a way that optimizes its spectral and spatial content with respect to the endmembers and the spatial features reflecting the *a priori* knowledge about the scene. The second stage of the CCSS-FS approach is shown schematically in Fig. 4.2.

It is clear that the CCSS-FS algorithm described above and the approaches surveyed in Section 4.1 share the same idea of "stacked" or "concatenated" spectral and spatial features. Implicit in this class of methods is also the assumption that the spatial and the spectral features are essentially separable. We remind that the key distinction between these approaches and the CCSS-FS is in the use of the CCFS algorithm at the first stage of the spatio-spectral feature selection. This means that in the CCSS-FS the spectral features are optimized with respect to a given set of endmembers, as opposed to methods in which the first principal components of the total image hypercube $\mathbf{u}$ are used.

4.3 Separable scene and sensor models

In this Section we examine the notion of separability in MS and HS imagery that is implicit in the definition of the CCSS-FS. For this purpose it is more convenient to work with continuous scene and sensing models. In particular, we assume that the scene can be described by a continuous function of the spatial coordinates x, y and the spectral variable λ:

$$p(x, y, \lambda) \in C^0; \qquad p : X \times Y \times \Lambda \to \mathbb{R}\,. \tag{4.4}$$

In (4.4) X, Y and Λ are closed intervals in $\mathbb{R}$.

A single sensor operation in a focal plane array (FPA) containing $I \times J$ pixels is modeled by a bounded linear functional that can be identified with a continuous function

$$F_{i,j,k}(x, y, \lambda) : X \times Y \times \Lambda \to \mathbb{R}\,, \tag{4.5}$$

where $1 \le i \le I$ and $1 \le j \le J$ refer to the spatial position of the pixel within the FPA and $1 \le k \le K$ refers to the band number of the sensor. Each $F_{i,j,k}$ is assumed to have a compact support:

$$\mathrm{supp}(F_{i,j,k}(x, y, \lambda)) \subset X_i \times Y_j \times \Lambda_k,$$

where

$$X_i = [i-1, i]\Delta x, \quad Y_j = [j-1, j]\Delta y, \quad \Lambda_k = [\underline{k}, \bar{k}]\Delta\lambda\,,$$

and $\Delta x = 1/(I-1)$, $\Delta y = 1/(J-1)$ and $\Delta\lambda = 1/(K-1)$.

A hyperspectral image $\mathbf{u} = \{u_k(i,j)\} \in \mathbb{R}^{I\times J\times K}$ is obtained by integration of the functional $F_{i,j,k}(x, y, \lambda)$ and the scene $p(x, y, \lambda)$:

$$\begin{aligned} u_k(i,j) = \langle F_{i,j,k}, p\rangle \triangleq \int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j,k}(x, y, \lambda)p(x, y, \lambda)dxdyd\lambda\,, \\ 1 \le i \le I, \;\; 1 \le j \le J, \;\; 1 \le k \le K. \end{aligned} \tag{4.6}$$

We make the following definitions:

***Def. 1*:** Scene $p(x, y, \lambda)$ is separable *iff* there exist C^0 functions $q : X \times Y \to \mathbb{R}$ and $r : \Lambda \to \mathbb{R}$ such that $p(x, y, \lambda) = q(x, y)r(\lambda)$.

***Def. 2*:** Sensor $F_{i,j,k}(x, y, \lambda)$ is separable *iff* there exist bounded linear functionals $F_{i,j} : X \times Y \to \mathbb{R}$ and $F_k : \Lambda \to \mathbb{R}$ such that $F_{i,j,k}(x, y, \lambda) = F_{i,j}(x, y)F_k(\lambda)$ for all $1 \leq i \leq I$, $1 \leq j \leq J$ and $1 \leq k \leq K$.

***Def. 3*:** A MS or HS image $\mathbf{u}$ is separable *iff* there exist a a matrix $U \in \mathbb{R}^{I \times J}$ and a vector $V \in \mathbb{R}^K$ such that $\mathbf{u} = U \otimes V$, i.e., $u_k(i, j) = U(i, j)V(k)$ for all $0 \leq i \leq I$, $0 \leq j \leq J$ and $0 \leq k \leq K$.

In general, a hyperspectral image $\mathbf{u}$ is not necessarily separable, even if the sensor itself is separable. This can be seen from the following calculation:

$$\begin{aligned} u_k(i,j) &= \int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j,k}(x,y,\lambda)p(x,y,\lambda)dxdyd\lambda \qquad (4.7)\\ &= \int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j}(x,y)F_k(\lambda)p(x,y,\lambda)dxdyd\lambda \\ &= \int_{X_i}\int_{Y_j} F_{i,j}(x,y)\left(\int_{\Lambda_k} F_k(\lambda)p(x,y,\lambda)d\lambda\right)dxdy \\ &= \int_{X_i}\int_{Y_j} F_{i,j}(x,y)g_k(x,y)dxdy, \end{aligned}$$

where $g_k(x, y) = \int_{\Lambda_k} F_k(\lambda)p(x, y, \lambda)d\lambda$. The function g_k implicitly depends on the given spectral band. As a result, the last integral in (4.7) cannot be represented as a product of two numbers $U(i, j)$ and $V(k)$ that depend only on the spatial and the spectral dimensions, respectively. The following result provides further information about the separability of the hyperspectral image.

Lemma 3 *A MS/HS image* $\mathbf{u}$ *is separable iff both the sensor* $F_{i,j,k}$ *and the scene* $p(x, y, \lambda)$ *are separable.*

Proof. Let us first assume that both $F_{i,j,k}$ and $p(x, y, \lambda)$ are separable. Then

$$\begin{aligned} u_k(i,j) &= \int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j,k}(x,y,\lambda)p(x,y,\lambda)dxdyd\lambda \\ &= \int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j}(x,y)F_k(\lambda)q(x,y)r(\lambda)dxdyd\lambda \\ &= \left(\int_{X_i}\int_{Y_j} F_{i,j}(x,y)q(x,y)dxdy\right)\left(\int_{\Lambda_k} F_k(\lambda)r(\lambda)d\lambda\right) \\ &= U(i,j)\,V(k). \end{aligned}$$

Let us now assume that $\mathbf{u}$ is separable, i.e., $u_k(i, j) = U(i, j)V(k)$ for all i, j, k. Then $U(i, j)$ and $V(k)$ can be represented as:

$$U(i,j) = \int_{X_i}\int_{Y_j} H_{i,j}(x,y)q(x,y)dxdy \tag{4.8}$$

and

$$V(k) = \int_{\Lambda_k} K_k(\lambda)r(\lambda)dxdy \tag{4.9}$$

On the other hand

$$u_k(i,j) = \int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j,k}(x,y,\lambda)p(x,y,\lambda)dxdyd\lambda\,.$$

Thus,

$$\begin{aligned} &\int_{X_i}\int_{Y_j}\int_{\Lambda_k} F_{i,j,k}(x,y,\lambda)p(x,y,\lambda)dxdyd\lambda \\ &\qquad = \int_{X_i}\int_{Y_j} H_{i,j}(x,y)q(x,y)dxdy \int_{\Lambda_k} K_k(\lambda)r(\lambda)dxdy\,. \end{aligned} \tag{4.10}$$

Eq.(4.10) implies that

$$F_{i,j,k}(x,y,\lambda) = H_{i,j}(x,y)K_k(\lambda)\,, \quad p(x,y,\lambda) = q(x,y)r(\lambda). \qquad \square$$

In what follows we examine some conditions on p under which the separable model is a good approximation. Assume that $p(x, y, \lambda) \in C^k$. Then we can expand $p(x, y, \lambda)$ about λ_0 using Taylor series as follows:

$$p(x, y, \lambda) = p(x, y, \lambda_0) + p_\lambda(x, y, \lambda_0)(\lambda - \lambda_0) + p_{\lambda\lambda}(x, y, \lambda_0)\frac{(\lambda - \lambda_0)^2}{2} + O(\Delta\lambda^2)\,.$$

We further assume that $p_{\lambda\lambda}(x, y, \lambda_0)(\lambda - \lambda_0)^2/2$ is small compared to

$$p(x, y, \lambda_0) + p_\lambda(x, y, \lambda_0)(\lambda - \lambda_0).$$

This assumption holds in the cases when $(\lambda - \lambda_0)^2$ is small and when $p_{\lambda\lambda}(x, y, \lambda_0)$ is small. In such a case, the scene can be approximated only by the linear terms in the Taylor expansion as:

$$p(x, y, \lambda) \approx p_0(x, y) + p_1(x, y)(\lambda - \lambda_0)\,.$$

Hence, in such a case $p(x, y, \lambda)$ is approximately separable in the spatial and spectral domains. From this analysis we can conclude that separable images can be thought of as superpositions of a static (bias) scene p_0 and a scene that varies linearly with the spectral frequency λ.

4.4 Applications

In this section we apply the CCSS-FS algorithm to an AHI[1] hyperspectral imagery in the context of supervised classification. First, by means of visual inspection, three main endmember categories of buildings (B), roads (R) and ground/vegetation (G) were identified in the scene area captured by the image in Fig. 4.3. The representative spectral signatures were determined following the approach adopted in [12] by

[1]We remind the reader that the AHI sensor contains 256 spectral bands in the range 7–11.5 μm, with 0.1 μm spectral resolution for each spectral band. Further details on the AHI system and related data acquisition and calibration issues can be found in [6].

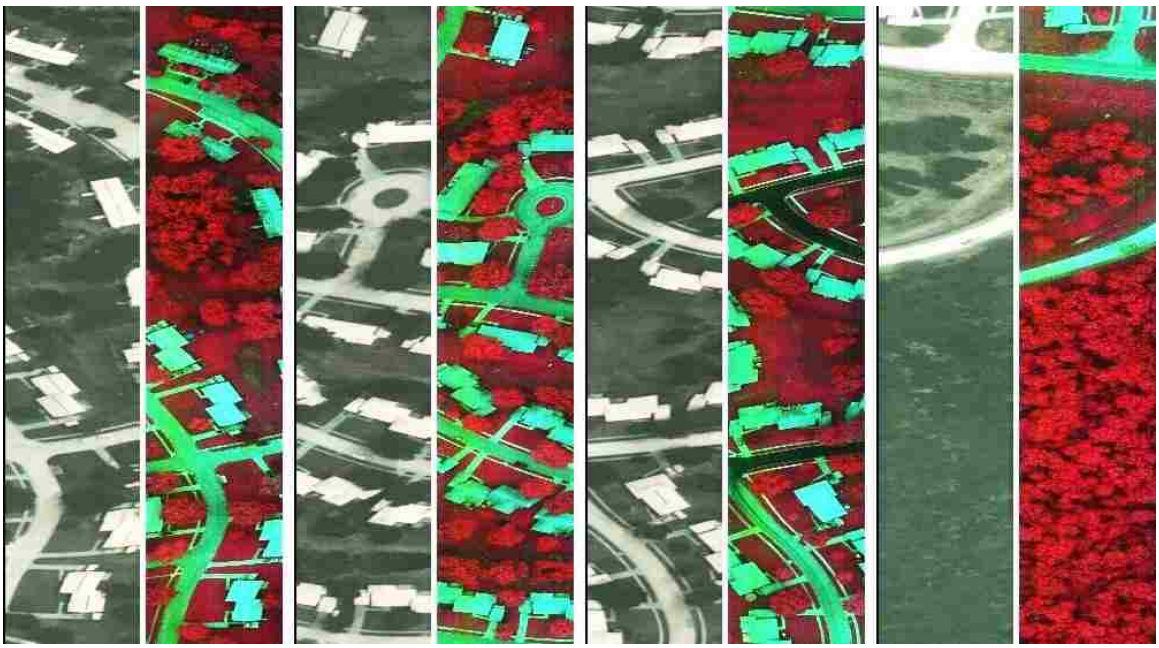


Figure 4.3: The AHI test flight image taken on July 26nd, 2004.

calculating the mean of spatially uniform regions that visually correspond to each designated endmember category. In addition, we selected two representative testing sets, which capture different parts of the scene shown in Fig. 4.3, and which contain all classes of interest. Endmember extraction was followed by spectral feature selection, where the CCFS algorithm from Chapter 2 was used to determine the three most informative directions $\{\widetilde{f}_{\mathrm{B}}, \widetilde{f}_{\mathrm{R}}, \widetilde{f}_{\mathrm{G}}\}$ in the AHI spectral space with respect to the three endmembers, and in the presence of noise. In particular, the performance of the CCSS-FS was investigated for four average signal-to-noise ratios (SNR) values in the range 10 to 60dB.

AHI spectral bands were approximated uniformly by piecewise linear functions with peaks at the center frequencies and base widths of 0.1 μm. After the three superposition bands $\{\widetilde{f}_{\mathrm{B}}, \widetilde{f}_{\mathrm{R}}, \widetilde{f}_{\mathrm{G}}\}$ for each SNR were determined, they were applied to the spectral content of each pixel in the testing data sets. Next, various spatial

feature-extraction techniques were applied to every pixel within the three superposition image planes in order to obtain the spectrally enhances spatial features. We used spatial masks of the form given in (4.2), each operating on 5×5 pixels' areas, i.e., $N = 5$ in (4.2). Specifically, in the second stage of the CCSS-FS algorithm we applied the 'Mean', 'Unsharp', 'Gaussian', 'Laplacian', 'Sobel', 'Log', 'Prewitt', 'Median' and 'Variance' [2] spatial masks.

The so-created spatio-spectral features were then appended to the three canonical features, as shown in Fig. 4.2, for a total of 30 spectrally enhanced spatial features. Different subsets of all 30 spectrally enhanced spatial features for both testing sets were used as an input to an Euclidean-distance classifier. In every testing case, the results of the classification are presented in the form of thematic maps.[2]

Figure 4.4 compares the thematic maps created from the classification results based on different subsets of spectrally enhanced spatial features. For an SNR of 10dB, these results show a noticeable improvement in the classification accuracy when the three superposition features are extended with 'Mean', 'Median' and 'Gaussian' spatial features, as shown in Fig. 4.4 (i) and (v). The investigation of different combinations of spectrally enhanced spatial features has shown that the combination presented in Fig. 4.4 (v) gives consistently, for different SNR values, the best classification results.

Figures 4.5 and 4.6 compare the performance of the CCSS-FS algorithm when the spectral features are selected by the CCFS and the noise-adjusted projection pursuit (NAPP) algorithms [65, 20] for the SNR cases of 10 and 20dB respectively. In this particular application, we observed similar performance with the CCFS spectral fea-

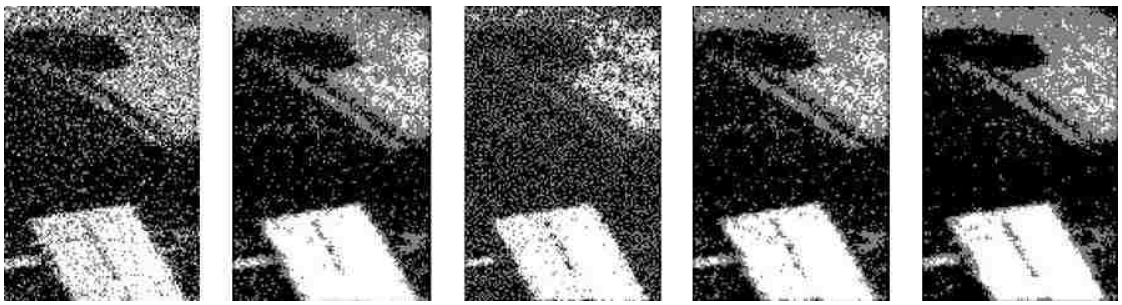
[2]Thematic maps are primarily designed to show a theme, a single spatial distribution or a pattern, using a specific map type. These maps show the distribution of a feature over a limited geographic area. Each map defines a partitioning of the area into a set of closed and disjoint regions, each includes all the points with the same feature value.

tures and the NAPP spectral features for both 10 and 20dB cases. The performance of both algorithms is similar again with the addition of spatially enhanced spectral features during the the classification stage.

The optimally performing combination of spectrally enhanced spatial features was used to investigate the sensitivity of the CCSS-FS algorithm with respect to the initial subsets of the AHI bands used during the first, CCFS, stage. We compare the performance of the CCSS-FS based on CCFS superposition features selected from subsets of 50 consecutive and 50 uniformly spaced AHI spectral bands. Cases (i) and (iii) in Fig. 4.7 show the classification performance based only on the superposition spectral features, when the superposition spectral features were selected from a set of 50 uniformly spaced AHI bands, and for SNR values 10dB and 20dB. The cases (ii) and (iv) in Fig. 4.7 show the classification performance based on the optimally performing combination of spectrally enhanced spatial features. Comparing these results with the results presented in Fig. 4.5 and 4.6 for the CCFS case, it is clear that uniformly spaced bands exhibit much lower sensitivity to noise compared to consecutive spectral bands, where the spectral correlation due to the bands' overlap is significant. As we can see for the case of SNR of 20dB, the results presented in Fig 4.7 are already in the limits of the noiseless case. For an SNR of 10dB, however, as shown in Fig 4.7, cases (i) and (ii), the extension of the canonical features with optimal selection of spatial features removed most of the classification errors.

4.5 Conclusions

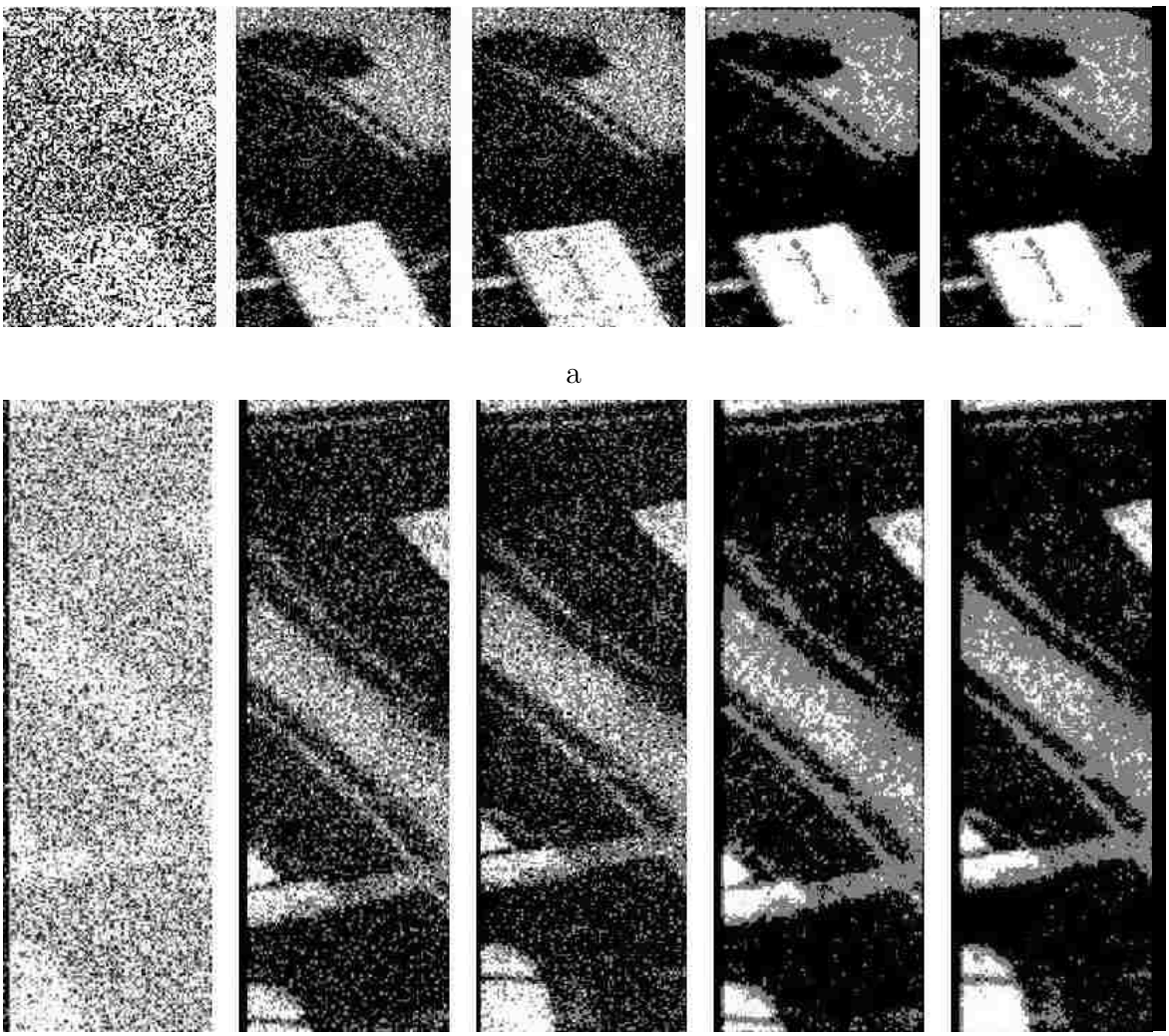
In this Chapter we have developed a simple methodology for the integration of spatial and spectral contents during feature selection and classification of hyperspectral imagery. The methodology is built upon the extension of the CCFS framework to a sequential spatio-spectral feature-selection (CCSS-FS) approach that utilizes the



(i) (ii) (iii) (iv) (v)

Figure 4.4: CCSS-FS thematic maps at SNR of 10dB. Left to right: (i) three CCFS features; (ii) three CCFS features extended with three Mean spatial features; (iii) three CCFS features extended with three Laplacian spatial features; (iv) case (ii) extended with three Gaussian spatial features; and (v) case (iv) extended with three Median spatial features.

idea of spectrally enhanced spatial features. Our studies show that inclusion of contextually appropriate spatial features, extracted for each canonical superposition band, can lead to noticeable improvement in the classification accuracy for low and medium SNR cases. Additionally, more significant improvement is observed when the first stage of the CCSS-FS, i.e., the spectral feature selection using the CCFS, is performed on a set of consecutive spectral bands that exhibit higher spectral correlation compared to uniformly spaced spectral bands. Our results confirm the potential of combining spectral superposition bands with a proper spatial feature extraction to utilize efficiently imagery from MS and HS sensors.

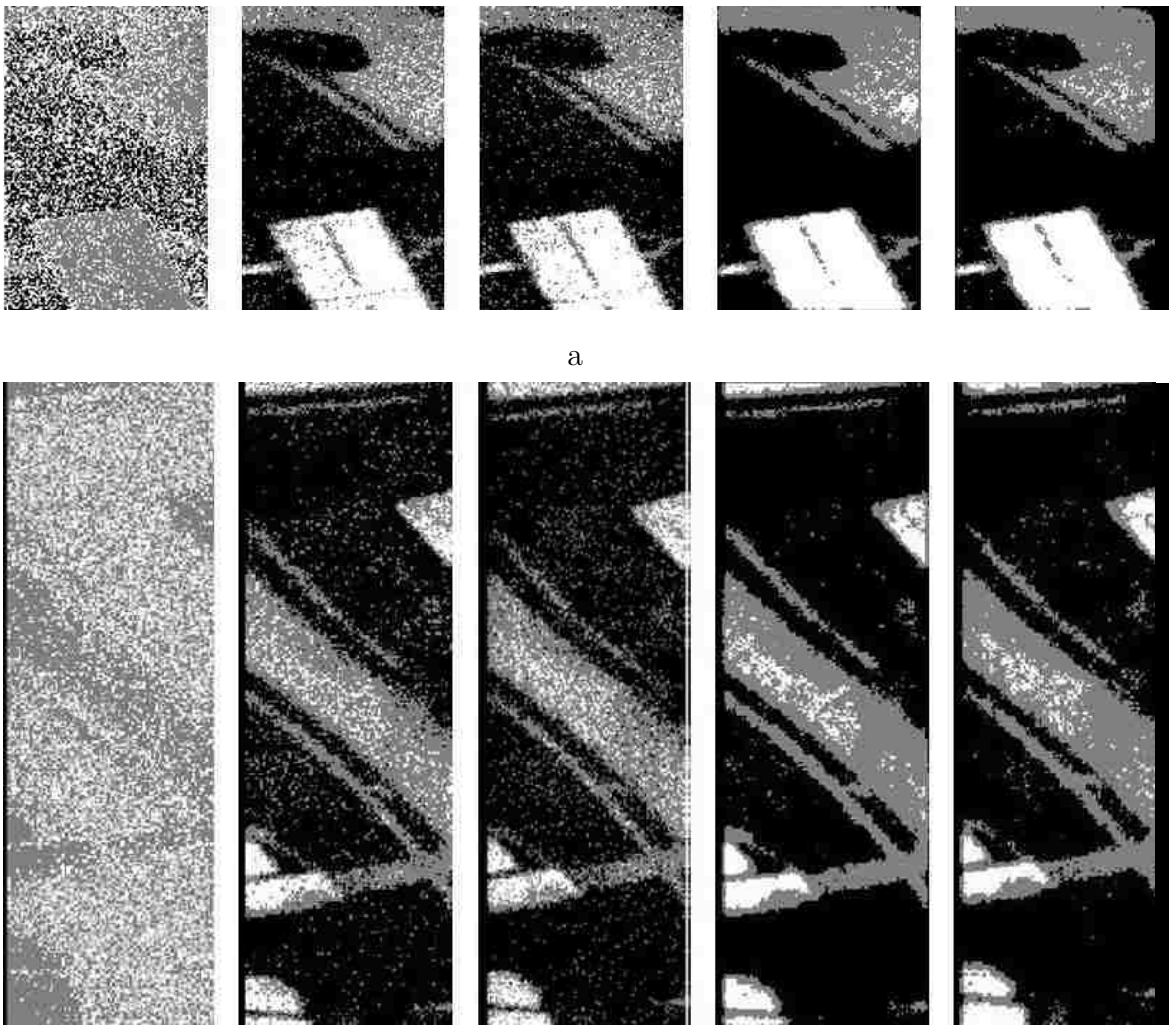


a

b

(i) (ii) (iii) (iv) (v)

Figure 4.5: Thematic maps based on the classification results for classes building, road and vegetation/ground using three superposition features selected by the CCFS and the noise-adjusted PP algorithms from a subset of 50 consecutive AHI bands and for SNR of 10dB. (a) Testing image 1; (b) Testing image 2. Left to right: (i) image classification without noise compensation; (ii) three CCFS features; (iii) three noise-adjusted PP features; (iv) case (ii) extended with Mean, Median and Gaussian spatial features; (v) case (iii) extended with Mean, Median and Gaussian spatial features.

