8-2008

Integration of Spatial and Spectral Information for Hyperspectral Image Classification

Zheng Du
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Zheng Du entitled "Integration of Spatial and Spectral Information for Hyperspectral Image Classification." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Engineering.

Seong G. Kong, Myong K. Jeong, Major Professor

We have read this dissertation and recommend its acceptance:

Hairong Qi, Mongi A. Abidi

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Zheng Du entitled "Integration of Spatial and Spectral Information for Hyperspectral Image Classification." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Engineering.

Seong G. Kong
Major Professor

Myong K. Jeong
Co-Advisor

We have read this dissertation
and recommend its acceptance:

Hairong Qi

Mongi A. Abidi

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the
Graduate School

(Original signatures are on file with official student records.)

Integration of Spatial and Spectral Information for Hyperspectral Image Classification

A Dissertation
Presented for the
Doctor of Philosophy
Degree

The University of Tennessee, Knoxville

Zheng Du
August 2008

Acknowledgement

I would like express my sincerest appreciation and gratitude to my advisor Dr. Seong G. Kong and Dr. Myong K. Jeong for their consistent supports, stimulating suggestion, and thoughtful directions while this work was being completed. Their guidance and constant encouragement have been invaluable.

I would like to thank my committee members, Dr. Hairong Qi and Dr. Mongi A. Abidi who provided additional insight and took the time to read my dissertation. You all had something valuable to offer.

I would also like to acknowledge many of my colleges and friends who sat with me over the years to discuss such matters related to this body of work. They all helped, encouraged, mentored and supported me in numerous ways.

Last and certainly not least, I would like to thank my family for their unconditional supports. None of this would have been possible without the love and support of them. I am forever grateful.

Abstract

Hyperspectral imaging has become a powerful tool in biomedical and agriculture fields in the recent years and the interest amongst researchers has increased immensely. Hyperspectral imaging combines conventional imaging and spectroscopy to acquire both spatial and spectral information from an object. Consequently, a hyperspectral image data contains not only spectral information of objects, but also the spatial arrangement of objects. Information captured in neighboring locations may provide useful supplementary knowledge for analysis. Therefore, this dissertation investigates the integration of information from both the spectral and spatial domains to enhance hyperspectral image classification performance.

The major impediment to the combined spatial and spectral approach is that most spatial methods were only developed for single image band. Based on the traditional single-image based local Geary measure, this dissertation successfully proposes a Multidimensional Local Spatial Autocorrelation (MLSA) for hyperspectral image data. Based on the proposed spatial measure, this research work develops a collaborative band selection strategy that combines both the spectral separability measure (divergence) and spatial homogeneity measure (MLSA) for hyperspectral band selection task. In order to calculate the divergence more efficiently, a set of recursive equations for the calculation of divergence with an additional band is derived to overcome the computational restrictions.

Moreover, this dissertation proposes a collaborative classification method which integrates the spectral distance and spatial autocorrelation during the decision-making process. Therefore, this method fully utilizes the spatial-spectral relationships inherent in the data, and thus improves the classification performance.

In addition, the usefulness of the proposed band selection and classification method is evaluated with four case studies. The case studies include detection and identification of tumor on poultry carcasses, fecal on apple surface, cancer on mouse skin and crop in agricultural filed using hyperspectral imagery. Through the case studies, the performances of the proposed methods are assessed. It clearly shows the necessity and efficiency of integrating spatial information for hyperspectral image processing.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation

Hyperspectral imaging has been one of the main focuses in the imaging field over last 20 years. The development of hyperspectral imaging originated from the need 30 years ago to develop techniques for remote sensing of the Earth from space [1]. The idea was to study the Earth utilizing details not readily visible to humans [3]. These details are manifested in spectral regions beyond the narrow wavelength region accessible to the human eye, from the near infrared to the microwave and into the radiometric wavelength [2][4].

Hyperspectral imaging combines conventional imaging and spectroscopy to acquire both spatial and spectral information from an object. This type of imaging produces a three-dimensional image cube [5], [6] with two spatial dimensions (horizontal and vertical) and one spectral dimension. The spectral dimension contains spectral (or wavelength) information for each pixel on the hyperspectral image cube, and the spatial dimension records the spatial information for each pixel. Because of these combined features of imaging and spectroscopy, hyperspectral imaging can enhance and expand our capability to identify the objects present in the observed scene as well as their spatial distributions.

Hyperspectral imaging has been widely used in a number of areas, including environmental monitoring [7], land cover [8], agricultural analysis [9], and military applications such as ground target detection [10]-[14], and face recognition [15]. As hyperspectral cameras have become more and more accessible, hyperspectral imaging has been applied to biomedical and agricultural uses, as well. Since hyperspectral imaging offers a non-contact optical sensing technique for obtaining both spectral and

spatial information, it has become as effective tool in biomedical applications such as various cancer diagnostics [16]-[19] [56]-[58]and in agricultural application such as food safety inspection [17][39][42]-[51].

The goal of hyperspectral imaging in biomedical applications is to be able to detect and diagnose any form of diseases from its spectral signature. There are many different tools that can help doctors in their diagnoses, but in many cases the disease has already spread, and treatment may already have been compromised. Developing hyperspectral imaging as a medical imaging system that could diagnose pathology at an early stage could be a major improvement in the biomedical field, thus saving many lives.

In agriculture, the technique of hyperspectral imaging can help in the design of a non-destructive, automated safety inspection machine. An automated inspection system would liberate humans from the traditional hand manipulation of agricultural products, reducing energy, labor, and materials costs and improving the quality of product.

