Advanced Spectral-Spatial Processing Techniques for Hyperspectral Image Analysis

Tong Qiao

Department of Electronic and Electrical Engineering

University of Strathclyde

This thesis is submitted for the degree of

Doctor of Philosophy

February 2016

I would like to dedicate this thesis to
my loving parents,
who have been a constant source of support
and the engineering enlightenment to me.

谨以此论文献给
我亲爱的父亲母亲
感谢他们
作为我对于工程师这一行业的启蒙导师
和多年以来对我的支持与关爱

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Tong Qiao
February 2016

Acknowledgements

The past three years are an unforgettable period in my life. My sincere thanks are going to a lot of people, who have helped me to walk through my PhD journey.

First, I would like to thank my supervisor Dr. Jinchang Ren, for his kind guidance and support. I still could recall the time when he carefully revised my first paper and it is him who taught me how to be an effective learner during my PhD studies. Most importantly, I truly thank him for believing my potential when I feel discouraged sometimes. Additionally, I would also like to thank my second supervisor Prof. Stephen Marshall. As the leader of the hyperspectral imaging centre, he does provide a lot of efforts for improving the reputation of our research group. I would also like to take this opportunity to thank Dr. Peter Yuen and Dr. Stephan Weiss - my viva examiners, for their helpful comments and suggestions.

Besides, thanks Quality Meat Scotland (QMS) and University of Strathclyde for offering financial support for my PhD, which helps me focusing on my research without worrying too much. Special thanks also go to former QMS staffs, Prof. Charlotte Maltin and Dr. Cameron Craigie, for encouraging me all the time and providing valuable comments to my papers/thesis.

I also wish to thank my colleagues working in the hyperspectral imaging centre, especially Dr. Paul Murray (also as my viva convener), Dr. Jaime Zabalza and Dr. Timothy Kelman. Whenever I have problems either in the academic world or in real life, they are always keen to suggest possible solutions for me.

Last, my deepest gratitude goes to my family and my future family member, Mr. Yihan Liu, for their constant encouragement and belief in me. Every success in my life would have not been possible without you.

Abstract

Tong Qiao. *Advanced Spectral-Spatial Processing Techniques for Hyperspectral Image Analysis.* Ph.D. thesis, University of Strathclyde, 2016.

The main objective of this research is to design and implement novel spectral-spatial processing techniques for hyperspectral image analysis and applications. Although the high dimensionality of hyperspectral image data makes its transmission and storage difficult, the uncompressed data format is still preferred as it avoids compression loss which may degrade classification accuracy. In this thesis, a quality-assured lossy compression scheme based on a modified three dimensional discrete cosine transform is proposed. This novel technique is demonstrated to maintain the integrity of hyperspectral data without degrading the classification accuracy. Furthermore, this work has led to the creation of an effective spectral feature extraction technique which uses curvelet transform and singular spectrum analysis. In addition to this, an original classification framework which combines joint bilateral filtering and an improved sparse representation classifier is presented. Experimental results show that the proposed methodologies outperform most of the state-of-the-art feature extraction and classification techniques commonly employed in the hyperspectral community. This work also demonstrates that hyperspectral imaging combined with advanced signal processing is an effective technology for food quality control applications. For example, when applied to the challenge of performing hyperspectral imaging-based meat quality assessment, the techniques proposed in this work are shown to provide a more effective solution than conventional visible and near-infrared spectroscopic technology. Finally, this thesis provides the first set of results of assessing the quality of beef and lamb samples using an improved data regression technique. To sum up, the outcome of this thesis advances the hyperspectral imaging community by proposing several novel methodologies, and extensive experiments have been conducted to demonstrate their superiority.

Table of contents

List of figures **ix**

List of tables **xiii**

Acronyms / Abbreviations **xvi**

1 Introduction **1**

1.1 Aims and motivations 1

1.2 Main contributions of this research 4

1.3 Thesis overview 5

2 Background and Related Work **7**

2.1 Introduction 7

2.2 Hyperspectral image compression 7

2.3 Hyperspectral image classification 11

2.4 Meat quality assessment using spectroscopic technologies 15

2.5 Principles of related techniques 19

2.5.1 Principal component analysis 19

2.5.2 Support vector machine 21

2.6 Summary 24

3 Hyperspectral Imagery Compression Based on an Improved Three-Dimensional Discrete Cosine Transform Approach **25**

3.1 Introduction 25

3.2 Improved 3D DCT-based approach 26

3.2.1 From 2D DCT to 3D DCT 26

3.2.2 Proposed learning-based approach 28

3.3 Data sets and experimental setup 29

3.3.1 Data set preparation 30

3.3.2 Evaluation criteria 31

3.3.3 Comparison of different quantisation matrices 34

3.4 Results and discussion . . . 35
3.4.1 Results of the subjective assessment . . . 35
3.4.2 Results of the objective assessment . . . 37
3.4.3 Results of the quality-assured classification assessment . . . 41
3.4.4 Complexity . . . 43
3.5 Summary . . . 44

4 Hyperspectral Image Feature Extraction by Exploiting the Curvelet Domain Using Singular Spectrum Analysis 45
4.1 Introduction . . . 45
4.2 Background principles . . . 47
4.2.1 The curvelet transform . . . 47
4.2.2 Singular spectrum analysis . . . 50
4.3 The proposed approach: applying SSA in the curvelet domain . . . 52
4.4 Data sets and experimental setup . . . 56
4.5 Results and discussion . . . 58
4.5.1 Comparison with inspirational approaches . . . 59
4.5.2 Comparison with other state-of-the-art spectral feature extraction techniques . . . 68
4.5.3 Comparison with recent spectral-spatial classification methods 69
4.6 Summary . . . 70

5 Hyperspectral Image Classification Using Sparse Representation of Features Extracted via Bilateral Filtering 73
5.1 Introduction . . . 73
5.2 Background principles . . . 75
5.2.1 Bilateral filtering . . . 75
5.2.2 Sparse representation classification . . . 76
5.3 The proposed approach . . . 78
5.4 Data sets and experimental setup . . . 82
5.5 Results and discussion . . . 85
5.5.1 Evaluation of classifiers . . . 85
5.5.2 Comparison with typical state-of-the-art spectral-spatial classification frameworks . . . 85
5.6 Summary . . . 91

6 Meat Quality Assessment Using Hyperspectral Imaging 93
6.1 Introduction . . . 93

6.2 Materials and methods . . . 95
6.2.1 Data set A . . . 96
6.2.2 Data set B . . . 98
6.2.3 Data set C . . . 102
6.2.4 Data set D . . . 103
6.3 Results and discussion . . . 104
6.3.1 SVM for predicting beef quality based on VIS-NIR spectroscopy . . . 105
6.3.2 Comparison between VIS-NIR spectroscopy and HSI in predicting beef eating quality . . . 108
6.3.3 SSA for improving HSI-based beef eating quality evaluation 110
6.3.4 Prediction of lamb quality using HSI . . . 112
6.4 Summary . . . 113

7 Conclusions and Future Work 115
7.1 Conclusions . . . 115
7.2 Future work . . . 117

References 119

Appendix A Hyperspectral Remote Sensing Data Sets 136
A.1 Moffett Field scene 1 . . . 136
A.2 Indian Pines . . . 137
A.3 Salinas Valley . . . 138
A.4 Pavia University . . . 139

Appendix B List of Author's Publications 140
B.1 International journal publications . . . 140
B.2 International conference publications . . . 141
B.3 Journal publications under consideration . . . 142

List of figures

1.1 From grayscale image to hyperspectral image. 2

2.1 The schematic diagram of hyperspectral image classification and analysis process (reproduced from [1]). 11
2.2 The process of reshaping 3D hyperspectral image to 2D spectral data matrix. 19

3.1 Pseudo colour images of four experimental data sets, from left to right: Moffett Field scene 1 (after cropping and binning), Indian Pines (after cropping), Salinas Valley (after cropping) and Pavia University (after cropping). 31
3.2 Ground truth maps for the last three data sets including Indian Pines (left), Salinas Valley (middle) and Pavia University (right). 31
3.3 Rate-distortion curves of different quantisation matrices for four experimental data sets. 35
3.4 Difference images between original and reconstructed images based on 3D SPIHT and 3D DCT (bottom) at 0.1 bpppb. From left to right, the four columns correspond to Moffett Field scene 1, Indian Pines, Salinas Valley and Pavia University, respectively. 36
3.5 Difference images of Moffett Field scene 1 based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different compression bit rates of 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right). 37
3.6 SNR (dB) comparison between 3D SPIHT and 3D DCT compression performance for four experimental data sets. 38
3.7 SSIM comparison between 3D SPIHT and 3D DCT compression performance for four experimental data sets. 38
3.8 Spectral angle maps of Moffett Field scene 1 based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right). 39

3.9 Spectral angle maps of Indian Pines based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right). 39
3.10 Spectral angle maps of Salinas Valley based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right). 40
3.11 Spectral angle maps of Pavia University based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right). 40
3.12 Comparison of classification accuracies for three reconstructed data sets using 3D SPIHT and different 3D DCT-based compression algorithms with available ground truth, including Indian Pines (top), Salinas Valley (middle) and Pavia University (bottom). . . 42

4.1 Flowchart of the proposed methodology: applying SSA in the curvelet domain. 53
4.2 Illustration of a 128 by 128 image with 4 decomposition scales, where (a) is the first (finest) scale with 1 orientation, (b) is the second scale with 8 orientations, (c) is the third scale with 8 orientations, and (d) is the fourth (coarsest) scale with 1 orientation. 55
4.3 Classification results (overall accuracy in percentage) of Indian Pines obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data. 60
4.4 Classification results (overall accuracy in percentage) of Salinas Valley obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data. 61
4.5 Classification results (overall accuracy in percentage) of Pavia University obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data. 62
4.6 Band 199 of Indian Pines in (a) original data and (b) CT-SSA processed data. 63
4.7 Band 2 of Salinas Valley in (a) original data and (b) CT-SSA processed data. 63
4.8 Band 1 of Pavia University in (a) original data and (b) CT-SSA processed data. 64

5.1 The performance of joint bilateral filtering with different sets of parameters. (a) Input band image from the Indian Pines data set, (b) the first principal component, (c) filtered images with different range parameters, and (d) filtered images with different domain parameters. 80
5.2 Comparison of bilateral filtering and joint bilateral filtering on low-contrast image. (a) Input band image from the Indian Pines data set, (b) the first principal component, (c) output image using bilateral filtering, and (d) output image using joint bilateral filtering with the first principal component as the guidance image. 80
5.3 The flowchart of the proposed SS-JSRC method. 81
5.4 Band 70 of the cropped Pavia University data set (left) and the 8 classes ground truth map (right), with number of samples in parentheses. 83
5.5 Effect of the sparsity level K_0 and the threshold parameter n for Indian Pines with the proposed SS-JSRC method. 84
5.6 The Indian Pines data set (a) ground truth and classification maps together with corresponding overall accuracy obtained by (b) SSA-SVM, (c) EMAP-SVM, (d) SADL-SVM, (e) SVM-EPF, (f) BF-SVM, (g) JBF-SVM and (h) JBF-SS-JSRC. 88
5.7 The Pavia University data set (a) ground truth and classification maps together with corresponding overall accuracy obtained by (b) SSA-SVM, (c) EMAP-SVM, (d) SADL-SVM, (e) SVM-EPF, (f) BF-SVM, (g) JBF-SVM and (h) JBF-SS-JSRC. 91

6.1 A plot of 10 replicate spectra (350 – 1800 nm) collected from the *M. longissimus thoracis* of one carcass in data set A, showing the noise at the extremes of the spectra. 97
6.2 Schematic diagram of the HSI system: components 1-5 refer to the EMCCD camera, spectrograph and lens, halogen lamp, sliding track and scanning tray, respectively. 99
6.3 A plot of 10 replicate spectra (350 – 2500 nm) collected from the *M. longissimus thoracis* of one carcass in data set B, showing the noise at the extremes of the spectra. 100
6.4 HSI spectral feature extraction process. (a) The original hyperspectral image, b) the cropped hyperspectral image after ROI selection, and (c) the median reflectance spectrum of the ROI after the iterative pixel removal process. 101

6.5 The original median spectrum and the SSA processed median spectrum. 102
6.6 HSI median reflectance spectrum extraction process for lamb. (a) Pseudo colour image of one lamb sample, with ROI marked by the white frame, and (b) the median reflectance spectrum of the ROI after the iterative pixel removal process. 104

A.1 Band 100 of Moffett Field scene 1 data set. 136
A.2 Band 168 of Indian Pines data set (left), the original 16 classes ground truth map (middle) and the reduced 9 classes ground truth map (right), with number of samples in parentheses. 137
A.3 Band 50 of Salinas Valley data set (left) and the 16 classes ground truth map (right), with number of samples in parentheses. 138
A.4 Band 70 of Pavia University data set (left) and the 9 classes ground truth map (right), with number of samples in parentheses. 139

List of tables

3.1 Optimal classification parameters and overall classification accuracies (%) for original data sets Indian Pines, Salinas Valley and Pavia University. 41
3.2 Comparison of encoding time (in seconds) for Moffett Field scene 1, Indian Pines, Salinas valley and Pavia Univeristy data sets using 3D SPIHT and 3D DCT. 44

4.1 Mean class-by-class, average and overall accuracies (%) of ten repeated experiments on test samples of the original Indian Pines data set and SSA, CT-MLR, WT-MLR, WT-SSA and CT-SSA processed data sets with 10% of data used for training, followed by the standard deviation. 65
4.2 Mean class-by-class, average and overall accuracies (%) of ten repeated experiments on test samples of the original Salinas Valley data set and SSA, CT-MLR, WT-MLR, WT-SSA and CT-SSA processed data sets with 5% of data used for training, followed by the standard deviation. 66
4.3 Mean class-by-class, average and overall accuracies (%) of ten repeated experiments on test samples of the original Pavia University data set and SSA, CT-MLR, WT-MLR, WT-SSA and CT-SSA processed data sets with 4% of data used for training, followed by the standard deviation. 67
4.4 Mean overall accuracy (%) and the corresponding standard deviation of ten repeated experiments on original data sets, SSA, CT-MLR, WT-MLR, WT-SSA, CT-SSA, CT-SSA-PCA and some state-of-the-art spectral feature extraction techniques processed data sets, with dimensionality of features shown in parentheses. . 69

4.5 Comparison of class-by-class, average and overall accuracies (%) of the CT-SSA-SPP method and recent spectral-spatial classification methods, with 10% training rate on the Indian Pines data set. . . 70
4.6 Comparison of class-by-class, average and overall accuracies (%) of the CT-SSA-SPP method and recent spectral-spatial classification methods, with 2% training rate on the Salinas Valley data set. . . 71
4.7 Comparison of class-by-class, average and overall accuracies (%) of the CT-SSA-SPP method and recent spectral-spatial classification methods, with 4% training rate on the Pavia University data set. 71

5.1 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classifiers for the Indian Pines data set with 10% of data used for training, followed by the standard deviation. 86
5.2 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classifiers for the Pavia University data set with 5% of data used for training, followed by the standard deviation. 87
5.3 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classification frameworks for the Indian Pines data set with 10% of data used for training, followed by the standard deviation. 89
5.4 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classification frameworks for the Pavia University data set with 5% of data used for training, followed by the standard deviation. 90
5.5 Computing time (in seconds) for different feature extraction techniques. 91

6.1 Summary statistics of all quality attributes for 234 carcasses in data set A. 105
6.2 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set A, using SVM and PLSR in the calibration set with noise-removed spectra (495 – 1600 nm). 105
6.3 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set A, using SVM and PLSR in the validation set with noise-removed spectra (495 – 1600 nm). 106

6.4 Publications for predicting instrumental meat quality in beef *M. longissimus thoracis* using VIS-NIR spectroscopy. 107
6.5 Summary statistics of quality attributes for 75 heifers in data set A. 107
6.6 Summary statistics of quality attributes for 118 steers in data set A. 107
6.7 Summary statistics of quality attributes for 41 young bulls in data set A. 108
6.8 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in heifer of data set A using SVM with noise-removed spectra (495 – 1600 nm). 108
6.9 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in steer of data set A using SVM with noise-removed spectra (495 – 1600 nm). 109
6.10 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in young bull of data set A using SVM with noise-removed spectra (495 – 1600 nm). 109
6.11 Summary statistics of all quality attributes for 858 carcasses in data set B. 109
6.12 Performance for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set B, using VIS-NIR spectroscopy and HSI in the calibration set, with noise removed spectra. 110
6.13 Performance for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set B, using VIS-NIR spectroscopy and HSI in the validation set, with noise removed spectra. 110
6.14 Summary statistics of all quality attributes for 211 carcasses in data set C. 110
6.15 Performance comparison of original HSI spectra and SSA processed spectra in data set C for predicting beef eating quality attributes. 111
6.16 Best performance of SSA processed HSI spectra in data set C for predicting beef eating quality attributes. 111
6.17 Summary statistics of all quality attributes for 80 lamb samples in data set D. 112
6.18 Performance of HSI for predicting instrumental lamb quality parameters of data set D. 112

Acronyms / Abbreviations

1D	One-Dimentional
2D	Two-Dimensional
3D	Three-Dimentional
AA	Average Accuracy
AOAC	Association Of Analytical Communities
AVIRIS	Airborne Visible/InfraRed Imaging Spectrometer
BPPPB	Bits Per Pixel Per Band
C-DPCM	Clustered Differential Pulse Code Modulation
CALIC	Context-based Adaptive Lossless Image Coding
CDF	Cohen-Daubechies-Feauveau
CIE	Commission Internationale de l-Eclairage
CS	Compressed Sensing
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DPCM	Differential Pulse Code Modulation
DSC	Distributed Source Coding
DWT	Discrete Wavelet Transform
EEMD	Ensemble Empirical Mode Decomposition
EMAP	Extended Morphological Attribute Profile

EMCCD	Electron Multiplying Charge-Coupled Device
EMD	Empirical Mode Decomposition
EMP	Extended Morphological Profile
EPF	Edge-Preserving Filter
EVG	EigenValue Grouping
EZW	Embedded Zerotrees of Wavelet coefficients
FDCT	Fast Discrete Curvelet Transform
GLA	Generalised Lloyd Algorithm
HHT	Hilbert-Huang Transform
HSI	HyperSpectral Imaging
ICA	Independent Component Analysis
IFRF	Image Fusion and Recursive Filter
IID	Intrinsic Image Decomposition
IMF	Intrinsic Mode Function
JPEG	Joint Photographic Experts Group
JSM	Joint Sparsity Model
JSRC	Joint Sparse Representation Classification
KLT	Karhunen-Loève Transform
LBGA	Linde-Buzo-Gray Algorithm
LDA	Linear Discriminant Analysis
LMLL	Logistic regression and Multi-Level Logistic
LUT	LookUp Table
MAP	Morphological Attribute Profile
MGA	Multi-scale Geometric Analysis

MIRINZ	Meat Industry Research Institute of New Zealand
MLR	Multiple Linear Regression
MNF	Minimum Noise Fraction
MP	Morphological Profile
MSC	Multiplicative Scatter Correction
MSE	Mean Squared Error
MSI	MultiSpectral Imaging
MSSIM	Mean Structural SIMilarity
NIR	Near-InfraRed
NMF	Non-negative Matrix Factorisation
NWFE	Non-parametric Weighted Feature Extraction
OA	Overall Accuracy
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDF	Probability Density Function
PLSR	Partial Least Squares Regression
QL	Quality Level
RBF	Radial Basis Function
RMSE	Root Mean Squared Error
ROI	Region Of Interest
ROSIS	Reflective Optics System Imaging Spectrometer
RPD	Ratio of Performance Deviation
SADL	Spatial-Aware Dictionary Learning

SAM	Spectral Angle Mapper
SD	Standard Deviation
SE	Structure Element
SNR	Signal-to-Noise Ratio
SNV	Standard Normal Variate
SOMP	Simultaneous Orthogonal Matching Pursuit
SP	Subspace Pursuit
SPIHT	Set Partitioning In Hierarchical Trees
SRC	Sparse Representation Classification
SSA	Singular Spectrum Analysis
SSF	Slice Shear Force
SSIM	Structural SIMilarity
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SWIR	Short-Wavelength InfraRed
TCQ	Trellis Coded Quantisation
USFFT	Uequally-Spaced Fast Fourier Transform
VIS-NIR	VISible and Near-InfraRed
VQ	Vector Quantisation
WBSF	Warner-Bratzler Shear Force

Chapter 1

Introduction

1.1 Aims and motivations

Hyperspectral imaging (HSI) sensors, which capture hundreds of continuous bands in a broad spectral range covering visible, near-infrared (NIR) and beyond, have proven their success in many areas [2]. Typical examples include traditional applications in remote sensing and military surveillance [3] as well as newly emerging platforms for food quality analysis [4, 5], pharmaceutical studies [6] and forensics/security related applications [7–9].

Hyperspectral images can be considered as an extension of traditional digital images. Figure 1.1 uses the Indian Pines data set (Appendix A) [10] as an example and shows a set of images, each of a different modality but corresponding to the same scene. As can be seen, a colour image is formed by combining three grayscale images in *red*, *green* and *blue* (RGB) channels, respectively. With the help of filters/spectrograph, light can be separated across the electromagnetic spectrum and multispectral/hyperspectral images can then be captured. The exact nature of the data acquisition process depends on the design of the imaging system being used. In a multispectral image there are generally tens of wide spectral bands (usually less than 20) and there are often relatively large spectral gaps between bands [11]. For this reason, multispectral imaging (MSI) systems are usually designed for applications where prior knowledge exists with regard to the spectral regions from which useful information can be extracted. Compared to multispectral images, hyperspectral images provide an increased number of spectral bands as well as higher spectral resolution, allowing an improved ability for categorising different materials within the scene.

One drawback with the additional spectral information provided by the hyperspectral image is that the amount of data required to be stored/transmitted

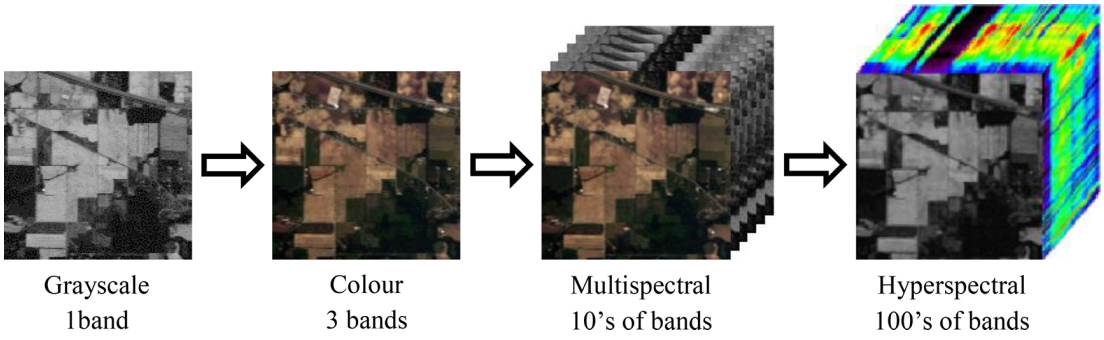


Fig. 1.1 From grayscale image to hyperspectral image.

is dramatticaly increased as a single hyperspectral image can potentially occupy hundreds of megabytes [12]. Considering that there are a lot of HSI-based remote sensing applications, hyperspectral images may need to be transmitted from satellite/aircraft to a ground station or simply stored on board. This makes data compression a necessary step for faster transmission and/or easier storage [13]. Since the data correlation between spectral bands is extremely high for hyperspectral images, traditional three-dimensional (3D) compression algorithms designed for isotropic 3D images might not achieve the best performance. Furthermore, in order to preserve the hyperspectral image quality for subsequent analysis, the development of compression algorithms has been focused on lossless compression schemes in the past. In recent years, an interest in lossy or nearly lossless compression for hyperspectral images has arisen. This is due to the fact that recently published advanced data compression algorithms allow a more balanced trade-off between data compression and the performance of end-user applications [14].

In addition to investigating efficient compression algorithms, another area of interest lies in developing techniques to automatically process and analyse hyperspectral images in order to efficiently understand their content. The spectral features for each pixel in the hyperspectral image, also named as the spectra, are associated with the spectral response of the light (reflection/absorption etc.) as captured by the HSI system at each wavelength. For different materials, the light reflection/absorption characteristics are usually unique along the electromagnetic spectrum. Therefore, the spectrum in each pixel of a hyperspectral image contains rich information about the captured scene, and this can be used for identifying and recognising a diverse range of materials and targets of interest that are present in the image. As a result, this has led to an active research area for interpreting the hyperspectral image – hyperspectral image classification.

Hyperspectral image classification, aiming at categorising image pixels into different land-cover classes according to their unique spectral information, has

become a popular research area in remote sensing for land mapping and land-cover analysis in recent years [15]. Since a huge amount of information is present in the hyperspectral image, it is impossible to perform this classification manually. However, by manually labelling a small number of pixels in the region of interest (ROI) (as opposed to the entire image), researchers found that this so called 'prior information' could be used to solve the problem for autonomous classification. This process is known as supervised classification, and the labelled pixels containing class labels as well as spectral signatures are known as labelled training data. Unfortunately, supervised classification of hyperspectral data is not straightforward and many factors can significantly influence the classification accuracy. First of all, the environmental effects present during image acquisition as well as the large spatial variability in the field of view introduce the possibility that spectra in the image may not directly match training spectra from the same category. Furthermore, different materials may share similar spectral signatures because of common chemical structures [16]. Another challenge posed by the hyperspectral image which needs to be efficiently addressed for supervised classification is the so called 'curse of dimensionality'. It is also known as the Hughes phenomenon [17], where the high number of spectral bands and low number of labelled training samples result in an decrease in the generalisation capability of the classifier [18]. Much effort has been spent by researchers aiming to overcome these aforementioned classification difficulties. Generally, these problems can be tackled in two steps, 1) by extracting effective features from spatial/spectral domain and 2) by constructing suitable classification models for raw/extracted features.

In addition to remote sensing-based applications, HSI has also been considered as a promising tool for food quality evaluation as it can detect chemical properties which cannot be observed using conventional imaging systems. Among all processes in the food industry, assessing the quality of meat and meat products is extremely important to consumers since meat is an essential source of protein for humans. Conventional meat quality analysis consists of manually assessing meat and scoring it in terms of: texture, flavour and visual appearance. This is a time-consuming, destructive and, in many cases, an inconsistent process conducted by humans using specific measuring instruments or reagents [19]. As such, this approach is prone to subjectivity and human error. Therefore, the meat industry is keen to develop modern, state-of-the-art techniques to help monitor and control the overall quality of their meat products in a fast and non-destructive way, in order to succeed in the extremely competitive marketplace [4]. In recent years, spectroscopy-based techniques have been developed for rapid quality assessment

of meat and meat products. However, these traditional approaches lack spatial information and therefore may fail to provide the most valuable or informative assessment results [20]. To this end, HSI systems which integrate both the imaging and spectroscopy techniques together have emerged to overcome this drawback of traditional spectroscopy. In spite of this, there is still no single and universally accepted methodology for analysing hyperspectral images in this context to assess the quality of meat.

The work presented in this thesis aims to develop novel data analysis techniques to address the aforementioned problems related to both remote sensing hyperspectral images and practical HSI-based applications, e.g. meat quality assessment. In summary, the following objectives are defined:

1. To develop an innovative data compression approach for hyperspectral image data, which is able to improve upon existing image compression methods for isotropic 3D images.

2. To define a novel feature extraction method that improves the subsequent classification accuracy when applied to hyperspectral images.

3. To investigate how to construct a suitable classification model for improved hyperspectral image classification.

4. To investigate the feasibility of using an HSI system to predict meat eating quality and to improve the prediction performance using various data analysis techniques.

1.2 Main contributions of this research

In this thesis, several contributions have been made to address the main objectives as listed in Section 1.1, and novel techniques have been developed for improving the performance of HSI-based analysis. The key contributions presented in this thesis are as follows:

1. An improved compression approach based on the 3D discrete cosine transform (DCT) is proposed for compressing hyperspectral image data where the novelty lies in the design of the quantisation matrix. Adapted originally from video compression, machine learning techniques have been applied for automatic prediction of the quantisation matrix, which provides a time-saving capability when choosing the correct quantisation matrix. Compared

with another popular compression technique based on the 3D discrete wavelet transform (DWT), more effective compression performance is achieved using the proposed methodology.

2. A novel spectral feature extraction method is proposed to improve the classification performance of the hyperspectral image. The method aims to apply the one-dimensional (1D) time series analysis technique, singular spectrum analysis (SSA), to the image transformed in the curvelet domain. Compared with a few state-of-the-art spectral feature extraction techniques, the proposed method provides more accurate classification results.

3. By investigating hyperspectral image classification using the joint sparsity representation classification (JSRC) method, an improved spectral similarity-based JSRC (SS-JSRC) method is proposed to overcome the weakness of the original JSRC. Additionally, joint bilateral filtering is proposed to be applied for extracting effective spatial features prior to applying the classification. By combining joint bilateral filtering and SS-JSRC together, the proposed classification framework shows its efficacy among several typical spectral-spatial methods for hyperspectral image classification.

4. The ability of HSI for evaluating meat quality non-destructively is investigated and several outcomes are achieved. Firstly, due to its excellent performance, a more suitable data prediction method based on support vector machines (SVM) replaces the commonly used partial least squares regression (PLSR). Secondly, the ability of HSI for predicting beef quality is compared with the traditional spectroscopy technique and the results demonstrate that HSI provides a more promising solution in the context of beef quality prediction. Furthermore, the aforementioned feature extraction technique SSA is applied to pre-process the hyperspectral data and this also leads to improved prediction accuracy. Finally, the HSI system is also tested for lamb quality assessment and offers valuable benchmarking results for other researchers working in this domain since there exists very little research in the area of lamb quality prediction using HSI.

1.3 Thesis overview

This remainder of this thesis is organised as follows:

Chapter 2 provides relevant background information on the research topics presented in this thesis by reviewing popular methods for hyperspectral image

compression, feature extraction and classification. A literature review has also been carried out in the area of spectroscopy-based technologies for meat quality assessment. In addition, the theoretical background and descriptions of two popularly used data analysis and machine learning techniques: principal component analysis (PCA) and SVM, which are used throughout the experiments in this thesis, are provided.

Chapter 3 presents an improved 3D DCT-based hyperspectral image compression approach, where the SVM is employed to determine the optimal weighting of an inter-band correlation parameter within the quantisation matrix that influences compression performance. In addition to conventional quantitative metrics for performance assessment, the classification accuracy on the compressed hyperspectral image is also adopted for quality-assured evaluation.

Chapter 4 describes a novel algorithm which exploits the properties of the curvelet transform and SSA to create a new feature extraction method for hyperspectral images. The curvelet transform and SSA are first introduced thoroughly. Then, a comprehensive set of experiments for assessing the performance of the proposed methodology are described in detail.

Chapter 5 presents a spectral-spatial classification framework for hyperspectral image data. Spatial features are extracted using joint bilateral filtering before being classified using the improved SS-JSRC. The proposed classification framework is benchmarked against several popular spectral-spatial classification methods to show its effectiveness.

Chapter 6 describes some practical HSI-based applications, where hyperspectral images of meat samples are used for non-destructive quality assessment. This work begins with choosing the right data prediction tool. Extensive experiments have been conducted on both beef and lamb samples. The HSI system is also compared with the popular spectroscopic system to demonstrate its superiority.

Chapter 7 presents a series of concluding remarks, and possible ideas and suggestions for future research topics are provided.

Chapter 2

Background and Related Work

2.1 Introduction

Following the motivations and objectives addressed in Chapter 1, background information and related work about hyperspectral image analysis are presented in this chapter. First, relevant literature of hyperspectral image compression, hyperspectral image classification and spectroscopic imaging-based meat quality assessment is reviewed. Then, principles of two popular techniques used throughout the thesis, principal component analysis (PCA) and support vector machine (SVM), are also introduced.

2.2 Hyperspectral image compression

Compression of hyperspectral images has been studied for a long time and is still considered as an active research area. According to different requirements in various applications, techniques for hyperspectral image compression can be divided into two categories, i.e. lossless and lossy compression [13]. For lossless compression, images are encoded without loss of information thus the original images can be fully recovered when decoded. Consequently, redundancy reduction within the data is achieved purely by examining its spatial and spectral distributions. Some widely used approaches for lossless compression include run-length encoding, predictive coding and entropy encoding [21]. Since it does not lose information, lossless compression can only achieve a very limited compression ratio of about 3:1 for hyperspectral images in some recent publications [22–24]. Although the performance of lossless compression does not usually meet the requirements for on-board systems [25], it is preferred for archiving the data and distributing to end-users [13]. Lossy compression, on the contrary, allows loss of information when

images are compressed. As a result, it helps to achieve a much higher compression ratio than lossless compression and appears more suitable for on-board systems, where users are allowed to tune parameters to obtain different rate-distortion performance according to their applications [13].

Although lossy compression can significantly reduce the amount of data for hyperspectral images, in most practical situations, lossless compression, which has high computational cost, is desired to avoid loss of information as the lossy compression may severely affect the accuracy for subsequent data analysis, i.e. land-cover analysis [12, 26]. In the following two paragraphs, some typical lossless compression techniques are consequently reviewed.

Adapted from 2D image compression algorithms, predictive coding is considered as the main trend for lossless compression of hyperspectral images. Thanks to the high inter-band correlation as well as the similarity of neighbouring pixels presented in the hyperspectral image, a predictor can be used to predict pixel values and the resulting prediction error is then coded using an entropy encoder, such as the popular arithmetic or Huffman encoder [27]. Therefore, efforts have been focused on designing optimal predictors for reducing data redundancy in the hyperspectral image. In the very first publication on lossless compression of hyperspectral images back to 1996, Roger and Cavenor [28] extended the standard differential pulse code modulation (DPCM) using several linear predictors which fell into spectral, spatial and spectral-spatial categories. As the variable-length coding method, both Rice coding and Huffman coding were chosen to encode the residues, where no codebook was required for Rice coding and adaptive codebooks were designed for Huffman coding based on the estimated instrument noise. Both coding methods could yield a relatively high compression ratio in the range from 1.6:1 to 2.0:1. As a variation of DPCM, the clustered DPCM (C-DPCM) was introduced in [29] and [30]. In the C-DPCM method, the spectral profiles of the hyperspectral image were clustered into different groups first. Then for each cluster, linear prediction was performed using separate sets of coefficients and the prediction error was coded using an adaptive entropy coder, with an achieved compression ratio of around 3:1.

Alternatively, Abrardo *et al.* proposed an error-resilient and low-complexity on-board lossless compression approach using distributed source coding (DSC) [31]. Instead of encoding the prediction error, each band was encoded separately according to its correlation with the previous band and joint decoding was required to decode each band iteratively. Although DSC suffered a slightly lower compression ratio of about 2.5:1 with respect to C-DPCM, its low complexity

made it possible to be implemented on board. Through reordering spectral bands to maximise inter-band correlation in the hyperspectral image, a more efficient scheme for lossless compression was presented in [32] and [33], with a higher compression ratio of about 3.3:1 achieved. In addition, there were also some other lossless compression approaches introduced, using techniques such as the lookup table (LUT) [34] and context-based adaptive lossless image coding (CALIC) [35]. Although higher compression ratios can be reached by such approaches, the cost of complex computations they require seems unaffordable for real-time applications.

Since redundancies within the hyperspectral image can be more effectively removed, lossy compression has been more widely explored by researchers, using typical approaches including vector quantisation (VQ) [36–39] and transform-domain approaches [40–42]. The procedure for VQ-based compression algorithm generally consists of two steps, which are codebook generation (training) and codevector matching (coding). In codebook generation step, similar vectors in the training set are clustered into a few groups and a representative vector named as a codevector is generated for each cluster using the generalised Lloyd algorithm (GLA) or the Linde-Buzo-Gray algorithm (LBGA) [43]. With the binary index assigned to each codevector in the codebook, the rest vectors can be referenced to the nearest codevectors in the coding step. In Chen's work [44], it was found that although it was simple, the traditional VQ approach was not able to compress the image at a low bit rate. Consequently, a new method based on Kronecker-product gain-shape VQ [45] was proposed for very low bit rate encoding of multispectral and hyperspectral images. In [39], a novel multivariate VQ approach was presented for hyperspectral image compression. In the codevector matching step, each spectrum was represented using a linear combination of two codevectors from the codebook and the reconstructed image distortion was reduced consequently.

By mapping the hyperspectral image into another space with more compacted energy, transform-based approaches are widely used for lossy compression at low compression bit rates. Typical transforms include Karhunen-Loève transform (KLT), discrete Fourier transform (DFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT). Although KLT is considered as the optimal transform with high compression ratio, its application is limited since it is rather complex and signal-adaptive [14]. For transform-domain compression, DCT and DWT are two commonly used techniques, which have been adopted in the Joint Photographic Experts Group (JPEG) and JPEG2000 standards, respectively. Basically, by applying them for hyperspectral image compression, these transforms help to remove correlation (redundancy) in both spatial and spectral domains,

followed by a quantiser and an entropy encoder. Usually, these approaches are employed to compress individual bands of the hyperspectral image and are taken as baseline techniques for benchmarking [46]. As a good approximation of KLT for Gauss-Markov processes but with less computational complexity, DCT represents the original image using a linear combination of some weighted basis functions that are related to its frequency components. An algorithm based on the 3D DCT was proposed by Abousleman *et al.* for hyperspectral image compression [47], employing the trellis coded quantisation (TCQ) scheme. It was shown that the 3D DCT system achieved much higher compression ratio than systems using DPCM, block truncation coding and various VQ schemes. More recently, Pan *et al.* proposed a low-complexity DSC scheme for hyperspectral image compression with the help of the 2D DCT, which significantly outperformed the DSC scheme applied in the wavelet domain [48].

As mentioned above, entropy coding should be applied to transformed coefficients. According to different entropy coding methods, compression algorithms based on DWT can be divided into two groups, which are zero-tree coding and context-based coding [49]. The most straightforward approach for zero-tree coding is the embedded zerotrees of wavelet coefficients (EZW), where the transformed wavelet coefficients can be successively quantised and encoded using the zerotrees. In [50], 3D EZW has shown its success for coding the hyperspectral image. By utilising zerotrees with higher degrees, Said and Pearlman extended EZW and introduced set partitioning in hierarchical trees (SPIHT) [42] with better performance. With the success of 3D SPIHT for video compression [40], it was also proposed for hyperspectral image compression in [51]. For context-based coding, the concept has been adopted in the JPEG2000 standard as the core entropy encoder. Annex N of Part 2 of JPEG2000 standard supports multi-component imagery compression which can be used for hyperspectral image compression. Even though Part 10 of JPEG2000 standard was designed for 3D images, it is not suitable for hyperspectral images because it requires the image as isotropic as possible. Many researchers have compared performances between 3D SPIHT and JPEG2000 multi-component. Fowler and Rucker have shown that JPEG2000 multi-component achieved the signal-to-noise ratio (SNR) of 0.1 to 0.3 dB higher than 3D SPIHT for several remote sensing images [12]. However, Tang and Pearlman have proved that 3D SPIHT yielded higher SNR, e.g. 1.5 to 3.5 dB, than JPEG2000 multi-component at various compression bit rates [51]. Nevertheless, 3D SPIHT provided comparable results to JPEG2000 multi-component as shown by Christophe *et al.* [2].

2.3 Hyperspectral image classification

Hyperspectral image classification is a complex problem comprising of several processes aiming at solving the difficulties addressed previously. The general supervised classification problem for the hyperspectral image is illustrated in Figure 2.1. Since features presented in the hyperspectral image are often redundant, an important early step to extract/select effective features is needed to maximise the separability between classes [1]. Once the feature extraction step is accomplished, choosing the specific classification algorithm to apply to those labelled training samples is the next task. A number of feature extraction techniques and classification algorithms have been proposed and explored for hyperspectral image classification, and an overview of such techniques is presented in following paragraphs.

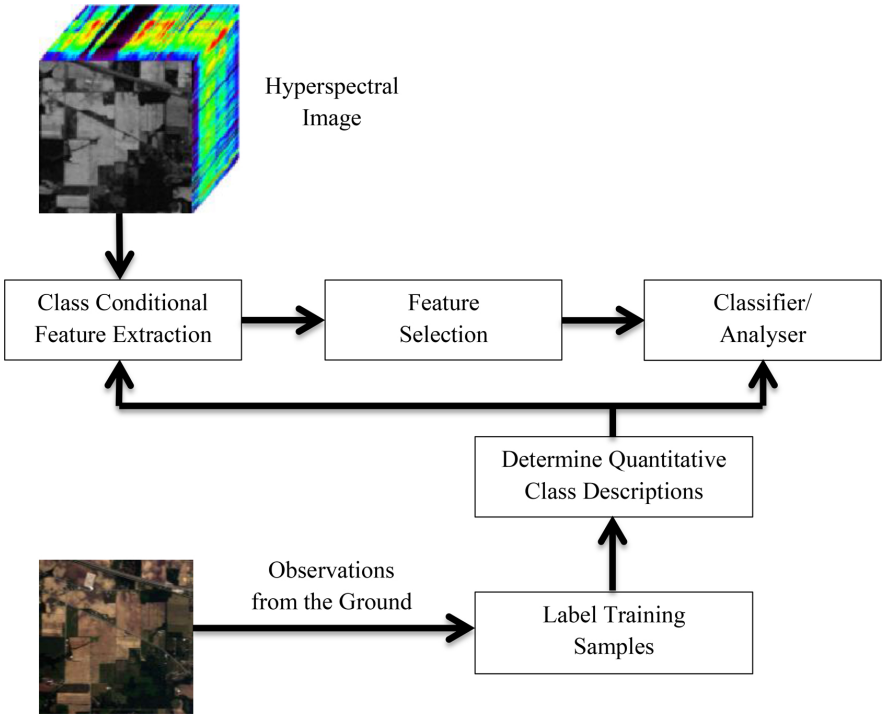


Fig. 2.1 The schematic diagram of hyperspectral image classification and analysis process (reproduced from [1]).

In addition to the curse of dimensionality, hyperspectral image also faces the problem of data redundancy due to its high inter-band correlation. Therefore, by mapping the high dimensional data into lower dimensional subspace without sacrificing the performance of following analysis, several unsupervised (if no prior information is available) and supervised (if training samples are available to obtain the projection matrix) feature extraction techniques have been proposed

for dimensionality reduction before classification. Dimensionality reduction could be considered as a linear or non-linear data transformation. Among unsupervised dimensionality reduction techniques, PCA [52, 53], also known as the discrete counterpart of the continuous KLT, is considered as the most widely applied linear approach. PCA is found to be able to transform the original data into a new feature space, where the variance in the original data is rearranged according to their magnitude, resulting in uncorrelated components in the new feature space [54]. In Section 2.5, background theories of PCA are explained in detail. The dimensionality reduction is consequently achieved by selecting the first a few principal components that are enough to represent most of the variance. In general, the first three principal components can preserve more than 90% energy of the original data [4]. It was found that for hyperspectral images, the conventional PCA is dominated by the visible and near-infrared bands since the variances of those spectral bands are higher than those of the remaining bands. To overcome the bias problem, Jia and Richards proposed the segmented PCA [54], achieving satisfactory results as well as reduced computational load. Other variations include the structured covariance PCA [55] and the folded-PCA [52], where both methods were proposed for fast computation. Independent component analysis (ICA) [56], which is famous for source separation, has also received considerable interest for dimensionality reduction of hyperspectral images, aiming at identifying a linear transformation that separates the original signal vector into a set of sources which are statistically independent. Similar approaches include minimum noise fraction (MNF) [57] and non-negative matrix factorisation (NMF) [58]. Using two cascaded PCA transformations, MNF projects the high-dimensional data into a new space consisting of two parts, the signal part and the noise-dominated part. This is also the reason why MNF is called the noise-adjusted PCA. By segregating the noise-dominated part in the data, dimensionality reduction can be achieved with MNF. Different from PCA or ICA which could only extract holistic features from the original image, NMF is a parts-based learning algorithm, decomposing a non-negative matrix into two non-negative matrices which are more intuitive and interpretable. One of the decomposed matrices is formed with a set of basis vectors, which can be utilised to project the original data into the lower dimensional subspace. The aforementioned linear approaches can be extended to non-linear ones through kernel methods at a cost of high computational load, leading to kernel PCA [59], kernel ICA [60], kernel MNF [61] and kernel NMF [62].