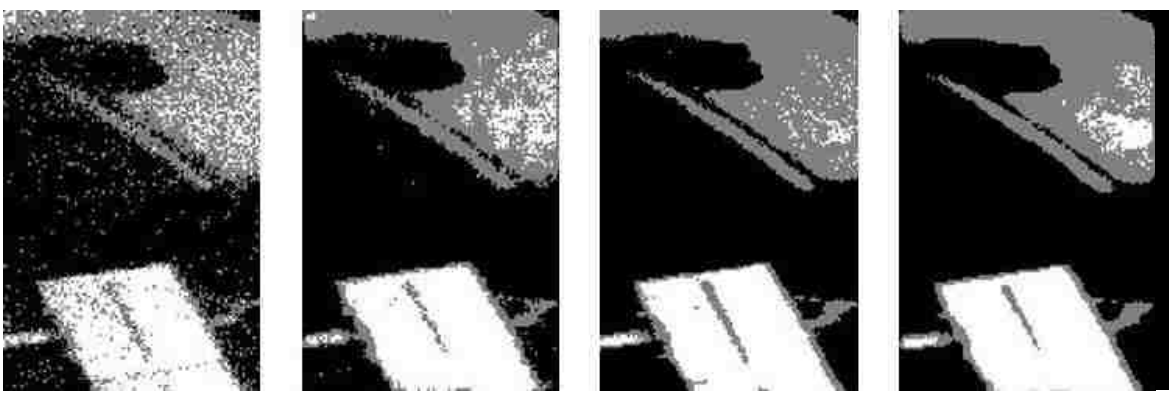


a

b

(i) (ii) (iii) (iv) (v)

Figure 4.6: Thematic maps based on the classification results for classes building, road and vegetation/ground using three superposition features selected by the CCFS and the noise-adjusted PP algorithms from a subset of 50 consecutive AHI bands and for SNR of 20dB. (a) Testing image 1; (b) Testing image 2. Left to right: (i) image classification without noise compensation; (ii) three CCFS features; (iii) three noise-adjusted PP features; (iv) case (ii) extended with Mean, Median and Gaussian spatial features; (v) case (iii) extended with Mean, Median and Gaussian spatial features.



(i) (ii) (iii) (iv)

Figure 4.7: Thematic maps based on the classification results for classes building, road and vegetation/ground. Left to right: (i) three CCFS features, selected from a subset of 50 uniformly spaced AHI bands, SNR of 10dB; (ii) case (i) extended with the optimal spatial features; (iii) same as case (i), but SNR of 20dB; (iv) case (iii) extended with the optimal spatial features.

Chapter 5

Joint Spatio-Spectral Feature Selection

In this chapter we turn attention to algorithms for *joint spatio-spectral* (JSS) feature selection. By "joint" we mean algorithms that simultaneously take into consideration spatial *and* spectral characteristics of a given multispectral (MS) or hyperspectral (HS) image. JSS feature selection offers unique opportunities for image processing and remote sensing because it allows us to take advantage of the correlation between spatial and spectral features. At the same time, development of JSS algorithms poses additional challenges. In this chapter we develop and verify, using AHI and DWELL imagery, a new approach for edge detection in HS and MS images, termed Spectral Ratio Contrast (SRC) edge detection algorithm that uses the concept of spectral ratio signatures.

The chapter is organized as follows. In Section 5.1 we review the relevant previous work on multi-color edge detection and the use of spectral ratio indices in image processing. The novel SRC edge detector and its implementation are presented in Section 5.2. In Section 5.3 we validate the SRC edge detector using AHI and DWELL

imagery data. For a benchmark, we consider the Multi Color Gradient (MCG) edge detector proposed in [1]. For the convenience of the readers, a brief summary of the MCG approach and its application to edge detection are presented in Appendix A. Our key findings are summarized in Section 5.4.

We recall the notation (4.1) for multi-color, MS, or HS images, introduced in Chapter 4. Specifically, such an image is a three-dimensional array of real numbers, also called an image *hypercube*:

$$\mathbf{u} \in \mathbf{R}^{I\times J\times K}; \quad \mathbf{u} = \{u_k(i,j) \in \mathbf{R} \,:\, 1 \le i \le I\,,\ 1 \le j \le J\,,\ 1 \le k \le K\}\,. \tag{5.1}$$

For fixed i and j, the k-dimensional vector $\mathbf{u}(i,j) = (u_1(i,j), \ldots, u_K(i,j))$ is called hyper-pixel. The indices i and j denote the spatial position of the hyper-pixel within a Focal Plane Array (FPA), and k is the band number. The value of $u_k(i,j)$ is referred to as the *intensity* of the k-th band. Alternatively, for a fixed k, the two-dimensional array $u_k(i,j)$ is the k-th image plane, or color slice, of the multi-color image. For MS images the number of bands K is typically between 10 and 15. For hyperspectral images K can be as high as several hundred bands.

5.1 Background and overview of relevant work

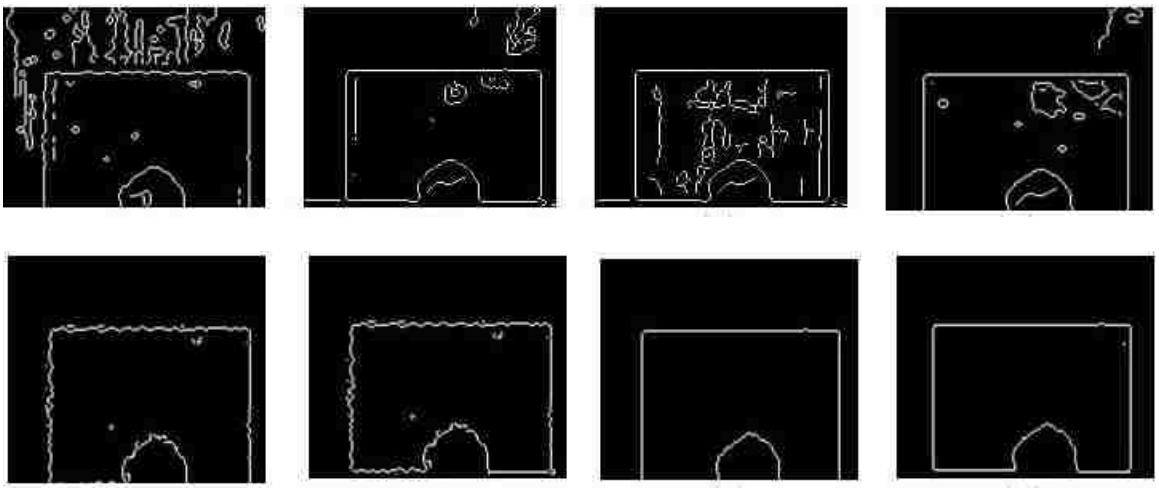
Image segmentation is one of the most important and difficult tasks in digital image processing. It represents a key stage of automated image analysis and interpretation. Segmentation algorithms for gray-scale images utilize basic properties of intensity values such as discontinuity and similarity [2, p.568]. For example, grayscale edges are defined as sets of pixels whose derivative values exceed a preset threshold. As a result, detection of discontinuities in gray-scale images is typically based on spatial masks which calculate a measure of the gray-level discontinuity for a specified (usually the central) pixel in the mask. The response of a mask at any pixel in the image is

given by a weighted sum of the gray-scale values [2, p.569].

Mathematically, a gray-scale image such as a broadband infrared (IR) image, can be modeled by a single-valued function $u(x, y)$ of the spatial coordinates (x, y). The pixels of the image are the values $u_{ij} = u(x_i, y_j)$ of that function at $N \times M$ uniform grid points $\{x_i, y_j\}$. Using this model it is easy to see that the action of a given mask on a gray-scale image can be thought of as a finite difference approximation of a differential operator acting on the function $u(x, y)$. Among the most popular gray-scale edge detectors are Canny [3], Sobel [4], and Perwitt [2], to name just few.

The Canny edge detector is considered to be one of the most robust gray-scale edge detectors. The Canny algorithm works in a multi-stage fashion [3], [78, Chapter 5]. First, the image is smoothed by Gaussian convolution. Then, a simple 2-D first derivative operator is applied to the smoothed image to identify the regions with high first-order spatial derivatives which give rise to "ridges" in the gradient magnitude image. The algorithm then tracks along the tops of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as non-maximal suppression. The tracking process exhibits hysteresis controlled by two thresholds. Tracking can only begin at a point on a ridge higher than the first threshold. Tracking then continues in both directions out from that point until the height of the ridge falls below the second threshold. This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments. Figure 5.1 compares the edge maps obtained by the Canny and the Sobel edge detectors applied to individual image planes obtained by the DWELL FPA.

In this chapter we consider image segmentation algorithms for multi-color images with particular emphasis on detection of multi-color edges. Switch from a gray-scale to a multi-color image significantly complicates edge detection. First, the standard definition of a gray-scale edge as a "ramp," or "ridge" between two regions [2, p.573] is not appropriate anymore, because a multi-color image has multiple image planes



(i) (ii) (iii) (iv)

Figure 5.1: Edge maps obtained by application of the Canny [3] and Sobel [4] gray-scale edge detectors to individual bands of the raw DWELL FPA imagery shown in Fig. 5.13 (left), consisting of limestone, granite and background. The bands are identified by their corresponding bias voltages as follows: (i) 0.3 V; (ii) 0.5 V; (iii) 0.7 V; (iv) 0.9 V. Top row: Canny edge detector. Bottom row: Sobel edge detector.

(channels) corresponding to different spectral bands. Depending on the scene, two distinct regions may exhibit the same intensity for one or more bands. In other words, with respect to these bands, the edge between the two regions is *iso-luminant*, i.e., it is characterized by a jump in *color* rather than *intensity*. It is clear that iso-luminant edges cannot be detected by a standard gradient operator because they do not exhibit an intensity "ramp" that can be estimated by the magnitude of this operator. This has been pointed out by Chan [28] who argues that gray-scale algorithms should not be applied directly[1] to multi-color images because intensity-based processing fails to

[1]Extension of other imaging techniques, based on differential operators, from gray-scale to multi-colored images faces similar difficulties. One example is Rudin-Osher-Fatemi's Total Variation de-noising method [79] for gray-scale images. In [28] Chan developed

detect iso-luminant edges.

This phenomenon is illustrated in Fig. 5.1 using the DWELL FPA imagery shown on Fig. 5.13, left. The image comprises of a background, granite and limestone classes, and is characterized by an almost iso-luminant edge between the granite and the limestone. The edge maps in Fig. 5.1 were obtained by the Canny and the Sobel edge detectors, applied to the individual bands in the DWELL FPA image. From the edge maps shown in the bottom row in Fig. 5.1 we can clearly see that the Sobel edge detector has missed the edge between granite and limestone in all image planes. The more sophisticated Canny edge detector picks this edge in some planes but not in all image planes.

Extension of differential edge detection to multi-color images has followed two principal paths [80]. A straightforward approach is to apply differential operators such as the gradient separately to each image plane and then somehow integrate this information to obtain edge and segmentation information. For example, one can apply a segmentation model to every image plane separately and then use bitwise logic operations to obtain segmentation for the multi-color image. Chan [34] points out that this can result in undesirable segmentation because information in separate channels is treated as independent whereas in actuality it is not.

Regarding edge detection, Sapiro [29] identifies three main drawbacks of the straightforward approach. First, edges can be defined by combinations of different planes and may be missing in some of the image planes. Examples are color images where iso-luminance areas show edges only for specific planes. Second, separate processing of image planes disregards the fact that, in general, information in them is highly correlated. Third, integration of information from separate image planes is not trivial and often is done in an ad hoc manner. In cases when an edge appears only in a subset of image planes there are no clear ways to integrate the

extensions of the TV norm that are applicable to multivalued images.

information.

A second approach to multi-color edge detection is to embed the variations of all color channels in a single measure, which is then used to obtain the edge maps [80]. Typically, this approach is developed by starting from a given gray-scale operator which is then consistently extended to multi-color images. By "consistently" we mean that the extended multi-color operator reduces to its original gray-scale prototype when applied to a single color image. Two representative examples of this approach are the *multi-color gradient* (MCG), proposed by Di Zenzo in [1], and the *morphological color gradient* (MoCG) of Evans and Liu [27].

The MCG operator [1] represents a consistent extension of the standard gradient operator to multi-color images. In contrast to the standard gray-scale gradient, which measures differences in the intensity values of the pixels in a mask, the MCG measures the local "steepness" of the multi-color image considered as a manifold embedded in a Euclidean space. A hyper-pixel belongs to a multi-color edge if the local steepness of the manifold, as measured by MCG, exceeds a given threshold. Because MCG edge detection utilizes simultaneously spatial and spectral information, it is an instance of a joint spatio-spectral image processing algorithm. We refer the reader to Appendix A for a brief summary of the MCG operator and MCG-based edge detection, which is used as the benchmark for the studies in Section 5.3.

Similarly, the MoCG operator [27] is a consistent extension of the morphological gray-scale gradient operator [81] to multi-color images. The latter is defined as the difference of the dilation and the erosion operators [74, 2], applied to a given structuring element γ. The starting point in [27] is the following equivalent form of the morphological gradient

$$\nabla u = \max |u(i_1, j_1) - u(i_2, j_2)| \quad \forall (i_1, j_1), (i_2, j_2) \in \gamma \tag{5.2}$$

where $u(i, j)$ is a gray-scale image. The idea of [27] is to replace the absolute value

operator in (5.2) by a vector norm; using the notation in (5.1) the proposed extension of (5.2) can be written as

$$\nabla \mathbf{u} = \max \|\mathbf{u}(i_1, j_1) - \mathbf{u}(i_2, j_2)\|_p \quad \forall (i_1, j_1), (i_2, j_2) \in \gamma \,, \tag{5.3}$$

where $\|\cdot\|_p$ is the p-th vector norm in the Euclidean space $\mathbf{R}^K$:

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{k=1}^{K} |\mathbf{x}_k|^p} \,.$$

The MoCG extension of (5.2) is another example of a joint spatio-spectral operator. Because for $K = 1$ any vector norm reduces to an absolute value of the single scalar component, it is clear that MoCG is indeed a consistent extension of (5.2).

The multi-color gradient and related ideas have been used with great success in digital image processing applications [26, 29, 25, 82], among others. However, as shown by the complexity estimate in Section 5.2.2, for multi-color images with large numbers of bands, such as hyperspectral imagery, computation of the multi-color gradient can be quite expensive. In this chapter we propose and develop an alternative joint spatio-spectral approach that utilizes information only from a few bands. Our main idea is to use the notion of *spectral ratio contrast*, i.e., band ratios, to define an edge signature (index) for an edge between two materials.

The edge signature represents a combination of spectral ratios calculated using bands that enhance the spectral contrast between the two materials. In conjunction with a spatial mask, the edge signatures give rise to a multispectral operator that can be viewed as a three-dimensonal extension of the mask, as shown in Fig. 5.2. In the extended mask, the third (spectral) dimension of each hyper-pixel can be chosen independently. *Such a mask does not operate in a single image plane but instead fuses information from multiple planes.*

We term this joint spatio-spectral approach Spectral Ratio Contrast (SRC) edge detection algorithm. SRC has two stages. The first stage is a training step which

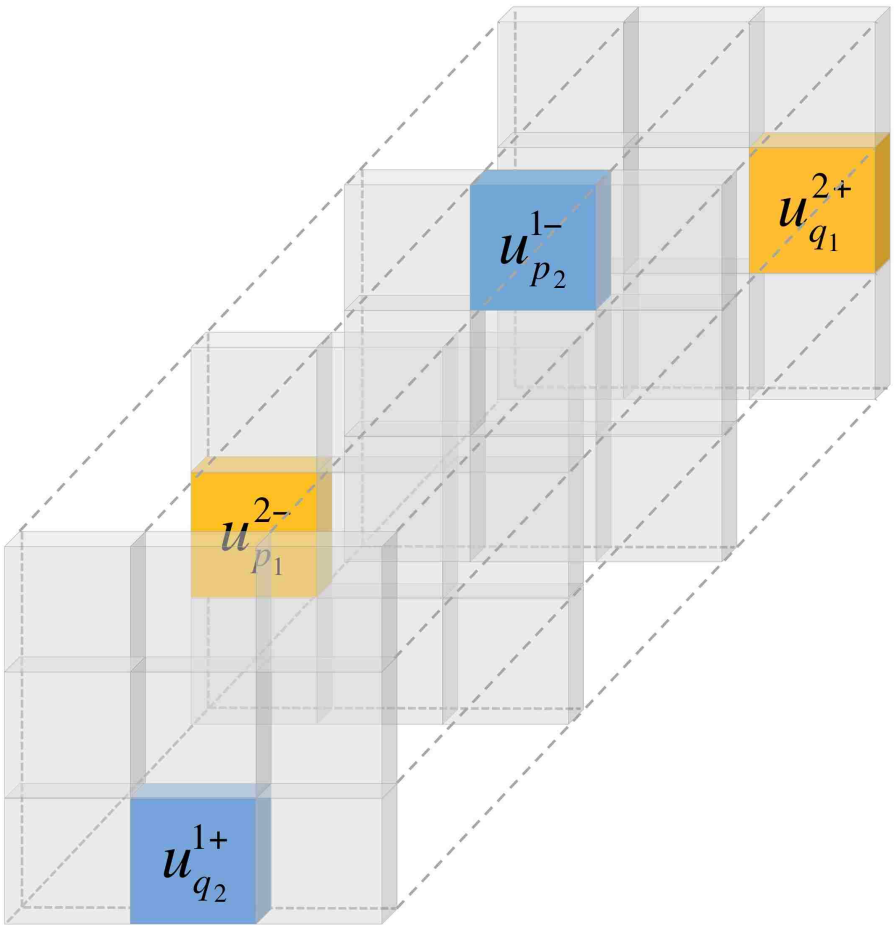


Figure 5.2: Fusion of spectral edge signatures with a spatial mask yields a non-separable joint spatio-spectral mask.

identifies the bands that maximize the spectral contrast between two given materials. The second stage is the feature extraction step using the three-dimensional mask defined by the bands selected at the training step. The presence of two stages in SRC is one important distinction from the MCG-based edge detection and other unsupervised edge detection algorithms. A second key difference is that SRC is not derivative-based, i.e., edge detection is effected by matching a given edge signature rather than by measuring the gradient magnitude.

Spectral ratios and cross-spectral ratios have been previously used for quantitative vegetation monitoring. Examples include the Normalized Difference Vegetation Index (NDVI) [35], the Soil-Adjusted Vegetation Index (SAVI) [36] and the Atmospherically Resistant Vegetation Index (ARVI) [37]. NDVI is perhaps the simplest of these indices and is defined by the formula

$$\text{NDVI} = (NIR - VIS)/(NIR + VIS)\,, \tag{5.4}$$

where NIR and VIS stand for near infrared and visible light reflected by vegetation [35]. Vegetation indices have demonstrated excellent capacity to distinguish vegetation areas from non-vegetation surfaces. From (5.4) we also see that an index compresses the data by a factor of two per ratio which is a significant added advantage of this approach.

Regional seismic discrimination [38, 39, 40, 41] and deblurring of noisy multichannel images [30] are two other applications where spectral ratios have been successfully utilized. In seismic applications spectral ratios are used to discriminate between natural events such as earthquakes and man-made events such as nuclear or chemical explosions. For example, [40] examines the use of five different spectral ratios based on combinations of various phase and frequency bands to this end. Their analysis identifies high-frequency ratio and short-period to long period Rayleigh wave ratio as two of the most consistent discriminants for separating earthquakes from explosions [40]. As a result, spectral ratios have found use in important applications such as monitoring of the comprehensive nuclear test-ban treaty adopted by the United Nations in 1996. We refer the reader to [41] for further details.

Application of spectral ratios to define multispectral operators for edge detection in this dissertation is novel and previously unexplored research direction. Besides the potential for significant data compression in HS and MS image processing, spectral ratios appear to be particularly well-suited for intelligent sensing algorithms using the DWELL sensor. Indeed, the training phase of the proposed SRC approach extracts

information about the most informative, with respect to edge detection, bias voltages (bands) in the sensed image. Image acquisition for a scene then can be carried out using only the relevant bias voltages, thereby significantly reducing the amount of data necessary for the image segmentation.

5.2 Spectral ratio contrast algorithm for edge detection

In this section we develop the joint spatio-spectral SRC approach for edge detection in MS and HS images. Given two distinct materials A and B, the main idea is to create a unique signature $\mathcal{E}_{\text{AB}}$ for the edge E_{AB} between A and B using the spectral ratios of their hyper-pixels. In other words, we seek those bands from A and B whose ratios can best discriminate E_{AB} from the rest of the spatial features in the scene.

We define E_{AB} as the set of hyper-pixels on the boundary between A and B. Thus, pixels belonging to E_{AB} are characterized by the existence of small neighborhoods containing elements from both A and B. Let

$$\mathbf{a} = (a_1, \ldots, a_K) \in \mathbf{R}^K \quad \text{and} \quad \mathbf{b} = (b_1, \ldots, b_K) \in \mathbf{R}^K$$

denote the representative hyper-pixels of A and B, respectively. For example, $\mathbf{a}$ and $\mathbf{b}$ may correspond to the class-means determined from a training set. In what follows, given a vector $\mathbf{c} \in \mathbf{R}^K$ we will use the notation

$$1/\mathbf{c} \triangleq (1/c_1, \ldots, 1/c_K)\,.$$

We define the *spectral ratio index* between A and B as the following $K \times K$ matrix:

$$\mathrm{A/B} \triangleq \mathbf{a}^T(1/\mathbf{b}) = \begin{pmatrix} \frac{a_1}{b_1} & \frac{a_1}{b_2} & \cdots & \frac{a_1}{b_K} \\ \frac{a_2}{b_1} & \frac{a_2}{b_2} & \cdots & \frac{a_2}{b_K} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{a_K}{b_1} & \frac{a_K}{b_2} & \cdots & \frac{a_K}{b_K} \end{pmatrix}. \tag{5.5}$$

The spectral ratio index of A is the matrix $\mathrm{A/A} \triangleq \mathbf{a}^T(1/\mathbf{a})$. Clearly, the diagonal of A/A is given by $(1, \ldots, 1)$.

The first step of the SRC approach is to identify a small subset of S ratios from (5.5), $S \ll K$, that can reliably discriminate the edge E_{AB} between A and B from other spatial features. We call such a subset *spectral ratio index* or *signature* of the edge E_{AB}. Succinctly, the edge signature of E_{AB} is a set

$$\mathcal{E}_{\mathrm{AB}} = \{(p_1, q_1, \rho_1), \ldots, (p_S, q_S, \rho_S)\,, \rho_s = a_{p_s}/b_{q_s}\,, 1 \leq S < K\}\,. \tag{5.6}$$

The integer S is referred to as the *length* of the edge signature. We remind that the goal is to find edge signatures with the shortest possible lengths so that $S \ll K$.