For hyperspectral imaging analysis in biomedical and agricultural applications, the main issues are band selection or dimensionality reduction and classification. Both of these applications require that the object can be processed in a timely manner with little or no time lag. This is referred as “real-time” processing. Band selection algorithms can reduce the data dimensionality without loss of critical information, making real-time processing possible. “Classification” is usually a name given to the process of grouping a large number of pixels into multiple classes. Classification of a hyperspectral image means to identify each pixel into multiple classes in the scene. Spectral classification methods produce satisfying results in many cases, but a serious limitation of such methods is that they assign pixels to classes only on their spectral similarity, without any consideration given to the spatial locations of the pixels. When the objects in a hyperspectral image do not have unique spectral signatures, the classification results they generated often display noisy or unrealistic features, such as isolated pixels assigned to a particular class. In this situation, additional information is required to distinguish them.

Hyperspectral data are collected by hyperspectral imaging spectrometers. The spectrometers regularize continuous scenes into a grid of equally sized and regularly spaced data in the form of pixels. Consequently, a hyperspectral image data not only contains information in the spectral domain, but also in the spatial domain in terms of the arrangement of objects in the scene. Information captured in neighboring locations may provide useful supplementary knowledge for analysis. Therefore, this dissertation investigates the integration of information from both the spectral and spatial domains to enhance hyperspectral image classification performance.

1.2 Problem Statement

In hyperspectral imagery, the electromagnetic spectrum is sampled at dozens, hundreds or even thousands of wavelength ranges in the visible and near infrared spectra. The result is a very detailed view of the spectral signature of the scene represented by a particular pixel. The additional information comes at a cost, however. The more features used for prediction, the more noise, redundancy, and model complexity can degrade accuracy. Practical considerations such as computation time, storage, and communication bandwidth must also be considered. These problems can be resolved by dimensionality reduction methods, which seek to remove redundant information and to keep only the information relevant to the applications. Among widely used dimensionality reduction methods, principal component analysis (PCA) [23] rearranges the data in terms of the significance measured by the eigenvalues of the data covariance matrix. While PCA is a powerful dimensionality reduction technique, it may not be optimal in terms of classification [24]. Other dimensionality reduction techniques in hyperspectral image processing include feature extraction and band selection. Feature extraction generates a reduced number of features by transforming the original data into a lower dimension with most information content preserved [25][26]. Band selection finds an optimal subset of spectral bands for dimensionality reduction in hyperspectral image processing without any loss of critical information. The transformed features generated by the feature

extraction methods can usually provide better discriminating power than the optimal bands selected by the band selection method. But feature extraction is not suitable for the applications discussed in this dissertation since it reduces the time and cost for the measurement equipment. Band selection is preferable in our applications because it can decrease the system cost and increase the processing speed.

Both supervised and unsupervised techniques have been developed for classifying hyperspectral images [20]. Although most classification techniques are suitable for classifying hyperspectral images by spectral information, their application to classification imaging is limited by the fact that they rely on the spectral properties of the data only, thus neglecting the spatial information related to the spatial arrangement of the pixels in the scene. The resulting classification is carried out without incorporating information on the spatially adjacent data, i.e., the data are managed not as an image but as a disarranged listing of spectral measurements, where the spatial coordinates can be randomly shuffled without affecting the analysis. However, one of the distinguishing properties of hyperspectral data, as collected by available imaging spectrometers, is the multivariate information coupled with a two-dimensional (2-D) pictorial representation amenable to image interpretation [21][22]. Consequently, a hyperspectral image data not only contain information regarding the spectral content, but also the spatial arrangement of objects in the scene under observation. An efficient classification procedure should consider both the spectral and spatial information when it assigns a pixel to a particular class.

1.3 Objective

The main objective of this research is to investigate the integration of spatial and spectral information for hyperspectral image classification. The major impediment to the combined spatial and spectral approach is that most spatial methods were only developed for single image band. By carefully investigating the existing spatial autocorrelation

measure, this dissertation successfully proposes a Multidimensional Local Spatial Autocorrelation (MLSA) for hyperspectral image data. This new measure is a fundamental improvement in the description of spatial autocorrelation for hyperspectral imaging and offers an excellent opportunity for combining spatial and spectral information in hyperspectral data analysis.

An important task in hyperspectral data analysis is to reduce the redundancy of the spectral and spatial information without losing any valuable details needed for the subsequent detection, discrimination, and classification processes. First, the use of divergence for band selection is mathematically characterized. A set of recursive equations for the calculation of divergence with an additional band is derived to overcome the computational restrictions. Then based on the proposed Multidimensional Local Spatial Autocorrelation (MLSA) measure, this research work develops a collaborative band selection strategy that combines both the spectral separability measure (divergence) and spatial homogeneity measure (MLSA) for hyperspectral band selection task. Moreover, this dissertation proposes a collaborative classification method to integrate the spectral and spatial information simultaneously for the classification process. The collaborative classification method consists of a spectral similarity term to measure the similarity of a given sample to a particular class and a spatial similarity term to measure how similar a pixel to its neighboring pixels. Therefore, this method fully utilizes the spatial-spectral relationships inherent in the data, and thus improves the classification performance for hyperspectral image data. Figure 1.1 illustrates the schematic diagram of proposed method.

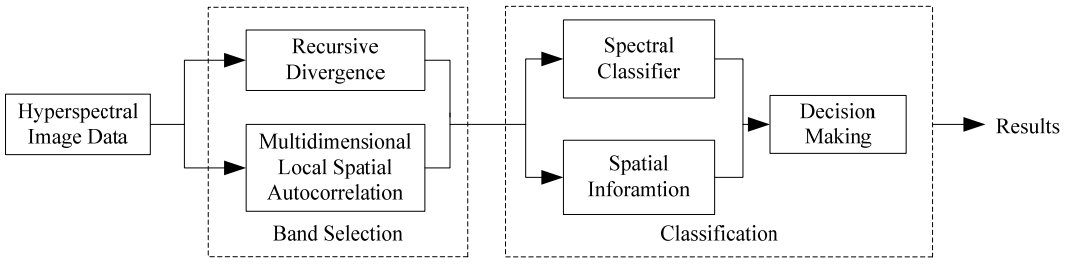


Figure 1.1: Schematic diagram of proposed method

1.4 Contributions

The primary goal of this study is to investigate how to efficiently integrate spatial and spectral information in hyperspectral image analysis. To this end, contributions include:

1. Proposing the Multidimensional Local Spatial Autocorrelation (MLSA) measure for hyperspectral image and deriving their statistic moment terms.
2. Developing a set of recursive equations for the calculation of divergence with an additional band to overcome the computational restrictions in band selection.
3. Proposing a collaborative band selection strategy that combines both the spectral separability measure (divergence) and spatial homogeneity measure (MLSA) for hyperspectral band selection.
4. Developing a collaborative classification method that considers both the spectral similarity and spatial homogeneity during the decision-making.
5. Evaluating the proposed band selection and classification method in four case studies, including detection and identification of tumor on poultry carcasses, fecal on apple surface, cancer on mouse skin and crop in agricultural filed using hyperspectral imagery.
6. Developing a MATLAB Toolbox for hyperspectral image analysis.

One of the reasons for the paucity of combined spatial and spectral approaches is that most spatial methods have only been developed for the single image band. The extension of classic spatial methods to hyperspectral image data is not straightforward. When such techniques are applied independently to each spectral image, there is a possibility for loss or corruption of information of the image due to the probability that new spectral constituents (not present in the original image) may be created as a result of processing the spectral image separately. This dissertation successfully ***proposes the Multidimensional Local Spatial Autocorrelation (MLSA) measure for hyperspectral image and derives its statistic moment terms.*** This new measure is a fundamental improvement in the description of spatial autocorrelation for hyperspectral imaging and

provides a great opportunity to combine spatial and spectral information in hyperspectral data analysis (Chapter 4.4.2).

The huge amount of hyperspectral image data often makes real-time computer processing a challenging task. This dissertation suggests a band selection method for hyperspectral images based on ***recursive divergence***. This method avoids transforming the original hyperspectral images to the feature space. Instead, it maximizes the class separability by considering the correlation information of spectral bands. This research ***mathematically characterizes the use of divergence for band selection***. Also, ***a set of recursive equations for the calculation of divergence with an additional band is derived*** to ***overcome the computational restrictions*** in real-time processing (Chapter 4.3).

In the spectral information-based band selection, a good subset will maximize the representation of the spectral separability. However, separability maximization does not guarantee a classification process that will produce either the best or the most accurate visual result. It is only when the spectral separability is spatially organized that regional variations become apparent in an image. This suggests that using only the spectral separability criterion cannot guarantee the most accurate results. This dissertation develops a ***collaborative band selection strategy that combines both spectral separability measure (divergence) and spatial homogeneity measure (MLSA) for hyperspectral band selection*** (Chapter 4.4.3).

Most techniques used for the analyzing hyperspectral images involve separate processing for extracting spatial and spectral information, and are not able to do extract both simultaneously. The spatial and spectral processing is done separately and their results combined. Thus, within each spectral and spatial domain of the processing, the individual processes are unable to make use of the information in the other domain, until the processing in each separate domain is complete, at which point their outputs can be combined. By using the proposed ***a collaborative classification method***, ***the spectral and spatial information of image can be combined simultaneously***. This method fully

utilizes the spatial-spectral relationships inherent in the data, and thus improves performance in image-analysis tasks (Chapter 5.3).

For biomedical and agricultural applications, hyperspectral imaging offers an instant, non-invasive diagnostic procedure based on the analysis of the spectral properties of the tissue. This dissertation evaluates the proposed band selection and classification method with ***four case studies***: ***poultry tumor detection, apple contamination detection, skin cancer detection and agricultural crop classification***. The performances of the proposed methods clearly show the necessity and efficiency of integrating spatial information for hyperspectral image processing. (Chapter 3, Chapter 4.5 and Chapter 5.4).

To help researchers using hyperspectral imagery, a MATLAB toolbox for hyperspectral image analysis is necessary. The Hyperspectral Image Toolbox ***provides rich visualization tools for display the 3D hyperspectral image***. In addition, it incorporates with ***both standard algorithms for hyperspectral image analysis and also this original work in hyperspectral band selection and classification***. (Appendix A)

2 Background

2.1 Introduction

In this chapter, we provide relative background for the research conducted by this dissertation. The chapter is organized as follows. Section 2.2 summarizes the history of hyperspectral imaging and several basic concepts for hyperspectral imaging. Section 2.3 describes the key characteristic of hyperspectral imaging and gives examples of spectral signatures. In section 2.4, two spectral classification methods, i.e. support vector machine (SVM) and Maximum likelihood classifier (MLC) are presented. These two classifiers will be used in experiments. Section 2.5 reviews the applications of hyperspectral imaging in biomedical and agriculture fields.

2.2 Hyperspectral Imaging

The development of hyperspectral imaging came from remote sensing of the Earth that started 30 years ago [1]-[4]. The idea was to study the Earth utilizing details not readily visible to humans. These details are manifested in spectrum regions beyond the narrow wavelength region accessible by eye, from the near infrared to the microwave, and in the radiometric wavelengths. With the launch of Landsat 1 in July 1975, the first multispectral remote sensing of the earth took place by NASA in black and white with no more than 10 spectral bands. Multispectral imaging allows the separation of different land cover types into thematic classes, based on differences in spectral reflectance or spectral emittance in the thermal region of the spectrum. The improvements of computer power has allowed for the creation of sensors with a higher number of spectral bands. The so-called hyperspectral sensors were used by the Airborne Imaging System (AIS 1, AIS 2) which were designed and built in the early 80s as part of a NASA Jet Propulsion

Laboratory (JPL) imaging spectrometry program. These new sensors could separate the signal into 128 contiguous bands in the spectral region from 1200 to 2400 nm. The next major step in the NASA program and in imaging spectrometry in general was the development of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Operational in 1987 and in current use, AVIRIS was designed to image 224 contiguous bands in the region from 400-2500 nm. Now, we are moving towards developing sensors with even better spatial resolution.

