Spectral-spatial Feature Extraction for Hyperspectral Image Classification

Jie Liang

A thesis submitted for the degree of
Doctor of Philosophy of
The Australian National University

January 2017

Except where otherwise indicated, this thesis is my own original work.

Jie Liang
12 January 2017

To my dear parents, Fengzhi Kang and Guangyong Liang.

Acknowledgments

I would thank my primary supervisor Dr. Jun Zhou. Without his help, I cannot finish this thesis. I received lots of help from him, including but not limited to following: discussing my research progress every week; helping me revise my conference and journal papers; purchasing new hyperspectral cameras to support my research; training me how to respond to reviews of my papers and how to review papers from other authors; confirming the chapter details of my thesis. Residing in a relatively new area, my research topic "spectral-spatial feature extraction for hyperspectral image classification" is quite challenging. I got stuck with my research frequently due to lack of datasets or standard evaluation methods. Jun and I had many discussions to solve those problems and explore new possibilities in the scope of my research topic. We believe that there are excellent opportunities with this topic, and it will draw much attention in the near future.

My panel chair, Associate Professor Hongdong Li, monitored my research progress and helped me complete a lot of paperwork when I was working as an external student of ANU. My co-supervisor, Associate Professor Henry Gardner, supported my research from Ph.D. application to the thesis finalization. My advisor Dr. Xavier Sirault introduced me to the advanced High-Resolution Plant Phenomics Centre and gave me opportunities to work with talented people there.

Professor Yongsheng Gao, although not in my supervising panel, actually worked as my supervisor as well. He brought me a lot of brilliant ideas and advices, not only for my research but also for my life, while I was visiting the Image Processing & Computer Vision Lab at Griffith University. He also granted me the opportunity to take part in the ARC Linkage Project on strawberry bruise detection via hyperspectral imaging. Rudi Bartels, Gilbert Eaton, and Lee Hamilton were my colleagues in the

linkage project, also helped me a lot in building the strawberry dataset. My lab partners Lei Tong, Ali Zia, and Yanyang gave me their selfless assistances for my research work. Besides that, Professor Yuntao Qian from Zhejiang University also worked as my advisor when he visited Griffith University in 2016. He inspired me a lot of ideas and helped a lot for my journal publication.

Jieqiong Hao, my girlfriend, entered my life at the toughest time of my Ph.D. studying. She kept encouraging me to insist on my research and accompanied me through a lot of difficulties. She acted as my life advisor and offered so many useful suggestions.

Back in my home country China, my uncles Guangyi Liang and Ming Liang, and aunts Junling Liang and Hong Fu greeted me now and then, comforting and encouraging me, and helped me take care of my families. Without them, I could not focus on my studying in Australia.

I would also thank my former supervisor Zhiqiang Zheng and Xiaohong Xu from National University of Defense Technology, where I got my bachelor degree. They offered me such an incredible opportunity to study abroad. They supported me during my application of Ph.D. and gave me the warm welcome every time visiting them in China. I appreciate their trust in me and recognition to my efforts.

The Chinese Scholarship Council and the Australian National University granted me the scholarship to live and study in Australia. ANU also supported me financially to attend academic conferences. These precious experiences enriched my life and expanded my vision. Griffith University also supplied me with all the research resources that I needed to finish my study.

At last, I would give my deepest thanks to my parents: Fengzhi Kang and Guangyong Liang, who raised me, educated me, pushed me to pursue a different life from their generation. Without them, I probably would only live my whole life in a small town, without knowing and enjoying the colorful world outside.

Abstract

As an emerging technology, hyperspectral imaging provides huge opportunities in both remote sensing and computer vision. The advantage of hyperspectral imaging comes from the high resolution and wide range in the electromagnetic spectral domain which reflects the intrinsic properties of object materials. By combining spatial and spectral information, it is possible to extract more comprehensive and discriminative representation for objects of interest than traditional methods, thus facilitating the basic pattern recognition tasks, such as object detection, recognition, and classification. With advanced imaging technologies gradually available for universities and industry, there is an increased demand to develop new methods which can fully explore the information embedded in hyperspectral images. In this thesis, three spectral-spatial feature extraction methods are developed for salient object detection, hyperspectral face recognition, and remote sensing image classification.

Object detection is an important task for many applications based on hyperspectral imaging. While most traditional methods rely on the pixel-wise spectral response, many recent efforts have been put on extracting spectral-spatial features. In the first approach, we extend Itti's visual saliency model to the spectral domain and introduce the spectral-spatial distribution based saliency model for object detection. This procedure enables the extraction of salient spectral features in the scale space, which is related to the material property and spatial layout of objects.

Traditional 2D face recognition has been studied for many years and achieved great success. Nonetheless, there is high demand to explore unrevealed information other than structures and textures in spatial domain in faces. Hyperspectral imaging meets such requirements by providing additional spectral information on objects, in completion to the traditional spatial features extracted in 2D images. In the second approach, we propose a novel 3D high-order texture pattern descriptor

for hyperspectral face recognition, which effectively exploits both spatial and spectral features in hyperspectral images. Based on the local derivative pattern, our method encodes hyperspectral faces with multi-directional derivatives and binarization function in spectral-spatial space. Compared to traditional face recognition methods, our method can describe distinctive micro-patterns which integrate the spatial and spectral information of faces.

Mathematical morphology operations are limited to extracting spatial feature in two-dimensional data and cannot cope with hyperspectral images due to so-called ordering problem. In the third approach, we propose a novel multi-dimensional morphology descriptor, tensor morphology profile (TMP), for hyperspectral image classification. TMP is a general framework to extract multi-dimensional structures in high-dimensional data. The n^{th}-order morphology profile is proposed to work with the n^{th}-order tensor, which can capture the inner high order structures. By treating a hyperspectral image as a tensor, it is possible to extend the morphology to high dimensional data so that powerful morphological tools can be used to analyze hyperspectral images with fused spectral-spatial information.

At last, we discuss the sampling strategy for the evaluation of spectral-spatial methods in remote sensing hyperspectral image classification. We find that traditional pixel-based random sampling strategy for spectral processing will lead to unfair or biased performance evaluation in the spectral-spatial processing context. When training and testing samples are randomly drawn from the same image, the dependence caused by overlap between them may be artificially enhanced by some spatial processing methods. It is hard to determine whether the improvement of classification accuracy is caused by incorporating spatial information into the classifier or by increasing the overlap between training and testing samples. To partially solve this problem, we propose a novel controlled random sampling strategy for spectral-spatial methods. It can significantly reduce the overlap between training and testing samples and provides more objective and accurate evaluation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Hyperspectral Imaging

Human vision has evolved to satisfy the requirement of living and played a significant role in finding food or avoiding dangers in the history [1]. We can sense not only the brightness of light but also the color. The trichromatic color vision results from three types of color photoreceptors which are sensitive to three different spectra in the visible light, corresponding to blue, green and red. Via cross-reference of these three colors, we can distinguish a number of color signals. Essentially, color is a vector instead of a scalar and color perception is a comparative sensory phenomenon. However, the human is still restricted to visible light and limited spectral resolution.

Researchers have developed different instruments to capture irradiance from objects in various wavelengths. For example, a spectrometer is used to measure the spectral irradiance at a single point, and a conventional RGB camera takes the integrated irradiance across the visible wavelength at a region of interest. Different from these two types of instruments, hyperspectral imager can obtain both spatial and spectral information simultaneously. A hyperspectral image usually contains tens or hundreds of continuous light wavelength indexed spectral bands, providing much higher spectral resolution than human vision or regular RGB cameras. The spectrum it covers is not only limited to the visible spectrum but also extends to ultraviolet or infrared light depending on the equipment characteristics. It has been known that objects consisting of different materials may emit, reflect, and absorb light and the proportion is a function of light wavelength or frequency [2]. Therefore, the irradi-

ance from objects by a hyperspectral camera can be used as a clue to estimate the physical or chemical properties of the objects.

Due to its high discriminative ability to identify and distinguish different materials, hyperspectral imaging has been widely used in remote sensing to analyze the earth surface for applications such as mining, military, agriculture, environment monitoring, etc. Moreover, with compact and high spatial resolution commercial instruments being developed, hyperspectral imaging has attracted increasing interests in close-range applications, e.g. food security, biomedicine, biometrics and quality control, etc. Another notable trend in the last decade is the introduction of hyperspectral imaging into computer vision applications. By providing extra information in the spectral domain, hyperspectral imaging has great potential to push forward research in some challenging problems in computer vision area, such as illumination estimation, super-resolution, camera sensitivity analysis [3, 4], scene analysis [5], document processing [6], object classification [7], etc.

Supporting a broad range of scope in research and applications, hyperspectral image classification covers broad topics in predicting the categories of targets, for instance, general object/scene classification, saliency detection, and image labeling, etc. The targets vary from single pixels, regions, objects to scenes. In remote sensing, hyperspectral image classification mainly focuses on pixel-level classification for land-cover classes identification and thematic map generation. In computer vision, hyperspectral images have not been widely used due to absence of large scale image data. Most image classification tasks are still using grayscale or color images which usually contain single object [8, 9, 10].

Though hyperspectral imaging gives access to more valuable and abundant data than grayscale and RGB images, it is necessary to extract and refine information for various analytical tasks. Feature extraction addresses this problem by providing compact and informative set of information to increase the efficiency and effectiveness of data storage and processing. Numerous feature extraction methods have been

developed in remote sensing and computer vision during the past decade. However, most of these methods focus on either pixel-wise spectral information in remote sensing or gray-scale spatial information in computer vision, in accordance with the characteristics of images available in two areas, i.e., high spectral resolution and low spatial resolution in remote sensing, and the opposite in computer vision.

For hyperspectral images obtained by advanced hyperspectral cameras, detailed spectral information and fine spatial resolution enable analysis of both material and structure of objects in a scene. However, most of the existing methods are not suitable for this task due to the constraints as mentioned above. Therefore, it is necessary to develop new techniques to exploit these underlying spatial and spectral information in hyperspectral images, thus addressing the limitation of human vision, computer vision, and remote sensing. Though some attempts have been made in both computer vision and remote sensing recently, there is still a huge gap between hyperspectral imaging and practical classification applications due to lack of effective spectral-spatial feature extraction methods. This thesis focuses on developing several novel methods based on spatial and spectral analysis, thus facilitating hyperspectral image classification.

1.2 Motivation

The demand for new spectral-spatial feature extraction methods is mainly from two perspectives, remote sensing, and computer vision. Hyperspectral images were firstly adopted in remote sensing when NASA Jet Propulsion Laboratory developed the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in 1987 [11]. In this application area, they are usually acquired by airborne or spaceborne sensors, capturing the spectral radiance of the earth surface. With advancement in sensor technology, the spatial resolution of hyperspectral images has significantly increased and is much higher than 29 years ago, describing detailed structures, shapes, and textures of the scene. Pixel analysis methods are widely adopted in traditional re-

mote sensing image analysis, which usually extract spectral information with band selection or dimensionality reduction without exploring spatial information. Due to random noises from data collection, transmission, and processing, pixel analysis may create pepper and salt like classification map instead of a spatially coherent map [2]. This type of method has ignored the fact that different objects with the same material may be distinguished with the aid of the spatial information.

Nevertheless, facing the challenging obstacles in pattern recognition with computational models, researchers have mainly focused on developing sophisticated feature extraction and machine learning approaches, take two influential works for example, Scale-Invariant Feature Transform (SIFT) [12] and Convolutional Neural Network (CNN) [9]. These methods are developed to match human vision system by exploring visual processing mechanism in the brain. Although tremendous success has been achieved, most methods are still limited to grayscale or RGB images analysis, meaning that there is no additional input into computational models than the human vision system.

Rather than relying on image processing and machining learning algorithms, hyperspectral imaging provides computer vision with new opportunities. First, hyperspectral images contain tens or hundreds of bands, thus dramatically increasing its discriminative ability to distinguish a large number of spectral responses beyond the trichromatic human vision system. Furthermore, some extra information beyond visible light range can be captured, for instances, fluorescence which is related to absorption of ultraviolet light, infrared light which can be used to detect plants, and the spectrum in short wave infrared light that water absorbs. The spectral response of materials is called spectral signature, with which the materials can be easily classified or detected.

Motivated by these facts, in this thesis, we explore the possibilities of fusing the spectral and spatial information by extracting the spectral-spatial features, thus facilitating the basic pattern recognition tasks, such as object detection, recognition,

and classification.

1.2.1 The advance of imaging technology

Hyperspectral imaging instruments were originally developed for airborne and satellite platforms so as to measure the spectral characteristics of land cover on earth surface. They were seldom available for close-range data collection due to their massive sizes and high prices. With new material, optical, electronics technologies applied to this area, new instruments keep emerging and becoming faster, cheaper, and more compact. Without much loss of spectral range or resolution, the spatial resolution of new hyperspectral cameras has been dramatically increased, making them comparable to conventional chromatic or monochrome cameras. Table 1.1 shows some examples of hyperspectral imaging instruments developed during the past years. We can observe that there are more and more high spatial and spectral resolution hyperspectral cameras available in the market.

1.2.2 Demand for new hyperspectral image processing methods

For the last decade, a series of methods have been developed to tackle the problems of the particular characteristics of hyperspectral images. Different mathematical formalisms have been built for typical tasks, such as classification, segmentation, spectral mixture analysis. However, most of such works still focus on the spectral domain. With high spatial resolution images available, it is inevitably demanded to develop new spectral-spatial methods for the emerging images. Meanwhile, hyperspectral imaging is still a new topic in the area of computer vision. The limited amount of existing research pays attention to the low level of image processing, for instance, how to obtain the hyperspectral image or videos with high spatial resolution, how to denoise the hyperspectral image and so on. There is a lack of basic methods for the hyperspectral image processing, especially the feature extraction methods.

[1]The spatial resolution of ground based camera depends on lens and object distance.

Table 1.1: Hyperspectral imaging instruments[1].

Name	Year	Producer	Spectral resolution	Spectral range	Spatial resolution	Application
Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)	1980s	JPL, NASA	10nm	400-2000nm	100m	remote sensing
Hyperion	2000s	NASA	10nm	400-2500nm	30m	remote sensing
Reflective Optics System Imaging Spectrometer (ROSIS)	2000s	German Aerospace Agency	4nm	430-860nm	1-6m	remote sensing
Hyperspectral Imaging Instruments	2013	BRIMROSE	5-20nm	400-1700nm	$\backsim$ 1cm	ground based
Pika	2015	RESONON	1.3-4.9nm	400-1700nm	$\backsim$1cm	ground based
Hyperspectral Cameras	2015	IMEC	10-15nm	400-1000nm	$\backsim$1cm	ground based

On the other hand, the hyperspectral imaging has broad application prospects in agriculture, industry, and military due to its higher discriminative ability than conventional cameras. But the lack of basic methods has constrained the analysis in the spectral domain. The various spatial approaches from the computer vision area have not been associated with the spectral methods. A lot of work needs to be done to build the foundation of hyperspectral imaging. It can be expected that there will be a trend that hyperspectral imaging plays a more important role in computer vision and remote sensing in near future, and the spectral-spatial feature extraction will make a difference to the prosperity of hyperspectral imaging.

1.2.3 Challenges

It is a challenging task to extract spectral-spatial features without introducing too much data complexity. The main challenges come from two perspectives: how to fuse spectral and spatial features and how to reduce redundant information in the extracted feature [13]. Several other factors also influence the spectral-spatial feature extraction process, including illumination, noise, and a large amount of data. We summarize these challenging problems as follows:

1. Huge data and high computation complexity. Hyperspectral images consist of tens or hundreds of bands. The size of a single hyperspectral image is usually much bigger than an RGB or grayscale image. To store, process and analyze such massive data is challenging for personal computers.

2. How to fuse spatial and spectral information. In hyperspectral images, spatial and spectral domain are completely different, showing distinct characteristics. Spatial information usually involves shape, structure, texture, edge, as well as contextual information. On the other hand, spectral information is extracted from the irradiance of object surface, showing unique object material related properties. In contrast to spatial edges and corners, reliable spectral information is usually located in a flat region which corresponds to a group of pixels

made of the same material. Due to their different characteristics, there is no straight way to fuse them to represent spectral-spatial features.

3. Curse of Dimensionality. In machine learning, the curse dimensionality refers to a phenomenon that when the dimensionality of feature space increases, the amount of data demanded to obtain a statistically reliable result grows exponentially with the dimensionality. This issue is because the number of parameters to be estimated will increase dramatically. Determining and optimizing such a large number of parameters in a high-dimensional space is problematic. If the number of training instances is much less than the dimension of the features, the learning algorithms tend to overfit. Hyperspectral image classification usually suffers from such a problem due to its high dimensional nature and limited labeled data. In many cases, the produced spectral-spatial features are of ultra high dimensionality [14, 15].

4. Influence of illumination and other external factors. Object material is not the only fact determining the spectral response of objects. Several factors also affect the imaging process, such as illumination, object geometry, surface smoothness, sensor sensitivity, etc.. Illumination plays an important role here. Objects illuminated by different light sources, such as sunlight, tungsten, halogen, and fluorescent, show varying spectral irradiance. Varying lighting condition substantially constrains the application of hyperspectral imaging in uncontrolled environment. Geometry and surface smoothness determine whether shadows or specularity will contaminate the spectral responses. Therefore, how to obtain consistent spectral reflectance under different conditions needs to be figured out.

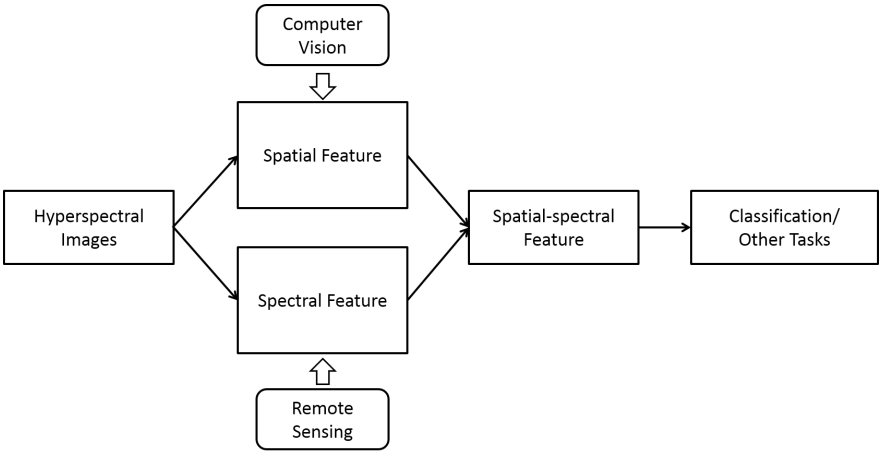


Figure 1.1: The framework of the proposed methodology.

1.3 Objective

The objective of this thesis is to develop novel spectral-spatial feature extraction solutions for hyperspectral image classification. Several goals shall be achieved, including effective fusing of spectral-spatial feature, achieving superior performance than spatial or spectral only features in several common tasks, such as remote sensing land cover classification, face recognition, salient region detection. For various tasks with different objectives, the developed features shall meet their unique needs. While spectral information is primary and spatial information is supplementary for remote sensing hyperspectral image classification, spatial structure is more important than spectral information for face recognition. Fig. 1.1 shows the framework of the proposed methodology. Spatial and spectral feature extraction methods originate from computer vision and remote sensing, respectively. The fused spectral-spatial features serve as the input to classification and other tasks. Our ultimate goal is to develop practical hyperspectral image descriptors and feature representation methods, so as to increase the classification accuracy, especially when available data are limited.

1.4 Contribution

In this thesis, we have developed three spectral-spatial feature extraction methods for saliency detection, hyperspectral face recognition, and remote sensing image classification. Furthermore, we also discuss the sampling strategy for the evaluation of spectral-spatial methods in remote sensing hyperspectral image classification.

The first method is for salient object detection in hyperspectral images. Object detection is an important task for many applications. While most traditional methods are pixel-based on hyperspectral images, many recent efforts have been putting on extracting spectral-spatial features. In this method, we extend Itti's visual saliency model to the spectral domain and introduce the spectral-spatial distribution based saliency model for object detection. This model enables the extraction of salient spectral features in the scale space, which is related to the material property and spatial layout of objects. To our knowledge, this is the first attempt to combine hyperspectral data with salient object detection. Several methods have been implemented and compared to show how color component in the traditional saliency model can be replaced by spectral information.

Traditional 2D face recognition has been studied for many years and achieved great success. Nonetheless, there is high demand to explore unrevealed information other than structure and texture in spatial domain in the faces. Hyperspectral imaging meets such requirements by providing additional spectral information on objects, in completion to the traditional spatial features extracted in 2D images. In the second approach, we propose a novel 3D high-order texture pattern descriptor for hyperspectral face recognition, which effectively exploits both spatial and spectral features in hyperspectral images. Based on the local derivative pattern, our method encodes the hyperspectral faces with multi-directional derivatives and binarization function in spectral-spatial space. Then a spectral-spatial feature descriptor is generated by applying a 3D histogram on the derivative pattern, which can be used to convert hyperspectral face images into vectorized representations. Compared to traditional face

recognition methods, our method can describe the distinctive micro-patterns which integrate both spatial and spectral information in faces.

The third approach is based on morphological profiles for remote hyperspectral image classification. Traditional mathematical morphology operations are limited to extracting spatial feature in two-dimensional data and are not able to be applied to hyperspectral image due to the so-called ordering problem. In this approach, we propose a novel multi-dimensional morphology descriptor, namely, tensor morphology profile (TMP). TMP is a general framework to extract multi-dimensional structures in high-dimensional data. The n^{th}-order morphology profile is proposed to work with the n^{th}-order tensor, which can capture the inner high order structures. By treating a hyperspectral image as a tensor, it is possible to extend morphology operations to high dimensional data so that this powerful tool can be used to analyze hyperspectral images with fused spectral-spatial information.

Spectral-spatial processing has been increasingly explored in remote sensing hyperspectral image classification. While extensive studies have focused on developing methods to improve the classification accuracy, experimental setting and design for method evaluation have drawn little attention. In the scope of supervised classification, we find that traditional experimental designs for spectral processing are often improperly used in the spectral-spatial processing context, leading to unfair or biased performance evaluation. This is especially the case when training and testing samples are randomly drawn from the same image - a practice that has been commonly adopted in the experiments. Under such a setting, the dependence caused by overlap between the training and testing samples may be artificially enhanced by some spatial information processing methods such as spatial filtering and morphological operation. Such an interaction between training and testing sets has violated data independence assumption that is abided by supervised learning theory and performance evaluation mechanism. Therefore, the widely adopted pixel-based random sampling strategy is not always suitable to evaluate spectral-spatial classification al-

gorithms because it 's hard to determine whether the improvement of classification accuracy is caused by incorporating spatial information into a classifier or by increasing the overlap between training and testing samples. To partially solve this problem, we propose a novel controlled random sampling strategy for spectral-spatial methods. It can substantially reduce the overlap between training and testing samples and provides more objective and accurate evaluation.

1.5 Thesis Outline

The rest of this thesis is structured as follows. Chapter 2 presents a literature review on hyperspectral imaging technology, feature extraction in both computer vision and remote sensing, and potential combination of spatial and spectral features. Chapter 3 introduces the proposed hyperspectral salient object detection method. In Chapter 4, we introduce 3D local derivative pattern for hyperspectral face recognition. Then, tensor morphological profile for hyperspectral image classification is elaborated in Chapter 5. After these three approaches, Chapter 6 discusses the sampling strategy for the evaluation of spectral-spatial methods in hyperspectral image classification. At last, a summary of this thesis and future work are given in Chapter 7.

1.6 List of Publications

1. Jie Liang, Jun Zhou, Yuntao Qian, Lian Wen, Xiao Bai, and Yongsheng Gao. On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–19, Nov 2016. (Corresponding to Chapter 6).

2. Jie Liang, Jun Zhou, Yuntao Qian, and Yongsheng Gao. Tensor morphology profile versus extended morphology profile: hyperspectral image classification with spatial-spectral descriptor. *Submitted to IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Aug 2016. (Corresponding to Chapter 5).

3. Changhong Liu, Jun Zhou, Jie Liang, Yuntao Qian, Hanxi Li, and Yongsheng Gao. Exploring structural consistency in graph regularized joint spectral-spatial sparse coding for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP(99):1–14, Aug 2016.

4. Jie Liang, Jun Zhou, Xiao Bai, and Yuntao Qian. Salient object detection in hyperspectral imagery. In *2013 IEEE International Conference on Image Processing*, pages 2393–2397. IEEE, Sep 2013. (Corresponding to Chapter 3).

5. Jie Liang, Ali Zia, Jun Zhou, and Xavier Sirault. 3D plant modelling via hyperspectral imaging. In *IEEE International Conference on Computer Vision 2013 and Computer Vision for Accelerated Bioscience*, pages 172–177. IEEE, Dec 2013.

6. Jie Liang, Jun Zhou, and Yongsheng Gao. 3D local derivative pattern for hyperspectral face recognition. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, volume 1, pages 1–6. IEEE, May 2015. (Corresponding to Chapter 4).

7. Jie Liang, Jun Zhou, and Yongsheng Gao. Tensor morphological profile for hyperspectral image classification. In *2016 IEEE International Conference on Image Processing*, pages 2197–2201, Sep 2016.

8. Jun Zhou, Jie Liang, Yuntao Qian, Yongsheng Gao, and Lei Tong. On the sampling strategies for evaluation of joint spectral-spatial information based classifiers. In *7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2015.

9. Ali Zia, Jie Liang, Jun Zhou, and Yongsheng Gao. 3D reconstruction from hyperspectral images. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 318–325. IEEE, Jan 2015.

10. Ali Zia and Jie Liang. 3D plant modelling using spectral data from visible to near infrared range. In *Computer Vision and Pattern Recognition in Environmental Informatics*. IGI Global Publication, 2016.

Chapter 2

Literature Review

2.1 Hyperspectral Imaging Technology

Before introducing hyperspectral imaging in more details, this section explains the terminology commonly used in related areas, for example, hyperspectral imager, hyperspectral imaging, multispectral imaging, spectral imaging, spectrometer, spectroscopy, and photography.

A hyperspectral imager is an instrument which can capture hyperspectral data/images, while hyperspectral imaging refers to the complete hyperspectral data processing including collection, measurement, analysis and interpretation of spectra. Both hyperspectral imaging and multispectral imaging are kinds of spectral imaging which capture the spectral information at every pixel on the image plane. The difference is that the former emphasizes high spectral resolution and continuous spectral range, while the latter contains only several bands without spectral continuity requirement. Spectral imaging is a combination of photography and spectroscopy. While photography projects scene and object on an electronic sensor or photographic film, spectroscopy studies the interaction between matter and radiated energy at a single point. Spectral imaging takes advantage of both photography and spectroscopy and captures both spatial and spectral information in a scene. Although the conventional camera can capture partial spectral information by separating light into three channels dominated by red, green and blue, the spectral resolution is rough and cannot cover spectrum out of the visible range compared to hyperspectral imaging. For the sake of conciseness and without confusion, we use "spectral imaging",

Figure 2.1: Brimrose hyperspectral imaging system.

"hyperspectral imaging" and "multispectral" interchangeably in the rest of the thesis.

The concept of hyperspectral imaging can trace back to 1980s when NASA's Jet Propulsion Laboratory developed a new remote sensing instrument AVIRIS which covers the wavelengths from 0.4 to 2.5 μm and produces more than two hundred spectral bands. Since then, a variety of hyperspectral systems have been developed, and they are no longer restricted to the remote sensing applications. Hyperspectral cameras capture images at narrow range wavelengths in the electromagnetic spectrum and thus the core issue to solve is how to split the spectrum and transfer light to the imaging sensor. So far, existing technologies such as liquid crystals, grating, light filters have been widely used as light separation tools. They are located either between lens and sensor or in front of the lens. Hyperspectral sensors can be categorized into line scan, push broom, and snapshot designs. The difference between these types is the order of obtaining spatial and spectral information. A snapshot system captures a grayscale image at a particular wavelength and then moves along the spectral dimension to receive the data cube, while a push broom or line scan system gets the complete spectrum at a spatial line and then moves along the other spatial dimension. Push broom system is widely adopted in remote sensing because airborne and satellite platforms fly in one direction which is required by the scan-

ning systems. On the contrary, snapshot cameras are more suitable for close-range computer vision applications where the complete spatial layout of objects and scenes shall be taken at a time. In Fig. 2.1, we show a Brimrose hyperspectral imaging system in the Spectral Lab at Griffith University.

Compared with conventional cameras, hyperspectral cameras have their limitations. First, the spatial resolution of images is relatively low at the cost of extra spectral bands. Furthermore, the imaging process is time-consuming, generally requiring the scene remain still during the exposure time which can last for several or tens of seconds. The exposure time is closely related to illumination condition. Since incident light is separated into narrow bands, a limited amount of light arrives at the sensor. Therefore, the generated images may have low signal to noise ratio. This can only be solved by either extending the exposure time or increase the light intensity when a sensor is given. Sometimes the band images may suffer from out of focus problem as the focus is normally tuned on a single band.

During the past several years, new hyperspectral imaging technologies have been developed so as to allow fast snapshot mechanism. For example, IMEC has designed a new series of hyperspectral cameras with snapshot mosaic technology[1]. The new camera is developed based on a large CMOS sensor wafer in which different pixels in an array are sensitive to different wavelengths. Therefore, this camera can capture real-time images and videos as traditional digital cameras, but with many more spectral channels. This advantage has dramatically increased the feasibility of hyperspectral cameras in real-world computer vision applications.

When data representation is concerned, a hyperspectral image is usually in the form of a data cube as shown in Fig. 2.2. It consists of two spatial dimensions X and Y, and a spectral dimension λ. Each pixel is a vector corresponding to the spectral response at the spatial location. At each wavelength, a grayscale band image can be acquired, representing the spatial distribution of the scene at this band.

[1]Products link: `http://www2.imec.be/be_en/research/image-sensors-and-vision-systems/hyperspectral-imaging.html`

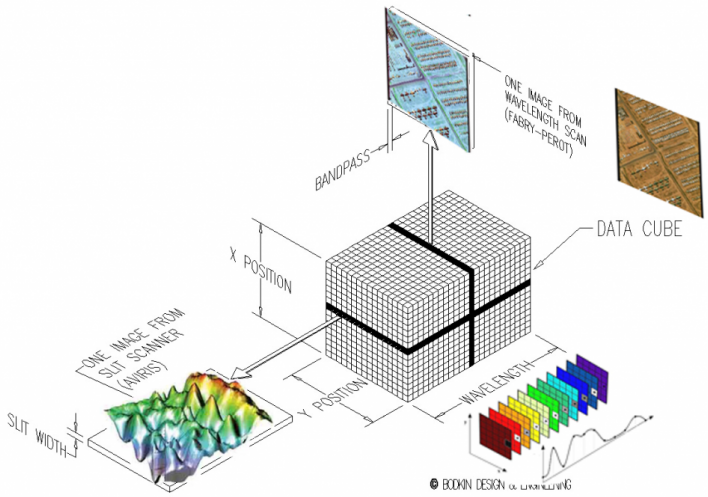


Figure 2.2: Representation of a hyperspectral image and meaning in different dimensions[2].

2.2 Radiometry for Hyperspectral Imaging

There has been a long history since researchers began to study the light and its interaction with matter. Radiometry is such a tool to measure light in the magnetic spectrum. The target of radiometry research covers different light sources such as sunshine and man-made lights, objects of different scales from man-made to earth and universe, and a variety of imaging devices including naked eyes, spectrometer, cameras and radio telescope. Among various imaging devices, spectrometers and cameras are particularly useful to measure the interactions of objects and light courses. Light contains abundant spectral information related to electromagnetic radiance. Imaging devices normally cover ultraviolet, visible and infrared light in the spectral range. Ultraviolet can be further divided into UVA, UVB, and UVC. Visible light is the spectral range which human vision can perceive. Infrared light covers near-infrared, short wave infrared, middle wave infrared, and long wave infrared light spectra. Different spectral ranges of light have unique characteristics and serve different functions. Fig. 2.3 shows the electromagnetic spectrum with respect to the

[2]This figure is from `http://www.bodkindesign.com/products-page/hyperspectral-imaging/hyperspectral-imaging/`

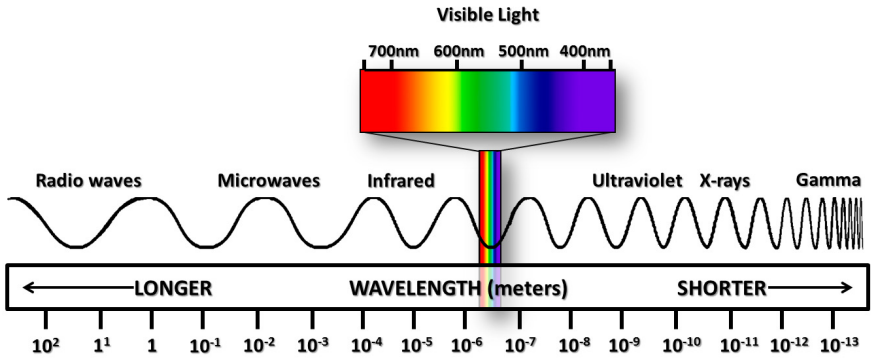


Figure 2.3: Electromagnetic spectrum[3].

wavelength.

Like the formation of normal images, hyperspectral imaging captures a proportion of spectrum within the electromagnetic spectrum. While a color image is normally formed by integration of light in red, blue, and green wavelengths, hyperspectral imaging is able to cover up to hundreds of bands in ultraviolet, visible and infrared ranges.

2.2.1 Spectral range, absorption, and materials

In physics, when an object interacts with light, it may absorb the incident radiation over a range of frequencies which is referred as spectrum absorption. It is primarily determined by the atomic and molecular composition of the material. Consequently, this property can be used to distinguish different elements and plays an important role in astronomy. For instance, each element has its own spectral atom spectrum. When scaling up, the molecule spectrum and material have photophysical effect. Different materials have distinct absorption properties. The wider range and higher resolution of the spectra a sensor can cover, the more materials it can distinguish [2, 16]. Across the whole electromagnetic spectrum, the visible bands only occupy a small proportion and a large amount information existing in other bands.

[3]This figure is borrowed from `http://www.ces.fau.edu/nasa/module-2/radiation-sun.php`

A lot of physical phenomena are connected with spectral reflectance or absorption. Plants, water/moisture, and DNA all have their own spectral signatures. In Table 2.1, we show some examples of materials which can be distinguished with specific spectral range.