Unlike unsupervised dimensionality reduction techniques, supervised methods require prior information of labelled training samples to achieve the discriminative projections. Linear discriminant analysis (LDA) is an standard supervised dimensionality reduction technique for pattern recognition, where the highest degree of class separability is obtained by maximising the Raleigh quotient, i.e. the ratio of the between-class scatter matrix to the within-class scatter matrix [4]. Since the conventional LDA requires the within-class scatter matrix to be non-singular, it might fail when there are not enough training samples for all the classes in the hyperspectral image. To overcome the ill-posed problem, a preliminary feature extraction step, such as PCA, is added to project the original image into a lower dimensional space before applying LDA so that the within-class scatter matrix is non-singular [4]. However, in this way, dimensionality reduction is actually achieved with PCA, and LDA only helps to maximise the class separability in the PCA-projected space. Another problem of LDA is that if the difference within the class-mean vectors is small, the selected features could be unreliable. If the distance between one mean vector and the others is large, its class will eclipse the others when computing the between-class scatter matrix. Therefore, non-parametric weighted feature extraction (NWFE) [63] was proposed to weight every sample to get the weighted mean, and new non-parametric between/within scatter matrix can be defined to obtain more features. Similar to the kernel methods, these supervised dimensionality reduction techniques also suffer from slow speed due to high computational cost.

Additionally, instead of reducing the dimensionality, feature extraction can also be achieved by signal decomposition, both in spectral and spatial domain. For spectral-based signal decomposition, an impressive technique, namely singular spectrum analysis (SSA) [64, 65], has recently demonstrated its ability in effective spectral feature extraction for HSI data, with a great improvement compared with another similar technique, empirical mode decomposition (EMD) [66, 67]. Being the fundamental part of the Hilbert-Huang transform (HHT), EMD is a non-linear and non-stationary signal decomposition method that decomposes the signal into a finite number of intrinsic mode functions (IMFs). Each IMF represents a zero-mean frequency-amplitude modulation component that is often related to a specific physical process so that EMD is not biased with any predetermined basis or subband filtering process. Although EMD has shown its efficacy in time series decomposition (e.g. speech recognition), results reported in [67] indicated that spectral feature extraction using EMD could potentially reduce the following classification accuracy. Consequently, a more robust and data-driven technique,

the ensemble EMD (EEMD) [68, 69] was proposed to alleviate the addressed problem. Besides, a major drawback of EMD is its heavy computational load, making it less attractive than SSA. In a similar way, SSA could decompose the spectral profiles of HSI data into a few components as well, including the trend, oscillations and noise components. It was demonstrated in [64] that compared with EMD, SSA was able to enhance the spectral feature extraction performance while reducing the computational complexity.

Although spectral feature extraction attracts much attention from researchers in the past, limited contribution has been found without incorporating the spatial context presented in the hyperspectral image. The extended 2D versions of EMD [67, 70] and SSA [71] have been applied to each individual band in the hyperspectral image and incredible improvement has been noticed with the help of the extracted spatial information. However, the computational complexity of 2D EMD is still a problem, let alone hundreds of bands in a hyperspectral image [16]. Besides, it is noticed the extracted spatial information using 2D-SSA can be over-smoothed with blurred features, leading to classification error especially around edges. A simple but effective way for effectively exploiting the contextual information is based on the filtering techniques, such as morphological operators and wavelet decompositions. The morphological profile (MP) constructed by opening/closing morphological operators using a series of structure elements (SEs) with different sizes has been developed as a valuable state-of-the-art technique for extracting spatial geometrical details for HSI data. As shown in [72] and [73], MP and morphological attribute profile (MAP) were applied to the first a few principal components achieved from a hyperspectral image, generating the extended MP (EMP) and the extended MAP (EMAP). In [74], texture features were effectively extracted using Gabor wavelets, leading to improved classification accuracy. Besides, wavelet decomposition has also been employed for image denoising, which can be combined with MP [16] and EMD [70] for better classification performance.

Furthermore, various machine learning techniques have been proposed for hyperspectral image classification, including maximum likelihood [75, 76] or Bayesian estimation methods [77, 78], artificial neural networks [79–81], and decision trees [82, 83] *et al.*. Maximum likelihood classification assumes that each class ω_i can be explained by the Gaussian distribution and is modelled with a class-conditional probability density function (PDF). The task is then to assign a given spectrum $\mathbf{x}$ to the class ω_i with the maximum posterior probability $p(\omega_i|\mathbf{x})$, which is equivalent to $p(\mathbf{x}|\omega_i)\,p(\omega_i)$ according to the Bayes rule. However, classification results based

on maximum likelihood could potentially be unreliable since the assumption of the class-conditional PDF is not true for all situations. Without making any assumption about the data frequency distribution within each class, artificial neural networks, inspired by the function of human brains, are able to perform model-free function estimation using non-linear mapping structures. The most widely used artificial neural network model for hyperspectral image classification is the back-propagating multi-layer perception, including one input layer, at least one hidden layer and one output layer. As an extension of the multi-layer perception model, deep learning, which comprises deep neural networks typically deeper than three layers, has recently demonstrated its success for hyperspectral image classification [84, 85]. In order to achieve superior performance, the artificial neural networks need to be carefully designed, where optimal parameters can be obtained using an extensive training phase. As a result, their applications are hindered by the complex configuration and the expensive computing time. As a fast and statistical assumption free method, decision trees employ a multi-stage approach for label assignment, where a chain of simple decisions rather than a single, complex decision are made based on the results of sequential tests. However, it was noticed that decision trees also suffer from the Hughes phenomenon [82], where the classification accuracy declines as the number of features increases. Among all approaches, SVM has demonstrated its success in supervised hyperspectral image classification [5, 18, 52] as it can handle large input spaces efficiently and it is less sensitive to the Hughes phenomenon compared to artificial neural networks [86]. In particular, details of SVM are explained separately in this chapter.

2.4 Meat quality assessment using spectroscopic technologies

Visible and near-infrared (VIS-NIR) spectroscopy has been recognised as a rapid, low cost and non-destructive tool for determining food chemical composition [87]. The visible spectrum is part of the electromagnetic spectrum which is visible to human eyes, lying between 390 to 700 nm [88]. The NIR and short-wavelength infrared (SWIR) spectrum, whose wavelength ranges from about 780 to 3000 nm, is just between the visible and intermediate infrared region [89]. Although the SWIR region is usually covered by the VIS-NIR spectroscopy as well, it is left out of the name for simplicity. Over the past thirty years, VIS-NIR spectroscopy has been applied in the meat industry for predicting the meat quality since the

demand and consumption of meat and meat products are generally high in most developed countries [90].

Compared with conventional approaches to assess meat quality, VIS-NIR spectroscopy has many advantages. In addition to being non-destructive, it requires minimal sample preparation, needs no reagents, and produces no waste, allowing assessing numerous meat properties at the same time [4]. It is also suitable for on-line application where the rapidity of measurement allows the determination of several different attributes simultaneously. The success of VIS-NIR spectroscopy has been demonstrated in [91] where the FOSS FoodScan NIR spectrophotometer was approved by the association of analytical communities (AOAC) for commercial analysis of moisture, fat and protein in meat and meat products. However, the quality evaluation method in [91] is destructive since it requires representative samples to be homogenised using grinding before collecting spectral profiles. Additionally, simple discrimination of different meat products using VIS-NIR spectroscopy has been proved feasible in late 1990s, e.g. for differentiating frozen and unfrozen beef [92], broiler and local chickens [93] as well as detecting minced beef adulteration [94, 95]. Moreover, VIS-NIR spectroscopy was found to be less reliable when predicting technological and sensory attributes of meat and meat products [90]. A thorough review on the application of VIS-NIR spectroscopy for predicting meat and meat products can be found in [90], with numerical prediction results reported.

According to [90], sensory attributes of meat generally include colour, shape, marbling, odour, flavour, juiciness, tenderness and firmness, among which flavour, juiciness and tenderness are three most significant parameters influencing the meat eating quality [96]. Besides, visual factors such as lean meat and fat colours are also adopted by consumers to infer meat eating quality at the time of purchase [97]. Generally, sensory attributes can be quantitatively represented by a number of technological parameters, including surface colour in the Commission Internationale de l-Eclairage (CIE) L*, a* and b* colour space, ultimate pH, as well as tenderness parameters, such as Warner-Bratzler shear force (WBSF) and slice shear force (SSF) [98–100].

The redness is often used as an indicator of freshness by consumers, and it can be affected by many factors during production, processing and packaging. Meat surface colour is commonly measured in the 3D L*a*b* colour space using a hand-held colourimeter calibrated against a reference tile. Take [101] as an example, beef surface colour was recorded with a Konica Minolta CR-400 chroma meter, where L* indicates colour lightness ranging from 0 (black) to 100 (white),

a* defines the green-red colour degree ranging from 0 (green) to 255 (red) and b* is the blue-yellow colour degree ranging from 0 (blue) to 255 (yellow). It was noticed that L* and b* are highly correlated to the absorbance spectra at 1230 – 1400 nm and 1600 – 1710 nm respectively, where the intramuscular fat is related to these wavelengths as well [99]. This finding was also confirmed in [102], concluding that meat with a higher intramuscular fat content appears lighter. As mentioned earlier, fat content can be successfully determined with VIS-NIR spectroscopy. Consequently, there should be no doubt that colour prediction of L* and b* was also found to be promising by means of VIS-NIR spectroscopy in most studies [98, 99, 103, 104]. On the contrary, the green-red colour degree a* is related to the water content as well as the concentration of myoglobin, which are usually corresponding to the visible range in the electromagnetic spectrum [90]. By selecting proper wavelengths for analysis, a* could also be predicted successfully [105].

Ultimate pH is the most commonly measured technological parameters for meat since it is related to most sensory traits. It was found that meat with high pH values is potentially darker, has more bacterial and less flavour [98]. Thus, early assessment of ultimate pH in a non-destructive way is necessary to detect spoiled products. It is well known that ultimate pH affects the light scattering properties of muscle tissue and it makes the accurate prediction of ultimate pH using VIS-NIR spectroscopy possible, as long as there is a wide range of reference data available when constructing the prediction model [90, 98, 106]. Tenderness is identified as the most significant factor influencing the result of a meat eating quality test, where the tenderness can be either measured with a trained sensory taste panel or quantitatively using laboratory-based shear force tests [107]. Most studies adopted WBSF as the tenderness representation [98, 99, 103, 108], while some researchers used SSF as a fast alternative to WBSF [109]. The performance of VIS-NIR spectroscopy in predicting meat tenderness was generally poor, probably due to lack of precision when measuring the shear force as well as a limited variability in the reference data [90].

While spatial information is usually ignored in a traditional single-point VIS-NIR spectroscopy, it is extremely important especially for heterogeneous meat products (e.g. meat with fat). This fundamental prerequisite drives the development of HSI system, which is also known as imaging spectroscopy. In the past two decades, a number of innovative implementations of the HSI technique have been reported as promising tools for quality and safety control of various meat products, though there are not as much as the reported one with VIS-NIR

spectroscopy. In [110], the HSI system was proven to be effective for identifying the faecal and ingesta contaminants on poultry carcasses. Rather than utilising all spectral bands, only two or three bands were employed for contamination detection in the experiments. For beef quality assessment, HSI was used to predict colour parameters [101, 111], ultimate pH [101], tenderness in terms of WBSF [111–113] and SSF [101], water-holding capacity [114] and fatty acid composition [115]. Similarly, HSI was also applied to pork samples for predicting several quality attributes in [116] and [117], including colour, ultimate pH and drip loss, while others tended to use the achieved hyperspectral images for categorising pork quality with the help of texture filters [118, 119]. Lamb quality assessment using the HSI system has also been investigated by predicting water content/drip loss [20, 120], fat content [120], protein content [120], ultimate pH [20], colour [20] and tenderness in terms of WBSF [121]. Besides, the potentiality of the HSI technique for fast detection of lamb adulteration has been studied in [122], where good predicting results are yielded, indicating the great promise of HSI.

Although meat quality assessment using both VIS-NIR spectroscopy and HSI has been studied a lot in the past, the majority of them estimated the prediction accuracy with the help of partial least squares regression (PLSR) [20, 101, 114, 117, 120], while some researchers used multiple linear regression (MLR) [111] and feed-forward neural networks [116]. Moreover, feature extraction was rarely applied to the hyperspectral image before the multivariate data analysis except for PCA to identify important spectral bands [110, 112, 114]. Therefore, choosing an effective feature extraction technique and a suitable regression model for the HSI data could be a possible solution to improve the prediction of meat quality parameters. For validating the robustness of the regression model, there are generally two ways: internal cross-validation and external validation. Although the leave-one-out cross-validation which entails removing and predicting each sample sequentially seems more popular among researchers as shown in [20, 101, 111, 112, 114, 117], whether it can be truly considered as a form of model validation is arguable as the future performance of the model on external samples is left unknown [97]. For this reason, the external validation method has become of interest to practical applications in the meat industry [123].

2.5 Principles of related techniques

Two data analysis and machine learning techniques, PCA and SVM, which are employed in the experiments throughout the current study, are introduced accordingly down below.

2.5.1 Principal component analysis

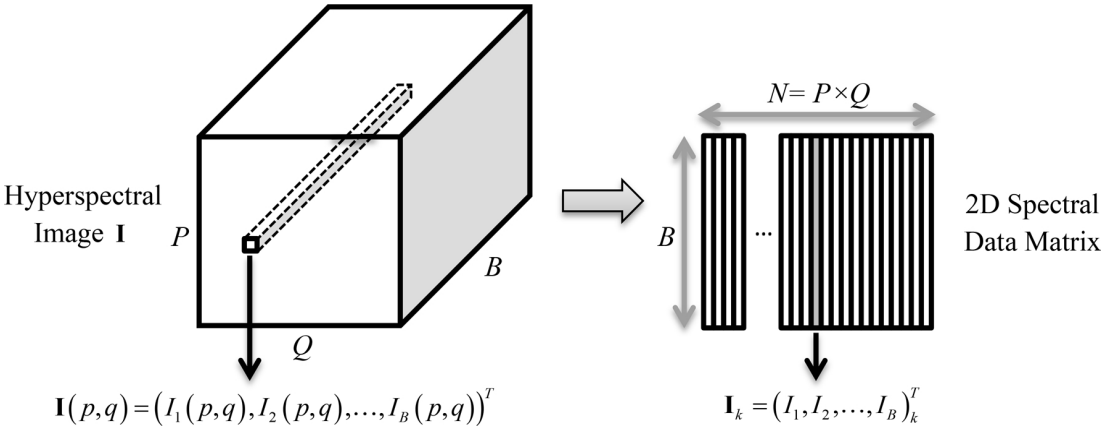


Fig. 2.2 The process of reshaping 3D hyperspectral image to 2D spectral data matrix.

There are generally three steps in the conventional PCA, which are covariance matrix calculation, eigenvalue decomposition of the covariance matrix and data projection. Assume a hyperspectral image with size of $P \times Q \times B$ is denoted as $\mathbf{I}(p,q) = (I_1(p,q), I_2(p,q), \ldots, I_B(p,q)) \in \mathbb{R}^B$ $(p = 1, 2, \ldots, P,\ q = 1, 2, \ldots, Q)$, where $\mathbb{R}$ represents the set of real numbers with the pixel intensity value $I_b(p,q)$ at all spectral channels $b = 1, 2, \ldots, B$ and P, Q refer to the spatial dimensions of the hypercube. The main purpose of PCA is to reduce the B-dimensional HSI cube to a m-dimensional cube (usually $m << B$), and the first step is to reshape the 3D image into a 2D spectral data matrix with size of $B \times N$, where $N = P \times Q$. Thus, the spectrum corresponding to an image pixel vector can be represented as $\mathbf{I}_k = (I_1, I_2, \ldots, I_B)_k^T$. The reshaping process is illustrated in Figure 2.2. Then, the B by B covariance matrix $\mathbf{C_I}$ for the image $\mathbf{I}$ is defined as:

$$\mathbf{C_I} = \mathrm{E}\left\{(\mathbf{I} - \mathrm{E}\{\mathbf{I}\})(\mathbf{I} - \mathrm{E}\{\mathbf{I}\})^T\right\}, \tag{2.1}$$

where $\mathrm{E}\{\cdot\}$ is the expectation operator. The covariance matrix can also be approximated via the following calculation,

$$\mathbf{C_I} = \frac{1}{N-1}\sum_{k=1}^{N}\Big((\mathbf{I}_k - \mathbf{I}_m)\,(\mathbf{I}_k - \mathbf{I}_m)^T\Big), \tag{2.2}$$

where $\mathbf{I}_m$ is the mean spectrum vector representing the average intensity value of the image across all bands and it is defined as below:

$$\mathbf{I}_m = \frac{1}{N}\sum_{k=1}^{N}(I_1, I_2, \ldots, I_B)_k^T. \tag{2.3}$$

The core step of PCA is based on eigenvalue decomposition of the obtained covariance matrix, which takes the form as Equation (2.4),

$$\mathbf{C_I} = \mathbf{P}^T\mathbf{D}\mathbf{P}, \tag{2.4}$$

where $\mathbf{D} = \mathrm{diag}\,\{\lambda_1, \lambda_2, \ldots, \lambda_B\}$ is a diagonal matrix composed of B eigenvalues of the covariance matrix $\mathbf{C_I}$ arranged in descending order, i.e. $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_B)$, and $\mathbf{P}$ is an orthogonal matrix of corresponding eigenvectors arranged in rows, i.e. $\mathbf{P} = (\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_B)^T$. Then the full PCA projection for $\mathbf{I}_k$ can be calculated as:

$$\mathbf{y}_k = \mathbf{P}\,(\mathbf{I}_k - \mathbf{I}_m)\,, \tag{2.5}$$

where $\mathbf{y}_k$ is the projected output of mean-shifted $\mathbf{I}_k$ so that the mean spectrum in each dimension is zero. This mean-shifting step is not necessary, but it makes the decorrelation of PCA more evident in subsequent stages. Since the full PCA projection keeps the highest variance in the first band and the lowest variance in the last band, the first a few projected bands would contain the majority of energy residing in the original hyperspectral image. By only keeping the first m rows in $\mathbf{P}$, the projected spectrum with reduced dimensionality can be achieved, as shown in Equation (2.6),

$$\mathbf{y}_k{}' = \begin{pmatrix} U_{(1,1)} & U_{(1,2)} & \cdots & U_{(1,B)} \\ U_{(2,1)} & U_{(2,2)} & \cdots & U_{(2,B)} \\ \vdots & \vdots & \ddots & \vdots \\ U_{(m,1)} & U_{(m,2)} & \cdots & U_{(m,B)} \end{pmatrix}(\mathbf{I}_k - \mathbf{I}_m)\,. \tag{2.6}$$

2.5.2 Support vector machine

Developed by Cortes and Vapnik in 1995 for binary classification problems, SVM [124] has been evolved into a powerful machine learning technique for high dimensional data classification and regression. Assume for a binary classification problem in a B-dimensional space, there are N training samples $\mathbf{x}_i \in \mathbb{R}^B$ with corresponding labels $y_i = \pm 1$. In order to avoid the situation that attributes with greater values dominate those in smaller numerical ranges, both training and test samples need to be normalised to the range $[-1, 1]$ or $[0, 1]$. If the data are separable by a hyperplane, there would be an infinite number of representations for the hyperplane. The goal of SVM is to separate the set of training vectors $\{(\mathbf{x}_i, y_i) \,|i \in [1, N]\}$ into two classes by finding an optimal hyperplane maximising the margin, which is the nearest distance between two vectors from different classes. The hyperplane H_p is named as the canonical hyperplane and it is defined as follows:

$$\mathbf{w} \cdot \mathbf{x} + b = 0, \ \forall \mathbf{x} \in H_p, \tag{2.7}$$

where $\mathbf{w} \in \mathbb{R}^B$ is the weight vector perpendicular to the hyperplane and $b \in \mathbb{R}$ is noted as the bias. For data points not on the canonical hyperplane, Equation (2.8) defines the distance from the data point $\mathbf{x}$ to the hyperplane H_p:

$$f(\mathbf{x}) = \frac{|\mathbf{w} \cdot \mathbf{x} + b|}{\|\mathbf{w}\|}. \tag{2.8}$$

For the canonical hyperplane, parameters $\mathbf{w}$ and b are constrained by Equation (2.9) to maximise the margin,

$$\min_{i \in [1,N]} |\mathbf{w} \cdot \mathbf{x}_i + b| = 1. \tag{2.9}$$

Training vectors $\mathbf{x}$ satisfying $\mathbf{w} \cdot \mathbf{x} + b = \pm 1$ are used for describing the canonical hyperplane, and they are therefore known as the support vectors.

When the training vectors are linearly separable, the canonical separating hyperplane satisfies the following constraint:

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \ \forall i \in [1, N]. \tag{2.10}$$

According to Equation (2.8), given two support vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ from different classes, the margin, which is the distance d between them on the direction

perpendicular to the canonical hyperplane, is calculated as:

$$d = \frac{|\mathbf{w} \cdot \mathbf{x}_1 + b| - |\mathbf{w} \cdot \mathbf{x}_2 + b|}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}. \tag{2.11}$$

Therefore, maximising the margin is equivalent to minimising $\|\mathbf{w}\|$, leading to the following quadratic optimisation problem:

$$\begin{aligned} &\text{minimise} \quad \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to} \quad y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) \geq 1, \ \forall i \in [1, N]\,. \end{aligned} \tag{2.12}$$

However, in general, the training data are not always linearly separable. To enable a generalised optimal separating hyperplane, non-negative slack variables ξ are introduced to address the misclassified data and the optimisation problem in Equation (2.12) is modified as:

$$\begin{aligned} &\text{minimise} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{subject to} \quad y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) \geq 1 - \xi_i, \ \forall i \in [1, N]\,, \end{aligned} \tag{2.13}$$

where the constant C is a pre-determined regularisation parameter that controls the amount of penalty. The optimisation problem in Equation (2.13) can be represented as its dual formulation utilising Lagrange multipliers and solved using quadratic programming accordingly. With the determined parameters $\mathbf{w}$ and b by solving Equation (2.13), the classification result y_t for a test sample $\mathbf{x}_t$ can be calculated as:

$$y_t = \operatorname{sgn}\left(\mathbf{w} \cdot \mathbf{x}_t + b\right). \tag{2.14}$$

It should be noted that the optimisation algorithm for solving Equation (2.13) only depends on the dot products between vectors $\mathbf{x}_i$. Hence, training vectors can be mapped into a high dimensional feature space with non-linear kernel functions and the linear separability of data is enhanced consequently. In this way, calculations can be carried out in the input space rather than the high dimensional feature space with reduced computational complexity. An inner product in the feature space holds an equivalent kernel in the input space, shown as below:

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \Phi\left(\mathbf{x}_i\right) \cdot \Phi\left(\mathbf{x}_j\right), \tag{2.15}$$

where $\Phi(\cdot)$ is the non-linear mapping function and $K(\cdot)$ is the kernel function satisfying the Mercer's condition [125]. The Gaussian radial basis function (RBF) is the most widely used kernel for hyperspectral image classification, with the form below:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \ \gamma > 0, \tag{2.16}$$

where γ is the width parameter for the RBF kernel. To train the classifier, a grid search using cross-validation is employed to look for optimal values of C and γ by trying exponentially growing sequences. For instance, $C = 2^{-10}, 2^{-9}, \ldots, 2^{14}, 2^{15}$ and $\gamma = 2^{-15}, 2^{-14}, \ldots, 2^{9}, 2^{10}$.

As mentioned in the beginning, SVM was originally designed for the binary classification problem. However, hyperspectral image classification is usually a K-class problem. A number of methods have been proposed, where the one versus one classification approach is adopted in a well-known publicly available SVM library – LIBSVM [126]. In this methodology, $K(K-1)/2$ binary classifiers are applied to each pair of classes and the test sample is assigned to a class label with the highest number of votes.

For the SVM regression, theories are similar to the classification, but the target values y_i for training samples $\mathbf{x}_i$ are real numbers rather than ± 1 any more. The optimisation problem for SVM regression has the formation as:

$$\begin{aligned} &\text{minimise} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{subject to} \quad |y_i - \mathbf{w}\cdot\mathbf{x}_i - b| \leq \varepsilon + \xi_i. \end{aligned} \tag{2.17}$$

The parameter ε serves as a threshold which limits that the deviation from the prediction to actual target value is ε at most. Therefore, this is also known as ε-SVM regression. For a test sample $\mathbf{x}_t$, the predicted target value y_t is calculated as:

$$y_t = \mathbf{w}\cdot\mathbf{x}_t + b. \tag{2.18}$$

Similarly, the non-linear mapping can be applied to SVM regression for better performance as well.

2.6 Summary

In this chapter, related work about hyperspectral image analysis are reviewed in three topics: hyperspectral image compression, hyperspectral image feature extraction and classification, and meat quality assessment using spectroscopic technologies. Popular techniques for hyperspectral image compression, including both lossless and lossy compression, are discussed. For the hyperspectral image classification problem, extracting effective features and constructing suitable classifiers are two important steps for improving the classification accuracy. State-of-the-art spectral/spatial feature extraction techniques as well as some typical classifiers are introduced. Additionally, literature review of meat quality assessment using both VIS-NIR spectroscopy and HSI system is carried out, with general meat quality measurement methods described. Furthermore, theoretic background of the most widely used feature extraction technique PCA and the most popular classifier SVM has also been reviewed in detail. These two techniques will be used throughout in this thesis.

Chapter 3

Hyperspectral Imagery Compression Based on an Improved Three-Dimensional Discrete Cosine Transform Approach

3.1 Introduction

Although the development of hyperspectral imaging (HSI) leads to success in many areas, a major associated problem is the huge volumes of data it contains, resulting in not only high cost for storage but also greater analysing/processing time and difficulty in data transmission, especially the one from satellites to ground [26]. As a consequence, compression and coding of the hyperspectral image is highly desired in this context.

The aim of this chapter is to propose a solution for quality-assured, low-cost and lossy compression of the hyperspectral image. Ideally, a high compression ratio should be achieved with very limited degradation on the accuracy for subsequent analysis and application. Despite the suggestion for lossless compression to preserve the quality of the hyperspectral image [12, 26], it is found that a moderate amount of data loss would not affect the image quality in many applications [37, 127]. Herein quality-assured is to reach a trade-off between a high compression ratio whilst minimising the effects to following data analysis, i.e. classification accuracy.

In this chapter, both discrete cosine transform (DCT) and discrete wavelet transform (DWT)-based three-dimensional (3D) approaches are investigated, and an improved 3D DCT-based approach is proposed. As mentioned in Chapter 2, 3D set partitioning in hierarchical trees (SPIHT) provides similar or even better compression performance than JPEG2000 multi-component scheme. Therefore, 3D SPIHT is chosen as the representation of 3D DWT-based approach in the following experiments and benchmarked with the proposed 3D DCT-based method.

Although it was found DWT usually outperforms DCT in this context [14, 128], better results from DCT were reported by others as well [48]. Therefore, this chapter also aims to evaluate these two approaches in terms of the compression performance and any side-effects on subsequent data analysis. Using four publicly available remote sensing hyperspectral data sets, interesting results are produced, and the efficacy and efficiency of the proposed approach for quality-assured hyperspectral image compression are fully validated.

The remaining part of this chapter is organised as follows. In Section 3.2, technical details of the proposed 3D DCT-based approach are presented. Section 3.3 discusses the experimental data sets and evaluation criteria used in the experiments. Results and evaluation on land-cover analysis are given in Section 3.4. Finally, some concluding remarks are drawn in Section 3.5.

3.2 Improved 3D DCT-based approach

By combining the two-dimensional (2D) spatial DCT and the one-dimensional (1D) spectral DCT together [49], 3D DCT is formed for more effective compression of the hyperspectral image as it can fully exploit the correlation in the 3D hypercube [46]. Furthermore, it allows fast access to each band image by partially decoding the compressed image in band groups. In this section, an improved 3D DCT approach with learning-based determination of optimal quantisation table is presented.

3.2.1 From 2D DCT to 3D DCT

For 2D DCT in JPEG standard, a block size of 8×8 is chosen for block-based compression. For 3D DCT used for hypercube compression, similarly, an $8 \times 8 \times 8$ sub-cube is applied. As a result, every 8 bands is grouped when they are compressed, which enables each band to be accessed by decoding the grouped 8 bands, i.e. partially decoding the whole compressed hypercube.

Let $f(x, y, \lambda)$ denote the sub-cube with $N = 8$, the 3D forward DCT is defined as follows [129],

$$\begin{cases} F(u, v, w) = \frac{c(u)c(v)c(w)}{(N/2)^{3/2}} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \sum_{\lambda=0}^{N-1} f(x, y, \lambda) \\ \cos\left(\frac{2x+1}{2N} u\pi\right) \cos\left(\frac{2y+1}{2N} v\pi\right) \cos\left(\frac{2\lambda+1}{2N} w\pi\right), \end{cases} \tag{3.1}$$

where $u, v, w = 0, 1, \ldots, N-1$ and $c(\cdot)$ satisfies

$$c(k) = \begin{cases} \frac{1}{\sqrt{2}} & , k = 0 \\ 1 & , k \neq 0 \end{cases}.$$

In order to acquire lossy compression, the 3D transformed coefficients must be quantised. The step of quantisation is the main source of loss in the DCT-based coding. Therefore, the quantiser should make the entropy of those quantised coefficients smaller. The quantisation process is defined in Equation (3.2), where Q is the 3D quantisation table and C is the quantised result,

$$C(u, v, w) = round\left(\frac{F(u, v, w)}{Q(u, v, w)}\right). \tag{3.2}$$

The 3D DCT quantisation table is formed with 512 quantitative values. After the transform, low frequency components are located near the coordinate $(0, 0, 0)$ and high frequency components are located near the coordinate $(7, 7, 7)$. High frequency components can be discarded after the quantisation stage since human vision systems are more sensitive to low frequency components [130]. However, JPEG standard has only defined the standard quantisation matrix for 2D DCT-based compression. Thus, the 3D DCT quantisation matrix can be designed on one's own choice, as long as values in the matrix are small at low frequency while large at high frequency.

In general, the quantisation matrix is the key for DCT-based compression as it affects the performance of coding. The multiplication [131] and the sum of the three coordinates [132], uvw and $u + v + w$, were used by some researchers to determine the quantisation matrix, respectively. In the proposed approach, however, the quantisation matrix [133] below is used, where the spectral component, w, is non-linearly weighted to reflect the inter-band correlation. In fact, inter-band correlation decreases when two bands are further apart, similar to correlation between video frames. As a result, the matrix from video compression can be used for coding of the hyperspectral image, shown in Equation (3.3),

$$Q(u, v, w) = \text{round}(u + v + kw + 3),\ u, v, w \in [0, 7], \tag{3.3}$$

where k stands for an inter-band correlation coefficient between spectral bands, which has 26 values of choice: 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5 and 8. Therefore, in total there are

26 quantisation tables, denoted from No.1 to No.26, where No.1 is used for weak inter-band correlation and No.26 is for strong inter-band correlation.

3.2.2 Proposed learning-based approach

For a trade-off between the quality of the reconstructed image and the compression bit rate achieved, a quality level ranging from 1 to 99 is decided, where a smaller number refers to poorer quality of the compressed image hence a higher compression ratio and a lower compression bit rate. By adjusting the quality level, the desired compression bit rate can be achieved.

The quality level of the quantisation matrix specified in Equation (3.3) is defined as 50 and Q_{50} is denoted as the standard quantisation matrix. For a quality level (QL) higher than 50 and lower than 50, the quantisation matrices are different as shown in Equation (3.4) [134],

$$Q = \begin{cases} \frac{100-\mathrm{QL}}{50} \times Q_{50}, & \mathrm{QL} > 50 \\ \frac{50}{\mathrm{QL}} \times Q_{50}, & \mathrm{QL} < 50 \end{cases} . \tag{3.4}$$

As mentioned before, every 8 bands forms a group to be compressed in the 3D DCT transform. Therefore, the value of k for each group should not be the same considering that the inter-band correlation for each group could be different. According to some preliminary experiments, the compression performance is sensitive to the choice of k if only a fixed value is applied for the whole hyperspectral image. Thus, the quantisation table in [133] with only a fixed k may not be an optimal solution. Additionally, the value of k in [133] is hand-picked by trial and error so that the process is not efficient enough. As a result, the support vector machine (SVM) is employed to predict k for each group in the proposed compression approach. In each group containing 8 spectral bands, if the correlation coefficient between every two bands is calculated, a vector containing 28 correlation coefficients can be obtained. In order to train the prediction model, a few spectral band groups from a hyperspectral image will be selected first. The next thing is to determine optimal values of k for selected groups. Quantisation tables with all values of k shown in Section 3.2.1 have been applied to compress every group at a bit rate of 0.5 bpppb (bits per pixel per band) respectively by adjusting the quality level. Therefore, for each group, there are 26 reconstructed images at the same compression bit rate. The lowest mean squared error (MSE) among 26 reconstructed images indicates the best compression performance, where the corresponding k is chosen as the optimal value of the quantisation table for that

particular group. Thus, in the training set for the SVM, for every instance, it contains one 'target value' k and 28 'attributes' (correlation coefficients). The goal is to construct a regression model to predict the 'target value' k for remaining band groups given their correlation coefficients only. The performance of the above mentioned two quantisation matrices in Section 3.2.1 and the proposed SVM prediction-based quantisation matrix is compared in Section 3.3.3.

Entropy coding is completed after the quantisation process, where coefficients at $(0, 0, 0)$ in each $8 \times 8 \times 8$ cube are referred to as DC coefficients whilst all others are named as AC coefficients. For the Huffman encoder adopted in the experiments, DC coefficients and AC coefficients are separately coded. Let DC_i and DC_{i-1} represent DC coefficients for block i and block $i-1$, respectively. Considering a high correlation of DC coefficients in adjacent blocks, differential coding can be employed to encode DC values for improved efficacy, where only the difference of DC_i and DC_{i-1} needs to be Huffman coded [134].

For AC coefficients, the scanning order is from low frequency to high frequency items, in the connection order with the 3D zigzag shape. Please note that the scanning order here is fixed, enabling fast access to designated items in a look-up table. On the contrary, DWT relies on SPIHT tree to determine the connection between coefficients, thus it is unfixed and time-consuming though it is expected to produce better results of coding.

Except for DC and AC coefficients, the inter-band correlation coefficients k for each compression group need to be coded as well. Similar to DC coefficients, values of k are differential coded.

At the decoding stage, Huffman codes are decoded [135], followed by the de-quantisation step and the inverse DCT transform. Then, the reconstructed image can be achieved. Similar to other block-based coding techniques, one of the most noticeable degradations of 3D DCT is the block artefact, which is a regular pattern of visible block boundaries [136]. However, such artefacts do not necessarily degrade the performance of subsequent image analysis, as demonstrated in next two sections.

3.3 Data sets and experimental setup

In total, four publicly available hyperspectral data sets are used for performance evaluation, and the data sets, experimental setup, and quantitative evaluation criteria are discussed in detail below.

3.3.1 Data set preparation

There are four hyperspectral data sets used in the experiments, including the airborne visible/infrared imaging spectrometer (AVIRIS) scenes Moffett Field scene 1, Indian Pines and Salinas Valley, as well as the reflective optics system imaging spectrometer (ROSIS) scene Pavia University, which are widely used in this context. Details of these data sets can be found in Appendix A.

The first data set, Moffett Field scene 1 [137], is cropped from the top-left corner for simplicity of compression, and 16 noisy bands are removed to form a hypercube with dimensions of $512 \times 512 \times 208$. The cropped hypercube is then binned to a smaller size of $128 \times 128 \times 208$ by using the average spectrum to represent each 4×4 block in the spatial domain. Therefore, the image size is dramatically reduced for easy processing using the 3D SPIHT algorithm. In the second data set Indian Pines [10], 20 bands affected by water absorption and noise are removed first [138]. After that, an additional 19 bands are discarded so that 185 bands are remained as suggested in [139]. Finally, the hypercube has a new size of $144 \times 144 \times 184$, by cropping the image to allow each dimension to be divided by 8 with no remainder, for the easy implementation of DCT in block-based compression. Similar to Indian Pines, for the third data set Salinas Valley [140], only 184 out of 224 bands are used for testing. Also, the image is cropped in spatial dimension, forming a new hypercube sized $144 \times 144 \times 184$, i.e. same as the second data set. The last data set Pavia University [141] is also further cropped to $144 \times 144 \times 96$ in the experiments for easy implementation of the coding algorithms.

As the 3D DCT-based compression proposed in this chapter is adapted from video compression whose maximum pixel value is 255, experimental hyperspectral data sets need to be normalised to the range of 0 to 255 first, and the normalised hyperspectral images are used both for the proposed 3D DCT-based compression and the 3D SPIHT-based compression. Although images are transformed from 12/16 bits to 8 bits per pixel, subsequent image analysis is not influenced as shown in Section 3.4. Another thing that needs to be addressed is that considering the test image size, the decomposition level of 3D DWT in the experiments is set to three for all data sets.

To illustrate the contents of four experimental data sets after essential cropping, pseudo colour images of each data set are shown in Figure 3.1. As can be seen, these images contain different natural/urban scenes, which correspond to various regions of interest. More importantly, there are ground truth maps available for

the last three data sets, as shown in Figure 3.2, which enables classification of pixels for quality-assured performance evaluation of both compression approaches.

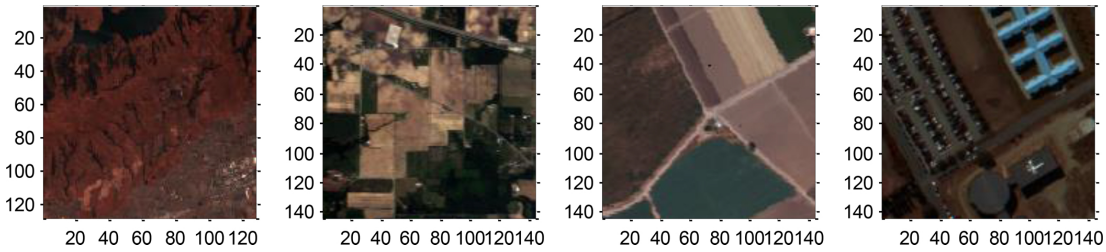


Fig. 3.1 Pseudo colour images of four experimental data sets, from left to right: Moffett Field scene 1 (after cropping and binning), Indian Pines (after cropping), Salinas Valley (after cropping) and Pavia University (after cropping).

Fig. 3.2 Ground truth maps for the last three data sets including Indian Pines (left), Salinas Valley (middle) and Pavia University (right).

In Figure 3.2, the ground truth for Indian Pines data set contains 16 classes, corresponding to alfalfa, corn, grass, hay, wheat, and so on. For Salinas Valley data set, the ground truth has 9 classes, including broccoli, lettuce, grapes, and so on. Finally, the Pavia University data set has 8 classes defined in the cropped ground truth, containing meadows, asphalt, soil and trees, etc. [138]. How these ground truth maps are used for the quality-assured performance evaluation is discussed in Section 3.3.2.

3.3.2 Evaluation criteria

For performance evaluation of both compression algorithms, several commonly used criteria are summarised. The first one is the compression bit rate, which is determined through dividing the compressed bit stream size by the original image size and its unit is bpppb.

Although a lower bit rate means a smaller number of data needed for the compressed hypercube, the degradation of the image quality is high. Therefore, a good compromise between the bit rate and the image quality is desired [139].

For quality assessment of the reconstructed image, both subjective and objective criteria are usually used and summarised below.

For subjective evaluation, it mainly relies on the judgement from some selected observers. First, visual capabilities of candidate observers are tested, and those qualified are chosen to assess the quality of the tested image by giving different ranks. Since human visual systems are highly adapted to extract structural information from the viewing area, subjective assessment is considered as the best way to evaluate images to be viewed by human beings. However, in practice, subjective evaluation is less preferable than objective assessments because the solution tends to be inconsistent, inconvenient, and time consuming [142], especially for the hyperspectral image that contains hundreds of bands. Besides, it is meaningless to compare a single band of the hyperspectral image because it is not typically designed to be observed by human viewers. Therefore, in the experiments, subjective assessment is completed by visually comparing two accumulated difference images between the reconstructed data set and the original data set under the same compression bit rate, thus a straightforward evaluation is achieved. The accumulated difference image I_{diff} for a B-band HSI is calculated as shown in Equation (3.5),

$$I_{diff}(i,j) = \sum_{k=1}^{B} |X(i,j,k) - Y(i,j,k)|, \tag{3.5}$$

where X and Y represent the original image and the reconstructed image, respectively.

Ideally objective evaluation approaches also prove good subjective assessments, and they are easier to apply as images could be automatically analysed for quality assessment without human involvement. Often, a full-reference approach is used for objective quality assessment, where a complete reference is supplied for comparison [143]. A simple statistical error metric called the signal-to-noise ratio (SNR) is adopted for this purpose. Under a given compression bit rate, SNR can be used to assess the rate-distortion performance in terms of reconstruction fidelity as detailed below.

Regarding the SNR, it is defined as the energy ratio between the original image and the noise, which is widely used for quality assessment. As SNR increases, the reconstructed image quality improves. However, it only has an approximate relationship with the human visual image quality because it compares the image pixel by pixel without taking the global content represented into consideration

[144]. The SNR is represented in decibels (dB), and its definition is given by

$$\mathrm{SNR} = 10 \log \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{K} X(i,j,k)^2}{\sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{K} \left(X\left(i,j,k\right) - Y\left(i,j,k\right)\right)^2}. \tag{3.6}$$

Despite the fact that SNR is simple to calculate, it is not perfectly matched to the perceived visual quality [145]. Therefore, another metric named as the structural similarity (SSIM) [146] is also applied to measure the similarity between two spectral bands. The formula for calculating SSIM between the original image X and the reconstructed image Y for each band is given in Equation (3.7), and the mean SSIM of all spectral bands will be used to compare the distortion,

$$\mathrm{SSIM}\left(X,Y\right) = \frac{\left(2\mu_X\mu_Y + C_1\right)\left(2\sigma_{XY} + C_2\right)}{\left({\mu_X}^2 + {\mu_Y}^2 + C_1\right)\left({\sigma_X}^2 + {\sigma_Y}^2 + C_2\right)}, \tag{3.7}$$

where μ is the mean intensity, C_1, C_2 are two constants, σ is the standard deviation and σ_{XY} can be estimated as

$$\sigma_{XY} = \frac{1}{MN-1} \sum_{u=1}^{MN} \left(\left(X_u - \mu_X\right)\left(Y_u - \mu_Y\right)\right). \tag{3.8}$$

To fully compare the similarity between original and compressed hyperspectral images, not only spatial difference should be addressed, but also the spectral integrity needs to be considered. Therefore, a spectral angle mapper (SAM) between original and reconstructed images is shown to illustrate it. The spectral angle of a single pixel is calculated as

$$\alpha\left(i,j\right) = \arccos\left(\frac{\sum_{k=1}^{B} \boldsymbol{\lambda}_X\left(i,j\right)\boldsymbol{\lambda}_Y\left(i,j\right)}{\sqrt{\sum_{k=1}^{B} {\boldsymbol{\lambda}_X}^2\left(i,j\right)}\sqrt{\sum_{k=1}^{B} {\boldsymbol{\lambda}_Y}^2\left(i,j\right)}}\right), \tag{3.9}$$

where $\boldsymbol{\lambda}$ represents the selected spectrum with B bands.

Although distortion measures shown above can give a general concept of how the reconstruction is distorted, they have little or even no correlation to the degradation in accuracy of subsequent image analysis [147]. Instead of looking at the rate-distortion performance, quality-assured assessments, such as hyperspectral image classification, can be adopted. Classification results from the original image and the reconstructed image are collected and compared to show how the compression and coding have affected the performance of subsequent data analysis.

Since rate-distortion performance is not always a good indicator for the quality of the reconstructed image [148], the quality-assured evaluation is particularly important for quality assessment of hyperspectral image compression.

With the available ground truth, SVM is employed for pixel-wise classification and it is applied to images compressed using 3D SPIHT and different 3D DCT techniques, respectively. By comparing classification results from images reconstructed by these techniques, a higher accuracy reflects better quality preserved from the quality-assured assessment point of view. The accuracy measurement used here is the overall accuracy (OA), which is the percentage of correctly classified labels among all pixels.