In order to extract spatial features such as edges, the edge signatures must be combined with a suitable spatial mask to obtain a joint spatio-spectral SRC mask; this is the second step of SRC. As in standard gray-scale image processing, the purpose of this mask is to compute the image response to a spatial structuring element. However, in SRC the spatial mask is used in a fundamentally different manner. Whereas in gray-scale edge detection the response is single-valued and represents a weighted average of the intensity values of the pixels in the structuring element, in SRC the response is multi-valued and returns the ratios of suitably defined pixel pairs from that structuring element.

Intuitively, in the SRC mask the band indices play the role of "weights" and summation of pixel intensities is replaced by their division according to a pair of band

indices. Therefore, starting from a given spatial mask we can define a joint-spatio-spectral SRC mask by retaining the spatial domain of the former and redefining its action in terms of spectral ratios corresponding to the bands from a given edge signature.

To illustrate the process of converting a spatial mask into a SRC mask, consider a gray-scale mask $\mathcal{M}$ whose response on a gray-scale image u is defined by the formula

$$\mathcal{M}(i,j) = \sum_{-N \leq m,n \leq N} \omega(m,n) u(i+m, j+n)\,, \tag{5.7}$$

where $u(i,j)$ is the pixel's intensity value at the center $\{i,j\}$ of the mask, $\omega(m,n)$ are the weights at offsets $\{m,n\}$ relative to the center of the mask, and N is the diameter of $\mathcal{M}$. We assume that $\mathcal{M}$ contains M distinct pixel pairs $\{u^{m-}(i,j), u^{m+}(i,j)\}_{m=1}^{M}$, centered at $\{i,j\}$, with weights $\{\omega^{m-}, \omega^{m+}\}_{m=1}^{M}$, so that the response of $\mathcal{M}$ can be written more simply as a sum over all distinct pixel pairs as follows:

$$\mathcal{M}(i,j) = \sum_{m=1}^{M} \left(\omega^{m-} u^{m-}(i,j) + \omega^{m+} u^{m+}(i,j)\right)\,. \tag{5.8}$$

To explain this notation, consider the 3×3 single-band image and the associated 3×3 mask shown in Fig. 5.3. For this mask the distinct pixel pairs centered at $\{i,j\}$ are given by

$$\{u(i,j-1), u(i,j+1)\};$$

$$\{u(i-1,j), u(i+1,j)\};$$

$$\{u(i-1,j-1), u(i+1,j+1)\};$$

$$\text{and} \quad \{u(i-1,j+1), u(i+1,j-1)\}\,,$$

respectively. Therefore, in this case $M = 4$ and we can make the following associa-

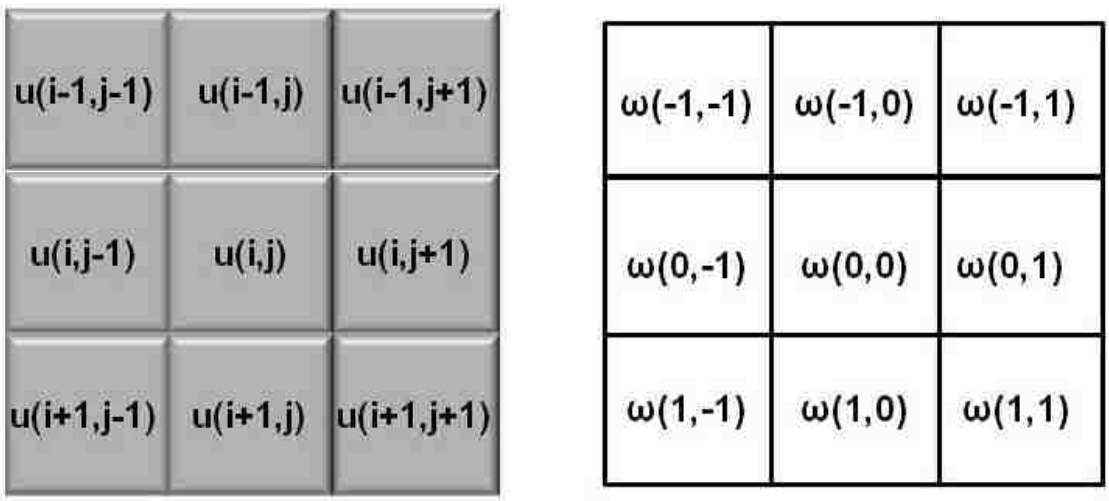


Figure 5.3: Left: 3 by 3 section of a single-band image; right: 3 by 3 spatial mask.

tion:

$$\begin{aligned}
\{u^{1-}(i,j), u^{1+}(i,j)\} &\rightarrow \{u(i,j-1), u(i,j+1)\} \\
\{u^{2-}(i,j), u^{2+}(i,j)\} &\rightarrow \{u(i-1,j), u(i+1,j)\} \\
\{u^{3-}(i,j), u^{3+}(i,j)\} &\rightarrow \{u(i-1,j-1), u(i+1,j+1)\} \\
\{u^{4-}(i,j), u^{4+}(i,j)\} &\rightarrow \{u(i-1,j+1), u(i+1,j-1)\}
\end{aligned} \tag{5.9}$$

Because the pixel pairs are used to define spectral ratios, the order of their elements is not important.

To define the response of the joint spatio-spectral SRC mask we discard the weights in (5.8) and combine the pixel pairs in $\mathcal{M}$ with the band indices from a given edge signature $\mathcal{E}_{\text{AB}}$. The result is a mapping $\mathcal{K}_{\text{AB}}$

$$\mathcal{K}_{\text{AB}} : \mathbf{R}^{I\times J\times K} \mapsto \mathbf{R}^{I\times J\times (M\times S)}, \tag{5.10}$$

from the three-dimensional MS or HS image cube with spatial dimensions $I \times J$ and a spectral dimension K into a smaller hypercube, with the same spatial dimensions $I \times J$ but with a reduced spectral dimension $MS \ll K$. Recall that M denotes the

number of distinct pairs of pixels involved in the mask and S is the length of the edge signature $\mathcal{E}_{\mathrm{AB}}$.

At every spatial location $\{i, j\}$, the response of $\mathcal{K}_{\mathrm{AB}}$ can be viewed as an $M \times S$ matrix of spectral ratios given by

$$\mathcal{K}_{\mathrm{AB}}(i,j) = \begin{pmatrix} \dfrac{u_{p_1}^{1-}(i,j)}{u_{q_1}^{1+}(i,j)} & \dfrac{u_{p_2}^{1-}(i,j)}{u_{q_2}^{1+}(i,j)} & \cdots & \dfrac{u_{p_s}^{1-}(i,j)}{u_{q_s}^{1+}(i,j)} \\ \dfrac{u_{p_1}^{2-}(i,j)}{u_{q_1}^{2+}(i,j)} & \dfrac{u_{p_2}^{2-}(i,j)}{u_{q_2}^{2+}(i,j)} & \cdots & \dfrac{u_{p_s}^{2-}(i,j)}{u_{q_s}^{2+}(i,j)} \\ \cdots & \cdots & \dfrac{u_{p_n}^{m-}(i,j)}{u_{q_n}^{m+}(i,j)} & \cdots \\ \dfrac{u_{p_1}^{M-}(i,j)}{u_{q_1}^{M+}(i,j)} & \dfrac{u_{p_2}^{M-}(i,j)}{u_{q_2}^{M+}(i,j)} & \cdots & \dfrac{u_{p_s}^{M-}(i,j)}{u_{q_s}^{M+}(i,j)} \end{pmatrix} \in \mathbf{R}^{M \times S} . \quad (5.11)$$

In the sequel, we denote the element (m, s) of the matrix in (5.11) by

$$\kappa_{\mathrm{AB}}^{m}(i,j,s) = \frac{u_{p_s}^{m-}(i,j)}{u_{q_s}^{m+}(i,j)} .$$

The sequence of spectral ratios in each row of the matrix $\mathcal{K}_{\mathrm{AB}}(i, j)$ is computed by using a distinct pixel pair and bands prescribed by a given edge signature. For example, if $\mathcal{K}_{\mathrm{AB}}$ is derived from the 3×3 mask shown in Fig. 5.3, then the first row corresponds to the ratios of the first pair of pixels $\{u^{1-}(i,j), u^{1+}(i,j)\}$, i.e., the horizontal pair $u(i, j-1)$ and $u(i, j+1)$, the second row corresponds to the second pair of pixels $\{u^{2-}(i,j), u^{2+}(i,j)\}$, i.e., the vertical pair $u(i-1, j)$ and $u(i+1, j)$, and so on.

To develop a better understanding of the joint spatio-spectral character of $\mathcal{K}_{\mathrm{AB}}$ consider an even simpler example in which the spatial mask comprises of the first two pairs in (5.9) and the edge signature has length 1, i.e., a case where $M = 2$ and $S = 1$. Therefore, at every spatial location $\{i, j\}$ the response $\mathcal{K}_{\mathrm{AB}}(i, j)$ is a 2×1 matrix. Computation of the row elements in this matrix is illustrated schematically

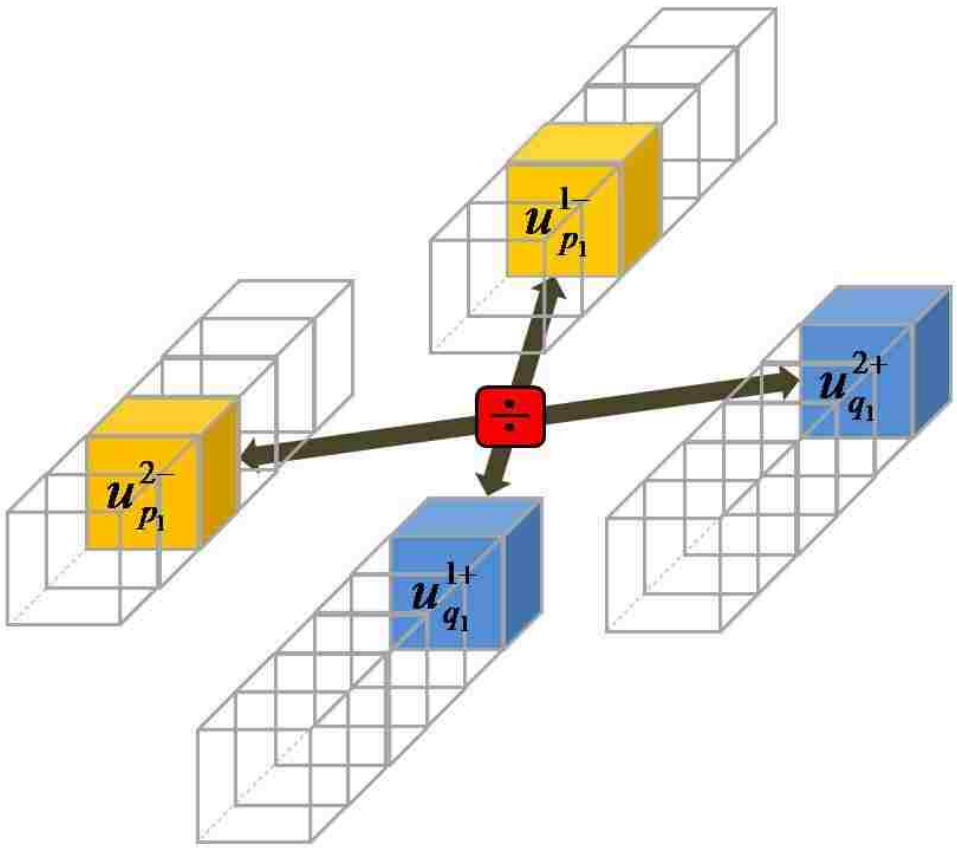


Figure 5.4: Computation of the joint spatio-spectral SRC mask $\mathcal{K}_{\mathrm{AB}}(i,j)$ for the first two pairs in (5.9) and edge signature $\mathcal{E}_{\mathrm{AB}}$ of length 1 ($M = 2$ and $S = 1$.)

in Fig. 5.4. This figure clearly shows that, unlike a conventional mask, $\mathcal{K}_{\mathrm{AB}}$ does not operate in a single image plane but instead fuses information from multiple planes and is not-separable into spectral and spatial components. The spectral plane for each element in $\mathcal{K}_{\mathrm{AB}}$ is selected through the edge signature $\mathcal{E}_{\mathrm{AB}}$, i.e., the indices $\{p, q\}$, in a way that enhances that feature.

We now proceed with the formulation of the training and feature extraction stages of the SRC approach and explain how $\mathcal{K}_{\mathrm{AB}}$ can be used to discriminate the edges between A and B. For clarity the selection criteria used to determine the edge signature

(5.6) are presented separately from the algorithm description.

5.2.1 Implementation of the SRC algorithm

For simplicity we describe the SRC approach assuming two distinct materials A and B, and then briefly discuss extension to more general cases. The algorithm has two stages: (1) a training step where one determines the appropriate edge signatures and defines the associated three-dimensional SRC mask, and (2) a feature-extraction step where the response of this mask on a given MS or HS image is used to effect the spatial feature extraction. Detailed description of each step follows.

Training stage

The three key components of this step are (1) selection of representative training data sets for materials A and B, (2) defining a selection criteria for determining the edge signature (5.6), and (3) selection of a spatial mask $\mathcal{M}$ used to define the SRC mask in (5.11). The second step is essential for the success of the SRC approach. In Sections 5.2.4–5.2.3 we propose two possible selection criteria that are used subsequently in the verification studies.

The training stage proceeds as follows. Let

$$\mathbf{a} = \{a_k(i,j) \in \mathbf{R} \,:\, i \in I(\mathrm{A}),\ j \in J(\mathrm{A}),\ 1 \le k \le K\}\,,$$

and

$$\mathbf{b} = \{b_k(i,j) \in \mathbf{R} \,:\, i \in I(\mathrm{B}),\ j \in J(\mathrm{B}),\ 1 \le k \le K\}$$

denote the training sets for classes A and B, respectively, where $I(\mathrm{A})$, $J(\mathrm{A})$, $I(\mathrm{B})$ and $J(\mathrm{B})$ are index sets that define the training samples. Using the training data we

compute the class means $\bar{\mathbf{a}} \in \mathbf{R}^K$ and $\bar{\mathbf{b}} \in \mathbf{R}^K$ according to

$$\bar{\mathbf{a}} = \frac{1}{\dim(\mathrm{A})} \left(\sum_{i \in I(\mathrm{A}), j \in J(\mathrm{A})} a_k(i,j) \right) ,$$

and

$$\bar{\mathbf{b}} = \frac{1}{\dim(\mathrm{B})} \left(\sum_{i \in I(\mathrm{B}), j \in J(\mathrm{B})} b_k(i,j) \right) ,$$

and the spectral ratio index A/B according to (5.5). Using the chosen selection criteria we identify a subset of A/B that defines the edge signature $\mathcal{E}_{\mathrm{AB}}$, and an associated tolerance value ϵ_{AB}. Finally, using the spatial mask $\mathcal{M}$ and the edge index $\mathcal{E}_{\mathrm{AB}}$ we define the SRC mask $\mathcal{K}_{\mathrm{AB}}$ according to (5.11). In summary, the output from the training step comprises of

1. The edge signature $\mathcal{E}_{\mathrm{AB}}$;

2. The associated tolerance value ϵ_{AB};

3. The SRC mask $\mathcal{K}_{\mathrm{AB}}$.

Feature extraction stage

At this step, we apply $\mathcal{K}_{\mathrm{AB}}$ to a given multispectral image $\mathbf{u} \in \mathbf{R}^{I \times J \times K}$. For each hyper-pixel $\mathbf{u}(i,j)$, located at position $\{i,j\}$ with respect to the FPA, we proceed as follows. First, we compute the $M \times S$ ratios giving the response $\mathcal{K}_{\mathrm{AB}}(i,j)$ at $\mathbf{u}(i,j)$, as defined in (5.11). Recall that M denotes the number of the distinct pixels pairs involved in the calculation of $\mathcal{K}_{\mathrm{AB}}(i,j)$ and S is the length of $\mathcal{E}_{\mathrm{AB}}$, i.e., the number of ratios used to discriminate the edge between A and B.

The response (calculated ratios) of the SRC mask at a given hyper-pixel $\mathbf{u}(i,j)$

is then used as an input to form an indicator matrix

$$\chi(i,j) = \begin{pmatrix} \delta_1^1 & \delta_2^1 & \cdots & \delta_S^1 \\ \delta_1^2 & \delta_2^2 & \cdots & \delta_S^2 \\ \cdots & \cdots & \delta_s^m & \cdots \\ \delta_1^M & \delta_2^M & \cdots & \delta_S^M \end{pmatrix} \in \mathbf{R}^{M \times S} \tag{5.12}$$

for that pixel. The elements δ_s^m of (5.12) are assigned the values of 0 or 1 according to the following rule:

$$\delta_s^m = \begin{cases} 1 \quad \text{if} \quad \begin{cases} \kappa_{\mathrm{AB}}^m(i,j,s) \in \rho_s + [-\epsilon_{\mathrm{AB}}, \epsilon_{\mathrm{AB}}], \text{ or} \\ (\kappa_{\mathrm{AB}}^m(i,j,s))^{-1} \in \rho_s + [-\epsilon_{\mathrm{AB}}, \epsilon_{\mathrm{AB}}] \end{cases} \\ 0 \quad \text{otherwise} \end{cases}, \tag{5.13}$$

where ϵ_{AB} is the tolerance determined at the training step. The use of the both $\kappa_{\mathrm{AB}}^m(i,j,s)$ and its reciprocal $(\kappa_{\mathrm{AB}}^m(i,j,s))^{-1}$ in (5.13) is required to account for the two possible material configurations at the m-th pixel pair $\{u^{m-}(i,j), u^{m+}(i,j)\}$. Specifically, the first ratio captures the case when $u^{m-}(i,j)$ is of class A and $u^{m+}(i,j)$ is of class B, whereas the reciprocal ratio is needed to account for the possibility that $u^{m-}(i,j)$ is of type B and $u^{m+}(i,j)$ is of type A. Therefore, the use of the two ratios removes dependence on the direction of the transition between A and B, and is similar to the use of the magnitude in the gradient operator to achieve its rotational invariance.

If the hyper-pixel pair $\{u_{p_s}^{m-}(i,j), u_{q_s}^{m+}(i,j)\}$ belongs to the same material type, then the test in (5.13) will force most, if not all, elements δ_s^m in the m-th row of $\chi(i,j)$ to zero. Conversely, if the hyper-pixels forming the pair are from the two different materials, either $\kappa_{\mathrm{AB}}^m(i,j,s)$ or its reciprocal $(\kappa_{\mathrm{AB}}^m(i,j,s))^{-1}$ will be close to the ratio ρ_s from the edge signature $\mathcal{E}_{\mathrm{AB}}$. As a result, the above test will set most if not all of the elements δ_s^m in the m-th row of $\chi(i,j)$ to one.

In summary, for a given pixel pair $\{u_{p_s}^{m-}(i,j), u_{q_s}^{m+}(i,j)\}$, the number of non-zeros

in the associated m-th row of the indicator matrix reveals the number of times the response of the SRC mask $\mathcal{K}_{\mathrm{AB}}(i,j)$ has matched, to within the specified tolerance, the spectral ratios from the edge signature $\mathcal{E}_{\mathrm{AB}}$. Because the pixel pairs used to form the response $\mathcal{K}_{\mathrm{AB}}(i,j)$ correspond to different edge orientations, e.g., horizontal, vertical or diagonal, the number of ones in each row indicates the strength of the particular edge direction at position $\{i,j\}$.

We use the information contained in the rows of $\chi(i,j)$ to define an edge map $\mathcal{F}_{\mathrm{AB}} : \mathbf{R}^{I\times J\times K} \mapsto \mathbf{R}^{I\times J}$, i.e., a mapping that assigns the value 1 to pixel location $\{i,j\}$ if $\mathbf{u}(i,j)$ belongs to an edge E_{AB}, and 0 otherwise, as follows. We classify $\mathbf{u}(i,j)$ as belonging to E_{AB} if the edge strength in at least one direction, as measured by the number of ones in the rows of $\chi(i,j)$, exceeds a specified integer tolerance value $\widetilde{S} \leq S$. If none of the edge strengths exceed $\widetilde{S}$, then $\mathbf{u}(i,j) \notin E_{\mathrm{AB}}$ and we set $\mathcal{F}_{\mathrm{AB}}(\mathbf{u}(i,j)) = 0$.

This criterion can be conveniently expressed in terms of the matrix infinity norm. For a given matrix $\mathbf{A} \in \mathbf{R}^{K\times L}$ the infinity norm $\|\mathbf{M}\|_\infty$ is defined as the maximum absolute row sum of $\mathbf{M}$:

$$\|\mathbf{M}\|_\infty = \max_{1\leq s\leq K} \sum_{l=1}^{L} |a_{kl}| \,.$$

Using this norm, we define the edge map $\mathcal{F}_{\mathrm{AB}} : \mathbf{R}^{I\times J\times K} \mapsto \mathbf{R}^{I\times J}$ as follows: given an integer $1 \leq \widetilde{S} \leq S$,

$$\mathcal{F}_{\mathrm{AB}}(\mathbf{u}(i,j)) = \begin{cases} 1 & \text{if} \quad \|\chi\|_\infty \geq \widetilde{S} \\ \\ 0 & \text{if} \quad \|\chi\|_\infty < \widetilde{S} \,. \end{cases} \tag{5.14}$$

The value of the integer parameter $\widetilde{S}$ can be used to adjust the sensitivity of the feature extraction step to, e.g., noise. For example, increasing $\widetilde{S}$ makes the algorithm less sensitive to noise but more restrictive. Conversely, decreasing $\widetilde{S}$ makes the algorithm less restrictive but more sensitive to noise. A similar effect can be

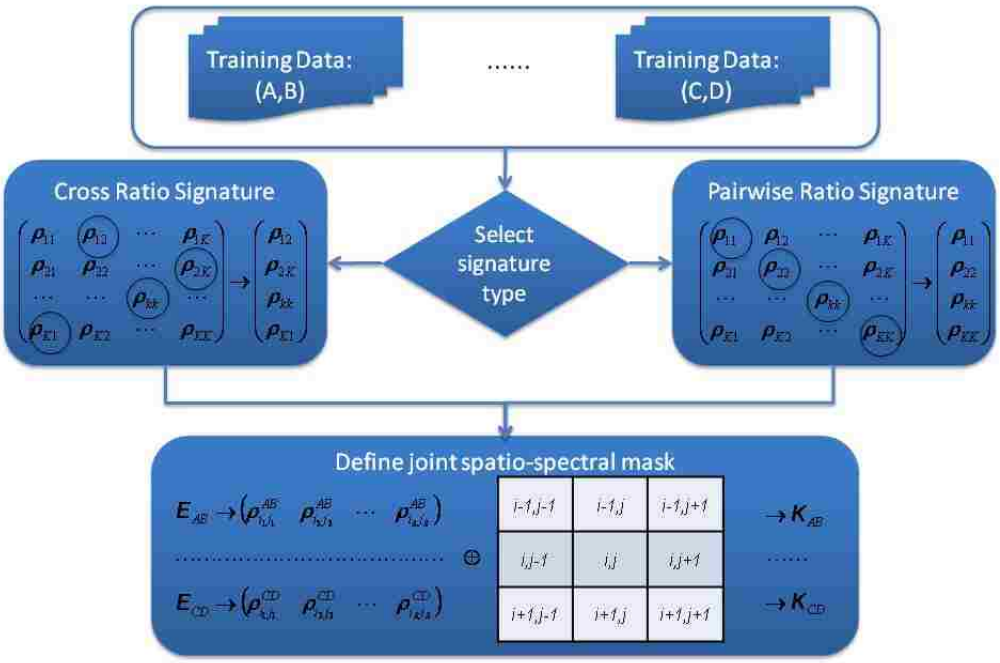


Figure 5.5: The training phase of the SRC algorithm.

achieved by setting $\widetilde{S} = S$ and increasing or decreasing the number of ratios in the edge signature $\mathcal{E}_{\mathrm{AB}}$.

Extension of the above approach to three or more materials is straightforward. We briefly describe the case of three distinct materials A, B and C. In this case there are three possible edge classes: between A and B, between A and C and between B and C. Accordingly, at the training stage we define three edge signatures, $\mathcal{E}_{\mathrm{AB}}$, $\mathcal{E}_{AC}$, and $\mathcal{E}_{BC}$, three tolerance values ϵ_{AB}, ϵ_{AC}, and ϵ_{BC}, and three joint spatio-spectral masks $\mathcal{K}_{\mathrm{AB}}$, $\mathcal{K}_{AC}$, and $\mathcal{K}_{BC}$.