Table 2.1: Examples of materials and most useful spectral range for recognition.

Materials	Spectral Range
Chlorophyll absorbtion/vegetation reflectance	510-970 nm
Oxygenation in blood	500-900 nm
Hydrocarbon organic compounds	750-950 nm
Water absorption	970, 1450,1850 nm
PVC/plastic recycling	1700-1900 nm
Minerals mapping	2300-2400 nm
CH4	2300 nm
CO2	2100,3500,4800 nm

2.3 Feature Extraction in Computer Vision and Remote Sensing

Where there is an image, there are features. Feature extraction is a major research topic in computer vision. It is widely used in image matching, tracking, segmentation, compression, recognition and so on [17]. The very basic features include edges, corners, and blob in an image. Features can be invariant to many conditions, such as the scale invariant features transform (SIFT) descriptors [12] and other rotation, transmission, and illumination invariant features [18, 19, 20]. These kinds of features usually help to look for correspondences between two images containing the same object but captured under different conditions. In image classification and recognition, image and objects are usually represented by a group of features, for example,

visual words [21]. Theoretically, a feature is a pattern which represents particular property of an image and sometimes even a robust presentation of image. In this chapter, I review some local features, texture features and color features that are relevant to my research, as well as how they are extracted and described.

2.3.1 Local image feature

Local features are an important cue to predict the shape of an object [17]. Other causes include sharp changes in albedo, surface orientation or illumination. Local features are very useful because it reflects the properties of images: the change of albedo results from the texture variant; the change of surface orientation shows the shape; the illumination change is caused by the movement of light source.

Local features can be extracted at points, edges or small patches in an image. Besides their locations, details about the regions where features locate can form descriptors for other computer vision steps. Depending on the properties of local feature and their corresponding patterns in images, local features can be used in various ways. For example, they may represent specific objects in some applications, such as road recognition which can be implemented by line detection in an image [22], and impurity identification in quality inspection of glasses by detecting blobs. Local features could also possess transform invariant property, which is one of the foundations of image registration. Last, a class of objects or images need general representation for recognition and classification purposes. A set of features can play such a role and act as abstract description.

2.3.2 Texture feature

Texture refers to repeated patterns in an image rather than the structure of an object. Texture features have been used in many tasks, for example, segmentation in which local texture representation describes similar points in a region, material recognition in which pooled texture representation is used, and shape estimating such as shape

from texture [17]. The reason why texture is important is that it describes property of materials and helps to identify an object. Although traditionally extracted from grayscale and color images, texture can be associated with spectral information which is covered in this thesis.

In order to represent a local texture, two factors shall be addressed: what texture element shall be extracted and how texture is distributed. An element of texture is usually called texton. Textures vary a lot so it is difficult to predict which textons exist in an image. A solution to this problem is to split textures into subelements like spots and bars. Then subelements are detected by applying a set of filters at various scales to an image so that each of subelement is represented by the vector of the corresponding filter. At last, the summary of the output maps provides a representation of texture in the image.

2.3.3 Color feature

Color is a kind of low spectral resolution information located in red, green and blue bands, which can be used for illumination estimation, saliency detection, image segmentation, specularity and shadow removal, etc. There are several color spaces, i.e. RGB color space, CIE XYZ color space, HSV color space. Each of them represents different information and can be used in various applications. However, color itself is not capable of distinguishing different materials due to metamerism [23], i.e., the same color may correspond to different spectral power distribution.

Color feature is traditionally used with other features because it contains low descriptive information of visual contents in an image. Here, the importance of reviewing color feature is due to its direct correlation with spectra in hyperspectral images. One aspect of color feature that we can learn from is the transform of color space. After RGB images are transformed to HSV (hue, saturation, and value) space, the new color feature maintains some good invariant properties which are more useful in classification and recognition than the original feature. Another aspect is

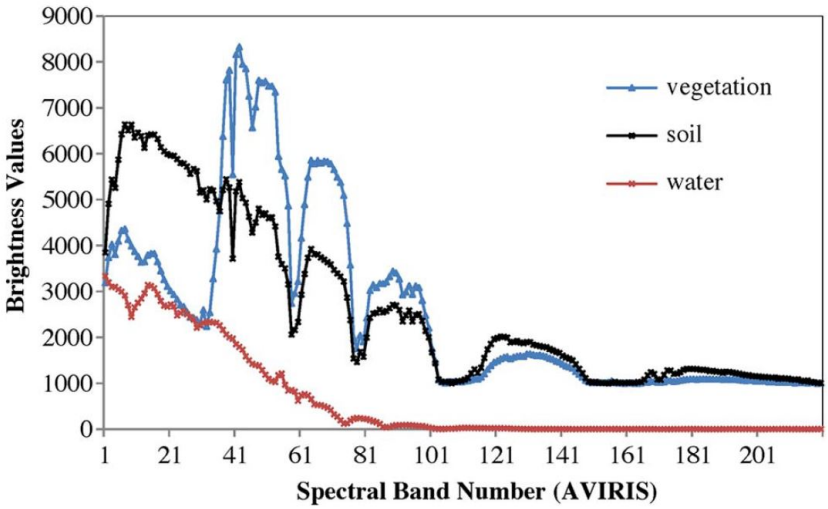


Figure 2.4: Vegetation, soil and water spectra recorded by AVIRIS [11].

the color histogram [24] which is a good color feature descriptor.

2.3.4 Spectral feature extraction in remote sensing

In remote sensing, researchers focus more on the spectral domain. It is because that, on one hand, the hyperspectral images consist of tens or hundreds of wavelength indexed bands which contain rich information of spectra; on the other hand, the spatial resolution of remote sensing image is quite low, usually measured by meters. As a result, researchers need to unmix a single pixel (which is to analyze the components of a pixel) to different components for image analysis, rather than considering shapes or textures in an image. Note that a pixel with a spectral vector is a feature itself. Therefore, most algorithms in this area are pixel based. The problem for feature extraction becomes how to use the feature efficiently rather than how to represent the feature. In order to solve this problem, a lot of research has been done in feature selection and feature extraction, including methods that are supervised and unsupervised, parametric and nonparametric, linear and nonlinear [25]. All of them try to seek the informative subspace of features.

There are a large number of bands in hyperspectral images but not all of them are

useful. Every material has its unique spectral absorption feature. However, discriminative information might only exist in a portion of wavelengths. For example, Fig. 2.4 shows that the difference of spectra between vegetation and soil mainly lies in the visible range but not the far infrared range. Even if the spectral reflectance is not similar, the spectral information may be redundant considering that many contiguous bands are highly correlated. Moreover, noises coming from imaging environment and instruments often contaminate some bands. Therefore, the raw hyperspectral bands shall be preprocessed for remote sensing applications. For the hyperspectral images captured on the ground as used in my research, this problem also needs to be considered.

A simple but useful method for dimensionality reduction is principal component analyze (PCA) [26]. Based on statistics of data, its objective is to find the most variant dimensions of data and map the original data on these dimensions so that the variance of data is maximally retained. Mathematically, given the original feature set X and its element x_i, each new feature y_i is calculated using the following equation:

$$y_i = w^T x_i \tag{2.1}$$

where the columns of w are the eigenvectors of the covariance matrix of the original feature set X. The method to calculate the covariance matrix is shown in Equation 2.2 where μ is the mean vector of X.

$$C = \frac{1}{N} \sum_{i,j} (x_i - \mu_i)^T (x_j - \mu_j) \tag{2.2}$$

As an unsupervised method, the feature extracted by PCA algorithm contains the largest variance but may not be optimal for classification. Therefore, supervised methods are employed to select features. With the training data, while parametric methods try to build models by measuring separability of two classes [2], nonparametric methods use label information directly to estimate the similarity between

classes [27].

2.4 Spectral-spatial Feature Extraction

Table 2.2: Summary of spectral-spatial feature extraction methods.

Year	Author	Feature/Classifier
2002	Q.Jackson et al.	Markov Random Field [28]
2003	D.A. Landgrebe	ECHO Classifier [29]
2005	Benediktsson et al.	Extended Morphological Profile [30]
2006	G. Camps-Valls et al.	Composite Kernels [31]
2009	Y. Tarabalka et al.	Spatial Regulation based on Segmentation [32]
2011	Y. Chen et al.	Dictionary-based Sparse Representation [33]
2011	M. Dalla Mura et al.	Extended Attribute Profile [34]
2011	L. Shen et al.	3D Gabor Wavelet [35]
2013	Y. Qian et al.	3D Discrete Wavelet Transform [15]
2013	L. Zhang et al.	Tensor Discriminative Locality Alignment [36]
2013	T. Fuai et al.	3D Gray Level Co-occurrence [37]
2014	W. Duan et al.	Superpixel Representation [38]
2014	R. Ji et al.	Spectral-Spatial Constraint [39]
2015	J. Liang et al.	3D Local Derivative Pattern [14]
2015	J. Li et al.	Multiple Feature Combination [40]

To our knowledge, existing spectral-spatial feature extraction methods are mainly developed for remote sensing and there has not been much research in computer vision on this topic. This is because computer vision based on hyperspectral imaging is still in its early research stage. Most computer vision tasks are accomplished on gray-scale images with spatial information heavily explored. Though color features are also widely used, it is less discriminative than spatial structures. Moreover, color is sometimes not reliable since it is easily influenced by the illumination condition, but the spatial information is not. For instance, edges can be consistently observed

under both sunlight and halogen light, but color is not. Due to its high spectral resolution, hyperspectral images provide much more information than color image, and its inner constraints can help to estimate the reflectance, therefore making them particularly useful for computer vision tasks [16].

The spatial information in a hyperspectral image may cover many factors. These include local information such as structures, textures, and contextual information, as well as global information such as geometric information. Due to the high dimensionality of this kind of feature, it is quite expensive to associate both spatial and spectral features at the same time. So an alternate way is to consider them separately.

The advantage of using hyperspectral data in land cover classification is that spectral responses reflect the properties of components on the ground surface [2]. Therefore, raw spectral responses can be used directly as the discriminative features of different land covers. At the same time, hyperspectral data also possesses the basic characteristic of conventional images - the spatial information which corresponds to where a pixel locates in the image. The spatial information can be represented in different forms, such as structural information including the size and shape of objects, textures which describe the granularity and patterns, and contextual information which can express the inter-pixel dependency [41]. This is also the foundation of developing spectral-spatial methods for hyperspectral image classification.

In general, spectral-spatial information can contribute to hyperspectral image classification through three ways. Firstly, in image preprocessing, it can be used for image denoising, morphology, and segmentation. Image denoising enables the reduction of random noises introduced from sensor, photon effects, and calibration errors. Several approaches have been exploited for this purpose, for example, smoothing filters, anisotropic diffusion, multi-linear algebra, wavelet shrinkage, and sparse coding methods [42]. In most cases, denoising can be done by applying a local filter with designed or learned kernel across the whole image. In mathematical morphology, operations are performed to extract spatial structures of objects according

to their spectral responses [43, 41]. Similar information is explored in image segmentation which groups spatially neighboring pixels into clusters based on their spectral distribution [44, 45].

Secondly, common usage of joint spectral-spatial information lies in the feature extraction stage. While traditional spectral features are extracted as responses at single pixel level in hyperspectral images, spectral-spatial feature extraction methods use spatial neighborhood to calculate features. Typical examples include texture features such as 3D discrete wavelet [15], 3D Gabor wavelet [46], 3D scattering wavelet[47], and local binary patterns [48]. Morphological profiles, alternatively, use closing, opening, and geodesic operators to enhance the spatial structures of objects [30, 49, 34]. Other spectral-spatial features include spectral saliency [50], spherical harmonics [51], and affine invariant descriptors [52]. Heterogeneous features can be fused using feature selection or reduction approaches [25].

Thirdly, some image classification approaches rely on spatial relation between pixels for model building. A direct way of doing so is calculating the similarity between a pixel and its surrounding pixels [53]. Markov random field, for example, treats hyperspectral image as dependent data and uses spectral information in the local neighborhood to help pixel class prediction [54, 45, 55]. Similar spatial structures are explored in conditional random fields [56], hypergraph modelling [39], and multi-scale analysis [57]. The spatial information can also be explored in constructing composite kernels in support vector machines [31]. Supervised learning approaches, such as K-nearest neighbors, linear discriminant analysis, Bayesian analysis, support vector machines, etc. are widely used in these classification tasks [58, 59]. Nonetheless, some approaches adopt semi-supervised or active learning strategies [60, 61].

2.4.1 Extended morphological profiles

Extended morphological profiles (EMP) is an extension of mathematical morphology for hyperspectral images. It solves the problem of mathematical morphology which

cannot be applied to hyperspectral images directly since there is no explicit ordering relationship in vectors. In other words, neither maximum nor minimum, which are necessary for basic morphological operations, is not defined for a set of vectors.

The details of this method and its variation can be found in a survey paper from Fauvel et al [41]. The spatial feature is extracted as follows

$$\Omega^{(n)}(I) = \left[o^{(n)}(I), ..., o^{(1)}(I), I, c^{(1)}(I), ..., c^{(n)}(I)\right] \tag{2.3}$$

where $o^{(n)}(I)$ and $c^{(n)}(I)$ are the opening and closing operations with a disk-shape structural element of size n, respectively. As different sizes of structuring elements are used, the morphological profile $\Omega^{(n)}(I)$ is capable of integrating multi-scale information. Before the feature extraction, a principle component analysis (PCA) step is applied to hyperspectral images to reduce the dimension of the data. Then the morphological profiles are obtained on each of the m primary components:

$$\hat{\Omega}_m^{(n)}(I) = \left[\Omega_1^{(n)}(I), \Omega_2^{(n)}(I), ..., \Omega_m^{(n)}(I)\right] \tag{2.4}$$

In the last step, the morphological profiles are stacked with the spectral response to form the spectral-spatial feature.

2.4.2 3D Gabor wavelet

Three dimensional Gabor filter [35] is the extension of two dimensional Gabor which is widely adopted in texture feature extraction. By tuning the scale and orientation, the Gabor Wavelet can extract multi-scale and multi-orientation features from images. A 3D Gabor filter with the frequency f and orientation (ϕ, θ) can be defined as following

$$G_{f,\phi,\theta}(x, y, b) = N \times exp(-(\frac{x'^2}{\sigma_x}) + (\frac{y'^2}{\sigma_y}) + (\frac{b'^2}{\sigma_b})) \times exp(j2\pi(xu + yv + bw)) \tag{2.5}$$

This function is a Gaussian kernel function modulated by a sinusoid plane wave. N is the normalise factor. $[x', y', b']^T$ is transformed signal which is coincide with orientation of the sinusoid. It can be calculated as

$$[x', y', b']^T = R \times [x - x_c, y - y_c, b - b_c]^T \tag{2.6}$$

where (x_c, y_c, b_c) is the central position for signals and R is the rotation matrix defined by (ϕ, θ). σ_x, σ_y and σ_b are width of Gaussian envelop in different axis. Regarding the sinusoid plane wave, w is the axis of the wave vector and $u - v$ is the perpendicular plane in frequency domain. $u = f \sin\phi \cos\theta, v = f \sin\phi \sin\theta, w = f \cos\phi$.

2.4.3 3D discrete wavelet transform

The three-dimensional wavelet texture features proposed by Qian et al [15] is a real 3D feature on hyperspectral images. In this method, hyperspectral data cube is treated as a a whole tensor and it is composed at different scales, frequencies and orientations by the 3D discrete wavelet transform. The definition of continuous wavelet transform is

$$\Psi_x^{\psi}(\tau, s) = \int x(t) \cdot \psi_{\tau,s}(t) dt \tag{2.7}$$

where $\psi_{\tau,s}$ is the basis functions (wavelet) with s and τ controlling the scale and translation, respectively. $x(t)$ is the original signal in the time domain. When it comes to discrete samples, discrete wavelet transform (DWT) is implemented by a series of filters in the frequency domain. Since hyperspectral images consist of three dimensions, 3D-DWT exploits the correlation along the wavelength axis, as well as along the spatial axes, so that both spatial and spectral structures of hyperspectral images can be adequately mapped into the extracted features.

The implementation of 3D-DWT has several steps. Firstly, the hyperspectral image is processed by a cascade of high pass filters and low pass filters. At each level, the data is decomposed into high-frequency part and low-frequency part. After three

levels of decomposition, the original data is separated into 15 sub-cubes $C_1, C_2, ..., C_{15}$ based on the bandwidth, such that each of the sub-cubes contains different scales of information. To further capture the spatial distribution of hyperspectral images, a mean filter is applied on the sub-cubes:

$$\hat{C}_n(x, y, .) = \frac{1}{9} \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} C(i, j, .) \tag{2.8}$$

where $C(i, j, .)$ is the spectral response at the position of (i, j). $[x-1, x+1]$ and $[y-1, y+1]$ define a 3 by 3 neighbourhood.

In order to keep the sub-cube and the original data cube at the same size, the filtered signals are not down-sampled as what the traditional DWT does. Then these sub-cubes are concatenated into the wavelet features. The multidimensional function is carried out along two spatial dimensions x and y, as well as the spectral dimension λ, respectively. The final concatenation work as the feature for the whole data cube and can be represented as:

$$f(x, y) = (\hat{C}_1^x, \hat{C}_2^x, ..., \hat{C}_{15}^x, \hat{C}_1^y, \hat{C}_2^y, ..., \hat{C}_{15}^y, \hat{C}_1^\lambda, \hat{C}_2^\lambda, ..., \hat{C}_{15}^\lambda) \tag{2.9}$$

where $f(x, y)$ is the 3D-DWT feature at location (x,y).

2.4.4 Tensor modeling

Recently, tensor decomposition methods have been employed in hyperspectral image classification. A tensor is a multidimensional array and number of dimensions is denoted as the order or mode. In Fig. 2.5, we show an example of third-order tensor in different formats including the cube, fibers, and slices [62]. Fibers are the higher-order analogue of matrix rows and columns. Fixing every index but one forms a fiber. Similarly, slices are two-dimensional sections of a tensor and fixing every index but two forms a slice. Hyperspectral images can be treated as a third-order tensor so that tensor methods can help analyze the hyperspectral data. A hyperspectral image can

be arranged as a three-dimensional array with the modes corresponding to spatial x spatial y and wavelength λ. Zhang et al. [36] represented the spectral-spatial feature of a pixel as a tensor. The feature was constructed by the spectra of the pixel itself as well as its k nearest neighbors. Then a tensor discriminative locality alignment (TDLA) was developed to transform the high-order tensor space to a low dimensional feature space via patch optimization which included the information of same and different classes in the training data. Its superiority over several unsupervised and supervised feature dimension reduction methods was proved on three commonly used hyperspectral datasets. Meanwhile, Velasco-Forero et al. [63] developed an additive morphological decomposition (AMD) which decomposed an image into two parts. While the first part was formulated by the summation of anti-extensive transformations and extensive transformations, representing all contrast and boundary information, the second part was related to the residue of anti-extensive transformations and extensive transformations with different scales, constituting a hierarchy of multiscale texture components. Then tensor modeling methods were employed to reduce the dimension of extracted features in which the spatial information was maintained.

2.4.5 Other spectral-spatial operations

Other spectral-spatial operations can also be used either before or after the classification to improve the classification accuracy. Image segmentation partitions an image into multiple groups of pixels based on their local properties. Naturally, it groups the nearby pixels together and embeds spatial information. When it applies to a hyperspectral image, the segmentation algorithms can make use of spectral information and create more accurate regions than RGB images. On these partitioned regions, feature extraction can combine spectral and spatial information. Extracted features can be the mean of the spectral responses in the region or other combined spatial and spectral features. After each pixel in a region is assigned a label with classification

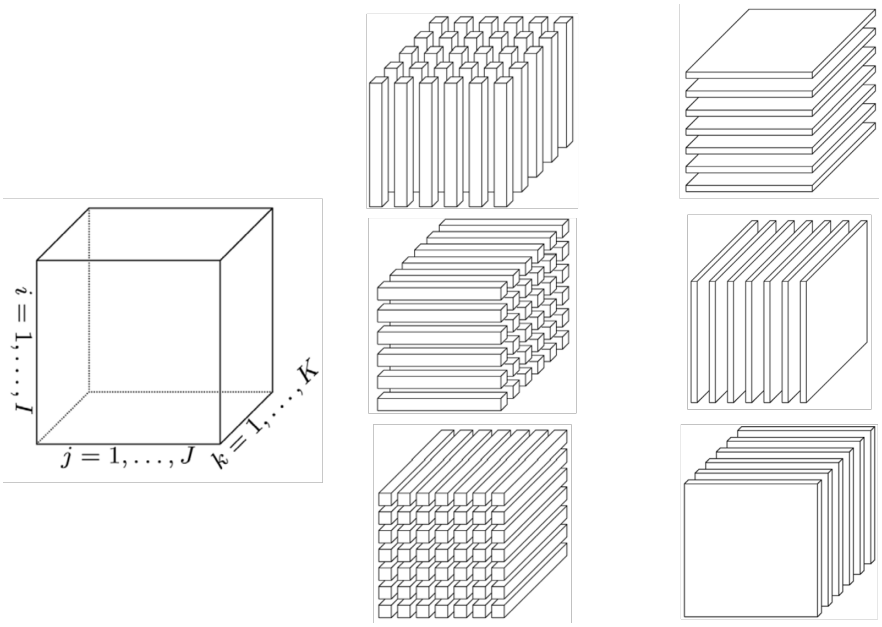


Figure 2.5: Tensor structure in different formats. From left to right: cube, fibers, slices (modified from [62]).

algorithms, major voting can be used to reassign the labels of pixel wise classification by counting the most frequent labels in the region so as to create a smooth classification map. Although this method is simple and straightforward to associate the spatial information with spectral information for hyperspectral image classification, the classification results heavily depend on the segmentation algorithms.

Alternatively, Markov Random Field (MRF) is popular as a post-processing step for hyperspectral image classification. It is a theory to model the distribution of a set of random variables in 2D space satisfying the Markov property that joint distribution of X is determined entirely by its local conditional distributions [64]. This property makes it suitable to describe the spatial context in images. In remote sensing, applying MRF in pixel wise classification takes advantage of the dependence among neighboring pixels to achieve a smooth classification thematic map [54, 65]. In this framework, the probability of the label of a pixel depends not only on its spectral response but also the labels of its neighboring pixels. Due to Markov-Gibbs equivalence, the maximization of the global posterior probability mass function $P(Y|X)$ is

equivalent to the minimization of the energy function $U(Y|X)$.

2.5 Classification Methods

Effective and efficient classifiers are very important for hyperspectral image classification. The inputs to classifiers are normally extracted features or image representation. Depending on the availability of the labeled data, classification methods can be categorized into unsupervised learning and supervised learning. Unsupervised classification methods can be used to discover groups of similar samples in the dataset based on statistics. Clustering is a typical unsupervised classification method. The most popular clustering algorithm is K-means clustering [66]. It adopts an iterative updating methods to partition data samples into k clusters in order to minimize the within class distance. On the contrary, supervised learning methods use labeled data to train a classifier and then apply it to unseen data. It is usually more reliable than unsupervised methods. In this thesis, we mainly focus on supervised learning methods.

For hyperspectral image classification, there are only limited labeled data available for training and testing due to lacking of benchmark datasets and high cost of ground truth data collection. Therefore, not all of supervised learning methods are suitable for such task, especially those require a large amount of training data such as deep learning approaches. We briefly introduce three typical methods commonly used in hyperspectral image classification: support vector machine, random forest, and extreme learning machine.

2.5.1 Support vector machine

Support vector machines (SVM) has been widely used in classification tasks [67]. It aims at estimating a hyperplane(s) to classify data into two classes such that the maximum margin between the hyperplane and training samples is achieved. The su-

[4]This figure is from `https://en.wikipedia.org/wiki/Support_vector_machine`

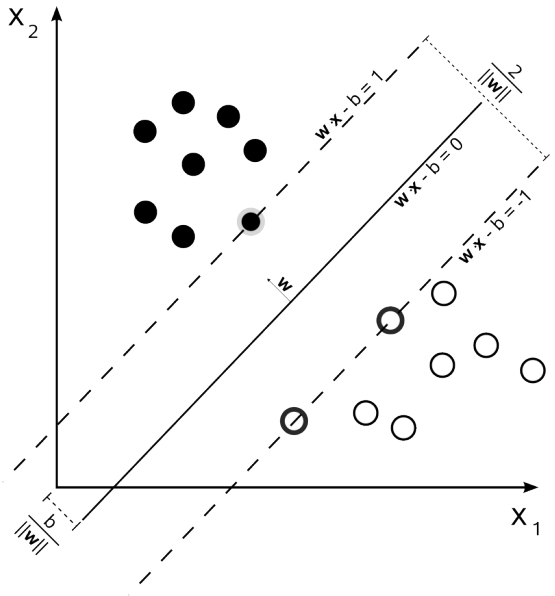


Figure 2.6: Maximum-margin hyperplane for a two-class SVM. White and black dots represent sample from class "+1" and "-1", respectively[4].

periority of SVM over the old statistical learning method is that it introduces the concept of geometrical margin that involves only a few training samples at the boundaries (support vectors), which makes it very suitable to deal with small training set problem. Fig. 2.6 shows an example to explain this concept. Given a binary classification problem in two-dimensional feature space X_1 and X_2. There are samples represented as white and black dots from class "+1" and "-1", respectively. In order to fully separate two classes, the optimal hyperplane should find those training samples near boundaries and maximize the distances between them and hyperplane. If the hyperplane can be defined as $w \cdot x - b = 0$, then the total distance equals to $\frac{2}{\|w\|}$.

Mathematically, the optimal hyperplane can be solved by the following linearly constrained optimization problem:

$$\begin{aligned} \underset{w,\ \xi}{\text{Minimize}} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\ \text{Subject to} \quad & y_i (w^T \phi(x_i) + b \geq 1 - \xi_i) \\ & \xi_i \geq 0 \end{aligned} \tag{2.10}$$

where ξ is slack variables introduced to account for the nonseparability of data, and the constant C is a regularization parameter which controls the penalty assigned to errors. For the linear non-separable problem, $\phi(x_i)$ is a kernel function, mapping data into a high dimensional space. The kernels can be polynomial, Gaussian radial basis function, and sigmoid function. A widely used Matlab toolbox for classification can be found at [68].

2.5.2 Random forest

Algorithm 1 Random forest algorithm

Require: Training set, the number of trees k, the number of features m
  Randomly sample a subset S of training set N times
  for each subset S_i in N **do**
    Build a decision tree T
    for each node n in T **do**
      Randomly select m features and split the tree
      Stop until the tree is fully grown and not pruned
    end for
  end for
  Combine all k trees by major voting to construct the classifier

Random forest is an ensemble learning method that combines multiple decision trees to construct more powerful classifier than any individual tree [69]. It has drawn increasing interest in hyperspectral image classification [70, 71, 72]. The advantage of this method is that it can avoid the overfitting problem even if the feature dimension is high. Unlike conventional decision trees which may overfit the training dataset, random forest uses multiple subsets of the training data and build multiple decision trees, thus to avoid the correlation of multiple trees. Furthermore, instead of using all feature variables, it only uses some of the features when splitting the trees. The algorithm is fast to implement and can deal with large-scale datasets with high dimension variables.

The procedure of random forest is showed in Algorithm 1. Two inputs need to be defined, the number of trees k and the number of features m. A new subset of

data is drawn from training set with replacement. A decision tree is learned from each subset. In each split of the procedure, only m randomly selected variables are considered. The iteration stops when the tree is fully grown. The output is an ensemble tree which is used to make prediction for new data.

2.5.3 Extreme learning machines

In recent years, neural network methods, especially deep learning, regain prosperity in the machine learning community. Huang et al [73] proposed extreme learning machine (ELM) which is a two-layer feedforward neural network that provides excellent classification performance without long training time. The characteristics of ELM make it qualified for the hyperspectral image classification in which only a limited training data are available, but massive data needs to be processed [38, 48].

ELM employs a hidden layer to connect the input and output layers where the input and output correspond to the feature and regression or classification results. Different from the networks trained using back-propagation, ELM does not update the weights between the input and hidden nodes and only learns the weights connecting to the output. It is claimed that any nonconstant piecewise continuous function can be approximated by such a model without tuning the hidden node parameters [74]. This mechanism makes it much faster than either SVM or deep learning.

Chapter 3

Salient Object Detection in Hyperspectral Imagery

3.1 Introduction

A hyperspectral image consists of tens or hundreds of contiguous narrow spectral bands. Each pixel in a hyperspectral image is a vector of spectral responses across the electromagnetic spectrum (normally in the visible to the near-infrared range). Such spectral responses are related to the material of objects in a scene that has been imaged, which provides valuable information for automatic object detection.

Due to its high dimensionality, traditional pattern recognition and computer vision technology cannot be directly applied to hyperspectral imagery. Most object detection methods for hyperspectral images are still pixel-based, i.e., performing pixel-wise detection and classification based on spectral signatures followed by post-processing to group pixels or segment regions from an image [7, 44]. In this manner, feature extraction is only performed in the spectral domain, but the spatial distribution of objects has not been fully explored. More recently, researchers have tried to use spectral-spatial structure modeling for hyperspectral image classification. Such efforts include Markov random field and conditional random field [56, 45], which introduces spatial information into classification steps using a probabilistic discriminative function with contextual correlation. Furthermore, multi-scale time-frequency signal analysis methods based on 3D discrete wavelet transform have also been in-

troduced for object detection and classification in remote sensing imagery [15].

Visual saliency is another type of approaches to extract multi-scale image features. The concept of saliency is from human attention model, which detects objects or regions in a scene that stand out with respect to their neighborhood [75]. As a consequence, saliency detection models are normally established on the trichromatic or greyscale images, which are visible to human eyes. When used for object detection in computer vision and robotics applications, saliency map is often constructed in a bottom-up manner. For example, Itti et al. calculated multi-scale differences of intensity, color, and orientation features, and linearly combined them to form the final saliency map [76]. Liu et al. formulated the saliency detection problem as a region of interest segmentation task [77]. Salient features were extracted at the local, regional and global levels, and were combined via learning with conditional random field. Similarly, many saliency detection methods try to detect image regions that are different from its neighborhood in the scale space, as reviewed in [75].

When applied to hyperspectral imagery, saliency model has been used for image visualization. Wilson et al. employed contrast sensitivity saliency to fuse different bands of hyperspectral remote sensing images so that it can be used for visual analysis [78]. Itti's model [76] has been combined with dimensionality reduction method to convert a hyperspectral image to a trichromatic image that can be displayed on computer screen [79]. Saliency has also been used to help edge detection and to predict eye fixation on hyperspectral images [80, 81].

Despite its success in object detection on RGB images, as far as we know, saliency has not been used for object detection in hyperspectral imagery. Therefore, the contribution of this chapter is to explore how salient regions can be extracted from hyperspectral images, and then be used for object detection. Compared with traditional pixel-level operations, this chapter introduces a novel region-based approach for hyperspectral object detection. We propose three methods based on Itti's saliency detection model which is a benchmark method. The first method converts a hy-

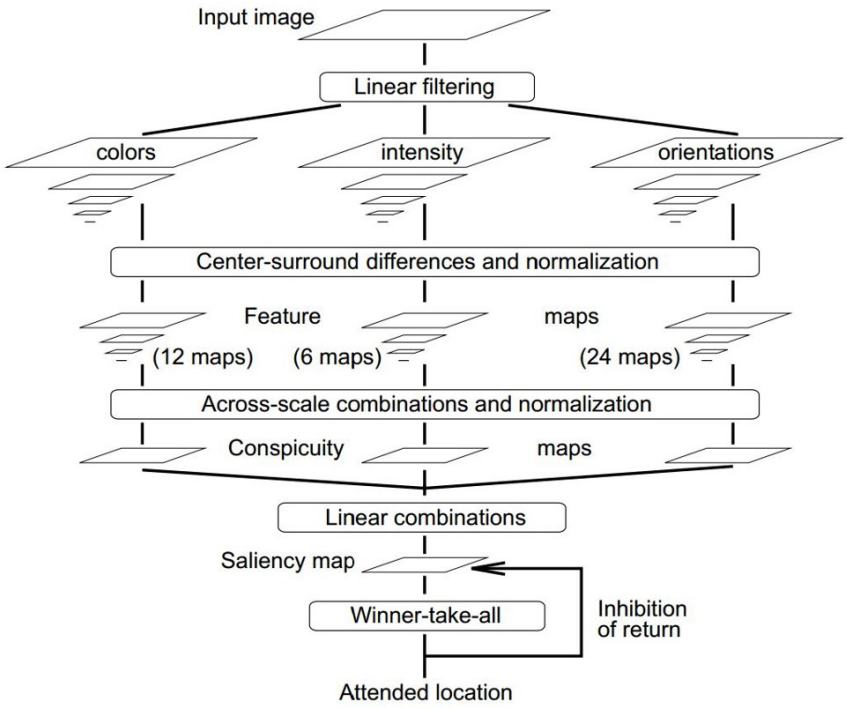


Figure 3.1: The architecture of Itti's saliency model [76].

perspectral image into an RGB image and applies Itti's model directly. The second method replaces the color double-opponent component with grouped band component. The third method directly uses the raw spectral signature to replace the color component. The last method creates a spectral-spatial distribution and uses it for saliency prediction.

3.2 Itti's Saliency Model

The saliency detection method proposed by Itti et al. mimics the behavior and structure of the early primate visual system [76]. It extracts three types of multi-scale features, including intensity, color, and orientation, and then computes their center-surround differences. These differences are linearly combined to form the final saliency map.

As shown in Fig. 3.1, this method is processed as follows. An input image is firstly smoothed using low-pass filters so as to generate nine spatial scales. Three types

of visual cues are then extracted from the intensity, color, and orientation features and each of them forms a conspicuity map. The intensity feature is obtained by averaging the RGB channel values at each pixel. By calculating the differences between each pair of fine and coarse scales, six intensity channels are generated. The second set of features are calculated from a set of color opponency between red, green and blue values against yellow value at each pixel. Center-surround differences for each pair of color opponent are then derived over three scales, which leads to 12 channels. The orientation features are generated using a set of even-symmetric Gabor filters. The dominant orientation at each pixel is recovered, whose center-surround differences are calculated at six scales and four orientations. This leads to 24 orientation channels. Channels of each type are then linearly combined to form three conspicuity maps in terms of the intensity, color and orientation. Finally, the mean of the conspicuity maps becomes the saliency map.