Hyperspectral imaging is an optical technique, which obtains the spectral information from an object to display it as an image. With conventional spectroscopy, we can measure the intensity of radiation at a certain wavelength at a single spot. We will then have the spectral signature of the object in a particular wavelength. The disadvantage of such a method is that we can only record the spectrum at a specific spot. Acquiring the spectral information over large areas is known as imaging spectroscopy. In that case we will record both spatial and spectral signature of the object of interest. HSI uses spectroscopy imaging in which the waveband is divided into a multitude of continuous bands.

There are various kinds of sensors in hyperspectral imaging, and each one of them analyzing a different feature of the material. Depending on the sensor, the information needed is extracted using the reflectance spectrum, the emission spectrum or the fluorescence spectrum. All of these spectrometric measurement techniques can be found in hyperspectral imaging in many to biomedical applications. In the next part, the fluorescence spectroscopy will be briefly introduced because of its widely use in the related applications.

2.3 Principle of Hyperspectral Imaging

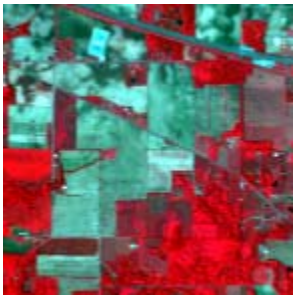
The key characteristic of hyperspectral imaging is the high spectral resolution that is provided over a large and continuous wavelength region. Spectral resolution is a measure

of the narrowest spectral feature that can be resolved by sensor. Spectral resolution might become the key parameter in identifying different materials. It was found that some minerals might have similar features at low-resolution spectrum. However, at high resolution, their spectrum might be quite different. Hyperspectral imaging differs from multispectral imaging and spectroscopic imaging by the number of bands in the spectral dimension or spectral resolution. For HSI, we are talking about over a hundred continuous narrow bands that extend from the visible (0.4μm to 0.7μm) region through the near-infrared (0.7μm to 2.5μm). Other sensors exploit the emissive properties of the materials in the mid wave and long wave infrared. The high resolution and continuity of such sensors can reveal important information and more properties for further classification.

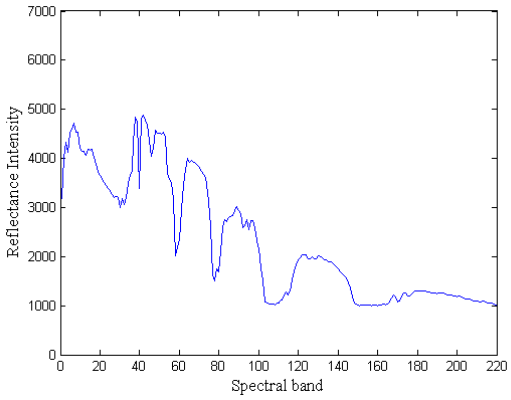
When we plot the response characteristics of a certain material type against wavelength, we define what is termed the spectral signature of that material. In hyperspectral image, the spatially co-registered pixels can be combined into a vector representing the spectral signature of materials. In principle, the spectral signature can be used to uniquely characterize and identify any given material over a sufficiently broad spectral band.

Figure 2.1 provides examples of spectral signatures for corn, grass, wheat and stone. This is a hyperspectral data set taken over an agricultural portion of NW Indiana in the early growing season of 1992. Data are delivered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), which features 224 spectral channels spaced about 10 nm apart in the spectral region from 0.4 to 2.45 μm at a spatial resolution of 20 m.

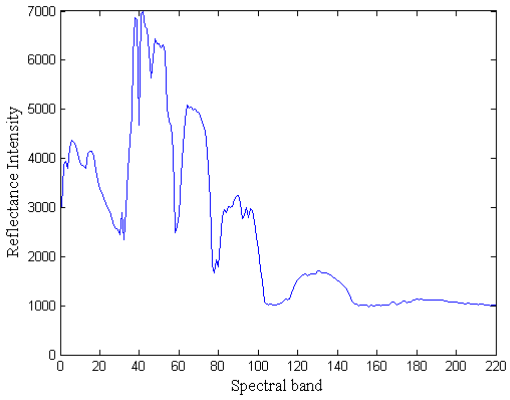
Data produced by hyperspectral imaging systems are essentially a three dimensional cube of data $\boldsymbol{H}(m,n,\lambda_k)$ (shown in Figure 2.2), where (m,n) (m = 0,1,…,M-1, n = 0,1,…,N-1) denotes the spatial coordinates for a pixel location in the image and λ_k (k = 1,2,…,K) denotes a spectral band (wavelength range). The value stored at $\boldsymbol{H}(m,n,\lambda_k)$ is the spectrometry response from the pixel (m,n) at a certain wavelength corresponding to the spectral bandλ_k. Each pixel picked from the hyperspectral imaging data can be



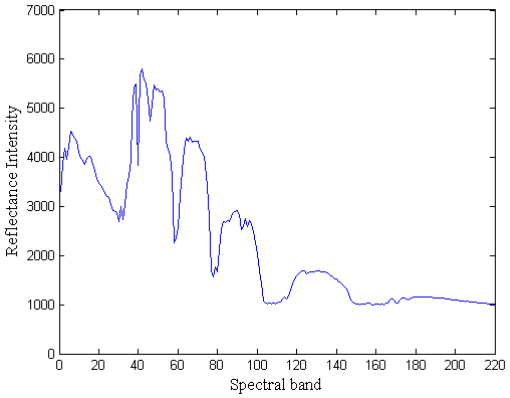
(a) Spectral image of the Indiana pine dataset