At the feature extraction stage we use the responses of $\mathcal{K}_{\mathrm{AB}}$, $\mathcal{K}_{AC}$, and $\mathcal{K}_{BC}$ to identify the hyper-pixels belonging to edge E_{AB} between A and B, edge E_{AC} between

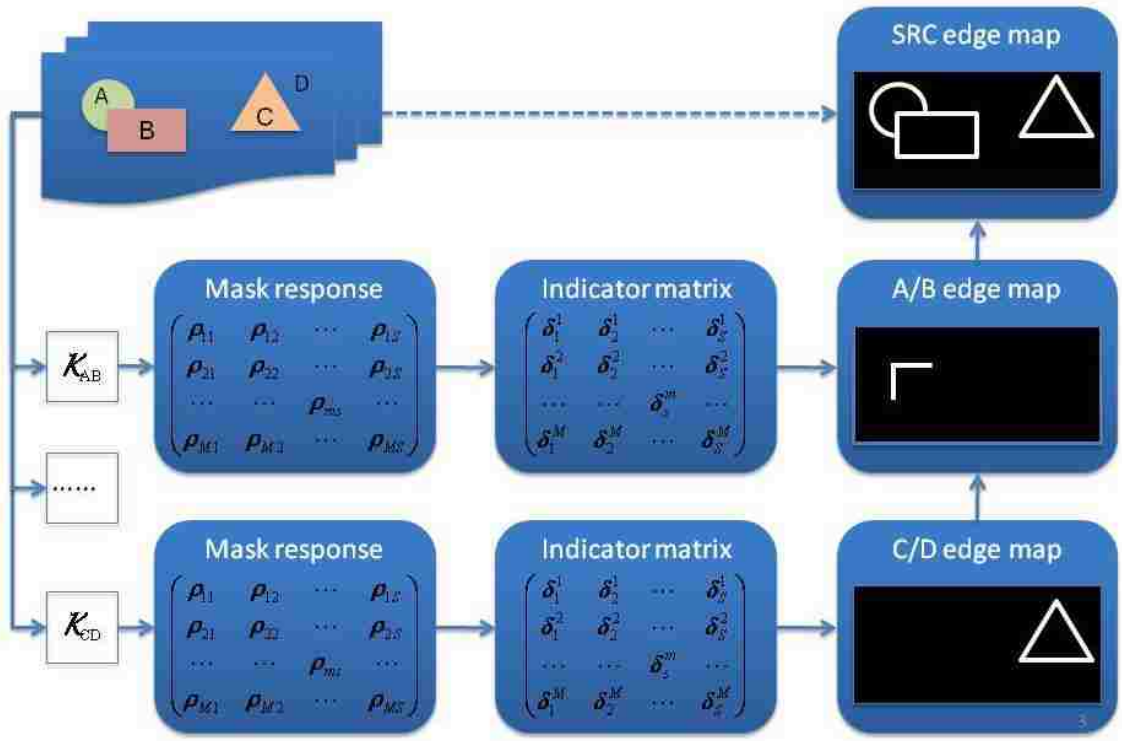


Figure 5.6: The feature extraction phase of the SRC algorithm.

A and C, and edge E_{BC} between B and C. The final image segmentation is obtained by the union of the three edges:

$$E_{ABC} = E_{\mathrm{AB}} \cup E_{AC} \cup E_{BC}\,. \tag{5.15}$$

The training and testing phases of the SRC algorithm are shown schematically in Fig. 5.5 and Fig. 5.6, respectively. Because the edge signatures and the associated tolerance values are determined independently for each pair of classes, the feature extraction depends only on the quality of the selection criteria used to obtain the edge signature for this pair and not on the strength of the edge, as measured by its MCG value. As a result, the SRC approach is particularly well suited to situations where the edge between two given materials is "weak," as measured by its MCG value,

compared to the edges between the other materials in the scene. For scenes that contain both "weak" and "strong" edges, as measured by their MCG gradients, an MCG-based feature extraction would require sophisticated locally adaptive threshold strategies to capture the "weak" edges. If a single, non-adaptive threshold is used, in order for MCG to pick the weak edges one has to increase the tolerance to a point where the noise level may become unacceptable. This conjecture is confirmed by the verification study in Section 5.3 using DWELL data corresponding to a scene with weak and strong edges.

Before we turn attention to strategies for selecting the edge signature from the spectral ratio index A/B, we estimate, in the next section, the complexity of the SRC algorithm. Then, in Sections 5.2.3–5.2.4 we discuss two approaches to determine the edge signatures in the SRC algorithm that will be used later in Section 5.3 to compare and contrast SRC with other feature extraction algorithms.

5.2.2 Complexity of the SRC algorithm

In this section we estimate the complexity of the feature extraction stage in the SRC algorithm and compare it with the cost of the MCG algorithm, described in Appendix A. The training stage is not included in the SRC complexity estimate because it is usually done off-line, before the actual image processing commences.

Consider an HS or an MS image given by a hypercube with dimensions $I \times J \times K$. We recall that $I \times J$ is the dimension of the image plane whereas K is the number of spectral bands, which ranges from tens for MS images, to hundreds for HS images.

The cost of the feature extraction stage of SRC comprises of (1) the cost to compute the response $\mathcal{K}_{\mathrm{AB}}(i,j)$ of the SRC mask at every pixel, including the reciprocal ratios $(\kappa^m_{\mathrm{AB}}(i,j,s))^{-1}$ needed in (5.13), (2) the cost to form the indicator matrix (5.12), and (3) the cost to compute the edge map according to (5.14). From these

three steps the first one is the costliest, as it involves floating point arithmetic. The second and the third steps require mostly integer arithmetic and logical operations that are faster than floating point arithmetic. Nonetheless, for simplicity, in the cost estimate we count all operations as being the same, without regard to the fact that some of them take less CPU time to execute.

Computation of the response $\mathcal{K}_{\mathrm{AB}}(i,j)$ at a single pixel (including the computation of the reciprocal ratios) takes $2MS$ operations. Thus, the total operation count for Step 1 is $2(I \times J) \times MS$. Forming the indicator matrix at Step 2 is based on (5.13) and requires at most four comparisons. Therefore, the cost of forming a single entry of $\chi(i,j)$ is 4 operations, the cost of forming $\chi(i,j)$ is $4MS$ and the cost of forming $\chi(i,j)$ for all pixels is $4(I \times J) \times MS$.

Finally, Step 3 requires computation of the matrix infinity norm of $\chi(i,j)$ for every pixel location $\{i,j\}$. Recall that the matrix infinity norm is the maximum absolute row sum of the matrix. Summing up the elements in a single row of $\chi(i,j)$ takes $S-1$ operations and so, computation of all row sums requires $M(S-1)$ operations. In the worst-case scenario, finding the largest of these sums takes $M-1$ comparisons. It follows that the per-pixel cost of Step 3 is $M(M-1)(S-1)$ operations, and the total cost of this step is $(I \times J) \times M(M-1)(S-1)$. To summarize, the total cost of the feature extraction step in SRC is shown in the Table 5.1.

We proceed to estimate the complexity of the MCG edge detector, as described in Section A.3 of Appendix A. For every pixel the cost of MCG comprises of (1) computation of the entries $g_{11}(i,j)$, $g_{12}(i,j)$ and $g_{22}(i,j)$ of the discrete first fundamental form $G(i,j)$, (2) computation of the eigenvalues of $G(i,j)$, and (3) computation of the monitor function and application of the threshold to compute the edge map.

Computation of the diagonal entries $g_{11}(i,j)$ and $g_{12}(i,j)$ takes the same amount of operations: one subtraction and one multiplication per term times $K-1$ additions

Step	Cost in floating point operations
Computation of the response $\mathcal{K}_{\mathrm{AB}}$	$2(I \times J) \times MS$
Forming of the indicator matrix $\chi(i,j)$	$4(I \times J) \times MS$
Computation of the edge map $\mathcal{F}_{\mathrm{AB}}$	$(I \times J) \times M(M-1)(S-1)$
Total operations	$(I \times J) \times (M^2(S-1) + 5MS + M)$

Table 5.1: Cost estimate of the SRC algorithm applied to an image hypercube with dimensions $I \times J \times K$.

to sum up all terms, for a total of $2(K-1)$ operations. The cost of the off-diagonal element is two subtractions and one multiplication per term times $K-1$ additions to sum up all terms, for a total of $3(K-1)$ operations. Therefore, the total cost to compute the first fundamental form for all hyper-pixels is $5(I \times J) \times (K-1)$.

Computation of the two eigenvalues of $G(i,j)$ can be done directly at a cost of 9 operations bringing the total cost of this step for the image to $9(I \times J)$ operations. Finally, the last step requires 3 operations per pixel: two operations to compute the monitor function and one comparison to apply the threshold. Thus, the total cost of this step is $3(I \times J)$ operations, and the total cost of the MCG edge detector, obtained by summing up the operation counts at each step, is shown in Table 5.2.

To appreciate the gains enabled by the SRC consider the case of the AHI data used in the verification studies in Section 5.3.1. In the experiments we utilize 200 out of the available 256 AHI bands. The longest and the shortest edge signatures used with the SRC algorithm have 5 and 1 ratios, respectively. The joint spatio-spectral mask $\mathcal{K}_{\mathrm{AB}}$ is generated using 4 pixel pairs. As a result, for the longest edge signature

Step	Cost in floating point operations
Computation of $G(i,j)$	$5(I \times J) \times (K-1)$.
Computation of the eignevalues $\lambda_{\pm}(i,j)$	$9(I \times J)$
Computation of the edge map	$3(I \times J)$
Total operations	$(I \times J) \times (5K+7)$

Table 5.2: Cost estimate of the MCG algorithm applied to an image hypercube with dimensions $I \times J \times K$.

$S = 5$, $M = 4$, and the SRC cost estimate is

$$4^2(5-1) + 5 \times 4 \times 5 + 4 = 168$$

floating point operations per pixel. In contrast, the cost estimate for the MCG is

$$5 \times 200 + 7 = 1007$$

floating point operations per pixel, i.e., a 6 fold increase over the cost of SRC. The difference is even more pronounced when the SRC is used with the single ratio signature for which $S = 1$. The SRC cost estimate reduces to

$$4^2(1-1) + 5 \times 4 \times 1 + 4 = 24$$

giving a 42 fold gain in performance over the MCG.

5.2.3 Pairwise ratio edge index

In this section we describe a strategy in which the selection of the spectral ratios forming the edge signature $\mathcal{E}_{\mathrm{AB}}$ is restricted to the diagonal elements of the spectral

ratio index A/B defined in (5.5). These ratios correspond to spectral bands with the same numbers, thus the term "pairwise ratios".

To motivate this strategy note that in the ideal case the spectral ratio index A/A for two identical materials has diagonal elements equal to one. Therefore, hyper-pixels from the same class will have nearly constant pairwise ratios close to 1, while hyper-pixels from different classes will have at least some of their pairwise ratios away from 1. We illustrate these observations using a sample of actual[2] AHI data. The AHI image, shown in Fig. 5.7 has three different types of materials: building (B), ground (G) and road (R).

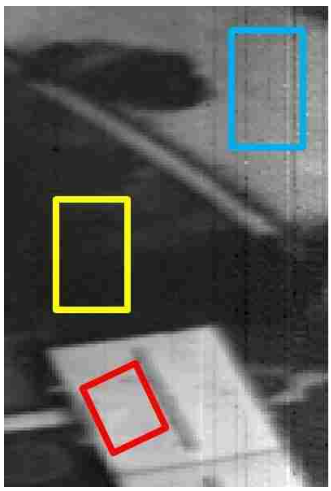


Figure 5.7: AHI training data: building (B), ground (G) and road (R) identified by red, yellow and cyan boxes, respectively.

Figure 5.8 shows several hyper-pixels from the B-class and their pairwise ratios, and Figure 5.9 shows the band ratios for the class-average hyper-pixels of the classes in Fig. 5.7. As expected, the pairwise ratios from the same class are clustered near the line $y = 1$, whereas hyper-pixels from different classes exhibit greater variation

[2]We use this data to define the training set for the AHI study in Section 5.3.

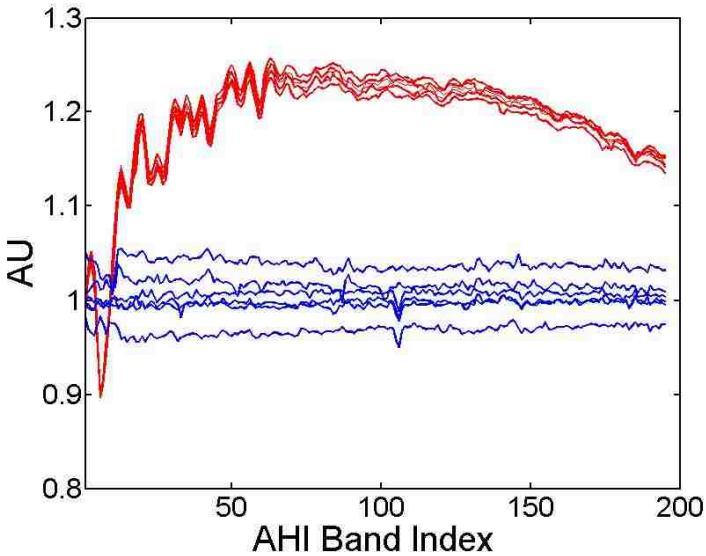


Figure 5.8: AHI training data. Red plots show a subset of hyper-pixels from the building class (B). Blue plots show band ratios between these hyper-pixels.

in their pairwise ratios.

From the plots in these figures it follows that variability of pairwise band ratios can play the same role for multi-color images as gray-scale discontinuity and similarity for gray-scale images. The objective is to define the pairwise ratio edge signatures using a small number of ratios ($S \ll K$) that can discriminate the edges between the different class types. Roughly speaking, we seek diagonal elements of the spectral ratio index A/B that best capture the "shape" of its diagonal.

One fairly straightforward approach that performs well in practice, is to use the maximum and the minimum pairwise ratios, i.e., set

$$\mathcal{E}_{\mathrm{AB}} = \{(k_{\max}, k_{\max}, \rho_{k_{\max}}), (k_{\min}, k_{\min}, \rho_{k_{min}})\} \ ,$$

where k_{min} and k_{max} are the band indices corresponding to the minimum and maximum diagonal elements of A/B, respectively. Note that for consistency with the

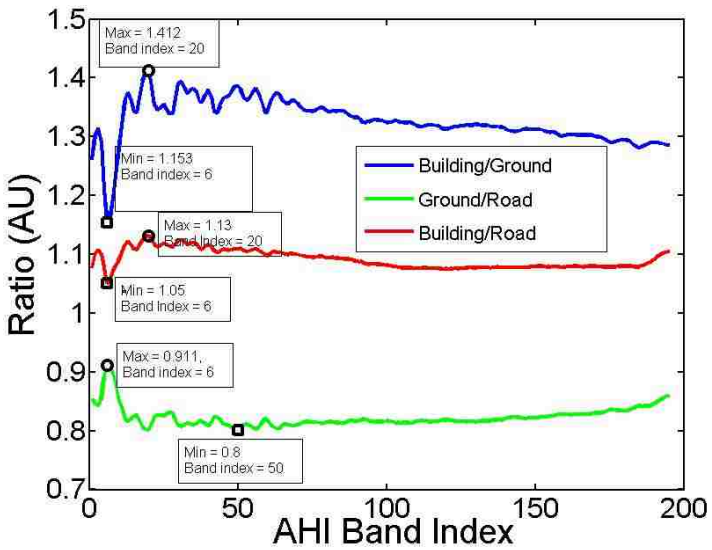


Figure 5.9: AHI training data: band ratios of class-average hyper-pixels.

notation in (5.6) both bands forming the ratio are stated explicitly even though they are the same.

The associated tolerances can be determined in several different ways. A good estimate can be obtained by computing the spread of the min and max ratios for the training data. For example, we can set

$$\Delta_{\min} = \max_{i,j} \frac{a_{k_{\min}}(i,j)}{b_{k_{\min}}(i,j)} - \min_{i,j} \frac{a_{k_{\min}}(i,j)}{b_{k_{\min}}(i,j)} \,;$$

$$\Delta_{\max} = \max_{i,j} \frac{a_{k_{\max}}(i,j)}{b_{k_{\max}}(i,j)} - \min_{i,j} \frac{a_{k_{\max}}(i,j)}{b_{k_{\max}}(i,j)} \,;$$

and then define

$$\epsilon_{\min} = \tau \Delta_{\min} \quad \text{and} \quad \epsilon_{\max} = \tau \Delta_{\max}$$

for some positive τ.

Application of the min-max pairwise selection strategy to the AHI training data shown in Fig. 5.7, yields the following set of edge signatures for the edges between

the B, G, and R classes:

$$\mathcal{E}_{\mathrm{BG}} = \{(6, 6, 1.153), (20, 20, 1.412)\} \tag{5.16}$$

$$\mathcal{E}_{\mathrm{BR}} = \{(6, 6, 1.050), (20, 20, 1.130)\} \tag{5.17}$$

$$\mathcal{E}_{\mathrm{RG}} = \{(50, 50, 0.800), (6, 6, 0.911)\} \tag{5.18}$$

From Fig. 5.9 we can see that $\mathcal{E}_{\mathrm{BG}}$, $\mathcal{E}_{\mathrm{BR}}$, and $\mathcal{E}_{\mathrm{RG}}$ capture distinctive shape characteristics of the pairwise ratios for these classes.

The 2-band pairwise min-max signatures defined in (5.16)–(5.18) are the simplest possible that capture the relevant variation in the band ratios. More complex pairwise signatures can be created by including additional band ratios in a recursive manner as follows. Starting from a basic 2-band min-max signature we can obtain a 3-band signature by adding the band corresponding to the next maximum ratio. Alternatively, we can obtain a 4-band signature by adding 2 bands for the next maximum *and* minimum ratios, or another 3-band signature by adding the band for the next minimum ratio. This process can be repeated until a satisfactory signature is obtained.

The comparative study in Section 5.3 shows that in the absence of noise performance of the simple edge signatures in (5.16)–(5.18) is comparable to that of MCG-based feature extraction. Our study also shows that sensitivity to noise can be further reduced by increasing the number of pairwise band ratios in the signature, i.e., the number $\widetilde{S}$ in (5.14).

Because the choice of ratios in the pairwise strategy is limited to the diagonal elements of the spectral ratio index A/B in (5.5), it is clear that in some cases this strategy may not perform as well as a more general strategy that allows one to choose from all possible ratio combinations in A/B. For example, if two materials have hyper-pixels that are translations of each other, i.e., differ by a constant intensity

factor, then the pairwise strategy will fail if the data is normalized. Likewise, a "weak" edge will have pairwise ratio close to one and so, its pairwise signature will be close to that of identical materials. Consequently, the pairwise approach will tend to miss weak edges. An alternative ratio selection strategy that is not restricted to the diagonal elements of A/B is presented in the next section.

5.2.4 Cross-ratio edge index

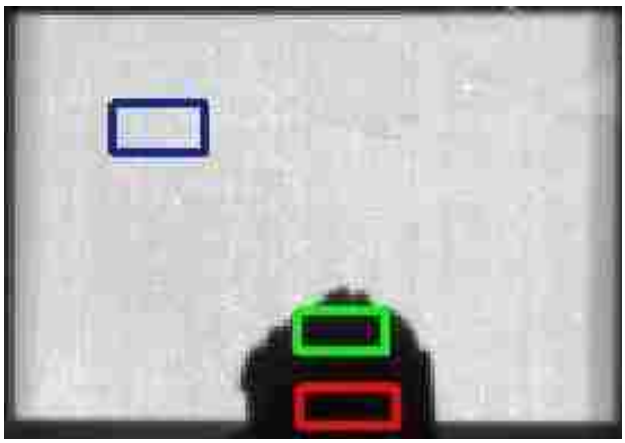


Figure 5.10: DWELL FPA training data: background (B), limestone (L) and granite (G) identified by blue, red and green boxes, respectively. The image corresponds to a bias voltage of 0.7 V.

In this section we propose a general cross-ratio strategy for determining the edge signature $\mathcal{E}_{\mathrm{AB}}$. In this strategy the bands for each ratio are selected in a two-step process and may be different, thus the term "cross-ratio." The key distinction between the cross-ratio approach and the pairwise approach in the last section is that now the choice of the ratios in the edge signature is not limited to the diagonal elements

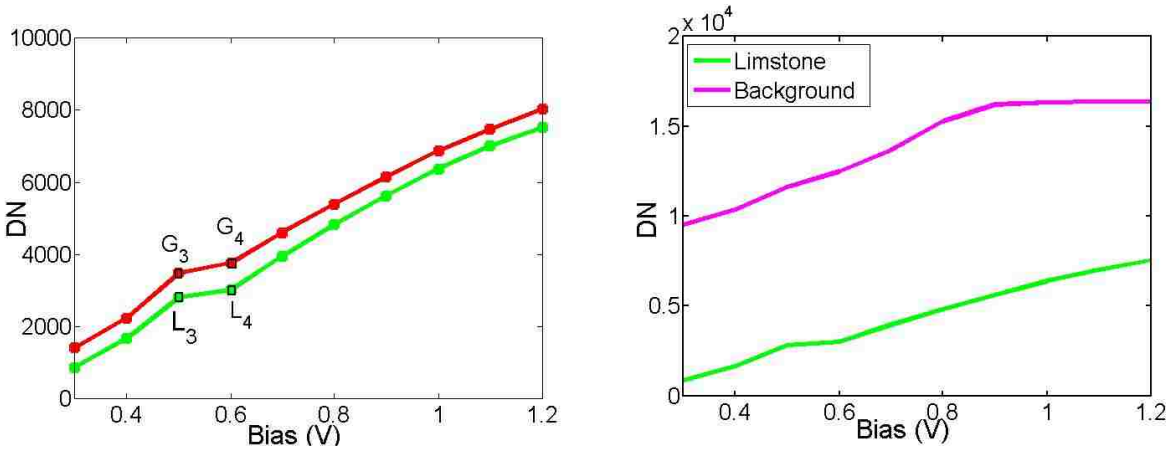


Figure 5.11: Class-averages for granite vs. limestone (left plot) and background vs. limestone (right plot).

of the spectral contrast matrix

$$\mathrm{A/B} = \begin{pmatrix} \frac{a_1}{b_1} & \frac{a_1}{b_2} & \cdots & \frac{a_1}{b_K} \\ \frac{a_2}{b_1} & \frac{a_2}{b_2} & \cdots & \frac{a_2}{b_K} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{a_K}{b_1} & \frac{a_K}{b_2} & \cdots & \frac{a_K}{b_K} \end{pmatrix}.$$

To motivate this strategy we use a sample of actual DWELL FPA data for a scene that contains both weak and strong edges. The DWELL image, shown in Fig. 5.10 was taken at bias voltage 0.7 V and has three different types of materials: limestone (L), granite (G) and background (B). Visual inspection of Fig. 5.10 reveals a very weak edge between the L and the G classes and a strong edge between the G and the B classes. Comparison of class-average hyper-pixels in Fig. 5.11 quantifies this observation. The representative hyper-pixels $\bar{\mathbf{g}}$ and $\bar{\mathbf{l}}$ for the G and the L classes, shown on the left, are nearly parallel and very close to each other, which means that for the DWELL FPA the two classes are nearly undistinguishable. Consequently, we

can write

$$\bar{g}_i \approx \bar{l}_i + c \quad \text{for all bias voltages (bands) } v_i \,,$$

for some constant C. Taking the difference of two pairwise ratios for bands i and j gives

$$\frac{\bar{g}_i}{l_i} - \frac{\bar{g}_j}{\bar{l}_j} \approx \frac{\bar{l}_i + c}{\bar{l}_i} - \frac{\bar{l}_j + c}{\bar{l}_j} = c \left(\frac{1}{\bar{l}_i} - \frac{1}{\bar{l}_j} \right) .$$

Because of the magnitudes of $\bar{l}_i$ the right hand-side above is small, i.e., the pairwise ratios of granite vs. limestone will have almost no significant variation across the bands. This can be viewed as an alternative characterization of a "weak" edge in terms of pairwise band ratios. On the other hand, the right plot in Fig. 5.10 shows high spectral contrast between the L and the G classes.

Motivated by this analysis we propose the following two-step procedure, which gives rise to edge signatures that enhance weak edges. Assume two materials A and B with representative hyper-pixels $\mathbf{a}$ and $\mathbf{b}$, respectively. Let $1 < S \leq K$ be a fixed integer. The case $S = 1$ obviously gives a pairwise ratio and for this reason will not be considered below.