3.3 Saliency Extraction in Hyperspectral Images

3.3.1 Spectral saliency

3.3.1.1 Hyperspectral to trichromatic conversion

What has hindered the adoption of saliency extraction by hyperspectral object detection is the large amount of bands in the spectral data. This makes the color component not able to be calculated directly. Furthermore, effective extraction of the intensity and texture saliency requires a grayscale image to be used. A direct solution to this problem is conversion of a hyperspectral image into a trichromatic image, which allows traditional saliency model to be applicable. As pointed out in [79], this can be achieved by dimensionality reduction, band selection, or color matching functions.

In this research, we have followed the method of Foster [82]. This method first converts the hyperspectral image to a CIE XYZ image. Given a hyperspectral image

$I(\lambda_i)$ for each of the bands λ_i, such conversion can be implemented by the following color matching function:

$$I_t = \sum_{i=1}^{N} I(\lambda_i) W_t(\lambda_i) \tag{3.1}$$

where N is the total number of bands, $t = \{X, Y, Z\}$ are the tristimulus component of the color space, and W_t comes from the spectral sensitivity curves of three linear light detectors that yield the CIE XYZ tristimulus values X, Y, and Z. This conversion is then followed by a further transform step to the sRGB color space [83], then Itti's method can be applied.

Figure 3.2: Spectral band group.

3.3.1.2 Spectral band opponent

Although the above method is straightforward, it does not take advantage of the high spectral resolution information provided by the hyperspectral image. Notice that the second conspicuity map in Itti's method is formed from RGB color channels, we shall be able to replace the color opponents with groups spectral bands that are approximately correspondence to these color channels. Furthermore, the yellow channel in Itti's method is extracted from RGB channel. For a hyperspectral image, we can group spectral bands to represent the original multichannel color information. To do so, we divide the bands into four groups with each group occupying approximately the same width of visible spectrum, as shown in Fig. 3.2. Then the original single value color component is replaced by a vector, so the double opponency can be computed as follows

$$Opp_1(c, s) = |(G1(c) - G3(c)) \ominus (G3(s) - G1(s))|_1 \tag{3.2}$$

$$Opp_2(c,s) = |(G2(c) - G4(c)) \ominus (G4(s) - G2(s))|_1 \tag{3.3}$$

where G1 to G4 are vectors whose entries are extracted from the corresponding group of spectral bands, c and s are two scales for the cross-scale difference computation, and $|.|_1$ is the 1-norm of a vector. $\ominus$ is the cross-scale center-surround difference operator as defined in [76]. Then Opp_1 and Opp_2 can replace the red/green and blue/yellow opponency in [76]. In this method, the intensity and orientation maps can be extracted from the grayscale image converted from the trichromatic image generated using the color matching function as described in section 3.3.1.1.

3.3.1.3 Spectral saliency with Euclidean distance and spectral angle distance

Further extending the method in section 3.3.1.2 allows the using of whole spectral responses for saliency detection. When replacing the color saliency with spectral saliency, the spectral relationship embedded in the spectral data can be fully explored. Following the general multi-scale operation, differences between the spectral responses and its neighborhood can be calculated. Regarding the measurement of vector distance, both Euclidean distance (EUD) and spectral angle distance (SAD) can be used to measure the similarity between two spectral vectors s_k and s_j. The EUD is computed via the following equation:

$$d_{\text{EUD}}(s_k, s_j) = \sqrt{\sum_{i=1}^{n} (s_k^i - s_j^i)^2} \tag{3.4}$$

where n denotes the dimension of the vector. We can calculate SAD by:

$$d_{\text{SAD}}(s_k, s_j) = \arccos\left(\frac{s_k^T s_j}{\|s_k\| \|s_j\|}\right) \tag{3.5}$$

After the distance between two spectral responses is defined, we follow the workflow of intensity feature extraction by calculating the differences between each pair of fine and coarse scales in hyperspectral images. In the original method, conspicuity

computed from intensity reflects the outstanding features from brightness. The above step leads to a set of center-surround spectral differences in the scale space. They can be combined into a spectral conspicuity map, which is used with the intensity and orientation conspicuity maps to form the final saliency map. The incorporation of spectral data suggests that not only the visual clue has been extracted, i.e., from the color and orientation contrast, but also the intrinsic material property of objects. This has provided visual saliency model with additional information beyond the capability of human and traditional camera vision. Furthermore, the SAD and Euclidean measures provide two pieces of spectral distance information that is useful for object detection.

3.3.2 Spectral-spatial distribution

The above features aim at replacing the color channel with spectral information and does not consider the spatial distribution of saliency objects. Based on Liu's paper [77], it is observed that the color component of salient object tends to concentrate in spatial domain but the background objects usually have color distributed across the image. Consequently, the variance of global distribution of color components can be used as a clue of the extent of saliency for an object. This characteristic of saliency suits hyperspectral images very well considering its high discriminative ability to the intrinsic material related properties of objects. Instead of estimating the color spatial distribution, we propose the spectral-spatial distribution to measure the saliency.

To estimate the spectral-spatial distribution, we firstly need to group the hyperspectral responses in an image into several representative components so as to reduce the parameters to be estimated. To do so, we employ hyperspectral unmixing approach to decompose the hyperspectral image. The unmixing theory makes the assumption that the spectral response at a pixel is a mixture of light scattered by substances located in the field of view [84]. Therefore, a hyperspectral image can be separated in a collection of spectral signatures, called endmembers, which represents

the pure materials, and a set of fractional abundances illustrating the percentage of each endmember. For instance, a hyperspectral image Y with n bands and m pixels is decomposed as:

$$Y^{n\times m} = M^{n\times b} \cdot R^{b\times m} + E^{n\times m} \tag{3.6}$$

where each column of M is an endmember with spectral response, each row of R is a vector of the abundance of each endmember at a pixel, b is the total number of endmembers, and E is the residue error.

Researchers have developed a number of unmixing methods such as N-FINDR [85], Pixel Purity Index, Vertex Component Analysis (VCA) [86], nonnegative matrix factorization (NMF) [87], etc. Among them, we adopt VCA for unmixing. In addition to its high efficiency, this method assumes the presence of at least one pure pixel per endmember in the hyperspectral image, which is consistent with the close-range hyperspectral image which has high spatial resolution [84]. Theoretically, VCA is based on geometrical orthogonal projection. It estimates the endmember one by one by projecting the data to orthogonal subspaces where new endmember corresponds to the extreme of the projection. Then it iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined until all endmembers are exhausted.

To help understand the unmixing process, we show the unmixing results of a hyperspectral scene in Fig. 3.3. The original hyperspectral image is converted to a color image shown in Fig. 3.4. For each endmember in Fig. 3.3, the first row shows it spectral responses, and the second row shows its abundance at each location. The brighter the abundance map is, the higher percentage of this particular endmember at the location. It can be seen that the endmembers relate to strong semantic meaning, i.e., roughly correspond to pink petals, white petals, specularity, leaves, and stamens.

The abundance map can be used to estimate the probability of each endmember at particular location so as to calculate the spectral-spatial distribution. The probability

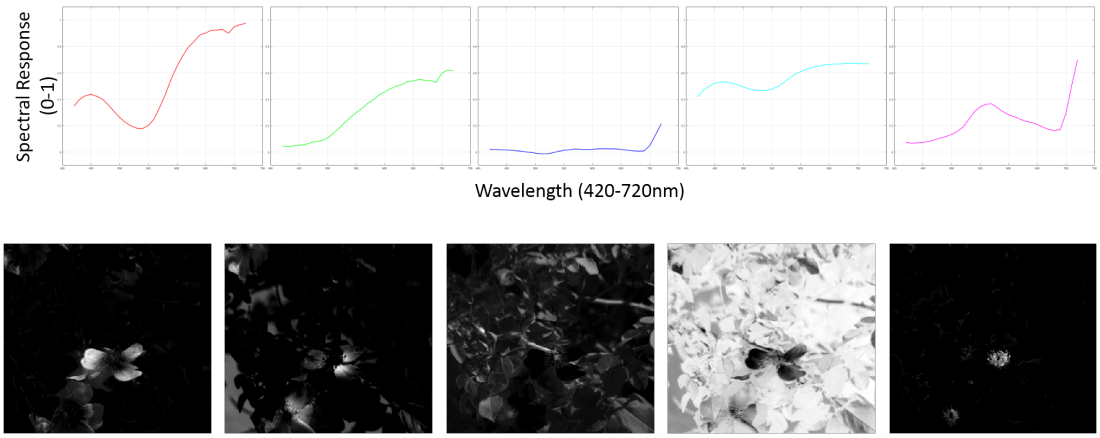


Figure 3.3: Unmixing results for a hyperspectral scene. The first row shows endmembers and the second row illustrates the corresponding abundance maps.

Figure 3.4: Conspicuity maps built from spectral-spatial distribution of a hyperspectral scene.

of an endmember c at a pixel u can be calculated as follows

$$p(c|(u_x, u_y)) = R(c, u_x + u_y * l); \tag{3.7}$$

where (u_x, u_y) is the coordinates of pixel u and l is the width of the image. $u_x + u_y * l$ represents the index of pixel after the pixels in the image are rearranged into a vector in the unmixing step. The variance of spatial distribution of component c is the

combination of variances along x and y dimensions which are calculated as

$$\begin{aligned}\sigma_x(c) &= \frac{1}{\mathcal{N}} \sum_i^n p(c|(u_x^i, u_y^i)) \cdot |u_x^i - \bar{m}_x(c)|^2 \\ \sigma_y(c) &= \frac{1}{\mathcal{N}} \sum_i^n p(c|(u_x^i, u_y^i)) \cdot |u_y^i - \bar{m}_y(c)|^2\end{aligned} \tag{3.8}$$

where $\mathcal{N}_x$ is the normalised factor which equals to $\sum_i^n p(c|(u_x^i, u_y^i))$. $\bar{m}_x(c)$ is the average x coordinate for component c and can be calculated as $\frac{1}{\mathcal{N}} \sum_i^n p(c|(u_x^i, u_y^i)) \cdot u_x^i$.

After combining $\sigma_x(c)$ and $\sigma_y(c)$ into $\sigma(c)$, we can calculate the conspicuity map from the spectral-spatial distribution. The conspicuity value at each pixel u is in accordance to the probability of u belonging to component c and the spatial variance of component c. To reduce the effect from the uncommon components of small objects near boundaries, we also weight the conspicuity value based on the distance between a pixel and the image center d_u. As a result, the conspicuity map based on spectral-spatial distribution is formulated as:

$$f(u) = \sum_c p(c|u)(1 - \sigma(c))(1 - D(c)) \tag{3.9}$$

where $D(c) = \sum_u p(c|u) d_u$, assigning less importance to the components near boundaries. Both $\sigma(c)$ and $D(c)$ are normalized to $[0,\ 1]$.

We draw the conspicuity map from the spectral-spatial distribution for a hyperspectral scene in Fig. 3.4. From Figs. 3.3 and 3.4, it can be observed that petals and stamens stand out in the conspicuity map as their spectral signatures concentrate in the spatial domain. On the contrary, leaves are suppressed due to the wide spreading of their corresponding spectral signatures.

Essentially, the spectral-spatial distribution reflects the material composition in a scene. The less frequent a material presents and more concentrated its spatial distribution is, it is more likely that material belongs to salient object. On one hand, the assumption of this method is consistent with our instinct. On the other hand, since it takes the advantage of hyperspectral imaging, it can extraction information

beyond human vision. Therefore, we expect this method can create more accurate estimation of the saliency map compared to the traditional methods based on color information.

3.4 Object detection

In the previous step, saliency map can be generated to highlight image regions that are different from their surrounding areas. To detect a region with a salient object, we binarize the saliency map using the optimal threshold recovery method [88]. This allows the pixels with low saliency values be removed. Then a set of morphological operations are used to fill the small holes in the connected components. Those small components that come from noisy clusters are removed.

The object detection follows a winner-take-all strategy, i.e., assuming that there is only one salient object per image. The remaining image regions that contain the highest value in the saliency map is selected as the one that contains the target object. It should be noted that this method can be easily extended to detect more than one objects by sequentially selecting regions in order of their highest saliency values.

3.5 Experiments

To compare the four salient object detection methods introduced in Section 3.3, we have performed experiments on close-range hyperspectral images from two online datasets. The first dataset was collected by Foster et al. [89, 82]. It contains in total 55 hyperspectral images of natural scenes, with 16 images available online for free access. The images consist of 33 bands, covering the spectral range from 400 to 720nm. Due to the low signal to noise ratio at two ends of the range, the first and last bands are removed for some scenes. The second dataset consists of 50 hyperspectral images collected at Harvard University [90], which includes images captured in both indoor and outdoor settings. The wavelength that this dataset covers is from 420 to

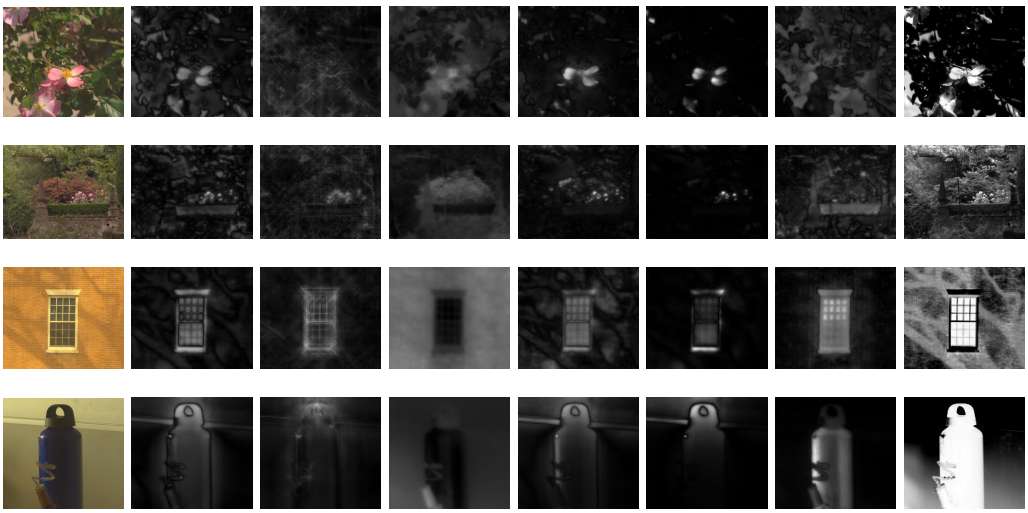


Figure 3.5: Conspicuity maps computed from different methods. From left to right: input image, conspicuity maps from intensity (I), orientation (O), trichromatic color (C), spectral band opponent (G), spectral Euclidean distance (E), spectral angle distance (A), and spectral-spatial distribution (S).

720nm, containing 31 bands.

It should be noted that these two datasets were not collected specifically for saliency object detection purpose. Therefore, in most of the images, it is hard to find salient objects or the scenes are cluttered with many objects. We have carefully selected image regions that contain salient objects in their surroundings from 13 images in these datasets for our experiments. To provide the ground truth, we have manually labeled the location of salient objects by bounding boxes. Because spectral saliency is not directly observable to human eyes, we have combined visual saliency on synthesized RGB images and domain knowledge on the object materials for judgment.

We have implemented both Itti's method and the proposed methods. The first method based on converted RGB images consists of three conspicuity maps computed from intensity, orientation, and color, respectively. We denote them as "I", "O" and "C". The second method based on grouped spectral band opponent is named "G". We have implemented two versions of the third method, i.e., using Euclidean distance of spectral response "E" and the spectral angle distance "A". The last method

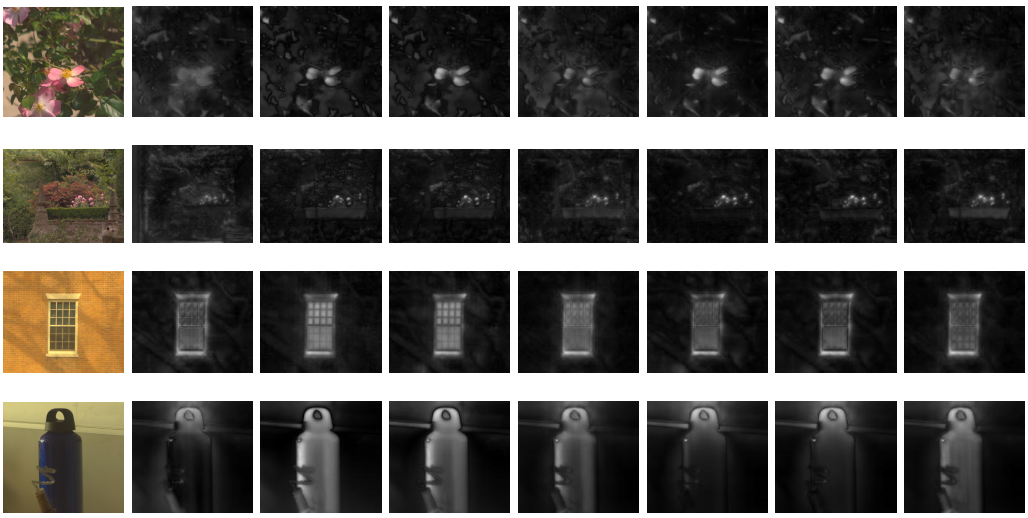


Figure 3.6: Saliency maps computed from different combinations of conspicuity maps. From left to right: input image, IOC, EAS, EASG, EOA, EOG, IOE and the method in [79].

employed is the spectral-spatial distribution based method and denoted as “S". We have calculated the above conspicuity maps for the all 13 scenes, four of which are shown in Fig. 3.5. From this figure, we have a few observations. Firstly, if comparing the third column with fourth to eight columns, the conspicuity maps from spectral bands are generally better than those from trichromatic color. The method “C" completely fails to highlight the objects in the third and fourth scenes. Furthermore, results from methods “E" and “A" are complementary to each other. They are able to emphasize different parts of the object in the scene. It is probably because they make use of different characteristic of distance measurements. At last, method “S" achieves excellent results. Although it mistakenly highlights the leaves on the top of the second scene, it successfully extracts the whole bottle in the fourth scene. Overall, this experiment indicates that the spectral bands based saliency detection methods perform better than the traditional one based on trichromatic color.

The saliency maps are usually generated by linearly combining several conspicuity maps. Here, we have tested a series of different combinations of conspicuity maps from intensity, orientation, color, group opponents, Euclidean distance, spectral an-

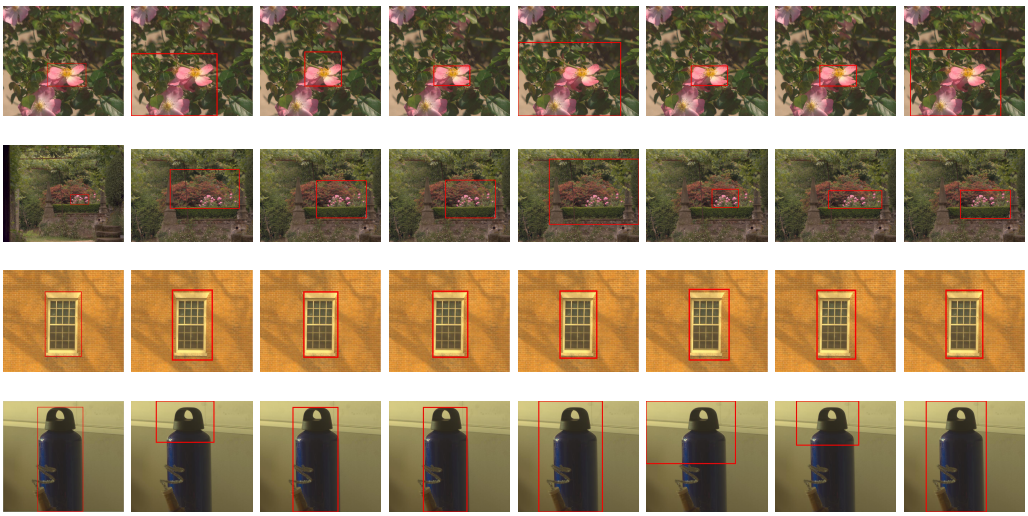


Figure 3.7: Salient object detection results. From left to right: manually labeled image, results computed from IOC, EAS, EASG, EOA, EOG, IOE and the method in [79].

gle distance and spectral-spatial distribution. For the sake of briefness and clearness, we denote these methods with the abbreviation introduced for the conspicuity map. For instance, the saliency map associating the conspicuity map from the intensity, orientation, and color is denoted as “IOC". The saliency map using the group opponents, Euclidean distance, spectral angle distance and spectral-spatial distribution is denoted as “EASG". Therefore, the methods tested include IOC, EAS, EASG, EOA, EOG, IOE. These methods are compared against an alternative method from Moan et al. [79]. The method from Moan et al. firstly combines spectral channels into red, green, and blue groups, then computes the spectral differences on each of the color groups to get the spectral saliency. The orientation saliency is extracted on the first principle component generated by principle component analysis to reduce the dimensionality of the hyperspectral image, while the intensity saliency is not used.

Fig. 3.6 and 3.7 display the saliency map and corresponding object detection results generated by each method under comparison. It can be observed that spectral information based methods can generate better saliency features than that based on color information. After combining all spectral-based methods, EASG achieves the

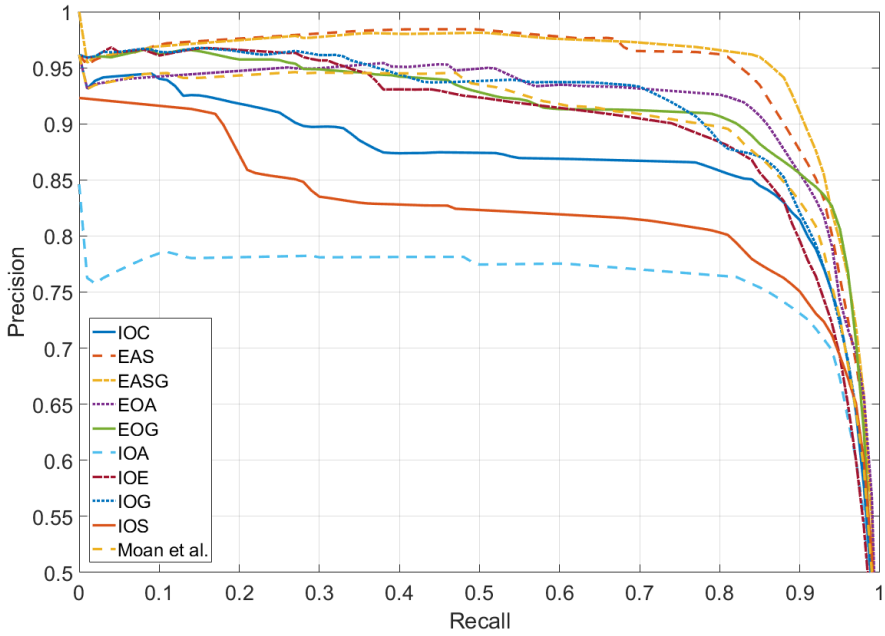


Figure 3.8: Precision recall curves computed from different saliency detection methods.

best result. When it comes to object detection, the spectral-based methods can detect the truth object more accurately, especially EAS and EASG. In contrast, IOA either includes plenty of background regions into the results or detects parts of the salient object. Moan's method, although achieved good results, also includes some background regions into the final results.

To provide quantitative analysis to the saliency object detection methods, we have calculated the precision and recall curve when different binarization thresholds are used. The results are shown in Fig. 3.8. The precision is the percentage of true object pixels out of all detected pixels. The recall calculates the percentage of true object pixels that have been detected. It can be seen that the spectral saliency methods are clearly better than the method based on color images. The performance of the spectral-based solutions are very close to each other, with the EASG option outperforming the others. Furthermore, we have calculated the F-Measure to evaluate different methods. F-Measure is a comprehensive used evaluating indicator to evaluate the quality of a saliency map. It is calculated as a weighted harmonic mean

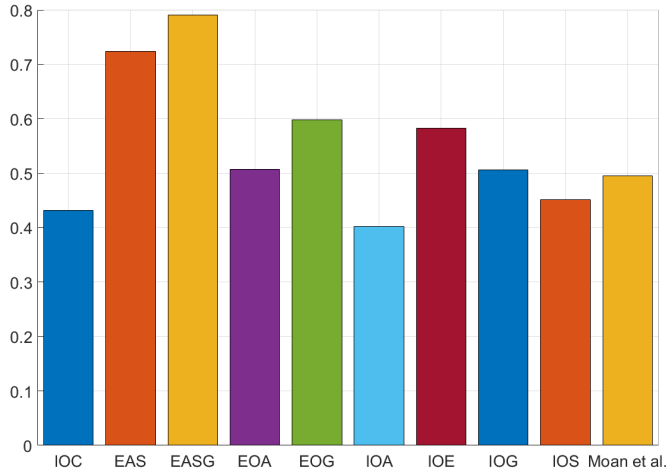


Figure 3.9: F-Measure computed from different saliency detection methods.

of precision and recall. For each scene, we can obtain an F-Measure value on the binary map of saliency map. Then the final F-Measure is computed as the average F-Measure value in the datasets. The results are shown in Fig. 3.9. It can be noted that EAS and EASG achieved significantly higher value than the rest of methods. It is consistent with our previous observations that the proposed features E,A,S and G have better prediction on the salient objects than traditional methods I, O and C. These experimental results show the advantage of combining different spectral and spatial based measures for saliency detection.

3.6 Conclusion

We have developed two strategies to generate saliency map from hyperspectral imagery for object detection. One is to extend Itti's visual attention model to hyperspectral images by replacing the color component with spectral saliency, which can be implemented by dividing visual spectrum into groups, or use the whole spectral responses. The second method makes use of variance of the spectral-spatial distribution of endmembers to estimate the saliency. The endmembers and their abundance maps are estimated using hyperspectral unmixing method. These methods allow extra information from spectral data to contribute to the traditional visual atten-

tion model. Experiments have shown the effectiveness of the proposed methods for salient object detection. In the future, we will apply the method to more hyperspectral data, for example, on remote sensing images. We will also incorporate other saliency models into the hyperspectral object detection tasks.

Chapter 4

3D Local Derivative Pattern for Hyperspectral Face Recognition

4.1 Introduction

Although a large number of techniques have been developed for face recognition by introducing new feature extraction and pattern recognition methods [91, 92, 93], existing methods still cannot meet the requirements from many real-world applications. This is partly due to the drawback of only using 2D spatial information for face recognition. To address this problem, one solution is to introduce sensors that can capture additional information other than radiance in the visible spectrum. Examples include range sensors such as Kinect and 3D scanners to capture three-dimensional geometrical information to improve the discrimination of individual faces under different orientation [94, 95, 96]. Another promising technique is to exploit spectral imaging devices to bring in the spectral responses of faces which are supposed contains more information than RGB or grayscale images. Such studies include infrared imaging, hyperspectral imaging [97, 98, 99], etc.

Studies have proved that human skin retains its own spectral properties due to the portion of melanin and hemoglobin [100]. For example, the differences in portion of melanin between African, Asian, and Caucasian lead to obvious distinction in the darkness of skin. Regarding individuals, this property could be influenced by the molecular composition related to tissue, blood, and organizational structure. Con-

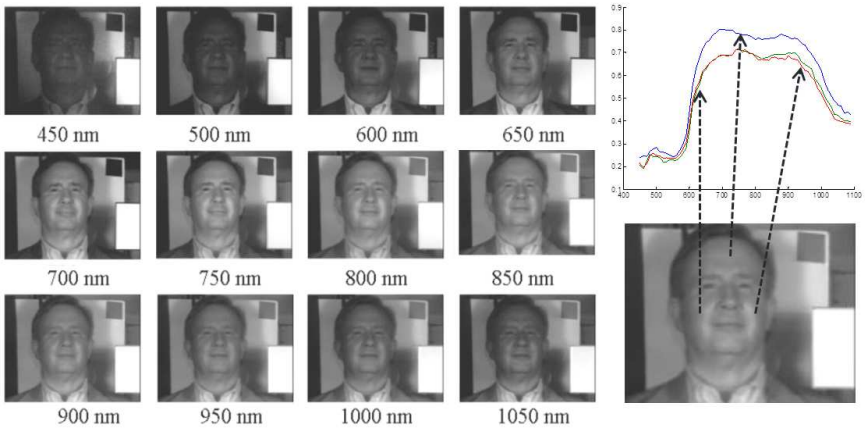


Figure 4.1: Hyperspectral face example. On the top right, the spectral responses on the forehead and cheeks are drawn with respect to the wavelength.

sequently, the spectral response of the skin has the potential of being an additional discriminative feature for face recognition. An example of hyperspectral face image is shown in Fig. 4.1 with some band images displayed. This example also shows that the spectral responses on the forehead and cheeks are different for a single subject and the spectral signature is more discriminative than the intensity at a single band.

Methods employing both spectral and spatial information for face recognition have been reported in recent years [101, 102, 103, 104, 105]. The challenges for integrating spectral-spatial features are in two aspects. The first one is the low quality of existing hyperspectral face images [102, 106]. Some spectral bands, which are important for face recognition, are corrupted by heavy noises and have very low signal to noise ratio. Most hyperspectral faces also suffer from the cross-band misalignment resulting from the offset of different wavelengths of light, movement of faces and blinking of eyes during the image capturing process which normally takes several seconds [102, 106]. These factors contaminate the spectral signature that can be extracted from faces. The second and a more serious problem is the totally different intrinsic attributes of spatial and spectral information. Direct concatenation of spatial and spectral features may not improve the recognition performance and lead to overfitting. [41].

Faced with these challenges, researchers have proposed different solutions. Pan et al. [101] pioneered the hyperspectral face recognition by extracting spectral features picked from typical face regions, such as forehead, cheeks, hair, and lips. Experiments proved that their method is robust to the orientation and expression variations. However, this method did not make use of any spatial information. In order to tackle this defect, an extended work integrated the spatial information by introducing a spectral eigenface method [107]. Di et al. [102] performed $(2D)^2$PCA on hyperspectral faces and transformed high dimension faces into subspace. A hyperspectral face dataset was collected to analyze this method in all bands, single band, and band subset with fusion. In the work of Shen and Zheng [103], a 3D Gabor Wavelet was employed to extract multi-scale and multi-orientation features from hyperspectral face images. This approach achieved a significantly higher accuracy than all aforementioned methods on the benchmark dataset adopted in [102]. Despite its high accuracy, this approach suffered from huge computation and memory costs. In more recent, Uzair et al. [104] developed a spectral-spatial feature extraction method based on 3D discrete cosine transform (3D DCT) to calculate low-frequency coefficients of hyperspectral face images. By using a partial least square regression model to perform face classification, this method achieved recognition rates higher than 90% on three datasets. Regarding spectral-spatial feature extraction, all these work did not completely solve the two challenges mentioned above. They either only used spectral signature or treated the hyperspectral data as an isotropic volume.

In this chapter, we propose a 3D high-order texture pattern descriptor based on local derivative pattern (LDP) for hyperspectral face recognition. LDP is a high-order derivative descriptor which provides a general framework for encoding directional features on 2D images [19]. It can be represented as a high-order and multi-direction derivative plus a special binarization function which acts as a denoising function. In this chapter, we introduce the 3D local derivative pattern (3D LDP) to extract the spectral-spatial information from hyperspectral face images. A 3D histogram is

constructed on the derivative pattern and then used as the feature descriptor for face recognition. We also compared our method with the 3D extension of local binary pattern (3D LBP) which is originally developed for dynamic texture analysis [108]. The superiority of our method is validated on two benchmark hyperspectral face datasets. The results show that the proposed method outperforms several state-of-the-art methods.

The rest of the chapter is organized as follows. Section 4.2 describes the proposed 3D LDP method and how the extracted texture feature is converted to a descriptor for each hyperspectral image. The implementation details are described in Section 4.3, followed by experiments and analysis in Section 4.4. Finally, the conclusions are drawn in Section 4.5.

4.2 3D Local Derivative Pattern

The proposed 3D LDP is a three-dimensional high-order texture descriptor. It analyzes the micro-patterns in three-dimensional data and encodes them into binary numbers. This method consists of two parts, a 3D directional derivative pattern and a special binarization function. The 3D directional derivative pattern describes the changes in multi-directions and curvature in high orders. The binarization function provides a general description on the consistency of two derivatives, which is suitable to extract detailed features in multi-dimensional data. Though hyperspectral images provide additional spectral information, there is large huge redundancy along the wavelength dimension. It is because that spectral responses of objects change smoothly across most wavelengths and the discriminative spectral information may only exist in several specific wavelengths. Therefore, 3D LDP has the potential to extract the spectral-spatial features in hyperspectral faces. After the 3D LDP features are calculated, they are converted into histogram to generate vectorized image representation.

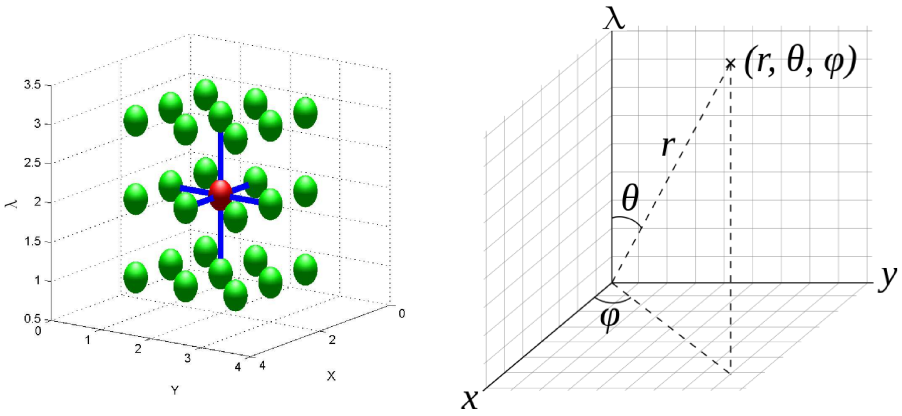


Figure 4.2: Coordinate systems in 3D space

4.2.1 Construction of local derivative pattern

Given a hyperspectral image represented as a cube, a Cartesian coordinate system and a spherical coordinate system can be defined. A $3 \times 3 \times 3$ example is shown in Fig. 4.2. In the 3D coordinate system, a point P is represented by its index of (x, y, λ) while angle/direction is expressed by the combination of inclination θ and azimuth φ. Given the 3D neighborhood of a central point, the first-order 3D directional derivative $I'_{(\theta,\varphi)}(x, y, \lambda)$ is defined as follows

$$\begin{aligned} I'_{(\theta,\varphi)}(x, y, \lambda) &= I(x + \triangle x, y + \triangle y, \lambda + \triangle\lambda) - I(x, y, \lambda) \\ \theta = \arctan(\frac{\sqrt{\triangle x^2 + \triangle y^2}}{\triangle\lambda}), &\varphi = \arctan(\frac{\triangle y}{\triangle x}) \end{aligned} \tag{4.1}$$

where $P = (x + \triangle x, y + \triangle y, \lambda + \triangle\lambda)$ is the nearest neighbor of the central point $P = (x, y, \lambda)$ in direction of (θ, φ). It should be noted that in the discrete 3D image space, θ and φ are also discrete which means $\theta \in \{k\pi/4 | k = 0, 1, 2, 3\}$ and $\varphi \in \{b\pi/4 | b = 0, 1, ..., 7\}$. Given the directional derivative calculated in a specific angle, higher order derivative can be calculated in the same angle.