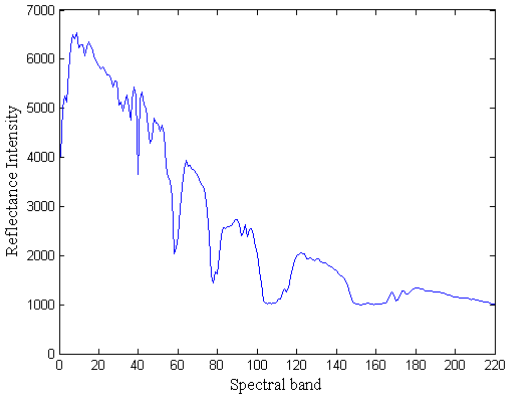
(b) Spectral signature of corn



(c) Spectral signature of grass



(e) Spectral signature of wheat



(f) Spectral signature of stone

Figure 2.1: Examples of spectral signature

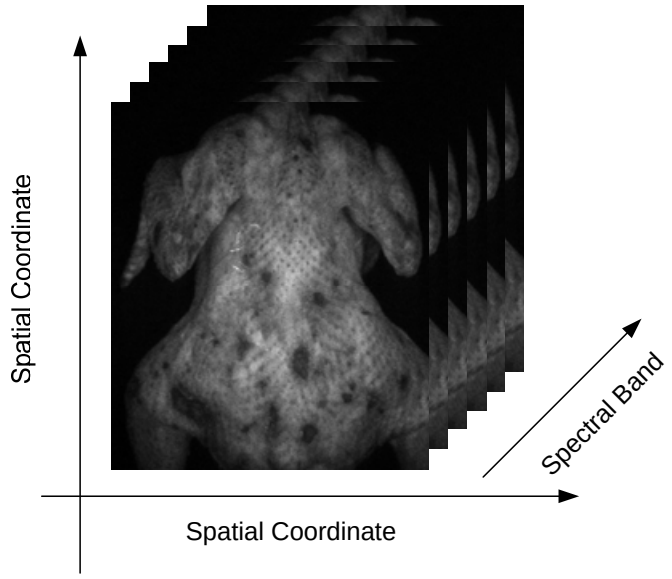


Figure 2.2: Representation of 3D nature of hyperspectral images

represented as a column vector $\boldsymbol{x}_i = (x_{i1}, \cdots x_{ik}, \cdots, x_{iK})^T$, with class label ω_l. For each pixel the spectrometry response levels at the entire spectral band in study are measured, so x_{ij} is the measurement of the normalized intensity value of the *k*-th spectral band for the *i*-th pixel.

2.4 Spectral Signature Classification

A hyperspectral image cube is made up of many, usually hundreds, of images that are spatially co-registered. Each of these images represents the reflected (absorbed, transmitted) energy of the materials within the pixel at different wavelengths and bandwidths. For any given material, the amount of energy that is reflected will vary with wavelength. This important property allows us to separate distinct material types based on their response values for a given wavelength. By comparing the spectral signature of different materials, we may be able to distinguish between them. For example in detection cancer metastases, by analyzing the spectral signature of human skin, we would like to know which cells are cancerous and which one are not. To do this we need to

classify each pixel of the spectral image into those that contains cancerous cells and those that do not contain cancerous cells.

In this part, Support vector machine (SVM) [31]-[33]and Maximum likelihood classifier (MLC) [34]-[36], two popular classifiers for hyperspectral image process, are introduced.

2.4.1 Support Vector Machine Classification

Support Vector Machines is a statistical learning theory introduced recently for regression and classification purpose [29][30]. In this section, the derivation and applications of SVM for high dimensionality data such as hyperspectral data will be discussed.

The idea of SVM is to find the optimal separation surface between two classes through an optimizing procedure that finds the support vectors that form the boundaries of the class. There are many different possibilities of separating the hyperplanes for a given set of data, but among them, there is only one way to find the hyperplane that maximizes the margin separation between the classes. SVM will find the linear classifier for an optimal separating hyperplane. In case there is no linear separation, the Kernel method is used to map the data with a non-linear transformation to a higher dimensional space and in that space it attempts to find a linear separation between classes where smaller number of training set can work.

Consider the case of classifying a set of linearly separating data. Assume a set of training vectors $\boldsymbol{x}_i$ that belong to two classes with the class label $y_i = \{+1, -1\}$ (i = 1, …, n). The data set is called linearly separable by a hyperplane $\boldsymbol{w}^T \boldsymbol{x} + b = 0$ if there exist a vector $\boldsymbol{w}$ and a scalar b such that

$$\begin{aligned} \boldsymbol{w}^T \boldsymbol{x}_i + b \geq +1 \quad & \text{if } y_i = +1 \\ \boldsymbol{w}^T \boldsymbol{x}_i + b \leq -1 \quad & \text{if } y_i = -1 \end{aligned} \tag{2.1}$$

which can be combined into an inequality:

$$y_i \left(\boldsymbol{w}^T \boldsymbol{x}_i + b \right) - 1 \geq 0 \ , \ i = 1, 2, \cdots, n \tag{2.2}$$

The problem reduces to determining the weight vector $\boldsymbol{w}$ and bias b that maximizes the margin of separation $2/\|\boldsymbol{w}\|$. The optimal hyperplane can be determined as the solution of a constrained optimization problem that minimizes the Lagrangian criterion function:

$$J\left(\boldsymbol{w}, b, \alpha\right) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{n} \alpha_i \left[1 - y_i \left(\boldsymbol{w}^T \boldsymbol{x}_i + b \right) \right], \quad \alpha_i \geq 0 \tag{2.3}$$

By differentiating the Lagrangian function with respect to $\boldsymbol{w}$ and b and setting to zeros leads to

$$\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i \tag{2.4}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{2.5}$$

The linearly constrained optimization problem can be translated into a dual problem that maximizes the following criterion function:

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\boldsymbol{x}_i^T \boldsymbol{x}_j \right) \tag{2.6}$$

subject to the constraints

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \ \text{ and } \ \alpha_i \geq 0 \ , \quad i = 1, 2, \cdots, n \tag{2.7}$$