The first step in the proposed strategy is to select the S bands $\{i_1, \ldots, i_S\}$ where the classes A and B have maximum separation. Given representative hyper-pixels $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ for A and B, respectively, we set

$$i_1 = \arg \max_{1 \leq i \leq K} |\bar{a}_i - \bar{b}_i| \,;$$

$$i_2 = \arg \max_{1 \leq i \leq K,\, i \neq i_1} |\bar{a}_i - \bar{b}_i| \,;$$

$$i_3 = \arg \max_{1 \leq i \leq K,\, i \neq i_1, i_2} |\bar{a}_i - \bar{b}_i| \,;$$

and so on. After the S bands $\{i_1, \ldots, i_S\}$ have been determined we proceed to compute the spectral ratios using *all possible* band combinations:

$$\rho_{pq} = \frac{\bar{a}_{i_p}}{\bar{b}_{i_q}} \,, \quad 1 \leq p, q \leq S \,.$$

Without loss of generality we may assume that all ratios are less than or equal to one; if $\rho_{pq} > 1$ for some p and q we simply replace it by its reciprocal.

At the second step we define $\mathcal{E}_{\mathrm{AB}}$ by selecting ratios that exhibit the strongest spectral contrast between the classes. To rank the ratios according to their spectral contrast we note that owing to the assumption that all $\rho_{pq} \leq 1$, the ratios closest to zero correspond to the strongest spectral contrast between any two bands. Thus we select the first pair of bands, $\{p_1, q_1\}$, as the pair corresponding to the smallest ratio:

$$\rho_1 = \rho_{p_1 q_1} = \arg \min_{1 \leq p,q \leq S} \rho_{pq} \,,$$

the second pair of bands $\{p_2, q_2\}$ as the pair corresponding to the next smallest ratio,

$$\rho_2 = \rho_{p_2 q_2} = \arg \min_{1 \leq p,q \leq S,\, p \neq p_1, q \neq q_1} \rho_{pq} \,,$$

and so on. To define the edge signature we choose the first R ratios:

$$\mathcal{E}_{\mathrm{AB}} = \{(p_1, q_1, \rho_1), \ldots, (p_R, q_R, \rho_R)\} \,. \tag{5.19}$$

To illustrate this selection criterion we apply it to define an edge signature $\mathcal{E}_{\mathrm{LG}}$ for the edge E_{LG} between limestone and granite from the DWELL training data shown in Fig. 5.10. For simplicity we choose the smallest possible number of bands ($S = 2$) for this approach. It is easy to see that

$$i_1 = 3 \quad \text{and} \quad i_2 = 4 \,,$$

as shown in Fig. 5.11. The two maximum separation bands in this case are consequitive because the class-averages $\bar{\mathbf{l}}$ and $\bar{\mathbf{g}}$ are monotone as functions of the band index; this may not be the case in general. The four possible band combinations yield the following four L/G ratios:

$$\rho_{33} = 0.8063, \quad \rho_{34} = 0.74502, \quad \rho_{43} = 0.8651, \quad \text{and} \quad \rho_{44} = 0.7994 \,.$$

The ranking of the ratios according to the strength of their spectral contrast gives

$$\rho_1 = \rho_{34}, \quad \rho_2 = \rho_{44}, \quad \rho_3 = \rho_{33}, \quad \text{and} \quad \rho_4 = \rho_{43}\,.$$

Using one ratio ($R = 1$) gives the following edge signature:

$$\mathcal{E}_{HG} = \{(3, 4, 0.74502)\} \tag{5.20}$$

for the edge between limestone and granite. The signature (5.20) is used in the DWELL studies in Section 5.3.

5.2.5 Implementation of the SRC mask

In this section we describe the implementation of the joint spatio-spectral mask $\mathcal{K}_{\text{AB}}$ used in the verification studies in Section 5.3. Selection of bands for the ratios in an edge signature (5.6) was already discussed in detail in Sections 5.2.3–5.2.4. In this section we focus on the fusing of the edge signature with a specific spatial mask $\mathcal{M}$.

For simplicity, we use the 3×3 spatial mask $\mathcal{M}$ shown in Fig. 5.3. Recall that for this mask $M = 4$, i.e., $\mathcal{M}$ has four distinct pixel pairs given by

$$\begin{aligned}
\{u^{1-}(i,j), u^{1+}(i,j)\} &\rightarrow \{u(i,j-1), u(i,j+1)\} \\
\{u^{2-}(i,j), u^{2+}(i,j)\} &\rightarrow \{u(i-1,j), u(i+1,j)\} \\
\{u^{3-}(i,j), u^{3+}(i,j)\} &\rightarrow \{u(i-1,j-1), u(i+1,j+1)\} \\
\{u^{4-}(i,j), u^{4+}(i,j)\} &\rightarrow \{u(i-1,j+1), u(i+1,j-1)\}
\end{aligned}$$

Let $\mathcal{E}_{\text{AB}}$ be an edge signature with length $S > 0$. Using $\mathcal{E}_{\text{AB}}$ in conjunction with the mask $\mathcal{M}$ yields a joint spatio-spectral mask $\mathcal{K}_{\text{AB}}$

$$\mathcal{K}_{\text{AB}} : \mathbf{R}^{I \times J \times K} \mapsto \mathbf{R}^{I \times J \times (4 \times S)} \tag{5.21}$$

whose response at a hyper-pixel $\mathbf{u}(i,j)$ is given by the $4 \times S$ matrix

$$\mathcal{K}_{\mathrm{AB}}(i,j) = \begin{pmatrix} \dfrac{u_{p_1}^{1-}(i,j)}{u_{q_1}^{1+}(i,j)} & \dfrac{u_{p_2}^{1-}(i,j)}{u_{q_2}^{1+}(i,j)} & \cdots & \dfrac{u_{p_s}^{1-}(i,j)}{u_{q_s}^{1+}(i,j)} \\ \dfrac{u_{p_1}^{2-}(i,j)}{u_{q_1}^{2+}(i,j)} & \dfrac{u_{p_2}^{2-}(i,j)}{u_{q_2}^{2+}(i,j)} & \cdots & \dfrac{u_{p_s}^{2-}(i,j)}{u_{q_s}^{2+}(i,j)} \\ \dfrac{u_{p_1}^{3-}(i,j)}{u_{q_1}^{3+}(i,j)} & \dfrac{u_{p_2}^{3-}(i,j)}{u_{q_2}^{3+}(i,j)} & \cdots & \dfrac{u_{p_s}^{3-}(i,j)}{u_{q_s}^{3+}(i,j)} \\ \dfrac{u_{p_1}^{4-}(i,j)}{u_{q_1}^{4+}(i,j)} & \dfrac{u_{p_2}^{4-}(i,j)}{u_{q_2}^{4+}(i,j)} & \cdots & \dfrac{u_{p_s}^{4-}(i,j)}{u_{q_s}^{4+}(i,j)} \end{pmatrix}$$

as defined in (5.11). The associated indicator matrix $\chi(i,j)$ defined from this output according to (5.13) measures the strength of the edges along the horizontal, vertical and two diagonal image axes.

Definition of (5.21) can be easily modified by adding or removing pixel pairs. For example, a simpler definition of $\mathcal{K}_{\mathrm{AB}}$ results from using only the first two pixel pairs $\{u^{1-}(i,j), u^{1+}(i,j)\}$ and $\{u^{2-}(i,j), u^{2+}(i,j)\}$ from $\mathcal{M}$. In this case $M = 2$, $\mathcal{K}_{\mathrm{AB}} : \mathbf{R}^{I \times J \times K} \mapsto \mathbf{R}^{I \times J \times (2 \times S)}$ and the response of $\mathcal{K}_{\mathrm{AB}}$ at a hyper-pixel $\mathbf{u}(i,j)$ is given by the $2 \times S$ matrix

$$\mathcal{K}_{\mathrm{AB}}(i,j) = \begin{pmatrix} \dfrac{u_{p_1}^{1-}(i,j)}{u_{q_1}^{1+}(i,j)} & \dfrac{u_{p_2}^{1-}(i,j)}{u_{q_2}^{1+}(i,j)} & \cdots & \dfrac{u_{p_s}^{1-}(i,j)}{u_{q_s}^{1+}(i,j)} \\ \dfrac{u_{p_1}^{2-}(i,j)}{u_{q_1}^{2+}(i,j)} & \dfrac{u_{p_2}^{2-}(i,j)}{u_{q_2}^{2+}(i,j)} & \cdots & \dfrac{u_{p_s}^{2-}(i,j)}{u_{q_s}^{2+}(i,j)} \end{pmatrix}.$$

The indicator matrix derived from this response measures the strength of the edges only along the horizontal and the vertical image axes.

5.3 Application of the SRC edge detection algorithm

The purpose of this section is to perform verification of the joint spatio-spectral SRC algorithm using real MS and HS imagery data. As a benchmark, we use edge maps obtained by the MCG edge detector. For definition of the MCG and its application to edge detection we refer to Appendix A. Also included for some data samples are edge maps obtained by the application of a standard gray-scale (GS) edge detector to individual HS or MS image bands. The study uses imagery from the AHI and DWELL FPA imagery acquired at the CHTM at UNM.

The objectives of the study are twofold. First, we aim to show that the new SRC algorithm, with either the pairwise ratio edge index described in Section 5.2.3 or the cross-ratio index from Section 5.2.4, is comparable with the benchmark MCG edge detector, and that in cases when the scene contains weak and strong edges, the SRC actually outperforms the MCG. We remind that such scenes represent challenge for magnitude-based edge detectors, which may require sophisticated locally adaptive threshold strategies to capture the "weak" edges. Our second objective is to compare and contrast the performance of the SRC with the pairwise and cross-ratio edge indices, especially for scenes that contain both weak and strong edges.

In our study we use both the raw sensor data, as well as normalized sensor data. Normalization of the DWELL FPA MS imagery and the AHI HS imagery in this section follows the same procedure as described in Chapter 3, Section 3.3.2.

5.3.1 Edge detection using AHI imagery

For this study we train the SRC algorithm using the AHI training data shown in Fig. 5.7. Recall that the scene contains three different classes: building (B), ground

Signature/Band	min	max	max 1	max 2	max 3
$\mathcal{E}_{\mathrm{BG}}$	6	20	31	51	64
$\mathcal{E}_{\mathrm{RG}}$	50	6	28	195	60

Table 5.3: Summary of the bands defining the pairwise ratio signatures for the edges between building and ground and ground and road classes using the AHI training data shown in Figure 5.12.

(G), and road (R). We first present results for the SRC algorithm with a pairwise ratio edge index determined according to the procedure described in Section 5.2.3.

SRC with pairwise ratio edge index

The SRC and the MCG edge detectors were tested on several configurations of the AHI imagery data. We first apply the algorithms to both normalized and non-normalized versions of the original AHI training and testing data sets. The experiments are then repeated for the AHI training and testing data perturbed with noise.

The pairwise ratio edge signatures for the SRC are determined as follows. We begin with a basic 2-band min-max signature and then add one by one the bands corresponding to the next three maximum ratios. This yields four different edge signatures with lengths ranging from 2 to 5. The bands used for the $\mathcal{E}_{\mathrm{BG}}$ (building and ground) and $\mathcal{E}_{\mathrm{RG}}$ (ground and road) edge signatures are summarized in Table 5.3. Figure 5.12 shows the corresponding pairwise ratios superimposed with the plots of the diagonal elements of the B/G and R/G matrices, defined in (5.5).

Figure 5.15 compares performance of the SRC and the MCG edge detectors for the

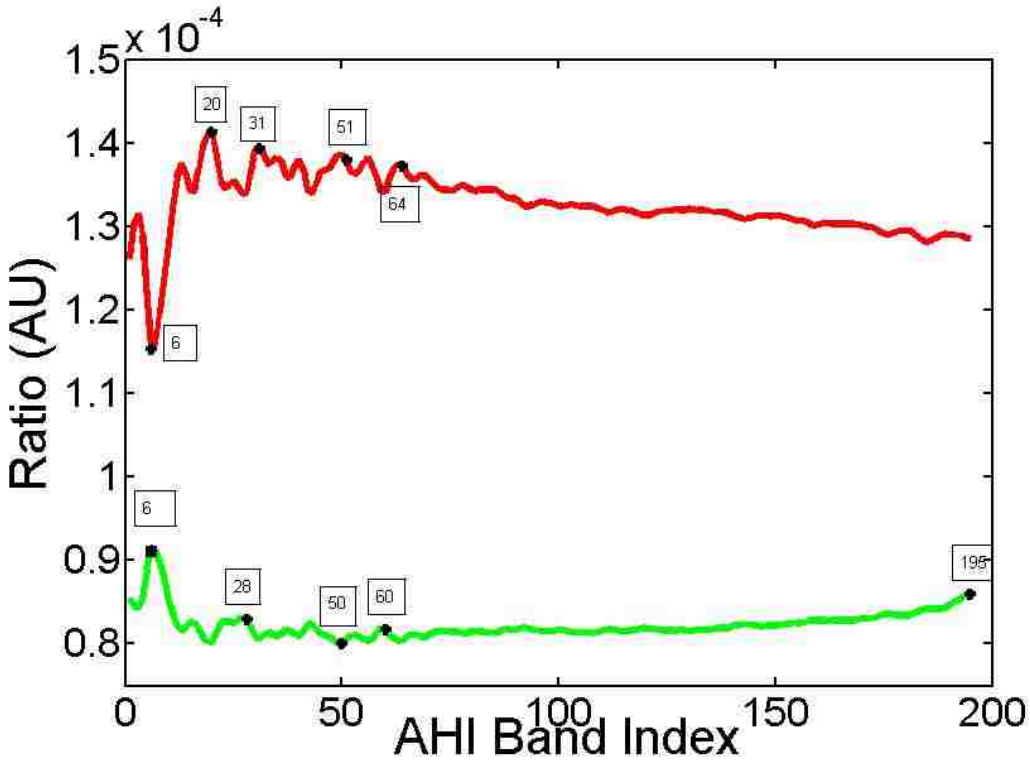


Figure 5.12: Pairwise ratio edge signatures $\mathcal{E}_{\mathrm{RG}}$ and $\mathcal{E}_{\mathrm{BG}}$ defined using the AHI training data: the green line shows the diagonal of the spectral ratio matrix R/G; the red line is the diagonal of the spectral ratio matrix B/G and the boxes show the band numbers used for $\mathcal{E}_{\mathrm{RG}}$ and $\mathcal{E}_{\mathrm{BG}}$.

AHI training data without the noise. In this particular setup, the same image is used both for training and testing the algorithms according to the following procedure. In the training phase, small uniform regions were identified for each class present in the scene, as shown in Fig. 5.7. Then, the trained SRC edge detector is tested over the entire scene using both raw and normalized AHI data. The first row in Fig. 5.15 shows the results obtained with the raw data and the second row shows the results with normalized AHI data. From the results presented in Fig. 5.15 we see that the SRC performs at least as well as the MCG algorithm. Moreover, when

the normalized data is used, the SRC gives slightly better edge maps compared to those obtained with the MCG.

Figure 5.16 compares the SRC and the MCG edge detectors for the unperturbed AHI testing data. The testing data is extracted from the same AHI test scene as the training set, but represents a different spatial region. In this study the SRC is applied with the same edge signatures as determined in the training phase. Figure 5.17 compares the edge maps derived by using the SRC with the 2-band (min,max) edge signature, the 5-band edge signature (min, max, max1, max2 and max3), and the MCG. From this figure we can conclude that in the absence of noise, the shortest and the longest edge signatures perform equally well.

Figure 5.18 compares the SRC with the 5-band signature, the MCG edge detector, and a gray-scale (GS) edge detector based on the standard gradient operator. The GS edge detector is applied to each one of the bands in the AHI testing image, and the best and the worst results are presented in Fig. 5.18. On the one hand, we see that it may be possible to find a band for which the standard GS operator will recover most of the essential spatial features in the image. On the other hand, we also see that if that single band is not carefully selected, then the standard GS algorithm may miss most of the important features in the edge map.

Our second study examines the performance of the SRC with different pairwise ratio edge signatures when the AHI testing data is perturbed with noise. The signal-to-noise ratio (SNR) used in this case is approximately 50dB, uniformly across all bands. The goal of this study is to corroborate the remark made at the end of Section 5.2.1 that the number of the ratios in the edge signature can be used to control the sensitivity of the SRC to the presence of noise. To this end, we apply the SRC starting with the shortest 2-band (min,max) signature and then gradually increase the length of the signature to 5. The bands from Table 5.3 are used to define a sequence of 4 edge signatures as follows: (min,max), (min,max,max1), (min,max,max1,max2) and

(min,max,max1,max2,max3).

Figure 5.19 clearly demonstrates that the sensitivity of the SRC algorithm to noise is noticeably reduced as the number of the bands in the edge signature is increased from 2 to 5. This conclusion is further confirmed by Figure 5.20, which compares the SRC with the 2-band and the 5-band edge signatures and the MCG edge detector.

One final and significant conclusion that can be gleaned from Fig. 5.15 and Fig. 5.16 is that the performance of the SRC does not deteriorate significantly when the AHI training and testing data are normalized. In contrast, the MCG edge map shows noticeable increase in the noise level for the normalized AHI data. This can be explained by noting that normalization smoothes the image manifold and so, the strength of the edges, as measured by their MCG values, decreases. At the same time, the spectral contrast, as measured by the band ratios, is affected to a lesser degree by the normalization, and in some cases may even improve. This behavior can be explained by noting that band ratios are almost invariant to band scaling.

SRC with cross-ratio edge index

In this series of experiments we apply the SRC with the cross-ratio edge signatures described in Section 5.2.4. We restrict attention to signatures with $S = 2$ and $R = 1$. Recall that S is the number of bands used and that R is the number of ratios selected from the set of all possible S^2 band ratios. Therefore, to define the edge signatures used in this section, we first select the two bands where the classes forming the edge exhibit the greatest separation. Then we form all possible $2^2 = 4$ ratios and choose the two bands that result in the strongest spectral contrast. The edge signatures obtained through this procedure contain a single ratio formed by these two, possibly different, bands.

Signature →	$\mathcal{E}_{\text{BG}}$		$\mathcal{E}_{\text{RG}}$		$\mathcal{E}_{\text{BR}}$	
Band ↓	raw	norm	raw	norm	raw	norm
band 1	50	6	50	6	20	20
band 2	20	7	49	7	19	19

Table 5.4: The cross-ratio edge signatures between classes B, G and R obtained by application of the strategy in Section 5.2.4 to the AHI training data shown in Figure 5.21.

We use this strategy to train the SRC for the AHI testing data shown in Fig. 5.7. The single ratio edge signatures obtained at the training phase are shown in Table 5.3. We note that none of the edge signatures in this table correspond to pairwise ratios, i.e., the ratios on the diagonals of the corresponding spectral contrast matrices. This means that edge signatures in Table 5.3 could not have been obtained by the pairwise strategy formulated in Section 5.2.3.

Figure 5.21 compares the edge maps of the SRC with the signatures in Table 5.4, and the MCG algorithm for normalized and non-normalized AHI training data. Results for the normalized and non-normalized AHI testing data are presented in Fig. 5.22. Except for the use of a different edge signature, the setup for these experiments is the same as before.

In the case of non-normalized AHI training and testing data we see that the SRC and the MCG produce virtually the same edge maps. When the data is normalized, the deterioration of the MCG edge map is much more noticeable than that of the SRC algorithm; this mirrors the earlier results shown in Fig. 5.15 and Fig. 5.16.

One important conclusion can be drawn from the results presented so far. By choosing bands with maximum separation and allowing unrestricted band combinations to form the ratios, we are able to define edge signatures with the minimal possible length which perform as well as the longer edge signatures whose ratios were restricted to the diagonal of the spectral contrast matrix. This is an important result because it enables potentially significant data compression with minimal loss of the ability to extract the spatial features.

5.3.2 Edge detection using DWELL FPA imagery

This section continues the verification of the SRC algorithm using data acquired with the DWELL FPA at the CHTM at the University of New Mexico. We use two DWELL FPA training and testing data sets shown in Fig. 5.13. The left plot in this figure is a representative image of the first data set which comprises of three distinct classes: background (B), granite (G) and limestone (L). The right plot in Fig. 5.13 shows a representative image of the second data set. The second data set contains four distinct classes: background (B), phyllite (P), granite (G) and limestone (L). Color photographs of the materials used in the study are shown in Fig. 5.14.

Following the procedure established in Section 5.3.1, the same images are used for the training and the testing of the SRC algorithm. Specifically, to train the SRC algorithm, small uniform regions were identified for each class present in the scene (the regions marked by the color boxes in Fig. 5.13), and used to determine the edge signatures. Then, the trained SRC edge detector is applied over the entire scene. As before, the benchmark edge maps are computed by the MCG edge detector.

The DWELL FPA images in Fig. 5.13 are characterized by the presence of strong edges between class B and the rest of the classes, and a weak edge between classes G and L. Therefore, our principal goal is to demonstrate the ability of the SRC

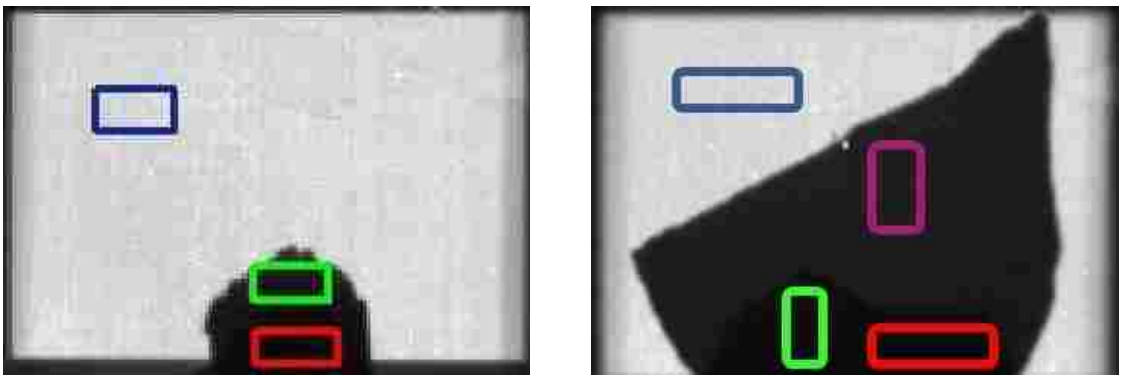


Figure 5.13: Images of the DWELL FPA training and testing data sets at bias voltage 0.7 V. The left image shows the first data set consisting of background (B), granite (G) and limestone (L) classes. The second data set consists of background (B), phyllite (P), granite (G) and limestone (L) classes, and is shown on the right. The DWELL FPA training data for the G, L, P and B classes is identified on the images by green, red, magenta and blue boxes, respectively.

algorithm to capture weak edges with the cross-ratio edge signatures. Accordingly, we begin with presentation of results for the SRC with cross-ratio signatures and then move on to discuss the performance of the SRC with the pairwise ratio edge signatures.

SRC with cross-ratio edge index

To obtain the cross-ratio edge signatures we follow the procedure described in Section 5.2.4. As in the case of the AHI data, we restrict attention to signatures with $S = 2$ and $R = 1$, where S denotes the number of maximally separated bands used to compute the ratios, and R is the number of ratios selected from all possible S^2 combinations. We remind the reader that in this approach the R ratios are selected according to the strength of their spectral contrast; thus, for $R = 1$ we select the

Figure 5.14: Photographs of the materials used in the DWELL study. Shown from left to right are the phyllite (P), granite (G), and limestone (L). Images courtesy of `www.geology.com`.

Signature →	$\mathcal{E}_{\mathrm{GB}}$		$\mathcal{E}_{\mathrm{LB}}$		$\mathcal{E}_{\mathrm{LG}}$	
Band ↓	raw	norm	raw	norm	raw	norm
band 1	6	1	6	1	3	9
band 2	7	10	7	10	4	10

Table 5.5: The cross-ratio edge signatures between the B, G, and L classes obtained by application of the strategy in Section 5.2.4 to the first DWELL FPA training data set, identified by the boxes on the left image in Fig. 5.13.

ratio with the strongest spectral contrast. Likewise, for each pair of classes we select the two bands where these classes exhibit the greatest separation.