In order to encode the derivative, our method employs a function to binarize the derivative between the central point and its neighbors. This function describes the

consistency of two neighboring derivatives. It is defined as follows

$$f(I'(P_c), I'(P_i)) = \begin{cases} 0 & \text{if } I'(P_c) \cdot I'(P_i) > 0 \\ 1 & \text{if } I'(P_c) \cdot I'(P_i) \leq 0 \end{cases} \tag{4.2}$$

where $I'(P_i)$ is the derivative calculated at the i^{th} neighbour of the central point P_c. When the derivatives are consistent, i.e., both are positive or negative, the result is 0, otherwise, the result is 1.

The directional derivative and binarization function play different roles and are independent to each other. The former extracts various distinctive spatial and spectral changes, while the latter quantizes the consistency between derivatives of the central point and its neighbors. From this point of view, the binarization function can be considered as another level of derivative on top of the directional derivative of the hyperspectral image, i.e., a special second order derivative of the raw image. Therefore, given point P_c, we denote such second order derivative as $D^2_{(\theta,\varphi)}(P_c)$, such that

$$D^2_{(\theta,\varphi)}(P_c) = f(I'_{(\theta,\varphi)}(P_c), I'_{(\theta,\varphi)}(P_i)) \tag{4.3}$$

After obtaining the binary derivative, these values are fed into an encoding system to generate a unique integer which encodes the local derivative pattern as follows

$$3DLDP^2_{(\theta,\varphi)} = \sum_{i=1}^{m} D^2_{(\theta,\varphi)} \times 2^{i-1} \tag{4.4}$$

where i indexes the i^{th} neighbour in the m neighbourhood of the central point.

When all directions are combined, the final second-order 3D LDP becomes

$$\begin{aligned} &3DLDP^2 = \\ &\textstyle\sum_{\theta\in A,\varphi\in B} \sum_{i=1}^{m} f(I'_{(\theta,\varphi)}(P_c), I'_{(\theta,\varphi)}(P_i)) \times 2^{(j-1)\times m+i-1} \end{aligned} \tag{4.5}$$

where $A = \{k\pi/4 | k = 0, 1, 2, 3\}$ is the set of angles for θ and $B = \{b\pi/4 | b = 0, 1, .., 8\}$ is the set of angles for φ in the discrete 3D space in Fig. 4.5. Different combinations of θ and φ determine different directions for derivative calculation, and j is the index for these combination.

4.2.2 n^{th}-order local derivative pattern

The 3D LDP can be extended to higher-order derivatives. This can be implemented by applying Equation 4.1 iteratively and then perform binarization function. Consequently, the n^{th}-order derivative of 3D LDP is constructed by calculating the directional derivative for $n-1$ times before applying the binarization function in Equation 4.2. Therefore, the n^{th}-order derivative is defined as

$$D^n_{(\theta,\varphi)}(P_c) = f(I^{n-1}_{(\theta,\varphi)}(P_c), I^{n-1}_{(\theta,\varphi)}(P_i)) \tag{4.6}$$

After the encoding step, the n^{th}-order 3D LDP code in the direction of (θ, φ) within the m neighborhood of P_c is calculated as

$$\begin{aligned} &3DLDP^n_{(\theta,\varphi)} = \\ &\textstyle\sum_{i=1}^{m} f(I^{n-1}_{(\theta,\varphi)}(P_c), I^{n-1}_{(\theta,\varphi)}(P_i)) \times 2^{(j-1)\times m+i-1} \end{aligned} \tag{4.7}$$

The 3D LDP code extracted in different directions can be combined into a final 3D LDP code as follows

$$\begin{aligned} &3DLDP^n = \\ &\textstyle\sum_{\theta\in A,\varphi\in B}\sum_{j=1}^{m} f(I^{n-1}_{(\theta,\varphi)}(P_c), I^{n-1}_{(\theta,\varphi)}(P_i)) \times 2^{(j-1)\times m+i-1} \end{aligned} \tag{4.8}$$

where the parameters have the same meaning as in Equation 4.5.

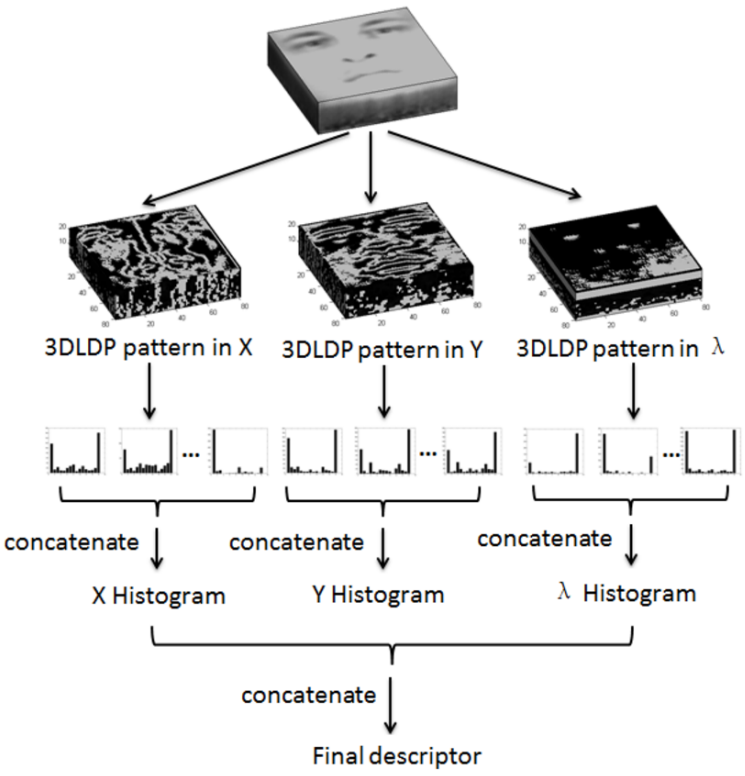


Figure 4.3: 3D LDP descriptor construction.

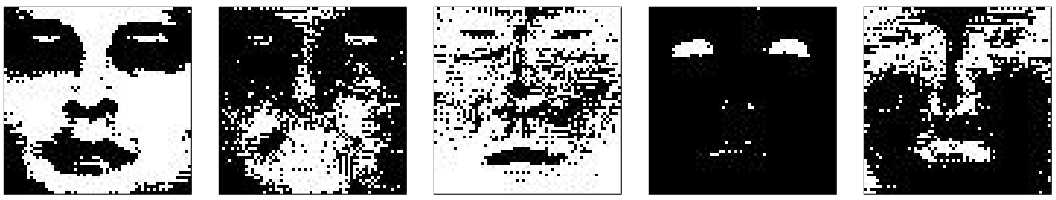


Figure 4.4: 3D LDP pattern in λ direction at different bands.

4.2.3 Construction of 3D LDP descriptor

Through the above procedure, each pixel is assigned with an integer at given direction (θ, φ). Such mixed-order derivative code contains discriminative texture feature of a local neighborhood in a hyperspectral image. To convert such texture feature into a vectorized descriptor for face recognition, statistical distribution of the 3D LDP feature in local regions is calculated and summarized using a histogram. Directly using the code generated by Equation 4.5 leads to very high dimensional and sparse

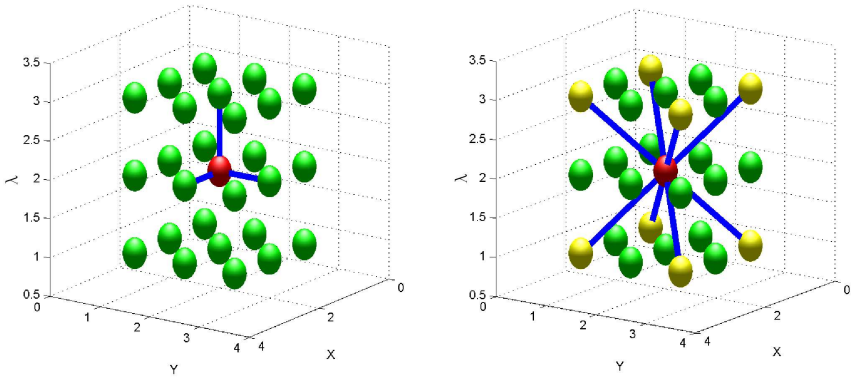


Figure 4.5: Left: directions of 3D LDP. Right: 8 neighbours chosen to code the pattern.

histogram, which is not suitable for efficient classification. In practise, the 3D LDP codes are constructed in different directions based on Equation 4.4. These codes are used to build histograms, and the length of each histogram is 2^m for each direction. The number of bins can be further shortened by merging the neighboring entries.

The hyperspectral image is also divided into small cubes. On each cube, a local histogram is generated in each direction. For each direction, the local histograms are concatenated, such that the resulting directional histogram embeds both spatial and spectral information at the region level. Then the directional histograms are merged into a final histogram which is the final descriptor of the hyperspectral face image. An example of 3D LDP descriptor extracted from hyperspectral face image is displayed in Fig. 4.3 and the pattern in *lambda* direction is shown in Fig. 4.4.

4.2.4 Hyperspectral face recognition

A distance measurement between hyperspectral face images is required for the final recognition step. In our method, histogram intersection is adopted to measure the similarity between two histograms

$$M_{12} = \sum_{i}^{K} min(H_1(i), H_2(i)) \tag{4.9}$$

where M_{12} is the histogram intersection between histogram H_1 and H_2. K is the length of the histogram. This measurement calculates the common area of two histograms. The higher the value is, the more similar two histograms are.

4.3 Implementation Details

In our method, we first filter the hyperspectral faces with a 3D Gabor filter. It is necessary because hyperspectral images often suffer from serious noises coming from both hardware and the imaging process. Directly applying 3D LDP will result in encoding a lot of noises in the spectral domain. The 3D Gabor filter reduces the influence of sensor noise and cross-band misalignment. It also enhances the discriminative patterns in the spatial and spectral domain. Different from the 3D Gabor Wavelet method in [103], filters with fixed scale are used.

In a 3D space, a pixel can have up to 26 direct neighboring pixels. Without losing generality, we make use of 3 directions and 8 neighbors in the cube in our implementation, as shown in Fig. 4.5. These three directions are $(0, 0)$, $(\pi/2, 0)$ and $(\pi/2, \pi/2)$, which correspond to positive x, positive y, and positive λ, respectively. The 8 neighbors are the eight corners of the cube which span both spatial and spectral dimensions. It should be mentioned that this is the most basic form and more directions and neighbors can be used to extract more detailed information. In the implementation, we calculate the second order 3D LDP. A summary of the second-order 3D LDP is given in Algorithm 2. Its extension to higher order is straightforward.

4.4 Experiments and Results

The experiments consist of two parts. In the first part, we compare the proposed methods with the state-of-the-art hyperspectral face recognition algorithms on two well-established hyperspectral face datasets. For the completeness of experiments, we also implement the 3D LBP in our test. The results show that our method outper-

Algorithm 2 Second-order 3D LDP

Require: Hyperspectral data cube S
  for each pixel in S **do**
    Calculate the derivative in x, y, and λ positive direction, obtaining Dx, Dy and $D\lambda$
  end for
  for each direction **do**
    for each pixel p_c in S **do**
      for each neighbour p_i around p_c **do**
        Apply Equation 4.2 on $D(p_i)$ and $D(p_c)$
      end for
      Calculate the LDP code using Equation 4.4
    end for
    for each local region in S **do**
      Build the histogram
    end for
    Concatenate the histogram in different local regions
  end for
  Concatenate the histogram in different directions

forms all other approaches. Then in the second part, we validate the usefulness of spectral information by comparing the proposed method with a state-of-the-art 2D face recognition method.

Hyperspectral face recognition is much less mature than work in 2D faces. There are very few datasets available. Some hyperspectral face datasets are not publicly available due to privacy or other reasons such as those collected and used in [101] and [104]. To our knowledge, there are only two publicly available datasets of hyperspectral face, i.e., Hong Kong Polytechnic University Hyperspectral Face Dataset (HK-PolyU) [102] and Carnegie Mellon University Hyperspectral Face Dataset (CMU dataset) [106]. As a consequence, our experiments were conducted on these two datasets.

4.4.1 Results on HK-PolyU hyperspectral face dataset

The first dataset is the Hong Kong Polytechnic University Hyperspectral Face Dataset (HK-PolyU) [102], which consists of 300 hyperspectral faces from 25 subjects. For

each subject, there are four sessions collected at two times with an average span of five months. Each session consists of frontal, right and left views with the neutral expression. Each image covers the visible wavelength from 400nm to 720nm with an interval of 10nm. The images are quite noisy at the both ends of the spectral range. So the first six and last 3 bands are removed in the experiment. Fig. 4.6 shows examples of two subjects taken in four sessions. It can be seen that the appearance of the same person varies a lot in different sessions. The bands of this dataset are not well registered and the pixel based spectral signature are contaminated. This explains why it cannot be used directly by 3D LDP.

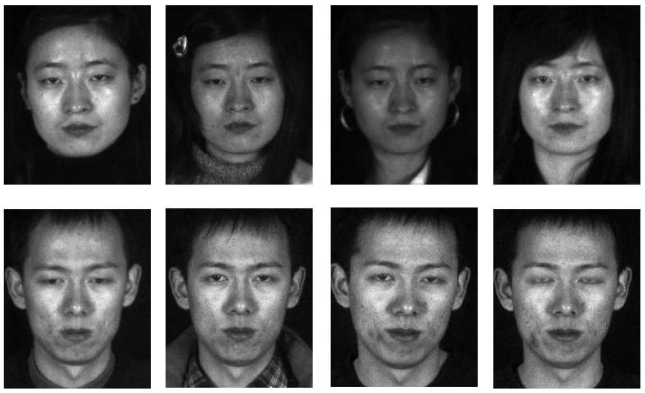


Figure 4.6: Examples from the HK-PolyU hyperspectral face dataset. Images are extracted from different sessions.

In order to evaluate the performance of 3D LDP, we compare it with several hyperspectral feature extraction methods. These methods are Spectral Feature [101], Spectral Eigenface [107], 2D PCA [102], 3D LBP [108], 3D Gabor Wavelet [103] and a newly developed method of 3D DCT [104]. We followed the experiment setting in [103]. The hyperspectral faces were cropped into 64×64 in spatial dimension and the position of eyes was manually aligned. Four sessions of 25 subjects were used as the test set. The gallery set was constructed by randomly selecting two sessions from the test set and the remaining faces were used as the probe set. For our method, 13 3D Gabor transforms (frequency = 0.25, $\theta = 0$, $\pi/4$, $\pi/2$, $3\pi/4$, $\varphi = 0$, $\pi/4$, $\pi/2$, $3\pi/4$) were used to preprocess the hyperspectral faces. We fixed the subregion size

to $8 \times 8 \times 8$ pixels to build the 3D histogram and set the derivative order to two. The 3D LBP uses the same settings as 3D LDP except that it takes 14 neighbors as introduced in [108]. In 3D Gabor Wavelet, the size of Gabor was $39 \times 39 \times 39$ and 52 filters were used as recommended in [103]. The 3D DCT method reported in [104] consists of a 3D DCT based feature extraction method and a partial least squares regression based classification method. We only implemented the feature extraction part using the code published by the authors of [104] because our method focuses on the spectral-spatial feature extraction. The face recognition in alternative methods was implemented by using the nearest neighbor classifier based on the Euclidean distance between probe and gallery images, while our method used the nearest neighbor classifier based on histogram intersection. We used the mean and standard deviation of rank-1 recognition rate of six possible combinations of the gallery and probe sets to evaluate the performance of each method.

Table 4.1: Recognition rate on the HK-PolyU Hyperspectral Face Dataset using Spectral Feature, Spectral Eigenface, 2D PCA, 3D DCT, 3D LBP, 3D Gabor Wavelet and the proposed 3D LDP.

Methods	Recognition Rate
Spectral Feature [101]	$45.35 \pm 3.87\%$
Spectral Eigenface [107]	$70.33 \pm 3.61\%$
2D PCA [102]	$71.00 \pm 3.16\%$
3D DCT [104]	$84.00 \pm 3.35\%$
3D LBP [108]	$88.80 \pm 1.79\%$
3D Gabor Wavelet [103]	$90.00 \pm 2.83\%$
3D LDP	$\mathbf{95.33 \pm 1.63\%}$

The experimental results are shown in Table 4.1. 3D LDP has achieved the highest recognition rate of $95.33 \pm 1.63\%$ on the HK-PolyU dataset, exceeding the second best 3D Gabor Wavelet by 5.33% in average. The 3D Gabor Wavelet feature extraction method has achieved similar results as reported in [103] and [104]. Just like LDP

performing better than LBP in 2D face recognition, the 3D version also shows the same trend in which the accuracy of 3D LDP is 6.53% higher than that of 3D LBP. The accuracy of 3D DCT is 84.00%, which is much lower than the results reported in [104]. This is partly due to the fact that the partial least square regression has not been used as the classifier. Compared with 3D LDP and 3D Gabor methods, 3D DCT only extracts low-frequency coefficients of DCT, which is a general representation of signal energy. In contrast, 3D LDP and 3D Gabor Wavelet extract multi-scale and multi-direction local textures, which capture more local information than 3D DCT. The results also show that pure spectral feature performs the worst because it has ignored the spatial information.

4.4.2 Results on CMU hyperspectral face dataset

The second hyperspectral face dataset is the Carnegie Mellon University Hyperspectral Face Dataset (CMU dataset) [106]. Examples of this dataset are shown in Fig. 4.7. Compared with the HK-PolyU dataset, the CMU dataset covers a wider range of spectrum from 450nm to 1090nm with an interval of 10nm. It includes data of 54 subjects taken in multiple sessions. Each session consists of four different illumination conditions which include all lights on, center light on, left light on and right light on. Each subject has 1 to 5 sessions taken at different times. Due to the low spectral sensitivity of hyperspectral camera, bands at the two ends of spectrum range have very low signal to noise ratio. In this experiment, faces under the condition of all lights on taken at different time were used as the test set. The gallery set was constructed by randomly sampling one face per subject from the test set and the rest faces were used as the probe set. This procedure repeated 10 times and the mean and standard deviation were used to evaluate the performance of different methods.

The gallery included 47 faces while the probe had 98 faces. Each face was cropped into 64×64 in spatial dimension and the position of eyes was manually aligned. The first 7 and last 2 noisy bands were removed. In this experiment, we adopted the same

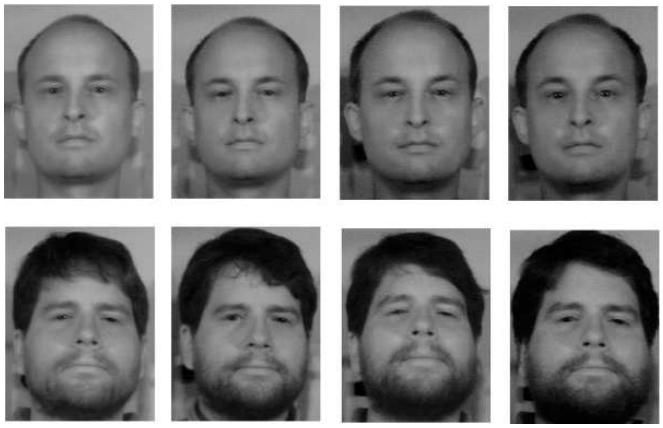


Figure 4.7: Example images on two subjects in the CMU hyperspectral face dataset. For each subject, the images are captured in four sessions with 700nm band displayed.

settings as on the HK-PolyU dataset for 3D LBP, 3D Gabor Wavelet, 3D DCT and 3D LDP.

Table 4.2: Recognition rates on the CMU Hyperspectral Face Dataset using Spectral Feature, Spectral Eigenface, 2D PCA, 3D DCT, 3D LBP, 3D Gabor Wavelet and the proposed 3D LDP..

Methods	Recognition Rate
Spectral Feature [101]	$38.18 \pm 1.89\%$
2D PCA [102]	$72.10 \pm 5.41\%$
Spectral Eigenface [107]	$84.54 \pm 3.78\%$
3D DCT [104]	$88.65 \pm 2.34\%$
3D LBP [108]	$92.16 \pm 3.52\%$
3D Gabor Wavelet [103]	$92.20 \pm 2.46\%$
3D LDP	$\mathbf{94.83 \pm 2.62\%}$

The experimental results are shown in Table 4.2. These results are consistent with those on the HK-PolyU dataset. The 3D LDP leads the performance among all the methods with an average recognition rate of 94.83%. The second and third methods are 3D Gabor Wavelet and 3D LBP. Regarding the efficiency, 3D DCT is the fastest

because it only produces one cube per face, while the 3D Gabor Wavelet is the worst which produces 52 cubes. The 3D LBP and 3D LDP are in the middle by creating 13 cubes per face. Overall, the 3D LDP efficiently extracts the spectral-spatial features and performs the best in recognition rate among all methods.

4.4.3 Further analysis of 3D LDP

In order to show that the spectral information provided by the hyperspectral images increases the accuracy of face recognition, we also conducted an experiment in which 3D LDP was compared with a 2D face recognition method. We chose 2D LDP for comparison because it uses the same rationale in feature extraction in 2D space. Because 2D LDP cannot be directly applied to hyperspectral images, we ran it on each individual band. Then average recognition rate and standard deviation from all bands were calculated. In 2D LDP, the subregion size was set to 8×8 pixels and the derivative order was set to two. The comparison was conducted on both datasets following the same settings as experiment A and B. The results in Table 4.3 show that 3D LDP performs significantly better than 2D LDP on both datasets. This proves the value of introducing spectral information for face recognition. Effective characterizing the spectral-spatial information has led to remarkable benefit in improving the accuracy of face recognition.

Table 4.3: Recognition rates of 3D LDP and 2D LDP on two datasets.

Methods	HKPolyU dataset	CMU dataset
2D LDP [19]	$86.25 \pm 9.62\%$	$90.27 \pm 3.51\%$
3D LDP	$95.33 \pm 1.63\%$	$94.83 \pm 2.62\%$

4.5 Conclusion

In this chapter, we have introduced a 3D texture descriptor for hyperspectral face recognition. It provides a framework for constructing multi-direction and multi-neighbourhood local derivative patterns. This pattern integrates the information from both spatial and spectral domains and reduces the negative influence of noise in the hyperspectral images. By building a 3D histogram on this feature, a spectral-spatial descriptor can be generated for hyperspectral face recognition. The proposed method has been tested on two publicly available hyperspectral face datasets and has been compared with several existing methods. The results prove that our method has outperformed the state-of-the-art methods in terms of accuracy. It is expected that the proposed method can also be used in other hyperspectral image classification applications.

Chapter 5

Tensor Morphological Profiles for Hyperspectral Image Classification

5.1 Introduction

In remote sensing, hyperspectral image classification addresses the problem of land-cover class identification and thematic map generation, which has extensive applications in precision agriculture, mine exploration and environment monitoring [2]. A classification system usually consists of several key steps, including data preprocessing, feature extraction, and prediction model building. Among these steps, feature extraction is of significant importance which aims to find the most compact and informative set of features and improve the accuracy and efficiency of classification tasks. In traditional pixel-wise feature extraction methods, the spectral response at a single pixel reflects the components of the ground object at the corresponding location. It usually works as the raw feature for hyperspectral image classification. Research on the raw spectral features has focused on extracting discriminative bands or a subset of bands selection for dimensionality reduction and class separation [29, 109, 110]. Although a well-trained classifier may have good generalization capability to unseen data with these features, the created thematic map often suffers from salt and pepper noises. Moreover, in images of urban areas, only relying on spectra might not be able to distinguish different classes made of similar materials. Therefore, it is necessary to introduce spatial information into hyperspectral image classification, so as to build

better models of local structures in the image and facilitate more accurate land-cover and object classification.

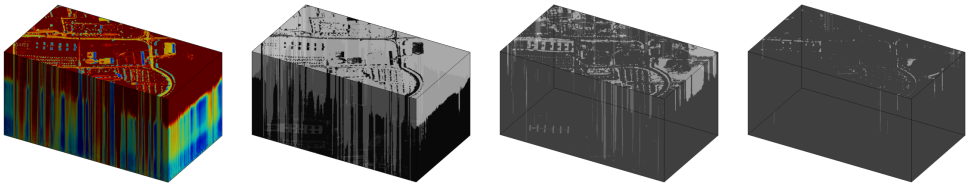
During the past several years, spectral-spatial feature extraction has attracted increasing attention in hyperspectral image classification [13, 41]. While traditional spectral features are extracted at the single pixel level in hyperspectral images, spectral-spatial feature extraction methods use spatial neighborhood to calculate features. Several typical image processing methods have been extended to multiple dimensions to extract spectral-spatial features in hyperspectral images, such as 3D discrete wavelet [15], 3D Gabor wavelet [35], 3D scattering wavelet[47], and local binary patterns [111]. Other novel spectral-spatial features based on other theories include spectral saliency [50], spherical harmonics [51], and affine invariant descriptors [52].

Alternatively, some researchers tend to extract spectral and spatial features separately. One of the highly cited methods is the extended morphology profile (EMP) [30, 49, 41]. Mathematical morphology is suitable to extract or suppress objects and structures in images and thus can be used for typical image processing tasks such as image filtering and image segmentation and image measurement. Benediktsson et al. [30] pioneered the application of mathematical morphology in hyperspectral image classification. Fauvel et al. [49] fused the spatial and spectral information by stacking the extended morphology profile and original spectral response vector, and then it was fed into an SVM classifier. Since then, EMP has become a baseline method to extract spectral-spatial features in hyperspectral image classification and several variations have been proposed [34, 41].

In spite of these efforts, mathematical morphology is still restricted to two-dimensional images and cannot cope with higher dimensional images directly. The reason is that morphological operations require assigning orders to pixels based on their values. In grayscale images, each pixel has an intensity value and thus they can be easily compared. A hyperspectral image, however, consists of multiple bands so that a pixel is

expressed in the form of a vector which represents the spectral responses at multiple wavelengths. Unfortunately, defining maximum or minimum among vectors is ambiguous, which causes the ordering of vectors a problem [112]. This leads to the difficulties in defining the basic set operations. Some researchers attempt to solve this problem by developing different vector ordering scheme, but these approaches have not been commonly accepted due to their suitability to particular tasks [113]. As a consequence, it is necessary to extend morphology to multivariate images in a simple way so that this powerful tool can be used to deal with the increasing number of multiple multivariate images such as Magnetic Resonance Imaging (MRI), (Hyperspectral Imaging) HSI and videos.

Recently, tensor decomposition methods have been employed in hyperspectral image classification. A tensor is a multidimensional array, and the order of a tensor is the number of indices needed to indicate an element in the array. A hyperspectral image can be arranged as a three-dimensional array with the modes corresponding to spatial x, spatial y and wavelength λ so that it can be treated as a third-order tensor. This suggests that tensor methods can help to analyze the hyperspectral data. Zhang et al. [36] represented a pixel's spectral-spatial feature as a tensor constructed by the spectra of the pixel itself as well as its k nearest neighbors. Then a tensor discriminative locality alignment method (TDLA) was developed to transform the high-order tensor space to a low dimensional feature space via patch optimization which includes the information from intraclass and interclass in the training data. Its superiority over several unsupervised and supervised feature dimensionality reduction methods has been proved on three commonly used hyperspectral datasets. Meanwhile, Velasco-Forero et al. [63] developed an additive morphological decomposition (AMD) which decomposes an image into two parts. The first part is formulated by the summation of anti-extensive transformations and extensive transformations, representing all contrast and boundary information. The second part is related to the residue of anti-extensive transformations and extensive transformations with dif-



(a) Raw cube (b) Binary cube (0.25) (c) Binary cube (0.5) (d) Binary cube (0.75)

Figure 5.1: The Pavia University dataset in tensor representation: (a) raw hyperspectral tensor rendered with false color; (b-d) binary hyperspectral tensors with thresholds of 0.25, 0.5 and 0.75. These images indicate that different classes are with different high-dimensional structures which can be extracted by mathematical morphology.

ferent scales, constituting a hierarchy of multiscale texture components. Then tensor modeling methods are employed to reduce the dimension of extracted features in which the spatial information is maintained. However, this method processed the hyperspectral image band by band in the AMD, resulting in breaking the correlation of spectral dimension.

In this chapter, we show that it is possible to design a multi-dimensional morphology which works better than two-dimensional morphology. Instead of treating an HSI as a two-dimensional image with vectorized values at each pixel, we borrow the concept of tensor modeling and design the morphology on the tensor structures so as to avoid the ordering problem. In Fig. 5.1, the hyperspectral image is represented as a tensor and different color indicates the value of a voxel. It can be noted that different classes are with distinguishable structures in both spatial and spectral dimensions. Such high-dimensional structure can be further investigated when the data cube is converted into binary data via a thresholding process. Each class consists of discriminative high dimensional binary structures mixed with isolated dots owing to random noises, which can be analyzed by set operations in mathematical morphology [114].

The main contributions of this chapter are as follows. Firstly, we develop a tensor morphology profile which fuses the spectral and spatial information in a better way

than the benchmark EMP method. Secondly, we design a framework to extend morphology to hyperspectral images by considering them as high-order tensors, thus avoiding the ordering problem of vector data. Thirdly, though exemplified on hyperspectral images, the proposed multivariate image feature extraction method is general in nature and can be applied to MRI and video analysis as well.

The rest of this chapter is organized as follows. Section 5.2 introduces some basic notations and background knowledge related to mathematical morphology, and explains the ordering problem of mathematical morphology for multivariate images. The detailed method description is provided in Section 5.3 which covers tensor modeling, multiple dimensional morphology as well as tensor morphology profile. It gives an in-depth analysis on the characteristics of spatial dimension and spectral dimension in hyperspectral images. The comparison experiments and results are presented in Section 5.4, which is followed by the conclusions and future work in Section 5.5.

5.2 Morphology in Multivariate Images

5.2.1 Notation and theoretical foundation

We have defined several symbols used in this chapter as shown in Table 5.1. Throughout the chapter, a set is denoted with capital letters. The scalars are denoted by lowercase letters and vectors are represented by boldface lowercase letters. Matrices are written in boldface capitals and high order tensors correspond to boldface Euler script letters.

Mathematical morphology was originally developed on two-dimensional images [41]. Then the target space of mathematical morphology evolves from images to Euclidean sets, and then reaches more general concept of complete lattice [112]. Given two sets of data, the spatial unit of images in two dimensions and the set consisting of intensity values, respectively, an image is a function f mapping the set of intensity values to

Table 5.1: List of symbols

Symbol	Description
$\mathbb{R}^n$	n dimensional Euclidean space
$\mathbb{Z}^n$	n dimensional integer grid
U	the set of spatial unit of an image, equivalent to a subset of $\mathbb{Z}^2$
u	the spatial position of a pixel (an element in U)
I	the set of intensity values
t_{max}	the maximum value in set I
f	the mapping function from the set of spatial unit to the set of intensity, representing an image
B	the structural element in mathematical morphology
ε	erosion operation
δ	dilation operation
γ	opening operation
ϕ	closing operation
Ω	morphological profiles
$\boldsymbol{\mathcal{T}}$	a tensor
g	the mapping function from the set of multidimensional unit to the set of intensity
X	the set of spatial and spectral unit of a hyperspectral image, indexed with x, y, b, equivalent to a subset of $\mathbb{Z}^3$
CS	cross section of thresholding

the spatial units. Therefore, the definition of different types of images mainly relies on the definition of set I. For example, a grayscale image can be defined as:

$$f : U \subset \mathbb{Z}^2 \rightarrow \{0, 1, ..., t_{max}\} \tag{5.1}$$

where t_{max} is the maximum value of image intensity. When t_{max} equals 1, f becomes a binary image.

After the object space is defined, it is possible to define the basic operations in mathematical morphology - erosion and dilation. They employ the structuring elements to enhance or alleviate structures based on the specific requirements from users. In terms of the operations set theories, the erosion and dilation on binary

images are defined as:

$$\begin{aligned} \varepsilon_B(U) &= \bigcap_{b\in B} U_b \\ \delta_B(U) &= \bigcup_{b\in B} U_b \end{aligned} \tag{5.2}$$

where U and B are image spatial set and structural element, and their entries are represented as u and b. Note that translation of set is used here, U_b means that set U is translated so that element b in B is the origin of U. When mathematical morphology is applied to grayscale images, the equation becomes:

$$\begin{aligned} [\varepsilon_B(f)](u) &= \min_{b\in B} f(u+b) \\ [\delta_B(f)](u) &= \max_{b\in B} f(u+b) \end{aligned} \tag{5.3}$$

Based on erosion and dilation, a series of tools can be developed such as opening and closing. These two operations can remove specific structures and noises without destroying the original primary structures. The opening is the erosion of an image followed by an dilation and the closing firstly dilates the image and then erodes the output. The mathematical definition is

$$\begin{aligned} \gamma_B(f) &= \delta_B[\varepsilon_B(f)] \\ \phi_B(f) &= \varepsilon_B[\delta_B(f)] \end{aligned} \tag{5.4}$$

5.2.2 Ordering problem in morphology on multivariate images

The above equations include operations such as union, intersection, maximum and minimum. These operations are straightforward in two-dimensional grayscale images. However, when applying to multivariate or multichannel data such as RGB or hyperspectral images, the mathematical morphology equations face difficulties in defining the maximum and minimum due to lack of ordering relationship. The ordering relationship has to satisfy the following conditions:

- 1. reflexivity, $A \leq A$;
- 2. anti-symmetry, $A \leq B$ and $B \leq A$ if and only if $A = B$;
- 3. transitivity, if $A \leq B$ and $B \leq C$, then $A \leq C$.