The Lagrange multipliers α_i's can be estimated using quadratic programming methods. The Karush-Kuhn-Tucker complementary conditions for primal optimization problem are

$$\alpha_i \left[1 - y_i \left(\boldsymbol{w}^T \boldsymbol{x}_i + b \right) \right] = 0 \ , \quad i = 1, 2, \cdots, n \tag{2.8}$$

Training samples $\boldsymbol{x}_i$ corresponding to nonzero Lagrange multipliers (α_i) are called support vectors. Support vectors lie on the class boundaries at the distance $1/\|\boldsymbol{w}\|$ from the hyperplane. All remaining samples in the training set but support vectors do not play

a role in finding optimal decision boundaries. The discriminant function corresponding to the optimal hyperplane depends both on the Lagrange multipliers and on the support vectors, i.e.,

$$f(\boldsymbol{x}) = \sum_{i \in SV} \alpha_i y_i \left(\boldsymbol{x}_i^T \boldsymbol{x} \right) + b^* \tag{2.9}$$

where SV denotes the set of support vectors. The bias can be represented by $b^* = 1 - \boldsymbol{w}^T \boldsymbol{x}_{sv}$ for $y_{sv} = +1$. The Lagrange multipliers behave as weights of each training sample according to its importance in determining the discriminant function.

For a non-linearly separable case, the input vectors are mapped to a higher dimensional feature space by a nonlinear function. Then the decision function for a two-class problem derived by the support vector classifier can be written as follows using a kernel function $K(\boldsymbol{x}, \boldsymbol{x}_j)$ of a new pattern $\boldsymbol{x}$ (to be classified) and a training pattern $\boldsymbol{x}_j$:

$$f(\boldsymbol{x}) = \text{sgn}(\sum_{j \in S} y_j \alpha_j K(\boldsymbol{x}, \boldsymbol{x}_j) + b) \tag{2.10}$$

Frequent choices of kernels include polynomial, radial basis, and sigmoid function.

2.4.2 Maximum Likelihood Classification

Maximum likelihood classification using Gaussian density function is often use in hyperspectral image discrimination. The method assumes that each information class ω_i is described by k spectral components, which are independent Gaussian random variables. Under this method, a hyperspectral image pixel is classified as belonging to information class l if [37]

$$b_k(\boldsymbol{x}) = \max\{g_l(\boldsymbol{x})\}, \; l = 1, 2, ..., L \tag{2.11}$$

$$g_l(\boldsymbol{x}) = \ln p(\boldsymbol{x} \mid \omega_l) + \ln p(\omega_l) \tag{2.12}$$

where L is the number of classes and $g_l(\boldsymbol{x})$ is a known discriminant function, based on the assumption that the classes have a multivariate normal density distribution. $p(\boldsymbol{x}|\omega_l)$ is the

likelihood and $p(\omega_l)$ is the posterior probability. The multivariate normal density in d dimension is written [37]:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}\right|^{1/2}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] \tag{2.13}$$

In the general multivariate normal case, we can write the discriminant function as [37]:

$$g_l(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}_l^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_l) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_l\right| + \ln p(\omega_l) \tag{2.14}$$

where $\boldsymbol{\mu}_l$ is the mean of class l and $\boldsymbol{\Sigma}_l$ is covariance matrix of class l.

2.5 Applications of Hyperspectral Imaging

2.5.1 Food Safety Inspection

There are more than 200 known diseases transmitted through food [38]. The food maybe carries various illnesses such as viruses, bacteria, parasites, toxins and metals. It may causes diseases ranging from mild gastroenteritis to life threatening neurological, hepatic, and renal syndromes. In the United States, food borne diseases have been estimated to cause 6 million to 81 million illnesses and up to 9,000 deaths each year.

The issue of food safety is very important, for example, contaminated apple juice came to the forefront in recent years with major outbreaks of Escherichia coli infections in people who drank unpasteurized apple juice or cider [40]. In addition, to monitoring for food contamination, it is necessary to detect which produce is satisfactory to be sold [39]. Nowadays, many food inspections are still done manually, workers are positioned along the conveyors to visually inspect the passing product and remove those with defects. Such inspections can be a labor-intensive job when inspecting large quantities of food [39].

However, ongoing changes in the food supply, the identification of new food borne diseases, and the availability of new surveillance data like hyperspectral imaging could reduce the occurrences of food poisoning. Rapid, noninvasive methods that can be implemented to assess hazardous conditions in food production would be substantially beneficial. Moreover, it will reduce the costs of energy, labor and materials. Hyperspectral imaging techniques may allow online measurements with such specificity [41].

2.5.1.1 Poultry Skin Tumor

The safety of our Nation's food supply is an issue faced by all individuals at every meal. One way to help inspectors in their quest to ensure the quality and safety of the Nation's food supply is to use automated machine vision to spot diseases and unwholesome food before it is shipped to store shelves. We saw previously that hyperspectral imaging system could be used to detect problems on fruits like apples, but it can also be extended to a range of other products like poultry skin. Identification and separation of poultry carcasses contaminated by feces and/or ingesta are very important to protect the consumer from a potential source of food poisoning when pathogens enter the food chain. The development of high speed and reliable inspection systems to ensure safe processing of poultry has become an important issue [47].

Despite the developments in this area, the machine-vision inspection technique with chickens presents more complex problems than apples. Apples are easier in part because they are more uniform in shape and surface texture than chickens. Still, there are uniformity problems, such as color differences from variety to variety and even within a single apple. The reflected light is analyzed by a computer and the differences between light shining on the chicken and light reflected are due to variations in external skin color texture, and chemical contents that are clues to problems.