For SVM-based multi-class classification, a publicly available library BSVM [149] is adopted. In the experimental setup, 50% of pixel samples in each class are used for training the SVM model and the remaining 50% of pixel samples are used for testing the model. The spectral data are linearly scaled within the range from 0 to 1 since normalisation is required before applying SVM for classification. As the most commonly used kernel function, the Gaussian radial basis function (RBF) kernel is selected as it generates higher classification results than other kernels [150]. Two important parameters, the regularisation parameter C and the width parameter γ, are optimally determined by ten-fold cross-validation, where a 'grid search' is conducted to look for optimal values of C and γ by trying exponentially growing sequences. Finally, the optimised model can be obtained from SVM for testing the classification performance, and relevant results are reported in Section 3.4.

3.3.3 Comparison of different quantisation matrices

As mentioned previously, the performance of the proposed SVM prediction-based quantisation matrix is compared the other two quantisation matrices used for 3D DCT-based compression in this subsection. For simplicity, the quantisation matrices using the multiplication and the sum of three coordinates are referred to as Matrix 1 [131] and Matrix 2 [132], respectively. The quantisation matrix in [133] with a fixed k is addressed as Matrix 3. Therefore, the proposed one is named as SVM predicted Matrix 3.

In the proposed approach, the first five groups (40 bands) of Moffett Field scene 1 data set are selected to train the SVM regression model. Optimal values of k are calculated for each group, which are 3, 5, 5, 5.5 and 3.5 respectively. The performance of the previously mentioned two quantisation matrices and the proposed SVM prediction-based quantisation matrix is compared using all

four data sets. Under the same compression bit rates from 0.1 to 1 bpppb, the rate-distortion curves with SNR are illustrated in Figure 3.3. As can be seen, thanks to the weighting of inter-band correlation, the SVM predicted Matrix 3 consistently outperforms Matrix 1 and Matrix 2 for all data sets, especially at higher compression bit rates.

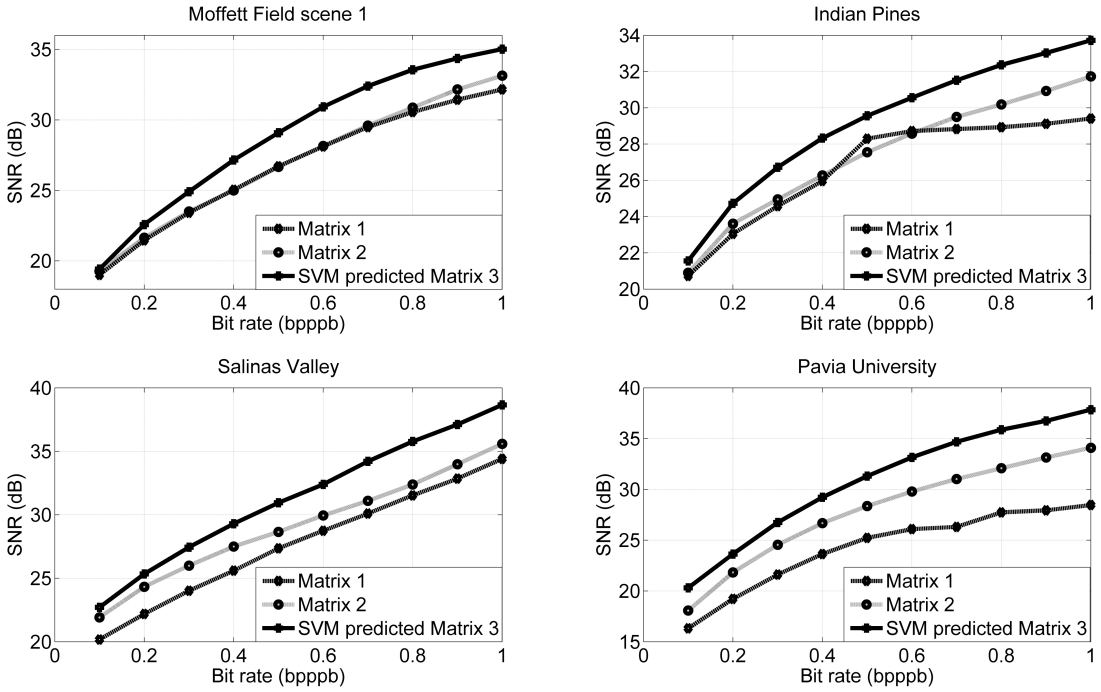


Fig. 3.3 Rate-distortion curves of different quantisation matrices for four experimental data sets.

3.4 Results and discussion

Using the four experimental data sets as described in the previous section, results from both subjective and objective assessments are analysed and compared in detail as follows.

3.4.1 Results of the subjective assessment

For all data sets, the compression bit rate is set as 0.1 bpppb and the accumulated difference images from 3D SPIHT and 3D DCT are compared in Figure 3.4, along with the average value of the intensity difference showing above each figure. Contrast of these difference images has been adjusted using histogram equalisation in order to be seen more clearly. Corresponding original images could be found in Figure 3.1 for reference.

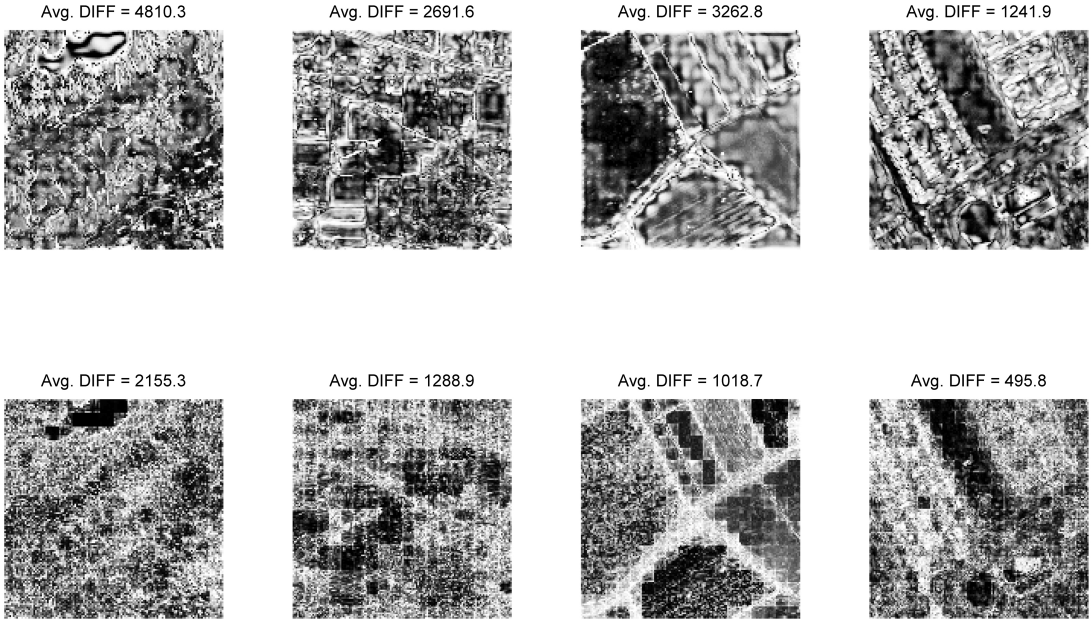


Fig. 3.4 Difference images between original and reconstructed images based on 3D SPIHT and 3D DCT (bottom) at 0.1 bpppb. From left to right, the four columns correspond to Moffett Field scene 1, Indian Pines, Salinas Valley and Pavia University, respectively.

As can be seen, the difference image could show some details at 0.1 bpppb, yet results from 3D SPIHT based on DWT seem to show more high-frequency components than those from 3D DCT approach, which can be proved by higher average difference values as well. Specifically, 3D SPIHT has over-smoothed high-frequency details of the image, while 3D DCT helps to preserve such details even under very low compression bit rates, though the block effect can be observed. In addition, those slight block artefacts with 3D DCT approach are due to the accumulated difference image, which can hardly be noticed in reconstructed images.

When the compression bit rate increases, due to limited space, only the Moffett Field scene 1 data set is compared. For other data sets, they are further evaluated using the quality-assured measurement as discussed in Section 3.4.3. Under various bit rates ranging within 0.2, 0.5, 0.8 and 1 bpppb, the contrast-adjusted difference images are shown in Figure 3.5 for comparison. Again, it can be seen that the improved 3D DCT approach produces better results at various bit rates. In addition, block effect caused artefacts are degraded at higher bit rates, and more detailed evaluation can be observed via quantitative results as given below.

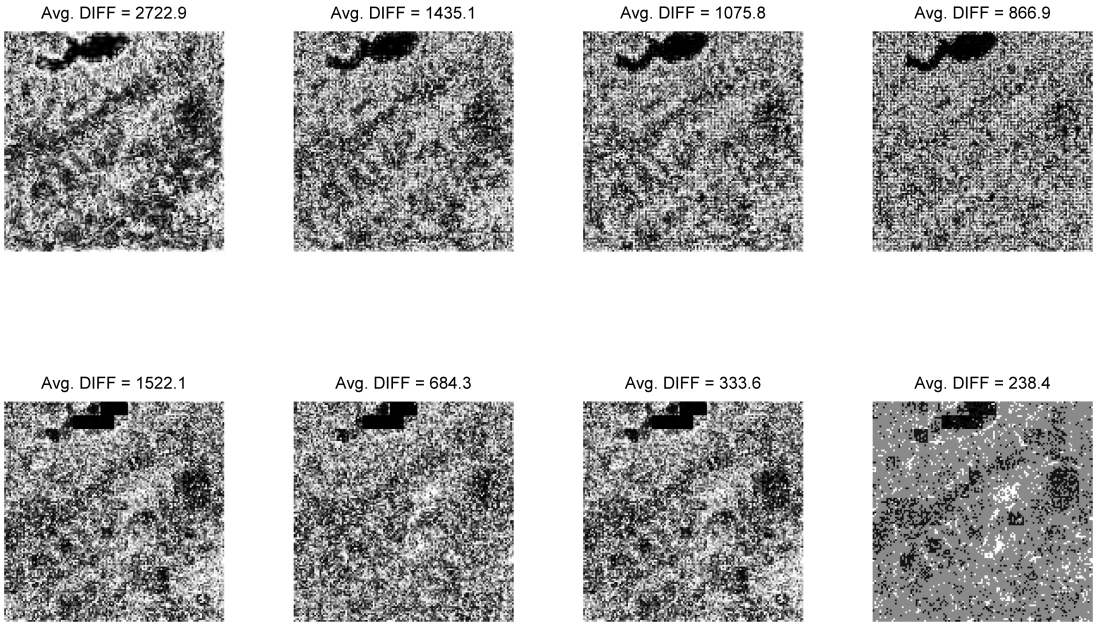


Fig. 3.5 Difference images of Moffett Field scene 1 based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different compression bit rates of 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right).

3.4.2 Results of the objective assessment

At a specific compression bit rate, the rate-distortion related metrics SNR and SSIM are employed for the objective assessment. A better compression approach is expected to produce higher values of SNR and SSIM in comparison with others. With the compression bit rate varying between 0.02 and 1 bpppb, results from the four experimental data sets are illustrated in Figure 3.6 and Figure 3.7 for evaluation.

For the four data sets, the proposed 3D DCT approach consistently outperforms the 3D SPIHT approach at different compression bit rates ranging from 0.02 to 1 bpppb, and an average gain of 5 – 8 dB in SNR is achieved. Regarding the SSIM, which is closer to human perception, 3D DCT also gives a more similar structure than 3D SPIHT.

Spectral angle maps for four data sets at compression bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb are shown in Figure 3.8 – 3.11, and an average spectral angle value is shown above each map. In order to show the difference more clearly, the contrast is adjusted for each data set and the colour bar is added below the spectral angle maps. It is noticed that with increased compression bit rate, average spectral angles of both 3D SPIHT and 3D DCT approach 0, while those of 3D DCT are much lower, which indicates the spectral distortion from the 3D DCT approach is less than that from the 3D SPIHT-based compression.

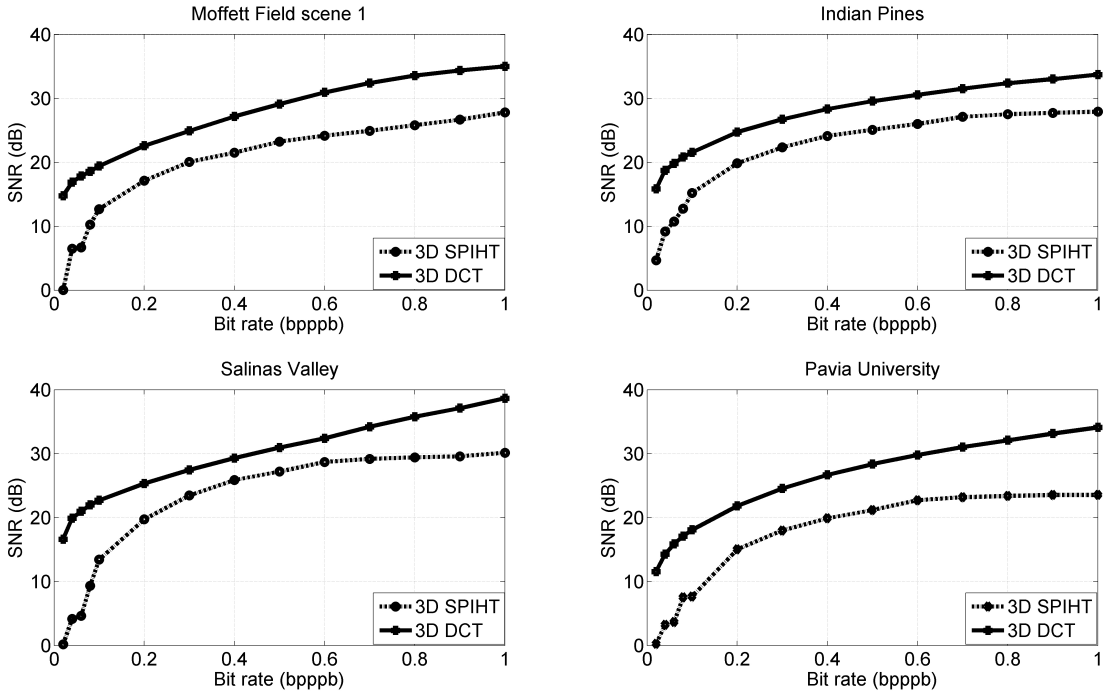


Fig. 3.6 SNR (dB) comparison between 3D SPIHT and 3D DCT compression performance for four experimental data sets.

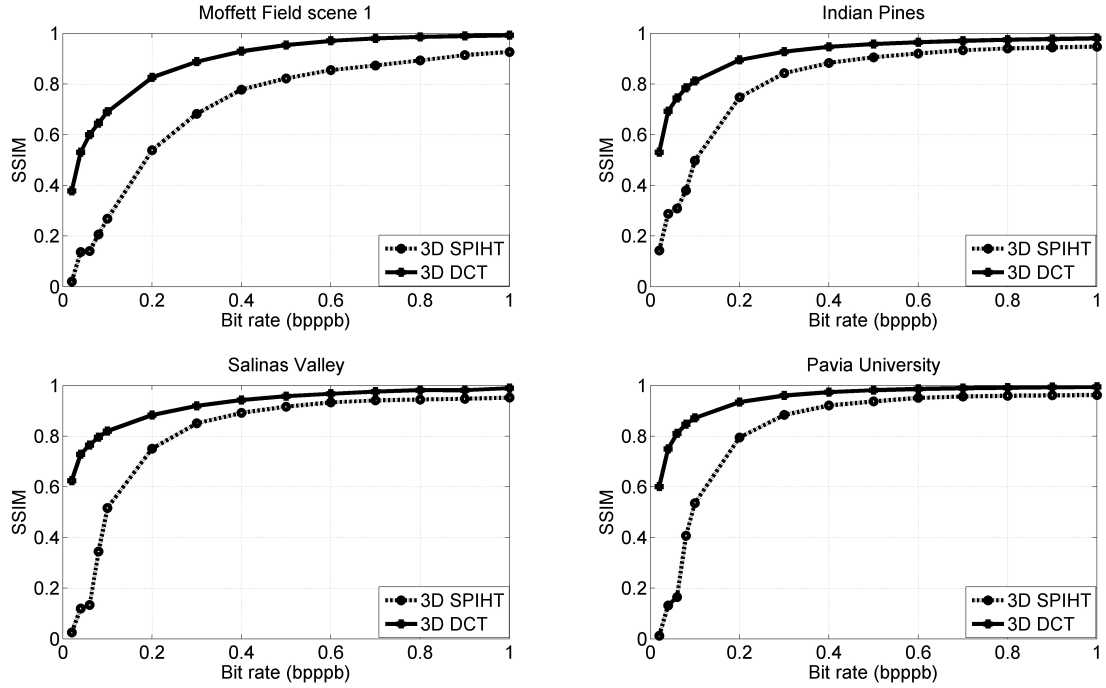


Fig. 3.7 SSIM comparison between 3D SPIHT and 3D DCT compression performance for four experimental data sets.

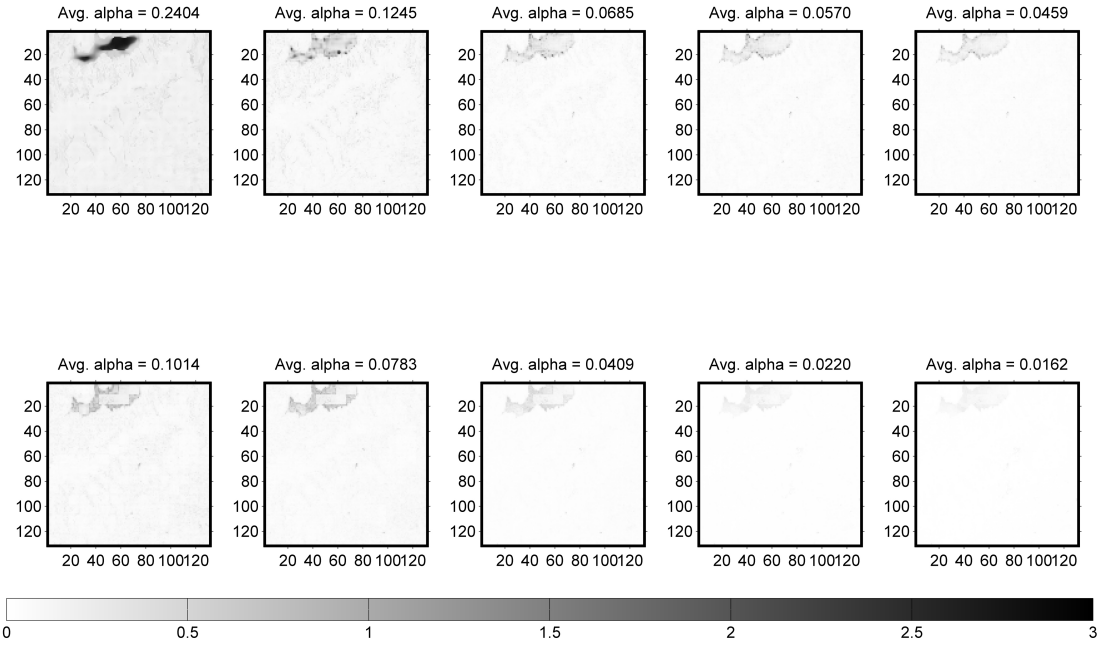


Fig. 3.8 Spectral angle maps of Moffett Field scene 1 based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right).

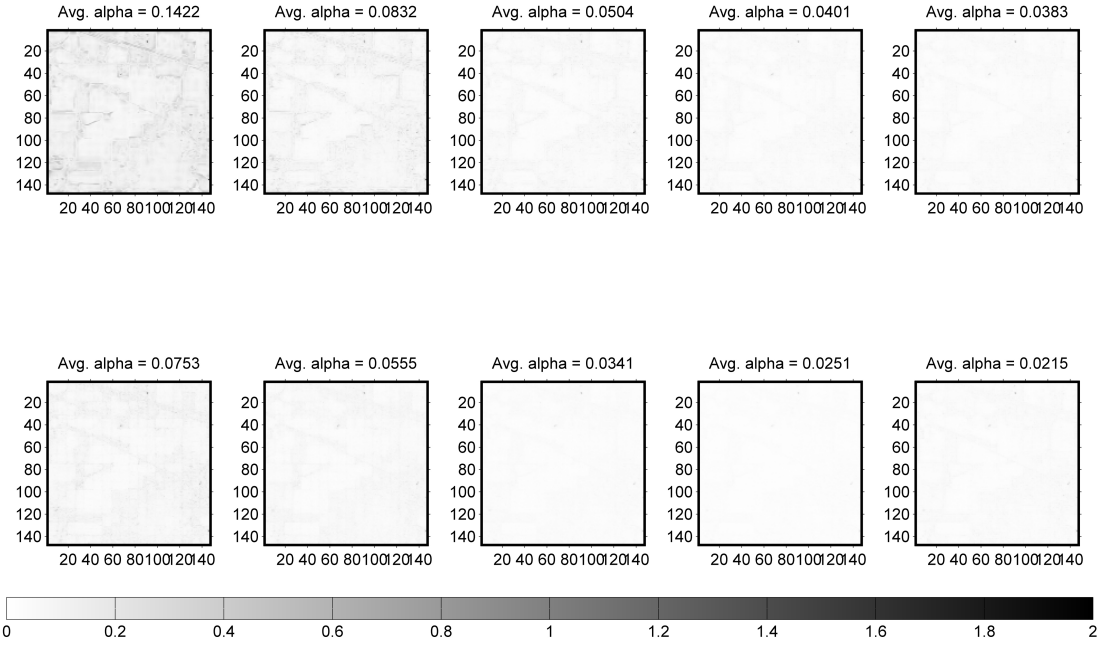


Fig. 3.9 Spectral angle maps of Indian Pines based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right).

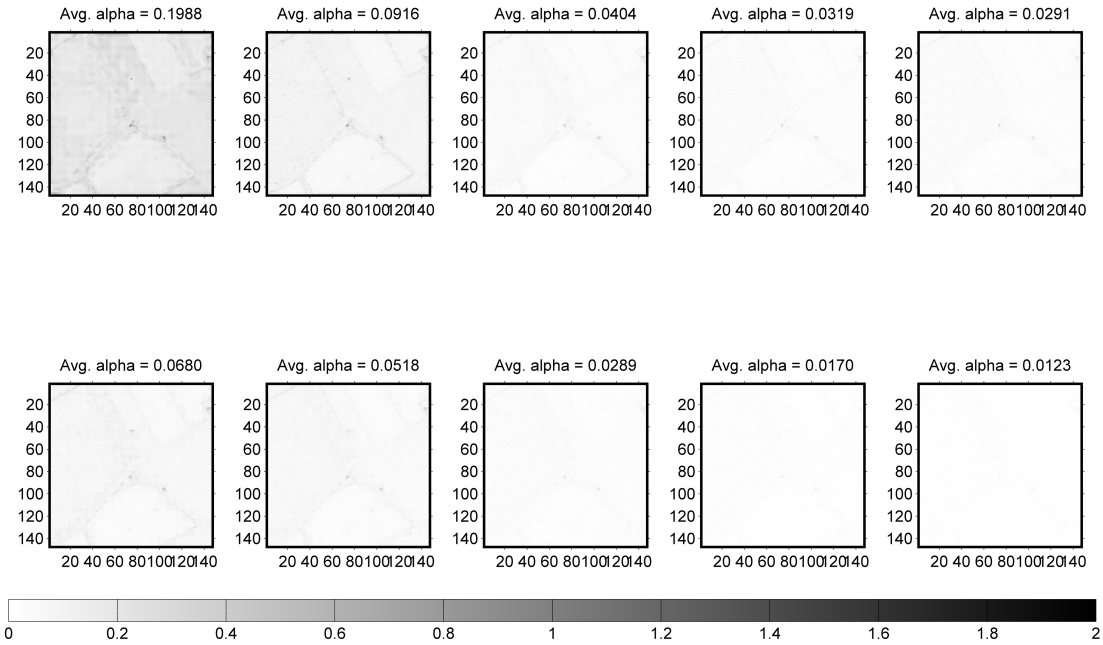


Fig. 3.10 Spectral angle maps of Salinas Valley based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right).

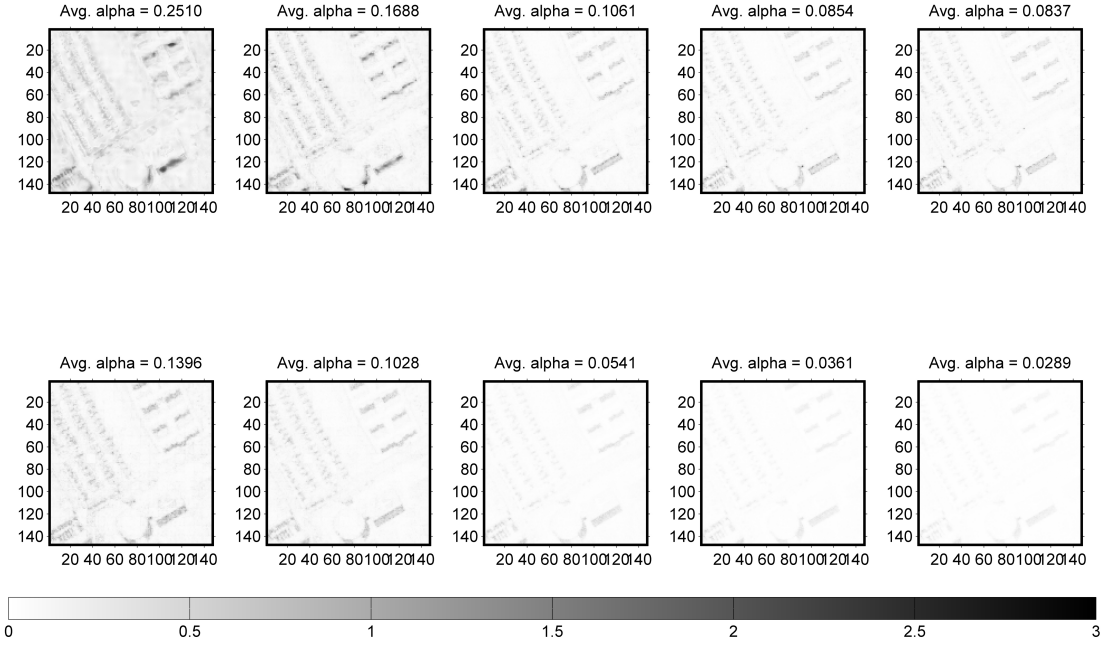


Fig. 3.11 Spectral angle maps of Pavia University based on 3D SPIHT (top) and 3D DCT (bottom) algorithms at different bit rates of 0.1, 0.2, 0.5, 0.8 and 1 bpppb respectively (from left to right).

3.4.3 Results of the quality-assured classification assessment

Since there is no ground truth available for the Moffett Field scene 1 data set, only the other three data sets, Indian Pines, Salinas Valley and Pavia University, are tested using the quality-assured assessment, where SVM-based classification is employed to measure the quality of subsequent data analysis.

For the three test data sets, optimal parameters determined by grid search-based cross-validation are summarised in Table 3.1. Using these optimal parameters, classification accuracies achieved with original images are also presented. In order to avoid error, the accuracy reported for each data set is an average one from ten random tests, where 50% of pixel samples in each class are used for training and the other 50% are used for testing.

After the optimal SVM model is learnt from the original image, it is applied to the reconstructed image for testing. A higher accuracy here indicates a better preservation of the image quality. Again, the test is carried out 10 times on randomly selected pixel samples and the average accuracy is then determined. For the three testing data sets at various compression bit rates, achieved overall classification accuracies from different compression approaches are plotted in Figure 3.12 for comparison.

Table 3.1 Optimal classification parameters and overall classification accuracies (%) for original data sets Indian Pines, Salinas Valley and Pavia University.

Data set	**Parameters**		**Classification accuracy (%)**
	C	γ	
Indian Pines	2^7	2^{-1}	93.68
Salinas Valley	2^9	2^{-1}	99.02
Pavia University	2^{10}	2^{-2}	98.56

As can be seen, classification accuracies for images reconstructed from the 3D DCT-based compression methods are consistently higher than those for images from the 3D SPIHT approach, especially at low compression bit rates from 0.02 to 0.5 bpppb. Besides, for 3D DCT-based compression methods, the proposed SVM predicted Matrix 3 is superior to other two matrices, which agrees with the results shown in Figure 3.3. When the bit rate becomes higher than 0.5 bpppb, classification accuracies on reconstructed images from the proposed approach are almost the same as those from original images. It means that the hypercube can be compressed at a moderate bit rate without degrading the quality of following data analysis. However, the compression has to be carried out using the proposed 3D DCT approach with the SVM prediction-based quantisation matrix rather

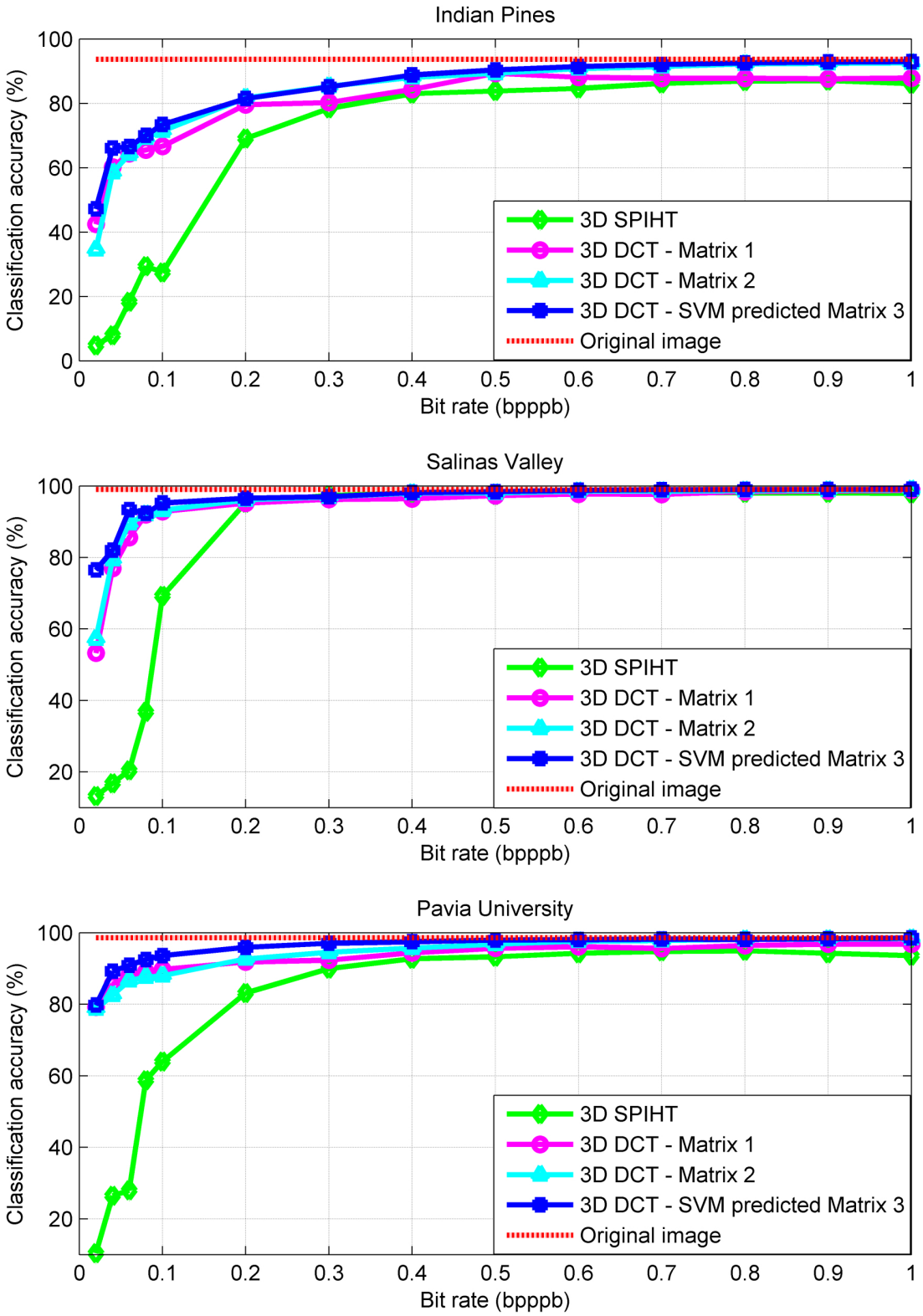


Fig. 3.12 Comparison of classification accuracies for three reconstructed data sets using 3D SPIHT and different 3D DCT-based compression algorithms with available ground truth, including Indian Pines (top), Salinas Valley (middle) and Pavia University (bottom).

than other methods, as those generate much worse results in Indian Pines and Pavia University even at a higher bit rate of 1 bpppb.

In comparison with the objective assessment, results from the quality-assured assessment are more consistent with the subjective assessment. For the bit rate lower than 0.2 bpppb, where reconstructed images are fuzzy, classification accuracies for images compressed using both techniques are low. Also, for images with poorer subjective evaluation, i.e. with 3D SPIHT compression approach, the classification error is higher. Besides, it is worth noting that it seems the block artefacts mentioned in Section 3.3 have little impact on the classification performance.

3.4.4 Complexity

For transform domain-based compression using 3D DCT and 3D SPIHT approaches, the computational complexity contains two parts, i.e. the transform itself and the following on sorting, quantisation and coding of coefficients. According to the analysis in [151], the computational complexity of 3D DWT is estimated as

$$C_{\mathrm{DWT}} = O\left(\frac{M}{2} LPB \frac{24}{7}\left(1 - 8^{-x}\right)\right), \tag{3.10}$$

where L is number of lines in the hyperspectral image, P is number of pixels per line, B is number of bands, x is the number of levels in the wavelet decomposition, and M is the length of the longest filter for the biorthogonal wavelet decomposition. For the CDF 9/7 filter used in the experiments, M is equal to 9.

The complexity of the fast 2D DCT is given as $O\left(N^2 \log N^2\right)$ for an $N \times N$ block [152]. Similarly, the fast 3D DCT complexity will be $O\left(N^3 \log N^3\right)$ for an $N \times N \times N$ block and N is 8 in the experiments as mentioned previously. Therefore, for an hyperspectral image with L lines, P pixels per line and B bands, the computational complexity for 3D DCT is determined as

$$C_{\mathrm{DCT}} = O\left(3LPB \log N\right). \tag{3.11}$$

Comparing Equation (3.10) and Equation (3.11), it is noticed that the computational complexity of 3D DWT is five times more than that of 3D DCT. In addition, the 3D DCT approach becomes more efficient in the second stage. This is because the sorting order in the 3D DCT approach is fixed and can be easily implemented using a lookup table for fast access and manipulation. On the con-

Table 3.2 Comparison of encoding time (in seconds) for Moffett Field scene 1, Indian Pines, Salinas valley and Pavia Univeristy data sets using 3D SPIHT and 3D DCT.

Data set	Coding scheme	Compression bit rates	
		0.2 bpppb	**0.5 bpppb**
Moffett Field scene 1	3D SPIHT	74.13	938.59
	3D DCT	4.04	10.74
Indian Pines	3D SPIHT	75.74	1093.23
	3D DCT	4.37	10.62
Salinas Valley	3D SPIHT	44.99	898.49
	3D DCT	3.19	9.18
Pavia University	3D SPIHT	23.66	272.17
	3D DCT	2.91	6.90

trary, the sorting order in the 3D SPIHT approach is determined by conditional comparison, hence requiring much more additional operations.

For the four experimental data sets, running time for the encoding stage at 0.2 and 0.5 bpppb is summarised in Table 3.2 for comparison. These experiments are carried out on a personal computer with an Intel Core i5-2400 CPU at 3.10 GHz. As can be seen, the proposed approach requires much less time than the 3D SPIHT approach, generating compressed images with better quality.

3.5 Summary

An improved 3D DCT-based approach is proposed, which appears to be a low-cost and quality-assured solution for lossy compression of the hyperspectral image. With the quantisation matrix determined using SVM via optimal weighting of inter-band correlation, the compression efficacy of the 3D DCT approach is further improved. In general, it generates a lower compression rate while successfully maintaining the quality of the image. The comparison between the proposed method and 3D SPIHT approach has suggested that 3D DCT has great potential to produce better compression performance as it preserves more high-frequency details. Consequently, the classification accuracy of the proposed methods achieved using SVM for land-cover analysis is also higher than that from 3D SPIHT as well as 3D DCT using other quantisation matrices. Besides, the classification-based quality-assured assessment is found to be consistent with visual inspections and distortion-based measurements, while it is difficult to gain a straightforward concept of image quality from the latter. Finally, it is suggested that, using the proposed approach, lossy compression of the hyperspectral image at a bit rate of no less than 0.8 bpppb is feasible as the degradation on image quality for subsequent analysis is negligible.

Chapter 4

Hyperspectral Image Feature Extraction by Exploiting the Curvelet Domain Using Singular Spectrum Analysis

4.1 Introduction

As mentioned previously, hyperspectral imaging (HSI) plays an important role in many areas. Among a number of HSI-based applications, one of the most widely applied area is remote sensing, where researchers could develop diverse algorithms for different tasks, e.g. target detection for military surveillance [3], data compression for faster transmission [39, 153] and data classification for land-cover analysis [52, 53, 64, 65, 71]. However, owing to the high spectral correlation within the hyperspectral image, data redundancy is inevitable. Furthermore, for remote sensing hyperspectral images, noise could be involved during the process of data acquisition and transmission. Therefore, effective de-noising and feature extraction of HSI data is crucial for remote sensing applications, in particular for supervised classification problems as discussed in this chapter.

It is assumed that directly applying de-noising techniques on the original image could probably remove fine features and noise at the same time, which is undesirable for following applications. To avoid this drawback, the multi-scale wavelet transform-based approaches are widely used for image de-noising, where the image is decomposed into a set of wavelet coefficients at different decomposition levels and noise could be removed from the low-energy channels in the transformed domain. A well-known algorithm called soft thresholding in the wavelet domain was proposed by Donoho in [154]. By removing small coefficients under a certain threshold and shrinking large coefficients, most unwanted noise could be discarded easily. For HSI data, Othman and Qian proposed a hybrid

spatial-spectral derivative-domain wavelet shrinkage approach based on the soft thresholding [155]. This algorithm works in the spectral derivative domain, in which the noise level is elevated and the signal regularity is different from the original domain of the hyperspectral image. Recently, Chen and Qian combined feature extraction and de-noising together, leading to a more effective de-noising method for HSI data using principal component analysis (PCA) and wavelet shrinkage [156]. In this approach, PCA transform is performed on the HSI data first and then a two-dimensional (2D) bivariate wavelet thresholding method is used to remove noise for low-energy PCA channels.

In this chapter, it is also proposed to combine feature extraction and de-noising approaches together for improving classification accuracies of remote sensing hyperspectral images. This work is inspired by [157] where the curvelet transform is applied to HSI data and the representation of noise-free signals in the curvelet domain is predicted using multiple linear regression (MLR). Although the wavelet transform has been widely applied for image de-noising, many studies have concluded that the wavelet transform cannot provide a good representation of anisotropic singularity [158–160]. In image processing, anisotropic singularities represent some significant parts of the image content, such as curves, edges or other details which help to interpret the image. For this reason, soft-thresholding directional curvelet coefficients that match image edges could achieve better noise reduction effect than the coefficients obtained in the wavelet domain. It has been proven that the curvelet transform could represent piecewise linear contours on multiple scales through few significant coefficients, leading to a better separation between geometric details and background noise [161]. Therefore, the curvelet transform is a good candidate for image de-noising and enhancement. After the curvelet transform, the coefficients of two adjacent bands still maintain the correlation similarity as the original hyperspectral image [162], which means spectral processing techniques could as well be applied in the curvelet domain. Feature extraction by MLR in [157] is achieved by estimating the noise-free band using all pixels in adjacent bands at the same time, while more accurate spectral feature extraction is proposed to be achieved by applying the above mentioned SSA pixel-wisely in this chapter. Besides, the denoising performance in [157] is assessed by adding noise to the original HSI data and comparing the mean squared error (MSE) and the mean structural similarity (MSSIM) between the original data and de-noised data, while the classification performance on the de-noised image is left unknown. Hence, the curvelet transform instead of the wavelet transform is adopted in the proposed approach and SSA is combined to extract

effective spectral features in the curvelet domain in order to improve classification accuracies for remote sensing hyperspectral images.

The chapter is organised as follows, the theory of the curvelet transform and SSA is introduced in Section 4.2. The proposed de-noising and feature extraction approach applied on HSI data is given in Section 4.3. Section 4.4 introduces the experimental setup, including data sets, the classifier and tuning of optimal parameters. Results and analysis are presented in Section 4.5. This chapter ends with a summary section.

4.2 Background principles

In this section, the theory of the curvelet transform and SSA presented in [64, 159, 163–166], which will be used in following sections, is reviewed.

4.2.1 The curvelet transform

Compared with the wavelet transform, the curvelet transform not only provides the multi-scale analysis on images, but also generates a better representation of anisotropic singularity. It was first proposed by Candès and Donoho in 1999, and their version is also called the first-generation curvelet transform [167]. However, this transform is rather complicated and needs at least four steps to complete, including a sub-band decomposition, smooth partitioning, re-normalisation and ridgelet analysis, where details can be found in [167]. A few years later, a simpler and faster version of the curvelet transform was developed, which is the second-generation transform named as the fast discrete curvelet transform (FDCT) [164]. Two forms of FDCT with the same computational complexity were proposed by Candès *et al.*, based on the unequally-spaced fast Fourier transform (USFFT) and the wrapping of specially selected Fourier samples, respectively. The curvelet transform used in this chapter is the one based on USFFT.

Let us assume there is a pair of smooth, non-negative and real-valued windows called the radial window $W(r)$ and the angular window $V(t)$, such that W is supported on $r \in (1/2, 2)$ and V is supported on $t \in [-1, 1]$. These windows will always satisfy the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j r\right) = 1, \; r \in (3/4, 3/2)\,; \tag{4.1}$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1,\ t \in (-1/2, 1/2). \tag{4.2}$$

For each scale j, the frequency window U_j is defined in the Fourier domain by

$$U_j(r,\theta) = 2^{-3j/4} W\left(2^{-j} r\right) V\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right), \tag{4.3}$$

where $\lfloor j/2 \rfloor$ is the integer part of $j/2$. Consequently the support of U_j is defined by the support of W and V, which will be a polar wedge.

A mother curvelet waveform $\varphi_j(x)$ is defined related to $U_j(\omega)$ ($U_j(r,\theta)$ will be abbreviated as $U_j(\omega)$), where the Fourier transform of φ_j is equal to U_j. Thus, all curvelets at scale 2^{-j} can be acquired by rotations and translations of φ_j, where the equispaced sequence of rotation angles are $\theta_l = 2\pi \cdot 2^{-\lfloor j/2 \rfloor} \cdot l$, with $l = 0, 1, \ldots$ such that $0 \leq \theta_l < 2\pi$, and the sequence of translation parameters are defined as $k = (k_1, k_2) \in \mathbb{Z}^2$. Then, the curvelets at scale 2^{-j}, orientation θ_l and position $x_k^{(j,l)} = R_{\theta_l}^{-1}\left(k_1 \cdot 2^{-j}, k_2 \cdot 2^{-j/2}\right)$ can be achieved, defined as

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}\left(x - x_k^{(j,l)}\right)\right), \tag{4.4}$$

where R_θ is the rotation by θ in radians and R_θ^{-1} is its inverse form,

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{4.5}$$

A curvelet coefficient of an element $f \in \mathbb{R}^2$ could be achieved by the inner product of f and a curvelet $\varphi_{j,l,k}$,

$$c(j,l,k) := \langle f, \varphi_{j,l,k} \rangle = \int_{\mathbb{R}^2} f(x)\, \overline{\varphi_{j,l,k}(x)} dx. \tag{4.6}$$

Discussion in this chapter on the digital curvelet transforms will always be in the frequency domain. Therefore, the curvelet coefficient could be re-expressed in the frequency domain via the Plancherel's theorem, where the integral of a function's squared modulus is equal to the integral of the squared modulus of its frequency spectrum,

$$\begin{aligned} c(j,l,k) &:= \tfrac{1}{(2\pi)^2} \int \hat{f}(\omega) \overline{\hat{\varphi}_{j,l,k}(\omega)} d\omega \\ &= \tfrac{1}{(2\pi)^2} \int \hat{f}(\omega)\, U_j(R_{\theta_l}\omega)\, e^{i\left\langle x_k^{(j,l)}, \omega \right\rangle} d\omega. \end{aligned} \tag{4.7}$$

However, in practice, instead of using the polar window defined in Equation (4.3), it is more often to use Cartesian equivalents. The Cartesian frequency window is shown as follows:

$$\tilde{U}_j(\omega) := \psi_j(\omega_1) V_j(\omega), \tag{4.8}$$

where $\psi(\omega_1) = \sqrt{\phi(\omega_1/2)^2 - \phi(\omega_1)^2}$ is defined as a bandpass profile in $\psi_j(\omega_1) = \psi(2^{-j}\omega_1)$ and $V_j(\omega) = V\left(2^{\lfloor j/2 \rfloor}\omega_2/\omega_1\right)$ with V still obeying Equation (4.2). By introducing the set of equispaced slopes $\tan\theta_l := l \cdot 2^{-\lfloor j/2 \rfloor}$ with $l = -2^{\lfloor j/2 \rfloor}, \ldots, 2^{\lfloor j/2 \rfloor} - 1$, for each $\theta_l \in [-\pi/4, \pi/4)$, the Cartesian window could be rewritten as:

$$\tilde{U}_{j,l}(\omega) := \psi_j(\omega_1) V_j(S_{\theta_l}\omega), \tag{4.9}$$

where S_θ is the shear matrix defined as below,

$$S_\theta := \begin{pmatrix} 1 & 0 \\ -\tan\theta & 1 \end{pmatrix}. \tag{4.10}$$

Similarly, curvelets should be re-defined in Cartesian form:

$$\tilde{\varphi}_{j,l,k}(x) = 2^{3j/4}\tilde{\varphi}_j\left(S_{\theta_l}^T\left(x - S_{\theta_l}^{-T}b\right)\right), \tag{4.11}$$

with $b := \left(k_1 \cdot 2^{-j}, k_2 \cdot 2^{-j/2}\right)$. Accordingly, the curvelet coefficient will be

$$c(j,l,k) = \int \hat{f}(S_{\theta_l}\omega)\tilde{U}_j(\omega) e^{i\langle b,\omega\rangle} d\omega. \tag{4.12}$$

Assume the input Cartesian arrays for an n by n image are in the form of $f[t_1, t_2]$, $0 \le t_1, t_2 < n$, a four-step implementation of FDCT via USFFT can be obtained as follows:

A. 2D FFT

Apply two-dimensional (2D) fast Fourier transform (FFT) to input arrays and obtain Fourier samples as:

$$\hat{f}[n_1, n_2] = \sum_{t_1,t_2=0}^{n-1} f[t_1, t_2] e^{-i2\pi(n_1 t_1 + n_2 t_2)/n}, \tag{4.13}$$

with $-n/2 \le n_1, n_2 < n/2$.