The single ratio edge signatures for the two data sets obtained at the training phase are shown in Table 5.5 and Table 5.6, respectively. What is by now a familiar situation, we note that none of the edge signatures in these tables correspond to pairwise ratios, i.e., they could not have been obtained by the pairwise strategy

Signature →	$\mathcal{E}_{\text{PL}}$		$\mathcal{E}_{\text{PB}}$		$\mathcal{E}_{\text{PG}}$		$\mathcal{E}_{\text{LG}}$	
Band ↓	raw	norm	raw	norm	raw	norm	raw	norm
band 1	5	9	6	1	5	9	5	10
band 2	4	10	7	10	4	10	6	9

Table 5.6: The cross-ratio edge signatures between the B, G, L and P classes obtained by application of the strategy in Section 5.2.4 to the second DWELL FPA training data set, identified by the boxes on the right image in Fig. 5.13.

formulated in Section 5.2.3.

Results for the first DWELL FPA data set are shown in Fig. 5.23 and Fig. 5.24. The edge maps in the first figure are computed using the raw DWELL FPA data. From the plots in the top row we see that the SRC with the cross-ratio signatures in Table 5.5 is capable of recovering both the strong edges between the B, G and L classes, as well as the weak edge $\mathcal{E}_{\text{LG}}$ between classes G and L. This should be contrasted with the plots at the bottom row in Fig. 5.23 which show that the MCG edge detector picks the weak edge only after its tolerance is increased to a degree that results in significant noise levels in the edge map.

From the bottom row in Fig. 5.24 we see that the MCG edge map deteriorates even further when the normalized DWELL FPA data is used. In addition to the noise, the weak edge is now smeared compared to the one obtained using the raw data. The SRC edge map also experiences some degree of degradation; however, the weak edge $\mathcal{E}_{\text{LG}}$ continues to be clearly identifiable.

Results for the second DWELL FPA data set are presented in Fig. 5.25 and

Fig. 5.26. The second data set is in a sense more challenging because the two classes with the weak edge are now positioned against a phyllite backdrop that is less contrasting than the background class. Nonetheless, from the top rows in Fig. 5.25 and Fig. 5.26, it is clear that the SRC with the cross-ratio signatures in Table 5.6 recovers the strong edges *and* the weak edge between classes L and G.

Interestingly enough, Fig. 5.26 shows that for the normalized DWELL data the resolution of the weak edge by the SRC is slightly improved compared to the raw data case. The SRC edge map in Fig. 5.26 is more noisy than in Fig. 5.25, but this should be contrasted with the MCG edge map, shown at the bottom row of Fig. 5.26, in which the weak edge is significantly smeared.

Finally, in Fig. 5.28 we compare the SRC edge detector with three gray-scale edge detectors applied to the individual bands of the first DWELL FPA data set. The purpose of this experiment is to demonstrate that the straightforward application of gray-scale edge detectors to individual bands may fail to recover the complete edge map. For this study we use native *Matlab*$^{\text{TM}}$ implementations of the Canny, Sobel and Prewitt edge detectors and DWELL FPA bands at 0.5, 0.7, 0.9 and 1.1 V.

From Fig. 5.28 we see that neither one of the three gray-scale edge detectors was able to identify the very weak edge between granite and limestone classes. Moreover, Sobel and Prewitt edge detectors also fail to capture the relatively weak edge between phyllite and granite and limestone. This, and the previous examples further demonstrate the potential of the spectral contrast ratio concept for segmentation of MS and HS images.

SRC with pairwise ratio edge index

The goal is to show that the pairwise ratio edge signature, which is satisfactory for scenes with edges of about the same strength, may become inadequate for some

scenes with weak and strong edges. To this end, we use the more challenging, second set of DWELL FPA training and testing data shown on the right plot in Fig. 5.13.

Because we aim to demonstrate the failure of the pairwise ratio signatures to recover weak edges, for the purposes of this study we can reuse all but the $\mathcal{E}_{\mathrm{LG}}$ edge signature from Table 5.6. To define a pairwise $\mathcal{E}_{\mathrm{LG}}$ for the non-normalized raw DWELL FPA data we proceed as follows.

Recall that increasing the number of the ratios in the edge signature makes the SRC edge detector more restrictive. Because we are already dealing with a weak edge, it makes sense to consider pairwise signatures $\mathcal{E}_{\mathrm{LG}}$ with the shortest possible length $S = 1$, i.e., signatures consisting of a single ratio selected from the diagonal of the spectral contrast matrix L/G. Using the raw DWELL FPA training data in Fig. 5.13 we find that the diagonal of L/G is given by

$$(0.9385, 0.9451, 0.9549, 0.9496, 0.9547, 0.9597, 0.9727, 0.9824, 0.9867, 0.9890)\,.$$

The minimum and maximum pairwise ratios are achieved at bands 1 and 10, respectively. However, the maximum value 0.9890 has almost no spectral contrast. Therefore, we choose the minimum value corresponding to band 1. For the sake of comparison we include two more signatures obtained by using the bands from Table 5.6 in a pairwise fashion. To summarize, for the raw DWELL FPA testing data we reuse the edge signatures $\mathcal{E}_{\mathrm{PL}}$, $\mathcal{E}_{\mathrm{PB}}$, and $\mathcal{E}_{\mathrm{PG}}$ defined in Table 5.6, and redefine $\mathcal{E}_{\mathrm{LG}}$ as follows:

$$\mathcal{E}^1_{\mathrm{LG}} = \{1, 1, 0.9385\}\,; \quad \mathcal{E}^5_{\mathrm{LG}} = \{5, 5, 0.9547\}\,; \quad \mathcal{E}^6_{\mathrm{LG}} = \{6, 6, 0.9597\}\,. \tag{5.22}$$

Our results are shown in Fig. 5.27. The top row in this figure compares the edge maps for edge E_{LG} obtained by the SRC algorithm with the three pairwise signatures defined in (5.22), and the original cross-ratio signature $\mathcal{E}_{\mathrm{LG}}$ from Table 5.6. The bottom row in Fig. 5.27 shows the complete SRC edge maps corresponding

to the same choices of $\mathcal{E}_{\mathrm{LG}}$. We remind that all other edge signatures are exactly the same as in Table 5.6.

An immediate conclusion that can be drawn from the plots in Fig. 5.27 is that, in at least some cases, the pairwise ratio signatures fail to recover weak edges. This conclusion can be quantified by comparing the pairwise signatures in (5.22) with the original definition from Table 5.6. There, to determine $\mathcal{E}_{\mathrm{LG}}$ we formed the four possible ratios for bands 5 and 6

$$\rho_{55} = 0.9547; \quad \rho_{56} = 0.7577; \quad \rho_{65} = 1/0.8270; \quad \text{and} \quad \rho_{66} = 0.9597\,.$$

and selected the one[3] with the strongest spectral contrast, i.e., the ratio ρ_{56}. The resulting cross-ratio

$$\mathcal{E}_{\mathrm{LG}} = \{5, 6, 0.7577\}$$

has by far the strongest spectral contrast from all ratios used in our study. Because the smallest diagonal element of L/G is already greater than ρ_{56}, it follows that for this particular data set no pairwise ratio can provide the same spectral contrast as the cross-ratio based on bands 5 and 6.

5.4 Conclusions

In this chapter we developed and tested a novel joint spatio-spectral approach for image segmentation of MS and HS images. The SRC algorithm differs from the existing edge detection methodologies for such images in several important ways. Unlike the multi-color gradient (MCG) [1] or the morphological color gradient (MoCG) [27] approaches, the SRC algorithm is *not an extension* of an existing gray-scale processing methodology, instead it is designed from the onset as a *dedicated* MS/HS

[3]We write ρ_{65} in terms of its reciprocal because the spectral contrast of each ratio is estimated by taking the smaller of the original value and its reciprocal.

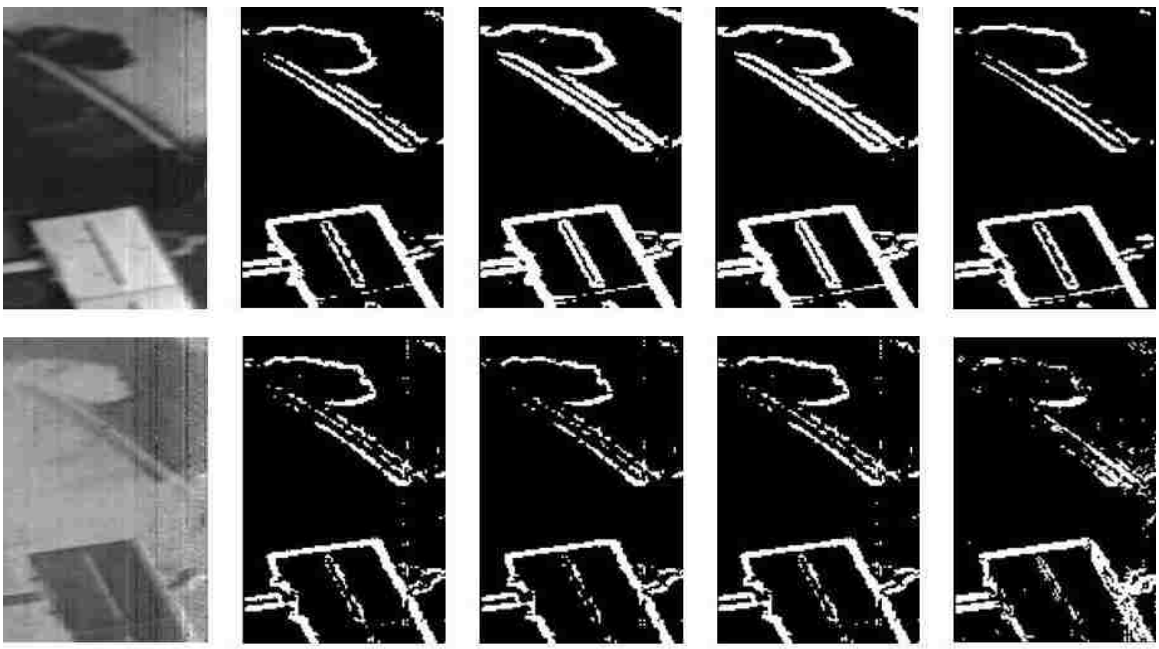
edge detection algorithm. In particular, the SRC algorithm fuses a given spatial mask with the spectral band ratios from a given edge signature into a non-separable, three-dimensional spatio-spectral mask.

Another equally important distinction is that our approach utilizes the concept of the spectral ratio contrast to estimate similarity and discontinuity in a given HS/MS image, rather than measuring the rate of change in the image, as in the MCG and MoCG algorithms. Among other things, by requiring only a few bands per edge signature, the use of spectral ratios in the SRC enables unprecedented levels of data compression at the feature extraction stage. This aspect of the SRC approach is particularly attractive for HS images, for which the number of bands can be as high as several hundred. One final distinction between the SRC and the approaches reviewed in Section 5.1 of this chapter, is that the SRC is a two-stage procedure comprising of a training phase and a feature extraction phase.

In Section 5.3 we carried out an extensive verification of the SRC algorithm using real HS and MS imagery collected by the AHI and the DWELL FPA sensors, respectively, and the MCG edge detector as a benchmark. For moderately difficult scenes in which the edges are of approximately the same strength, as measured by their MCG values, the SRC and the MCG edge detectors generate essentially identical edge maps. However, for more challenging imagery containing both "weak" and "strong" edges, the SRC edge detector outperforms the MCG edge detector by a wide margin. This provides a strong validation of the spectral ratio contrast concept by showing that with a suitably defined procedure to define the edge signatures, band ratios can reliably discriminate weak edges from the background noise and other spatial features in the image.

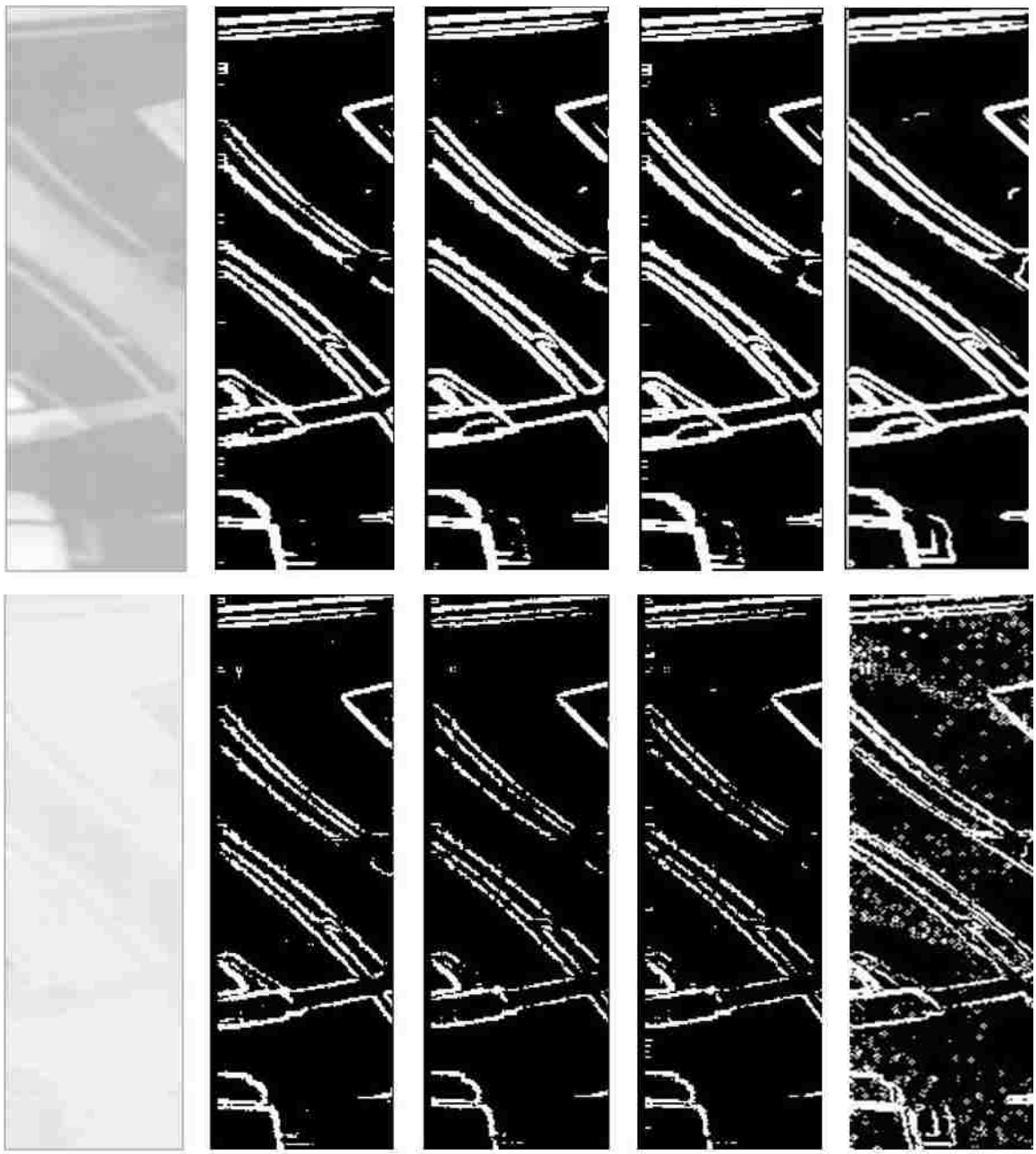
Last but not least, our study provides further validation of the fact that band-wise application of standard gray-scale edge detectors to MS and HS images may fail to produce complete edge maps due to the inability of intensity-based processing to

identify iso-luminant edges.



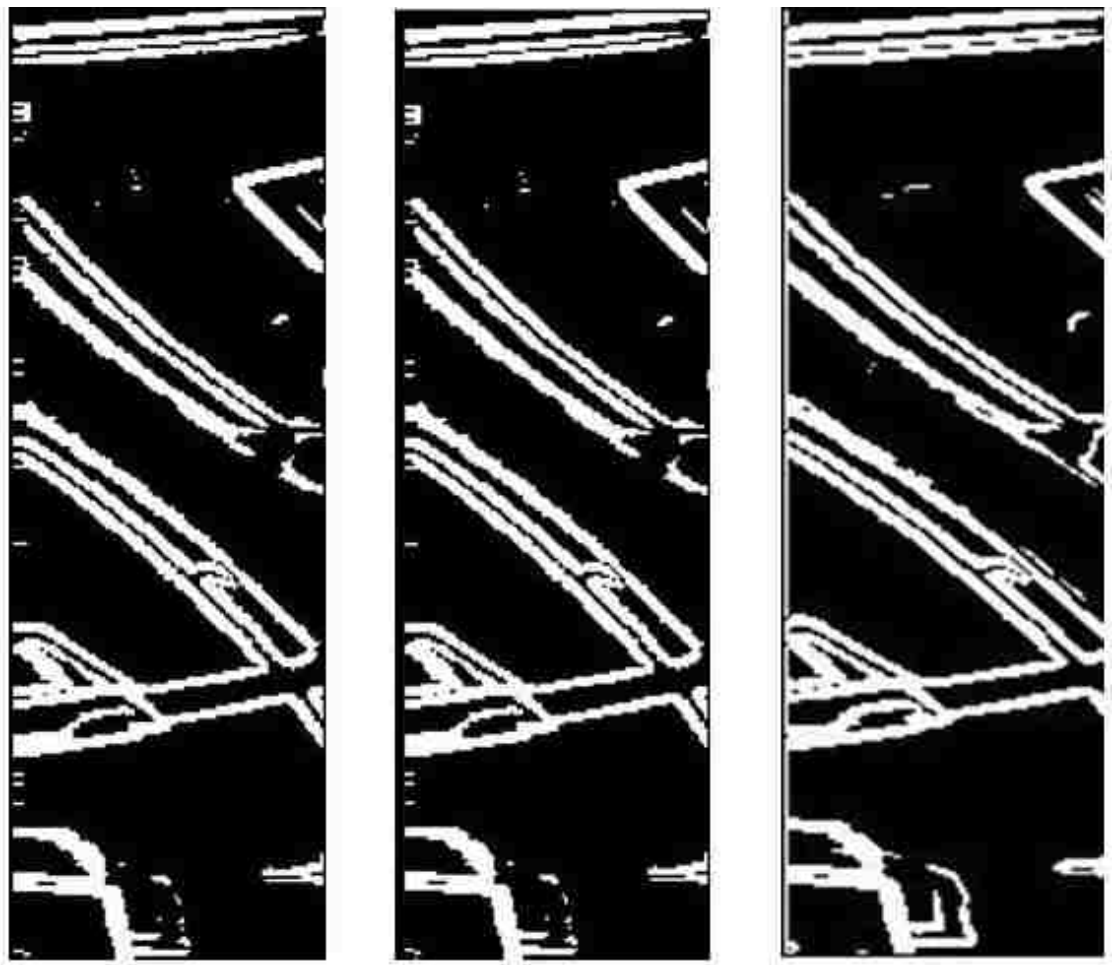
(i) (ii) (iii) (iv) (v)

Figure 5.15: Comparison between the SRC with the 2-band min-max signature and the MCG edge detectors for raw (top row) and normalized (bottom row) AHI training data: (i) AHI training image at band 10; (ii) SRC edges E_{BG}; (iii) SRC edges E_{RG}; (iv) combined SRC edge map; (v) MCG edge map.



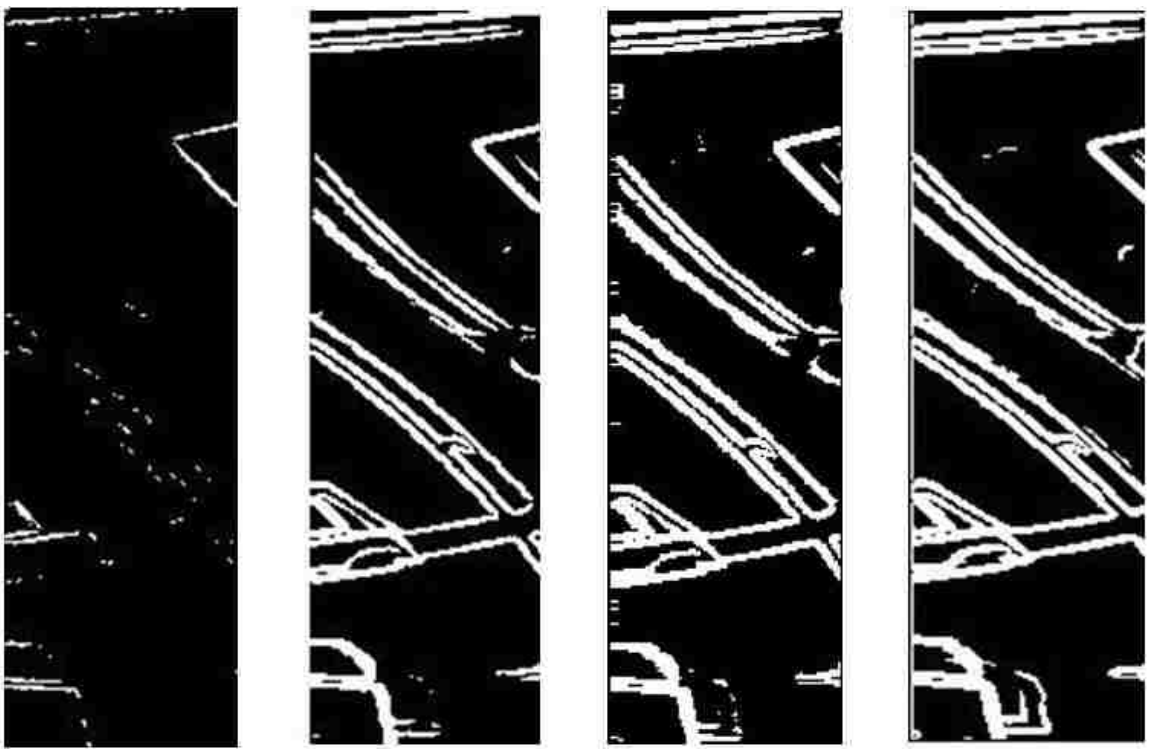
(i) (ii) (iii) (iv) (v)

Figure 5.16: Comparison between the SRC with the 2-band min-max signature and the MCG edge detectors for raw (top row) and normalized (bottom row) AHI testing data: (i) AHI test image at band 10; (ii) SRC edges E_{BG}; (iii) SRC edges E_{RG}; (iv) combined SRC edge map; (v) MCG edge map.



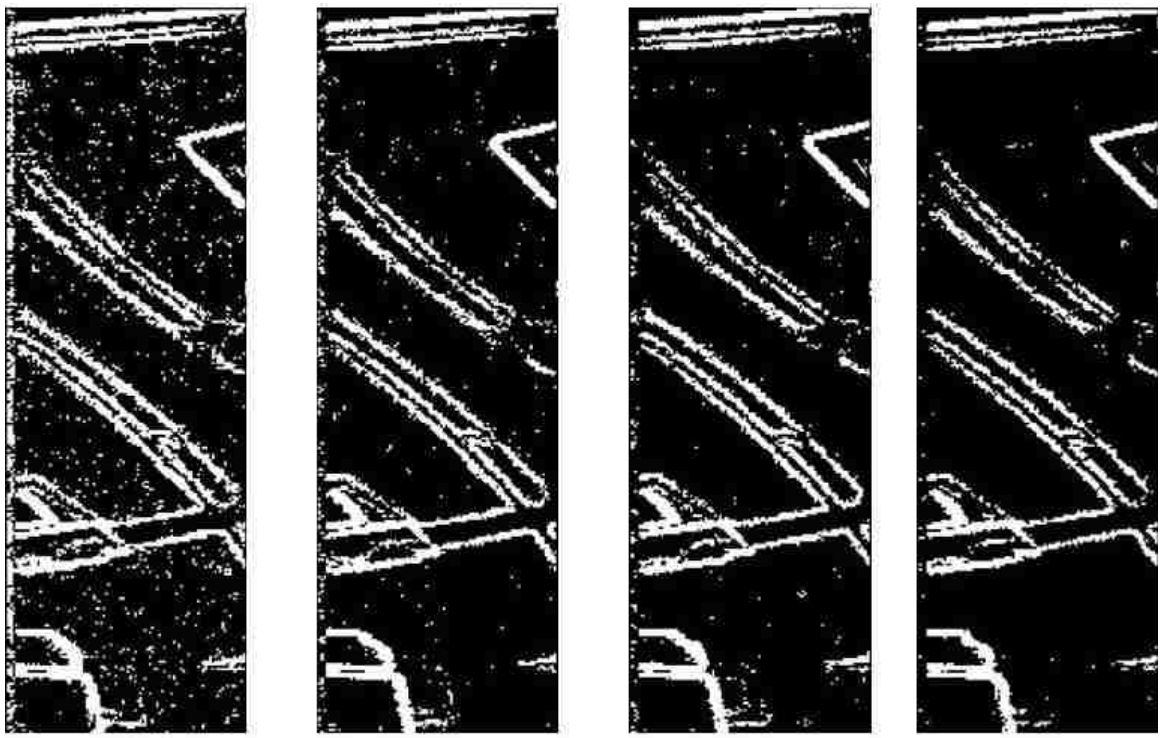
(i) (ii) (iii)

Figure 5.17: Comparison of the SRC and the MCG edge detectors using two different edge signatures in SRC: (i) SRC edges with the 2-band min-max ratio signature; (ii) SRC edge map with the 5-band signature defined by taking all 5 bands in Table 5.3; (iii) MCG edge map.