The set satisfying the above conditions is called a partially ordered set and the relation "$\leq$" is the partial order. A stricter concept total order also requires the property of totality, for any A and B, only one of $A < B$, $B < A$, $A = B$ is true.

The ordering problem is involved with multichannel or multivariate images. In this case, the set of intensity is a vector rather than a scaler. An image with m channels or bands can be denoted as:

$$f : U \subset \mathbb{Z}^2 \rightarrow \{I^m | I = \{0, 1, ..., t_{max}\}\} \tag{5.5}$$

where $I^m = [i_1, i_2, ..., i_m]$, representing the possible intensity values in the multivariate images. In other word, the function f maps each pixel at location u in image to a vector.

$$f(u) = [f_1(u), f_2(u), ..., f_m(u)]^T \tag{5.6}$$

Therefore, defining Equation 5.3 on the new object space is not straightforward. In particular, hyperspectral image usually consists of multiple bands so that a pixel is expressed in a vector which represents the spectral responses at different wavelengths. Applying mathematical morphology on hyperspectral images will have the same problem.

5.2.3 Extended morphological profile

The ordering problem in multivariate images can be solved through a marginal process, which ignores the inner-correlation between channels and processes each channel of an image separately. Since hyperspectral images consist of many bands, marginal process will surfer from heavy computational burden and redundant

information. Extended morphological profile employs principle component analysis (PCA) to transform the original hyperspectral images into fewer number of bands while maintaining the majority of spectral variance:

$$[I_1, I_2, ..., I_p] = PCA(f(x)) \tag{5.7}$$

where $I_1, I_2, ..., I_p$ are the first p principle components. Then a series of morphology operations can be applied on a single component to extract multi-scale spatial information with different sizes of structuring elements:

$$\Omega^{(n)}(I) = \left[\gamma^{(n)}(I), ..., \gamma^{(1)}(I), I, \phi^{(1)}(I), ..., \phi^{(n)}(I)\right] \tag{5.8}$$

where $\gamma^{(n)}(I)$ and $\phi^{(n)}(I)$ are the opening and closing operations with a disk-shape structural element of size n, respectively. Then the morphological profiles are obtained on each of the p primary components:

$$\hat{\Omega}_p^{(n)}(I) = \left[\Omega_1^{(n)}(I), \Omega_2^{(n)}(I), ..., \Omega_p^{(n)}(I)\right] \tag{5.9}$$

In the last step, the morphological profiles are stacked with the spectral response to form the spectral-spatial feature.

Two limitations of EMP exist in this process. Firstly, unsupervised dimensionality reduction methods cannot guarantee the completeness of spatial information associated with different classes. Applying PCA on hyperspectral images maintains the largest spectral variance at each pixel but it may not contain all spatial clues belonging to different classes of objects. Secondly, since EMP and spectral response are of different characteristics, directly stacking them into a representation leads to unbalanced spatial and spectral information to be extracted from the same pixel.

5.2.4 Vector morphology profile

Other than the marginal process, it is also possible to define an ordering relationship or ranking scheme on vectors, such as conditional ordering, partial ordering, and reduced ordering [113]. For instance, reduced ordering transforms the vectors to scalar values so that they can be ranked according to natural order. The mapping process can be realised based on either projection or distances. Given a vector $\mathbf{x} = (x_1, x_2, ..., x_p)$, the scalar value $h(\mathbf{x})$ can be obtained by using one of the followings:

$$h(\mathbf{x}) = \sum_{i=1}^{p} \lambda_i x_i \tag{5.10}$$

$$h(\mathbf{x}) = \phi(\mathbf{x}, \mathbf{x}_{ref}) \tag{5.11}$$

where λ_i is the weight parameter in the mapping function, ϕ is the distance measuring function, and $\mathbf{x}_{ref}$ is the predefined reference vector. Then the morphological profiles can be defined on the scalar values $h(\mathbf{x})$.

5.3 Proposed Approach

5.3.1 Tensor modeling

As mentioned earlier, it is convenient to model hyperspectral images as tensors. A tensor can be defined as a multidimensional array whose number of dimensions is the order or mode [62]. Formally, an N^{th} order tensor can be represented as $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times ... \times d_n}$. $1 \sim d_n$ is the range of elements in n^{th} dimension. Therefore, a hyperspectral image can be arranged as a third order tensor with modes corresponding to the columns, rows and wavelengths of images [36].

A tensor can also be transformed into sub-array data for the sake of interpretation. Fixing every index but one forms a fiber which is a vector. Similarly, fixing every index but two forms a slice. A hyperspectral image is a third-order tensor which

can be represented as $\mathcal{H} \in R^{m \times n \times b}$. m, n, b corresponds to the height and width of images, and the number of bands in the HSI. A spectral response is a fiber with fixed x and y dimension. A frontal slice, horizontal slice, lateral slice correspond to X-Y, X-λ, and Y-λ spaces, respectively. Denoting hyperspectral image as a tensor makes it easier to analyze its intrinsic properties from multiple aspects than the original form.

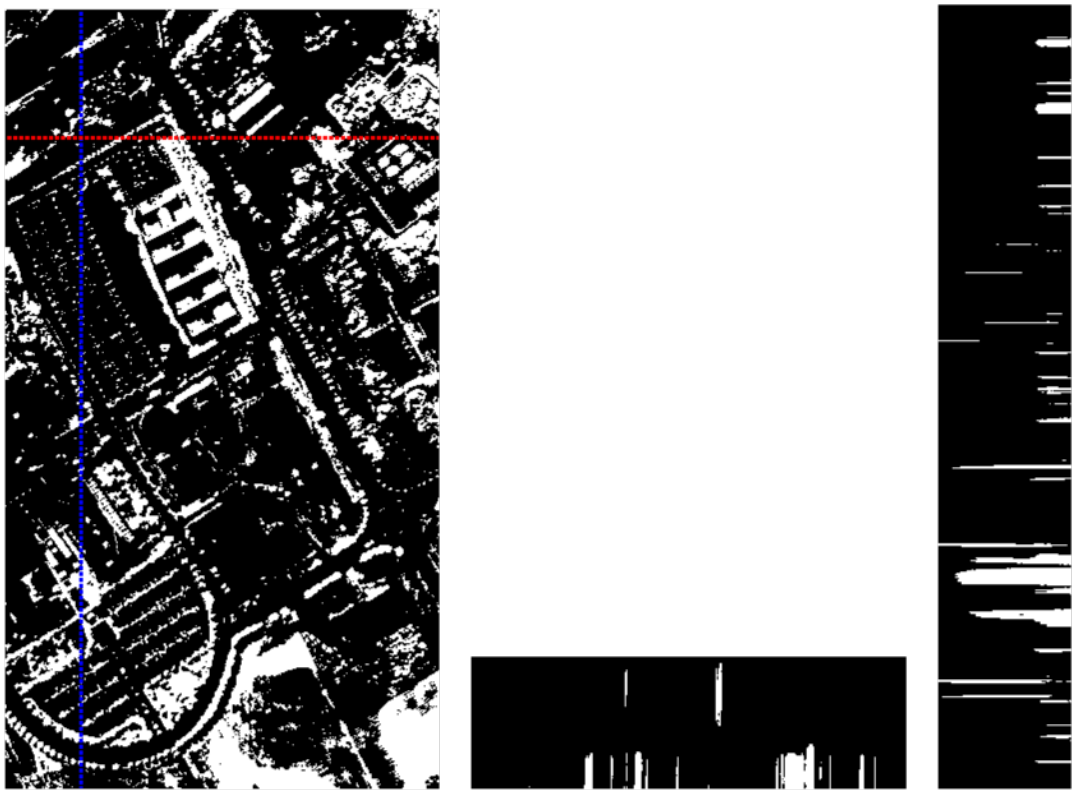
5.3.2 Multiple dimensional morphology

Morphology operations are primarily used for spatial feature extraction. After defining a hyperspectral image as a tensor, the structural information is not restricted to the spatial dimension. A fiber can either be a spectral response or spatial line. A slice can also reside in a spectral-spatial space. Theoretically, the morphology can be applied in any dimension or space in a tensor. But it is necessary to inspect whether it is reasonable to employ morphology to describe spectral responses.

Fig. 5.2 shows three truncation surfaces from a hyperspectral image after a thresholding process. They correspond to a tradition X-Y slice, an X-λ slice, and a Y-λ slice. Although X-λ and Y-λ are not images in practice, yet in the perspective of data analysis, they are similar to the X-Y slice and retain some sort of structures in the hyperspectral image. Figs. 5.2(a) and (b) also show some interesting phenomena. They depict the distributions of individual classes at spectral-spatial space, i.e. the X-λ and Y-λ distributions. It provides us with another point view of data distribution from which we can obtain something new. Specifically, those isolated holes and zigzags are likely owing to random noises. Furthermore, these 2D slices are still not sufficient to describe the complete view of the characteristics of each class. That is the reason we need to apply morphology to the whole tensor structures.

5.3.3 Tensor morphology profile

In perspective of object space for mathematical morphology, a tensor is a function rather than a subset of Euclidian space. It maps a subset of Euclidean space to a



(a) X-Y slice (b) X-λ slice (c) Y-λ slice

Figure 5.2: Two dimensional representation of the Pavia University dataset thresholded by 0.5: (a) X-Y slice at band 103; (b) X-λ slice at Y=100 (red line in X-Y slice); (c) Y-λ slice at X=60 (blue line in X-Y slice). These images reflect different shape structures and distributions.

range of intensity values. A grayscale tensor with mode n can be formalised as:

$$\mathcal{T} = g : X \subset \mathbb{Z}^{d_1 \times d_2 \times \ldots \times d_n} \rightarrow \{0, 1, ..., t_{max}\} \tag{5.12}$$

where X is the defined index domain for tensor and each element χ is a location of the original tensor $\mathcal{T}$. t_{max} is the maximum of tensor values. The difference between equation 5.5 and the definition here is that the former definition treats the data as two-dimensional sets plus $n-1$ dimensional intensity set. In contrast, the later one treats data as n dimensional data along with one-dimensional set accounting to the intensity. In this manner, the vector ordering issue can be avoided in the subsequent

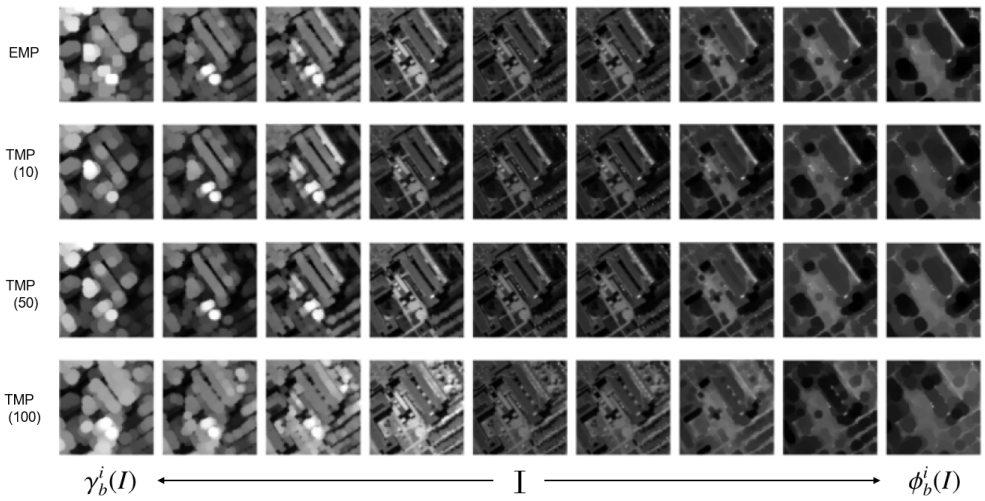


Figure 5.3: Extended morphological profiles versus tensor morphological profiles on a region of the Pavia University dataset. The second to fourth rows correspond to TMP at different bands. Different columns illustrate opening and closing with structural element of different sizes.

definition of morphology operations.

For the sake of understanding the practical meaning of mathematical morphology on a tensor, it can be further partitioned to multiple binary tensors so that the set operations can be applied. The thresholding process on an intensity set is defined as:

$$CS_h g(\chi) = \begin{cases} 0 & \text{if } g(\chi) < h \\ 1 & \text{if } g(\chi) \geq h \end{cases} \tag{5.13}$$

$CS_h g(x)$ is named as cross section which represents the output with a threshold of h. In order to contain the complete information of the original data, it is necessary to apply thresholding steps for multiple times and set $h = 0, 1, 2, ..., t_{max}$. Then, we can obtain the a series of binary tensors as introduced in Fig. 5.1. Interestingly, we can find special relationship among these cross sections:

$$CS_{t_{max}} g(\chi) \subseteq CS_{t_{max}-1} g(\chi) \subseteq ... CS_0 g(\chi) \tag{5.14}$$

Thus a grayscale tensor can be decomposed into the sum of its cross-sections:

$$\mathcal{T} = \sum_{h=1}^{t_{max}} CS_h\ g(\chi) = max\{h \mid CS_h\ g(\chi) = 1\} \tag{5.15}$$

The thresholding step not only plays an intermediate role to connect the grayscale tensor and its binary representation but also provides the practical meaning of applying mathematical morphology on a tensor. It is straightforward to apply the binary morphology to extract the high dimensional structures inside the binary tensor using following erosion and dilation:

$$\begin{aligned} \varepsilon_B(CS_h\ g(\chi)) &= \bigcap_{b\in B} CS_h\ g(\chi + b) \\ \delta_B(CS_h\ g(\chi)) &= \bigcup_{b\in B} CS_h\ g(\chi + b) \end{aligned} \tag{5.16}$$

This step is able to remove those isolated structures in tensors which are mainly caused by random noises. Then these binary high dimensional set based features can be combined to represent the feature of grayscale tensors based on Equations 5.15 and 5.16 .

$$\begin{aligned} \varepsilon_B(\mathcal{T}) &= \sum_{h=1}^{t_{max}} \left[\bigcap_{b\in B} CS_h\ g(\chi + b)\right] \\ \delta_B(\mathcal{T}) &= \sum_{h=1}^{t_{max}} \left[\bigcup_{b\in B} CS_h\ g(\chi + b)\right] \end{aligned} \tag{5.17}$$

Base on the threshold decomposition principle [112], the above equation is equivalent to

$$\begin{aligned} \varepsilon_B(\mathcal{T}) &= \min_{b\in B} g(\chi + b) \\ \delta_B(\mathcal{T}) &= \max_{b\in B} g(\chi + b) \end{aligned} \tag{5.18}$$

Here structural element B is a binary tensor with the same mode to f and it corresponds to the flat structural element in EMP. It is straightforward to derive opening

and closing:

$$\begin{aligned} \gamma_B(\mathcal{T}) &= \delta_B[\varepsilon_B(\mathcal{T})] \\ \phi_B(\mathcal{T}) &= \varepsilon_B[\delta_B(\mathcal{T})] \end{aligned} \tag{5.19}$$

At last, the tensor morphological profiles are obtained by applying opening and closing on a tensor with multiple structural elements of size n:

$$\Omega^{(n)}(\mathcal{T}) = \left[\gamma^{(n)}(\mathcal{T}), ..., \gamma^{(1)}(\mathcal{T}), \mathcal{T}, \phi^{(1)}(\mathcal{T}), ..., \phi^{(n)}(\mathcal{T})\right] \tag{5.20}$$

The output $\Omega^{(n)}(\mathcal{T})$ fuses the spectral-spatial information in high dimensions. Note that choosing a suitable structural element is crucial to TMP but it is beyond the scope of this chapter. To show the difference between EMP and TMP, an example on Pavia Univeristy is shown in Fig. 5.3. It can be noted that the profiles extracted by TMP have taken account of spectral information and vary at different bands.

5.3.4 Tensor structural element

Since no prior knowledge is known for target classes, the structural element has to cover as many properties of objects as possible such as multi-scale, multi-shape, and rotations. In EMP, we choose to use a disk-shaped structural element whose radius varies to cover the multi-scale information. When the mathematical morphology is extended to 3D dimension, the above condition remains valid in the form of 3D set theory. Instead of extracting 2D shape structures, 3D morphology enhances specific 3D structures based on the 3D structural element. Fig. 5.4 shows some common 2D and 3D shapes which can be used as the structural elements. Based on the design of 2D structural element, the first consideration is the sphere and it may extract multi-scale information in the hyperspectral images by varying radius. However, based on our experiment, it is not suitable for the spectral dimension which has different characteristics from the spatial dimension. In Fig. 5.2 (b) and (c), it can be noted that

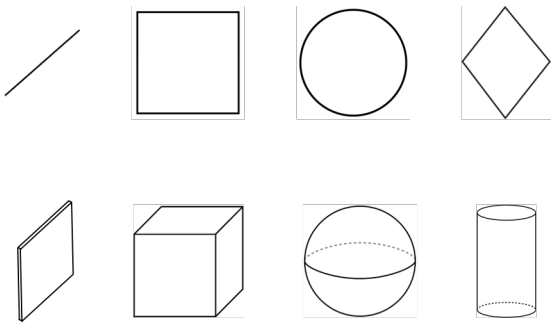


Figure 5.4: Examples of two dimensional structural elements (top row) and potential three dimensional structural elements (bottom row).

the spectral-spatial slices usually show bar-shape patterns because pixels in a local region are likely to share similar spectral responses. Therefore, based on the above analysis, we choose cylinder shape as the tensor structural element.

5.4 Experimental Analysis

We developed experiments to evaluate the effectiveness of our method. Three feature extraction methods were used in the experiments, including raw spectral response (SPE), extended morphological profile (EMP) and the proposed tensor morphological profile (TMP). While a disk-shaped structural element was used in EMP, a cylinder-shaped structural element was adopted in TMP. Experimentally, the radius of the cylinder was set to 1, 3, 5, 7 and the height was 3, 5, 9, 17, respectively. The shape of structural element is designed based on the characteristics of spatial and spectral dimensions. A linear SVM was adopted in the classification step. To guarantee the fairness of the experiments, the same setting were applied to different experiments including the classifier and the parameters of EMP and TMP.

5.4.1 Datasets

Three commonly used hyperspectral datasets were used in the experiments, including Pavia University, Pavia Center, and Washington DC datasets. The Pavia Univer-

sity dataset was captured at Pavia, northern Italy by the Reflective Optics System Imaging Spectrometer sensor (ROSIS). It has 103 spectral bands covering spectral range from 0.43 to 0.86 μm. The resolution is very high at 1.3 m per pixel. There are nine classes in the ground truth thematic map. Available training and testing samples are available with the dataset which can be found in [41]. Fig. 5.6 (a) and (b) shows the training map and testing map.

The Pavia Center dataset was also captured by the same ROSIS sensor. However, it contains seven different land cover classes from the Pavia University dataset. Both datasets were provided by the Telecommunications and Remote Sensing Laboratory, Pavia University, Italy.

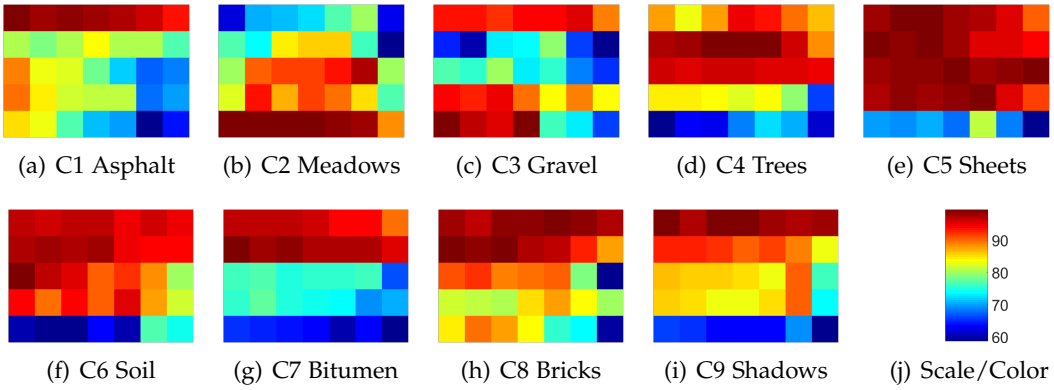
The Washington DC dataset was acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne over the Washington, DC Mall. The dataset contains 1280 scan lines with 307 pixels in each scan line. The number of spectral bands is 189 (21 bands are deleted due to water absorption), covering $0.4 - 2.4 \mu m$ range of the visible and infrared spectrum. Seven classes are considered in the experiment. This dataset comes from MultiSpec Project by Landgrebe [115]. The labeled pixels are divided into training set and testing set which are shown in Fig. 5.8 (a) and (b).

5.4.2 Parameters of cylinder-shaped structural element

To determine the influence of key parameters of cylinder-shaped tensor on image classification, we performed experiments on the Pavia University dataset. Then we calculated the classification accuracy for each category when the cylinder with different combinations of radius and height were used in the TMP. The radius and height were set to 1, 3, 5, 7, 9 and 1, 3, 5, 9, 17, 33, 65, respectively. The results are shown in Fig. 5.5. The classification accuracy is represented by color with red and blue corresponding to high and low values.

This figure shows some interesting results. While the category of Asphalt corre-

sponding to the roads is sensitive to small radius cylinders, the category of Meadows prefers large radius cylinders. For category of gravel, the most sensitive cylinder is with the smallest height. Since the signal in the third dimension of tensor may have distinct characteristics from other two dimensions, the structural element should not be isotropic. Therefore, the designed cylinder shape structural element in TMP are suitable to extract various information from different classes.



(a) C1 Asphalt (b) C2 Meadows (c) C3 Gravel (d) C4 Trees (e) C5 Sheets

(f) C6 Soil (g) C7 Bitumen (h) C8 Bricks (i) C9 Shadows (j) Scale/Color

Figure 5.5: Classification results for Class 1 to 9 on the condition of different combination of cylinder size in morphology. The radius of cylinder is represented by Y axis (from up to down): 1, 3, 5, 7, 9; the height of cylinder is represented by X axis (from left to right): 1, 3, 5, 9, 17, 33, 65. The color indicates the value of accuracy. The variance of color on different images shows that different classes are sensitive to different structural elements.

5.4.3 Experiment on the Pavia University dataset

The classified thematic maps on the Pavia University dataset are shown in Fig. 5.6 (c)-(d). From the figure, it can be seen that the spectral-spatial methods achieve much smoother results than raw spectral method. Furthermore, since TMP fuses the spectral and spatial information in high dimension and extracts the inner structures, it successfully classifies a few samples when EMP failed on the meadows. In Table 5.3, we also show the statistical results of each class, along with overall accuracy, average accuracy and Kappa coefficient. Some interesting phenomena can be observed from this table. SPE achieves the highest accuracies for Gravel, Bitumen, and Shadow

which do not consist of fixed spatial structures. EMP performs better than two other methods for Asphalt and Bricks and TMP obtains the better results on the rest of classes including Meadows, Trees, and Sheets. When only comparing two morphological methods, it can be found that TMP performs better than EMP on most classes, which means that TMP can extract more discriminative feature than EMP by fusing the spectral-spatial information in morphology operations.

5.4.4 Experiments on the Pavia Center dataset

Similarly, we show the training and testing map on the Pavia Center dataset in Fig. 5.7 (a) and (b). The classification maps are shown in Fig. 5.7 (c)-(d). Table 5.4 shows the results for class 1 to 9 in terms of classification accuracy, overall accuracy, average accuracy and kappa coefficient. In this table, SPE achieves the highest accuracies for Asphalt and Bare Soil. The results of TMP are comparable with EMP for most classes. However, TMP 10.4% and 9.6% achieves higher performance than EMP and SPE for class Bricks, respectively. Based on the overall accuracy and average accuracy, TMP achieves the highest accuracy and EMP performs better the SPE. This is consistent with the previous experimental result.

5.4.5 Experiments on the Washington DC Mall dataset

The classification results on the Washington DC Mall dataset are shown in Table 5.5. The results are consistent with two previous experiments. TMP achieves the highest accuracy on most of the classes. In terms of overall accuracy (OA) and average accuracy (AA), TMP still performs better than EMP.

From above experimental analysis, the results indicate that TMP consistently performs better than EMP. This is mainly due to that TMP has associated the spectral information during the morphology operation, thus extracting more discriminative information than EMP.

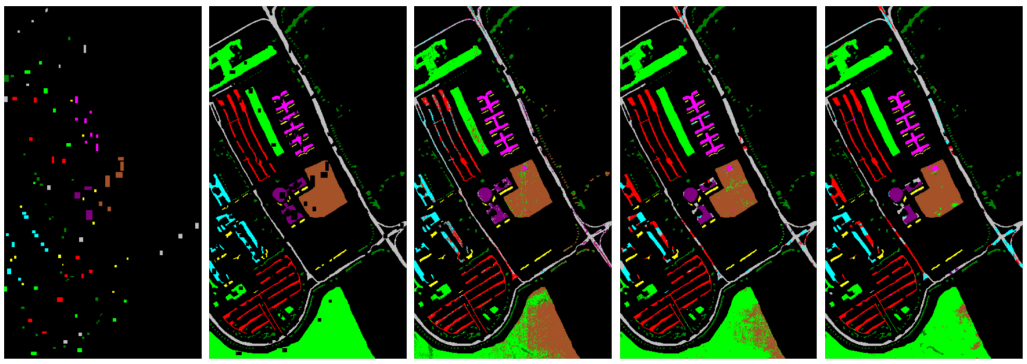
5.5 Conclusion

In this chapter, a tensor morphological profile has been introduced to extend mathematical morphology to high dimensional data. It partially solves the ordering problem in traditional mathematical morphology and can be used to extract spectral-spatial features with morphological operations. By associating the spectral information in the morphological profile, the designed structural element can extract variable information in different classes. Experiments validate that tensor morphological profile consistently performs better than the original extend mathematical morphology on different datasets.

Although the proposed method is simple regarding the application of tensor modeling and morphology operations, it provides a new insight to develop tensor-based morphology operations such that different morphology operations can be applied to hyperspectral images directly. This idea can be extended to other feature extraction methods bearing the similar properties such as extended attribute profiles. Furthermore, the tensor structural element is with high flexibility. By developing different structural elements, it would be promising to improve the performance of tensor morphological profile in hyperspectral image classification and create possibilities in other computer vision tasks such as Magnetic Resonance Imaging or videos processing.

Table 5.3: Overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) on the Pavia University dataset when different feature extraction methods were used: spectral feature (SPE), extended morphological profiles (EMP) and tensor morphological profiles (TMP).

Class	SPE	EMP	TMP
C1 Asphalt	81.0	**86.8**	77.5
C2 Meadows	59.1	88.0	**94.3**
C3 Gravel	**74.3**	39.3	47.8
C4 Trees	78.5	97.2	**99.0**
C5 Sheets	99.5	99.5	**99.7**
C6 Soil	93.7	**95.8**	**95.8**
C7 Bitumen	**87.3**	75.8	75.0
C8 Bricks	84.1	**96.7**	94.6
C9 Shadows	**99.4**	90.1	91.7
OA	73.3	88.0	**89.7**
AA	84.1	85.5	**86.2**
κ	0.67	0.84	**0.86**

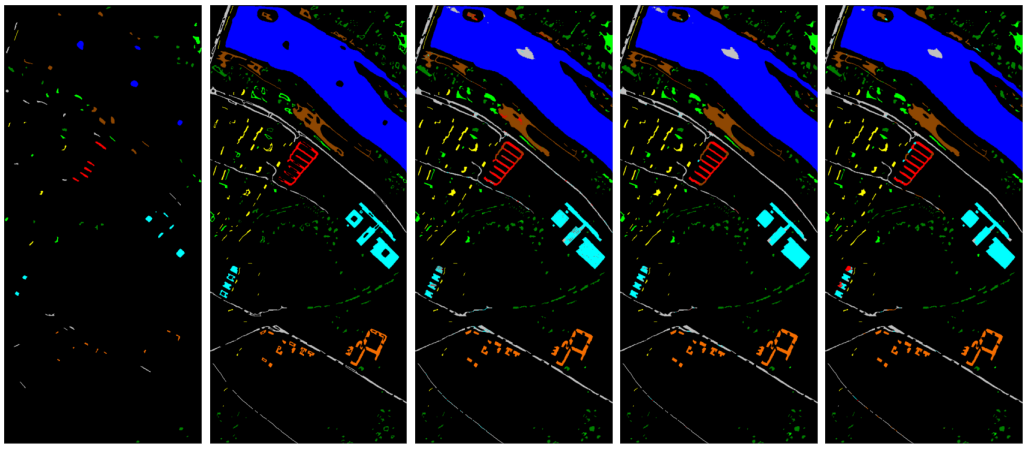


(a) Training map (b) Testing Map (c) SPE (d) EMP (e) TMP

Figure 5.6: Classification results on the Pavia University dataset with different methods: (a)training map; (b)testing map; (c)spectral feature (SPE);(d)extended morphological profiles (EMP) and (e)tensor morphological profiles (TMP).

Table 5.4: Overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) on the Pavia Center dataset when different feature extraction methods were used: spectral feature (SPE), extended morphological profiles (EMP) and tensor morphological profiles (TMP).

Class	SPE	EMP	TMP
C1 Water	98.5	98.6	**99.1**
C2 Trees	94.6	**96.9**	95.0
C3 Asphalt	**95.4**	94.0	93.8
C4 Bricks	72.5	71.7	**82.1**
C5 Bitumen	94.1	**98.3**	97.9
C6 Tiles	94.2	97.8	**97.9**
C7 Shadows	93.6	**95.6**	94.0
C8 Meadows	99.8	**100.0**	**100.0**
C9 Bare Soil	**99.9**	**99.9**	99.6
OA	97.0	97.7	**98.0**
AA	93.6	94.8	**95.5**
κ	0.95	**0.96**	**0.96**

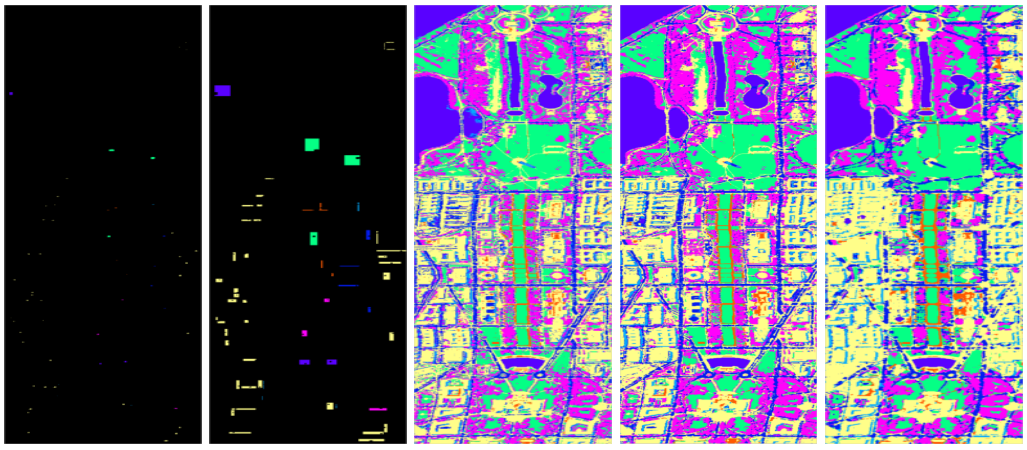


(a) Training map (b) Testing Map (c) SPE (d) EMP (e) TMP

Figure 5.7: Classification results on the Pavia Center dataset with different methods: (a)training map; (b)testing map; (c)spectral feature(SPE);(d)extended morphological profiles (EMP) and (e)tensor morphological profiles (TMP).

Table 5.5: Overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) on the Washington DC Mall dataset when different feature extraction methods were used: spectral feature (SPE), extended morphological profiles (EMP) and tensor morphological profiles (TMP).

Class	SPE	EMP	TMP
C1 Roofs	95.1	**95.9**	95.2
C2 Roads	97.3	99.7	**100.0**
C3 Grass	93.1	99.8	**100.0**
C4 Trail	94.0	94.0	**95.6**
C5 Water	99.9	**100.0**	99.8
C6 Shadows	97.2	**99.4**	98.1
C7	91.5	95.7	**96.5**
OA	94.6	97.6	**97.8**
AA	95.4	97.8	**97.9**
κ	0.93	**0.97**	**0.97**



(a) Training map (b) Testing Map (c) SPE (d) EMP (e) TMP

Figure 5.8: Classification results on the Washington DC dataset with different methods: (a)training map; (b)testing map; (c)spectral feature(SPE);(d)extended morphological profiles (EMP) and (e)tensor morphological profiles (TMP).

Chapter 6

On the Sampling Strategy for the Evaluation of Spectral-Spatial Methods in Hyperspectral Image Classification

6.1 Introduction

Spectral-spatial processing has attracted increasing attentions during the past several years. Bringing spatial information into traditional single pixel based spectral analysis leads to better modeling of local structures in the image and facilitates more accurate land-cover and object classification. While a large portion of the hyperspectral remote sensing community have focused their research on improving classification accuracy by developing a variety of spectral-spatial methods [116, 33, 41, 65], few attention has been paid to experimental settings. Evaluation of hyperspectral image classification methods requires careful design of experiments such as appropriate benchmark data sets, sampling strategy to generate training and testing data, and appropriate and fair evaluation criteria [116, 117]. In the scope of supervised classification, we find that traditional experimental designs for spectral processing are often improperly used in the context of spectral-spatial processing, leading to unfair or biased performance evaluation. This is particularly the case when training and test-

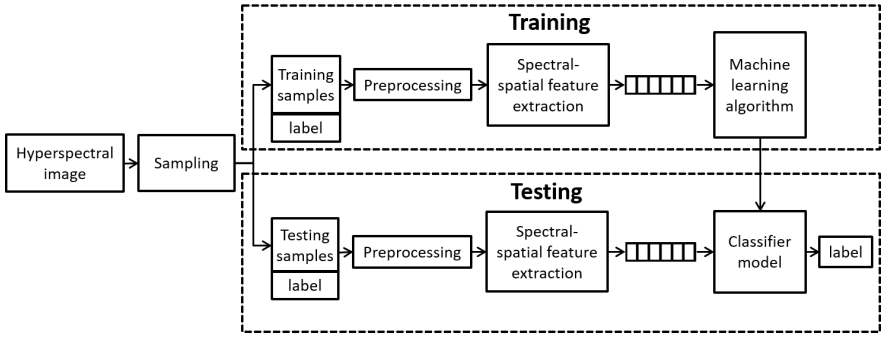


Figure 6.1: Framework of a supervised hyperspectral image classification system that uses spectral-spatial features.

ing samples are randomly drawn from the same image/scene which is a common setting in the hyperspectral image classification research due to limited availability of benchmark data and high cost of ground truth data collection.