B. Interpolation

For each pair of scale and angle (j, l), Fourier samples $\hat{f}[n_1, n_2]$ are interpolated to get new values $\hat{f}[n_1, n_2 - n_1 \tan \theta_l]$ for $n_1, n_2 \in P_j$. P_j is a set below:

$$P_j = \{(n_1, n_2) : n_{10} \leq n_1 \leq n_{10} + L_{1,j}, n_{20} \leq n_2 \leq n_{20} + L_{2,j}\}, \tag{4.14}$$

where $L_{1,j}$ and $L_{2,j}$ are the length and width of a rectangle and (n_{10}, n_{20}) is the index of the pixel at the bottom-left of the rectangle.

C. Multiplication

Multiply the interpolated samples with the frequency window $\tilde{U}_j$ and obtain

$$\tilde{f}_{j,l}[n_1, n_2] = \hat{f}[n_1, n_2 - n_1 \tan \theta_l] \tilde{U}_j[n_1, n_2]. \tag{4.15}$$

D. Inverse 2D FFT

The last step is to apply inverse 2D FFT to $\tilde{f}_{j,l}$ to get the discrete curvelet coefficients:

$$C^D(i, j, k) = \sum_{n_1, n_2 \in P_j} \hat{f}[n_1, n_2 - n_1 \tan \theta_l] \tilde{U}_j[n_1, n_2] e^{i2\pi(k_1 n_1 / L_{1,j} + k_2 n_2 / L_{2,j})}. \tag{4.16}$$

Of all four steps, the discrete curvelet transform via USFFT requires $O(n^2 \log n)$ flops for computation and $O(n^2)$ for storage.

4.2.2 Singular spectrum analysis

As a well-established approach, SSA has been applied for time series analysis and forecasting widely used in different areas, including mathematics, economics and even biomedical engineering [168]. SSA shares the same theoretical foundations as PCA. Both of them are able to decompose the original time series into a linear combination of a new orthogonal basis, which includes eigenvectors generated from the diagonalisation of the data correlation matrix [169]. The main capability of SSA is that it can decompose the original series into some interpretable components, such as the trend, oscillations and unstructured noise [64]. The SSA algorithm consists of two stages including the decomposition and the reconstruction, and it is briefly explained as follows.

A. Decomposition

Assume that $\mathbf{X}$ is a non-zero one-dimensional (1D) series vector with length N, i.e. $\mathbf{X} = (x_1, x_2, \cdots, x_N)$. Given a window length L $(1 < L < N)$, the original series $\mathbf{X}$ is mapped to K lagged vectors, $\mathbf{X}_i = (x_i, x_{i+1}, \cdots, x_{i+L-1})^T$ for $i = 1, 2, \cdots, K$, where $K = N - L + 1$. The window length should be chosen properly depending on different applications. Then the trajectory matrix is formed as:

$$\mathbf{T} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_K \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix}. \tag{4.17}$$

It could be noted from Equation (4.17) that $\mathbf{T}$ has equal elements x_{ij} on the anti-diagonals where $i + j =$ constant. Therefore $\mathbf{T}$ is a Hankel matrix with size of $L \times K$.

The next step is to compute the singular value decomposition (SVD) of the trajectory matrix $\mathbf{T}$. First, eigenvalues of $\mathbf{T}\mathbf{T}^T$ are calculated and sorted in the descending order, i.e. $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d)$, where d is the rank of $\mathbf{T}$. Let the corresponding eigenvectors be $(\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_d)$, and the result of SVD is shown in Equation (4.18),

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_d, \tag{4.18}$$

where $\mathbf{T}_i = \sqrt{\lambda_i}\mathbf{U}_i{\mathbf{V}_i}^T$ $(i = 1, 2, \cdots, d)$ and it is called the elementary matrix with rank 1, and $\mathbf{V}_i = \mathbf{T}^T\mathbf{U}_i/\sqrt{\lambda_i}$ are often referred to the principal components of the matrix $\mathbf{T}$. Generally, the contribution of the elementary matrix $\mathbf{T}_i$ to the trajectory matrix $\mathbf{T}$ is determined by the ratio of each eigenvalue and the sum of all eigenvalues, as shown in Equation (4.19),

$$\eta_i = \lambda_i / \sum_{i=1}^{d} \lambda_i. \tag{4.19}$$

B. Reconstruction

The first step of reconstruction is grouping, where the set of indices $\{1, 2, \cdots, d\}$ is divided into m disjoint subsets $I_1, I_2, \cdots, I_m$. Assume $I = \{i_1, i_2, \cdots, i_p\}$ is one group in the disjoint subsets, the trajectory matrix corresponding to this particular group I can be acquired as $\mathbf{T}_I = \mathbf{T}_{i_1} + \mathbf{T}_{i_2} + \cdots + \mathbf{T}_{i_p}$. Similarly, the resultant trajectory matrices can be computed for each group and consequently

Equation (4.18) is expanded to Equation (4.20):

$$\mathbf{T} = \mathbf{T}_{I_1} + \mathbf{T}_{I_2} + \cdots + \mathbf{T}_{I_m}. \tag{4.20}$$

The last step is diagonal averaging, which hankelises each grouped matrix $\mathbf{T}_I$ and then transforms it into a new series with length N. Assume $\mathbf{Y}_1 = (y_1, y_2, \cdots, y_N)$ is the transformed 1D series of $\mathbf{T}_{I_1}$, elements in $\mathbf{Y}_1$ can be calculated as Equation (4.21),

$$y_k = \begin{cases} \frac{1}{k} \sum_{j=1}^{k} y^*_{j,k-j+1}, & 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{j=1}^{L^*} y^*_{j,k-j+1}, & L^* \leq k \leq K^* \\ \frac{1}{N-k+1} \sum_{j=k-K^*+1}^{N-K^*+1} y^*_{j,k-j+1}, & K^* < k \leq N \end{cases}, \tag{4.21}$$

where $L^* = \min(L, K)$, $K^* = \max(L, K)$, $y^*_{j,k-j+1}$ refers to the elements in $\mathbf{T}_{I_1}$, $y^*_{j,k-j+1} = y_{j,k-j+1}$ if $L < K$ and $y^*_{j,k-j+1} = y_{k-j+1,j}$ if $L \geq K$. Then, the initial series $\mathbf{X} = (x_1, x_2, \cdots, x_N)$ is decomposed into m series, like shown in Equation (4.22),

$$\mathbf{X} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_m. \tag{4.22}$$

Therefore, the original series vector could be reconstructed by only using the first or the first a few groups generated from its eigenvalues and the rest could be discarded as noise. However, there is no general rule for grouping. Like the window length L, one can choose the eigenvalue grouping (EVG) value depending on their applications. If all decomposed components are involved in the EVG, the reconstructed series will be the same as the original series.

4.3 The proposed approach: applying SSA in the curvelet domain

As briefly mentioned in Section 4.1, the proposed de-noising and feature extraction approach will be applied on hyperspectral images and the flowchart is shown in Figure 4.1 for reference. Hyperspectral sensors have a relatively high spectral resolution that could generate hundreds of observation channels [18]. The obtained three-dimensional (3D) HSI data cube can be regarded as a stack of 2D images of the same scene corresponding to different wavelengths [157], and the correlation between each two adjacent bands is fairly high. The first step is to perform the

curvelet transform on each band of the hyperspectral image and a few image stacks are generated after the transform as shown in Figure 4.1. By applying the curvelet trasnform on each band, the band correlation property can be preserved so that SSA is able to exploit the spectral signature. The next step is to apply SSA on the spectral dimension for each detail image stack for further feature extraction and de-noising. After that, the de-noised detail images at each band are gathered according to their original location, followed by the inverse curvelet transform to get the de-noised hyperspectral image.

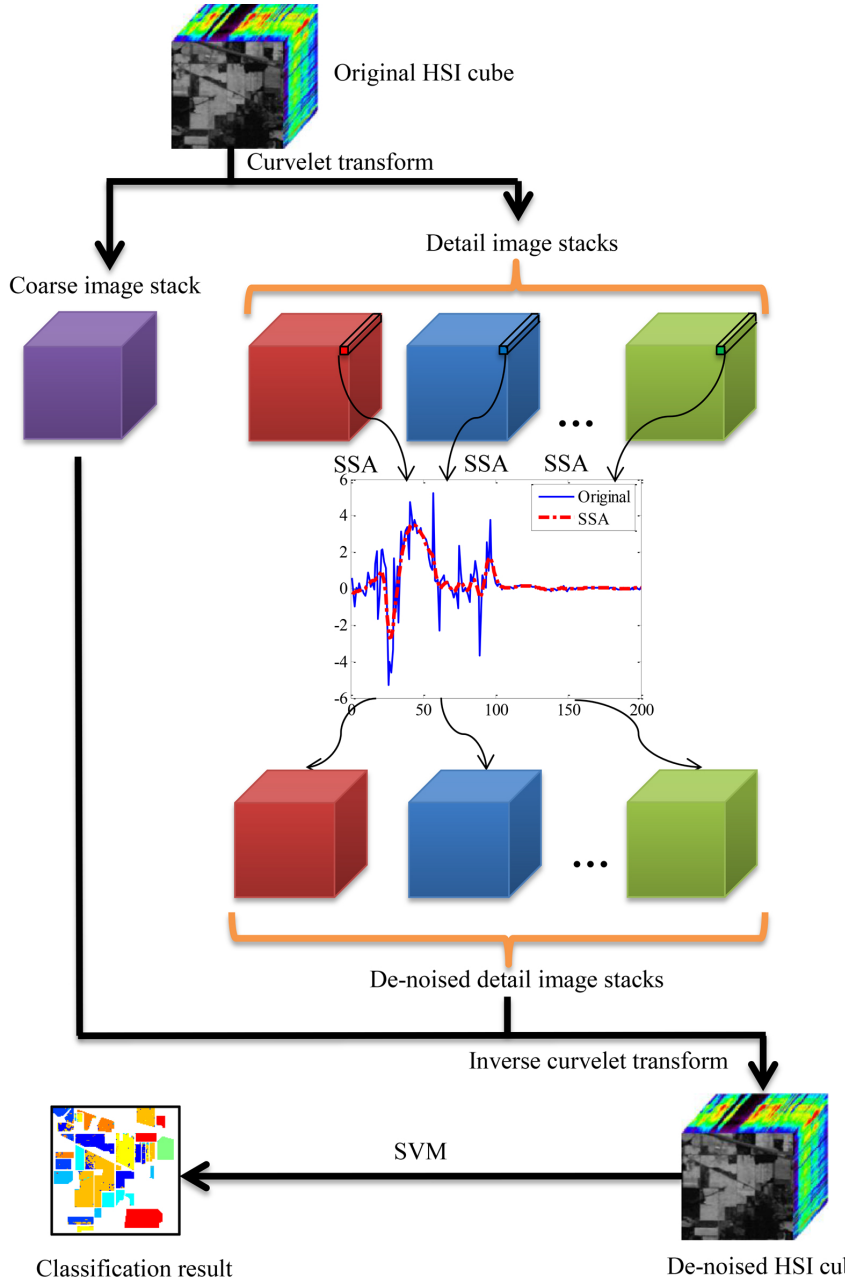


Fig. 4.1 Flowchart of the proposed methodology: applying SSA in the curvelet domain.

In this chapter, a hyperspectral image with size of $P \times Q \times B$ is denoted as $\mathbf{I}(p,q) = (I_1(p,q), I_2(p,q), \ldots, I_B(p,q)) \in \mathbb{R}^B$ $(p = 1, 2, \ldots, P,\ q = 1, 2, \ldots, Q)$, where $\mathbb{R}$ represents the set of real numbers with the pixel intensity $I_b(p,q)$ at all sensor channels $b = 1, 2, \ldots, B$ and P, Q are the spatial dimensions of the hyperspectral image. The 2D curvelet transform via USFFT employed in this chapter is completed using the toolbox CurveLab (version 2.1.2) [170]. Similar to the wavelet transform, the curvelet transform could also decompose the image into a course image and several detail images. Just like most image processing algorithms, the curvelet transform requires the processed image to be a square whose dimension is a power of two. If the size of the original image is not a power of two, pixels with a value of zero are padded to the next larger power of two. Given the zero-padded HSI data set $\mathbf{I}\_pad(m,n) = (I\_pad_1(m,n), I\_pad_2(m,n), \ldots, I\_pad_B(m,n))$ $(m, n \in [1, N])$ consisting of B bands where each band has N^2 pixels, the curvelet transform is performed on a band image I_pad_b to obtain the curvelet coefficients corresponding to that band c_b^D:

$$c_b^D = \mathrm{CT}(I\_pad_b), \tag{4.23}$$

where CT stands for the discrete curvelet transform operation explained in Section 4.2.1. The decomposition results of the transform can be regarded as a superposition in the form as follows:

$$c_b^D = C_{b,J} + \sum_{j=1}^{J-1} W_{b,j}, \tag{4.24}$$

where $C_{b,J}$ is the coarse version of the original band image with low frequency contents and $W_{b,j}$ stands for the detail band image at scale 2^{-j} containing high frequency contents. With an N by N image, the default number of decomposition scales is $J = \log_2(N) - 3$ as set in the CurveLab toolbox. Take a band image with size of 128 by 128 for instance, the number of decomposition scales J will be four. According to the settings in the toolbox, there are eight orientations of the curvelet coefficients in the second and third scales, starting from the top-left wedge and increasing in a clockwise fashion, but for the coarsest and finest scales, there is only one direction. Thus, Equation (4.24) is rewritten as below,

$$c_b^D = C_{b,J,1} + W_{b,1,1} + \sum_{j=2}^{J-1} \sum_{l=1}^{8} W_{b,j,l}, \tag{4.25}$$

where $W_{b,1,1}$ represents the finest scale with one orientation. The schematic diagram of four decomposition scales corresponding to a 128 by 128 image is shown in Figure 4.2. Then after applying the curvelet transform to all bands, there will be a coarse image stack and 17 detail image stacks.

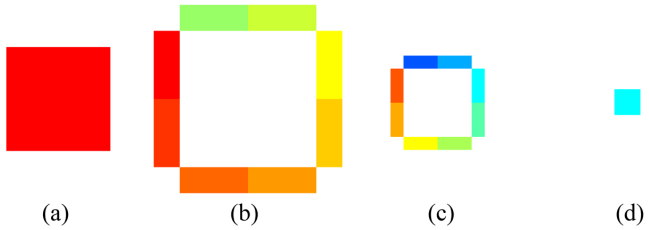


Fig. 4.2 Illustration of a 128 by 128 image with 4 decomposition scales, where (a) is the first (finest) scale with 1 orientation, (b) is the second scale with 8 orientations, (c) is the third scale with 8 orientations, and (d) is the fourth (coarsest) scale with 1 orientation.

As suggested by the commonly adopted wavelet de-noising rule, noise is removed by thresholding only wavelet coefficients of detail sub-bands, while keeping low frequency wavelet coefficients unchanged [171]. Therefore, in the proposed approach, the coarse image stack stays unaltered as well. For each detail image stack, SSA is applied to the spectral dimension for smoothing the spectral profile, followed by the inverse curvelet transform to get the de-noised HSI data cube. Given a detail image stack at scale 2^{-j} and orientation l, which is denoted as $W_{b,j,l}(x,y)$ $(b = 1, 2, \ldots, B,\ x = 1, 2, \ldots, X,\ y = 1, 2, \ldots, Y)$ containing $X \times Y$ coefficients in each band, the spectral series vector for one pixel can be constructed as $\mathbf{W}_{j,l}(x,y) = (W_{1,j,l}(x,y), W_{2,j,l}(x,y), \ldots, W_{B,j,l}(x,y))$. Then, the smoothing process on the spectral dimension by SSA could be achieved, denoted as below,

$$\mathbf{Y}_{j,l}(x,y) = \mathrm{SSA}_{L,EVG}\left(\mathbf{W}_{j,l}(x,y)\right), \tag{4.26}$$

in which SSA is the operation of the singular spectrum analysis listed in Equation (4.18) – (4.22), L and EVG are the parameters of SSA, and $\mathbf{Y}_{j,l}(x,y) = (Y_{1,j,l}(x,y), Y_{2,j,l}(x,y), \ldots, Y_{B,j,l}(x,y))$ is the smoothed spectral feature in the curvelet domain. After collecting all smoothed spectra for each detail stack, the inverse discrete curvelet transform is performed on the smoothed curvelet coefficients band by band and the de-noised HSI data cube is achieved, shown as Equation (4.27),

$$I\_d_b = \mathrm{ICT}\left(C_{b,J,1} + Y_{b,1,1} + \sum_{j=2}^{J-1}\sum_{l=1}^{8} Y_{b,j,l}\right), \tag{4.27}$$

where ICT stands for the inverse curvelet transform operation. The last step is to crop the de-noised HSI cube to its original size. The pseudo code of the proposed curvelet and SSA approach is summarised in Algorithm 1.

4.4 Data sets and experimental setup

In this section, the experimental data sets will be introduced as well as the choice of the classification model and parameters.

Three publicly available and widely used HSI remote sensing data sets are employed to evaluate experimental performance in this chapter, including the airborner visible/infrared imaging spectrometer (AVIRIS) natural scenes Indian Pines [10] and Salinas Valley [140], as well as the reflective optics system imaging spectrometer (ROSIS) urban scene Pavia University [141].

Details of all data sets are explained in Appendix A. For Indian Pines, the reduced 9 classes ground truth map is used. Additionally, due to atmospheric water absorption, bands 104–108, 150–163 and 220 do not contain useful information and are consequently removed to prevent from decreasing the classification accuracy, resulting in a reduced data set with 200 spectral bands. Similarly, for Salinas Valley, water absorption bands ([108–112], [154–167] and 224) are removed as well, leaving 204 spectral bands. Pavia University is kept the same as described in Appendix A.

The spatial area of all data sets is not in a square, hence, before applying the curvelet transform, Indian Pines, Salinas Valley and Pavia University are zero-padded to the size of $256 \times 256 \times 200$, $512 \times 512 \times 204$ and $1024 \times 1024 \times 103$ respectively. Then, when the de-noising and feature extraction process is finished, reconstructed images are cropped to their original sizes for the following performance evaluation.

In the context of supervised classification, a variety of methods have been developed for HSI data classification problems, including the famous artificial neural networks [79–81], multinomial logistic regression [172], as well as the widely used support vector machine (SVM) which shows its outstanding performance in many papers [5, 18, 52, 86]. The properties of SVM make it an effective tool for HSI data classification problems which are influenced by the Hughes phenomenon. In this chapter, a publicly available SVM library called LIBSVM [126] is adopted and the Gaussian radial basis function (RBF) kernel is chosen as it outperforms other kernel functions for HSI data in terms of classification accuracy [18]. Optimal values of the regularisation parameter C and the width

Algorithm 1 The Curvelet and SSA algorithm

procedure CURVELETSSA(I, L, EVG) ▷ $P \times Q \times B$ HSI data set I, SSA window size L and grouping parameter EVG

  if $P \neq Q$ and P, Q are not a power of two **then**
    zero-pad each band of I to the next larger power
    of two to get a new data cube I_pad with size of
    $N \times N \times B$
  else
    $I\_pad \leftarrow I$
  end if

  for $b \leftarrow 1, B$ **do**
    apply curvelet transform to I_pad_b to get the
    coarse image $C_{b,J}$ and detail images $W_{b,j,l}$
  end for

  for all detail image stacks **do**
    for $x \leftarrow 1, X$ **do**
      for $y \leftarrow 1, Y$ **do**
        $\mathbf{W}_{j,l}(x,y) \leftarrow ((W_{1,j,l}(x,y), \ldots, W_{B,j,l}(x,y))$
        $\mathbf{Y}_{j,l}(x,y) \leftarrow \mathrm{SSA}(\mathbf{W}_{j,l}(x,y), L, EVG)$
      end for
    end for
  end for

  substitute original detail curvelet coefficients with
  SSA-processed detail coefficients

  for $b \leftarrow 1, B$ **do**
    apply inverse curvelet transform to $C_{b,J}$ and
    $\mathbf{Y}_{b,j,l}$ to get the denoised band image I_d_b
  end for

  crop I_d_b to the original size $P \times Q \times B$
end procedure

parameter γ of the Gaussian RBF kernel are tuned with ten-fold cross-validation of the exponentially increasing sequence $\{2^{-10}, 2^{-9}, \ldots, 2^{10}\}$ on training samples for both parameters.

4.5 Results and discussion

In this section, the classification results based on SVM for the three de-noised hyperspectral data sets discussed in the previous section are revealed. The classification performance is evaluated by the overall accuracy (OA) and the average accuracy (AA), where OA refers to the percentage of all pixels that are correctly labelled and AA stands for the average percentage of correctly labelled pixels for each class. Considering for C classes the number of correctly classified pixels for each class is N_i and the total number of pixels for each class is T_i, the OA and AA can be mathematically defined as Equation (4.28) and Equation (4.29),

$$OA = \frac{\sum_{i=1}^{C} N_i}{\sum_{i=1}^{C} T_i}, \tag{4.28}$$

$$AA = \frac{1}{C} \sum_{i=1}^{C} \frac{N_i}{T_i}. \tag{4.29}$$

Since the proposed methodology is inspired by the feature extraction approach simply using SSA [64] and the de-noised approach in [157] which uses MLR applied in the curvelet domain, it is compared with them in subsection 4.5.1. Besides, the curvelet transform is similar to the wavelet transform as a multi-scale geometric analysis (MGA). Therefore, SSA and MLR are also applied in the wavelet domain for comparison in the same subsection. For the wavelet transform-based approaches, the hyperspectral data are also zero-padded and the number of decomposition scales is set the same as the curvelet transform, where the Cohen-Daubechies-Feauveau (CDF) 9/7 wavelet is adopted. For convenience, the proposed approach is denoted as CT-SSA and other comparison approaches are named as SSA, CT-MLR (MLR applied in the curvelet domain), WT-MLR (MLR applied in the wavelet domain) and WT-SSA (SSA applied in the wavelet domain). The computational cost required for these methods is also discussed in this part.

Then, the CT-SSA method as well as its extended version CT-SSA-PCA, are further compared with some state-of-the-art spectral feature extraction techniques in subsection 4.5.2, including PCA, LDA, NMF and ensmeble EMD (EEMD) [69], to show the efficacy of the CT-SSA method.

The proposed approach only takes into account the spectral information of the hyperspectral image, while ignoring the important spatial information. For this reason, in subsection 4.5.3, a simple yet powerful spatial post-processing technique is applied to the classified ground truth map and it is compared with several recent spectral-spatial classification methods listed in [173].

4.5.1 Comparison with inspirational approaches

The SSA-based feature extraction approach mentioned in [64] has two parameters, which are the window size L and the grouping parameter EVG. By trial and error, optimal settings for L and EVG are set as 5 and the first component for both Indian Pines and Salinas Valley. For Pavia University, L is set as 4 and EVG is still set as the first component. Same settings of SSA are used for CT-SSA and WT-SSA as well.

For the Indian Pines data set, in each class out of 9 classes, 10% of pixels are randomly chosen into the training set with the rest allocated to the test set. As Salinas Valley is a high resolution HSI data set, spectra with this scene are more separable than those with Indian Pines. Therefore, a lower percentage of samples in each class (5%) is used for training the classification model of SVM to give a more interpretable result. A recent publication proposed to use a fixed number (200) from each class in the Pavia University data set as training samples [174], which account for 4% of the whole labelled pixels. Therefore, 4% of pixels are also randomly chosen from Pavia University for training SVM in the following experiments. As shown in Figure 4.3 – 4.5, compared with those inspirational methods, the CT-SSA method always presents better de-noising and feature extraction performance in terms of highest overall classification accuracies.

Additionally, the capability of the proposed CT-SSA method in removing noise is compared in particular. One noisy band is taken from each data set and the same band after being processed by CT-SSA is compared visually in Figure 4.6 – 4.8. It can be observed that the CT-SSA method works for both high-noise (Figure 4.6 and Figure 4.7) and low-noise (Figure 4.8) cases, where the noise in the corrupted band is effectively suppressed and local details of the original image can be kept simultaneously.

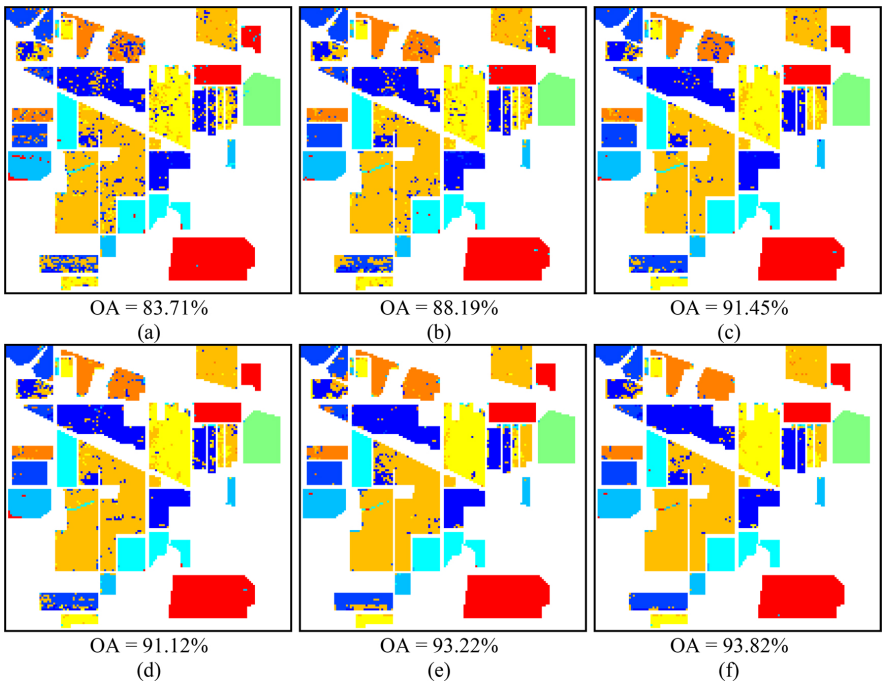


Fig. 4.3 Classification results (overall accuracy in percentage) of Indian Pines obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data.

In order to avoid errors and to get a more consistent result, ten repeated experiments are carried out based on randomly chosen training and test samples, and average classification accuracies are calculated, with numerical results shown in Table 4.1 – 4.3 for three data sets. It can be noticed that compared with the original image, OA using the CT-SSA method is increased with a percentage of 10.02%, 4.44% and 0.89% respectively for three experimental data sets, achieving an impressive improvement over other methods. For both Indian Pines and Salinas Valley, in most cases, the proposed method presents the highest classification accuracy, especially in class 15 of Salinas Valley, the accuracy is improved by over 20% compared with the raw image. Although it was found that the classification of hyperspectral urban data is a challenge problem without combining spatial and spectral information together [175], both AA and OA are slightly improved for Pavia University using the proposed spectral processing method CT-SSA. In comparison with CT-MLR, WT-MLR yields slightly worse results in both AA and OA. Same situations occur for WT-SSA and CT-SSA, where the latter presents slightly higher AA and OA than the former. It proves that the curvelet transform does have advantages over the wavelet transform for extracting geometric details in some ways, but the performance of the curvelet transform is largely dependent on the data set. Despite that CT-SSA is inspired by CT-MLR, SSA is applied to

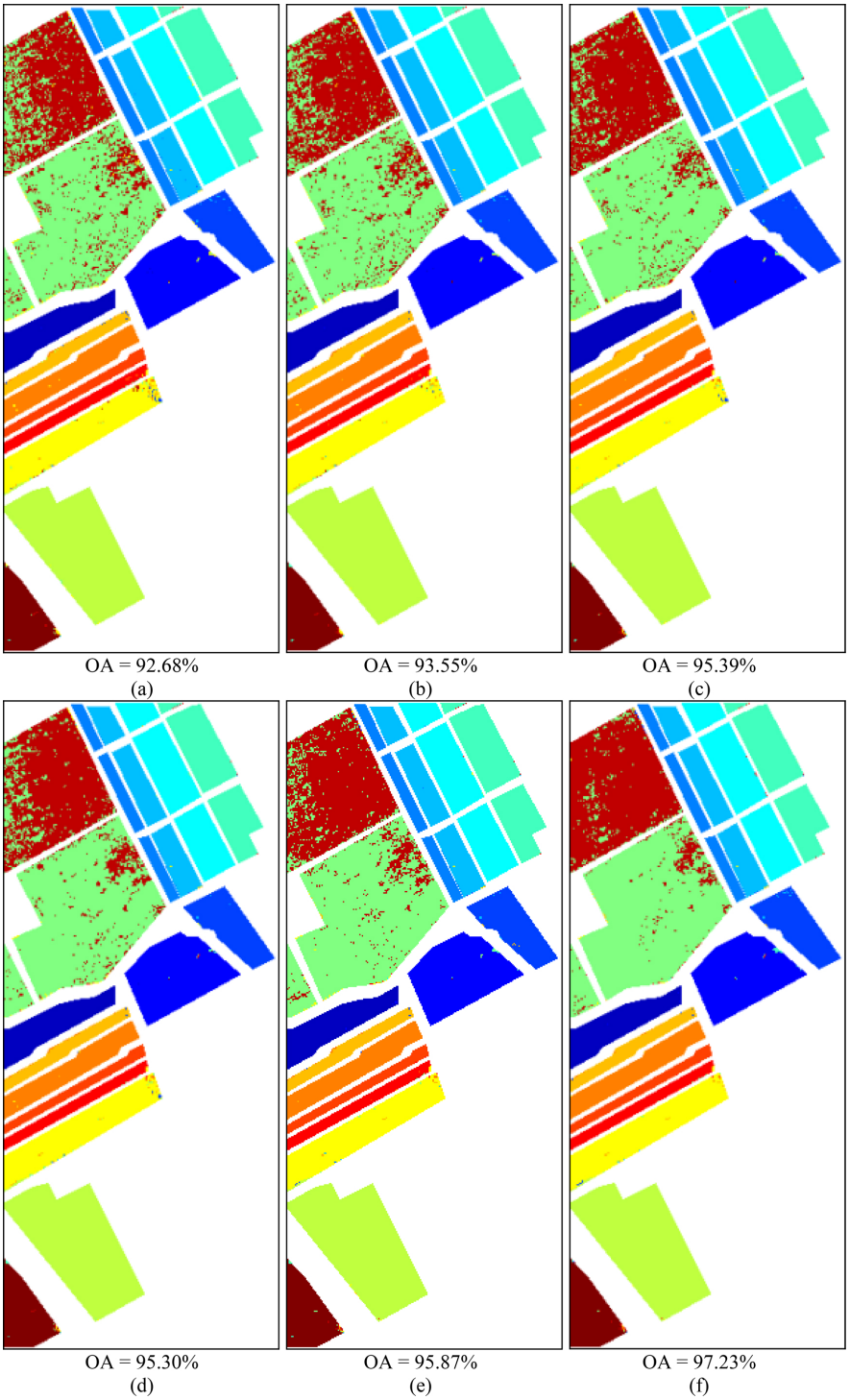


Fig. 4.4 Classification results (overall accuracy in percentage) of Salinas Valley obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data.

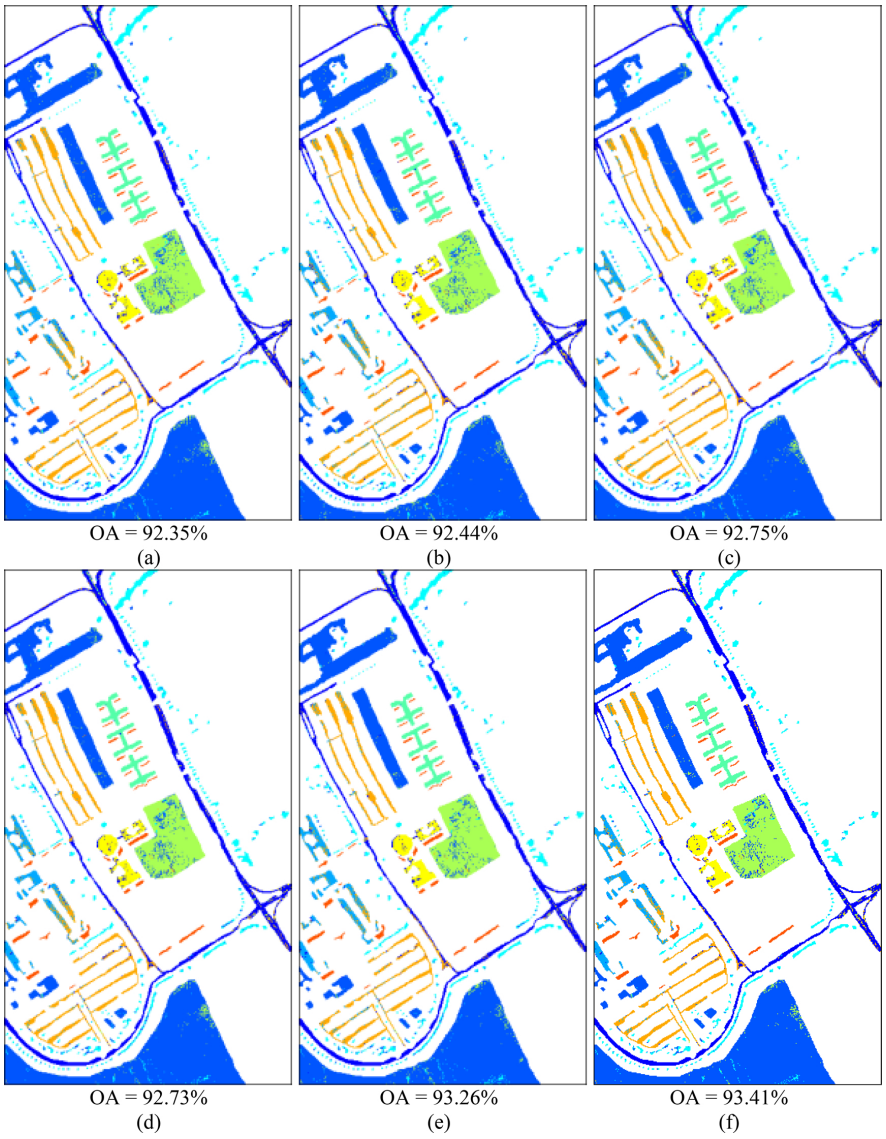
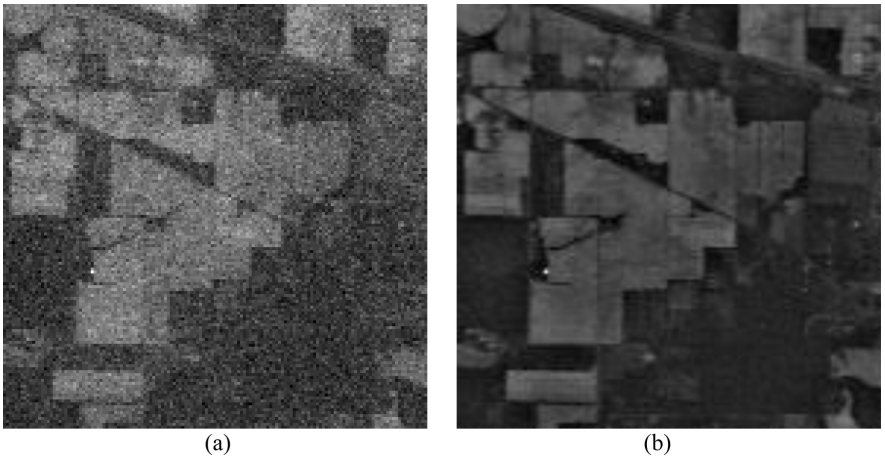
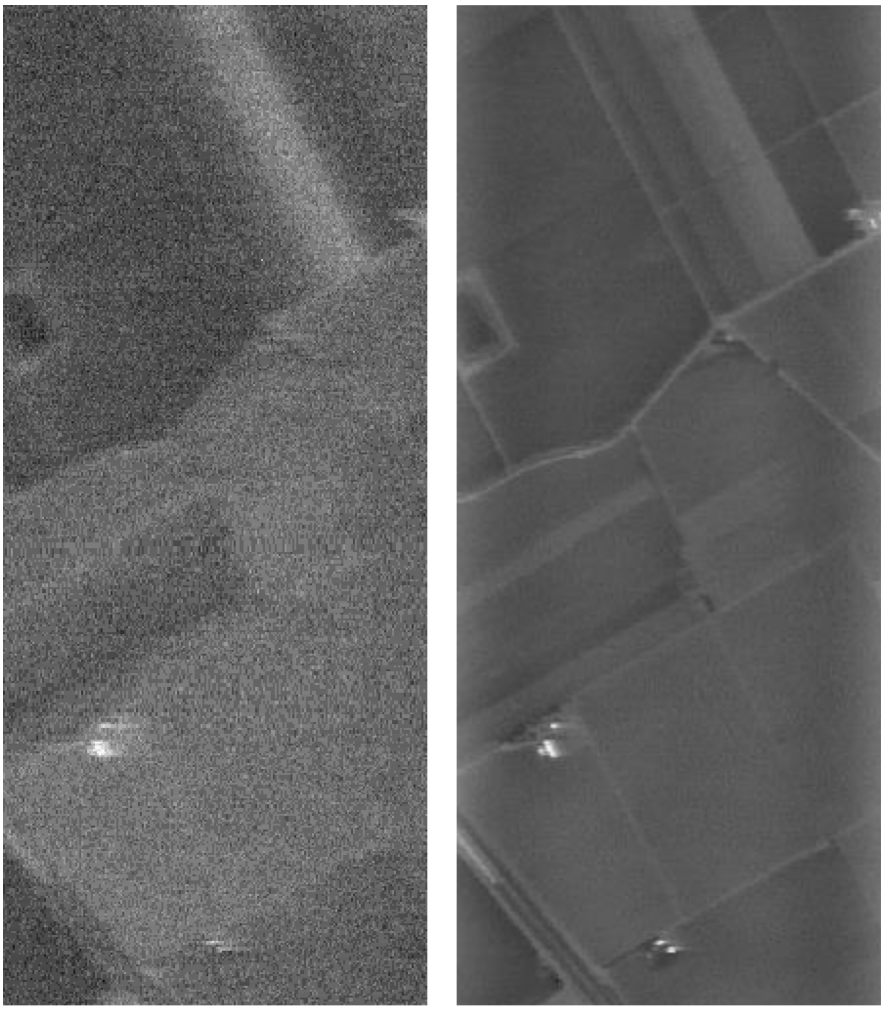


Fig. 4.5 Classification results (overall accuracy in percentage) of Pavia University obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data.



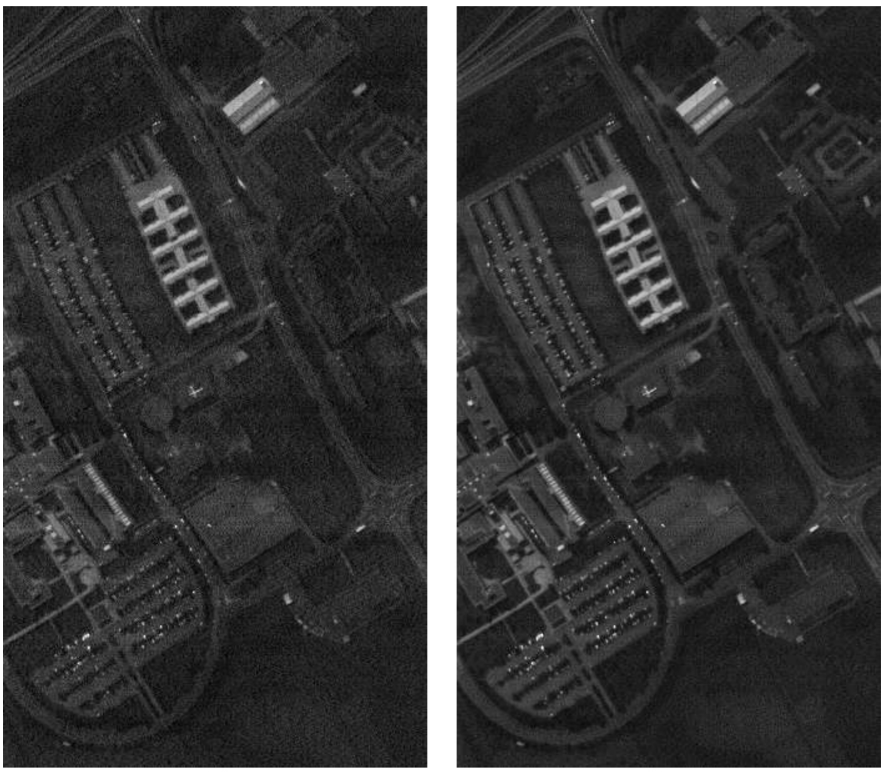
(a) (b)

Fig. 4.6 Band 199 of Indian Pines in (a) original data and (b) CT-SSA processed data.



(a) (b)

Fig. 4.7 Band 2 of Salinas Valley in (a) original data and (b) CT-SSA processed data.



(a) (b)

Fig. 4.8 Band 1 of Pavia University in (a) original data and (b) CT-SSA processed data.

detail curvelet coefficients pixel-wisely while MLR is applied to each detail image stack all at once, which is the reason why CT-SSA could extract more effective and detailed features for improving the classification accuracy.

For an $N \times N$ block, the computational complexity for the curvelet transform is given as $O\left(N^2 \log N\right)$ [164], while the computation of the wavelet transform requires $O\left(\frac{3}{2}\left(m_H + m_L\right)\left(1 - \frac{1}{4^J}\right)N^2\right)$ flops for a block with the same size, where J is the decomposition level and m_H, m_L are the lengths of the filter [176]. With the CDF 9/7 filter used in the experiments, the lengths of the filter are both 9, i.e. $m_H = m_L = 9$. Therefore, when J is higher than 2, the computational complexity for the wavelet transform can be approximated as $O\left(27N^2\right)$, which is much higher than that of the curvelet transform. Besides, it should be noted that the processing time for MLR and SSA is largely dependent on the number of detail coefficient stacks and the number of total pixels in those stacks. However, it is not a problem for SSA, as fast implementation of SSA [65] has been proposed without degrading the feature extraction performance. For the Indian Pines data set, the computing time of the proposed CT-SSA method (using the original SSA implementation) is about 5 min on a personal computer with an Intel Core i5-2400 CPU at 3.10 GHz using Matlab 2014a (Mathworks). Hence, it is reasonable to

Table 4.1 Mean class-by-class, average and overall accuracies (%) of ten repeated experiments on test samples of the original Indian Pines data set and SSA, CT-MLR, WT-MLR, WT-SSA and CT-SSA processed data sets with 10% of data used for training, followed by the standard deviation.