Chao et al. [48] initiated to introduce hyperspectral images on the detection of chicken skin tumors. They applied principle component analysis (PCA) on the hyperspectral reflectance image to select three useful wavelength bands. Through analyzing the eigenimage, they manually chose three bands which gave good contrast between tumor and good skin regions. These three wavelengths were then used in a real-time 3-band multispectral imaging system, which was used to image 60 tumorous and 20 normal chickens. The ratioed images captured by this multispectral imaging system were then divided into regions of interest (ROI's) classified as either tumorous or normal by a veterinarian. Statistic features, such as skewness, kurtosis and coefficient of variation, for each ROI were extracted for use as inputs to fuzzy classifiers. The fuzzy classifiers were able to separate normal from tumorous skin with increasing accuracies as more features were used. They obtained classification rates of 91% and 86% for normal and tumorous skin tissue regions, respectively (44 of 51 test set tumors were correctly detected).

In [49], Nakariyakul et al studied a feature selection method to detect chicken skin tumors using hyperspectral reflectance image. Two sets of spectral bands were selected to detect the central lesion regions and the outer thickened-skin regions of the tumors from normal skin. The two detection results were fused to reduce false alarms.

In the particular case of skin tumors, Fletcher and Kong [50] discovered that because tumors are different than skin discoloration in shape rather than in color, using a non-invasive hyperspectral imaging by fluorescence was better than using reflectance.
Kim et al [47] also used hyperspectral fluorescence image to detect chicken skin tumors. In their research, they manually chose three features: the maximum response in bands 20-25, the slope of the response in bands 10-20, and the ratio of the maximum response in bands 40-45 over the maximum response in bands 20-25, thus 23 bands in total. These features were used to train a fuzzy classifier. Their research obtained 76% detection rate (31 of 41 skin tumors were detected) in ten HS image sets with 12 false alarm regions in total.

Kong et al [17]also investigated using hyperspectral fluorescence image on detecting chicken tumors. Before applied PCA, the spatial HS data were compressed using a discrete wavelet transform. Then the spectral responses of the data were reduced by PCA. Two were manually selected: the average of the responses in bands 20-24, and the difference between the average response in bands 20-24 and the average in bands 44-48, which are ten bands in total. A fuzzy classifier is trained to classify each pixel to tumor, normal skin, or background. Morphological processing and median filtering were used as postprocessing to reduce false alarms. There were able to achieve a detection rate of 82%.

In our initial work on detecting chicken skin tumors [51], a spectral band selection method for feature dimensionality reduction is proposed. Hyperspectral fluorescence imaging offers an instant, noninvasive inspection method for detecting biomedical abnormalities. However, the huge amount of hyperspectral image data often makes real-time computer processing a challenging task. Our research suggests a band selection method of hyperspectral images based on the recursive divergence for the automatic detection of poultry carcasses. This method avoids transforming the original hyperspectral images to the feature space. Instead, it maximizes the class separability by considering the correlation information of spectral bands. In that paper, we mathematically characterize the use of divergence for band selection. Also, a set of recursive equations for the calculation of divergence with an additional band is derived to overcome the computational restrictions in real-time processing. A support vector machine is used as a classifier for tumor detection. Our proposed recursive divergence approach gives 90.6% detection rate, which is within the industry-accepted accuracy of 90–95%, while achieving the computational saving for real-time processing.

2.5.1.2 Fecal Contamination

Fecal contamination detection on produce has become increasingly important since the Food and Drug Administration (FDA) averred fecal contamination as a major source of

human pathogens. Hyperspectral imaging devices for anomaly detection on produce and meat has been a significant breakthrough in reducing to case of contamination.

The introduction of hyperspectral imaging as made it possible to detect and classify with a high level of accuracy these problems related to food safety. Kim et al. [43][44][45][46] implemented a hyperspectral imaging device based on fluorescence spectroscopy for detection of Escherichia coli contamination from animal feces on apple surface. The hyperspectral imaging system was design to work in the visible and the near infrared regions of the spectrum by using fluorescence spectroscopy. The idea is to use a machine vision system that is much quicker and more accurate than the human eye and without requiring anyone to handle the fruit. The imaging device can detect dirt, fly specks, fungi, rot, and other diseases, all of which can cause fruit to harbor more bacteria, in addition creating obvious quality problems. The result of their experiment proves that hyperspectral fluorescence techniques can be used efficiently to detect fecal contamination on apple surfaces for commercial purposes.

2.5.1.3 Fruit Defect Inspection

The United States packs over 220 million boxes of apples each year. Apples with rot, injury, disease, serious bruising and other defects must be removed before waxing to prevent cross-contamination and reduce subsequent processing cost. Although some aspects of the packing process have been automated, a key step of the apple packing process, the defect inspection process, is still done by hand.
Cheng [39] integrates hyperspectral imaging, real time machine vision, dual-spectral sensing, and pattern recognition techniques for automatic defect inspection in fresh produce. In particular, these technologies were applied for cucumber chilling damage inspection and online apple defect detection.

Lu [42] investigates the potential of near-infrared (NIR) hyperspectral imaging for detecting bruises on apples in the spectral region between 900 nm and 1700 nm. An NIR

hyperspectral imaging system was developed and a computer algorithm was created to detect both new and old bruises on apples. Experiments were conducted to acquire hyperspectral images from Red Delicious and Golden Delicious apples over a period of 47 days after bruising. Results showed that the spectral region between 1000 nm and 1340 nm was most appropriate for bruise detection.

2.5.2 Cancer Detection

Cancer remains a primary cause of human mortality. Treatment, though improving, is held back by late or inadequate detection of cancerous and precancerous tissue. If the detection methodology lacks sufficient precision, surgical removal of cancerous tissue can be incomplete. Hyperspectral, noninvasive imaging examines human tissue with extremely high spectral resolution, detecting phenomena in extremely narrow bands of emission, thereby greatly increasing image resolution. With high resolution and multiple bandwidths, very subtle differences in signature characteristics of tissues can be identified.