Fig. 6.1 shows a typical spectral-spatial hyperspectral image classification system built on a supervised learning scheme. Training and testing samples are drawn from an image dataset following a specific sampling strategy. After image preprocessing which may involve spectral-spatial operations, feature extraction step fuses the spectral and spatial information to explore the most discriminative features for different classes. The extracted features are used to train a classifier that minimizes the error on the training set. In the testing step, the learned classifier is used to predict the classes of testing samples based on the extracted features. The testing error is given by comparing the predicted labels with the ground truth, which can be used as a performance indicator for image preprocessing, feature extraction and classification methods.

In the experimental setting, the sampling strategy plays an important role in the classifier learning and evaluation. Given a dataset including a hyperspectral image and its land-cover classes or other ground truth data, in most cases, training and testing samples are not given in advance. A sampling strategy has to be employed to create the training and testing sets [118, 2, 119]. Random sampling is a natural choice since it treats all labeled data equally and each sample would be selected with

the same probability. However, by this method, some classes with a small number of labeled samples may have much fewer selected samples than expectation. Therefore, a more sophisticated sampling method, stratified random sampling, is often used [2]. To guarantee each class having sufficient samples, it firstly groups those labeled samples into subsets based on their class labels, and then random sampling is carried out within each subset. In terms of the number of training samples in each subset, it normally requires that the proportion of each group should be the same as in the population. Then the rest of the samples are employed as testing samples in the testing step. This method is very simple to implement, reproducible, and of statistical significance. To the best of our knowledge, majority of existing hyperspectral image classification methods adopted this option in the experimental setting [33, 45, 41, 15, 57]. In the following sections, we refer to the stratified random sampling as random sampling.

Before proceeding to the issue of random sampling, we have to re-affirm some basic principles for supervised learning. Under a statistical learning framework, a common assumption for inference purpose is that random variables are independent and identically distributed (*i.i.d.*). The identical condition implies that training and testing samples are generated from the same data distribution. The independent condition requires that the occurrence of each sample do not affect the probability of other samples. *i.i.d.* shall hold for data in different forms, for example, both raw spectral responses and extracted features. Most supervised hyperspectral image classification approaches assume that data are *i.i.d.*. Pixels in the same class shall have similar spectral responses or spectral-spatial features so that a trained classifier can be generalized to predict the labels of unseen samples. However, the independent assumption does not always hold if the training and testing samples are not carefully selected.

In general, arbitrary samples selected from a population by random sampling can be seen roughly independent from each other, or at least independent between the

sets of training and testing samples. However, for hyperspectral images, the random sampling is usually undertaken on the same image. Consequently, those randomly selected training samples spread over the image and the testing samples will locate adjacent to them. Then the independence assumption would become jeopardized due to the spatial correlation between training and testing samples. This is not a problem for the traditional pixel-based spectral analysis methods in which no spatial information is used. However, when it comes to the spectral-spatial methods, the training and testing samples would inevitably interact with each other, and thus the dependence caused by overlap or partial overlap between the training and testing data could result in exaggerated classification accuracy. To be more specific, the information from the testing set could have been used in the training step through spatial operations, leading to a biased evaluation result. The sampling problem was originally noticed by Friedl et al. [117], who referred to overlap as auto-correlation. Zhen et al. [120] compared the influence of different sampling strategies to the classification accuracy. However, none of these work has given theoretical analysis on the problems and provided an effective solution. Therefore, it is necessary to revisit the sampling strategy and data dependence issue for supervised hyperspectral image classification, especially those based on spectral-spatial processing. In-depth discussion on this issue can be made from both experiment and the computational learning theory points of view.

In this chapter, we study the relationship between sampling strategies and the spectral-spatial processing in hyperspectral remote sensing image classification, when the same image is used for training and testing. We find that the experimental setting with random sampling makes data dependence on the whole image be increased by some spectral-spatial operations, and in turn increases the dependence between training and testing samples[1]. To address this problem, we propose an alternative controlled random sampling strategy to alleviate the side effect of traditional random

[1]For the sake of conciseness and without confusion, we use “dependence between training and testing data" and “data dependence" interchangeably in the rest of the chapter.

sampling on the same hyperspectral image. This leads to a fairer way to evaluate the effectiveness of spectral-spatial methods for hyperspectral image classification.

In summary, the contributions of this work are in three aspects. Firstly, we point out that the traditional random sampling from the same image experimental setting is not suitable for supervised spectral-spatial classification algorithms. This helps to re-examine the performance evaluation of various spectral-spatial classification methods. Secondly, we find that under the random sampling setting, spectral-spatial methods can enhance the data dependence and improve the classification accuracy. We give a theoretical explanation to this phenomenon via computational learning theory. Finally, we propose a novel controlled random sampling strategy which can greatly reduce the overlap between training and testing samples caused by spatial processing, such that more objective and accurate evaluation can be achieved.

The rest of this chapter is organized as follows. Section 6.2 reviews the spectral-spatial processing methods that has been commonly used for hyperspectral image classification. Section 6.3 provides an in-depth analysis on the dependency between training and testing samples. The spatial information embedded in the spectral-spatial processing under the experimental setting with random sampling is examined. Section 6.4 analyzed the overlap between neighboring training and testing samples caused by spatial operations. Such overlap increases the dependence between training and testing samples, which may lead to mistakenly using of the testing data in the training process. Section 6.5 discusses the relationship among spectral-spatial processing, data dependence and classification accuracy via computational learning theory. A new sampling strategy is proposed in Section 6.6 which reduces the influence of overlap between training and testing data. To prove its advantage over random sampling, a series of experiments are developed and results are presented in Section 6.7. At last, the conclusions are drawn in Section 6.8.

6.2 Spectral-spatial Processing in Hyperspectral Image Classification

The advantage of using hyperspectral data in land cover classification is that spectral responses reflect the properties of components on the ground surface [2]. Therefore, raw spectral responses can be used directly as the discriminative features of different land covers. At the same time, hyperspectral data also possesses the basic characteristic of conventional images - the spatial information which corresponds to where a pixel locates in the image. The spatial information can be represented in different forms, such as structural information including the size and shape of objects, textures which describe the granularity and patterns, and contextual information which can express the inter-pixel dependency [41]. This is also the foundation of the development of most spectral-spatial methods for hyperspectral image classification.

In general, spectral-spatial information can contribute to hyperspectral image classification through three ways. Firstly, in image preprocessing, it can be used for image denoising, morphology, and segmentation. Image denoising enables the reduction of random noises introduced from sensors, photon effects, and calibration errors. Several approaches have been exploited for this purpose, for example, smoothing filters, anisotropic diffusion, multi-linear algebra, wavelet shrinkage, and sparse coding methods [42]. In most cases, denoising can be done by applying a local filter with designed or learned kernel across the whole image. In mathematical morphology, operations are performed to extract spatial structures of objects according to their spectral responses [43, 41]. Similar information is explored in image segmentation, which groups spatially neighboring pixels into clusters based on their spectral distribution [44, 45].

Secondly, common usage of joint spectral-spatial information lies in the feature extraction stage. While traditional spectral features are extracted as responses at single pixel level in hyperspectral images, spectral-spatial feature extraction methods use spatial neighborhood to calculate features. Typical examples include tex-

ture features such as 3D discrete wavelet [15], 3D Gabor wavelet [46], 3D scattering wavelet[47], and local binary patterns [48]. Morphological profiles, alternatively, use closing, opening, and geodesic operators to enhance spatial structures of objects [30, 49, 34]. Other spectral-spatial features include spectral saliency [50], spherical harmonics [51], and affine invariant descriptors [52]. Heterogeneous features can be further fused using feature selection or reduction approaches [25].

Thirdly, some image classification approaches rely on spatial relation between pixels for model building. A direct way of doing so is calculating the similarity between a pixel and its surrounding pixels [53]. Markov random field, for example, treats hyperspectral image as dependent data and uses spectral information in the local neighborhood to help pixel class prediction [54, 45, 55]. Similar spatial structures are explored in conditional random fields [56], hypergraph modelling [39], and multi-scale analysis [57]. The spatial information can also be explored in constructing composite kernels in support vector machines [31]. While supervised learning approaches, such as K-nearest neighbors, linear discriminant analysis, Bayesian analysis, support vector machines, etc. are widely used in these classification tasks [58, 59], some approaches adopt semi-supervised or active learning strategies [60, 61].

6.3 Spatial Information Embedded in Random Sampling

Random sampling makes the training and testing samples spread over the image, embedding plenty of underlying spatial information. In this section, we point out that the embedded spatial information will mistakenly influence the classifier learning and evaluation. We exploit this problem in a specific/extreme way, by which a hyperspectral image classification task can even be done without spectral information.

In many benchmark hyperspectral datasets, pixels in the same class are not distributed randomly in an image. On the contrary, they tend to exist in continuous regions and follow a certain spatial distribution, especially when objects in the same

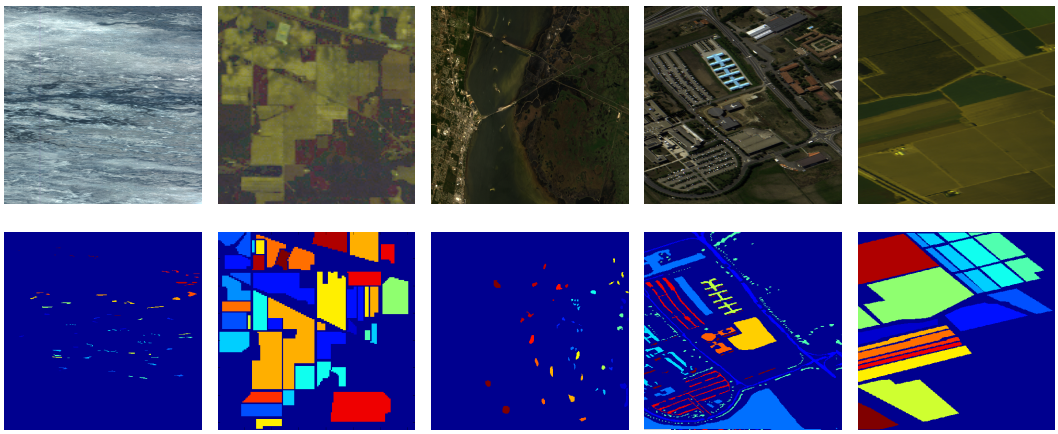


Figure 6.2: Three band false color composite (top row) and ground truth labels (bottom row) of five commonly used hyperspectral datasets. From left to right: Botswana, Indian Pines, Kennedy Space Center, Pavia University, and Salinas Scene.

materials present in the scene. Fig. 6.2 shows the false color composite and ground truth maps of five commonly used hyperspectral datasets, i.e., Botswana, Indian Pines (Indian), Kennedy Space Center (KSC), Pavia University (PaviaU), and Salinas scene (Salinas) [121]. In these images, there are strong dependencies between the spatial locations of pixels and land cover classes. This results in the potential using of the spatial structure and distribution of each single class. In most cases, if random sampling is used for selecting training and testing samples in the same image, the class label of a testing sample can be easily inferred only by its spatial relation to the training samples. This can be exemplified by Fig. 6.3, in which 5%, 10% and 25% of training data are sampled from the Indian Pines and Pavia University datasets. When it comes to 25% sampling rate, the spatial distribution of training samples (last column) is similar to the shape of the ground truth map (first column) in the spatial domain.

To show the extent that the classification accuracy is impacted by spatial information, we performed experiments on five benchmark datasets in Fig. 6.2. In the experiment, a nonlinear support vector machine (SVM) was employed because the land cover classes are not linearly separable in the spatial domain. The spatial coor-

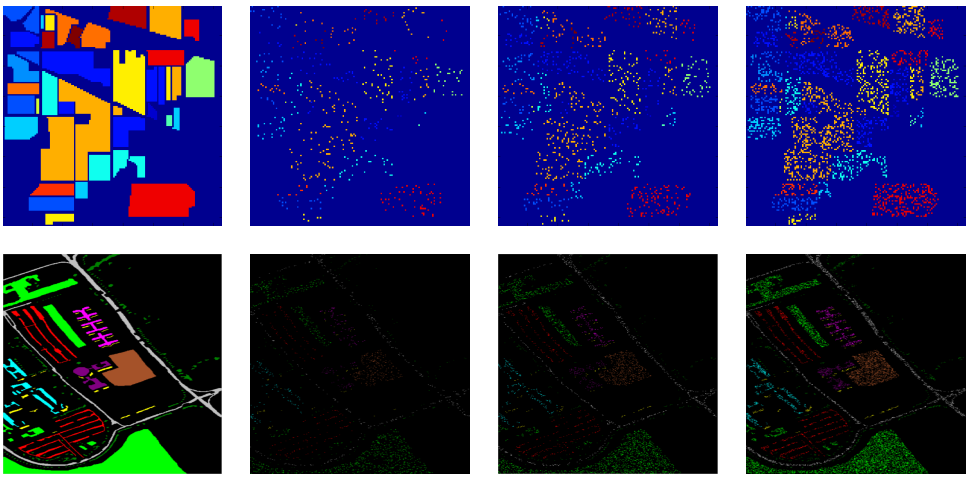


Figure 6.3: Random sampling strategy on Indian Pines (top row) and Pavia University (bottom row) datasets. From left to right: the ground truth map, training set with 5% sampling rate, training set with 10% sampling rate, training set with 25% sampling rate.

dinates were used as the spatial feature and no spectral information was included. The parameters of the SVM were learned via five-fold cross validation. Three sampling rates were explored, i.e. 5%, 10%, and 25% to generate the training data from all labeled samples, while the rest of labeled data served as the testing samples. In contrast to the spatial feature, the traditional spectral feature based methods were also implemented in which we followed the same setting as the spatial method.

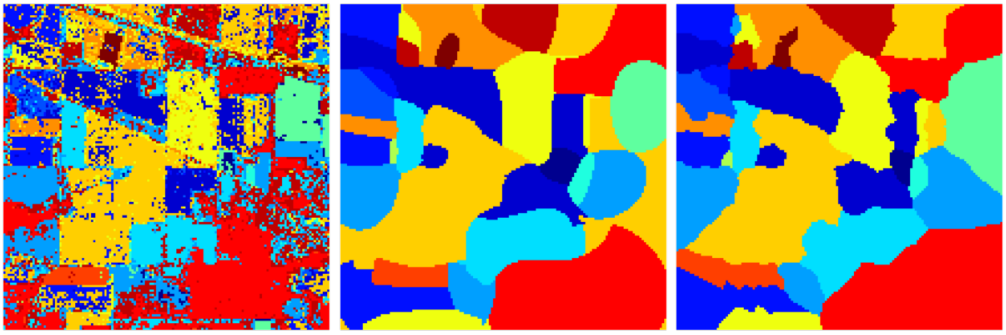
Each test was repeated ten times in the experiment with random generation of training and testing samples. The overall classification accuracies (OA), average accuracies (AA) and Kappa Coefficient (κ) are shown in Table 6.1 for different methods. The comparison between accuracies using spectral feature with SVM (*Spe*) and spatial feature with SVM (*Spa1*) shows some surprising results. Classification accuracy based on pure spatial feature has significantly outperformed the counterpart using pure spectral feature in all cases. In terms of overall accuracy, the spatial method achieves more than 93.8% accuracy on all datasets when only 5% of training samples are used, while the spectral method has only around $75.5\% - 93.2\%$ in accuracy. When the sampling rate becomes 25%, the accuracy almost reaches 100% for the spa-

Table 6.1: Overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) on five hyperspectral datasets when different feature/classifier combinations were used: spectral feature with SVM (Spe), spatial feature with SVM (Spa1) and spatial feature with KNN (Spa2).

Dataset	OA			AA			κ		
	Spe	Spa1	Spa2	Spe	Spa1	Spa2	Spe	Spa1	Spa2
Botswana (%5)	89.1	**93.8**	93.3	89.0	**93.8**	92.9	0.873	**0.933**	0.928
Botswana (%10)	91.9	**98.1**	97.7	92.7	**97.9**	97.5	0.913	**0.979**	0.975
Botswana (%25)	94.9	**99.7**	99.7	95.3	99.6	**99.7**	0.944	0.996	**0.997**
Indian (%5)	75.5	**95.5**	95.1	67.7	**92.1**	90.5	0.718	**0.949**	0.944
Indian (%10)	81.0	**98.0**	97.6	76.5	**97.1**	94.9	0.783	**0.977**	0.972
Indian (%25)	87.0	**99.7**	99.4	84.6	**99.5**	98.7	0.851	**0.996**	0.993
KSC (%5)	87.6	98.1	**98.8**	81.6	97.5	**98.5**	0.862	0.979	**0.987**
KSC (%10)	90.3	99.6	**99.8**	85.4	99.2	**99.7**	0.892	0.995	**0.998**
KSC (%25)	93.4	99.9	**100.0**	89.6	99.9	**100.0**	0.927	1.000	1.000
PaviaU (%5)	93.2	96.4	**96.9**	91.3	90.1	**93.3**	0.910	0.952	**0.958**
PaviaU (%10)	94.2	97.3	**98.7**	92.3	91.8	**96.8**	0.923	0.964	**0.982**
PaviaU (%25)	95.3	98.0	**99.7**	94.0	93.4	**99.2**	0.941	0.973	**0.996**
Salinas (%5)	93.1	**99.9**	99.2	96.2	**99.8**	98.1	0.923	**0.999**	0.991
Salinas (%10)	94.1	**99.9**	99.7	97.1	**99.9**	99.4	0.934	**0.999**	0.997
Salinas (%25)	95.3	**100.0**	99.9	97.8	**100.0**	**100.0**	0.948	**1.000**	**1.000**

tial feature, which agrees with the perceptual intuition in Fig. 6.3. Essentially, these phenomena are caused by the random sampling strategy on the same image. The results also show that higher sampling rate leads to increase of classification accuracy on all datasets.

In another point of view, the spatial classification can also be exploited in the local neighborhood. Since the training samples spread uniformly in the image, it would be easy to find the nearest training sample for any testing samples that belong to the same class. An experiment was designed to test how the local information con-



(a) (b) (c)

Figure 6.4: Classification maps of the Indian Pines (including the unlabeled pixels) using only spectral or spatial features: (a) *Spe*, (b) *Spa1* and (c) *Spa2*.

tributes the classification. We employed the K-nearest neighbor (KNN) classifier and set the parameter K to 1. The results are displayed in Table 6.1 under the columns of *Spa2*. It can be seen that the performance of *Spa2* is comparable to the spatial method Spa1 on all datasets, which has significantly outperformed the spectral method on all datasets. It should be noted that in the KNN classification, predicting the label of testing samples is only based on the nearest training pixels in their spatial neighborhood. This is similar to the mechanism of some spectral-spatial methods which also make use of the local spatial neighborhood information but in a different way. This experiment further proves that the training data provide too much information on the spatial domain for the classification task.

While classification based on spatial coordinates seems to perform better than the spectral information, it is infeasible in real applications in which unlabeled pixels are involved. Those unlabeled pixels are prone to be classified into its nearby class, thus producing a thematic map dramatically different from the ground truth. To exemplify this phenomenon, Fig. 6.4 shows the classification maps of the Indian Pines including the unlabeled pixels with 10% sampling rate. Although *Sp1* and *Sp2* achieve higher classification accuracy than *Spe*, their classification maps are far away from the ground truth map. Therefore this method is not acceptable in reality. In summary, these two experiments show that random sampling from the same image

makes an underestimated amount of spatial information be embedded in the training set and the testing set. It is natural to raise the concern that they would interact with each other if spatial processing is applied to the image.

6.4 Overlap between Training and Testing Data from the Same Image

The spectral-spatial methods make use of the spatial information in different forms and different ways as introduced in Section 6.2. When it comes to the random sampling strategy, a more severe problem may happen in the spectral-spatial analysis, especially at the feature extraction stage. When only spectral responses are used, feature extraction is performed at single pixel, without exploring its spatial neighborhood. Therefore, random sampling strategy provides a statistical solution for data splitting and there is no explicit overlap between training and testing samples. However, the spectral-spatial methods usually exploit information from neighborhood pixels. This is normally implemented by a sliding window with a specific size, for example, 3×3, 5×5 and so on. In each window, a kernel or filter is used to extract discriminative information. Since the training and testing samples are drawn from the same image, their features are almost certain to overlap in the spatial domain due to the shared source of information.

Fig. 6.5 shows the extent of overlap between training and testing data on the Indian Pines dataset. In the figure, the white dots show the locations of training samples, and the surrounding white squares cover a 3×3 region used for spectral-spatial feature extraction. The testing samples, however, may just lie in the square and has its own surrounding regions. This brings about a shared region between features extracted from the training and testing data such that they interact with each other and lose the mutual independence. It is also evident that a larger filter leads to more overlap areas. An example is shown in Fig. 6.6 in which a 3×3 and

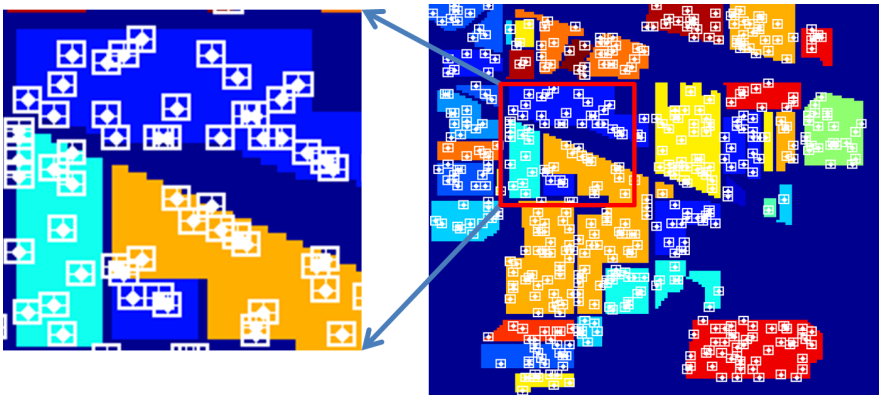
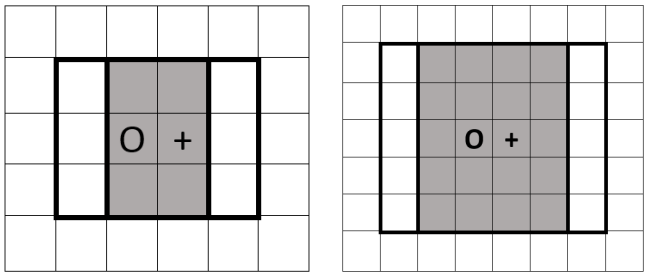


Figure 6.5: Overlap between training and testing data on Indian Pines dataset under 5% sampling rate.



(a) (b)

Figure 6.6: The regions for feature extraction from a training sample (O) and a testing sample (+) overlap with each other, as represented in gray color. The proportion of overlap is $\frac{2}{3}$ and $\frac{4}{5}$ for (a) 3×3 sliding window and (b) 5×5 sliding window, respectively.

5×5 window will result in $\frac{2}{3}$ and $\frac{4}{5}$ of overlap for adjacent training and testing samples, respectively.

Such overlap leads to using of the testing data for training purpose and gives significant advantages to the spectral-spatial feature extraction approaches. This violates the basic principle of supervised learning that training and testing data shall not interact with each other. Depending on how feature is extracted, benefit of testing data may be explicit, for example when the spectral-spatial feature is extracted by concatenating the spectral responses of pixels in a neighborhood, or implicit, for example, by extracting texture features based on spatial frequency analysis such as

discrete wavelet transform.

6.4.1 Experiment with a mean filter based spectral-spatial method

In order to estimate how the overlap impacts the accuracy of spectral-spatial method with random sampling strategy, an experiment was carried out on the Indian Pines dataset. In this experiment, a linear SVM classifier was used to facilitate further comparison. The features were constructed by applying a mean filter to calculate the mean of the spectral responses in a neighborhood of the hyperspectral images, which was mathematically formulated as follows:

$$f(x,y) = \frac{1}{MN} \sum_{i=x-\frac{M}{2}}^{x+\frac{M}{2}} \sum_{j=y-\frac{N}{2}}^{y+\frac{N}{2}} S(i,j) \tag{6.1}$$

where M and N are the width and height of neighborhood surrounding (x,y). In the experiment, we set M and N both from 1 up to 27 with an interval of 2. $S(i,j)$ represents the spectral response at location (i,j) and $f(x,y)$ is the feature extracted on location (x,y) which contains both spectral and spatial information. This process can be considered as one of the simplest approaches to extract spectral-spatial features.

When the size of the neighborhood is 1×1, this reduces to extracting spectral features only. Larger size of window results in more overlap. The calculated rate of testing samples covered by the neighborhood of training samples is shown in Fig. 6.7. When 5% training data are sampled, 30.9% testing samples are covered by the 3×3 regions used to extract training features. When random sampling rate increases to 25%, the extent of overlap becomes 86.4%. The rise of sampling rate leads to rapid increase of overlap. Furthermore, when the size of filter grows, the overlap rate also increases rapidly. Eventually, when the overlap rate reaches 100%, all testing samples are used in the training process.

The experiment was repeated ten times. In each time, the indices of the training and testing pixels were randomly generated. Features were generated using different

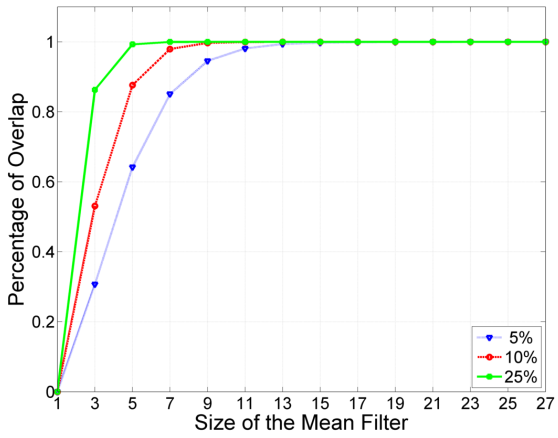


Figure 6.7: Overlap of training and testing data on the Indian Pines with different size filters.

settings of filter size and sampling rate. Under each setting, the same training, and testing samples were used for fair comparison. The overall classification accuracies are shown in Fig. 6.8. A significant increase of the classification accuracy can be observed when spatial information is added to the spectral information. When the size of neighborhood increases, more testing data contribute to the training step, therefore the classification accuracy increases. It is also interesting to see that after the neighborhood increases to a specific size, the accuracy stops growing and tends to be stable. This is probably because that when the neighborhood becomes too large, unlabeled data or samples from other classes are involved in the feature extraction, which neutralizes the benefits of overlap.

6.4.2 Non-overlap measurement

Other than overlap, the increase of classification accuracy also owes to the better discriminative capability of spectral-spatial features. With larger filter size, the feature includes more spatial information. To demonstrate how the spatial neighborhood influences the effectiveness of spectral-spatial feature, we performed another experiment on those testing samples not overlapped with the training data.

Following the same setting as the previous experiment, we removed the testing

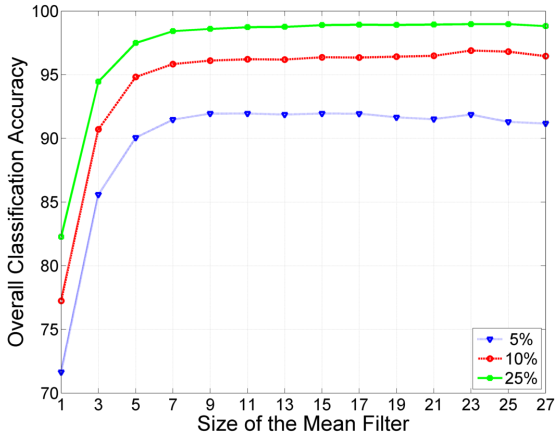


Figure 6.8: Classification accuracies on the Indian Pines using a simple mean filter with different filter sizes.

Table 6.2: Classification accuracies on all testing samples and non-overlapped testing samples.

Filter Size	1	3	5	7	9	11
All samples (5%)	72.1	86.1	90.2	91.4	92.1	92.3
Non-overlap (5%)	72.1	82.9	83.2	79.1	71.6	68.0
All samples (10%)	77.4	90.4	94.5	95.9	96.1	96.4
Non-overlap (10%)	77.4	86.2	84.8	77.9	65.5	NaN
All samples (25%)	82.4	94.6	97.5	98.3	98.5	98.7
Non-overlap (25%)	82.4	87.9	80.6	NaN	NaN	NaN

samples that were covered by the training set and only test on the remaining samples. Table 6.2 shows the comparison of classification accuracy on all testing samples and non-overlap testing samples. The results show that when testing on non-overlap testing samples, the accuracy is improved when the neighborhood information is initially introduced by the 3×3 mean filter. However, when a larger size filter is used, the accuracy of non-overlap testing samples does not increase and even decrease [2]. The decrease could be caused by the fact that the non-overlap testing samples are

[2]In Table 6.2, the null values are due to the absence of non-overlapped testing samples.

easily influenced by the samples from other classes in the neighborhood. In contrast, the classification accuracy with overlapped testing samples has remarkable improvement when a larger filter size is used.

Based on the above analysis, under the random sampling strategy, some filter-based spectral-spatial feature extraction methods would make the training and testing samples overlap and then interact with each other. Subsequently, in the training process, information from testing samples are included to train the classifier, which in return is used to classify the testing samples in the testing step. Although this kind of methods improves the classification results, they are not desired because they violate the basic assumption of supervised learning and their generalization is questionable. So far we have only analyzed a special case of spectral-spatial methods, it would be interesting to extend the analysis to a broader scope. Next, we try to discuss the data dependence and its impact on classification results by computational learning theory.

6.5 Data Dependence and Classification Accuracy

Computational learning theory aims to analyze computational complexity, feasibility of learning, and performance bound [122]. A widely known computational learning framework is the probably approximately correct (PAC) learning which estimates the sample complexity based on the required generalization error, probability of inference and complexity of a space of functions. Another classic theory is the Vapnik-Chervonenkis theory (VC theory). One of its functions is to bound the generalization ability of learning processes which is usually represented as the testing error $R(h)$.

Before introducing the computational learning theory, some basic learning concepts shall be firstly introduced in the scope of *i.i.d.* data. In computational learning, instead of considering classification accuracy, a more general term, generalization error bound, is usually derived to describe the ability of learning algorithms to predict unseen data. For a binary classification problem, given a hypothesis $h \in H$ where H

are all hypotheses, a target hypothesis c, and a sample set $S = (x_1, x_2, ..., x_m)$ following a distribution D, the empirical error (training error) $\hat{R}(h)$ and the generalization error (testing error) $R(h)$ can be defined as:

$$\hat{R}(h) = \frac{1}{m} \sum_{i=1}^{m} \; l(h(x_i), c(x_i)) \tag{6.2}$$

$$R(h) = \mathop{\mathbb{E}}_{x \in D} \; l(h(x), c(x)) \tag{6.3}$$

where l is the error function and $\mathbb{E}$ is the expectation in statistics. $R(h)$ describes the expectation of error considering the whole distribution D.

Despite that the empirical error $\hat{R}(h)$ can be calculated once the training data S, its label $c(x_i)$ and the hypothesis h are known, the generalization error cannot be estimated directly. In practice, simply decreasing $\hat{R}(h)$ by building a complex classification model may not always minimize $R(h)$ because it may lead to overfitting. In order to bound $R(h)$, more factors have to be considered. Based on PAC learning, the generalization bound can be calculated as:

$$R(h) \leq \hat{R}(h) + \frac{1}{m}(log|H| + log\frac{1}{\delta}) \tag{6.4}$$

which means that given training data of size m and hypothesis complexity $|H|$, the inequality of generalization holds with probability no less than $1 - \delta$. This definition conforms to our understanding of learning that more training data leads to better learning outcome. Based on the inequality, the generalization bound can be tightened by increasing the training sample size m or by decreasing the probability $1 - \delta$ which is equivalent to the inference confidence. The complexity of hypothesis is determined by the learning models.

When the hypothesis sets are infinite, the above bound is uninformative. In order to impose generalization bound for infinite cases, the Redemacher complexity is introduced to measure the hypothesis complexity [123]. Specifically, it measures the

variety of a set of functions by estimating the degree to which a hypothesis can fit random noise. The Rademacher complexity based generalization bound on *i.i.d.* data samples is defined as:

$$R(h) \leq \hat{R}(h) + \hat{\mathfrak{R}}_s(H) + 3\sqrt{\frac{log\frac{2}{\delta}}{2m}} \tag{6.5}$$

where $\hat{\mathfrak{R}}_s(H)$ is the empirical Rademacher complexity. $1-\delta$ is the probability or confidence and m is the training sample size. $\hat{\mathfrak{R}}_s(H)$ can be estimated by growth function or VC-dimension [122].

Even though these models provide generalization bounds for different learning algorithms, they are all based on the *i.i.d.* assumption. For *non-i.i.d.* data, the generalization bound has not been fully studied due to the lack of statistical model for dependent data. However, *i.i.d.* does not always hold in practice. In general, the samples in a hyperspectral image are not *i.i.d.*, as the samples are spatially overlapping to each other in the image. The data dependence will inevitably happen no matter how carefully the sampling strategy is designed.

In recent years, researchers begin to develop new learning theories on this topic. Among all kinds of *non-i.i.d.* data, some data types possess the property of asymptotic independence, which is weaker than independence but stronger than dependence, for instances, time series signal [124]. In order to define this kind of data, mixing condition is used to define explicitly the dependence of the future signal on the past signal based on decay. A commonly used model in *non-i.i.d.* scenario is the stationary β-mixing model [125]. Suppose events A and B are generated from a time sequence $\alpha_{t\in(-\infty,+\infty)}$ with an interval k, the definition of β-mixing coefficient is

$$\beta(k) = \sup_m \mathbb{E}_{B\in\alpha_{-\infty}^{m}} \left[\sup_{A\in\alpha_{m+k}^{+\infty}} |Pr(A|B) - Pr(A)| \right] \tag{6.6}$$

This equation defines the dependence coefficient as the supremum of the difference between the conditional probability $Pr(A|B)$ and probability $Pr(A)$ when choosing

arbitrary moment m which separates event A and B. The sequence α is β-mixing if $\beta(k) \to 0$ when $k \to +\infty$. It implies that the dependence coefficient $\beta(k)$ decreases with the increase of interval k.