Class	Method					
	Original	**SSA**	**CT-MLR**	**WT-MLR**	**WT-SSA**	**CT-SSA**
2	75.44 ± 2.09	80.02 ± 2.38	86.76 ± 1.35	86.15 ± 2.20	87.61 ± 1.97	**89.58 ± 3.15**
3	69.92 ± 5.11	77.01 ± 5.34	88.26 ± 3.16	89.64 ± 2.84	**93.19 ± 1.42**	92.38 ± 2.43
5	90.46 ± 3.41	91.31 ± 1.34	93.33 ± 2.10	92.60 ± 2.08	93.17 ± 2.06	**94.76 ± 1.94**
6	98.11 ± 0.86	97.79 ± 0.67	98.71 ± 0.95	98.54 ± 0.88	98.89 ± 0.95	**98.95 ± 0.45**
8	99.26 ± 0.42	99.63 ± 0.23	99.61 ± 0.35	99.77 ± 0.36	99.54 ± 0.48	**99.84 ± 0.11**
10	72.35 ± 2.59	80.87 ± 4.74	86.22 ± 1.52	85.49 ± 2.56	**90.20 ± 2.33**	89.45 ± 2.29
11	85.10 ± 2.32	87.52 ± 1.73	92.01 ± 1.52	91.65 ± 1.65	**94.59 ± 2.02**	93.96 ± 0.95
12	75.21 ± 3.96	82.42 ± 5.23	85.99 ± 2.16	87.83 ± 2.80	**91.92 ± 2.00**	90.83 ± 2.73
14	98.27 ± 0.52	98.37 ± 0.62	99.17 ± 0.35	99.07 ± 0.57	98.93 ± 1.04	**99.54 ± 0.26**
AA	84.90 ± 1.18	88.35 ± 1.01	92.34 ± 0.41	92.30 ± 0.55	94.23 ± 0.29	**94.48 ± 0.32**
OA	84.11 ± 0.87	87.54 ± 0.83	91.93 ± 0.50	91.76 ± 0.44	93.87 ± 0.38	**94.13 ± 0.28**

Table 4.2 Mean class-by-class, average and overall accuracies (%) of ten repeated experiments on test samples of the original Salinas Valley data set and SSA, CT-MLR, WT-MLR, WT-SSA and CT-SSA processed data sets with 5% of data used for training, followed by the standard deviation.

Class	Method					
	Original	**SSA**	**CT-MLR**	**WT-MLR**	**WT-SSA**	**CT-SSA**
1	99.02 ± 0.50	99.07 ± 0.61	**99.48 ± 0.44**	99.40 ± 0.41	99.32 ± 0.37	99.35 ± 0.33
2	**99.85 ± 0.06**	99.71 ± 0.20	99.68 ± 0.20	99.72 ± 0.20	99.68 ± 0.24	99.69 ± 0.27
3	99.07 ± 0.41	99.37 ± 0.49	99.39 ± 0.33	**99.67 ± 0.08**	99.65 ± 0.19	99.59 ± 0.22
4	99.21 ± 0.34	**99.44 ± 0 .34**	99.32 ± 0.28	99.30 ± 0.41	99.09 ± 0.69	99.28 ± 0.33
5	98.79 ± 0.42	98.52 ± 0.75	**99.07 ± 0.66**	98.84 ± 0.60	98.98 ± 0.55	99.04 ± 0.32
6	99.80 ± 0.08	99.74 ± 0.17	99.80 ± 0.09	**99.98 ± 0.14**	99.83 ± 0.16	99.81 ± 0.08
7	99.62 ± 0.18	99.48 ± 0.23	99.52 ± 0.17	99.51 ± 0.23	99.59 ± 0.18	**99.62 ± 0.15**
8	88.05 ± 1.47	87.90 ± 1.20	91.28 ± 0.82	90.62 ± 0.77	92.26 ± 0.74	**94.60 ± 0.75**
9	99.65 ± 0.22	99.72 ± 0.14	99.60 ± 0.42	99.54 ± 0.21	99.86 ± 0.09	**99.88 ± 0.09**
10	94.98 ± 1.20	95.19 ± 0.80	95.36 ± 1.30	96.01 ± 1.11	**97.35 ± 1.17**	97.19 ± 1.02
11	97.63 ± 1.01	97.02 ± 2.10	97.84 ± 1.42	97.62 ± 2.44	97.78 ± 2.58	**98.58 ± 0.30**
12	99.53 ± 0.28	99.64 ± 0.19	99.78 ± 0.15	99.50 ± 0.80	99.79 ± 0.20	**99.92 ± 0.06**
13	98.09 ± 1.11	98.45 ± 1.40	98.43 ± 0.86	**98.84 ± 0.72**	98.52 ± 0.93	98.54 ± 0.95
14	94.56 ± 2.12	95.70 ± 1.99	**96.86 ± 1.45**	95.89 ± 2.14	94.83 ± 3.80	96.46 ± 1.62
15	71.73 ± 3.21	76.16 ± 1.85	82.70 ± 2.88	83.17 ± 1.26	87.12 ± 1.73	**92.49 ± 0.86**
16	98.34 ± 0.28	98.13 ± 0.51	**98.72 ± 0.32**	98.48 ± 1.18	98.50 ± 0.97	98.71 ± 0.26
AA	96.12 ± 0.11	96.45 ± 0.22	97.30 ± 0.28	97.25 ± 0.31	97.63 ± 0.44	**98.30 ± 0.13**
OA	92.91 ± 0.12	93.49 ± 0.25	95.18 ± 0.33	95.11 ± 0.21	96.08 ± 0.34	**97.35 ± 0.16**

Table 4.3 Mean class-by-class, average and overall accuracies (%) of ten repeated experiments on test samples of the original Pavia University data set and SSA, CT-MLR, WT-MLR, WT-SSA and CT-SSA processed data sets with 4% of data used for training, followed by the standard deviation.

Class	Method					
	Original	**SSA**	**CT-MLR**	**WT-MLR**	**WT-SSA**	**CT-SSA**
1	93.05 ± 0.69	92.50 ± 1.11	93.27 ± 0.99	92.27 ± 0.66	92.81 ± 0.66	**94.17 ± 0.91**
2	97.63 ± 0.38	97.47 ± 0.54	97.74 ± 0.54	97.65 ± 0.39	**98.10 ± 0.27**	97.92 ± 0.21
3	76.45 ± 1.65	76.21 ± 2.02	79.05 ± 2.79	77.51 ± 2.78	**80.48 ± 0.84**	79.63 ± 1.59
4	93.15 ± 1.63	92.88 ± 1.42	93.40 ± 1.41	92.69 ± 1.99	92.43 ± 2.26	**93.97 ± 1.66**
5	99.08 ± 0.35	99.03 ± 0.29	99.11 ± 0.41	**99.15 ± 0.32**	99.01 ± 0.24	99.10 ± 0.31
6	86.19 ± 1.91	85.20 ± 1.56	86.99 ± 1.02	86.46 ± 1.53	88.10 ± 1.16	**88.11 ± 1.67**
7	83.07 ± 3.70	85.45 ± 2.14	83.55 ± 2.09	85.47 ± 2.99	**86.59 ± 2.94**	85.62 ± 1.72
8	87.43 ± 2.52	87.75 ± 1.44	87.66 ± 2.40	88.04 ± 1.97	88.02 ± 0.58	**88.08 ± 1.90**
9	99.84 ± 0.12	99.77 ± 0.23	99.85 ± 0.16	**99.85 ± 0.13**	99.76 ± 0.12	99.84 ± 0.11
AA	90.59 ± 0.49	90.69 ± 0.41	91.18 ± 0.36	91.01 ± 0.54	91.70 ± 0.58	**91.82 ± 0.32**
OA	92.96 ± 0.32	92.77 ± 0.37	93.34 ± 0.20	93.05 ± 0.30	93.67 ± 0.21	**93.85 ± 0.28**

believe that with fully optimised CurveLab toolbox and the fast implementation of SSA, the time cost of the proposed method should be much lower.

4.5.2 Comparison with other state-of-the-art spectral feature extraction techniques

Although the curvelet transform is applied on the spatial domain of the hyperspectral image, the dominating part of the proposed de-noising and feature extraction method, SSA, is applied on the spectral dimension. As a matter of fact, the proposed CT-SSA approach actually belongs to the 1D spectral processing technique for hyperspectral images. Therefore, a few state-of-the-art 1D spectral feature extraction and dimensionality reduction techniques, including PCA, LDA, NMF, are compared with the proposed approach for further performance assessment. The NMF algorithm is realised using the NMF Matlab toolbox [177]. Additionally, as an upgraded version of EMD, EEMD is also included for comparison as it overcomes the drawback of EMD and therefore it outperforms EMD [69]. EEMD is achieved by sifting an ensemble of white noise-added signal and it generates a series of intrinsic mode functions (IMFs), where the sum of them represents the processed signal. For this part of the experiment, the fast EEMD Matlab toolbox [66] is adopted. The performance of PCA on the proposed approach for further feature extraction is also included for comparison.

The number of resulting features using PCA on both original data and CT-SSA processed data is tested within 5 – 50 in a step of 5 [71], while the dimensions of extracted features are tested from 2 to 10 for LDA and NMF. Best results with the highest classification accuracy are chosen for comparison. With regard to EEMD, it is suggested the input white noise level ε_0 should be in the range of 0.1 – 0.4 and the number of ensemble N_E should be the order of 100 [68]. Therefore, same parameters are kept as [69], where $\varepsilon_0 = 0.2$, $N_E = 200$ and the number of IMFs is 7. Experiments are repeated ten times as well and comparison results are given in Table 4.4.

It can be seen that those dimensionality reduction techniques could only give similar results compared with original data sets, even though they could speed up the classification process. As given in [64], with 10% training rate of Indian Pines, the highest classification OA achieved by EMD was 75.49%. Although EEMD has made an improvement of over 8% on this data set, it is still incomparable to the proposed CT-SSA approach. For Indian Pines and Salinas Valley, adding an extra PCA step leads to further improvement in the classification accuracy while reducing the classification time with less number of features. Even though

Table 4.4 Mean overall accuracy (%) and the corresponding standard deviation of ten repeated experiments on original data sets, SSA, CT-MLR, WT-MLR, WT-SSA, CT-SSA, CT-SSA-PCA and some state-of-the-art spectral feature extraction techniques processed data sets, with dimensionality of features shown in parentheses.

Method	Data set		
	Indian Pines (Training = 10%)	**Salinas Valley (Training = 5%)**	**Pavia University (Training = 4%)**
Original	84.11 ± 0.87 (200)	92.91 ± 0.12 (204)	92.96 ± 0.32 (103)
SSA	87.54 ± 0.83 (200)	93.49 ± 0.25 (204)	92.77 ± 0.37 (103)
CT-MLR	91.93 ± 0.50 (200)	95.18 ± 0.33 (204)	93.34 ± 0.20 (103)
WT-MLR	91.76 ± 0.44 (200)	95.11 ± 0.21 (204)	93.05 ± 0.30 (103)
WT-SSA	93.87 ± 0.38 (200)	96.08 ± 0.34 (204)	93.67 ± 0.21 (103)
CT-SSA	94.13 ± 0.28 (200)	97.35 ± 0.16 (204)	**93.85 ± 0.28 (103)**
CT-SSA-PCA	**95.75 ± 0.46 (50)**	**98.33 ± 0.15 (50)**	91.78 ± 0.40 (10)
PCA [53]	82.65 ± 0.53 (50)	92.78 ± 0.17 (40)	92.93 ± 0.39 (50)
LDA [178]	87.03 ± 0.43 (9)	91.62 ± 0.19 (10)	92.06 ± 0.53 (10)
NMF [177]	77.43 ± 1.15 (10)	92.64 ± 0.31 (10)	91.71 ± 0.25 (10)
EEMD [66]	83.78 ± 0.84 (200)	93.20 ± 0.09 (204)	93.22 ± 0.20 (103)

the extra step of PCA has limited performance on the Pavia University data set, the proposed CT-SSA method still outperforms other spectral feature extraction techniques.

4.5.3 Comparison with recent spectral-spatial classification methods

For the majority of supervised classifiers, the spectral feature-based classification chain is adopted. However, they are not able to incorporate spatial dependencies presented in the original scene into the classification process [179]. Several recent publications have demonstrated that the integration of spectral and spatial information could be beneficial to hyperspectral image classification problems, where some spectral-spatial classification methods, including the extended morphological profile (EMP) [180], the logistic regression and multi-level logistic (LMLL) [181], the edge-preserving filter (EPF) [182], image fusion and recursive filter (IFRF) [183] and the intrinsic image decomposition (IID) [173], have been proven to be superior to the spectral feature-based classification method. Therefore, a post-processing step is added to the proposed CT-SSA approach for increasing the spatial consistency in the classification result, where a $T \times T$ spatial window is applied around each central pixel in the classification map and the final classified label is decided in favour of the class which appears most in the window. The window shape could be designed conforming to different scenes, but to make it simple, the square window is used in this chapter.

As can be noticed in ground truth maps in Figure A.2 – A.4, Indian Pines has a smaller spatial area for some classes compared with Salinas Valley and Pavia University. Therefore, the spatial post-processing (SPP) window size T is chosen as 5 for Indian Pines, 9 for Salinas Valley and 9 for Pavia University, respectively. Results using EMP, LMLL, EPF, IFRF and IID are cited directly from [173]. Since only 9 classes in Indian Pines are involved in the experiment of the proposed method, the rest 7 classes from their results are eliminated accordingly and AA and OA are recalculated as well. For Salinas Valley, they used 2% in each class for training. Therefore, the experiment under the same training rate has been re-performed with other parameters unchanged. Numerical results are shown in Table 4.5 – 4.7, where the proposed method with spatial post-processing (CT-SSA-SPP) gives relatively higher results than other spectral-spatial classification methods in most cases.

Table 4.5 Comparison of class-by-class, average and overall accuracies (%) of the CT-SSA-SPP method and recent spectral-spatial classification methods, with 10% training rate on the Indian Pines data set.

	Method					
Class	**EMP [180]**	**LMLL [181]**	**EPF [182]**	**IFRF [183]**	**IID [173]**	**CT-SSA-SPP (T = 5)**
2	86.63	90.64	95.03	95.41	**96.90**	96.18
3	89.74	87.23	95.68	96.07	97.20	**99.63**
5	93.23	96.39	**98.39**	98.07	97.88	97.08
6	97.50	99.90	98.90	98.62	99.40	**100**
8	**100**	**100**	99.49	**100**	**100**	**100**
10	86.18	93.49	82.09	95.26	**95.49**	95.45
11	93.82	89.75	95.93	98.42	98.03	**99.38**
12	87.09	**98.60**	92.81	96.72	97.23	97.32
14	99.59	92.45	99.11	99.75	99.76	**99.96**
AA	92.64	94.27	95.27	97.59	97.99	**98.33**
OA	92.45	92.57	95.07	97.56	97.89	**98.40**

4.6 Summary

In this chapter, two powerful tools, SSA and the curvelet transform, are combined for hyperspectral image de-noising and feature extraction. By applying SSA in the curvelet domain, noise can be smoothed from the decomposed signals, resulting in more effective feature extraction. Inspired by applying MLR in the curvelet domain and another method of solely applying SSA on the hyperspectral image, the proposed method takes advantages of those approaches. Using SVM on three publicly available data sets for classification performance assessment, i.e.

Table 4.6 Comparison of class-by-class, average and overall accuracies (%) of the CT-SSA-SPP method and recent spectral-spatial classification methods, with 2% training rate on the Salinas Valley data set.

Class	Method					
	EMP [180]	**LMLL [181]**	**EPF [182]**	**IFRF [183]**	**IID [173]**	**CT-SSA-SPP (T = 9)**
1	99.93	99.83	**100**	99.99	**100**	**100**
2	99.60	99.96	**100**	**100**	99.97	**100**
3	96.75	99.80	98.36	99.89	99.85	**100**
4	98.25	99.22	97.17	97.32	97.80	**99.93**
5	99.14	98.75	**99.96**	99.94	**99.96**	99.73
6	99.95	99.66	99.98	**100**	99.98	**100**
7	99.87	99.55	99.98	99.70	99.90	**100**
8	94.79	85.97	91.78	**99.34**	99.07	98.15
9	99.41	99.69	99.65	**99.99**	99.86	99.92
10	95.10	94.68	93.88	**99.81**	98.80	99.69
11	96.39	98.47	98.80	99.26	99.98	**100**
12	98.79	**100**	99.99	99.90	99.99	99.99
13	98.87	98.48	**100**	99.28	**100**	**100**
14	97.06	98.60	**99.19**	97.96	97.51	97.15
15	86.67	78.78	85.50	97.07	**99.52**	99.04
16	97.31	99.35	**99.99**	99.95	97.92	**99.99**
AA	97.37	96.92	97.76	99.34	99.38	**99.60**
OA	96.11	93.55	95.59	99.27	**99.45**	99.39

Table 4.7 Comparison of class-by-class, average and overall accuracies (%) of the CT-SSA-SPP method and recent spectral-spatial classification methods, with 4% training rate on the Pavia University data set.

Class	Method					
	EMP [180]	**LMLL [181]**	**EPF [182]**	**IFRF [183]**	**IID [173]**	**CT-SSA-SPP (T = 9)**
1	98.61	94.36	98.07	97.25	99.62	**100**
2	98.84	97.79	98.38	99.68	99.88	**100**
3	95.09	88.10	98.39	95.53	**98.92**	98.57
4	96.64	98.05	**98.45**	96.52	97.54	95.91
5	98.00	99.83	99.51	99.90	99.89	**100**
6	88.33	99.74	95.35	98.72	99.91	**99.97**
7	93.51	98.14	97.57	96.69	99.59	**100**
8	97.48	93.77	93.72	94.40	**98.66**	97.29
9	99.91	**99.92**	98.23	92.54	94.84	90.37
AA	96.27	96.63	97.63	96.80	**98.76**	98.01
OA	96.82	96.81	97.98	98.07	**99.41**	99.19

Indian Pines, Salinas Valley and Pavia University, the proposed CT-SSA method significantly improves OA by as much as 10% compared with the result of the original Indian Pines data set. In addition, noise could be effectively removed by checking the band image visually. Compared with the inspirational approaches and some state-of-the-art spectral feature extraction techniques, the proposed method always stands out with the highest classification accuracy. Furthermore, it is noticed that with the spatial post-processing step, another 4% improvement could be added to the OA of Indian Pines, while for Salinas Valley and Pavia University, classification results can be even improved to nearly 100%. It is worth noting that by adding the simple spatial post-processing step, the performance of the proposed de-noising and feature extraction method in terms of the classification accuracy is either comparable with or even higher than some recent spectral-spatial classification methods.

Chapter 5

Hyperspectral Image Classification Using Sparse Representation of Features Extracted via Bilateral Filtering

5.1 Introduction

As discussed in Chapter 4, hyperspectral image classification has been a relatively active topic in remote sensing. In order to achieve high classification accuracy, sufficient training samples are needed to cope with the curse of dimensionality, i.e., the Hughes phenomenon [18]. However, sometimes this requirement is infeasible for classes with a limited number of samples. Besides, due to the noise present in the hyperspectral image, the same material may have different spectral profiles, whereas different categories may share similar spectral signatures [16]. The aforementioned problems can significantly influence the classification result. Much effort has been put into overcoming these problems by extracting more effective features from spatial/spectral domain and constructing suitable classification models for raw/extracted features as indicated in Chapter 2.

In recent years, sparse representation has received considerable attention in fields of signal/image processing, computer vision and pattern recognition. As first proposed in [184] for face recognition, the test image can be sufficiently represented by samples from the same category via sparse representation. Sparse representation is an indispensable prerequisite of the compressed sensing (CS) framework [185–187], which demonstrates that most natural signals can be precisely reconstructed by only a few measured values. Thus, the characteristics of sparse representation make it an outstanding tool for image classification.

In [188], sparse representation classification (SRC) was applied for hyperspectral image classification by Chen *et al.* and they have proposed a joint sparse

representation classification (JSRC) method by assuming that all pixels around the central test pixel in the neighbourhood window shares a common sparsity pattern. Therefore, the hyperspectral image classification problem can be solved by a joint sparsity model (JSM) [189, 190] and the classification accuracy is greatly enhanced compared with the pixel-wise SRC method since the spatial information is incorporated as well. However, in the JSRC method, all pixels around the central test pixel in the neighbourhood window are considered equally when constructing the JSM, which is inappropriate for an area composed of more than one class. To address this problem, Zhang *et al.* [191] proposed a nonlocal weighted joint sparse representation classification (NLW-JSRC) method to determine the weighted contribution of neighbouring pixels. In the NLW-JSRC method, the weight of each neighbouring pixel is calculated by comparing the structural similarity between an image patch around that pixel and the patch around the central test pixel. In this way, both the spectral information between two pixels and the geometrical details between two nonlocal image patches are compared, leading to improved classification accuracy. Besides, sparse representation has also been applied successfully for target detection in hyperspectral images, where an adaptive spatial support was proposed for improved performance in [3].

Another family of frameworks to solve those classification difficulties is based on the post-processing of preliminary classification results. In both [182] and [192], authors proposed to use the edge-preserving filter for smoothing the pixel-wise classification map under the smoothness assumption. Results showed that the filtering-based post-processing could improve the classification accuracy significantly.

Edge-preserving filtering is a popular topic in the area of image processing. As a representative of the edge-preserving filter family, bilateral filter has been successfully applied for hyperspectral image analysis, such as hyperspectral image visualisation [193]. In [194], a vector bilateral filtering-based approach was proposed for hyperspectral image enhancement. However, its application on hyperspectral image feature extraction is rare. In this chapter, a spectral-spatial classification framework based on bilateral filter and sparse representation is proposed. It is believed that only with the contextual information is insufficient in the JSRC method since there is no spatial/spectral feature extraction prior to the classification. Therefore, in the proposed method, the spatial feature is extracted using a joint bilateral filter with the first principal component as the guidance image. With the first principal component as the guidance image, the quality of extracted features can be effectively enhanced in the low-contrast spectral bands.

For the classification, although it is claimed that NLW-JSRC could achieve higher classification accuracies than JSRC [191], the choice of the Gaussian kernel and the computational complexity could be a problem. Hence, a spectral similarity-based JSRC (SS-JSRC) method, which utilises the statistics within a neighbourhood window to exclude irrelevant pixels, is proposed as the improved classifier.

The layout of this chapter is as follows. The background principles of bilateral filter and SRC are discussed in Section 5.2. Section 5.3 explains the proposed feature extraction and classification framework, followed by the experimental data sets and setup in Section 5.4. Classification results on two hyperspectral images are given in Section 5.5, and conclusion is drawn in Section 5.6.

5.2 Background principles

Bilateral filtering and SRC are two basic concepts used in the proposed classification framework. In this section, background principles are briefly introduced as below.

5.2.1 Bilateral filtering

The bilateral filtering proposed by Tomasi and Manduchi [195] in 1998 is a non-linear filtering technique that combines domain filtering and range filtering techniques together to smooth images while preserving edges. This filtering technique replaces the pixel value at one location with a weighted sum of nearby pixel intensities based on their geometric closeness and their photometric similarity. Given an input image $f(x)$, using a continuous representation notation as in [195], the bilateral-filtered image $h(x)$ is achieved mathematically as follows:

$$h(x) = \frac{1}{k(x)} \int_{-\infty}^{\infty} f(\xi)\, c(\xi, x)\, s(f(\xi), f(x))\, d\xi, \tag{5.1}$$

where $c(\xi, x)$ is a function that measures the geometric closeness (in the domain) between the neighbourhood centre x and a neighbouring pixel ξ. The function $s(f(\xi), f(x))$ calculates the photometric similarity (in the range) between x and ξ and the normalisation factor is given by:

$$k(x) = \int_{-\infty}^{\infty} c(\xi, x)\, s(f(\xi), f(x))\, d\xi. \tag{5.2}$$

To make it simple, both the closeness function $c(\xi, x)$ and the similarity function $s(f(\xi), f(x))$ in a bilateral filter can be represented by Gaussian functions

[195]. For image processing, it is more common to use discrete representations. Thus, the discrete version of bilateral filtering for an input image I at the location (i, j) can be expressed as [193]:

$$I_{BF}(i,j) = \frac{1}{k(i,j)} \sum_{(p,q)\in w} G_{\sigma_d}(i-p, j-q)\, G_{\sigma_r}(I(i,j) - I(p,q))\, I(p,q), \quad (5.3)$$

with the normalisation factor

$$k(i,j) = \sum_{(p,q)\in w} G_{\sigma_d}(i-p, j-q)\, G_{\sigma_r}(I(i,j) - I(p,q)), \quad (5.4)$$

where (p, q) represents the location of pixels in the $(2\sigma_d + 1) \times (2\sigma_d + 1)$ local window w around pixel at (i, j). The kernels for domain and range filtering are defined as $G_{\sigma_d}(i-p, j-q) = \exp\left(-\left((i-p)^2 + (j-q)^2\right)/2{\sigma_d}^2\right)$ and $G_{\sigma_r}(I(i,j) - I(p,q)) = \exp\left(-(I(i,j) - I(p,q))^2/2{\sigma_r}^2\right)$. The value of σ_d determines the neighbourhood window under consideration and the value of σ_r defines the weight of contribution from neighbouring pixels [195].

5.2.2 Sparse representation classification

The task of supervised classification is to determine the correct category of test samples based on a set of labelled training samples from M distinct classes [184, 188]. Assume there are N_m training samples from the mth class in a B-dimensional hyperspectral data cube, denoted as $\mathbf{A}^m = \left[\begin{array}{cccc} \mathbf{a}_1^m & \mathbf{a}_2^m & \cdots & \mathbf{a}_{N_m}^m \end{array}\right] \in \mathbb{R}^{B\times N_m}$, then any test sample $\mathbf{x} \in \mathbb{R}^B$ from the same class could approximately be represented by a linear combination of the training samples associated with class m:

$$\mathbf{x} = \alpha_1^m \mathbf{a}_1^m + \alpha_2^m \mathbf{a}_2^m + \cdots + \alpha_{N_m}^m \mathbf{a}_{N_m}^m = \mathbf{A}^m \boldsymbol{\alpha}^m, \quad (5.5)$$

where $\boldsymbol{\alpha}^m$ is an N_m-dimensional coefficient vector with only a few non-zero entries in the sparsity model. Since normally the label for a test sample is unknown, a new matrix $\mathbf{A}$ should be defined by concatenating all training samples from all classes:

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{A}^1 & \mathbf{A}^2 & \cdots & \mathbf{A}^M \end{array}\right] \in \mathbb{R}^{B\times N}, \quad (5.6)$$

where $N = \sum_{m=1}^{M} N_m$. Then, the test sample $\mathbf{x}$ can be approximated as:

$$\mathbf{x} = \mathbf{A}^1\boldsymbol{\alpha}^1 + \mathbf{A}^2\boldsymbol{\alpha}^2 + \cdots + \mathbf{A}^M\boldsymbol{\alpha}^M = \mathbf{A}\boldsymbol{\alpha}, \tag{5.7}$$

where $\boldsymbol{\alpha}$ is an N-dimensional sparse coefficient vector with most of entries as zero.

The task of hyperspectral image SRC is to determine the sparse vector $\boldsymbol{\alpha}$ and to assign a label to the test sample according to $\boldsymbol{\alpha}$. The sparse solution of $\boldsymbol{\alpha}$ can be obtained by solving the following optimisation problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \mathbf{A}\boldsymbol{\alpha} = \mathbf{x}, \tag{5.8}$$

where $\|\cdot\|_0$ is the ℓ^0-norm representing the number of non-zero entries in a vector. The optimisation problem in Equation (5.8) can be also modified as minimising the reconstruction error with a pre-defined sparsity level K_0 [188]:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\mathbf{A}\boldsymbol{\alpha} - \mathbf{x}\|_2 \quad \text{subject to} \quad \|\boldsymbol{\alpha}\|_0 \leq K_0. \tag{5.9}$$

The problem in Equation (5.9) is a NP-hard problem that could be solved by greedy pursuit algorithms, such as orthogonal matching pursuit (OMP) [196] or subspace pursuit (SP) [197]. In this chapter, it is solved using OMP where the support set is augmented by one index for each iteration until K_0 non-zero entries are selected, and the computational complexity is in the order of $O(BNK_0)$. With the obtained sparse vector $\boldsymbol{\alpha}$, the label of $\mathbf{x}$ could be assigned to the class with lowest residue:

$$\text{Class}(\mathbf{x}) = \arg \min_{m=1,2,\ldots,M} r^m(\mathbf{x}), \tag{5.10}$$

where the residue r^m $(m = 1, 2, \ldots, M)$ is calculated for each class separately as follows:

$$r^m(\mathbf{x}) = \|\mathbf{x} - \mathbf{A}^m\hat{\boldsymbol{\alpha}}^m\|_2. \tag{5.11}$$

For hyperspectral images, the spectral profiles of neighbouring pixels in a small window are usually highly correlated. By assuming pixels within the neighbourhood window lie in a common subspace, Chen *et al.* [188] proposed the JSRC method for hyperspectral images. Consider that the investigated test sample $\mathbf{x}_1$ is located at the centre of the neighbourhood window, the sample matrix for the window with T pixels can be modelled as $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_T \end{bmatrix} \in \mathbb{R}^{B\times T}$.

Applying the joint sparsity model, the sample matrix $\mathbf{X}$ can be represented by:

$$\mathbf{X} = \begin{bmatrix} \mathbf{A}\boldsymbol{\alpha}_1 & \mathbf{A}\boldsymbol{\alpha}_2 & \cdots & \mathbf{A}\boldsymbol{\alpha}_T \end{bmatrix} = \mathbf{A} \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_T \end{bmatrix} = \mathbf{AS}, \tag{5.12}$$

where $\mathbf{S} \in \mathbb{R}^{N \times T}$ is the set of sparse coefficient vectors for all pixels within the neighbourhood window. The sparsity support for each pixel is the same, but the associated coefficients are different. Similarly, the task of JSRC is to obtain the sparse matrix $\mathbf{S}$ by solving the following optimisation problem:

$$\hat{\mathbf{S}} = \arg\min \|\mathbf{S}\|_{\text{row},0} \quad \text{subject to} \quad \mathbf{AS}{=}\mathbf{X}. \tag{5.13}$$

The symbol $\|\cdot\|_{\text{row},0}$ stands for the number of non-zero rows in the denoted matrix. In the same way as Equation (5.9), the optimisation problem in Equation (5.13) can be rewritten as:

$$\hat{\mathbf{S}} = \arg\min \|\mathbf{AS} - \mathbf{X}\|_F \quad \text{subject to} \quad \|\mathbf{S}\|_{\text{row},0} \leq K_0, \tag{5.14}$$

where $\|\cdot\|_F$ is the Frobenius norm. In this chapter, a greedy pursuit algorithm named as simultaneous OMP (SOMP) [189] is used to solve the NP-hard problem in Equation (5.14). Given the solved sparse matrix $\mathbf{S}$, the investigated test pixel centred at the neighbourhood window can be assigned to the class with the smallest residue, as shown below:

$$\text{Class}(\mathbf{x}_1) = \arg \min_{m=1,2,\ldots,M} r^m(\mathbf{X}) = \arg \min_{m=1,2,\ldots,M} \left\|\mathbf{X} - \mathbf{A}^m \hat{\mathbf{S}}^m\right\|_F. \tag{5.15}$$

5.3 The proposed approach

The spectral-spatial classification framework proposed in this chapter combines feature extraction and classification processes together. The effective feature extraction step is achieved by joint bilateral filtering with the first principal component as the guidance image, and the filtered image is classified using the SS-JSRC method. In the following part, details about the proposed classification framework is revealed.

For basic bilateral filtering introduced in Section 5.2.1, the Gaussian kernel for range filtering is calculated based on the image intensities of a small patch. If the pixel at (p, q) in the image patch has a similar intensity with the 'to-be-smoothed' pixel at (i, j), the weight of pixel (p, q) will be large. This is how the edge information in the image is preserved using bilateral filtering.

However, for some hyperspectral images, the contrast in a few spectral bands is low. If the range filtering kernel is calculated using those bands, the edge information will be lost in those bands after bilateral filtering. Therefore, in this chapter, it is proposed to use a joint bilateral filter with the first principal component as the guidance image since the first principal component after principal component analysis (PCA) accounts for as much of the variability in the hyperspectral image as possible. In this case, Equation (5.3) should be modified as follows:

$$I_{JBF}(i,j,b) = \frac{\sum\limits_{(p,q)\in w} G_{\sigma_d}(i-p,j-q)\, G_{\sigma_r}(I_{PC}(i,j) - I_{PC}(p,q))\, I(p,q,b)}{\sum\limits_{(p,q)\in w} G_{\sigma_d}(i-p,j-q)\, G_{\sigma_r}(I_{PC}(i,j) - I_{PC}(p,q))}, \tag{5.16}$$

where I_{PC} represents the first principal component (PC), b is the spectral band index and I_{JBF} is the output of joint bilateral filtering (JBF).

Figure 5.1 (a) and (b) are one input band image and the first principal component from the Indian Pines data set [10], while (c) and (d) show the filtered images with different sets of parameters. Before applying the joint bilateral filter, the the pixel intensities of the hyperspectral image are normalised to the range [0,1]. In Figure 5.1 (c), the kernel parameter of domain filter σ_d is set as 2 and the kernel parameter of range filter σ_r is changed from 0.1 to 0.4. As can be noticed, the performance of joint bilateral filtering gradually approaches Gaussian blurring with an increased σ_r. Hence, the range parameter σ_r is chosen as 0.1 for the edge-preserving effect and the domain parameter σ_d is changed from 1 to 4 for performance comparison in Figure 5.1 (d). Judging from the filtered image quality, a higher domain parameter leads to smoother features in the result. Therefore, the domain parameter σ_d is chosen as 4 in following experiments.

As mentioned above, joint bilateral filtering is applied to the hyperspectral image instead of the normal bilateral filtering since the filtered image quality in low-contrast spectral bands can be improved with the guidance image. Figure 5.2 (a) shows an example of low-contrast spectral band from the same data set, and the first principal component is shown in (b). The result with normal bilateral filtering and the one with joint bilateral filtering using the same set of parameters ($\sigma_d = 4, \sigma_r = 0.1$) are shown in Figure 5.2 (c) and (d), respectively. It is notable that the edge information is kept or even enhanced using joint bilateral filtering with the first principal component as the guidance image.

With the extracted spatial features using joint bilateral filtering, the hyperspectral image is ready for classification. As mentioned in Section 5.2.2, all pixels

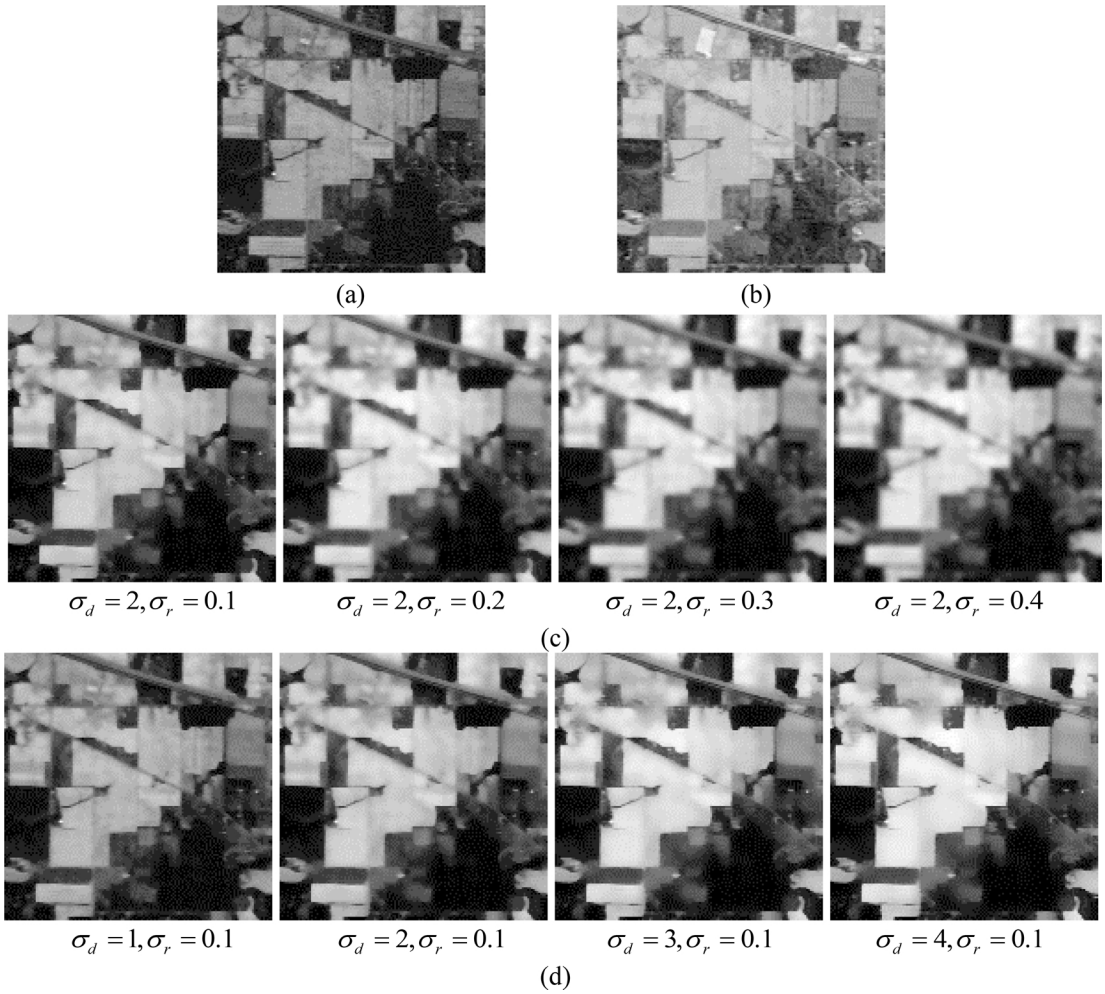


Fig. 5.1 The performance of joint bilateral filtering with different sets of parameters. (a) Input band image from the Indian Pines data set, (b) the first principal component, (c) filtered images with different range parameters, and (d) filtered images with different domain parameters.

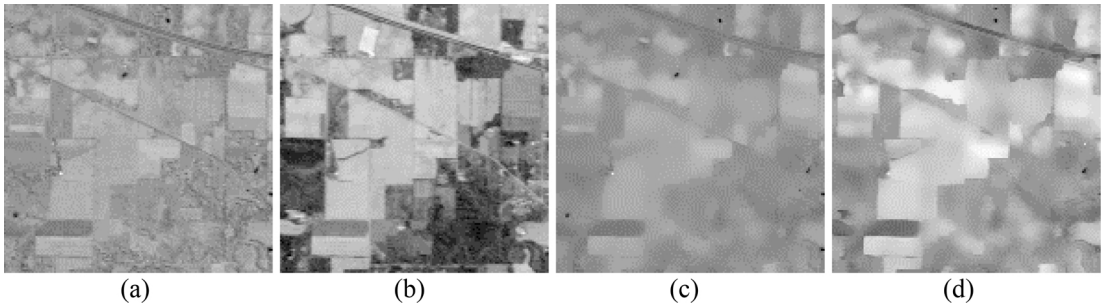


Fig. 5.2 Comparison of bilateral filtering and joint bilateral filtering on low-contrast image. (a) Input band image from the Indian Pines data set, (b) the first principal component, (c) output image using bilateral filtering, and (d) output image using joint bilateral filtering with the first principal component as the guidance image.

in a neighbourhood window are considered equally in the JSRC method. However, the window may contain pixels from different categories around class edges, which is against the assumption of JSM that neighbouring pixels from the same category share a common sparsity pattern. Therefore, the SS-JSRC method is proposed for getting rid of irrelevant pixels within the neighbourhood window and the flowchart is illustrated in Figure 5.3.

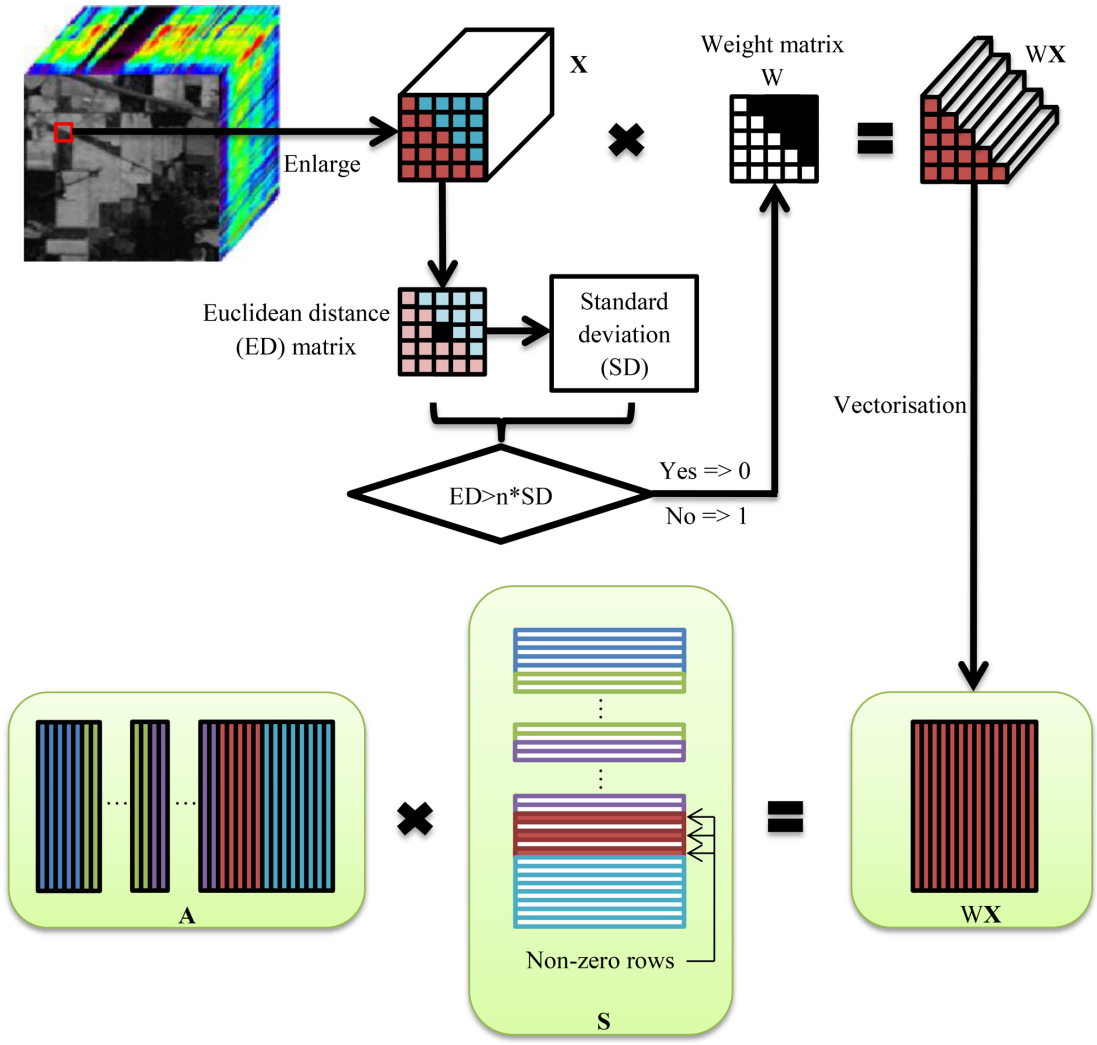


Fig. 5.3 The flowchart of the proposed SS-JSRC method.