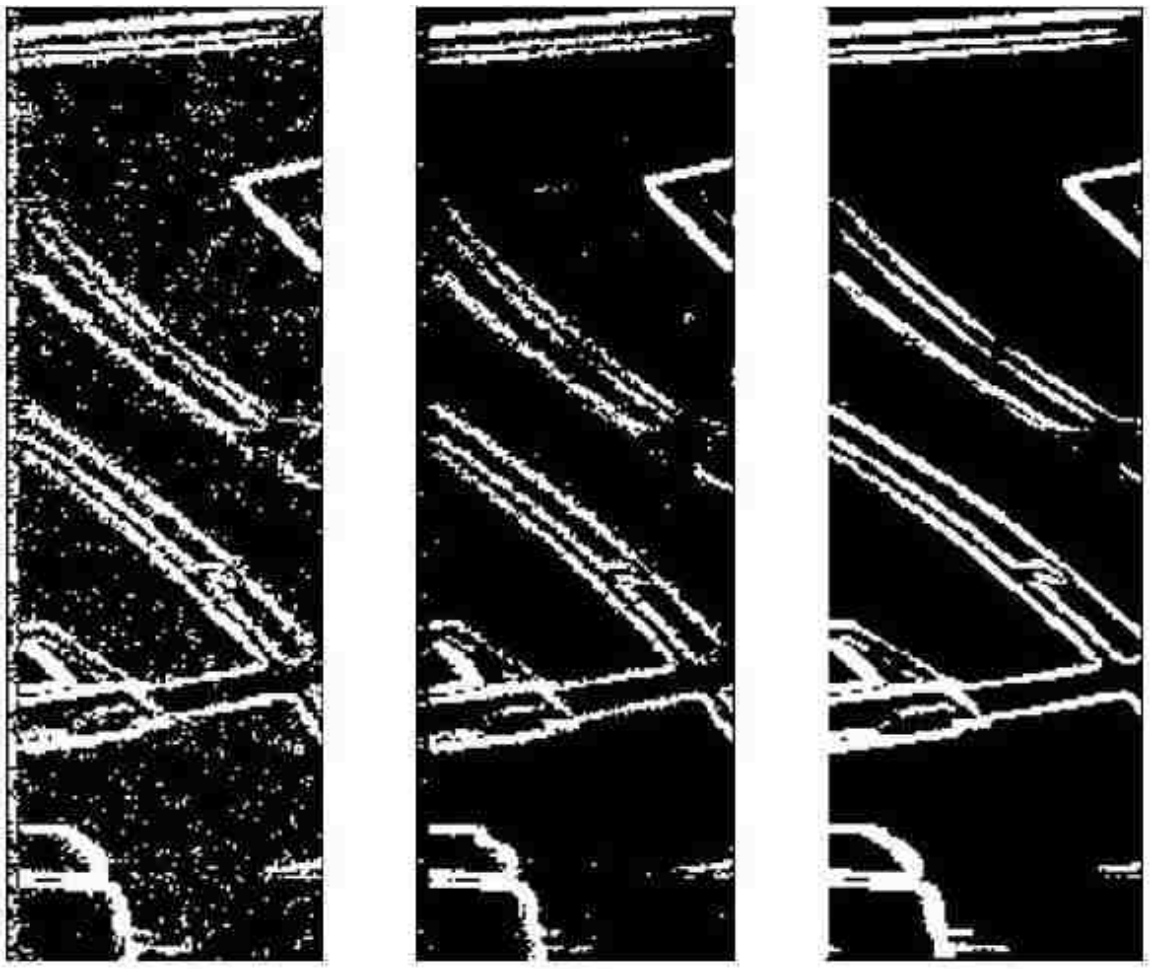
(i) (ii) (iii) (iv)

Figure 5.18: Comparison between the Sobel [4] edge detector applied band-wise and the SRC with the 5-band signature defined by taking all 5 bands in Table 5.3, and the MCG edge detectors: (i) Sobel edge map for the worst performing band; (ii) Sobel edge map for the best performing band; (iii) SRC edge map; (iv) MCG edge map.



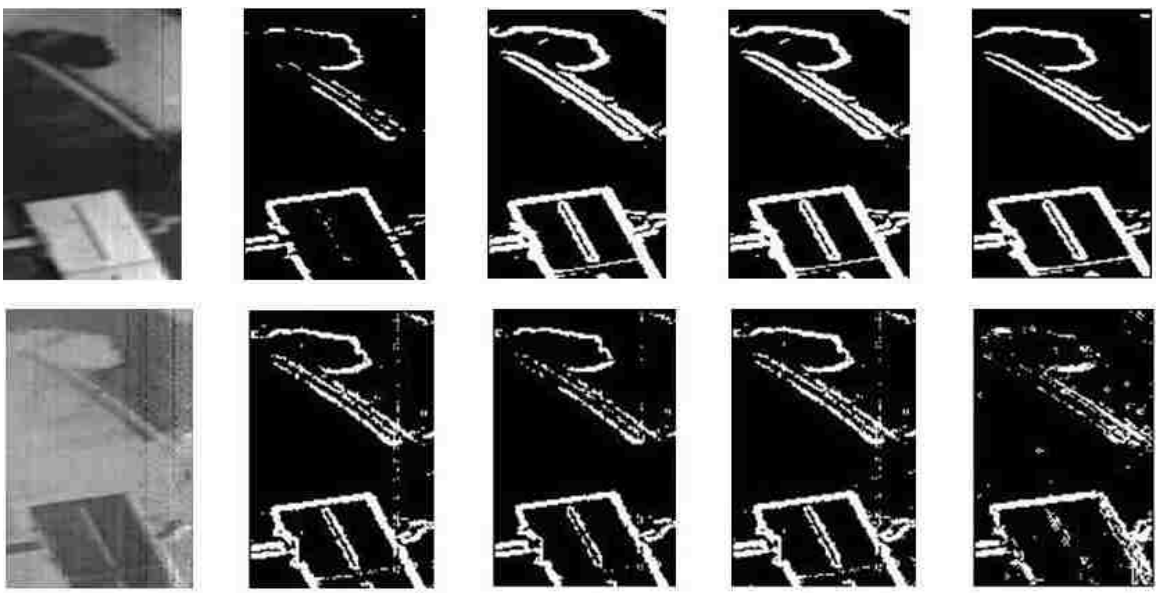
(i) (ii) (iii) (iv)

Figure 5.19: Progression in the performance of the SRC edge detector for noisy AHI testing data with an SNR of 50dB, when the number of the pairwise ratios in the edge signatures is gradually increased: (i) SRC edge map with 2-band min-max signature; (ii) SRC edge map with a 3-band signature defined by taking the first 3 bands in Table 5.3; (iii) SRC edge map with a 4-band signature defined by taking the first 4 bands in Table 5.3; (iv) SRC edge map with a 5-band signature defined by taking all 5 bands in Table 5.3.



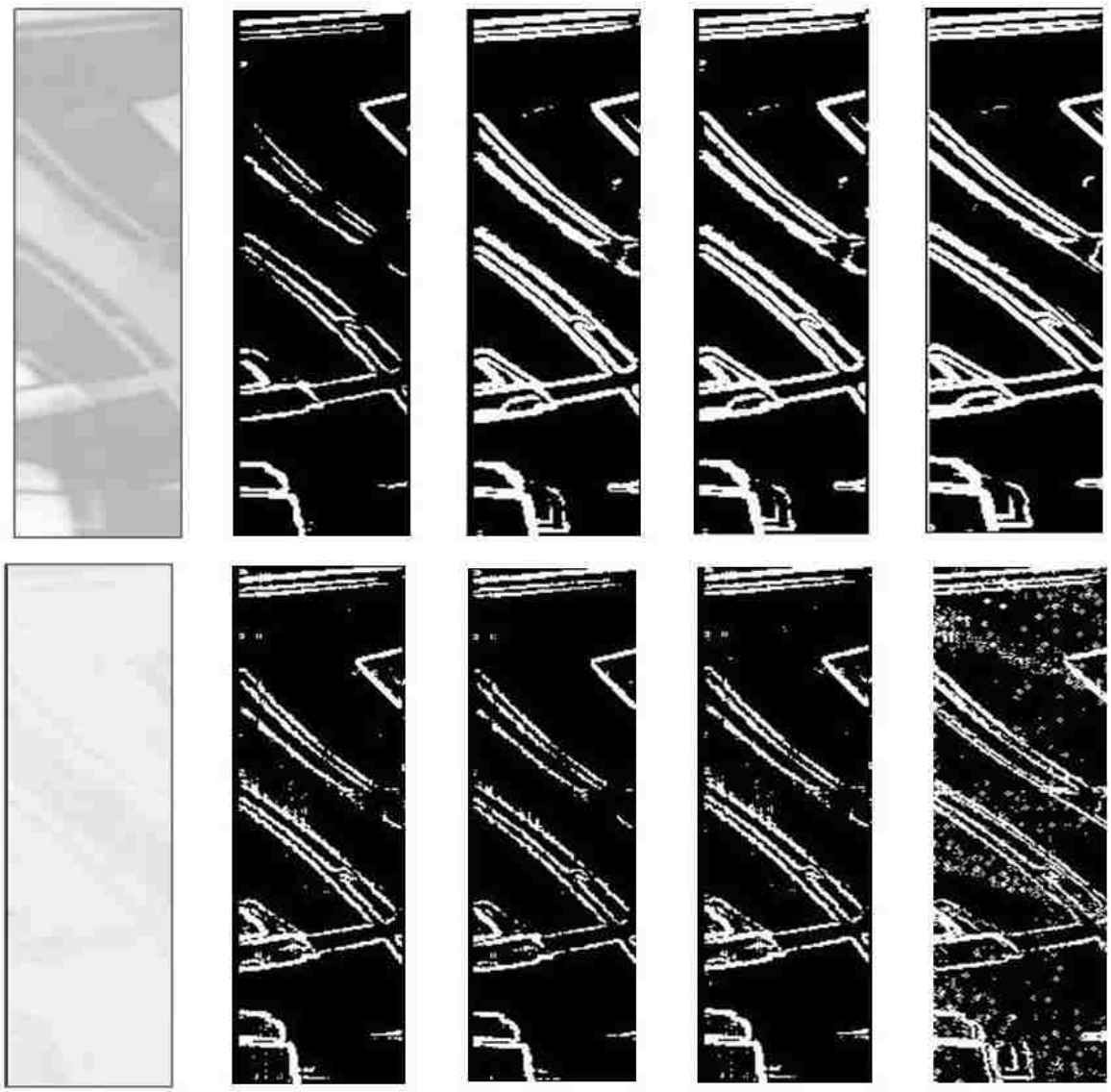
(i) (ii) (iii)

Figure 5.20: Comparison between the SRC and the MCG edge detectors for noisy AHI testing data with an SNR of 50dB: (i) SRC edge map with the 2-band min-max signature; (ii) SRC edge map with the 5-band signature defined by taking all 5 bands in Table 5.3; (iii) MCG edge map.



(i) (ii) (iii) (iv) (v)

Figure 5.21: Comparison between the SRC edge detector with the cross-ratio signatures defined in Table 5.4, and the MCG edge detector for raw (top row) and normalized (bottom row) AHI training data: (i) AHI training image at band 10; (ii) edges E_{BG}; (iii) edges E_{RG}; (iv) Combined SRC edge map; (v) MCG edge map.



(i) (ii) (iii) (iv) (v)

Figure 5.22: Comparison between the SRC edge detector with the cross-ratio signatures defined in Table 5.4, and the MCG edge detector for raw (top row) and normalized (bottom row) AHI testing data: (i) AHI test image at band 10; (ii) edges E_{BG}; (iii) edges E_{RG}; (iv) Combined SRC edge map; (v) MCG edge map.

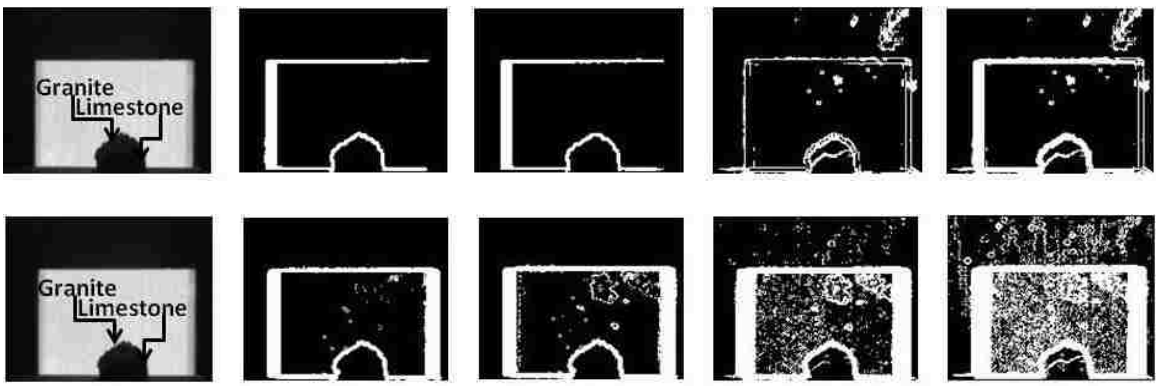
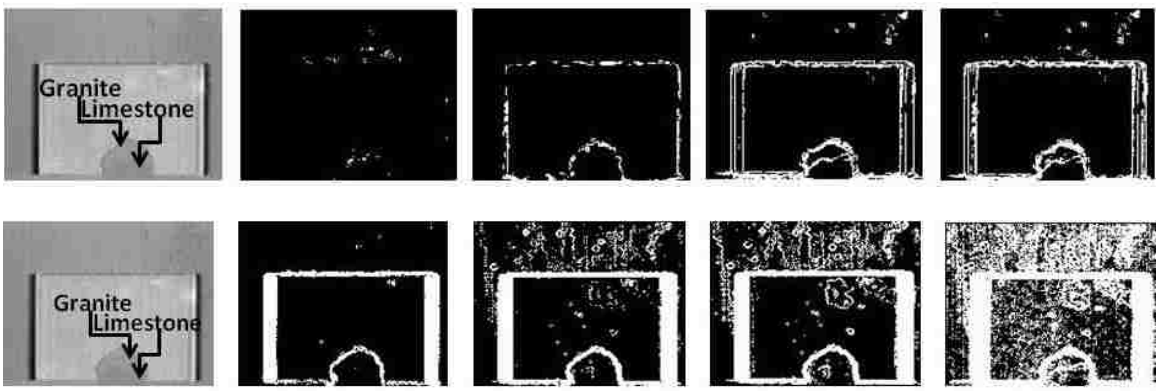
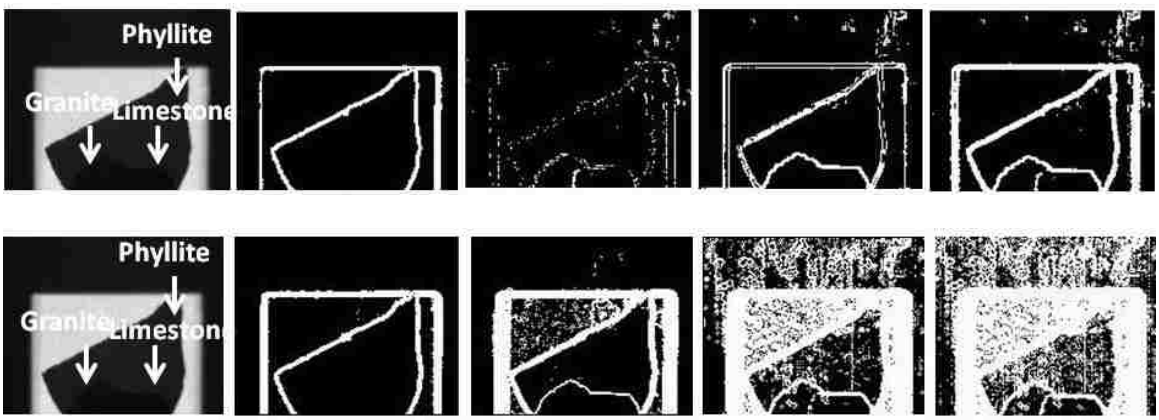


Figure 5.23: Comparison between the SRC edge detector (top row) with the cross-ratio signatures in Table 5.5, and the MCG edge detector (bottom row) for the DWELL's first testing data set consisting of limestone, granite and background; *raw data.* Top row: (i) the DWELL test image at 0.7 V; (ii) edges E_{GB}; (iii) edges E_{LB}; (iv) edges E_{LG}; (v) combined SRC edge map. Bottom row: (i) the DWELL test image at 0.7 V; (ii–v) MCG edge maps for a sequence of increasingly permissive tolerances.



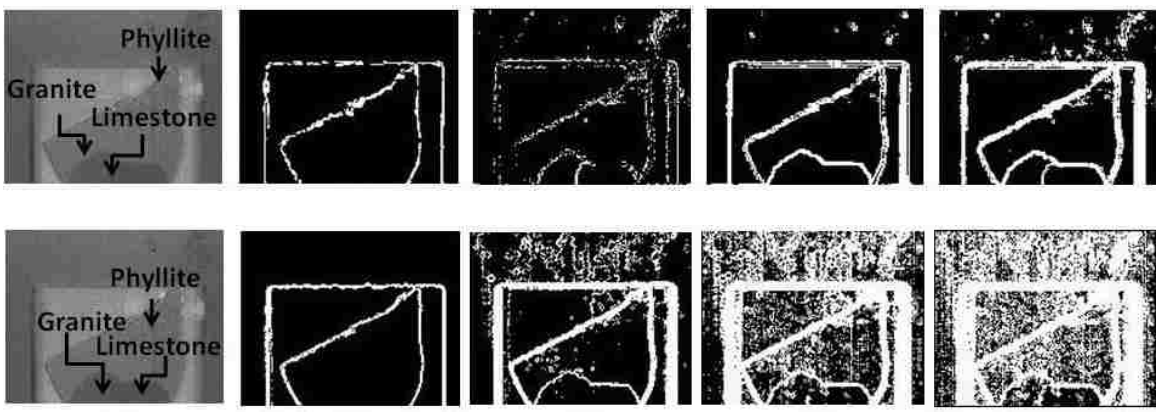
(i) (ii) (iii) (iv) (v)

Figure 5.24: Comparison between the SRC edge detector (top row) with the cross-ratio signatures in Table 5.5, and the MCG edge detector (bottom row) for the DWELL's first testing data set consisting of limestone, granite and background; *normalized data*. Top row: (i) the DWELL test image at 0.7 V; (ii) edges E_{GB}; (iii) edges E_{LB}; (iv) edges E_{LG}; (v) combined SRC edge map. Bottom row: (i) the DWELL test image at 0.7 V; (ii–v) MCG edge maps for a sequence of increasingly permissive tolerances.



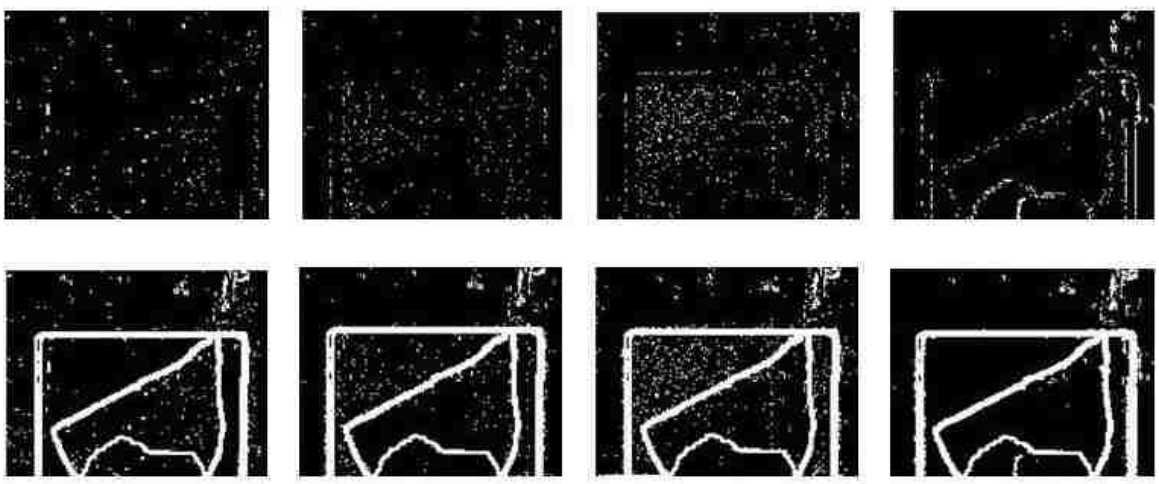
(i) (ii) (iii) (iv) (v)

Figure 5.25: Comparison between the SRC edge detector (top row) with the cross-ratio signatures in Table 5.6, and the MCG edge detector (bottom row) for the DWELL's second testing data set consisting of phyllite, limestone, granite and background; *raw data*. Top row: (i) the DWELL test image at 0.7 V; (ii) edges E_{PB}; (iii) edges E_{LG}; (iv) edges E_{PG}; (v) combined SRC edge map. Bottom row: (i) the DWELL test image at 0.7 V; (ii–v) MCG edge maps for a sequence of increasingly permissive tolerances.



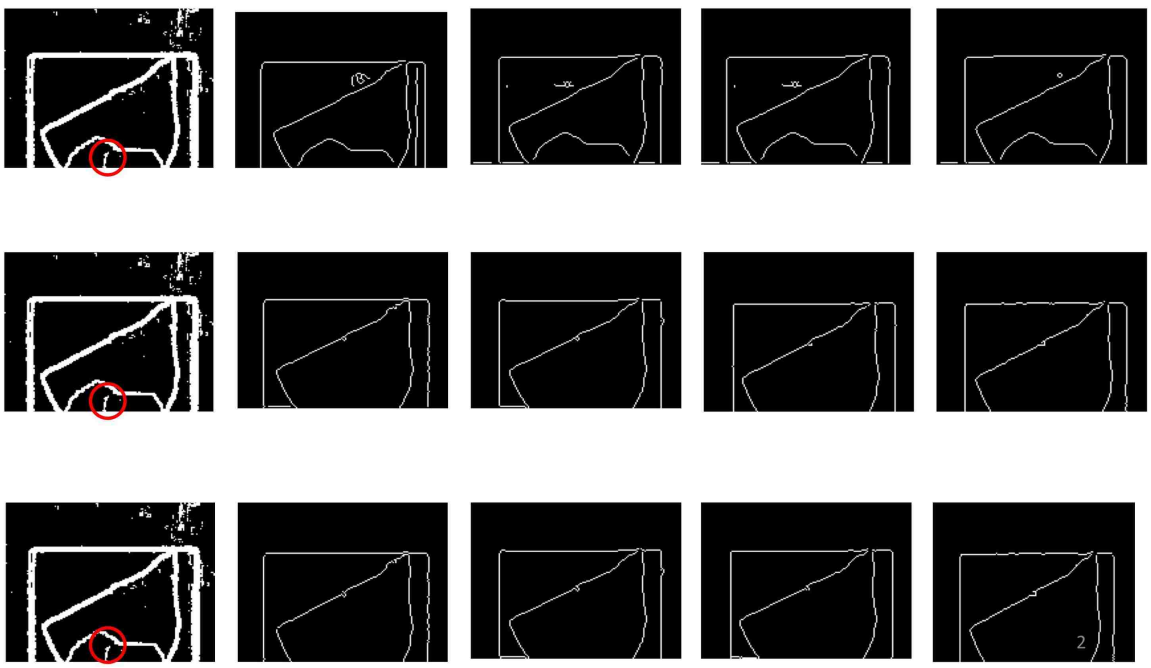
(i) (ii) (iii) (iv) (v)

Figure 5.26: Comparison between the SRC edge detector (top row) with the cross-ratio signatures in Table 5.6, and the MCG edge detector (bottom row) for the DWELL's second testing data set consisting of phyllite, limestone, granite and background; *normalized data*. Top row: (i) the DWELL test image at 0.7 V; (ii) edges E_{PB}; (iii) edges E_{LG}; (iv) edges E_{PG}; (v) combined SRC edge map. Bottom row: (i) the DWELL test image at 0.7 V; (ii–v) MCG edge maps for a sequence of increasingly permissive tolerances.