Several learning models have already been derived on stationary β-mixing data, such as VC-dimension bound[125], PAC learning [126] and Rademacher complexity [127]. In this work, the Rademacher complexity based generalization bound is employed since it associates the generalization bounds with β-mixing coefficient. It uses a technique to transferring the original dependent data to independent blocks. Let 2μ be the number of blocks and each block contains k consecutive points, then the size of sample $m = 2\mu k$. The original bound in Equation 6.5 is extended to β-mixing data as follows:

$$R(h) \leq \hat{R}(h) + \hat{\mathfrak{R}}_s(H) + 3M\sqrt{\frac{log\frac{2}{\delta - 4(\mu-1)\beta(k)}}{2\mu}} \tag{6.7}$$

where M is the bound of a set of hypothesis H.

Compared to the *i.i.d.* case, this bound is not only related to the training error $\hat{R}(h)$, empirical Rademacher complexity $\hat{\mathfrak{R}}_s(H)$, and probability δ, but also relies on the β-mixing coefficient $\beta(k)$ which implies the degree of dependence among data. Considering the impact of β-mixing coefficient to the bound, this equation can be further simplified as:

$$R(h) \leq f(\beta(k)) + C \tag{6.8}$$

where $f(\beta(k))$ is a monotonically decreasing function. As a result, the generalization bound is tightened when the $\beta(k)$ increases, i.e. the dependence among data is enhanced.

Applying learning theory to hyperspectral image classification is challenging due to the complex statistical characteristic of hyperspectral images. To our knowledge, similar work is very rare. In the following experiments, we show that hyperspectral images share the same properties of β-mixing data.

Spectral feature extracted at image pixels often have strong dependence to their surrounding regions [65]. However, it is still questionable whether such dependence decreases with the increasing distance between the central pixel and its neighboring pixels. In addition, since a hyperspectral image is a three-dimensional data, how the dependence is related to the spatial direction is still unknown. To check how the dependence varies with the distance, we performed a simple statistical analysis on the Indian Pines dataset. Here, the dependence between two pixels X and Y is approximated by the linear correlation coefficient of their spectral responses:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma X \sigma Y} \tag{6.9}$$

where cov and σ represent the covariance and standard deviation, respectively. A random location was firstly selected on this image, then the correlation coefficient ρ was calculated between the pixel and its neighborhood pixels with different distances. We have calculated the result on a 9×9 patch. As expected, it does not show clear pattern at a single random pixel. Then we took into account the pixels centered at all locations and calculated the mean of correlation coefficients in all patches based on the following:

$$c(x_i, y_i) = \frac{1}{M \times N} \sum_{\hat{x} \in M, \hat{y} \in N} \rho(S(\hat{x}, \hat{y}), S(x_i, y_i)) \tag{6.10}$$

where (x_i, y_i) is the neighbourhood pixels of the central pixel $(\hat{x}, \hat{y})$ in the patch. For a 9×9 patch, $x_i = \hat{x} - 4, \hat{x} - 3, ..., \hat{x} + 4$ and $y_i = \hat{y} - 4, \hat{y} - 3, ..., \hat{y} + 4$. M and N are the width and height of the whole image. $S(x, y)$ represents the spectral response at location (x, y). The statistical result is shown in Fig. 6.9. In the center of the patch, the intensity is one due to self-correlation. It clearly shows that with the increasing interval, the correlation coefficient gradually drops in all directions. This is consistent with the characteristic of β-mixing.

Now we can safely assume that hyperspectral images are β-mixing, and explore

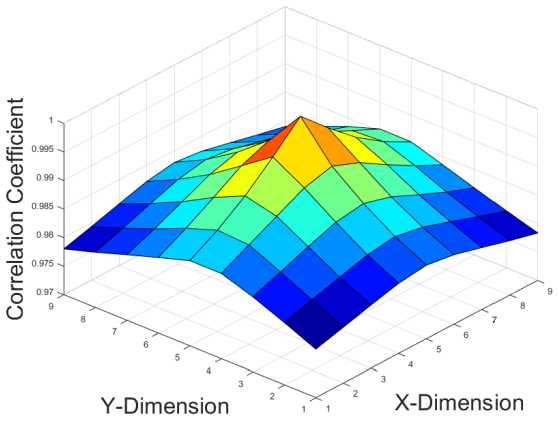


Figure 6.9: The statistics on the correlation coefficients between a central pixel and its 9×9 neighbourhood pixels of an image from the Indian Pines dataset.

the relationship between the generalization bound with data dependence. Based on Equation 6.8, the bound is inversely related to β-mixing. As a consequence, the classification accuracy can be increased by enhancing the dependence between training and testing data. Recall that in the experiment with a mean filter based spectral-spatial method (Fig. 6.8), the accuracy increases with larger filters. It would be interesting to measure the pixel correlation to see how the mean filter influences the data dependence. Similarly, Equation 6.1 was used to calculate the mean spectral responses in a sliding window and the sizes of the windows were set to 1×1, 3×3, 5×5, 7×7, 9×9, 11×11, respectively. Then the statistics on the correlation coefficients on a 9×9 patch was calculated from the filtered images based on Equation 6.9 and Equation 6.10. Note when the 1×1 was used, this corresponds to using the original image without filtering. The results are shown in Fig. 6.10 in which we only draw the correlation coefficients along X dimension. The results with different filters are represented by distinctive curves. The bottom curve with 1×1 filter can be seen as the projected illustration along X dimension from Fig. 6.9. From this new figure, two trends can be observed. Firstly, all curve drops continuously when the distance increases, which means that the processed data agree with the properties of β-mixing. Secondly, at the same distance, the larger the filter is, the stronger the dependence

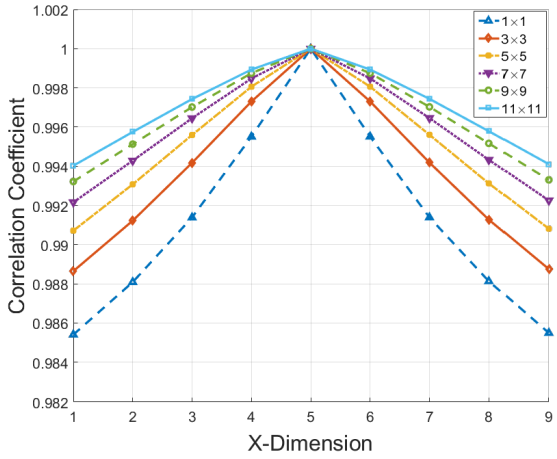


Figure 6.10: Pixel correlations along X dimension in an image from the Indian Pines dataset after a mean filter with different sizes was applied.

between the central pixel and its adjacent pixels becomes. Therefore, the overlap enhances the data dependence which tightens the error bound of the final classification results.

The theories presented above have explained why mean filter improves the classification accuracy, and can be extended to other spectral-spatial operations that increase the data dependence. It should be noted that the above analysis is built on the assumption of random sampling for performance evaluation. Under such experimental setting, the improvement of classification accuracy comes from not only incorporating spatial information into a classifier but also enhancing the dependence between training and testing data. The former is the main purpose of algorithm performance evaluation and the later should be avoided.

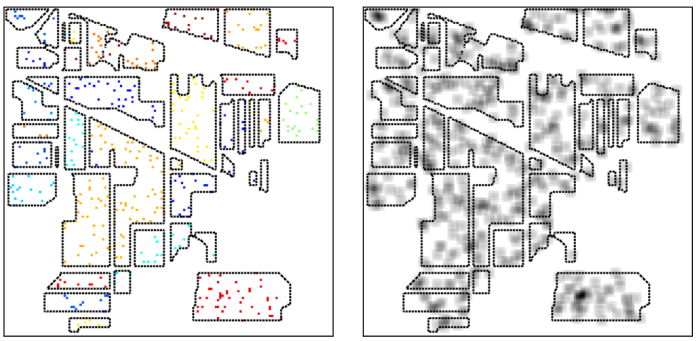
6.6 A Controlled Random Sampling Strategy

Following the discussion in previous sections that random sampling from the same image is not suitable for evaluation the spectral-spatial methods, it is necessary to develop a new sampling strategy to separate the training and testing sets without overlap. It would be perfect if we could perform training and testing on two dif-

Algorithm 3 Controlled Random Sampling Strategy

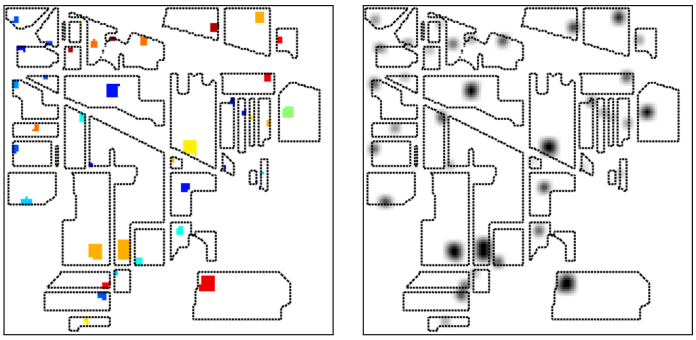
Require: Hyperspectral Image I and sampling rate s
  for each class c in I **do**
    Selects all unconnected partitions P in the class c
    for each partition p in P **do**
      Count the number of samples n_p in the partition
      Calculate the number of training samples n_t in the partition by $n_t = n_p \times s$
      Randomly select a seed point q in the partition
      Applying the region-growing algorithm to extend q to a region r whose size is equal to n_t
    end for
    Combine these regions r to form training samples R_c
  end for
  Combine the training samples R_c and their corresponding class labels to get the whole training set R

ferent images. Unfortunately, this is still infeasible in most cases due to the limited availability of benchmark datasets and high cost of ground truth data collection. Therefore, without changing much the current experimental setting, the goal is to significantly reduce the extent of data overlap and make the evaluation fair enough. Based on our analysis, the main problem of random sampling is that it makes the training and testing samples spatially adjacent to each other, leading to their overlap in the subsequent spatial operations. On the other hand, as a classical method, it has advantages such as simplicity, reproducibility, and statistical significance. As a result, the new sampling strategy should satisfy the following requirements. Firstly, it shall avoid selecting samples homogeneously over the whole image so that the overlap between training and testing set can be minimized. Secondly, those selected training samples should also be representative in the spectral domain, meaning that it shall adequately cover the spectral data variation in different classes. There is a paradox between these two properties, as the spatial distribution and the spectral distribution are coupling with each other. The first property tends to make the training samples clustered so that it generates less overlap between the training and testing data. However, the second property prefers training samples being spatially distributed as random sampling does and covering the spectral variation in different



(a) Random sampling

(b) Random sampling after Gaussian filtering



(c) Controlled random sampling

(d) Controlled random sampling after Gaussian filter

Figure 6.11: Overlap between the training and testing data under different sampling strategies before (first column) and after (second column) a Gaussian filter is applied.

regions of the image. Therefore, a good trade-off has to be achieved by the new sampling strategy. Thirdly, because there is no prior knowledge, we do not know which samples are more important than the others. Therefore the new method shall possess the property of randomness.

Here we propose a controlled random sampling strategy to achieve a compromise of the above considerations. Similar to random sampling, a pre-defined proportion of samples in each class is to be randomly selected as the training samples and the rest data serve as the testing samples. Those training samples shall be concentrated locally and dispersed globally. We borrow the idea of region growing to create region-shape training samples [128]. On the ground truth map, the seed points are randomly selected from different partitions of classes to make the training samples

disperse globally and randomly. The partition referred here is a group of connected pixels with the same labels. For each class, there are usually several partitions distributed on the map, corresponding to the land cover of the same category at different locations. Then controlled random sampling proceeds with three steps. Firstly, it selects the unconnected partitions for the each class and counts the samples in each partition. This step is to find the spatial distribution of each class and make sure that the selected training samples in the next step cover the spectral variance at the most extent. Secondly, for each partition, the training samples are generated by extending region from the seed pixel. In terms of region growing, it expands in all directions and takes account of 8-connected neighborhood pixels. All the adjacent pixels of seed pixels are examined and if they are within the same class, they work as the new seed points. This process is repeated until the sum selected points reaches a pre-defined number which is proportional to the number of pixels in the corresponding partition. This guarantees that the total number of training samples meet the pre-defined proportion of the whole data population. Thirdly, after the above steps are applied to all classes, those samples in the grown regions with their labels are chosen as the training samples and the rest of pixels work as the testing samples. In case when there are more partitions than the required training samples, partitions are again randomly sampled. A summary of this strategy is given in Algorithm 3.

In Fig. 6.11, we demonstrate different degrees of overlap between training and testing samples under random sampling and controlled random sampling strategies, after a Gaussian filter is applied. In the left column of the figure, the training and testing data are represented by colored dots and white regions in each partition. Applying the Gaussian filter creates the gray regions in the right column of the figure, representing the overlap between the training and testing data. It can be noticed that all the training samples are impacted by the testing data under random sampling. On the contrary, for controlled random sampling, only training samples at the edges of the training regions are influenced by the testing data. This figure

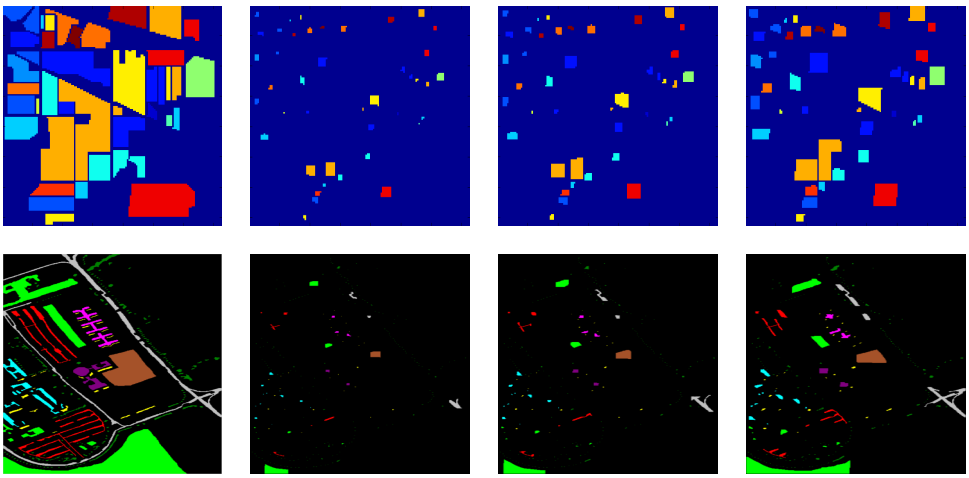


Figure 6.12: Controlled random sampling strategy on the Indian Pines (top row) and Pavia University (bottom row) datasets. From left to right: the ground truth map, training set with 5%, 10%, and 25% sampling rates, respectively.

clearly shows that the overlap from controlled random sampling is significantly less than that from the traditional random sampling.

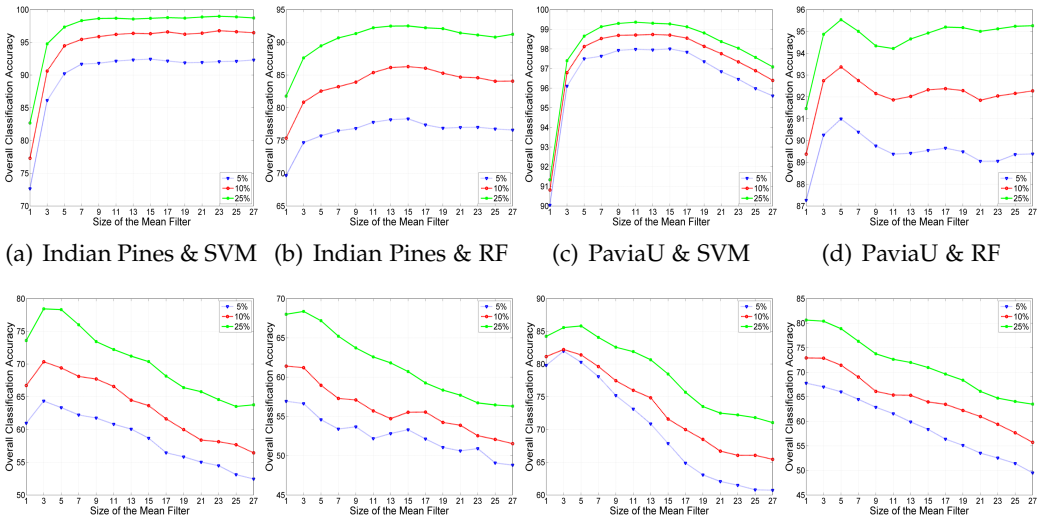
To further illustrate how the controlled random sampling works with real datasets, examples on Indian Pines and Pavia University are given in Fig. 6.12 with 5%, 10%, and 25% sampling rates. Compared to the random sampling strategy in Fig. 6.3, it can be observed that the spatial structure of each class can no longer be inferred from the training data as random sampling does. In the meantime, the training samples are still distributed across the whole image and a wide range of spectral variances are covered. Though this approach cannot completely eliminate overlap between the training and testing data, the influence of testing data in the training stage can be greatly reduced to limited pixels at the boundaries of each training region. The experimental setting with the proposed sampling method can help us more accurately and objectively evaluate the performance of spectral-spatial methods.

6.7 Experiments

To prove the usability and advantage of the proposed controlled random sampling against random sampling, we have developed a series of experiments to test these two strategies when they are used to evaluate spectral-spatial operations in different stages of image classification. In the preprocessing step, we adopted a mean filter and a Gaussian filter as examples of smoothing and denoising operations. Then, we performed experiments with raw spectral feature to examine the effectiveness of the proposed sampling method when evaluating the spectral responses without spatial processing. Furthermore, two spectral-spatial feature extraction methods, i.e. 3D discrete wavelet and morphological profiles, were compared using two sampling methods. Finally, we compared the performance of two sampling strategies at varying sampling rates and discussed the limitation of the proposed method. In order to make the experiments more convincing, we adopted two widely used supervised classifiers, support vector machine (SVM) and random forest (RF) [71] to validate our results. The SVM was implemented using the LIBSVM package [68], and the RF was implemented using the well-known Weka 3 data mining toolbox [129]. We present results on five benchmark datasets, i.e., Botswana, Indian Pines (Indian), Kennedy Space Center (KSC), Pavia University (PaviaU), and Salinas scene (Salinas).

6.7.1 Evaluation of spectral-spatial preprocessing method

The spectral-spatial preprocessing step contributes to classification by improving the quality of hyperspectral images, reducing random noises, and enhancing specific features. By varying the parameters of mean filter and Gaussian filter, their influence to the classification accuracy under two sampling strategies can be analyzed. We undertook experiments on both Indian Pines and Pavia University datasets with SVM and RF, respectively. The results with mean filter are shown in Fig. 6.13. When traditional random sampling is used, the accuracy on the Indian Pines dataset increases with larger filter size when SVM and RF are adopted (Fig. 6.13(a) and (b)). For the Pavia



(a) Indian Pines & SVM (b) Indian Pines & RF (c) PaviaU & SVM (d) PaviaU & RF

(e) Indian Pines & SVM (f) Indian Pines & RF (g) PaviaU & SVM (h) PaviaU & RF

Figure 6.13: Classification accuracies vary with the size of mean filter on the Indian Pines and Pavia University (PaviaU) datasets under random sampling (first row) and controlled random sampling (second row) strategies.

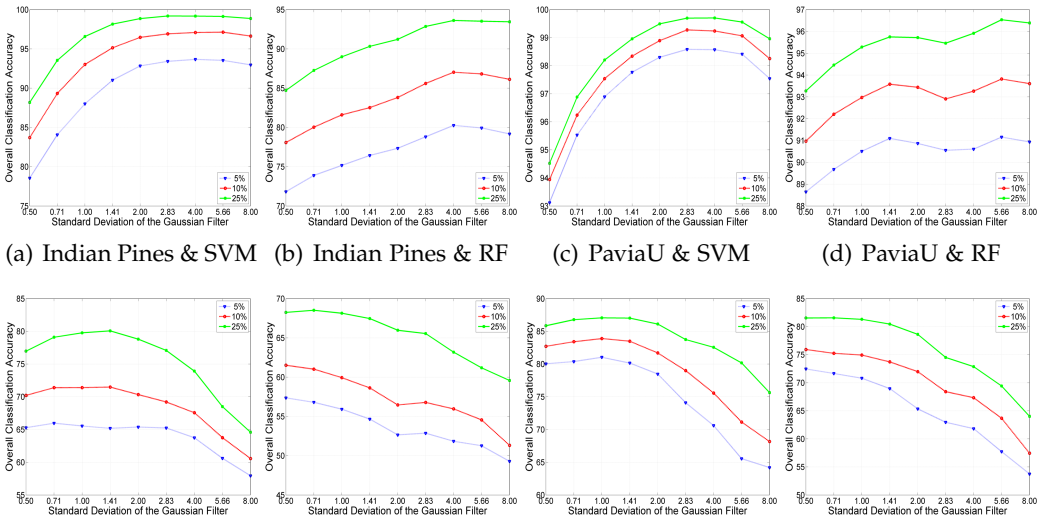
University dataset (Fig. 6.13(c) and (d)), the accuracies also increase with larger filter sizes but decrease when the size reaches a specific value, which is slightly different from the results on the Indian Pines image. The reason may be that Pavia University has higher spatial resolution and interacts with filters in more complex way than the low spatial resolution Indian Pines data. The results confirm that using a mean filter with relative large size can increase the classification accuracy, up to 92.4% on Indian Pines and 98.0% on Pavia University. Essentially, it is mainly because larger filters lead to more overlap between the training and testing data. In contrast, when adopting the controlled random sampling strategy, the classification accuracy firstly improves marginally, but then dramatically drops with larger size filters. This is consistent with our expectation in evaluating the influence of spectral-spatial operations rather than the data dependence. Therefore, the proposed sampling method successfully avoids the problem of random sampling.

We then performed an experiment with Gaussian filter under the same settings to compare two sampling strategies. Among different denoising and smoothing ap-

proaches, Gaussian filter is a basic but effective tool to reduce the random noise in hyperspectral images. It works as a low-pass filter whose standard deviation controls the shape of filter and sets the threshold to remove the corresponding high-frequency signal. The larger the stand deviation is, the lower frequency the signal can be preserved and the image be more smoothed. We applied a Gaussian filter on each band of hyperspectral images with a range of standard deviations. The size of filters varies with the standard deviation so that the smoothing effect decays to nearly zero at the boundaries of filtering masks. Then the smoothed image was fed into the classifier. This experiment was repeated ten times and the mean of overall accuracy was used as the evaluation criterion. The standard deviation ranged from 2^{-1} to 2^{3} with an interval of 0.5 on the exponential term.

We plot the classification accuracy as a function of the standard deviation in Fig. 6.14 for random sampling and controlled sampling method, respectively. From Fig. 6.14 (a)-(d), we can see that the accuracy continuously increases until a specific point when Gaussian filter with larger standard deviation is used with random sampling strategy. This is consistent with the observation on the mean filter. We can assume that the Gaussian filter influences the data dependence to varying extents under different standard deviations, such that the classification accuracy is impacted by the filter parameter. This is also consistent with our earlier analysis that when data dependence is increased, the classification error bound will be tightened. However, this is not desired when evaluating a preprocessing method for image classification as we would like to know what is the actual contribution from the operation itself.

Compared to the random sampling, the controlled random sampling presents a different trend between the accuracy and standard deviation. The accuracy firstly improves marginally and then becomes stable or drops. This indicates that smoothing with an appropriate Gaussian operator can remove noises, and thus contribute to the final image classification. However, when the standard deviation of Gaussian is very large, too strong image smoothing does not help much for the discrimination



(a) Indian Pines & SVM (b) Indian Pines & RF (c) PaviaU & SVM (d) PaviaU & RF

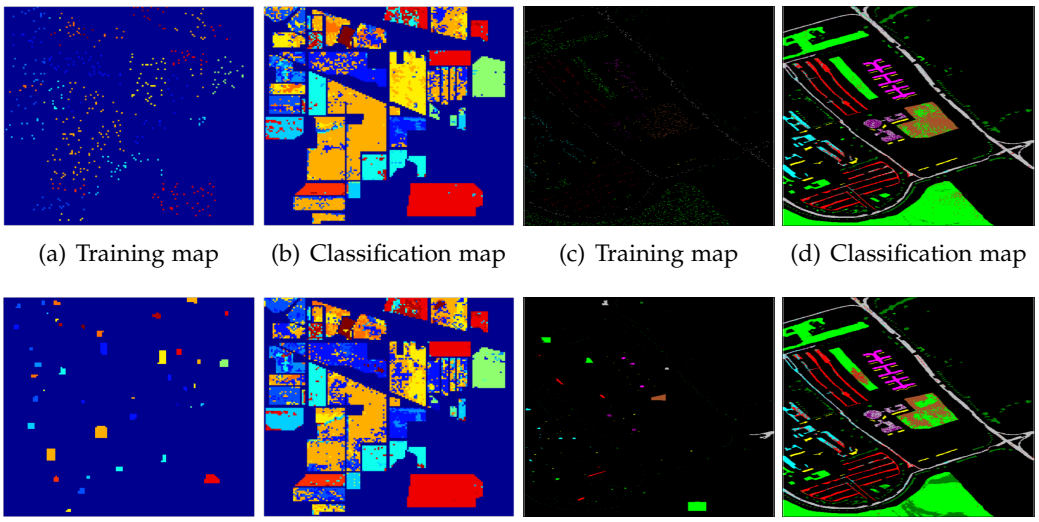
(e) Indian Pines & SVM (f) Indian Pines & RF (g) PaviaU & SVM (h) PaviaU & RF

Figure 6.14: Classification accuracies vary with the standard deviation of Gaussian filter on the Indian Pines and Pavia University datasets under random sampling (first row) and controlled random sampling (second row) strategies.

of different classes since it may mix the training data with unlabeled data at boundaries of image regions, thus losing its adaptability. Under the new sampling strategy, Gaussian filter is able to improve the classification but not very significantly and the training and testing data dependence caused by overlap is no longer the dominant factor to the classification. Overall, these two experiments prove that the proposed sampling strategy can neutralize the improper benefit gained from enhancement of dependence between training and testing data.

6.7.2 Raw spectral feature

We then performed an experiment to compare two sampling strategies when raw spectral features were used on the benchmark datasets. The objective of this experiment is to examine the effectiveness and objectiveness of the proposed sampling method compared to random sampling. As mentioned in Section 6.1, there is no issue with the experimental setting of random sampling when evaluating a pixel based spectral feature. But we still do not know whether the proposed sampling method is



(a) Training map (b) Classification map (c) Training map (d) Classification map

(e) Training map (f) Classification map (g) Training map (h) Classification map

Figure 6.15: Training/classification maps on the Indian Pines and Pavia University datasets under random sampling (first row) and controlled random sampling (second row) strategies, when raw spectral features are used.

qualified in such a task.

In the experiment, only the raw spectral features were used without any spatial processing. Other settings were the same as the previous experiment such as the classifiers, repetition of experimental runs, etc. The overall accuracy and standard deviation under random sampling and the controlled random sampling strategies(*) are reported in Tables 6.3 and 6.4 for SVM and RF, respectively. Following observations can be made from the results. Firstly, higher sampling rate leads to increase of classification accuracy on all datasets. This is the same and expected for both sampling methods. Secondly, the standard deviation of the accuracy from the proposed sampling strategy is much higher than that of the random counterpart. This is due to the distinction of training data generated from the random seeds each time. Lastly, there is a reduction on the classification accuracy when the proposed sampling strategy is used. This is due to the fact that variations in the same class data in different regions are less sufficiently captured as some of them may not be included in the training samples when the proposed sampling strategy is used. The difference of

accuracies is more evident on Indian Pines, Pavia University and Salinas datasets as these scenes include large blocks of regions in the same class, which leads to more benefits from spectral variation covered by random sampling strategy. For further illustrating this phenomenon, the classification maps on the Indian Pines and Pavia University under two sampling strategies are shown in Fig. 6.15. Compared to random sampling, those testing samples far away from the training regions are easily misclassified under controlled random sampling.

Despite the differences, this does not affect a fair evaluation of different algorithms with the proposed sampling strategy. In this experiment, assuming that the goal is to evaluate SVM and RF, it can be concluded from the results that SVM is a preferred classifier since it generates higher classification accuracy. Therefore, although the new sampling strategy has made the hyperspectral image classification a more challenging problem and forces more rigorous evaluation to the feature extraction and classification approaches, it is still qualified in evaluating the algorithms in hyperspectral image classification.

6.7.3 Spectral-spatial features

Now we turn our attention to testing the proposed sampling strategy with two typical spectral-spatial feature extraction methods, i.e., 3D discrete wavelet transform (3D-DWT) and morphological profile. 3D-DWT is a typical example of filter-based methods. The morphological profile is a widely adopted spatial feature extraction method, including a number of variations for hyperspectral image classification.

6.7.3.1 3D discrete wavelet transform

The discrete wavelet transform is derived from the wavelet transform which is a mathematical tool for signal analysis. Unlike Fourier transform, the advantage of wavelet transform is that the transformed signal provides time-frequency representation for the non-stationary signal, meaning that we cannot only know whether a

frequency component exists but also when it happens in a signal.

The experimental results under random sampling strategy and controlled random sampling strategy(*) are shown in Table 6.3 and Table 6.4 for SVM and RF, respectively. As expected, controlled random sampling strategy leads to lower accuracy compared to random sampling strategy on all datasets. An interesting observation can be obtained by comparing these results with the results on the raw spectral feature in Tables 6.3 and 6.4. On the one hand, 3D-DWT performs better than raw spectral feature under both sampling methods. This indicates that the proposed method confirms that 3D-DWT is able to extract more discriminative information than raw spectral feature. On the other hand, under experimental setting with random sampling, 3D-DWT significantly improves the accuracy on all datasets over raw spectral feature. However, when it is tested with the proposed controlled sampling strategy, the improvement cannot reach the same level of significance, especially on Indian Pines, Pavia University, and Salinas datasets. It means that 3D-DWT does not perform significantly better than the raw spectral features as expected when the advantage of introducing information from the testing data into the training stage is eliminated.

6.7.3.2 Morphological profile

To further analyze this issue, we undertook experiments on the mathematical morphology feature. Morphological operations apply the structuring elements to the image, making it possible to enhance or alleviate structures based on the specific requirements from users. The basic operators include erosion and dilation which expands and shrinks the structures, respectively. Combining them results in the opening (erosion-dilation) and closing (dilation-erosion) operations. These two processes can remove specific structures and noises without destroying the original primary structures in the image. The results of morphological processing are called morphological profiles. The morphological profile based feature extraction method is

able to explore the structures of objects based on the contrast and size of objects in the images, therefore, it has been widely studied for hyperspectral image classification [30, 34].

The classification results with two sampling strategies are shown in Table 6.3 and Table 6.4. Similar to the results on 3D-DWT, although the morphological profile feature has achieved better performance than the raw spectral method when tested with random sampling strategy, the improvement is not as significant as when controlled random sampling is used. This is mainly because that spectral-spatial method does not take much advantage of the overlapped information between training and testing samples under the proposed method.

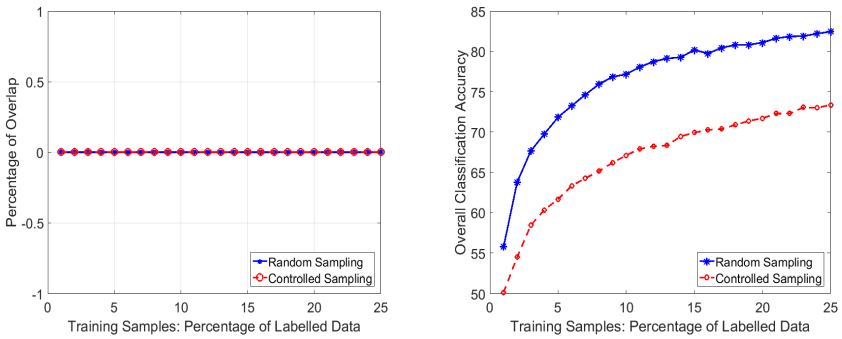
Directly comparing two completely different spectral-spatial methods may not make much sense since different features are more suitable to extract features on specific datasets or sensitive to specific classifiers. Here we analyze the results from another point of view, which may explain the advantage of the proposed sampling over random sampling. In Table 6.3, 3D-DWT achieves higher accuracy than EMP on both Indian Pines and Pavia University datasets when random sampling is adopted. When adopting the new sampling strategy, 3D-DWT still performs slightly better than EMP on the Indian Pines dataset, but EMP performs significantly better than 3D-DWT on the Pavia University dataset. This is consistent with the fact that the morphology method is capable of extracting more spatial structures than 3D-DWT on the dataset with high spatial resolution [30]. Under the proposed sampling method, the properties of the spectral-spatial method can be more accurately reflected and evaluated in the experiments. This is impossible under random sampling because the classification result is strongly misled by the overlap between training and testing samples. Overall, the proposed sampling strategy reveals more real discriminative ability of the spectral-spatial methods, which is the purpose of the evaluation.

6.7.4 Relationship between traditional random sampling and the proposed method

In previous experiments, we have used at least 5% of labeled data as training samples. In recent years, some researchers have used very few labeled samples to construct the training set. In such case, the potential overlap between training set and testing set would not be significant. To show how the number of training samples influence two sampling strategies, especially when a small number of training samples were used, we developed an experiment to measure the performance of random sampling and controlled sampling when 1% to 25% of labeled samples were used as the training samples. In the experiment, if the number of samples in the training set was less than 1 for some small classes, we set the size of training set to 1. Regarding the feature extraction, we employed the raw spectral response and a mean filter with size of 5×5 for the spectral-spatial method, respectively. Other settings were the same as previous experiments, such as classifier, repeating times, and so on. To measure the differences and similarities between two sampling methods, we have calculated both overall classification accuracy and the percentage of overlap[3], and explore the relevance of overlap and classification results under different sampling rates.