As shown in Figure 5.3, an investigated test sample $\mathbf{x}_1$ is located at the centre of a small neighbourhood window in the Indian Pines data set and the data within the window is the sample matrix $\mathbf{X}$ in Equation (5.12). With the enlarged window, it can be noticed that the centre test pixel belongs to class red. However, pixels from class blue, which have different spectral characteristics from the centre test pixel, are involved in the sample matrix as well. To remove these pixels from class blue, a Euclidean distance (ED) matrix is obtained first. For every neighbouring

pixels in the window, its Euclidean distance to the centre test pixel is calculated and the standard deviation (SD) of this matrix can be achieved. Then, a threshold based on the standard deviation is set and pixels with distances lower than the threshold are considered similar to the centre test pixel as shown in Equation (5.17).

$$W = \begin{cases} 0, & ED > n * SD \\ 1, & ED < n * SD \end{cases} . \tag{5.17}$$

Then the joint sparsity model can be applied with the reduced sample matrix $W\mathbf{X}$ and the classified label for the centre test pixel can be achieved as follows:

$$\hat{\mathbf{S}} = \arg\min \|\mathbf{AS} - W\mathbf{X}\|_F \quad \text{subject to} \quad \|\mathbf{S}\|_{\text{row},0} \leq K_0, \tag{5.18}$$

$$\text{Class}\left(\mathbf{x}_1\right) = \arg \min_{m=1,2,\ldots,M} r^m\left(W\mathbf{X}\right) = \arg \min_{m=1,2,\ldots,M} \left\|W\mathbf{X} - \mathbf{A}^m\hat{\mathbf{S}}^m\right\|_F. \tag{5.19}$$

For details of the SOMP algorithm which is used to solve Equation (5.18), please refer to [188].

5.4 Data sets and experimental setup

In this section, the experimental data sets will be introduced as well as the parameter setting in the proposed classification framework.

Two publicly available and widely used HSI remote sensing data sets are employed to evaluate experimental performance in this chapter, including the airborne visible/infrared imaging spectrometer (AVIRIS) natural scenes Indian Pines [10] as well as the reflective optics system imaging spectrometer (ROSIS) urban scene Pavia University [141].

Details of these data sets can be found in Appendix A. For Indian Pines, the 16 classes ground truth map is used. Besides, bands 104–108, 150–163 and 220, which do not contain useful information due to atmospheric water absorption, are consequently removed to prevent from decreasing the classification accuracy, resulting in a reduced data set with 200 spectral bands. For Pavia University, a small portion is cropped from the original data set for following experiments in order to save computing time. One band from the cropped Pavia University data set and its corresponding ground truth map containing 8 classes are shown in Figure 5.4.

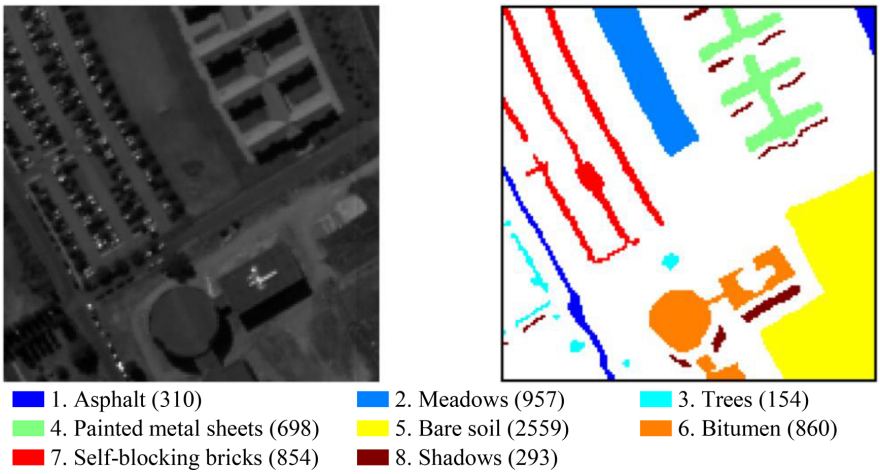


Fig. 5.4 Band 70 of the cropped Pavia University data set (left) and the 8 classes ground truth map (right), with number of samples in parentheses.

Experiments are repeated ten times using randomly selected training and test samples to avoid errors. For Indian Pines, 10% of the labelled samples from each class are chosen for training the classifier, while a smaller percentage, 5%, is used for Pavia University. The classification performance is evaluated using the overall accuracy (OA), the average accuracy (AA) as well as the kappa coefficient (κ). The kappa coefficient, also named as the kappa statistics, provides a standard for overall classification performance by comparing the agreement against the one occurring by chance and it is defined mathematically as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{5.20}$$

where p_o is the proportion of all correctly-labelled samples in the test set, i.e. OA, and p_e is the proportion agreement occurring by chance. The value of κ ranges from -1 to 1, where -1 stands for complete disagreement and 1 stands for perfect agreement.

In order to verify the efficacy of the proposed SS-JSRC method, it is first compared with the classical classifier SVM with the radial basis function (RBF) kernel [18], SRC [188], JSRC [188] and NLW-JSRC [191]. Optimal parameters for SVM are achieved using ten-fold cross-validation. It is suggested to choose a larger neighbourhood window size for images containing large homogeneous regions in [188]. As can be noticed that there are large homogeneous regions in both data sets, the neighbourhood window size is consequently chosen as 9 $\times$ 9. Since [191] did not provide the Gaussian kernel parameter σ used for the nonlocal weighting scheme, a value of 2 is chosen for the experiments in this chapter, while keeping other parameters the same. The sparsity level K_0 for all

sparsity representation-based methods are tested from $K_0 = 5$ to $K_0 = 40$. In Figure 5.5, the effect of the sparsity level K_0 and the threshold parameter n in the SS-JSRC method is revealed using the Indian Pines data set. As can be seen, the overall classification accuracies are generally high when the threshold parameter n is 2. Therefore, the value of n is set as 2 for following experiments.

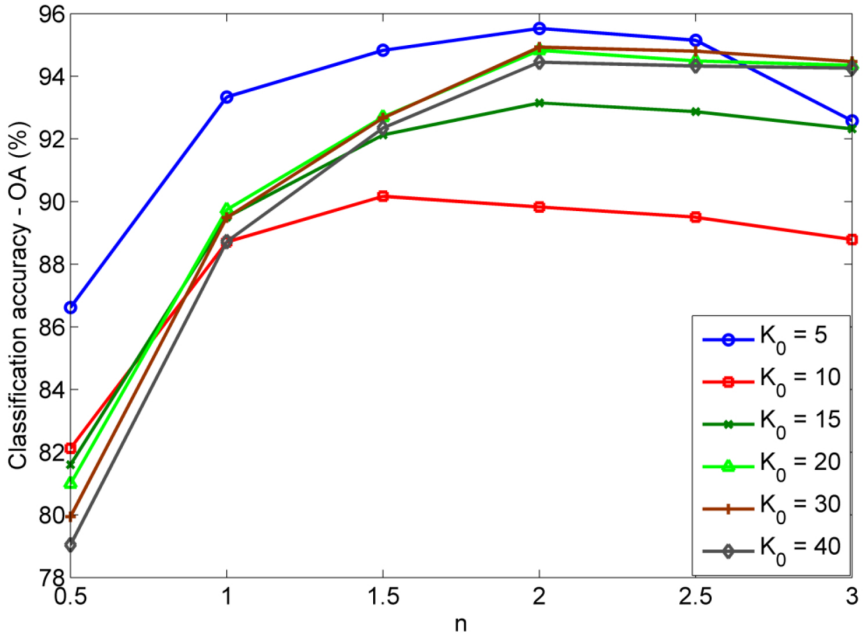


Fig. 5.5 Effect of the sparsity level K_0 and the threshold parameter n for Indian Pines with the proposed SS-JSRC method.

After the evaluation of classifiers, the proposed classification framework (JBF-SS-JSRC) using the joint bilateral filtering together with the SS-JSRC method is compared with a few state-of-the-art spectral-spatial classification frameworks. For the bilateral filter, the domain parameter σ_d and the range parameter σ_r are set as 4 and 0.1 respectively as discussed in Section 5.3. The first comparison method is SSA-SVM [71], where 2D-SSA is applied on each band for spatial feature extraction, followed by SVM for classification. Another popular method EMAP-SVM [73], where spatial features are achieved using a few attributes constructed on components extracted by PCA and classified by SVM, is included for comparison as well. The third comparison method proposes a spatial-aware dictionary learning (SADL) approach and the learnt sparse coefficients are classified using SVM [198]. Instead of only comparing with the 'feature extraction and classification' family, the proposed framework is also compared with the 'classification and post-processing' family. The method in [182], which utilises an edge-preserving filter for post-processing the initial SVM-classified result, is adopted and represented as SVM-EPF. In order to show that the joint bilateral filtering is superior to

the normal bilateral filtering, features learnt with these two filters are classified using SVM for comparison as well and these two approaches are abbreviated as BF-SVM and JBF-SVM respectively.

5.5 Results and discussion

With the two data sets and the experimental setup discussed in Section 5.4, experiments have been carried out and results are shown in following subsections.

5.5.1 Evaluation of classifiers

First, results from the SS-JSRC method are compared with those using SVM, SRC, JSRC and NLW-JSRC. The mean value and the standard deviation with ten experiments for Indian Pines and Pavia University are given in Table 5.1 and Table 5.2, together with chosen value for the sparsity level K_0.

It can be seen that in most cases, classifiers containing contextual information (JSRC, NLW-JSRC and SS-JSRC) outperform pixel-wise classifiers (SVM and SRC), which proves that the robustness of the joint sparsity model. Although authors in [191] claimed that NLW-JSRC helps to improve the classification accuracies over JSRC, it is impossible to achieve the optimal results without being given the right parameter for the Gaussian kernel, not to mention the heavy computational complexity when calculating the weight parameters. Comparing the proposed SS-JSRC method with the JSRC method, the overall accuracy has increased from 94.58% to 95.19% for Indian Pines and 96.79% to 97.47% for Pavia University, which seems not a big improvement. However, it is notable that the improvement of the average accuracy is impressive since the results of those classes with limited training samples are enhanced due to the simple pixel removal process in the SS-JSRC method. For example, class 9 in the Indian Pines data set is a narrow strip as shown in the ground truth map. With the JSRC method, pixels from both class 3 and class 6 are included in the 9×9 window, while irrelevant pixels can be removed using the SS-JSRC method successfully.

5.5.2 Comparison with typical state-of-the-art spectral-spatial classification frameworks

After the comparison of different classifiers, the proposed classification framework (JBF-SS-JSRC) is further compared with several other state-of-the-art classification frameworks under the same experimental setting used previously. The classification

Table 5.1 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classifiers for the Indian Pines data set with 10% of data used for training, followed by the standard deviation.

	Classifier				
Class	**SVM**	**SRC** ($K_0 = 5$)	**JSRC** ($K_0 = 30$)	**NLW-JSRC** ($K_0 = 30$)	**SS-JSRC** ($K_0 = 5$)
1	72.86 ± 17.64	66.33 ± 7.28	83.88 ± 5.65	81.02 ± 6.60	**91.75 ± 4.53**
2	85.78 ± 1.58	64.83 ± 1.98	**94.67 ± 1.40**	91.85 ± 2.88	94.65 ± 1.26
3	79.43 ± 1.95	62.85 ± 3.22	92.30 ± 2.45	85.38 ± 2.80	**92.39 ± 1.55**
4	74.93 ± 11.13	46.02 ± 7.26	**93.27 ± 3.21**	82.80 ± 6.20	83.46 ± 7.04
5	94.36 ± 2.27	90.16 ± 1.84	92.98 ± 2.29	91.48 ± 3.06	**94.59 ± 2.30**
6	97.59 ± 0.89	94.70 ± 1.20	99.17 ± 0.45	**99.51 ± 0.29**	96.91 ± 1.23
7	80.44 ± 11.27	84.78 ± 8.26	52.64 ± 21.54	68.70 ± 11.37	**86.96 ± 7.94**
8	98.34 ± 1.01	97.75 ± 1.32	99.96 ± 0.14	**100 ± 0**	99.41 ± 0.12
9	71.11 ± 19.91	52.78 ± 11.19	3.33 ± 4.68	44.45 ± 17.76	**77.78 ± 21.60**
10	83.37 ± 3.68	72.09 ± 2.59	86.83 ± 1.25	83.82 ± 4.04	**92.04 ± 2.06**
11	88.64 ± 1.08	76.92 ± 1.49	**97.48 ± 0.76**	96.83 ± 0.88	96.38 ± 1.16
12	84.27 ± 3.36	57.65 ± 2.69	86.62 ± 3.38	88.25 ± 3.06	**93.16 ± 3.65**
13	99.01 ± 0.87	97.91 ± 1.31	99.16 ± 0.62	99.32 ± 0.82	95.33 ± 3.40
14	95.41 ± 1.30	93.75 ± 1.79	99.30 ± 0.36	**99.61 ± 0.35**	99.44 ± 0.50
15	61.52 ± 3.84	44.91 ± 3.30	**97.69 ± 1.17**	94.36 ± 2.17	93.10 ± 3.77
16	90.12 ± 2.16	90.35 ± 4.92	94.94 ± 2.01	**96.82 ± 2.15**	96.47 ± 2.22
OA	87.76 ± 0.19	76.12 ± 0.49	94.85 ± 0.48	93.32 ± 0.64	**95.19 ± 0.44**
AA	84.82 ± 1.76	74.61 ± 0.99	85.89 ± 1.50	87.76 ± 1.36	**92.74 ± 1.37**
$\boldsymbol{\kappa}$	0.8604 ± 0.0021	0.7274 ± 0.0056	0.9412 ± 0.0054	0.9236 ± 0.0074	**0.9451 ± 0.0050**

Table 5.2 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classifiers for the Pavia University data set with 5% of data used for training, followed by the standard deviation.

	Classifier				
Class	**SVM**	**SRC** ($K_0 = 5$)	**JSRC** ($K_0 = 10$)	**NLW-JSRC** ($K_0 = 5$)	**SS-JSRC** ($K_0 = 10$)
1	65.92 ± 4.68	48.20 ± 4.72	77.35 ± 11.32	64.56 ± 4.78	**80.31 ± 8.20**
2	96.16 ± 1.40	94.04 ± 1.80	**99.76 ± 0.34**	99.55 ± 0.42	98.29 ± 1.03
3	94.66 ± 4.76	97.26 ± 1.99	90.68 ± 6.41	**98.50 ± 0.78**	94.32 ± 2.31
4	99.44 ± 0.51	99.73 ± 0.18	**99.98 ± 0.05**	99.88 ± 0.15	99.73 ± 0.09
5	96.36 ± 1.46	91.51 ± 1.53	**99.54 ± 0.73**	97.68 ± 0.66	98.56 ± 0.45
6	91.71 ± 1.73	86.11 ± 1.39	97.53 ± 1.69	96.51 ± 1.08	**98.63 ± 0.69**
7	90.80 ± 1.95	81.53 ± 3.53	98.75 ± 1.27	96.67 ± 1.25	**98.98 ± 1.05**
8	**96.33 ± 1.82**	93.42 ± 2.27	71.19 ± 5.86	83.13 ± 3.59	91.84 ± 1.68
OA	93.90 ± 0.95	88.97 ± 0.47	96.79 ± 0.66	95.74 ± 0.32	**97.47 ± 0.51**
AA	91.42 ± 1.10	86.47 ± 0.81	91.85 ± 1.63	92.06 ± 0.82	**95.08 ± 1.23**
$\boldsymbol{\kappa}$	0.9221 ± 0.0120	0.8594 ± 0.0058	0.9588 ± 0.0086	0.9457 ± 0.0041	**0.9677 ± 0.0065**

maps of various methods for Indian Pines and Pavia University are shown in Figure 5.6 and Figure 5.7. Detailed numerical results over ten repetitions are shown in Table 5.3 and Table 5.4, respectively.

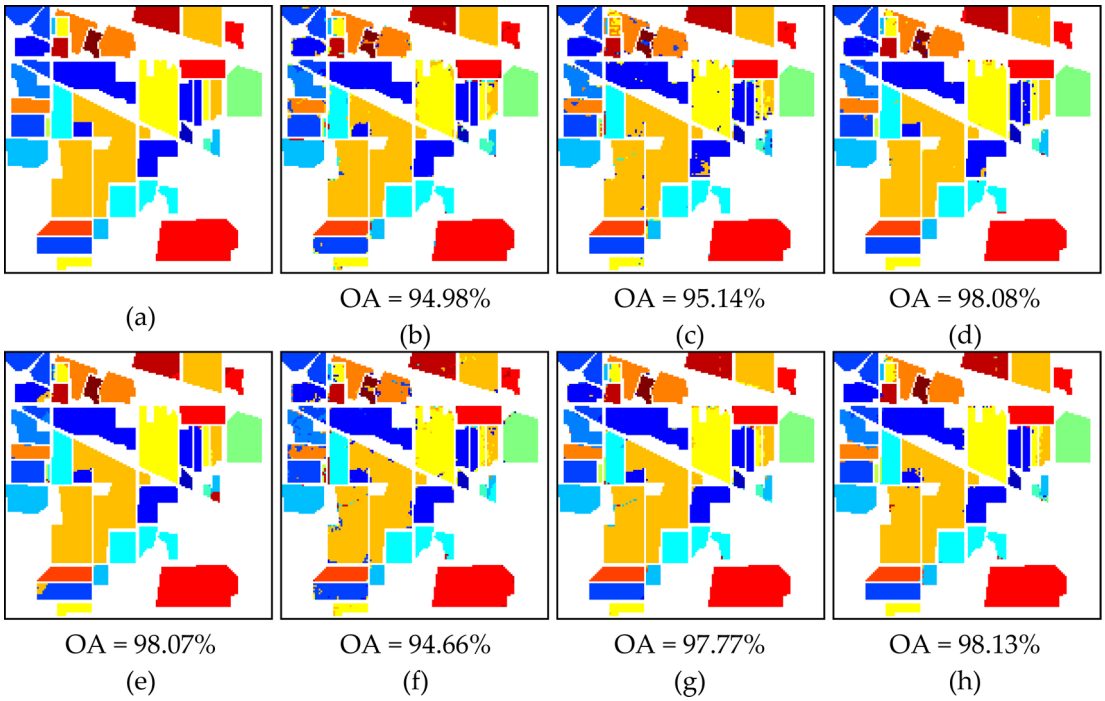


Fig. 5.6 The Indian Pines data set (a) ground truth and classification maps together with corresponding overall accuracy obtained by (b) SSA-SVM, (c) EMAP-SVM, (d) SADL-SVM, (e) SVM-EPF, (f) BF-SVM, (g) JBF-SVM and (h) JBF-SS-JSRC.

As shown in Table 5.3 and Table 5.4, four main findings can be identified: 1) by comparing results from BF-SVM and JBF-SVM, it is found that joint bilateral filtering with the first principal component as the guidance image yields better spatial feature extraction performance than normal bilateral filtering, which is due to the fact that the edge information is enhanced with the former. Take the Indian Pines data set in Figure 5.6 (f) and (g) as an example, there are more misclassified pixels around edges with the normal bilateral filtering. 2) Results achieved with JBF are higher than some feature extraction techniques, such as SSA and SADL, under the same classifier. Occasionally results of EMAP are slightly higher than those of JBF. However, its computation is extremely slow. The computing time of these feature extraction techniques is recorded on a personal computer with an Intel Core i5-2400 CPU at 3.10 GHz using Matlab 2014a (Mathworks) and results are presented in Table 5.5. 3) The proposed classification framework JBF-SS-JSRC outperforms the one using JBF-SVM, which proves the finding in Section 5.5.1 again that SS-JSRC is superior to the pixel-wise classifier SVM. 4) For both data sets, the performance of the proposed JBF-SS-JSRC framework is slightly better

Table 5.3 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classification frameworks for the Indian Pines data set with 10% of data used for training, followed by the standard deviation.

	Classification framework						
Class	**SSA-SVM**	**EMAP-SVM**	**SADL-SVM**	**SVM-EPF**	**BF-SVM**	**JBF-SVM**	**JBF-SS-JSRC** ($K_0 = 30$)
1	78.96 ± 12.86	87.92 ± 4.79	88.33 ± 8.29	**100 ± 0**	86.04 ± 6.22	87.08 ± 8.61	98.54 ± 1.72
2	96.45 ± 0.82	90.30 ± 1.27	95.12 ± 1.23	97.40 ± 1.20	94.66 ± 0.78	**97.67 ± 0.65**	97.40 ± 0.71
3	93.91 ± 2.04	97.08 ± 1.35	96.33 ± 2.27	98.35 ± 0.57	93.99 ± 1.22	**98.55 ± 0.62**	97.72 ± 1.11
4	89.09 ± 2.07	92.14 ± 4.86	89.00 ± 4.08	92.99 ± 5.61	88.29 ± 2.32	95.24 ± 2.40	**95.67 ± 3.55**
5	93.74 ± 1.33	94.68 ± 2.00	97.18 ± 1.90	**99.40 ± 0.43**	96.40 ± 2.76	96.71 ± 2.31	97.52 ± 1.92
6	98.66 ± 0.79	98.82 ± 0.85	**99.09 ± 1.00**	98.62 ± 1.38	98.23 ± 0.77	97.75 ± 0.71	98.02 ± 0.40
7	72.17 ± 23.83	95.65 ± 2.90	**100 ± 0**	95.75 ± 13.44	96.09 ± 1.37	94.78 ± 3.43	95.22 ± 3.21
8	98.32 ± 0.99	99.46 ± 0.29	**100 ± 0**	99.34 ± 2.08	99.07 ± 0.23	99.61 ± 0.34	99.93 ± 0.15
9	77.78 ± 23.13	76.67 ± 15.00	**95.00 ± 10.62**	80.00 ± 42.16	75.00 ± 18.75	88.89 ± 15.93	88.89 ± 9.44
10	92.77 ± 1.91	89.69 ± 2.38	**97.16 ± 1.57**	93.39 ± 4.56	93.10 ± 1.81	95.68 ± 1.83	96.14 ± 1.14
11	95.41 ± 1.45	97.73 ± 0.51	**98.70 ± 0.93**	96.04 ± 1.15	94.38 ± 1.20	97.45 ± 1.03	98.65 ± 0.73
12	92.84 ± 3.61	89.97 ± 4.06	93.19 ± 3.05	**98.76 ± 0.53**	90.45 ± 2.21	95.38 ± 1.87	96.29 ± 1.85
13	98.58 ± 1.88	99.42 ± 0.16	98.74 ± 0.67	**100 ± 0**	98.21 ± 0.71	98.53 ± 1.05	99.11 ± 0.79
14	98.36 ± 0.62	99.32 ± 0.34	99.30 ± 0.79	96.65 ± 2.17	97.65 ± 1.24	98.93 ± 0.50	**99.93 ± 0.05**
15	91.35 ± 5.55	96.93 ± 1.55	97.95 ± 2.23	97.43 ± 3.04	92.34 ± 5.03	94.68 ± 2.71	**98.54 ± 0.77**
16	94.00 ± 6.71	89.77 ± 9.40	96.47 ± 4.37	**98.93 ± 0.38**	92.12 ± 6.86	95.76 ± 2.30	97.65 ± 0.55
OA	95.28 ± 0.64	95.36 ± 0.36	97.31 ± 0.40	96.93 ± 0.64	94.81 ± 0.39	97.29 ± 0.35	**98.05 ± 0.24**
AA	91.40 ± 2.20	93.47 ± 1.62	96.35 ± 0.84	96.44 ± 3.32	92.67 ± 1.85	95.79 ± 1.15	**97.20 ± 0.55**
$\boldsymbol{\kappa}$	0.9462 ± 0.0073	0.9471 ± 0.0041	0.9694 ± 0.0045	0.9650 ± 0.0073	0.9409 ± 0.0044	0.9696 ± 0.0035	**0.9778 ± 0.0027**

Table 5.4 Mean class-by-class, average, overall accuracies (%) and the kappa coefficient of different classification frameworks for the Pavia University data set with 5% of data used for training, followed by the standard deviation.

	Classification framework						
Class	**SSA-SVM**	**EMAP-SVM**	**SADL-SVM**	**SVM-EPF**	**BF-SVM**	**JBF-SVM**	**JBF-SS-JSRC** ($K_0 = 5$)
1	77.95 ± 4.92	**99.73 ± 0.50**	94.52 ± 3.37	99.45 ± 1.05	98.44 ± 1.88	97.18 ± 2.83	99.63 ± 0.95
2	91.57 ± 1.91	99.83 ± 0.08	98.84 ± 0.92	99.64 ± 0.96	99.83 ± 0.19	99.72 ± 0.33	**99.85 ± 0.49**
3	97.08 ± 2.46	89.66 ± 4.17	97.53 ± 3.53	**98.46 ± 2.80**	95.48 ± 2.61	98.36 ± 1.17	97.95 ± 2.35
4	99.78 ± 0.29	99.41 ± 0.29	99.74 ± 0.19	**100 ± 0**	99.47 ± 0.61	99.37 ± 0.39	99.91 ± 0.15
5	96.62 ± 1.09	**99.95 ± 0.06**	99.23 ± 0.78	99.94 ± 0.12	99.64 ± 0.36	99.60 ± 0.23	99.93 ± 0.03
6	94.85 ± 1.26	**99.93 ± 0.13**	97.77 ± 2.44	99.37 ± 1.27	98.00 ± 2.49	97.80 ± 2.09	99.89 ± 0.35
7	85.60 ± 2.59	99.06 ± 0.52	98.37 ± 0.55	96.38 ± 3.16	98.09 ± 1.40	98.22 ± 0.75	**99.62 ± 0.21**
8	98.70 ± 1.16	**100 ± 0**	99.93 ± 0.15	99.89 ± 0.17	95.76 ± 3.77	97.19 ± 2.55	96.58 ± 2.74
OA	93.79 ± 0.73	99.52 ± 0.13	98.70 ± 0.45	99.28 ± 0.60	98.92 ± 0.37	98.94 ± 0.44	**99.67 ± 0.06**
AA	92.77 ± 0.69	98.45 ± 0.56	98.24 ± 0.68	99.14 ± 0.64	98.09 ± 0.53	98.43 ± 0.69	**99.19 ± 0.17**
$\boldsymbol{\kappa}$	0.9209 ± 0.0094	0.9939 ± 0.0017	0.9835 ± 0.0057	0.9909 ± 0.0077	0.9863 ± 0.0048	0.9865 ± 0.0057	**0.9958 ± 0.0007**

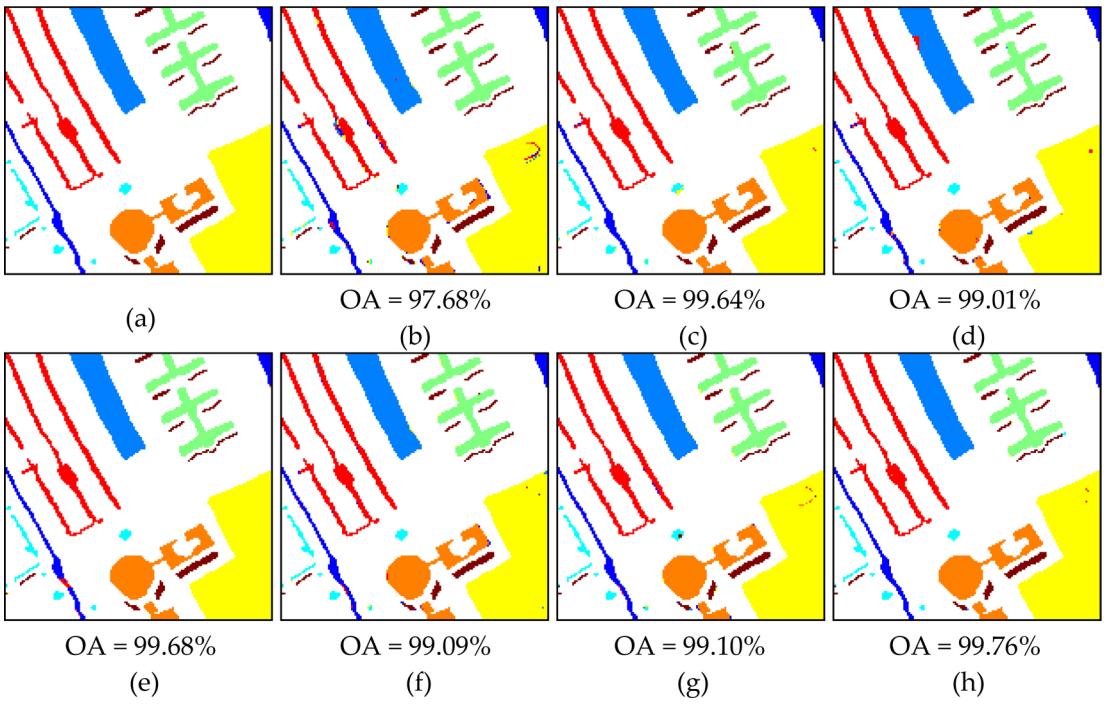


Fig. 5.7 The Pavia University data set (a) ground truth and classification maps together with corresponding overall accuracy obtained by (b) SSA-SVM, (c) EMAP-SVM, (d) SADL-SVM, (e) SVM-EPF, (f) BF-SVM, (g) JBF-SVM and (h) JBF-SS-JSRC.

than other state-of-the-art classification frameworks. For instance, as observed from Table 5.3, the overall accuracy of JBF-SS-JSRC is 2.77%, 2.69%, 0.74% and 1.12% higher than that of SSA-SVM, EMAP-SVM, SADL-SVM and SVM-EPF, respectively. All in all, the obtained results confirm the strong superiority of joint bilateral filtering and SS-JSRC over other popular spectral-spatial classification frameworks.

Table 5.5 Computing time (in seconds) for different feature extraction techniques.

Data sets	**Feature descriptor**				
	SSA	**EMAP**	**SADL**	**BF**	**JBF**
Indian Pines	22	637	77	66	4
Pavia University	12	423	9	36	3

5.6 Summary

In this chapter, efforts have been focused on developing effective algorithms for sparse representation-based classification of hyperspectral images. The task is divided into two parts, i.e. extracting effective features and constructing suitable classifiers. To this end, joint bilateral filtering with the first principal component

as the guidance image, is proposed to extract spatial features without blurring the edge information, where its superiority has been proven by comparing with the normal bilateral filtering. In addition, an alternative classifier SS-JSRC, which is an updated version of the JSRC method, is proposed for improved classification capabilities. With the SS-JSRC method, a simple yet powerful pixel removal process based on the statistics within the neighbouring window, is added before applying the joint sparsity model. Results show that both the classification accuracy and the kappa coefficient are improved over those of JSRC and NLW-JSRC since irrelevant pixels have been effectively removed using SS-JSRC, especially for classes with a small spatial area. By combining joint bilateral filtering and the SS-JSRC method together, the classification performance of the proposed approach is shown to outperform several typical spectral-spatial classification approaches for both Indian Pines and Pavia University data sets.

Chapter 6

Meat Quality Assessment Using Hyperspectral Imaging

6.1 Introduction

Serving as a very important source of nutrition, the quality of meat and fish influences the repurchase behaviour of consumers [123]. In order to meet the quality standard of target consumers, control process must be carried out on meat products before entering the market. Considering that food quality control requires non-destructive real-time monitoring on the production line, visible and near-infrared (VIS-NIR) spectroscopy was first established as a fast and promising tool for multi-constitute food quality analysis. In recent years, the ability of VIS-NIR spectroscopy for food quality assessment has been investigated by many researchers. Due to its characteristics of being rapid and non-destructive, VIS-NIR spectroscopy is widely applied to determine the composition of a variety of agricultural products, which has proved its feasibility especially in meat industry [123, 199].

However, with limited spatial information, sometimes internal constituent gradients within food products could not be captured by VIS-NIR spectroscopy successfully, leading to discrepancies between predicted and measured composition [199]. Therefore, multispectral imaging (MSI) system, combining images at a small number of narrow wavebands, was developed to overcome the above mentioned drawbacks and has demonstrated its success for detection of meat quality [19, 200].

By combining hundreds of continuous bands at different wavelengths, another similar platform named as the hyperspectral imaging (HSI) system, has received considerable attention in recent years as an updated version of the MSI system. Unlike VIS-NIR spectroscopy which could only attain the reflectance spectrum for a small area of an object, HSI is able to collect the spectral information for

more pixels over a large spatial area on the object. Due to the fact that HSI could acquire more details than VIS-NIR spectroscopy and MSI, there is a growing trend to investigate its ability for fast, non-destructive and non-reagent based meat quality prediction. In this chapter, beef and lamb are chosen as representatives of muscle food since beef contributes most to the meat market in EU with 8,000,000 tonnes annual consumption and similar levels of production [19], while lamb also plays an important role in UK agriculture, contributing over 10% of total livestock output [201].

Meat eating quality is related to many chemical and physical properties. It is found that three key factors, including tenderness, juiciness, and flavour, contribute to the consumer perception of satisfactory meat quality [96]. Surface colour and ultimate pH, which are related to juiciness and flavour, can influence meat shelf life and determine retail selection of consumers [202]. However, tenderness is usually accepted as the most critical issue that affects the repurchasing behaviour of customers [203], where it is commonly represented quantitatively by Warner-Bratzler shear force (WBSF) or slice shear force (SSF).

Objective approaches as such offered by spectral technologies to determine meat quality may be useful, but the analytical algorithms used remain to be optimised. For predicting the above-mentioned meat quality traits, various statistical regression approaches can be used to construct calibration equations, including multiple linear regression (MLR), partial least squares regression (PLSR), principal component regression (PCR) and neural networks [204]. Among these approaches, PLSR is the most common method to construct the calibration models from VIS-NIR spectra for the prediction of sensory parameters in meat [99, 104, 107, 108]. The PLSR model establishes a relationship between independent spectral variables $\mathbf{X}$ and the dependent quality trait variable Y, where data in $\mathbf{X}$ is projected onto a few latent vectors called PLS components [205]. However, PLSR is based on the linear algorithm so that the best performance can only be achieved when there is a linear relationship between spectra and quality trait values [20], which might not be the case in every situation. In contrast, the support vector machine (SVM), another popular machine learning approach, is based on the statistical learning theory and is effective in the prediction of both linear and non-linear data [206]. Advantages of SVM over other methods are that it does not require large quantities of samples to construct the calibration model and seems unaffected by sample outliers [125]. Although SVM has been successfully used in various disciplines, it has not been applied extensively to the analysis of meat quality parameters based on spectroscopic technologies. To this end, the performance of

PLSR and SVM as tools for meat quality analysis using VIS-NIR spectroscopy is evaluated first in this chapter. To the best of my knowledge, there is no research comparing the performance of VIS-NIR spectroscopy and HSI on the same data set. Hence, the second objective of this chapter is to determine whether HSI could outperform VIS-NIR spectroscopy using the same batch of beef samples.

Besides, usually for both VIS-NIR spectroscopy and HSI, pre-processing of the spectral profile is needed, to eliminate undesired effects and noise produced during data collection process. Common pre-processing techniques, especially for VIS-NIR spectra, include calculating derivatives, standard normal variate (SNV) and multiplicative scatter correction (MSC) [207]. However, using derivatives of spectra may even enhance the noise and lead to more difficult spectral interpretation. For SNV and MSC, it is required to apply these transformations to all spectra as the corrected spectra would be more accurate if more spectra were involved, which is infeasible in the abattoir. In practice, ideally a prediction model based on HSI will be installed in the abattoir production line, and for every single piece of beef steak, the model should predict the quality assessment result in a real-time manner. Therefore, it is necessary to use a pre-processing technique that can be applied to every single spectrum itself, without considering other spectra. The third objective of this chapter is to demonstrate that the previously mentioned singular spectrum analysis (SSA) could be regarded as an optimal pre-processing step in de-noising HSI spectra of beef, where it will not be restricted by the number of HSI samples. Last but not least, the feasibility of HSI to facilitate abattoirs classifying carcasses into different grading levels is also tested on lamb since limited research has been conducted using HSI to predict eating quality traits of lamb samples.

The remaining sections of this chapter are organised as follows. In Section 6.2, different data sets used in the experiment and data processing methods are introduced. Results of four objectives are revealed in Section 6.3, followed by summary in Section 6.4.

6.2 Materials and methods

There are four data sets involved in total, corresponding to the experiment for each objective in this chapter. For simplicity, these data sets are named as data set A, B, C and D respectively. Details of data acquisition, quality reference measurement and data processing methods are explained down below.

6.2.1 Data set A

Data set A is acquired on beef samples using VIS-NIR spectroscopy. A total of 234 cattle below 30 months of age, comprising 75 heifers, 118 steers, and 41 young bulls, are randomly selected in a Scottish abattoir over a 10-week period. The mean hot carcass weight is 332.6 kg, which is similar to the average carcass weight (341.7 kg) of steers, heifers, and young bulls slaughtered in UK in January 2014 [208]. This data set is acquired as part of another experiment, where PLSR was used for beef quality prediction [97].

All 234 carcasses are quartered between the 10th and 11th ribs at 48 hours post mortem into hind and forequarters. An approximate 2.5 cm piece of beef sample containing the *M. longissimus thoracis*, associated muscles, and subcutaneous fat is removed from the 11th rib section of each carcass at quartering. After 2 min of blooming as advised in [203], 10 replicate scans are randomly collected from different sites on the cut surface using an ASD QualitySpec Pro VIS-NIR spectrometer with a fibre-optic probe (wavelength λ = 350 – 1800 nm at 1 nm intervals). A calibration reference is required to determine corresponding wavelengths where absorption occurs and it is achieved by using a white calibration tile that reflects nearly 100% of the radiation at all VIS-NIR wavelengths [89]. Dividing the reflected spectrum by the white reference spectrum, the reflectance R can be calculated as

$$R = \frac{I}{I_0}, \tag{6.1}$$

where I and I_0 are intensities of the reflected spectrum and the white reference, respectively. For VIS-NIR spectroscopy to measure meat and meat products, reflectance is usually converted to absorbance (A) by logarithm transform, linearising the relationship between the concentration of an absorbing compound and the absorption spectrum [107], as shown in Equation (6.2).

$$A = \log_{10}\left(\frac{1}{R}\right). \tag{6.2}$$

Once the spectral information on the steak is acquired, ultimate pH (pH_u) is determined with a Testo 205 pH meter. The colour of the *M. longissimus lumborum* is measured in the L*a*b* scale with a Minolta CR-410 colourimeter after allowing the surface to bloom for 45 min, where L* represents the lightness, a* is the green-red indicator, and b* is the blue-yellow indicator. Steaks are then sealed in plastic bags, transported at 4 °C to the laboratory, and stored overnight

at 3 °C. On the following day (3 days post mortem), steaks are cooked on a clam-shell grill until the internal temperature reached 71 °C, measured using a stainless thermometer (Hanna HI-98509 Check-temp 1). Once cooked, a 50 × 10 mm slice of steak is sheared orthogonal to the muscle fibre axis using a Lloyd TA-plus texture analyser fitted with a flat blunt-end blade. Then, the peak force is extracted from the force deformation curve as the SSF.

As mentioned above, 10 replicate spectra are acquired for each sample and they are plotted in Figure 6.1. It is shown in the plot that excessive noise exists in the extreme parts of the spectra, which is illustrated by dotted lines. Removing the noise results in 495 – 1600 nm as the working spectra plotted in solid lines. The median spectrum of 10 replicates is used to represent the spectral profile of each steak, followed by transforming the reflectance to absorbance.

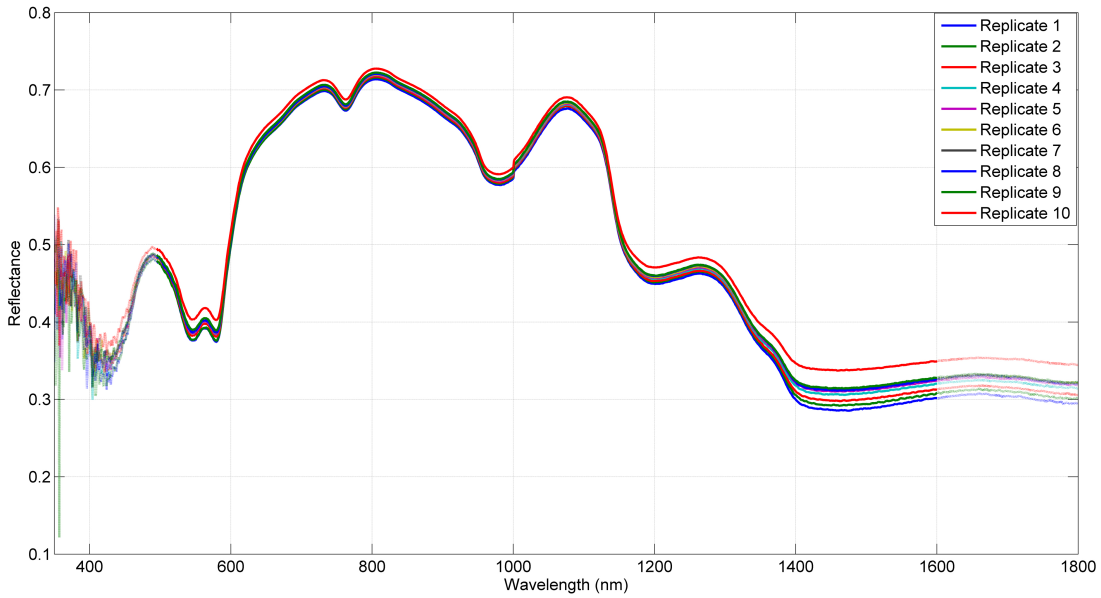


Fig. 6.1 A plot of 10 replicate spectra (350 – 1800 nm) collected from the *M. longissimus thoracis* of one carcass in data set A, showing the noise at the extremes of the spectra.

The performance of SVM is compared with PLSR, where both experiments are carried out using Matlab. Similar to the data set construction method used for PLSR in [97], steak samples are split into calibration (training) and validation (test) data sets. Being sorted in ascending order separately for each trait, every 4th sample is then selected into the validation set, with the intervening three samples being allocated to the calibration set, as recommended in [209]. Accordingly, 75% of the whole data set is assigned to the calibration set and the remaining 25% is assigned to the validation set. In this way, it makes sure that the validation set is a representative of the calibration set with similar mean and standard deviation (SD). After the calibration data used for modelling are defined, the number of

PLS components is then chosen based on the full leave-one-out cross-validation, where each sample is removed, predicted, and replaced in a sequential manner [210].

For VIS-NIR spectral data, the high dimensionality could increase the computational complexity for SVM regression [18]. Therefore, feature selection and reduction before SVM is expected, in order to reduce the data dimensionality by mapping the data to a new space where the differentiability is higher in a subset of transformed features than in any subset of the original data [4]. The feature reduction technique used in the experiment is principal component analysis (PCA) and the number of reduced dimensionality is decided by trial and error. With reduced dimensionality, it is easy for SVM to construct the prediction model from the calibration set. The grid-search is conducted with four-fold cross-validation to optimise the regression model.

6.2.2 Data set B

Data set B is acquired on beef samples using both VIS-NIR spectroscopy and HSI system. A total of 858 cattle are collected in four commercial abattoirs across Scotland. In each abattoir, over 200 beef samples are randomly selected in the production line. Allowing for 48 hours of ageing, an approximate 2.5 cm piece of steak sample containing the *M. longissimus thoracis* is removed from each carcass. After blooming for 2 min, hyperspectral images of steak samples are collected using a push-broom HSI system (Gilden photonics) with wavelength ranging from 400 to 862.90 nm at a spectral resolution of about 2.5 nm. Figure 6.2 shows a schematic diagram of the imaging system.

The HSI system, which consists of an 12-bit Andor Luca electron multiplying charge-coupled device (EMCCD) camara and a Specim V8E spectrograph, could acquire the high spatial and spectral resolution information of the investigated object. Samples are illuminated with an Armley 150 W halogen lamp and the exposure time is set as 6 ms. For each pixel, light is reflected from the target sample onto the entrance slit (width of 30 μm and length of 11.84 mm) and dispersed into a spectrum by the grating in the spectrograph. With the help of a Zolix KSA 11-200S4N motorised stage set at a scanning speed of 10 deg/s for the screw, the HSI system could achieve the undistorted hyperspectral image in the push-broom way, which only scans a single line of the investigated target at a time. The push-broom technique allows rapid and stable image acquisition, which is ideal for production line measurement in meat industry.

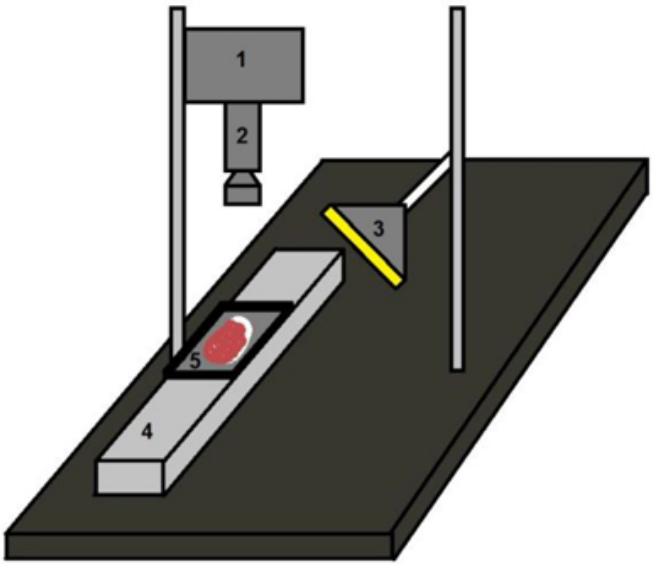


Fig. 6.2 Schematic diagram of the HSI system: components 1-5 refer to the EM-CCD camera, spectrograph and lens, halogen lamp, sliding track and scanning tray, respectively.