(i) (ii) (iii) (iv)

Figure 5.27: Comparison between the performance of the pairwise and the cross ratio edge signatures in the SRC edge detector. Top row shows the edge map for the weak edge between classes L and G obtained with the following edge signatures defined in (5.22) and Table 5.6: (i) edges E_{LG} using $\mathcal{E}^1_{\mathrm{LG}}$; (ii) edges E_{LG} using $\mathcal{E}^5_{\mathrm{LG}}$; (iii) edges E_{LG} using $\mathcal{E}^6_{\mathrm{LG}}$; (iv) edges E_{LG} using $\mathcal{E}_{\mathrm{LG}}$ from Table 5.6. Bottom row: (i–iv) complete SRC edge maps for the same choices of $\mathcal{E}_{\mathrm{LG}}$.



(i) (ii) (iii) (iv) (v)

Figure 5.28: Comparison between the performance of the SRC edge detector with the signatures defined in Table 5.6 and the Canny (top row), Sobel (middle row) and Prewitt (bottom row) gray-scale edge detectors applied to individual DWELL FPA bands: (i) SRC edge map; (ii) gray-scale edge maps at 0.5 V; (iii) gray-scale edge maps at 0.7 V; (iv) gray-scale edge maps at 0.9 V; (v) gray-scale edge maps at 1.1 V. The weak edge between granite and limestone is indicated by the red circle in column (i).

Appendix A

Multi-Color Gradient Edge Detector

In this Appendix we briefly review the definition of the multi-color gradient [1] and explain the idea of the MCG-based edge detector. The MCG algorithm is used as a benchmark for the validation of the performance of the PSS edge detection-algorithm developed in this dissertation and described in Chapter 5.

A.1 Definition of the multi-color gradient

Recall that a multi-color image is a three-dimensional array (5.1) of real numbers. To explain the definition of the multi-color gradient operator it is convenient to regard the algebraic data representation (5.1) as sampling of a differentiable multi-valued function

$$\mathbf{u} : \Omega \mapsto \mathbb{R}^K, \quad \mathbf{u}(x_1, x_2) = (u_1(x_1, x_2), \ldots, u_k(x_1, x_2)) \,. \tag{A.1}$$

In (A.1) $\Omega = [0,1]^2$ is the image plane. The coordinate functions $u_k : \Omega \mapsto \mathbb{R}$ correspond to the image planes of the multi-color image generated by the bands of the sensor.

The data model (A.1) is also referred to as *vector-valued or* multi-color image; see [28, 25]. In this model the spatial position of the pixel is assumed to be continuously varying in Ω while the band is discrete.

We define the multi-color gradient (MCG) operator following the ideas of Di Zenzo [1] and Cumani [26]. These authors provide extension of the standard gradient to multi-color images of the form (A.1) by treating the image as a manifold in $\mathbb{R}^K$. This approach has been used in [25, 29] for extension of anisotropic diffusion methods for image processing to multi-color images. We first review the definition of MCG using the functional image model (A.1) and then use this definition to motivate an MCG operator for the discrete image model (5.1).

The MCG operator [1] for (A.1) is defined as follows. Assume that the Jacobian

$$(D\mathbf{u})_{ki} = \frac{\partial u_k}{\partial x_i}, \quad 0 \le k \le K, \quad i = 1, 2$$

is rank-2 for all $(x_1, x_2) \in \Omega$. Then the map $\mathbf{u} : \Omega \mapsto \mathbb{R}^K$ defines a two-dimensional manifold (surface) in $\mathbb{R}^K$. This surface is parameterized by the pixel coordinates, i.e., Ω serves as parameterization domain.

Let $P \in \Omega$ and $Q = P + dP$ where $dP = (dx_1, dx_2)$ is infinitesimal displacement. The differential

$$d\mathbf{u} = \sum_{i=1}^{2} \frac{\partial \mathbf{u}}{\partial x_i} dx_i \,,$$

where

$$\frac{\partial \mathbf{u}}{\partial x_i} = \left(\frac{\partial u_1}{\partial x_i} \cdots, \frac{\partial u_k}{\partial x_i} \right) ,$$

estimates the "jump" (the difference) in the image values between the two pixels P and Q. It follows then that the variation in the image can be measured by the

Euclidean norm of $d\mathbf{u}$:

$$|d\mathbf{u}|^2 = \left(\sum_{i=1}^{2} \frac{\partial \mathbf{u}}{\partial x_i} dx_i \right)^2 = \sum_{i,j=1}^{2} \frac{\partial \mathbf{u}}{\partial x_i} \cdot \frac{\partial \mathbf{u}}{\partial x_j} dx_i dx_j \,. \tag{A.2}$$

Using the notation

$$g_{ij} = \frac{\partial \mathbf{u}}{\partial x_i} \cdot \frac{\partial \mathbf{u}}{\partial x_j} = \sum_{k=1}^{K} \frac{\partial u_k}{\partial x_i} \cdot \frac{\partial u_k}{\partial x_j}$$

we see that (A.2) defines a quadratic form

$$g(dx_1, dx_2) = \sum_{i,j=1}^{2} (g_{ij}) dx_i dx_j \,.$$

This form is known as the *first fundamental form* of the manifold $\mathbf{u}(x_1, x_2)$.

A.2 Application of the MCG to edge detection

Intuitively, the first fundamental form g measures the "steepness" of the manifold. Therefore, it can be used to estimate the rate of change in a multi-color image. This is precisely the idea of MCG edge detection exploited in [1, 26].

Specifically, if $(dx_1, dx_2) = (\cos\theta, \sin\theta)$ is a unit vector, then $g(dx_1, dx_2)$ measures the rate of change of the image in the direction of this vector. The fastest change will occur along directions parallel to the eigenvector corresponding to the maximum eigenvalue of the matrix g_{ij}. Direct computation shows that

$$\lambda_{\pm} = \frac{1}{2} \left(g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2} \right) .$$

The eigenvectors are $\mathbf{v}_{\pm} = (\cos\theta_{\pm}, \sin\theta_{\pm})$ with

$$\theta_{+} = \frac{1}{2} \arctan(2g_{12}/(g_{11} - g_{22})) \,; \quad \theta_{-} = \theta_{+} + \pi/2 \,.$$

Following Sapiro [25] we call θ_{+} the direction of maximal change and λ_{+} the maximal rate of change. θ_{-} is the direction of minimal change and λ_{-} is the rate

of minimal change. If, at a point (x_1, x_2) λ_+ is much larger than λ_-, it follows that the multi-color image changes rapidly in the direction of θ_+ and so an edge can be placed in the perpendicular direction. On the other hand, if $\lambda_+ \approx \lambda_-$, the image changes at the same rate in all directions and no edge exists.

It follows that edges can be detected by a monitor function $f(\lambda_+ - \lambda_-)$ that attenuates the difference between the maximal and minimal rates of change [25]. The use of the first fundamental form to define a multi-color gradient operator ensures consistency with the standard gray-scale edge detection because in the limit case of a single band ($K = 1$) the MCG operator yields the standard gradient. Indeed, if $K = 1$ we have that $\mathbf{u}(x_1, x_2) = u(x_1, x_2)$ and the components of the first fundamental form are given by

$$g_{ij} = \frac{\partial u}{\partial x_i} \cdot \frac{\partial u}{\partial x_j} .$$

A direct calculation shows that $\lambda_+ = |\nabla u|^2$, and $\lambda_- = 0$. Therefore, an multi-color edge detector based on, e.g.,

$$f(\lambda_+ - \lambda_-) = |\lambda_+ - \lambda_-|^{1/2}$$

is consistent extension of a gray-scale edge detector based on the magnitude of the standard gradient.

A.3 Implementation of the MCG edge detector

To implement an edge detector for the discrete multi-color image (5.1) we view this rank-3 tensor as a discrete sample of the continuous image function (A.1). Accordingly, specialization of MCG to (5.1) requires numerical approximation of the partial derivatives of the image planes $u_k(x_1, x_2)$. We restrict attention to implementation using a 3×3 mask and central differences along the coordinate axes to approximate the partial derivatives. In this case, the elements of the first fundamental form are

approximated by

$$g_{11}(i,j) = \sum_{k=1}^{K} (u_k(i,j-1) - u_k(i,j+1))^2 \,,$$

$$g_{22}(i,j) = \sum_{k=1}^{K} (u_k(i-1,j) - u_k(i+1,j))^2 \,,$$

and

$$g_{12}(i,j) = \sum_{k=1}^{K} (u_k(i,j-1) - u_k(i,j+1))(u_k(i-1,j) - u_k(i+1,j)) \,,$$

respectively, and the discrete first fundamental form at pixel (i,j) is given by

$$G(i,j) = \begin{pmatrix} g_{11}(i,j) & g_{12}(i,j) \\ g_{12}(i,j) & g_{22}(i,j) \end{pmatrix} .$$

Let $\lambda_+(i,j)$ and $\lambda_-(i,j)$ denote the eigenvalues of $G(i,j)$. We implement the MCG edge detector for (5.1) using the following monitor function

$$\Delta\lambda(i,j) = |\lambda_+(i,j) - \lambda_-(i,j)|^{1/2} \,.$$

Specifically, the super-pixel at (i,j) belongs to an edge if

$$\Delta\lambda_{i,j} > \tau$$

where τ is a positive threshold parameter.

References

[1] S. Di Zenzo, "A note on the gradient of a multi-image," *Comput. Vision Graph. Image Process.*, vol. 33, no. 1, pp. 116–125, 1986.

[2] R. Gonzalez and R. Woods, *Digital Image Processing.* Upper Saddle River, New Jersey, 07458: Prentice Hall, 2001.

[3] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, Nov. 1986.

[4] I. Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," 1968, talk presented at the Stanford Artificial Project.

[5] "NASA/JPL Airborne Visible Infrared Imaging Spectrometer," http://aviris.jpl.nasa.gov/html/aviris.

[6] P. G. Lucey, T. J. Williams, J. L. Hinrichs, M. E. Winter, D. Steutel, and E. M. Winter, "Three years of operation of AHI: the University of Hawaii's Airborne Hyperspectral Imager," *SPIE Infrared Technology and Applications*, vol. XXVII, pp. 112–120, 2001.

[7] R. Vaughan, W. Calvin, and J. Taranik, "SEBASS hyperspectral thermal infrared data: surface emissivity measurement and mineral mapping," *Remote Sensing of Environment*, vol. 85, pp. 48–63, April 2003.

[8] S. Krishna, "Optoelectronic properties of self-assembled InAs/InGaAs quantum dots," *III-V Semiconductor Heterostructures: Physics and Devices*, vol. 3438, pp. 234–242, 2003.

[9] R. V. Shenoi, R. S. Attaluri, A. Siroya, J. Shao, Y. D. Sharma, A. Stinz, T. E. Vandervelde, and S. Krishna, "Low-strain InAs/InGaAs/GaAs quantum dots-in-a-well infrared photodetector," *J. Vac. Sci. Technol. B*, vol. 26, pp. 1136–1139, 2008.

[10] S. Krishna, M. M. Hayat, J. S. Tyo, S. Raghvan, and Ü. Sakoglu, "Spectrally adaptive quantum dot infrared sensors for focal plane arrays," U.S. Patent No. 7,217,951, 2007.

[11] S. Krishna, "Quantum dots-in-a-well infrared photodetectors," *Journal of Physics D: Applied Physics*, vol. 38, pp. 2142–2150, 2005.

[12] B. S. Paskaleva, M. M. Hayat, Z. Wang, J. S. Tyo, and S. Krishna, "Canonical correlation feature selection for sensors with overlapping bands: Theory and application," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 46, pp. 3346–3358, 2008.

[13] W.-Y. Jang, B. Paskaleva, M. M. Hayat, and S. Krishna, "Spectrally adaptive nanoscale quantum dot sensors," *Wiley Handbook of Science and Technology for Homeland Security*, 2008.

[14] B. Paskaleva, W.-Y. Jang, M. M. Hayat, and S. Krishna, "Application of the canonical-correlation feature-selection algorithm to tunable quantum-dot detectors: Material discrimination and gas detection," *International Symposium on Spectral Sensing Research (ISSSR 08)*, 2008, June 23-27, Hoboken, New Jersey.

[15] W.-Y. Jang, M. M. Hayat, S. Bender, Y. D. Sharma, J. Shao, and S. Krishna, "Performance enhancement of an algorithmic spectrometer with quantum-dots-

in-a-well infrared photodetectors," *International Symposium on Spectral Sensing Research (ISSSR 08)*, 2008, June 23-27, Hoboken, New Jersey.

[16] J. H. Friedman, "Exploratory projection pursuit," *J. American Statistical Association*, vol. 82, pp. 249–266, 1987.

[17] A. Ifarraguerri and C. Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2529–2538, Nov. 2000.

[18] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[19] J. Friedman and W. Stuetzle, "Projection pursuit regression," *J. Amer. Statist. Asso.*, vol. 76, pp. 817–823, 1981.

[20] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 26, pp. 65–74, 1988.

[21] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological prifiles," *IEEE Trans. on Geosci. and Remote Sens. Lett.*, vol. 42, pp. 480–491, 2005.

[22] F. DellAcqua, P. Gamba, A. Ferrari, J. A. Palmason, and J. A. Benediktsson, "Exploiting spectral and spatial information in hyper-spectral urban data with high resolution," *IEEE Trans. on Geosci. and Remote Sens. Lett.*, vol. 1, pp. 322–326, 2004.

[23] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classication of hyperspectral data from urban areas using morpho-logical preprocessing and independent component analysis," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS05)*, 2005, pp. 176 –179.

[24] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classication with spectral-based kernels," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS03)*, vol. 1, 2003.

[25] G. Sapiro and D. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *Image Processing, IEEE Transactions on*, vol. 5, no. 11, pp. 1582–1586, Nov 1996.

[26] A. Cumani, "Edge detection in multispectral images," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 1, pp. 40 – 51, 1991.

[27] A. N. Evans and X.-U. Liu, "A morphological gradient approach to color edge detection," *IEEE Trans. on Image Processing*, vol. 15, pp. 1454–1463, 2006.

[28] P. Blomgren and T. F. Chan, "Color tv: Total variation methods for restoration of vector valued images," *IEEE Trans. Image Processing*, vol. 7, pp. 304–309, 1998.

[29] G. Sapiro, "Color snakes," *Comput. Vis. Image Underst.*, vol. 68, no. 2, pp. 247–253, 1997.

[30] U. Al-Suwailem and J. Keller, "Multichannel image identification and restoration using continuous spatial domain modeling," *Image Processing, International Conference on*, vol. 2, p. 466, 1997.

[31] Ü. Sakoglu, J. S. Tyo, M. M. Hayat, S. Raghavan, and S. Krishna, "Spectrally adaptive infrared photodetectors using bias-tunable quantum dots," *J. Optical Society of America B*, vol. 21, pp. 7–17, 2004.

[32] Ü. Sakoglu, M. M. Hayat, J. S. Tyo, P. Dowd, S. Annamalai, K. T. Posani, and S. Krishna, "Statistical adaptive sensing by detectors with spectrally overlapping bands," *Applied Optics*, vol. 45, pp. 7224–7234, 2006.

[33] W.-Y. Jang, M. M. Hayat, J. S. Tyo, R. S. Attaluri, T. E. Vandervelde, Y. D. Sharma, R. Shenoi, A. Stintz, E. R. Cantwell, S. Bender, and S. Krishna, "Demonstration of bias controlled algorithmic tuning of quantum dots in a well mid-infrared detectors," *IEEE J. Quantum, Electronics*, vol. 45, pp. 5537–5540, 2009.

[34] B. Sandberg and T. F. Chan, "Logic operators for active contours on multi-channel images," in *UCLA Department of Mathematics CAM Report*, 2002, pp. 02–12.

[35] J. Weier and D. Herring, "Measuring vegetation (NDVI and EVI)," http://earthobservatory.nasa.gov/Features/MeasuringVegetation/.

[36] A. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sensing of Environment*, vol. 25, pp. 53–70, 1988.

[37] Y. Kaufman and D. Tanre, "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS," in *Proc. IEEE Int. Geosci. and Remote Sensing Symp.* New York: IEEE, 1992, pp. 261–270.

[38] K. Mayeda, A. Hofstetter, J. O'Boyle, and W. Walter, "Stable and transportable regional magnitudes based on coda-derived moment-rate spectra," *Bull. Seismol. Soc. Am.*, vol. 93, pp. 224–239, 2003.

[39] K. Mayeda, L. Malagnini, and W. Walter, "A new spectral ratio method using narrow band coda envelopes: Evidence for non-self-similarity in the Hector Mine sequence," *Geophys. Res. Lett.*, vol. 34, p. L11303, 2007.

[40] H. Hartse, S. Taylor, W. Phillips, and G. Randall, "Regional seismic discrimination in central Asia with emphasis on Western China," Los Alamos National Laboratory, Tech. Rep. LA-UR 96-2390, 1996.

[41] H. J. Patton and B. J. Mitchell, *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Regional Wave Propagation and Crustal Structure.* Birkhuser Verlag, 2001.

[42] M. Lennon and G. Mercier, "Noise-adjusted non orthogonal linear projections for hyperspectral data analysis," in *Proc. IEEE IGARSS'03*, vol. 6, 2003, pp. 3760–3762.

[43] J. Jiang, K. Mi, R. McClintock, M. Razeghi, G. J. Brown, and C. Jelen, "Demonstration of 256x256 focal plane array based on Al-free GaInAs-InP QWIP," *IEEE Photon. Technol. Lett*, vol. 15, pp. 1273–1275, 2003.

[44] S. Krishna, S. Ragahavan, G. Winckel, A. Stinz, G. Ariawansa, S. G. Matsik, and A. Perera, "Three-color InAs/nGaAs quatudots-in-a-well detectors," *Applied Physics Letters*, vol. 83, pp. 2745–2747, 2003.

[45] J. Topol'ancik, S. Pradhan, P. C. Yu, S. Chosh, and P. Bhattacharya, "Electrically injected photonic edge-emmiting quantum-dot light source," *IEEE Photon. Technol. Lett*, vol. 16, pp. 960–962, 2004.

[46] K. T. Posani, V. Thripati, S. Annamalai, N. Weirs-Einstein, S. Krishna, P. Perahia, O. Crisafulli, and O. J. Painter, "Nanoscale quantum dot infrared sensors with photonic crystal cavity," *Applied Physics Letters*, vol. 88, pp. 151 104–1–151 104–3, 2006.

[47] P. Bhattacharya, *Semiconductor Optoelectronic Devices.* Prentice Hall, 1997.

[48] L. O. Jimenez and D. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 37, pp. 2653–2667, 1999.

[49] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, pp. 626–634, 1999.

[50] Z. Wang, B. S. Paskaleva, J. S. Tyo, and M. M. Hayat, "Canonical correlations analysis for assessing the performance of adaptive spectral imagers," *Proc. SPIE*, vol. 5806, pp. 23–34, 2005.

[51] Z. Wang, J. S. Tyo, and M. M. Hayat, "Bi-orthogonal geometrical model for spectral sensors with correlated bands," *J. Optical Society of America A*, vol. 29, pp. 2864–2870, 2007.

[52] W. Rudin, *Real and Complex Analysis*. McGraw-Hill, Inc, 1986.

[53] J. Dauxois and G. M. Nkiet, "Canonnical analysis of two Euclidean subspaces and its application," *Elsevier Science Inc.*, vol. 27, pp. 354–387, 1997.

[54] A. Björck and G. H. Golub, "Numerical methods for computing the angles between linear subspaces," *Math. Comp.*, vol. 27, pp. 579–594, 1973.

[55] A. V. Knyazev and M. E. Argentati, "Principal angles between subspaces in a A-based scalar product: Algorithms and perturbation estimates," *SIAM J. Sci. Comput.*, vol. 23, pp. 2009–2041, 2002.

[56] B. Paskleva, M. M. Hayat, M. M. Moya, and R. J. Fogler, "Multispectral rock-type separation and classification," *Proc. SPIE*, vol. 5543, pp. 152–163, 2004.

[57] B. S. Paskaleva, M. M. Hayat, J. S. Tyo, Z. Wang, and M. Martinez, "Feature selection for spectral sensors with overlapping noisy spectral bands," *Proc. SPIE*, vol. 6233, pp. 623 329–1–623 329–7, 2006.

[58] S. W. Ruff, P. R. Christensen, P. W. Barbera, and D. L. Anderson, "Quantitative thermal emission spectroscopy of minerals: A laboratory technique for measurement and calibration," *J. Geophys. Res.*, vol. 102, pp. 14 899–14 913, 1997.

[59] F. D. Palluconi and G. R. Meeks, "Thermal infrared multispectral scanner (TIMS): An investigators guide to TIMS data," *Jet Propul.Lab., Pasadena, CA*, 1985.

[60] K. C. Feely and P. R. Christensen, "Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks," *J. Geophys. Res.*, vol. 104, pp. 24 195–24 210, 1999.

[61] W. B. Clodius, P. G. Weber, C. C. Borel, and B. W. Smith, "Multispectral thermal imaging," *Proc. SPIE*, vol. 3438, pp. 234–242, 1998.

[62] "Advanced Spaceborn Thermal Emission and Reflection Radiometer," http://asterweb.jpl.nasa.gov.

[63] R. O. Duda, P. E. Hart, and D. G. Strok, *Pattern Classification*. John Wiley and Son, 2000.

[64] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1984.

[65] C. B. Akgül, "Projection pursuit for optimal visualization of multivariate data," Ph.D. dissertation, Bogazici University, Istanbul, 2003. [Online]. Available: http://www.tsi.enst.fr/akgul/oldprojects/qli

[66] D. Landgrebe, "Information extraction principles and methods for multispectral and hyperspectral image data," *Information Processing for Remote Sensing*, vol. 82, pp. 3–38, 1999.

[67] J. W. Boardman, "Analysis and understanding and visualization of hyperspectral data as convex sets in n-space," *Proc. SPIE*, vol. 2480, pp. 14–22, 1995.

[68] M. Winter, "Fast autonomous spectral endmembers determination in hyperspectral data," *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, vol. II, pp. 337–344, Vancouver, B.C., Canada, 1999.

[69] C. Kwan, B. Ayhan, G. Chen, J. Wang, J. Baohong, and C.-I. Chang, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 44, pp. 409–419, 2006.

[70] J. Andrews, W.-Y. Jang, J. E. Pezoa, Y. D. Sharma, S. J. Lee, S. K. Noh, M. M. Hayat, S. Restaino, S. W. Teare, and S. Krishna, "Demonstration of a bias tunable quantum dots-in-a-well focal plane array," *Infrared Phys. Technol.*, vol. in press, 2009.

[71] E. Varley, M. Lenz, S. J. Lee, J. S. Brown, D. A. Ramirez, A. Stintz, and S. Krishna, "Single bump, two-color quantum dot camera," *Applied Physics Letters*, vol. 91, 2007.

[72] A. Milton, F. Barone, and M. Kruer, "Influence of nonuniformity on infrared focal plane array performance," *Opt. Eng. (Bellingham)*, vol. 24, pp. 855–862, 1985.

[73] J. A. Benediktsson, J. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–490, Mar. 2005.

[74] P. Soille, *Morphological Image Analysis: Principles and Applications*. Germany: Springer-Verlag, 2003.

[75] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Treans. Geosci. and Remote Sensing*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[76] P. Gamba, F. Dell'Acqua, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Arnasson, "Exploiting spectral and spatial informationin the hyperspectral urban data with high resolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 322–326, Oct. 2004.

[77] J. C. Tilton, "Image segmentation by region growing and spectral clustering with natural convergance criterion," *In Proceedings of IGARSS 1998*, pp. 1766–1768, 1998.

[78] E. Davies, *Machine Vision: Theory, Algorithms and Practicalities.* Academic Press, 1990.

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

[80] L. Lucchese and S. K. Mitra, "Color image segmentation: a state-of-the-art survey," *Proc. Ind. Nat. Sci. Acad.*, vol. 67, pp. 207–221, 2001.

[81] J.-F. Rivest, P. Soille, and S. Beucher, "Morphological gradients," *J. Electron. Imag.*, vol. 2, pp. 326–336, 1993.

[82] P. J. Toivanen, J. Ansamki, J. P. S. Parkkinen, and J. Mielikinen, "Edge detection in multispectral images using the self-organizing map," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 2987 – 2994, 2003.