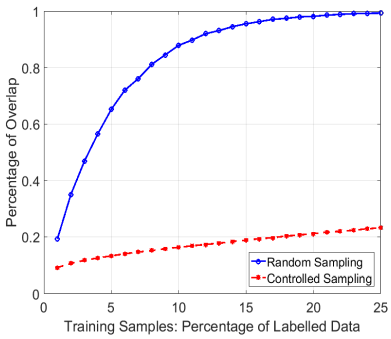
The results are drawn in Fig. 6.16 where the first and second rows represent the results with the raw spectral response and after applying a 5×5 mean filter, respectively. In Fig. 6.16 (a), it is expected that no overlap between training and testing data exists for both two sampling methods when the pure spectral feature is exploited. Therefore, the classification results should not be biased. Fig. 6.16 (b) shows that with more training data available, the classification accuracy gradually increases. The rising rates under two sampling strategies show similar trends and there is no apparent difference in tendency. However, if considering the results with the 5×5 filter (spectral-spatial feature) in the second row, the overlap and accuracy grow very differently under two sampling approaches. Firstly, when the sampling

[3]The percentage of overlap is defined as the percentage of the testing samples covered by the training features against all the testing samples, as described in Section 6.4.

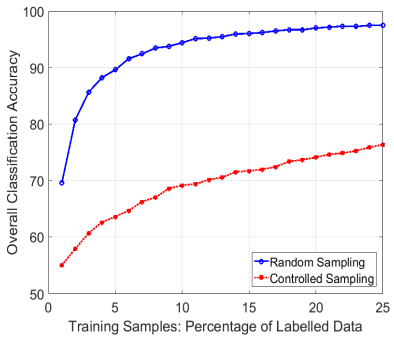


(a) Overlap

(b) Accuracy



(c) Overlap



(d) Accuracy

Figure 6.16: Overlap of training and testing data and classification accuracy on the Indian Pines dataset under random sampling and controlled random sampling strategies, when a different percentage of labeled data were used as the training samples. The first row shows results of raw spectral responses and the second row shows the results of a 5×5 mean filter.

rate is quite low (1%), the overlap between training samples and testing sampling for both methods are similar, which indicates that the proposed method is close to random sampling in the degree of overlap and cannot fully remove the overlap. With the sampling rate increasing, the overlap of random sampling dramatically increases to 1 (100%) before the sampling rate reaches 25%. In contrast, the overlap of controlled sampling increases very slowly, only reaching to 0.2 with the sampling rate of 25%. This figure demonstrates again that the proposed method can greatly reduce the overlap compared to the traditional random sampling, especially when the scale of training set becomes large. In Fig. 6.16 (d), it can be observed that the classification accuracy under random sampling increases much faster than that

of controlled sampling when the sampling rates rise, which is quite similar to the trend on the overlap in Fig. 6.16 (c). It reveals that under random sampling, the classification result is much more biased due to the overlap when the number of training samples increases. In contrast, the classification results are not much affected under the proposed method when comparing Fig. 6.16 (d) with (b).

From this experiment, we have the following conclusions. Firstly, when the training set is very small, there is not much overlap between training and testing set when spectral-spatial operations are involved. No matter the training set is generated by the traditional random sampling or the proposed method, there will only be a few testing samples involved in the training stage. Therefore, the evaluation is not much biased. Secondly, with the scale of the training set growing, the overlap between training and testing samples dramatically increases under the random sampling strategy whereas the proposed one can suppress such increase. Then the proposed method would be much less biased than the traditional one. Conversely, when the size of the training set decreases, the proposed method becomes closer to the traditional sampling strategy in the degree of overlap and objectiveness. In particular, when only one sample is used for each class, two sampling methods are equivalent to each other.

6.8 Conclusion

This chapter presented a comprehensive study on the influence of the widely adopted sampling strategy for performance evaluation of the spectral-spatial methods in hyperspectral image classification. We point out that random sampling has some problems because it has ignored the overlap and spatial dependency between training and testing samples when they are selected from the same image. Based on the *non-i.i.d.* characteristic of hyperspectral image data, we proved that the improvement of classification accuracy by some spectral-spatial methods are partly due to the enhancement of dependence between training and testing data, compared with

sole spectral information based methods. This phenomenon is more obvious when a large amount of training data are used. An alternative controlled random sampling strategy is proposed to alleviate these problems. It has been proved that although the proposed method cannot completely remove the overlap when the training set is very small, it will suppress the increase of overlap when the training set grows. This new strategy provides a better way to evaluate the effectiveness of spectral-spatial operations and the corresponding classifiers.

Finally, it should be noted that the aim of this chapter is not to criticize the spectral-spatial methods themselves or the exploration of spatial information. The concern is only on the widely adopted evaluation approach, or more strictly s-peaking, on the experimental setting. Under the experimental setting with random sampling, the performance evaluation may be not equally fair and unbiased for all spectral-spatial methods. This is especially the case for the practice that training and testing are performed on the same image. This problem is ultimately due to the lack of labeled hyperspectral data that are available for method evaluation. Therefore, a more urgent task for the research community is to build more benchmark datasets to facilitate future spectral-spatial hyperspectral image analysis research.

Table 6.3: Classification accuracies (overall accuracy and standard deviation) using raw spectral feature, 3D discrete wavelet transform (3D-DWT) and extended morphological profile (EMP) with random sampling and controlled random sampling(*), when linear SVM is adopted.

Dataset	**Raw Spectral Feature**			**3D-DWT**			**EMP**		
	%5	%10	%25	%5	%10	%25	%5	%10	%25
Botswana	88.6 (1.4)	92.2 (1.1)	95.0 (0.6)	96.2 (0.7)	97.7 (0.5)	99.4 (0.3)	95.5 (2.0)	98.3 (0.6)	99.5 (0.2)
Botswana*	87.4 (1.4)	90.7 (0.5)	93.0 (0.4)	95.1 (1.4)	95.9 (1.3)	96.6 (0.8)	95.4 (1.2)	96.6 (0.7)	97.6 (0.5)
Indian	72.5 (0.7)	77.1 (0.8)	82.4 (0.4)	88.1 (0.8)	93.7 (0.7)	97.9 (0.3)	83.0 (1.1)	88.2 (0.7)	92.4 (0.4)
Indian*	63.8 (2.2)	68.2 (1.7)	75.0 (1.6)	65.2 (3.0)	69.9 (3.0)	79.1 (1.7)	64.8 (2.8)	69.2 (2.7)	77.2 (2.8)
KSC	76.1 (0.9)	80.4 (0.9)	86.3 (0.7)	87.7 (1.8)	91.8 (0.6)	96.4 (0.5)	76.9 (1.0)	83.3 (0.8)	89.1 (0.5)
KSC*	73.8 (2.1)	78.5 (1.1)	83.8 (0.7)	81.6 (2.2)	83.7 (2.6)	87.9 (1.1)	72.3 (3.6)	78.3 (2.7)	84.5 (1.5)
PaviaU	89.9 (0.2)	90.7 (0.3)	91.3 (0.2)	97.8 (0.1)	98.6 (0.1)	99.3 (0.1)	97.0 (0.2)	97.6 (0.1)	98.1 (0.1)
PaviaU*	80.9 (3.9)	82.7 (4.0)	84.5 (4.3)	84.8 (3.2)	86.4 (3.4)	89.2 (3.6)	87.4 (3.1)	89.5 (1.7)	91.7 (1.9)
Salinas	92.4 (0.1)	92.8 (0.1)	93.1 (0.1)	96.4 (0.1)	97.3 (0.1)	98.3 (0.1)	94.5 (0.3)	95.0 (0.2)	95.3 (0.1)
Salinas*	81.8 (2.7)	81.6 (3.8)	83.0 (3.3)	80.9 (3.0)	82.2 (3.6)	83.4 (3.4)	83.5 (1.4)	85.0 (2.4)	84.8 (2.4)

Table 6.4: Classification accuracies (overall accuracy and standard deviation) using raw spectral feature, 3D discrete wavelet transform (3D-DWT) and extended morphological profile (EMP) with random sampling and controlled random sampling(*), when RF is adopted.

Dataset	**Raw Spectral Feature**			**3D-DWT**			**EMP**		
	%5	%10	%25	%5	%10	%25	%5	%10	%25
Botswana	82.4 (1.6)	85.4 (0.8)	88.7 (0.5)	90.5 (1.8)	94.3 (1.1)	97.7 (0.5)	90.4 (1.6)	94.2 (1.0)	97.3 (0.9)
Botswana*	80.5 (1.6)	82.8 (2.0)	85.9 (1.4)	88.3 (1.5)	90.7 (1.4)	93.5 (1.1)	88.4 (1.6)	91.6 (1.2)	94.6 (1.1)
Indian	70.5 (0.9)	75.6 (0.9)	81.4 (0.5)	75.1 (1.4)	81.6 (0.8)	89.7 (0.5)	81.1 (1.5)	88.0 (0.9)	93.7 (0.6)
Indian*	56.7 (1.5)	61.3 (2.5)	66.7 (2.9)	57.4 (2.1)	61.4 (2.1)	67.5 (1.6)	64.4 (1.5)	69.6 (2.7)	76.2 (3.0)
KSC	82.9 (0.6)	86.6 (0.8)	90.1 (0.3)	82.2 (1.3)	88.7 (0.7)	92.9 (0.6)	87.1 (1.0)	91.8 (1.1)	95.5 (0.5)
KSC*	77.1 (2.4)	79.8 (1.7)	84.5 (1.7)	74.8 (2.2)	79.1 (2.1)	84.6 (1.6)	80.6 (2.8)	84.4 (1.5)	89.8 (1.8)
PaviaU	87.3 (0.4)	89.3 (0.2)	91.4 (0.1)	92.4 (0.3)	94.1 (0.2)	96.1 (0.1)	95.5 (0.2)	97.1 (0.2)	98.5 (0.1)
PaviaU*	71.2 (5.0)	73.6 (4.3)	81.4 (2.1)	75.8 (3.6)	79.0 (3.2)	83.4 (2.2)	78.2 (5.2)	80.9 (4.5)	88.2 (2.3)
Salinas	90.3 (0.1)	91.5 (0.2)	93.0 (0.1)	93.0 (0.3)	94.2 (0.2)	95.7 (0.1)	94.9 (0.3)	96.4 (0.2)	97.7 (0.2)
Salinas*	79.0 (2.6)	80.9 (3.9)	84.1 (2.6)	77.6 (2.9)	80.6 (2.6)	83.8 (1.8)	82.0 (1.6)	84.4 (2.8)	86.6 (2.6)

Chapter 7

Conclusion

7.1 Summary

With the development of advanced imaging instruments, hyperspectral imaging provides an alternative way to analyze and tackle traditional problems in remote sensing and computer vision. To employ abundant spectral and spatial information from hyperspectral images, it is necessary to develop a series of methods to extract highly comprehensive and discriminative representation of interested objects, thus facilitating basic pattern recognition tasks, such as object detection and recognition, and image classification.

Existing methods from remote sensing and computer vision focus on either grayscale based spatial feature or pixel wise spectral feature. Although research has been done on fusing the spatial and spectral information, there is high demand on developing novel, effective, and efficient spectral-spatial feature extraction methods for various applications from industry and society.

This thesis introduces three novel spectral-spatial feature extraction methods for hyperspectral image classification. Each method focuses on a fundamental topic in computer vision or remote sensing. These methods are derived from traditional two-dimensional approaches and then extended to hyperspectral images. It is worthwhile to point out that the spatial or spectral information to be extracted varies depending on different tasks or objectives. Spectral saliency extends salient object detection to spectral domain, making the saliency detection beyond human vision. 3D local derivative pattern extracts the spectral-spatial textures and improves the recognition

rate for hyperspectral image recognition. Tensor morphological profile provides a new method based on mathematical morphology for remote sensing image classification. Apart from feature extraction methods, in the end, we propose the controlled random sampling strategy, with which the performance of different spectral-spatial feature extraction methods can be fairly evaluated. The significance of the proposed feature extraction methods, as well as the sampling strategy, are summarized as follows.

Chapter 3 addresses the problem of salient object detection with hyperspectral images. As far as we know, it is the first attempt to incorporate spectral responses into saliency detection. The color bands in the classic Itti's model are replaced with the spectral responses by a variety of methods. We further incorporated spatial distribution of spectral responses into the saliency model, which has successfully detected the salient regions in close-range hyperspectral images. Though this method is still within the framework of Itti's human vision inspired model, it has proved the feasibility of using hyperspectral imaging for saliency detection.

Chapter 4 explores novel spectral-spatial feature for face recognition with hyperspectral imaging. We developed a three-dimensional local derivative pattern to characterize hyperspectral images and encode the spectral response in a similar way with the spatial pattern. Then each hyperspectral face is represented as a histogram of the extracted spatial and spectral binary patterns. The problem of this method is that the encoding in spectral dimension is very sensitive to noises, and it only contains the turning points without covering the full information embedded in the spectral curves. Although applying a Gabor filter before the feature extraction can partially solve the problem, there is space to further improve this method.

Chapter 5 introduces an innovative spectral-spatial feature extraction method based on tensor modeling and mathematical morphology. The proposed tensor morphological profile extracts not only spatial structures but also spectral information. However, this method generates a high-dimensional feature, weakening its effective-

ness in hyperspectral image classification due to the curse of dimension. A promising solution is to use the tensor decomposition to reduce the feature dimension and remove those correlated features. Despite this drawback, this framework is quite encouraging since a variety of morphology operations can be used for hyperspectral directly, such as the morphological attribute profile.

At last, Chapter 6 points out that the traditional random sampling approach is not suitable to evaluate the spectral-spatial classifier in remote sensing hyperspectral image classification because the overlap between training and testing samples may bias the classification accuracy in the evaluation. Thus the actual advantages of the spectral-spatial operations cannot be fairly assessed. With the proposed controlled random sampling, newly developed methods can be evaluated properly. This work will be quite beneficial to the research community and promote the evolution of spectral-spatial feature extraction methods for hyperspectral image classification.

7.2 Future work

In this thesis, the proposed feature extraction methods are hand-crafted features without learning procedure, thus limiting them to specific applications. This is a result of missing large-scale hyperspectral image datasets. To our knowledge, existing close-range datasets only consist of general indoor or outdoor scenes whose scale is not big enough to support large-scale image classification or machine learning. Building a hyperspectral image dataset usually takes significantly longer time than conventional image dataset collection. The hyperspectral images have to be captured one by one with hyperspectral instruments by the researcher, rather than crawling from the internet. Furthermore, due to its working principle, capturing a hyperspectral image requires stricter environments in order to obtain high quality images.

Faced with this challenge and opportunity, we have already embarked on build large-scale hyperspectral image datasets for various computer vision applications,

including strawberry quality analysis, object detection, and video analysis. Taking strawberry quality analysis as an example, it is a collaborating research project with a local strawberry farm in Queensland, Australia, as part of a Linkage Project supported by the Australian Research Council. The goal is to improve the quality of packed strawberry sold to supermarkets by building a grading system, which automatically rates the strawberry with consideration of ripeness, size, bruise, etc. Among these defects, bruise detection is quite difficult from RGB or grayscale images as the fresh bruise is invisible in the visible spectrum. We are in the process of building a dataset containing both good and bruised strawberries using both hyperspectral cameras and chromatic cameras. We expect that building these datasets will greatly benefit the research community in computer vision and pattern recognition.

Furthermore, salient object detection via hyperspectral imaging is of great potential in many applications. Current saliency detection methods are mainly built via mimicking the human vision system so as to detect outstanding structures, textures or colors. In the natural environment, animals and insects tend to camouflage themselves, making them difficult to be found by the visual system. However, when considering their spectral responses, these targets can be treated as the salient object and detected based on their spectral differences with the neighboring environment. We will continue to build hyperspectral saliency datasets, and explore new methods and applications of this research.

Last but not least, object tracking in open environments is of increasing interest and it is also a challenging task in computer vision. A problem in the task comes from the fact that some objects may be easily mixed with complex background scenes in both shape and color. For example, a pedestrian dressed in green clothes may not be easily detected from trees. By developing spectral-spatial methods for hyperspectral image based tracking, high discriminative ability can be achieved for tracking systems to distinguish targets from background. This may also handle partial occlusion problems by matching spectral signatures of objects, thus alleviating its dependence

on spatial features.

Overall, there are tremendous opportunities for researchers to explore hyperspectral imaging with typical computer vision tasks. Spectral-spatial feature extraction will provide essential tools to make these work successful and benefit research, industry, and society.

Bibliography

1. T. King, “Human color perception, cognition, and culture: Why "red" is always red,” *IS&T-The Society for Imaging Science and Technology*, vol. 5667, no. 1, pp. 234–242, 2005. (cited on page 1)

2. J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 3rd ed., 1999. (cited on pages 1, 4, 19, 24, 26, 73, 98, 99, and 102)

3. S. Han, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, “Camera spectral sensitivity estimation from a single image under unknown illumination by using fluorescence,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 805–812, 2012. (cited on page 2)

4. R. Kawakami, H. Zhao, R. T. Tan, and K. Ikeuchi, “Camera spectral sensitivity and white balance estimation from sky images,” *International Journal of Computer Vision*, vol. 105, no. 3, pp. 187–204, 2013. (cited on page 2)

5. I. Marin-Franch and D. Foster, “Estimating information from image colors: An application to digital cameras and natural scenes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 78–91, 2013. (cited on page 2)

6. S. Joo Kim, F. Deng, and M. S. Brown, “Visual enhancement of old documents with hyperspectral imaging,” *Pattern Recognition*, vol. 44, no. 7, pp. 1461–1469, 2011. (cited on page 2)

7. Z. Fu, A. Robles-Kelly, and J. Zhou, “MILIS: Multiple instance learning with

instance selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 958–977, 2011. (cited on pages 2 and 37)

8. K. Mikolajczyk and C. Schmid, "Performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005. (cited on page 2)

9. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012. (cited on pages 2 and 4)

10. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. (cited on page 2)

11. R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998. (cited on pages 3 and 23)

12. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. (cited on pages 4 and 20)

13. A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, no. SUPPL. 1, 2009. (cited on pages 7 and 74)

14. J. Liang, J. Zhou, Y. Gao, J. Liang, J. Zhou, and Y. Gao, "3D Local Derivative Pattern for Hyperspectral Face Recognition," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, pp. 1–6, IEEE, may 2015. (cited on pages 8 and 25)

15. Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2276–2291, 2013. (cited on pages 8, 25, 27, 29, 38, 74, 99, and 103)

16. A. A. Robles-Kelly and C. P. Huynh, *Imaging Spectroscopy for Scene Analysis*. Springer, 2013. (cited on pages 19 and 26)

17. D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002. (cited on pages 20, 21, and 22)

18. T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006. (cited on page 20)

19. B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533–544, 2010. (cited on pages 20, 57, and 70)

20. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2564–2571, 2011. (cited on page 20)

21. F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. I, pp. 604–610, 2005. (cited on page 21)

22. H. Kong, J.-y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 1–10, 2009. (cited on page 21)

23. D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *Journal of the Optical Society of America A*, vol. 23, no. 10, p. 2359, 2006. (cited on page 22)

24. M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991. (cited on page 23)

25. X. Jia, B.-C. Kuo, and M. Crawford, "Feature mining for hyperspectral image classification," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 676–697, 2013. (cited on pages 23, 27, and 103)

26. V. Z. V. Zubko, Y. J. K. Y. J. Kaufman, R. I. B. R. I. Burg, and J. V. M. J. V. Martins, "Principal Component Analysis of Remote Sensing of Aerosols Over Oceans," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 730–745, 2007. (cited on page 24)

27. B. Paskaleva, M. Hayat, Z. Wang, J. Tyo, and S. Krishna, "Canonical correlation feature selection for sensors with overlapping bands: Theory and application," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 10, pp. 3346–3358, 2008. (cited on page 25)

28. Q. Jackson and D. A. Landgrebe, "Adaptive bayesian contextual classification based on markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2454–2463, 2002. (cited on page 25)

29. D. Landgrebe, *Signal theory methods in multispectral remote sensing*. Wiley, 2003. (cited on pages 25 and 73)

30. J. Benediktsson, J. Palmason, and J. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans-*

actions on Geoscience and Remote Sensing, vol. 43, no. 3, pp. 480–491, 2005. (cited on pages 25, 27, 74, 103, and 131)

31. G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite Kernels for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, pp. 93–97, jan 2006. (cited on pages 25, 27, and 103)

32. Y. Tarabalka, J. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2973–2987, 2009. (cited on page 25)

33. Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 3973–3985, Oct 2011. (cited on pages 25, 97, and 99)

34. M. Dalla 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," *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 3, pp. 542–546, 2011. (cited on pages 25, 27, 74, 103, and 131)

35. L. Shen and S. Jia, "Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 5039–5046, 2011. (cited on pages 25, 28, and 74)

36. L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor Discriminative Locality Alignment for Hyperspectral Image Spectral-spatial Feature Extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, pp. 242–256, jan 2013. (cited on pages 25, 31, 75, and 82)

37. T. Fuai and J.-S. Lai, "Feature Extraction of Hyperspectral Image Cubes Using

Three-Dimensional Gray Level Cooccurrence," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 6, pp. 3504–3513, 2013. (cited on page 25)

38. W. Duan, S. Li, and L. Fang, "Spectral-Spatial Hyperspectral Image Classification Using Superpixel and Extreme Learning Machines," *Pattern Recognition*, pp. 159–167, 2014. (cited on pages 25 and 36)

39. R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, and X. Li, "Spectral-spatial constraint hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1811–1824, 2014. (cited on pages 25, 27, and 103)

40. J. Li, X. Huang, P. Gamba, J. M. B. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1592–1606, 2015. (cited on page 25)

41. M. Fauvel, Y. Tarabalk, J. Benediktsson, J. Chanussot, and J. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013. (cited on pages 26, 27, 28, 56, 74, 77, 89, 97, 99, and 102)

42. M. Ye, Y. Qian, and J. Zhou, "Multi-task sparse nonnegative matrix factorization for joint spectral-spatial hyperspectral imagery denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2621–2639, 2015. (cited on pages 26 and 102)

43. S. Velasco-Forero and J. Angulo, "Spatial structures detection in hyperspectral images using mathematical morphology," in *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, 2010. (cited on pages 27 and 102)

44. Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and

Markov random fields," *Pattern Recognition*, vol. 43, no. 7, pp. 2367–2379, 2010. (cited on pages 27, 37, and 102)

45. J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 809–823, 2012. (cited on pages 27, 37, 99, 102, and 103)

46. S. Jia, L. Shen, and Q. Li, "Gabor feature-based collaborative representation for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1118–1129, 2015. (cited on pages 27 and 103)

47. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2467–2480, 2015. (cited on pages 27, 74, and 103)

48. W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681–3693, 2015. (cited on pages 27, 36, and 103)

49. M. Fauvel, J. A. J. 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, pp. 3804–3814, nov 2008. (cited on pages 27, 74, and 103)

50. J. Liang, J. Zhou, X. Bai, and Y. Qian, "Salient object detection in hyperspectral imagery," in *20th IEEE International Conference on Image Processing*, pp. 2393–2397, 2013. (cited on pages 27, 74, and 103)

51. F. Nina-Paravecino and V. Manian, "Spherical harmonics as a shape descriptor for hyperspectral image classification," in *Proc. SPIE 7695, Algorithms and Tech-*

nologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI, p. 76951E, 2010. (cited on pages 27, 74, and 103)

52. P. Khuwuthyakorn, A. Robles-Kelly, and J. Zhou, "Affine invariant hyperspectral image descriptors based upon harmonic analysis," in *Machine Vision Beyond the Visible Spectrum* (R. Hammoud, G. Fan, R. McMillan, and K. Ikeuchi, eds.), Springer, 2011. (cited on pages 27, 74, and 103)

53. H. Pu, Z. Chen, B. Wang, and G. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7008–7022, 2014. (cited on pages 27 and 103)

54. Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, 2010. (cited on pages 27, 32, and 103)

55. L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Geoscience and Remote Sensing Letters*, vol. 53, no. 3, pp. 1490–1503, 2015. (cited on pages 27 and 103)

56. P. Zhong and R. Wang, "Learning conditional random fields for classification of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1890–1907, 2010. (cited on pages 27, 37, and 103)

57. L. Fang, S. Li, X. Kang, and J. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7738–7749, 2014. (cited on pages 27, 99, and 103)

58. J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and

J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013. (cited on pages 27 and 103)

59. P. Legendre, "Spatial autocorrelation: Trouble or new paradigm?," *Ecology*, vol. 74, no. 6, pp. pp. 1659–1673, 1993. (cited on pages 27 and 103)

60. J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 844–856, 2013. (cited on pages 27 and 103)

61. L. Wang, S. Hao, Y. Wang, Y. Lin, and Q. Wang, "Spatial-spectral information-based semisupervised classification algorithm for hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 8, pp. 3577–3585, 2014. (cited on pages 27 and 103)

62. T. G. Kolda, B. W. Bader, and S. N. Laboratories, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009. (cited on pages 30, 32, and 82)

63. S. Velasco-Forero and J. Angulo, "Classification of hyperspectral images by tensor modeling and additive morphological decomposition," *Pattern Recognition*, vol. 46, no. 2, pp. 566–577, 2013. (cited on pages 31 and 75)

64. S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Springer, 2009. (cited on page 32)

65. G. Moser, S. Serpico, and J. Benediktsson, "Land-cover mapping by Markov modeling of spatial-contextual information in very-high-resolution remote sensing images," *Proceedings of the IEEE*, vol. 101, pp. 631–651, March 2013. (cited on pages 32, 97, and 117)

66. J. A. Hartigan, *Clustering Algorithms*. New York, NY, USA: John Wiley & Sons, Inc., 99th ed., 1975. (cited on page 33)

67. V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995. (cited on page 33)

68. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`. (cited on pages 35 and 124)

69. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. (cited on page 35)

70. M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005. (cited on page 35)

71. P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006. (cited on pages 35 and 124)

72. V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, 2012. (cited on page 35)

73. G.-B. Huang, Q.-y. Zhu, C.-k. Siew, G.-b. H. Ã, Q.-y. Zhu, C.-k. Siew, G.-B. Huang, Q.-y. Zhu, and C.-k. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006. (cited on page 36)

74. G. B. Huang, "What are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt's Dream and John von Neumann's Puzzle," *Cognitive Computation*, vol. 7, no. 3, pp. 263–278, 2015. (cited on page 36)

75. A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013. (cited on page 38)

76. L. Itti, C. Koch, and E. Niebur, "A Model of Saliency Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998. (cited on pages 38, 39, and 42)

77. T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011. (cited on pages 38 and 43)

78. T. A. Wilson, "Perceptual-based image fusion for hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, 1997. (cited on page 38)

79. S. L. Moan, A. Mansouri, J. Hardeberg, and Y. Voisin, "Saliency in spectral images," in *Proceedings of the 17th Scandinavian conference on Image analysis*, pp. 114–123, 2011. (cited on pages 38, 40, 49, and 50)

80. C. V. Dinh, R. Leitner, P. Paclik, M. Loog, and R. P. W. Duin, "SEDMI: Saliency based edge detection in multispectral images," *Image and Vision Computing*, vol. 29, no. 8, pp. 546–556, 2011. (cited on page 38)

81. A. Garcia-Diaz, V. Leborán, X. R. Fdez-Vidal, and X. M. Pardo, "On the relationship between optical variability, visual saliency, and eye fixations: A computational approach," *Journal of Vision*, vol. 12, no. 6, p. 17, 2012. (cited on page 38)

82. D. H. Foster, S. M. C. Nascimento, and K. Amano, "Information limits on neural identification of colored surfaces in natural scenes.," *Visual neuroscience*, vol. 21, pp. 331–336, 2004. (cited on pages 40 and 47)

83. M. Anderson, R. Motta, and S. C. M. Stokes, "A Standard Default Color Space for the Internet - sRGB," *Color and Imaging Conference*, vol. 8, pp. 238–245, 1996. (cited on page 41)

84. J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012. (cited on pages 43 and 44)

85. M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, vol. 3753, no. July, pp. 266–275, 1999. (cited on page 44)

86. J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005. (cited on page 44)

87. L. Tong, J. Zhou, X. Bai, and Y. Gao, "Dual graph regularized NMF for hyperspectral unmixing," in *2014 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2014*, 2015. (cited on page 44)

88. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. (cited on page 47)

89. S. M. C. Nascimento, F. P. Ferreira, and D. H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *Journal of the Optical Society of America A*, vol. 19, no. 8, p. 1484, 2002. (cited on page 47)

90. A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in

Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 193–200, 2011. (cited on page 47)

91. W. Zhao, R. Chellappa, P. Phillips, and a. Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003. (cited on page 55)

92. X. Zou, J. Kittler, and K. Messer, "Illumination Invariant Face Recognition: A Survey," in *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pp. 1–8, 2007. (cited on page 55)

93. X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009. (cited on page 55)

94. A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," *Audio- and Video-Based Biometrie Person Authentication*, vol. 2688, no. 2688, pp. 62–70, 2003. (cited on page 55)

95. K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006. (cited on page 55)

96. B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna, "Using Kinect for face recognition under varying poses, expressions, illumination and disguise," in *Proceedings of IEEE Workshop on Applications of Computer Vision*, pp. 186–192, 2013. (cited on page 55)

97. S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition - A review," 2005. (cited on page 55)

98. S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007. (cited on page 55)

99. Y. Zheng and A. Elmaghraby, "A brief survey on multispectral face recognition and multimodal score fusion," in *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2011*, pp. 543–550, 2011. (cited on page 55)

100. T. Igarashi, K. Nishino, and S. K. Nayar, "The Appearance of Human Skin: A Survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 1, pp. 1–95, 2007. (cited on page 55)

101. Z. P. Z. Pan, G. E. Healey, M. Prascad, and B. J. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 12, pp. 1552–1560, 2003. (cited on pages 56, 57, 65, 66, 67, and 69)

102. W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 40, no. 6, pp. 1354–1361, 2010. (cited on pages 56, 57, 65, 66, 67, and 69)

103. L. Shen and S. Zheng, "Hyperspectral face recognition using 3D gabor wavelets," in *Proceedings of the International Conference on Pattern Recognition*, pp. 1574–1577, 2012. (cited on pages 56, 57, 64, 66, 67, and 69)

104. M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition using 3D-DCT and partial least squares," in *Proceedings of the British Machine Vision Conference*, pp. 57.1–57.10, 2013. (cited on pages 56, 57, 65, 66, 67, 68, and 69)

105. M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiospectral information fusion and PLS regression," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1127–1137, 2015. (cited on page 56)

106. L. J. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," Tech. Rep. CMU-RI-TR-02-25, Robotics Institute, 2002. (cited on pages 56, 65, and 68)

107. Z. Pan, G. Healey, and B. Tromberg, "Comparison of spectral-only and spectral/spatial face recognition for personal identity verification," *EURASIP Journal on Advances in Signal Processing*, pp. 8:1–8:6, 2009. (cited on pages 57, 66, 67, and 69)

108. G. Zhao and M. Pietik??inen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007. (cited on pages 58, 66, 67, and 69)

109. C. Lee and D. a. Landgrebe, "Feature Extraction Based on Decision Boundaries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 388–400, 1993. (cited on page 73)

110. D. A. Kuo Bor-Chen; Landgrebe, "A Robust Classification Procedure Based on Mixture Classifier and Nonparametric Weighted Feature Extraction," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, no. 11, pp. 2486–2494, 2002. (cited on page 73)

111. W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681–3693, 2015. (cited on page 74)

112. P. Soille, *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc., jan 2003. (cited on pages 75, 77, and 86)

113. E. Aptoula and S. Lefèvre, "A comparative study on multivariate mathematical morphology," *Pattern Recognition*, vol. 40, pp. 2914–2929, nov 2007. (cited on pages 75 and 82)

114. P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A Survey on Spectral-Spatial Classification Techniques Based on Attribute Profiles," *IEEE Transactions*

on Geoscience and Remote Sensing, vol. 53, pp. 2335–2353, may 2015. (cited on page 76)

115. D. Landgrebe and L. Biehl, "MultiSpec: A freeware multispectral image data analysis system." `https://engineering.purdue.edu/~biehl/MultiSpec/`. (cited on page 89)

116. D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, pp. 823–870, Jan. 2007. (cited on page 97)

117. M. A. Friedl, C. Woodcock, S. Gopal, D. Muchoney, A. H. Strahler, and C. Barker-Schaaf, "A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data," *International Journal of Remote Sensing*, vol. 21, pp. 1073–1077, jan 2000. (cited on pages 97 and 100)

118. S. V. Stehman and R. L. Czaplewski, "Design and analysis for thematic map accuracy assessment: Fundamental principles," *Remote Sensing of Environment*, vol. 64, no. 3, pp. 331 – 34, 1998. (cited on page 98)

119. G. Foody, "Status of land cover classification accuracy assessment," *Remote Sensing of Environment*, vol. 80, no. 1, pp. 185–201, 2002. (cited on page 98)

120. Z. Zhen, L. J. Quackenbush, S. V. Stehman, and L. Zhang, "Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification," *International Journal of Remote Sensing*, vol. 34, no. 19, pp. 6914–6930, 2013. (cited on page 100)

121. "Hyperspectral remote sensing scenes." `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`. Accessed: 2015-01-28. (cited on page 104)

122. M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. MIT Press, 2012. (cited on pages 113 and 115)

123. P. P. L. Bartlett and S. Mendelson, "Rademacher and Gaussian Complexities: Risk Bounds and Structural Results," *Journal of Machine Learning Research*, vol. 3, no. 3, pp. 463–482, 2002. (cited on page 114)

124. D. J. McDonald, C. R. Shalizi, and M. Schervish, "Estimating beta-mixing coefficients.," *JMLR workshop and conference proceedings*, vol. 15, pp. 516–524, 2011. (cited on page 115)

125. B. Yu, "Rates of Convergence for Empirical Processes of Stationary Mixing Sequences," *Annals of Probability*, vol. 22, no. 1, pp. 94–116, 2016. (cited on pages 115 and 116)

126. M. Vidyasagar, *Learning and Generalisation: With Applications to Neural Networks*. Springer Science & Business Media, 2013. (cited on page 116)

127. M. Mohri and A. Rostamizadeh, "Rademacher Complexity Bounds for Non-I.I.D. Processes," in *Advances in Neural Information Processing Systems 21*, pp. 1097–1104, 2008. (cited on page 116)

128. R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994. (cited on page 121)

129. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, pp. 10–18, Nov. 2009. (cited on page 124)