Before the image collection starts, a spectral calibration procedure has to be conducted at two extreme illuminating conditions, using a white tile that reflects nearly 100% of the radiation at all working wavelengths and also the cap of lens to get a dark image. These steps make sure that the sample reflectance can be separated from the system response [112]. The following equation shows how the calibrated reflectance image is achieved,

$$R = \frac{I - B}{I_W - B}, \tag{6.3}$$

where R is the calibrated reflectance image at one spectral band, I is the raw image at that band, B is the dark reference image and I_W is the white reference image.

After the hyperspectral image collection, VIS-NIR spectra of data set B are collected by a VIS-NIR spectrometer (ASD Labspec Pro) with wavelength from 350 nm to 2500 nm. The spectrometer is fitted with an internally-illuminated fibre optic probe (active scanning diameter $\oslash$ = 68 mm). Due to the fact that only a small area can be collected by the VIS-NIR spectrometer at a time, similarly, scans are replicated in 10 random places for each steak to capture the maximum variation across the sample.

Each steak is divided into 2 halves after data collection, labelled and vacuum packaged. One half is aged for an additional 5 days at -1 ^{o}C and the other half is aged for an additional 12 days under commercial conditions. Thus, steak samples

are aged for 7 days and 14 days in total. Before the quality reference measurement, samples are defrosted at the ambient temperature for 24 hours. On the next day, ultimate pH is determined with a calibrated Hanna meat pH meter (HI 99163), followed by the tenderness measurement. Steaks are cooked on a clam-shell grill until the internal temperature reaches 71 °C. Once cooked, a slice of steak is sheared orthogonal to the muscle fibre axis using a Tenderscot tenderometer (Pentland Precision Engineering Ltd) and the peak force is extracted as SSF during the shearing process. In summary, there are four beef quality parameters in total for each steak from data set B, which are pH7, pH14, SSF7 and SSF14 respectively.

Similar to data set A, excessive noise can be also noticed in the extreme parts of VIS-NIR spectra for data set B as shown by dotted lines in Figure 6.3. After removing the noise, the working wavelength for VIS-NIR spectra (plotted by solid lines) is 501 – 2200 nm and the median spectrum is used to represent the sample spectral profile.

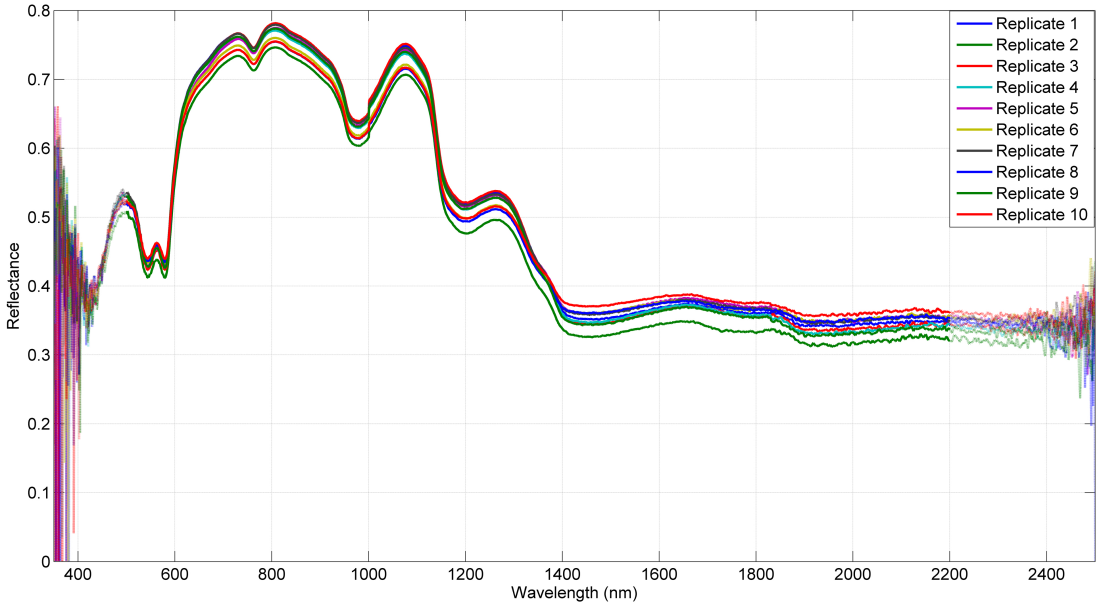


Fig. 6.3 A plot of 10 replicate spectra (350 – 2500 nm) collected from the *M. longissimus thoracis* of one carcass in data set B, showing the noise at the extremes of the spectra.

There are originally 250 spectral bands for each acquired hyperspectral image. However, due to the fact that the camera and the spectrograph do not have the same spectral range, not all bands contain useful information. By removing redundant data in the hyperspectral image, 200 spectral bands are left for the following analysis, with working wavelength ranging from 399.62 nm to 862.90 nm.

In order to improve the regression performance and processing speed, a region of interest (ROI) with size of 200 × 100 is selected from the lean part on each hyperspectral image, illustrated in Figure 6.4 (a) and (b). Even though most of fat area is discarded after ROI selection, there are still a few pixels that have different spectral features from pure lean pixels, i.e. dead pixels. In this context, an iterative pixel removal process [211] is applied to the ROI to exclude these dead pixels. For all spectra in the ROI, the Euclidean distances to the median spectrum and the standard deviation of these distances are calculated. Then those pixels with distances higher than five standard deviations are removed from the ROI. It is assumed that 100 times iterations should be enough for removing dead pixels in the current experiment. After these pre-processing steps, a median reflectance spectrum in Figure 6.4 (c) can be achieved.

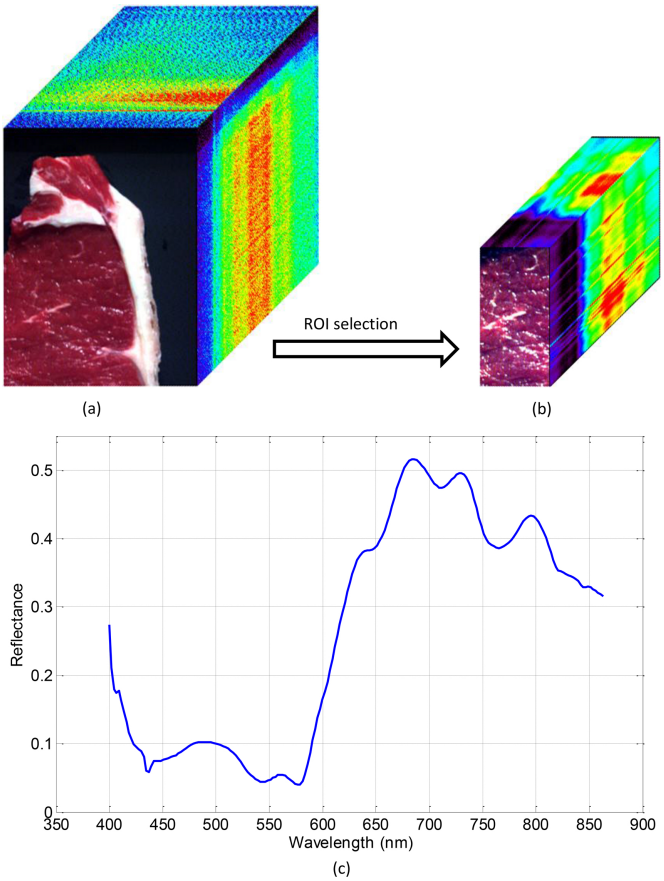


Fig. 6.4 HSI spectral feature extraction process. (a) The original hyperspectral image, b) the cropped hyperspectral image after ROI selection, and (c) the median reflectance spectrum of the ROI after the iterative pixel removal process.

For both VIS-NIR spectra and HSI spectra, reflectance needs to be converted to absorbance as mentioned previously in Equation (6.2). Similarly, samples in data set B are also split into two sets, where 75% is used for calibration and 25%

is used for validation. PCA is applied for dimensionality reduction and a four-fold cross-validation is adopted by SVM to optimise parameters for the calibration set to avoid over-fitting.

6.2.3 Data set C

Data set C is used for testing the performance of SSA as an optimal pre-processing step in de-noising HSI spectra of beef. Therefore, in order to save computing time, a subset from data set B is chosen for the experiment. There are over 200 samples in data set C, which are all from one abattoir. HSI data collection, quality reference measurement and spectra extraction steps are the same as described in Section 6.2.2. The only difference is that SSA is applied on the extracted HSI spectra for further noise removal. The algorithm of SSA can be found in Chapter 4.

There are two important parameters that affect the performance of SSA, which are the window length L and the eigenvalue grouping (EVG). The window length determines the number of decomposed series after SSA, while the eigenvalue grouping defines how many decomposed series are used for reconstruction in order to remove noise. In the experiment, only the first decomposed series is used for spectrum reconstruction and then the only parameter to be adjusted is the window length L. As an example, SSA is applied to the same spectrum shown in Figure 6.4 (c) with the window length set as 5, and the result is shown in Figure 6.5. Even though a subtle difference can be noticed in the plot, it is demonstrated that SSA does improve the prediction performance in the result section. After effective feature extraction by SSA, reflectance spectra are converted to absorbance spectra. Data splitting method is the same as data set B, and PCA is also applied before model construction by SVM.

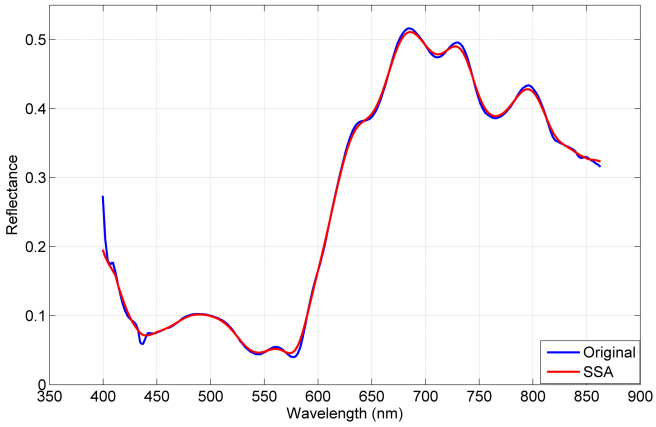


Fig. 6.5 The original median spectrum and the SSA processed median spectrum.

6.2.4 Data set D

Data set D is acquired on lamb samples using the HSI system. A total of 80 lamb samples are purchased from a commercial abattoir in Scotland, regardless of sex, age and breed groups. Storing at 3±2 °C for 7 days after being slaughtered, left lamb loins are removed from packaging. Allowing blooming for 2 min, hyperspectral images are collected using the HSI system (Gilden photonics) mentioned in Section 6.2.2.

Right after imaging, ultimate pH is determined by probing into the lamb sample directly using a calibrated Hanna pH meter with a glass electrode (HI 99613). The surface colour is measured in the L*a*b* scale with a Konica Minolta CR-410 colourimeter, where the machine is set to take three scans and the average value is calculated as the final reading. Then, each lamb sample is labelled, vacuum packaged and frozen at -30 °C to prevent from further ageing. The tenderness measuring standard defined by Meat Industry Research Institute of New Zealand (MIRINZ) named as MIRINZ shear force (SF) [212] is measured as the tenderness reference for data set D. On the night before tenderness measurement, samples are placed in plastic bags and sub-merged in a water bath until reaching an internal temperature of 70 °C and then chilled in the fridge. On the following day, ten sub-samples with 10 mm × 10 mm cross section are prepared parallel with the muscle fibre axis. All sub-samples are then sheared orthogonal to the fibre axis with a Tenderscot tenderometer (Pentland Precision Engineering Ltd). As suggested by the MIRINZ protocol, a wedge shape tooth blade is attached in the tenderometer for imitating the chewing behaviour of human beings. Similarly, the peak force is extracted during each shear process and the average value of ten measurements is considered as the final MIRINZ SF to minimise error. In total, there are 5 lamb quality related parameters for each sample, which are pH_u, L*, a*, b* and MIRINZ SF.

Similarly, hyperspectral images of lamb samples need to be calibrated first using Equation (6.3). Then redundant bands are removed, leaving 170 spectral bands with working wavelength ranging from 469.47 nm to 862.90 nm. Also, ROI with size of 200 × 100 pixels (approximately 100 × 50 mm^2) is mainly selected from the lean part and the iterative pixel removal process mentioned for data set B is applied. Finally, a median spectrum is achieved as illustrated in Figure 6.6. Reflectance is then converted to absorbance and SSA is applied to absorbance spectra for further feature extraction.

Eighty lamb samples are then split into the calibration set (75%) and the validation set (25%), where prediction models are learnt from the calibration set

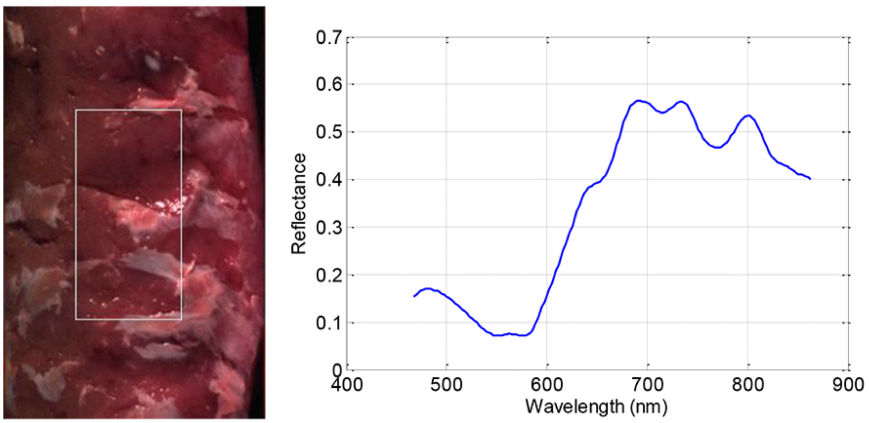


Fig. 6.6 HSI median reflectance spectrum extraction process for lamb. (a) Pseudo colour image of one lamb sample, with ROI marked by the white frame, and (b) the median reflectance spectrum of the ROI after the iterative pixel removal process.

and verified on the validation set. In this way, the ability of HSI for predicting quality of unknown lamb samples could be tested.

6.3 Results and discussion

In this section, results of four objectives mentioned in Section 6.1 are shown. For all experiments, the performance of prediction is evaluated by the coefficient of determination (R^2), root mean squared error (RMSE) as well as the ratio of performance deviation (RPD). Equations are given below,

$$R^2 = 1 - \frac{\sum_{k=1}^{n} (y_k - f_k)^2}{\sum_{k=1}^{n} (y_k - \bar{y})^2}, \tag{6.4}$$

$$RPD = \frac{SD}{RMSE} = \frac{SD}{\sqrt{\frac{\sum_{k=1}^{n} (y_k - f_k)^2}{n}}}, \tag{6.5}$$

where y_k is the measured quality trait value, f_k is the predicted value, $\bar{y}$ is the mean of measured trait value and n is the number of samples. The coefficient of determination ranges from 0 to 1, where 0 represents a poor correlation between predicted trait values and measured trait values and 1 stands for a high degree of correlation. For RPD, values higher than 2 indicates that the calibration model is

suitable to use and values higher than 3 indicate a model adequate for analytical purposes [20].

6.3.1 SVM for predicting beef quality based on VIS-NIR spectroscopy

The ability of SVM for predicting quality parameters, including surface colour, ultimate pH and SSF, using VIS-NIR spectroscopy, is tested on data set A. Results are given in this section, as well as results from PLSR for comparison. The influence of animal gender on beef quality prediction is also investigated.

Table 6.1 shows the mean, range and SD of measured values of quality parameters analysed in data set A for the calibration and validation sets, respectively. SVM models are developed and cross-validated on the calibration set and tested on the validation set without carrying out any spectral pre-treatment on the data.

Table 6.1 Summary statistics of all quality attributes for 234 carcasses in data set A.

Trait	Calibration set				Validation set			
	n	Mean	Range	SD	n	Mean	Range	SD
L*	175	37.38	30.64 – 45.39	2.54	59	37.30	27.65 – 44.32	2.78
a*	175	24.55	18.06 – 32.05	2.41	59	24.46	14.88 – 31.40	2.67
b*	175	9.05	3.70 – 13.99	1.96	59	9.01	2.74 – 13.95	2.06
pH_u	175	5.52	5.20 – 6.74	0.23	59	5.52	5.18 – 6.49	0.23
SSF	175	123.37	72.77 – 329.39	42.46	59	123.61	69.83 – 312.00	44.75

In Table 6.2 and Table 6.3, results from the calibration set and the validation set using both SVM and PLSR approaches are given for comparison. The number of principal components used for SVM and the number of partial least squares components used for PLSR are denoted as PC and PLSC in tables. It can be seen that, for calibration results, SVM always shows a better performance than PLSR, where $R^2{}_{val}$ for SSF is more than double of that with PLSR.

Table 6.2 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set A, using SVM and PLSR in the calibration set with noise-removed spectra (495 – 1600 nm).

Trait	Calibration using SVM			Calibration using PLSR		
	PC	R^2_{cal}	$RMSE_{cal}$	PLSC	R^2_{cal}	$RMSE_{cal}$
L*	10	0.83	1.05	11	0.83	1.04
a*	45	0.70	1.34	10	0.55	1.60
b*	25	0.67	1.13	10	0.64	1.17
pH_u	50	0.90	0.07	13	0.77	0.11
SSF	10	0.89	14.66	5	0.16	38.79

Table 6.3 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set A, using SVM and PLSR in the validation set with noise-removed spectra (495 – 1600 nm).

Trait	Validation using SVM				Validation using PLSR			
	PC	$\mathbf{R^2_{val}}$	$\mathbf{RMSE_{val}}$	$\mathbf{RPD_{val}}$	**PLSC**	$\mathbf{R^2_{val}}$	$\mathbf{RMSE_{val}}$	$\mathbf{RPD_{val}}$
L*	10	0.80	1.27	2.19	11	0.76	1.37	2.03
a*	45	0.64	1.71	1.56	10	0.60	1.71	1.56
b*	25	0.54	1.42	1.45	10	0.53	1.43	1.44
pH_u	50	0.74	0.12	1.92	13	0.67	0.13	1.77
SSF	10	0.20	42.40	1.06	5	0.08	43.39	1.03

The prediction of quality attributes in beef *M. longissimus thoracis* using VIS-NIR spectroscopy by other researchers are shown in Table 6.4. Results obtained by SVM in the current experiment are similar to (or even higher than) those in Table 6.4 for predicting L*, which means the lightness L* can be successfully predicted by SVM. For the prediction of a* and b*, results achieved by SVM have an RPD value between 1.45 and 1.56, which are much higher than those from [98, 99]. For the prediction of pH_u, R^2 and RPD in the present study are only lower than those presented in [98], probably due to the fact that the research in [98] was conducted on 30 bulls, and meat from young bulls commonly has high ultimate pH [97]; thus a greater number of high pH readings in a data set aid in the development of a robust calibration model.

Most studies using VIS-NIR spectroscopy to assess beef tenderness adopt WBSF as the texture benchmark, while few published reports have attempted to predict SSF in beef. Results from SVM in the present study are quite low compared with those from WBSF-based studies, possibly due to the tight quality specification of the commercial cattle used. The calibration SD in the current data set is only 16.8% of the difference between maximum and minimum values of the SSF parameter, indicating that the data set is lacking a sufficiently large variation, which could be a possible reason for poor prediction.

As shown in [98], prediction results obtained in young bulls seem to be quite favourable. Therefore, the prediction performance with different genders is investigated in the present study using SVM as well. The splitting method for calibration and validation data sets is the same as introduced previously, and the associated mean, data range and SD of measured values for all parameters analysed in both sets are compared in Table 6.5 – 6.7, respectively. Despite occasional large differences in terms of the SD measurement caused by outliers, data used for calibration and validation in each gender are statistically consistent in terms of the mean and the data range covered. Moreover, the data distribution

Table 6.4 Publications for predicting instrumental meat quality in beef *M. longissimus thoracis* using VIS-NIR spectroscopy.

Sex	Age (months)	Weight (kg)	Trait	R^2_{val}	RPD_{val}
			Reference [98]		
Bull	9 – 11	90 – 150	L*t0 (No blooming)	0.80	2.22
			L*t60 (Blooming for 1 h)	0.75	2.07
			a*t0 (No blooming)	0.23	1.14
			a*t60 (Blooming for 1 h)	0.29	0.90
			b*t0 (No blooming)	0.27	1.17
			b*t60 (Blooming for 1 h)	0.46	1.37
			pH_u	0.91	3.17
			Reference [99]		
Steer	53 ± 0.5	813 ± 10.9	L*	0.59	1.24
			a*	0.01	0.98
			b*	0.35	1.16
			pH_u	0.41	1.12
Sex N/A	11.6 ± 0.2	442.5 ± 8.55	L*	0.87	2.17
			a*	0.71	1.58
			b*	0.90	2.51
			pH_u	0.47	1.26
			Reference [100]		
Steer	18 – 24	190 – 220	pH_u	0.83	N/A
			Reference [109]		
Steer and heifer	19.5 ± 0.9	335 ± 30.3	SSF3 (3 days post mortem)	0.31	1.25
			SSF14 (14 days post mortem)	0.23	1.14

for heifers is very similar to that for steers, although young bulls appear to have lower values in a*, b* and L*. In addition, mean ultimate pH values for three genders are very close, whereas the associated mean SF values are quite different.

Table 6.5 Summary statistics of quality attributes for 75 heifers in data set A.

Trait	Calibration set for heifers				Validation set for heifers			
	n	Mean	Range	SD	n	Mean	Range	SD
L*	56	37.68	32.61 – 45.39	2.35	19	37.36	30.64 – 42.89	2.67
a*	56	25.15	19.11 – 31.40	2.23	19	24.84	18.06 – 29.29	2.49
b*	56	9.50	5.24 – 13.54	1.66	19	9.35	5.18 – 12.52	1.79
pH_u	56	5.46	5.23 – 5.87	0.12	19	5.44	5.20 – 5.63	0.10
SSF	56	122.59	77.56 – 312.00	43.14	19	119.62	77.50 – 223.88	38.03

Table 6.6 Summary statistics of quality attributes for 118 steers in data set A.

Trait	Calibration set for steers				Validation set for steers			
	n	Mean	Range	SD	n	Mean	Range	SD
L*	88	37.69	32.22 – 44.32	2.35	30	37.64	32.11 – 43.91	2.53
a*	88	24.89	20.05 – 31.03	2.20	30	24.82	19.25 – 29.75	2.33
b*	88	9.29	4.97 – 13.99	1.85	30	9.24	3.80 – 13.95	2.09
pH_u	88	5.50	5.20 – 6.37	0.19	30	5.50	5.18 – 6.15	0.19
SSF	88	114.25	69.83 – 258.85	33.10	30	114.67	69.83 – 258.85	37.17

Table 6.7 Summary statistics of quality attributes for 41 young bulls in data set A.

Trait	Calibration set for young bulls				Validation set for young bulls			
	n	Mean	Range	SD	n	Mean	Range	SD
L*	30	36.07	30.76 – 42.44	2.78	11	35.93	27.65 – 43.03	4.03
a*	30	22.48	18.73 – 24.90	1.77	11	22.73	14.88 – 32.05	4.11
b*	30	7.51	3.70 – 10.49	1.81	11	7.75	2.74 – 13.71	2.79
pH_u	30	5.65	5.28 – 6.49	0.35	11	5.70	5.24 – 6.74	0.47
SSF	30	150.91	86.82 – 273.03	49.06	11	156.64	72.77 – 329.39	71.70

Prediction results using SVM on three noise-removed data sets with different genders are shown in Table 6.8 – 6.10. For colour parameters and ultimate pH, results agree with those in Table 6.2 and Table 6.3. However, the variation in these results for SSF is high. First, the calibration SSF accuracy for heifers is the best and achieves over 0.96 in R^2 measurement, followed by 0.72 for steers, whereas the result for young bulls is less than 0.3 due to large variations with the particular data set. Second, the highest validation R^2 of SSF achieved is approximately 0.27 from heifers, which is quite close to results from [109]. Results obtained from steers are low ($R^2{}_{val} = 0.06$), even though the calibrated R^2 is over 0.7. In contrast, the validation R^2 for young bulls achieves 0.15, although the calibration R^2 is less than 0.3. In addition, comparing RPD from three gender groups, it is noticed that the prediction performance of young bull quality traits is much better than those of heifer and steer, especially in predicting colour parameters including L*, a* and b*, which corroborates high prediction accuracies obtained in bulls reported in [98].

Table 6.8 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in heifer of data set A using SVM with noise-removed spectra (495 – 1600 nm).

Trait	PC	Calibration for heifers		Validation for heifers		
		R^2_{cal}	$RMSE_{cal}$	R^2_{val}	$RMSE_{val}$	RPD_{val}
L*	20	0.78	1.12	0.74	1.46	1.83
a*	35	0.73	1.21	0.29	2.21	1.13
b*	15	0.52	1.34	0.45	1.39	1.29
pH_u	35	0.64	0.08	0.47	0.08	1.25
SSF	2	0.96	9.85	0.27	40.91	0.93

6.3.2 Comparison between VIS-NIR spectroscopy and HSI in predicting beef eating quality

The ability of VIS-NIR spectroscopy and HSI in predicting beef eating quality is compared on data set B. Statistics of the calibration set and the validation set are

Table 6.9 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in steer of data set A using SVM with noise-removed spectra (495 – 1600 nm).

Trait	PC	Calibration for steers		Validation for steers		
		R^2_{cal}	$RMSE_{cal}$	R^2_{val}	$RMSE_{val}$	RPD_{val}
L*	25	0.87	0.92	0.89	1.01	2.50
a*	35	0.73	1.15	0.59	1.46	1.60
b*	30	0.75	0.96	0.50	1.47	1.42
pH_u	20	0.90	0.06	0.59	0.12	1.58
SSF	2	0.72	22.58	0.06	40.35	0.92

Table 6.10 Performance of VIS-NIR spectroscopy for predicting instrumental meat quality in young bull of data set A using SVM with noise-removed spectra (495 – 1600 nm).

Trait	PC	Calibration for young bulls		Validation for young bulls		
		R^2_{cal}	$RMSE_{cal}$	R^2_{val}	$RMSE_{val}$	RPD_{val}
L*	4	0.81	0.21	0.94	0.98	4.11
a*	25	0.98	0.30	0.63	2.57	1.60
b*	4	0.51	1.27	0.54	1.97	1.42
pH_u	10	0.63	0.22	0.71	0.29	1.62
SSF	2	0.30	41.22	0.15	63.92	1.12

given in Table 6.11. As can be seen, both sets have similar mean, SD and range for four quality traits.

Table 6.11 Summary statistics of all quality attributes for 858 carcasses in data set B.

Trait	Calibration set				Validation set			
	n	Mean	Range	SD	n	Mean	Range	SD
SSF7	644	139.57	46.97 – 369.21	46.53	214	139.54	68.71 – 304.80	45.74
SSF14	644	141.04	60.77 – 316.40	43.93	214	141.05	68.13 – 297.51	43.53
pH7	644	5.58	5.34 – 6.37	0.11	214	5.58	5.42 – 6.25	0.10
pH14	644	5.63	5.43 – 6.46	0.11	214	5.63	5.46 – 6.33	0.11

Prediction results for VIS-NIR spectroscopy and HSI for both calibration and validation set are shown in Table 6.12 and Table 6.13 respectively. The number of principal components is tested from 5 to 50 in a step of 5, and the best result is given here. Comparing values of R^2 in Table 6.12 and Table 6.13, for the majority of parameters, HSI offers better prediction performance than VIS-NIR spectroscopy. Even though for SSF14, VIS-NIR spectroscopy yields slightly higher R^2 than HSI, its RPD is still lower than that of HSI.

Results shown here seem not quite robust since the current research aims to develop a tool on the production line with 2-day post mortem beef to predict eating quality at 7 days and 14 days post mortem, which may result in lower prediction accuracy. During the ageing process, meat quality parameters could be affected by biochemical and structural changes. Even though that is the case,

Table 6.12 Performance for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set B, using VIS-NIR spectroscopy and HSI in the calibration set, with noise removed spectra.

Trait	Calibration with VIS-NIR spectra			Calibration with HSI spectra		
	PC	R^2_{cal}	$RMSE_{cal}$	PC	R^2_{cal}	$RMSE_{cal}$
SSF7	25	0.35	39.14	45	0.39	37.80
SSF14	25	0.47	34.91	50	0.59	30.61
pH7	20	0.59	0.07	40	0.74	0.18
pH14	5	0.42	0.09	45	0.70	0.06

Table 6.13 Performance for predicting instrumental meat quality in beef *M. longissimus thoracis* of data set B, using VIS-NIR spectroscopy and HSI in the validation set, with noise removed spectra.

Trait	Validation with VIS-NIR spectra				Validation with HSI spectra			
	PC	R^2_{val}	$RMSE_{val}$	RPD_{val}	PC	R^2_{val}	$RMSE_{val}$	RPD_{val}
SSF7	25	0.10	43.86	1.04	45	0.11	43.37	1.05
SSF14	25	0.19	39.76	1.09	50	0.18	39.56	1.10
pH7	20	0.35	0.08	1.25	40	0.43	0.08	1.25
pH14	5	0.36	0.09	1.22	45	0.44	0.08	1.38

preliminary results do prove that HSI is more promising in predicting beef eating quality than VIS-NIR spectroscopy. One thing worth of noting is that the HSI system used in the experiment is in visible range, which means that prediction results might be higher if an NIR HSI system is used.

6.3.3 SSA for improving HSI-based beef eating quality evaluation

Data set C is a subset of data set B and it is used for testing the performance of SSA as a pre-processing step for hyperspectral images. Statistics of data set C are shown in Table 6.14. Some samples have been exempted from the experiment for predicting ultimate pH because of missing value, resulting in different sample numbers for SSF and pH.

Table 6.14 Summary statistics of all quality attributes for 211 carcasses in data set C.

Trait	Calibration set				Validation set			
	n	Mean	Range	SD	n	Mean	Range	SD
SSF7	159	131.46	46.97 – 299.54	48.18	52	130.73	69.41 – 285.62	45.69
SSF14	159	132.23	63.35 – 291.56	42.91	52	131.32	73.61 – 239.82	39.91
pH7	154	5.63	5.44 – 6.37	0.13	51	5.63	5.46 – 6.34	0.14
pH14	154	5.69	5.46 – 6.46	0.14	51	5.69	5.48 – 6.41	0.14

By trial and error, the number of principal components is set to 30. In that case, the data dimensionality has been reduced dramatically while enough

features have been preserved. The prediction performance has been focused on the external validation set and results comparing original spectra and SSA processed spectra for the validation set are shown in Table 6.15. The window length L is tuned to achieve the best performance for each quality trait. As highlighted in bold, except for limited improvement on SSF14, SSA could always stand out prominently, which demonstrates that it is an excellent feature extraction and de-noising pre-processing procedure for HSI spectral profiles.

Table 6.15 Performance comparison of original HSI spectra and SSA processed spectra in data set C for predicting beef eating quality attributes.

Trait	Original HSI spectra		SSA processed HSI spectra		
	$\mathbf{R^2_{val}}$	$\mathbf{RPD_{val}}$	**L**	$\mathbf{R^2_{val}}$	$\mathbf{RPD_{val}}$
SSF7	0.19	1.10	2	**0.33**	**1.21**
SSF14	0.10	1.03	2	0.11	1.02
pH7	0.42	1.25	3	**0.45**	**1.28**
pH14	0.28	1.12	7	**0.34**	**1.21**

If both the number of principal components and the window length L of SSA are adjusted by conducting a grid search, the regression performance with the validation set can be improved even further. Chosen parameters and results for SSA processed HSI spectra are shown in Table 6.16.

Table 6.16 Best performance of SSA processed HSI spectra in data set C for predicting beef eating quality attributes.

Trait	PC	L	$\mathbf{R^2_{val}}$	$\mathbf{RPD_{val}}$
SSF7	30	2	0.33	1.21
SSF14	20	6	0.20	1.10
pH7	45	2	0.58	1.43
pH14	45	3	0.49	1.36

As can be seen, for predicting ultimate pH values, both R^2 and RPD indicate that HSI could be a promising tool for the meat industry. However, results for predicting SSF are not very satisfying, which might be due to the fact that SSF is not a standard quantitative metric of meat tenderness, regardless of inconsistent measurements within the ground truth. Similar results have been achieved by other researchers as well. Referring to Table 6.4, SSF has been predicted using VIS-NIR spectroscopy in [109], giving R^2 of 0.31 (RPD = 1.25) and 0.23 (RPD = 1.14) for SSF3 and SSF14 respectively. But, it should be noted that comparison between studies are not fully reliable due to different samples used. Nevertheless, the potential of HSI in providing additional information that could help to improve prediction of meat quality attributes should not be denied.

6.3.4 Prediction of lamb quality using HSI

Data set D contains hyperspectral images of 80 lamb samples and it is used for testing the ability of HSI to predict lamb quality attributes. Summary statistics for all quality attributes for data set D are given in Table 6.17, where SF represents MIRINZ SF for short.

Table 6.17 Summary statistics of all quality attributes for 80 lamb samples in data set D.

Trait	Calibration set				Validation set			
	n	Mean	Range	SD	n	Mean	Range	SD
pH_u	60	5.59	5.46 – 5.90	0.07	20	5.61	5.48 – 6.08	0.12
L*	60	41.47	36.69 – 46.64	1.83	20	41.83	38.74 – 49.63	2.31
a*	60	23.28	20.44 – 25.37	0.91	20	23.77	22.28 – 25.71	0.83
b*	60	7.77	5.89 – 9.15	0.58	20	7.85	7.02 – 9.17	0.57
SF	60	51.05	28.17 – 92.03	13.32	20	53.89	33.69 – 117.65	18.80

The number of principal components is tried from 5 to 50 in a step of 5, and both raw HSI absorbance spectra and SSA processed absorbance spectra are tested. Combinations of best prediction results are shown in Table 6.18, where 'A' stands for raw absorbance spectra and 'A + SSA' means SSA processed absorbance spectra.

Table 6.18 Performance of HSI for predicting instrumental lamb quality parameters of data set D.

Trait	Pre-treatment	PC	Calibration set		Validation set		
			R^2_{cal}	$RMSE_{cal}$	R^2_{val}	$RMSE_{val}$	RPD_{val}
pH_u	A	45	0.54	0.06	0.38	0.11	1.07
L*	A	15	0.83	0.67	0.77	1.34	1.72
a*	A + SSA	40	0.95	0.22	0.48	0.76	1.09
b*	A + SSA	15	0.83	0.34	0.26	0.50	1.13
SF	A + SSA	40	0.82	6.76	0.41	14.90	1.26

As the study is to test the ability of HSI in predicting unknown lamb quality for on-line use, results of the external validation set are particularly important. For the prediction of ultimate pH, R^2_{val} is higher than those reported in [201] using VIS-NIR spectroscopy (R^2_{val} = 0.03 – 0.19). For colour parameters and MIRINZ SF prediction, limited research has been found on lamb. Nevertheless, results achieved in the current study agree with those predicted with beef samples using VIS-NIR spectroscopy [20, 98, 99, 103, 213], whose average R^2_{val} values are 0.76 and 0.44 for L* and a*. However, result of b* in the experiment is poorer than that of others (R^2_{val} = 0.57), which may be due to variation between different samples. Similarly, the MIRINZ SF is compared with WBSF predicted by others

on beef samples. It is found that the average ${R^2}_{val}$ of their research is 0.30, which is lower than than the one in the current study (${R^2}_{val} = 0.41$). The RPD values of the resulting models did not give as much information as R^2. A possible reason is that all lamb samples are purchased from the abattoir fulfilling market standards. Therefore, the standard deviation of those quality traits is low and the range of values is also limited.

In the present study some colour parameters are found to be more correlated with the hyperspectral data, while ultimate pH and MIRINZ SF are less predictable. As mentioned previously, ultimate pH and SF contribute to tenderness, juiciness and flavour of lamb. However, as admitted by authors in [201], these complex quality parameters could be affected by many factors during production process, and their variation has multiple biological causes, which also lowers the prediction performance with HSI.

6.4 Summary

In this chapter, firstly, VIS-NIR spectroscopy (350 – 1800 nm) is investigated for predicting beef quality parameters, including colour, ultimate pH and SSF. Due to the relatively high dimensionality of VIS-NIR spectra, PCA is applied for feature extraction and data reduction, followed by SVM-based modelling and prediction. Applying constructed models to external validation data sets, predictions of L*, a*, b*, pH_u and SSF in terms of R^2 are obtained as 0.80, 0.64, 0.54, 0.74 and 0.20 on the validation set, respectively, where the corresponding RPD values are 2.19, 1.56, 1.45, 1.92 and 1.06. Although the predicted R^2 values for SSF are quite low in comparison with other published data, overall, results achieved by SVM are higher than those achieved by PLSR. It indicates that the prediction is strongly dependent on the data acquired, but SVM can play an important role in quantitatively predicting beef quality with VIS-NIR spectroscopy. Besides, it is found that quality traits of young bulls can be better predicted than those from heifers and steers, especially for the lightness parameter L*, which indicates great potential for improved quality prediction using additional clues such as carcass gender.

Secondly, this chapter compares the ability of VIS-NIR spectroscopy (350 nm – 2500 nm) and HSI (400 nm – 862.90 nm) for predicting beef eating quality, including SSF and pH at 7 days and 14 days post mortem. Results suggest that compared with VIS-NIR spectroscopy, HSI may offer more additional information that could

help to improve prediction of beef quality attributes, with an improvement for both R^2 and RPD in most cases.

Thirdly, in this chapter, a relatively new time series analysis method SSA is proposed to pre-process the HSI spectra. By decomposing the original spectra and reconstructing the spectra using the most significant components, SSA demonstrates its ability in removing noise and improving the prediction accuracy for HSI-based beef eating quality evaluation.

Last but not least, this chapter also employs the same HSI system to predict quality parameters for lamb samples because of its superiority demonstrated previously. The study has shown that HSI offers an alternative choice for non-destructive measurement of lamb quality traits. Even though the prediction performance for some complex quality traits is low and requires further improvement, HSI is still proved to be superior to conventional VIS-NIR spectroscopy by comparing with other studies. Most importantly, this part has provided a benchmarking baseline for other researchers, since limited research has been conducted in the area of HSI-based lamb quality prediction.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The general objective of this thesis is to develop advanced spectral-spatial techniques for effective analysis of hyperspectral images. These include various approaches for image compression, feature extraction and data classification as presented in Chapters 3, 4 and 5, respectively. The last main chapter focuses on the practical applications of the HSI technology, i.e. evaluation of meat eating quality. The main contributions of the thesis are summarised in detail as follows.

1. In Chapter 3, an improved three-dimensional (3D) discrete cosine transform (DCT)-based approach is proposed for compressing the hyperspectral image, making possible for faster transmission and easy storage of the data while not degrading the performance of following data analysis. Using the quantisation table adapted from video compression, support vector machine (SVM)-based machine learning is applied for optimally predicting the inter-band correlation parameter in the quantisation table, which is possible due to the regular pattern presented in highly correlated spectral bands. The proposed method is compared with another state-of-the-art compression technique named as 3D set partitioning in hierarchical trees (SPIHT), which is based on 3D discrete wavelet transform (DWT). Comprehensive results show that the proposed 3D DCT approach outperforms 3D SPIHT under both subjective and objective evaluation metrics. Besides, a quality-assured evaluation criterion using classification accuracy is tested on reconstructed hyperspectral images from the two 3D compression approaches. The proposed method is proved to be more suitable as the degradation on reconstructed image quality is negligible with a bit rate of no less than 0.8 bpppb.

2. Chapter 4 presents a de-noising and feature extraction technique for hyperspectral images, which can be treated as a pre-processing step before hyperspectral image classification for improving the classification accuracy. Two well known techniques, the curvelet transform and singular spectrum analysis (SSA), are combined together for this purpose. Both as the multi-scale geometric analysis, the curvelet transform provides a better representation of curves or edges in the image than the wavelet transform, leading to a better separation between geometric details and background noise. By transforming the hyperspectral image into the curvelet domain, noise could be effectively reduced through the decomposition and reconstruction processes of SSA in the spectral dimension. Compared with several typical state-of-the-art spectral feature extraction techniques, the proposed methodology proves its superiority with the best classification performance. With a further spatial post-processing step added, the classification framework with the proposed spectral feature extraction technique is either comparable or better than some recent spectral-spatial classification methods considering the classification accuracy.

3. In Chapter 5, a novel classification framework for hyperspectral images based on the joint bilateral filter and sparse representation classification (SRC) is proposed. Employing the first principal component as the guidance image for the joint bilateral filter, spatial features can be extracted without blurring the edge information and the image quality is enhanced in bands with low contrast. For this reason, the performance of joint bilateral filter is proven to be better than that of the normal bilateral filter. Besides, the spectral similarity-based joint SRC (SS-JSRC) is proposed to overcome the weakness of the traditional JSRC method. By combining the joint bilateral filtering and SS-JSRC together, the superiority of the proposed classification framework is demonstrated by comparing with several typical spectral-spatial classification approaches, with higher results in both classification accuracy and kappa coefficient.

4. Chapter 6 focuses on investigating the feasibility of the hyperspectral imaging (HSI) system for predicting meat eating quality. The study starts with a similar system, the visible near-infrared (VIS-NIR) spectroscopy, to determine the suitable analysis technique for multivariate data. Although the prediction is strongly dependent on the acquired data, it is found SVM is able to achieve better prediction performance than the commonly used

partial least squares regression (PLSR). Therefore, SVM is chosen as the data prediction technique through the study. Besides, the study also compares the performance of VIS-NIR spectroscopy and HSI for predicting beef eating quality. Results show that HSI is more promising in this context with the help of the spatial information presented in the data. Additionally, the aforementioned feature extraction technique SSA is also applied to the beef HSI data. By effectively removing the noise presented in the data set using SSA, the prediction accuracy for HSI-based beef quality evaluation can be improved further more. Last, the same HSI technique is tested on lamb samples for quality evaluation. Since limited research has been conducted in the area of HSI-based quality prediction, the last part of the study offers a useful comparison reference for other researchers.

7.2 Future work

Followed by the conclusions of those innovative techniques proposed for hyperspectral image analysis in this thesis, the potential directions for future research are pointed out in this section.

1. Hyperspectral image compression has always been an active research area, and researchers have proposed various compression approaches successfully. However, in particular for hyperspectral images, there is a lack of effective evaluation criteria for reconstructed data. Future work can focus on applying combination of spatial and spectral information for saliency-enabled object-based performance assessment of reconstructed hyperspectral images, especially for quality-assured analysis. Besides, 3D DCT is achieved by applying 1D DCT sequentially in different dimensions. After applying the first 1D DCT in one dimension, the energy is contracted and useful information is located in the low frequency area. By only applying the second 1D DCT in the low frequency area, followed by the third 1D DCT applied in the same way, the whole transform speeds up. Therefore, a more efficient 3D DCT algorithm could be designed in the future.

2. The proposed curvelet-based feature extraction technique only employs one-dimensional (1D) SSA for de-noising in the spectral domain. In the future, two-dimensional (2D) SSA can be applied to the spatial domain of the curvelet-transformed coefficients as 2D SSA is proven to be a more effective feature extraction technique than 1D SSA recently. Furthermore, a

simple square window is used for post-processing in the experiments, while a shape-adaptive window could be designed for better performance.

3. For the proposed SS-JSRC method, the optimal classification is achieved by looking for the minimum reconstruction residue. Some researchers have found that histograms of sparse codes can be used for successful object detection [214], where it is possible to apply this new feature for hyperspectral image classification as well.

4. In the study of HSI-based meat quality assessment, the HSI system used is in the visible range due to the limit of available experimental instrument. For future work, the performance of near-infrared HSI systems should be examined since many chemical bonds related to meat quality parameters are located in the near-infrared wavelength range. Additionally, meat sample in the current study are bought from commercial abattoirs with high quality standards, which means the range of quality reference is limited. In order to build a more robust regression model, a wider range of reference data should be used in the future study.

References

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002.

[2] E. Christophe, C. Mailhes, and P. Duhamel, "Hyperspectral image compression: adapting SPIHT and EZW to anisotropic 3-D wavelet coding," *IEEE Transactions on Image Processing*, vol. 17, no. 12, pp. 2334–2346, 2008.

[3] C. Zhao, X. Li, J. Ren, and S. Marshall, "Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery," *International Journal of Remote Sensing*, vol. 34, no. 24, pp. 8669–8684, 2013.

[4] D.-W. Sun, *Hyperspectral Imaging for Food Quality Analysis and Control*. Elsevier, 2010.

[5] T. Kelman, J. Ren, and S. Marshall, "Effective classification of Chinese tea samples in hyperspectral imaging," *Artificial Intelligence Research*, vol. 2, no. 4, pp. 87–96, 2013.

[6] Y. Roggo, A. Edmond, P. Chalus, and M. Ulmschneider, "Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms," *Analytica Chimica Acta*, vol. 535, no. 1-2, pp. 79–87, 2005.

[7] G. Payne, N. Langlois, C. Lennard, and C. Roux, "Applying visible hyperspectral (chemical) imaging to estimate the age of bruises," *Medicine, Science and the Law*, vol. 47, no. 3, pp. 225–232, 2007.

[8] L. Brewer, J. Ohlhausen, P. Kotula, and J. Michael, "Forensic analysis of bioagents by X-ray and TOF-SIMS hyperspectral imaging," *Forensic Science International*, vol. 179, no. 2-3, pp. 98–106, 2008.

[9] K. Gill, J. Ren, S. Marshall, S. Karthick, and J. Gilchrist, "Quality-assured fingerprint image enhancement and extraction using hyperspectral imaging," in *Proceedings of Fourth International Conference on Image for Crime Detection and Prevention*, London, UK, 2011.

[10] AVIRIS image Indian Pines. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec

[11] F. Vagni, "Survey of hyperspectral and multispectral imaging technologies," North Atlantic Treaty Organisation, Tech. Rep., 2007.

[12] J. Fowler and J. Rucker, "Three-dimensional wavelet-based compression of hyperspectral imagery," in *Hyperspectral Data Exploitation: Theory and Applications*. Wiley-Interscience, 2007, pp. 379–407.

[13] E. Christophe, "Hyperspectral data compression tradeoff," in *Optical Remote Sensing*. Springer, 2011, pp. 9–29.

[14] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Transform coding techniques for lossy hyperspectral data compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1408–1421, 2007.

[15] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. S110–S122, 2009.

[16] Z. He, Q. Wang, Y. Shen, and M. Sun, "Kernel sparse multitask learning for hyperspectral image classification with empirical mode decomposition and morphological wavelet-based features," *IEEE Transactions on Geosecience and Remote Sensing*, vol. 52, no. 8, pp. 5150–5163, 2014.

[17] G. P. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[18] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[19] E. Z. Panagou, O. Papadopoulou, J. M. Carstensen, and G.-J. E. Nychas, "Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage," *International Journal of Food Microbiology*, vol. 174, pp. 1–11, 2014.

[20] M. Kamruzzaman, G. ElMasry, D.-W. Sun, and P. Allen, "Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis," *Analytica Chimica Acta*, vol. 714, pp. 57–67, 2012.

[21] E. Magli, "Multiband lossless compression of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1168–1178, 2009.

[22] Y. Liang, J. Li, and K. Guo, "Lossless compression of hyperspectral images using hybrid context prediction," *Optics Express*, vol. 20, no. 7, pp. 8199–8206, 2012.

[23] J. Zhang, H. Li, and C. W. Chen, "Distributed lossless coding techniques for hyperspectral images," *IEEE Jounal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 977–989, 2015.

[24] J. Wu, W. Kong, J. Mielikainen, and B. Huang, "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2194–2198, 2015.

[25] S.-E. Qian, M. Bergeron, I. Cunningham, L. Gagnon, and A. Hollinger, "Near lossless data compression onboard a hyperspectral satellite," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 42, no. 3, pp. 851–866, 2006.

[26] E. Christophe, D. Léger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2103–2114, 2005.

[27] D. Valsesia and E. Magli, "A novel rate control algorithm for onboard predictive coding of multispectral and hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6341–6355, 2014.

[28] R. Roger and M. C. Cavenor, "Lossless compression of AVIRIS images," *IEEE Transactions on Image Processing*, vol. 5, no. 5, pp. 713–719, 1996.

[29] J. Mielikainen and P. Toivanen, "Clustered DPCM for the lossless compression of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2943–2946, 2003.

[30] J. Mielikainen and B. Huang, "Lossless compression of hyperspectral images using clustered linear prediction with adaptive prediction length," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1118–1121, 2012.

[31] A. Abrardo, M. Barni, E. Magli, and F. Nencini, "Error-resilient and low-complexity onboard lossless compression of hyperspectral images by means of distributed source coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1892–1904, 2010.

[32] J. Zhang and G. Liu, "An efficient reordering prediction-based lossless compression algorithm for hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 283–287, 2007.

[33] C. Huo, R. Zhang, and T. Peng, "Lossless compression of hyperspectral images based on searching optimal multibands for prediction," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 339–343, 2009.

[34] J. Mielikainen and P. Toivanen, "Lossless compression of hyperspectral images using a quantized index to lookup tables," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 474–478, 2008.

[35] X. Wu and N. Memon, "Context-based lossless interband compression-extending CALIC," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 994–1001, 2000.

[36] M. Mendenhall and E. Merényi, "Relevance-based feature extraction for hyperspectral images," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 658–672, 2008.

[37] S.-E. Qian, A. Hollinger, M. Dutkiewicz, H. Tsang, H. Zwick, and J. Freemantle, "Effect of lossy vector quantization hyperspectral data compression on retrieval of red-edge indices," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1459–1470, 2001.

[38] S.-E. Qian, "Fast vector quantization algorithms based on nearest partition set search," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2422–2430, 2006.

[39] X. Li, J. Ren, C. Zhao, T. Qiao, and S. Marshall, "Novel multivariate vector quantization for effective compression of hyperspectral imagery," *Optics Communications*, vol. 332, pp. 192–200, 2014.

[40] B.-J. Kim, Z. Xiong, and W. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-d spiht)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374–1387, 2000.

[41] W. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient, low-complexity image coding with a set-partitioning embedded block coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 11, pp. 1219–1235, 2004.

[42] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[43] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Springer Science & Business Media, 2012, vol. 159.

[44] C.-C. Chen, "On the selection of image compression algorithms," in *Proceedings of Fourteenth International Conference on Pattern Recognition*, vol. 2. IEEE, 1998, pp. 1500–1504.

[45] G. Canta and G. Poggi, "Kronecker-product gain-shape vector quantization for multispectral and hyperspectral image coding," *IEEE Transactions on Image Processing*, vol. 7, no. 5, pp. 668–678, 1998.

[46] M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Springer Science & Business Media, 2002.

[47] G. P. Abousleman, M. W. Marcellin, and B. R. Hunt, "Compression of hyperspectral imagery using the 3-D DCT and hybrid DPCM/DCT," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 1, pp. 26–34, 1995.

[48] X. Pan, R. Liu, and X. Lv, "Low-complexity compression method for hyperspectral images based on distributed source coding," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 2, pp. 224–227, 2012.

[49] M. Pickering and M. Ryan, "Efficient spatial-spectral compression of hyperspectral data: Analysis of hyperspectral image data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1536–1539, 2001.

[50] A. Bilgin, G. Zweig, and M. W. Marcellin, "Three-dimensional image compression with integer wavelet transforms," *Applied Optics*, vol. 39, no. 11, pp. 1799–1814, 2000.

[51] X. Tang and W. Pearlman, "Three-dimensional wavelet-based compression of hyperspectral images," in *Hyperspectral Data Compression*. Springer, 2006, pp. 273–308.

[52] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 93, no. 7, pp. 112–122, 2014.

[53] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging," *IEEE Signal Processing Magazine.*, vol. 31, no. 4, pp. 149–154, 2014.

[54] X. Jia, J. Richards *et al.*, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 538–542, 1999.

[55] J. Zabalza, J. Ren, J. Ren, Z. Liu, and S. Marshall, "Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging," *Applied Optics*, vol. 53, no. 20, pp. 4440–4449, 2014.

[56] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Transactions of Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.

[57] U. Amato, R. M. Cavalli, A. Palombo, S. Pignatti, and F. Santini, "Experimental approach to the selection of the components in the minimum noise fraction," *IEEE Transactions of Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 153–160, 2009.

[58] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 12, pp. 4173–4185, 2008.

[59] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for feature reduction in hyperspectrale images analysis," in *Proceedings of the Seventh Nordic Signal Processing Symposium.* Reykjavik, Iceland: IEEE, 2006, pp. 238–241.

[60] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *The Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2003.

[61] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 612–624, 2011.

[62] Y. Gu, Q. Wang, H. Wang, D. You, and Y. Zhang, "Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2739–2751, 2015.

[63] B.-C. Kuo and D. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.

[64] J. Zabalza, J. Ren, W. Zheng, S. Marshall, and J. Wang, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1886–1890, 2014.

[65] J. Zabalza, J. Ren, Z. Wang, H. Zhao, J. Wang, and S. Marshall, "Fast implementation of singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2845–2853, 2015.

[66] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, "On the computational complexity of the empirical mode decomposition algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 400, pp. 159–167, 2014.

[67] B. Demir and S. Ertürk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4071–4084, 2010.

[68] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[69] Z. He, Y. Shen, Q. Wang, and Y. Wang, "Optimized ensemble EMD-based spectral features for hyperspectral image classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1041–1056, 2014.

[70] B. Demir, S. Ertürk, and M. K. Güllü, "Hyperspectral image classification using denoising of intrinsic mode functions," *IEEE Geoscience and Remote Sensong Letters*, vol. 8, no. 2, pp. 220–224, 2011.

[71] J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4418–4433, 2015.

[72] M. Fauvel, J. A. Benediktsson, J. chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.

[73] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 542–546, 2011.

[74] O. Rajadell, P. Garcia-Sevilla, and F. Pla, "Spectral–spatial pixel characterization using Gabor filters for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 860–864, 2013.

[75] J. A. Richards and X. Jia, "Using suitable neighbors to augment the training set in hyperspectral maximum likelihood classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 774–777, 2008.

[76] S. Kuching, "The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis," *Journal of Computer Science*, vol. 3, no. 6, pp. 419–423, 2007.

[77] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 217–255, 2008.

[78] Q. Jackson, D. Landgrebe *et al.*, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2454–2463, 2002.

[79] D. L. Civco, "Artificial neural networks for land-cover classification and mapping," *International Journal of Geographical Information Science*, vol. 7, no. 2, pp. 173–186, 1993.

[80] H. Bischof and A. Leonardis, "Finding optimal neural networks for land use classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 1, pp. 337–341, 1998.

[81] L. Bruzzone and D. F. Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 1179–1184, 1999.

[82] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sensing of Environment*, vol. 86, no. 4, pp. 554–565, 2003.

[83] C.-C. Yang, S. O. Prasher, P. Enright, C. Madramootoo, M. Burgess, P. K. Goel, and I. Callum, "Application of decision tree technology for image classification using remote sensing data," *Agricultural Systems*, vol. 76, no. 3, pp. 1101–1117, 2003.

[84] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[85] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015.

[86] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[87] B. G. Osborne, "Near-infrared spectroscopy in food analysis," *Encyclopedia of Analytical Chemistry*, 2000.

[88] C. Starr, *Biology: Concepts and Applications.* Wadsworth Publishing Company, Belmont, California, 2005.

[89] C. A. Roberts, J. Workman, and J. B. Reeves, *Near-infrared Spectroscopy in Agriculture*. American Society of Agronomy, Madison, Wisconsin, 2004, vol. 44.

[90] N. Prieto, R. Roehe, P. Lavín, G. Batten, and S. Andrés, "Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review," *Meat Science*, vol. 83, no. 2, pp. 175–186, 2009.

[91] S. Anderson, "Determination of fat, moisture, and protein in meat and meat products by using the FOSS FoodScan near-infrared spectrophotometer with FOSS artificial neural network calibration model and associated database: collaborative study," *Journal of AOAC International*, vol. 90, no. 4, pp. 1073–1083, 2007.

[92] K. Thyholt and T. Isaksson, "Differentiation of frozen and unfrozen beef using near-infrared spectroscopy," *Journal of the Science of Food and Agriculture*, vol. 73, no. 4, pp. 525–532, 1997.

[93] H. Ding, R.-J. Xu, and D. K. O. Chan, "Identification of broiler chicken meat using a visible/near-infrared spectroscopic technique," *Journal of the Science of Food and Agriculture*, vol. 79, no. 11, pp. 1382–1388, 1999.

[94] H. Ding and R.-J. Xu, "Differentiation of beef and kangaroo meat by visible/near-infrared reflectance spectroscopy," *Journal of Food Science*, vol. 64, no. 5, pp. 814–817, 1999.

[95] ——, "Near-infrared spectroscopic technique for detection of beef hamburger adulteration," *Journal of Agricultural and Food Chemistry*, vol. 48, no. 6, pp. 2193–2198, 2000.

[96] L. E. Jeremiah and D. M. Phillips, "Evaluation of a probe for predicting beef tenderness," *Meat Science*, vol. 55, no. 4, pp. 493–502, 2000.

[97] C. R. Craigie, "An investiagetion into the use of video image analysis (VIA) and visible-near infrared (NIR) spectroscopy for carcase evaluation," Ph.D. dissertation, Massey University, 2012.

[98] S. Andrés, A. Silva, A. L. Soares-Pereira, C. Martins, A. Bruno-Soares, and I. Murray, "The use of visible and near infrared reflectance spectroscopy to predict beef *M. longissimus thoracis et lumborum* quality attributes," *Meat Science*, vol. 78, no. 3, pp. 217–224, 2008.

[99] N. Prieto, S. Andrés, F. J. Giráldez, A. Mantecón, and P. Lavín, "Ability of near infrared reflectance spectroscopy (NIRS) to estimate physical parameters of adult steers (oxen) and young cattle meat samples," *Meat Science*, vol. 79, no. 4, pp. 692–699, 2008.

[100] K. Rosenvold, E. Micklander, P. W. Hansen, R. Burling-Claridge, M. Challies, C. Devine, and M. North, "Temporal, biochemical and structural factors that influence beef quality measurement using near infrared spectroscopy," *Meat Science*, vol. 82, no. 3, pp. 379–388, 2009.

[101] G. ElMasry, D.-W. Sun, and P. Allen, "Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef," *Journal of Food Engineering*, vol. 110, no. 1, pp. 127–140, 2012.

[102] L. Fiems, S. De Campeneere, S. De Smet, G. Van de Voorde, J. Vanacker, and C. V. Boucqué, "Relationship between fat depots in carcasses of beef bulls and effect on meat colour and tenderness," *Meat Science*, vol. 56, no. 1, pp. 41–47, 2000.

[103] B. Leroy, S. Lambotte, O. Dotreppe, H. Lecocq, L. Istasse, and A. Clinquart, "Prediction of technological and organoleptic properties of beef *Longissimus thoracis* from near-infrared reflectance and transmission spectra," *Meat Science*, vol. 66, no. 1, pp. 45–54, 2004.

[104] Y. Liu, B. G. Lyon, W. R. Windham, C. E. Realini, T. D. D. Pringle, and S. Duckett, "Prediction of color, texture, and sensory characteristics of beef steaks by visible and near infrared reflectance spectroscopy. a feasibility study," *Meat Science*, vol. 65, no. 3, pp. 1107–1115, 2003.

[105] D. Cozzolino, N. Barlocco, A. Vadell, F. Ballesteros, and G. Gallieta, "The use of visible and near-infrared reflectance spectroscopy to predict colour on both intact and homogenised pork muscle," *LWT-Food Science and Technology*, vol. 36, no. 2, pp. 195–202, 2003.

[106] D. Cozzolino and I. Murray, "Effect of sample presentation and animal muscle species on the analysis of meat by near infrared reflectance spectroscopy," *Journal of Near Infrared Spectroscopy*, vol. 10, no. 1, pp. 37–44, 2002.

[107] S. R. Rust, D. M. Price, J. Subbiah, G. Kranzler, G. G. Hilton, D. L. Vanoverbeke, and J. B. Morgan, "Predicting beef tenderness using near-infrared spectroscopy," *Journal of Animal Science*, vol. 86, no. 1, p. 211, 2008.

[108] R. Rodbotten, B.-H. Mevik, and K. I. Hildrum, "Prediction and classification of tenderness in beef from non-invasive diode array detected NIR spectra," *Journal of Near Infrared Spectroscopy*, vol. 9, no. 3, pp. 199–210, 2001.

[109] N. Prieto, D. W. Ross, E. A. Navajas, G. R. Nute, R. I. Richardson, J. J. Hyslop, G. Simm, and R. Roehe, "On-line application of visible and near infrared reflectance spectroscopy to predict chemical–physical and sensory characteristics of beef quality," *Meat Science*, vol. 83, no. 1, pp. 96–103, 2009.

[110] B. Park, K. C. Lawrence, W. R. Windham, and D. P. Smith, "Performance of hyperspectral imaging system for poultry surface fecal contaminant detection," *Journal of Food Engineering*, vol. 75, no. 3, pp. 340–348, 2006.

[111] J. Wu, Y. Peng, Y. Li, W. Wang, J. Chen, and S. Dhakal, "Prediction of beef quality attributes using VIS/NIR hyperspectral scattering imaging technique," *Journal of Food Engineering*, vol. 109, no. 2, pp. 267–273, 2012.

[112] G. K. Naganathan, L. M. Grimes, J. Subbiah, C. R. Calkins, A. Samal, and G. E. Meyer, "Visible/near-infrared hyperspectral imaging for beef tenderness prediction," *Computers and Electronics in Agriculture*, vol. 64, no. 2, pp. 225–233, 2008.

[113] K. Cluff, G. K. Naganathan, J. Subbiah, R. Lu, C. R. Calkins, and A. Samal, "Optical scattering in beef steak to predict tenderness using hyperspectral

imaging in the VIS-NIR region," *Sensing and Instrumentation for Food Quality and Safety*, vol. 2, no. 3, pp. 189–196, 2008.

[114] G. ElMasry, D.-W. Sun, and P. Allen, "Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging," *Food Research International*, vol. 44, no. 9, pp. 2624–2633, 2011.

[115] K.-i. Kobayashi, Y. Matsui, Y. Maebuchi, T. Toyota, and S. Nakauchi, "Near infrared spectroscopy and hyperspectral imaging for prediction and visualisation of fat and fatty acid content in intact raw beef cuts," *Journal of Near Infrared Spectroscopy*, vol. 18, no. 5, p. 301, 2010.

[116] J. Qiao, N. Wang, M. Ngadi, A. Gunenc, M. Monroy, C. Gariepy, and S. Prasher, "Prediction of drip-loss, pH, and color for pork using a hyperspectral imaging technique," *Meat Science*, vol. 76, no. 1, pp. 1–8, 2007.

[117] D. F. Barbin, G. ElMasry, D.-W. Sun, and P. Allen, "Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging," *Analytica Chimica Acta*, vol. 719, pp. 30–42, 2012.

[118] J. Qiao, M. O. Ngadi, N. Wang, C. Gariépy, and S. O. Prasher, "Pork quality and marbling level assessment using a hyperspectral imaging system," *Journal of Food Engineering*, vol. 83, no. 1, pp. 10–16, 2007.

[119] L. Liu, M. Ngadi, S. Prasher, and C. Gariépy, "Categorization of pork quality using gabor filter-based hyperspectral imaging technology," *Journal of Food Engineering*, vol. 99, no. 3, pp. 284–293, 2010.

[120] M. Kamruzzaman, G. ElMasry, D.-W. Sun, and P. Allen, "Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression," *Innovative Food Science & Emerging Technologies*, vol. 16, pp. 218–226, 2012.

[121] ——, "Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging," *Food chemistry*, vol. 141, no. 1, pp. 389–396, 2013.

[122] M. Kamruzzaman, D.-W. Sun, G. ElMasry, and P. Allen, "Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis," *Talanta*, vol. 103, pp. 130–136, 2013.

[123] J. Weeranantanaphan, G. Downey, P. Allen, and D.-W. Sun, "A review of near infrared spectroscopy in muscle food analysis: 2005–2010," *Journal of Near Infrared Spectroscopy*, vol. 19, no. 2, pp. 61–104, 2011.

[124] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[125] C. J. C. Burges, "Geometry and invariance in kernel based methods," in *Advances in Kernel Methods—Support Vector Learning*. The MIT Press, 1999, pp. 89–116.

[126] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines (version 3.20)." [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[127] X. Tang and W. Pearlman, "Lossy-to-lossless block-based compression of hyperspectral volumetric data," in *Proceedings of 2004 International Conference on Image Processing*, vol. 5. IEEE, 2004, pp. 3283–3286.

[128] Z. Xiong, K. Ramchandran, M. Orchard, and Y.-Q. Zhang, "A comparative study of DCT- and wavelet-based image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 5, pp. 692–695, 1999.

[129] D. Adjeroh and S. Sawant, "Error-resilient transmission for 3D DCT coded video," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 178–189, 2009.

[130] M.-T. Sun and I.-M. Pao, "Statistical computation of discrete cosine transform in video encoders," *Journal of Visual Communication and Image Representation*, vol. 9, no. 2, pp. 163–170, 1998.

[131] M. Lee, R. Chan, and D. Adjeroh, "Quantization of 3D-DCT coefficients and scan order for video compression," *Journal of visual communication and image representation*, vol. 8, no. 4, pp. 405–422, 1997.

[132] B.-L. Yeo and B. Liu, "Volume rendering of DCT-based compressed 3D scalar data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 1, pp. 29–43, 1995.

[133] H. Tang, W. Sun, B. Gao, and F. Zhang, "Research on quantization and scanning order for 3-D DCT video coding," in *Proceedings of 2012 International Conference on Computer Science and Electronics Engineering*, vol. 1. IEEE, 2012, pp. 200–204.

[134] G. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[135] R. Hashemian, "Direct Huffman coding and decoding using the table of code-lengths," in *Proceedings of 2003 International Conference on Information Technology: Coding and Computing*. IEEE, 2003, pp. 237–241.

[136] Y. Luo and R. Ward, "Removing the blocking artifacts of block-based DCT compressed images," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 838–842, 2003.

[137] AVIRIS image Moffett Field scene 1. [Online]. Available: http://aviris.jpl.nasa.gov/data/free_data.html

[138] A. Chakrabarty, O. Choudhury, P. Sarkar, A. Paul, and D. Sarkar, "Hyperspectral image classification incorporating bacterial foraging-optimized spectral weighting," *Artificial Intelligence Research*, vol. 1, no. 1, pp. 63–83, 2012.

[139] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2297–2307, 2010.

[140] AVIRIS image Salinas Valley. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[141] ROSIS image Pavia University. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[142] T.-J. Chen, K.-S. Chuang, J. Wu, S. C. Chen, M. Hwang, and M.-L. Jan, "Quality degradation in lossy wavelet image compression," *Journal of Digital Imaging*, vol. 16, no. 2, pp. 210–215, 2003.

[143] B. J. Erickson, "Irreversible compression of medical images," *Journal of Digital Imaging*, vol. 15, no. 1, pp. 5–14, 2002.

[144] S. Winkler and P. Mohandas, "The evolution of video quality measurement: from PSNR to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.

[145] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[146] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[147] M. R. Pickering and M. J. Ryan, "An architecture for the compression of hyperspectral imagery," in *Hyperspectral Data Compression*, G. Motta, F. Rizzo, and J. A. Storer, Eds. New York: Springer, 2006, pp. 1–34.

[148] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Quality assessment for hyperspectral imagery: Comparison between lossy and near-lossless compression," in *Proceedings of Fortieth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2006, pp. 1902–1906.

[149] C. Hsu and C. Lin, "BSVM multiclass classifier 2.08." [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/bsvm/

[150] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural networks*, vol. 17, no. 1, pp. 113–126, 2004.

[151] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Progressive 3-D coding of hyperspectral images based on JPEG 2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 125–129, 2006.

[152] S. Boussakta and H. O. Alshibami, "Fast algorithm for the 3-D DCT-II," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 992–1001, 2004.

[153] T. Qiao, J. Ren, M. Sun, J. Zheng, and S. Marshall, "Effective compression of hyperspectral imagery using an improved 3D DCT approach for land-cover analysis in remote-sensing applications," *International Journal of Remote Sensing*, vol. 35, no. 20, pp. 7316–7337, 2014.

[154] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[155] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 397–408, 2006.

[156] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 973–980, 2011.

[157] D. Xu, L. Sun, J. Luo, and Z. Liu, "Analysis and denoising of hyperspectral remote sensing image in the curvelet domain," *Mathematical Problems in Engineering*, vol. 2013, 2013.

[158] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.

[159] J.-L. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[160] M. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

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

[162] L. Sun and J. Luo, "Junk band recovery for hyperspectral image based on curvelet transform," *Journal of Central South University of Technology*, vol. 18, pp. 816–822, 2011.

[163] J.-L. Starck, F. Murtagh, E. Candès, and D. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *IEEE Transactions on Image Processing*, vol. 12, no. 6, pp. 706–717, 2003.

[164] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[165] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Berlin, Germany: Springer-Verlag, 2013.

[166] N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*. London, UK: CRC Press, 2001.

[167] E. Candès and D. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*. Nashville, TN: Vanderbilt University Press, 1999, pp. 105–120.

[168] S. Kouchaki, S. Sanei, E. Arbon, and D.-J. Dijk, "Tensor based singular spectrum analysis for automatic scoring of sleep EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 1, pp. 1–9, 2015.

[169] W. Pereira, S. Bridal, A. Coron, and P. Laugier, "Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancellous bone specimens," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 51, no. 3, pp. 302–312, 2004.

[170] E. Candès, L. Demanet, and L. Ying. CurveLab toolbox 2.1.2. [Online]. Available: http://www.curvelet.org/papers.html

[171] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[172] P. Zhong and R. Wang, "Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 688–705, 2011.

[173] X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *Journal of Central South University of Technology*, vol. 53, no. 4, pp. 2241–2253, 2015.

[174] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6663–6674, 2015.

[175] J. A. Benediktsson, J. A. 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. 480–491, 2005.

[176] A. Ignat, "Combining features for texture analysis," in *Proceedings of Sixteenth International Conference on Computer Analysis of Images and Patterns*. Springer, 2015, pp. 220–229.

[177] Y. Li and A. Ngom, "The non-negative matrix factorization toolbox for biological data mining," *Source Code for Biology and Medicine*, vol. 8, no. 1, pp. 1–15, 2013.

[178] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 184–198, 2013.

[179] M. Rojas, I. Dópido, A. Plaza, and P. Gamba, "Comparison of support vector machine-based processing chains for hyperspectral image classification," in *Satellite Data Compression, Communications, and Processing VI*. International Society for Optics and Photonics, 2010, p. 78100B.

[180] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.

[181] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3947–3960, 2011.

[182] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2666–2677, 2014.

[183] ——, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3742–3752, 2014.

[184] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[185] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[186] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[187] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[188] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.

[189] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing–Special Issue on Sparse Approximations in Signal and Image Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[190] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[191] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2056–2065, 2014.

[192] K. Schindler, "An overview and comparison of smooth labeling methods for land-cover classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4534–4545, 2012.

[193] K. Kotwal and S. Chaudhuri, "Visualization of hyperspectral images using bilateral filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2308–2316, 2010.

[194] H. Peng and R. Rao, "Hyperspectral image enhancement with vector bilateral filtering," in *Proceedings of Sixteenth IEEE International Conference on Image Processing*. IEEE, 2009, pp. 3713–3716.

[195] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of Sixth International Conference on Computer Vision*. IEEE, 1998, pp. 839–846.

[196] J. Tropp, A. C. Gilbert *et al.*, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[197] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[198] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 527–541, 2015.

[199] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging–an emerging process analytical tool for food quality and safety control," *Trends in Food Science & Technology*, vol. 18, no. 12, pp. 590–598, 2007.

[200] B. S. Dissing, O. S. Papadopoulou, C. Tassou, B. K. Ersbøll, J. M. Carstensen, E. Z. Panagou, and G.-J. Nychas, "Using multispectral imaging for spoilage detection of pork meat," *Food and Bioprocess Technology*, vol. 6, no. 9, pp. 2268–2279, 2013.

[201] S. Andrés, I. Murray, E. A. Navajas, A. V. Fisher, N. R. Lambe, and L. Bünger, "Prediction of sensory characteristics of lamb meat samples by near infrared reflectance spectroscopy," *Meat Science*, vol. 76, no. 3, pp. 509–516, 2007.

[202] L. E. Jeremiah, Z. L. Carpenter, and G. C. Smith, "Beef color as related to consumer acceptance and palatability," *Journal of Food Science*, vol. 37, no. 3, pp. 476–479, 1972.

[203] S. D. Shackelford, T. L. Wheeler, and M. Koohmaraie, "On-line classification of us select beef carcasses for *longissimus* tenderness using visible and near-infrared reflectance spectroscopy," *Meat Science*, vol. 69, no. 3, pp. 409–415, 2005.

[204] M. Prevolnik, M. Candek-Potokar, and D. Skorjanc, "Ability of NIR spectroscopy to predict meat chemical composition and quality - a review," *Czech Journal of Animal Science*, vol. 49, no. 11, pp. 500–510, 2004.

[205] H. Abdi, "Partial least square regression (PLS regression)," in *Encyclopedia for Social Science Research Methods*. SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2003, pp. 792–795.

[206] I. Steinwart and A. Christmann, *Support Vector Machines*. Springer Science & Business Media, 2008.

[207] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.

[208] S. Thompson, "United Kingdom slaughter statistics," Department for Environment, Food & Rural Affairs, York, Tech. Rep., 2014.

[209] P. Williams, K. Norris *et al.*, *Near-infrared Technology in the Agricultural and Food Industries.* American Association of Cereal Chemists, Inc., 1987.

[210] T. Naes, T. Isaksson, T. Fearn, and T. Davies, *A User Friendly Guide to Multivariate Calibration and Classification.* NIR publications, 2002.

[211] J. Burger and P. Geladi, "Hyperspectral NIR image regression part II: dataset preprocessing diagnostics," *Journal of Chemometrics*, vol. 20, no. 3-4, pp. 106–119, 2006.

[212] A. G. Frazer, "New Zealand tenderness and local meat quality mark programmes," in *Proceedings of the 43rd International Congress of Meat Science and Technology*, Auckland, New Zealand, 1997, pp. 37–51.

[213] A. Cecchinato, M. De Marchi, M. Penasa, A. Albera, and G. Bittante, "Near-infrared reflectance spectroscopy predictions as indicator traits in breeding programs for enhanced beef quality," *Journal of animal science*, vol. 89, no. 9, pp. 2687–2695, 2011.

[214] X. Ren and D. Ramanan, "Histograms of sparse codes for object detection," in *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 2013, pp. 3246–3253.

[215] M. K. Griffin and H.-h. K. Burke, "Compensation of hyperspectral data for atmospheric effects," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 29–54, 2003.

[216] G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis.* Chichester, UK: Wiley, 2009.

[217] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 674–677, 2007.

Appendix A

Hyperspectral Remote Sensing Data Sets

A.1 Moffett Field scene 1

The Moffett Field scene 1 data set [137], was taken by the airborne visible/infrared imaging spectrometer (AVIRIS) at the jet propulsion laboratory from the national aeronautics and space administration (NASA) [138]. It was acquired above the Moffett Field area in California at the southern end of San Francisco Bay in 1997. This image contains 224 continuous spectral bands of a hilly ridge and an urban area, ranging from 400 to 2500 nm in the complete visible-near infrared-short wave infrared (VIS-NIR-SWIR) spectrum[215]. The spatial area of this data set covers 512 lines and 614 samples with a radiometric resolution of 12 bits. Since there is no ground truth for this data set, only a sample band (band 100) is shown in Figure A.1 for reference.

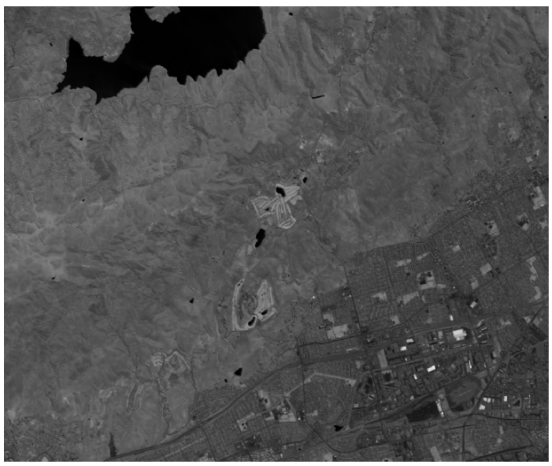


Fig. A.1 Band 100 of Moffett Field scene 1 data set.

A.2 Indian Pines

The Indian Pines data set [10], also known as 92AV3C, was collected over the Indian Pines test site in north-western Indiana, USA, in the early 1990s by AVIRIS as well [138, 139]. The spatial size of this data set contains 145 × 145 pixels of an agriculture area, with a spatial resolution of 18 m [138]. There are 224 spectral reflectance bands covering from 400 to 2500 nm, where the nominal spectral resolution is 10 nm and the radiometric resolution is 16 bits. The scene consists of two-thirds agriculture, and one-third forest or other natural perennial vegetation, with 16 classes in the original ground truth map [216]. However, some classes have insufficient samples for training the classification model. As suggested by other researchers [18, 64, 217], sometimes 7 out of 16 classes are discarded for more consistent results, leaving 9 classes (2-3, 5-6, 8, 10-12 and 14) instead. One random band image (band 168) and the original/reduced ground truth maps of Indian Pines are shown in Figure A.2.

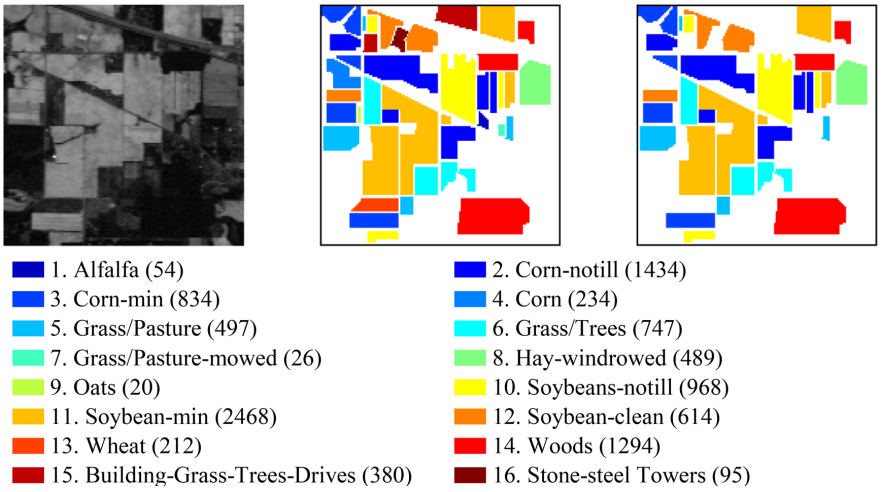


Fig. A.2 Band 168 of Indian Pines data set (left), the original 16 classes ground truth map (middle) and the reduced 9 classes ground truth map (right), with number of samples in parentheses.

A.3 Salinas Valley

The Salinas Valley data set [140], was gathered at Salinas Valley, California in 1988 by AVIRIS at low altitude, leading to a high spatial resolution of 3.7 m per pixel [138]. The full scene, which includes vegetables, bare soils and vineyard fields, comprises 512 lines by 217 samples with 224 spectral bands. The spectral resolution and the radiometric resolution of this data set are the same as Indian Pines, which are 10 nm and 16 bits respectively. A random band (band 50) and the ground truth map of Salinas Valley containing 16 classes are shown in Figure A.3.

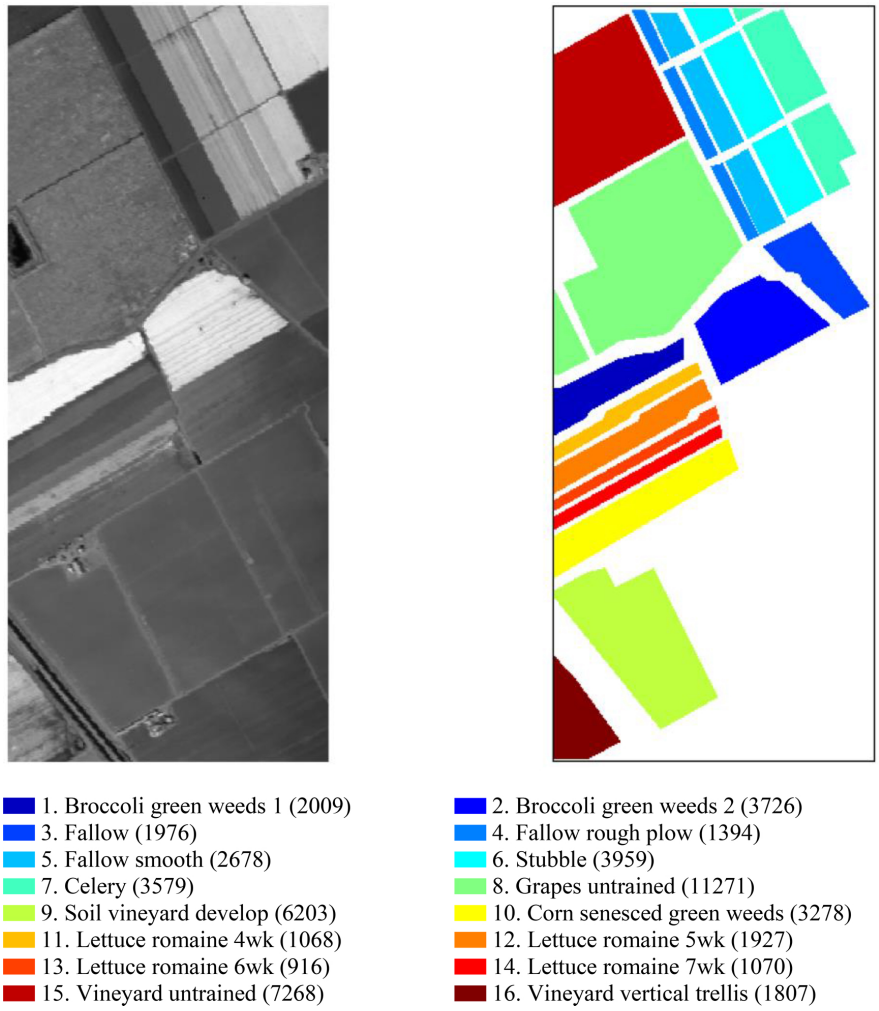


Fig. A.3 Band 50 of Salinas Valley data set (left) and the 16 classes ground truth map (right), with number of samples in parentheses.

A.4 Pavia University

The Pavia University data set [141], was collected using the reflective optics system imaging spectrometer (ROSIS) during a flight campaign over Pavia district in north Italy, with a spatial resolution of 1.3 m per pixel [175]. There are originally 115 bands with a spectral coverage ranging from 430 to 860 nm. However, 12 channels have been removed due to noise, leaving 103 bands with 610 × 340 pixels per band. Nine classes of interest are provided in the ground truth map, including urban, soil and vegetation features, as shown in Figure A.4.

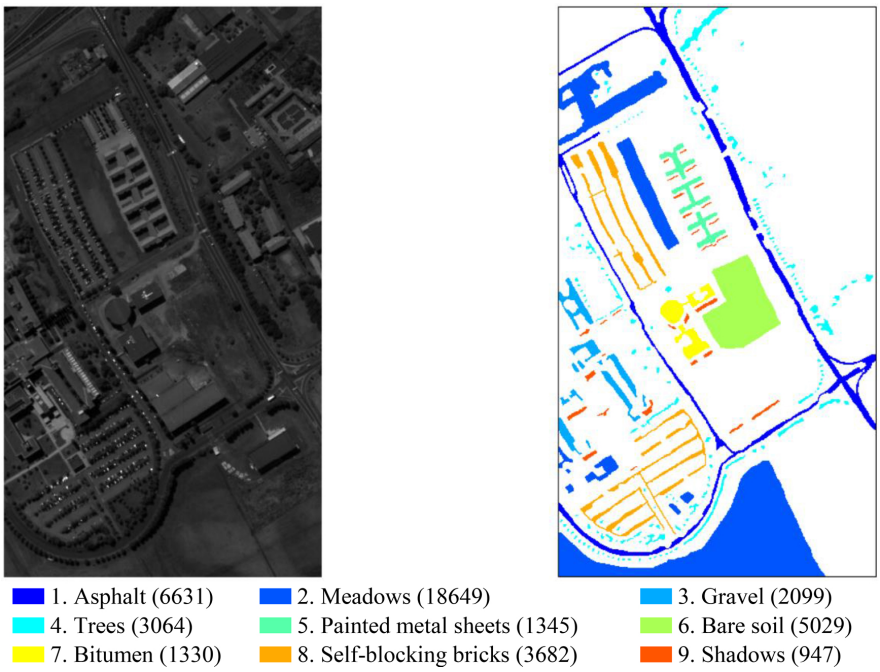


Fig. A.4 Band 70 of Pavia University data set (left) and the 9 classes ground truth map (right), with number of samples in parentheses.

Appendix B

List of Author's Publications

B.1 International journal publications

1. **T. Qiao**, J. Ren, M. Sun, J. Zhang, and S. Marshall, "Effective compression of hyperspectral imagery using an improved 3D DCT approach for land-cover analysis in remote-sensing application," *International Journal of Remote Sensing*, vol. 35, no. 2, pp. 7316–7337, 2014.

2. **T. Qiao**, J. Ren, C. Craigie, J. Zabalza, C. Maltin, and S. Marshall,"Quantitative prediction of beef quality using visible and NIR spectroscopy with large data samples under industry condition," *Journal of Applied Spectroscopy*, vol. 82, no. 1, pp. 137–144, 2015.

3. **T. Qiao**, J. Ren, C. Craigie, J. Zabalza, C. Maltin, and S. Marshall,"Singular spectrum analysis for improving hyperspectral imaging based beef eating quality evaluation," *Computers and Electronics in Agriculture*, vol. 115, pp. 21–25, 2015.

4. **T. Qiao**, J. Ren, Z. Yang, C. Qing, and J. Zabalza,"Visible hyperspectral imaging for lamb quality prediction," *TM - Technisches Messen*, vol. 82, no. 12, pp. 643–652, 2015.

5. X. Li, J. Ren, C. Zhao, **T. Qiao**, and S. Marshall,"Novel multivariate vector quantization for effective compression of hyperspectral imagery," *Optics Communications*, vol. 332, pp. 192–200, 2014.

B.2 International conference publications

1. **T. Qiao**, J. Ren, C. Craigie, S. Marshall, and C. Maltin, "Effective compression of hyperspectral imagery using improved three dimensional cosine transform," in *2013 6th UK, Europe, China Millimeter Waves and THz Technology Workshop (UCMMT)*, Rome, Italy, 2013, pp. 1–2.

2. **T. Qiao**, J. Ren, C. Craigie, J. Zabalza, C. Maltin, and S. Marshall, "Comparison between near infrared spectroscopy and hyperspectral imaging in predicting beef eating quality," presented at the Hyperspectral Imaging Conference, Coventry, UK, 2014.

3. **T. Qiao**, J. Ren, J, Zabalza, and S. Marshall, "Prediction of lamb eating quality using hyperspectral imaging," in *2nd International Conference on optical Characterization of Materials*, Karlsruhe, Germany, 2015, pp. 15–25.

4. S. Marshall, T. Kelman, **T. Qiao**, P. Murray, and J. Zabalza, "Hyperspectral imaging for food applications," in *2015 European Signal Processing Conference (EUCIPCO)*, Nice, France, 2015.

B.3 Journal publications under consideration

1. **T. Qiao**, J. Ren, Z. Wang, J. Zabalza, S. Li, J. Benediktsson, and S. Marshall, "Hyperspectral image feature extraction by exploiting the curvelet domain using singular spectrum analysis," submitted to *IEEE Transactions on Geoscience and Remote Sensing*.

2. **T. Qiao**, J. Ren *et al.*, "Joint bilateral filtering and spectral similarity-based sparse representation for effective feature extraction and classification in hyperspectral imaging," under preparation.