2.5.2.1 Skin Cancer

The skin is the body's largest organ and it protects the body against heat, light, injury, and infection. Healthy skin cells grow, divide, and replace themselves in an orderly controlled way. Sometimes, however, normal cells lose the ability to divide and grow normally, and grow out of control to form tumors. Tumors can be benign or malignant. Benign tumors are not cancer, while malignant tumors are cancer. The three types of skin cancer are basal cell carcinoma, squamous cell carcinoma and malignant melanoma. Skin cancer is one of the most common cancers in the United Sates and its incidence is increasing dramatically. According to the statistics of the American Cancer Society, in the US, more than 1 million Americans will be diagnosed with non-melanoma skin cancer every year, and 59,580 persons will be diagnosed with melanoma in 2005 [52]. The melanoma

mortality rates increased from 2.0 per 100,000 in 1969 to 3.0 in 1999 [53]. Yet the skin cancer would be almost 100 percent curable if it were detected early and treated. Unfortunately there are very few options for the medical clinician to diagnosis skin cancer. The most definitive test has been a biopsy. A biopsy is the removal of a sample of tissue from the body for examination. The tissue will be examined under a microscope to assist in diagnosis. The biopsy is an invasive diagnostic technique, which requires both trained professionals and significant waiting time. Because of the accessibility of the skin, it is possible to obtain the optical properties of the tumor tissues [54].

Our initial work [18][19][55] present hyperspectral fluorescence imaging and a support vector machine for detecting skin tumors. A hyperspectral image contains spatial information measured at a sequence of individual wavelength across a sufficiently broad spectral band at high-resolution spectrum. Fluorescence is a phenomenon where light is absorbed at a given wavelength and then is normally followed by the emission of light at a longer wavelength. Fluorescence generated by the skin tissue is collected and analyzed to determine whether cancer exists. Oak Ridge National Laboratory developed an endoscopic hyperspectral imaging system capable of fluorescence imaging for skin cancer detection. That hyperspectral imaging system captures hyperspectral images of 21 spectral bands of wavelength ranging from 440 nm to 640 nm. Experiments show that the SVM classification with spatial filtering achieves high skin tumor detection accuracies.

2.5.2.2 Cervical Cancer

Causal factors for cervical cancer (known also as cervical intraepithelial neoplasia [CIN]) may include carcinogens, multiple cell mutations, viruses, and multiple causal factors. The human papillomavirus (HPV) is an identified risk factor for CIN. Optic technology that is based upon biochemistry and structure is ideal for diagnosing CIN. With sufficient spatial resolution, spectral imaging can reveal the increased vascularity in subsurface vessels of the cervix that contain hemoglobin often found in cases of cervical cancer. [56] provided evidence that a hyperspectral system detected cervical cancer precursors at a

rate greater than that obtained by a simultaneously collected Pap smear. Although this study used a small test sample that scanned only the ectocervix, rather than the endocervix as well, the researchers showed that spectral sensing has the potential to discern grade of disease.

2.5.2.3 Colon Cancer

The colon is the upper part of the large intestine tube while the rectum is the lower part of this tube. Practically, colon or rectum cancer is characterized as separate cancer instances. Colorectal or bowel cancer is a composite name for colon and rectum cancer. It is the uncontrolled growth of tissue cells in either the colon or rectum which causes the colorectal cancer. According to a recent publication1, over 34,000 new cases of colorectal cancer are diagnosed each year. Yet 80% of colorectal cancer cases can be treated if caught at an early stage. New improved screening and diagnosis methods could potentially save thousands more lives each year.

Rajpoot [57][58][16] studied the classifier performance for a hyperspectral colon tissue cell classification system. It was shown that considerably high classification accuracy could be achieved for their tissue cell classification system by selecting optimal set of parameters for the Gaussian kernel.

2.5.3 Precision Farming

Hyperspectral imaging will have a particular enabling impact on precision agriculture [61]. Precision agriculture is the technique of managing each part of a field differently and in the most effective way. Information on the performance and potential production of each part of the field is collected, monitored, and analyzed so that informed management decisions can be made [59]. The result is potentially increased yields with less input and reduced impact on the environment. Goel et al. [60] reported that there is a potential of using hyperspectral airborne remote sensing in the visible and near-infrared

regions for the detection of weed infestation in corn fields. In [9], Yang, et al. examined airborne hyperspectral imagery for mapping grain sorghum yield variability as compared with yield monitor data. Hyperspectral images were acquired using a CCD camera-based hyperspectral imaging system from two grain sorghum fields during the 2000 growing season, and yield data were also collected from the fields using a yield monitor. Results from this study demonstrate that airborne hyperspectral imagery can be a useful data source for mapping crop yield variability.

3 Hyperspectral Image Datasets for Case Study

In this dissertation, four case studies using hyperspectral imagery are investigated, namely detection and identification of tumor on poultry carcasses, fecal on apple surface, cancer on mouse skin and crops in an agricultural filed. All data in these case studies are real-world data. Especially the first three datasets are exclusive, and the last one is a public data set. In this chapter, a brief description of the three hyperspectral datasets used is given. Section 3.1 introduces the hyperspectral chicken data for poultry tumor detection. Section 3.2 introduces the dataset used for detecting fecal on apple surface. The dataset used for detecting cancer on mouse skin is described in Section 3.3. Last, the famous Indian pine dataset is briefly described in Section 3.4.

3.1 Hyperspectral Imaging for Poultry Tumor Detection

3.1.1 Background

Machine vision systems have been widely used for inspection and quality control in automated production processes. Poultry carcasses with pathological problems must be identified and removed from food processing lines to meet the requirement of high standards of food safety. Traditionally, trained human inspectors carry out the inspection and examine a small number of representative samples from a large production run. Manual inspection and classification of agricultural products can be a highly repetitive and tedious task. Human inspectors are often required to examine 30-35 poultry samples per minute. Such working conditions can lead to repetitive motion injuries, distracted attention and fatigue problems, and result in inconsistent quality. Rapid, non-invasive machine vision inspection methods for assessing hazardous conditions in food production would provide a substantial benefit in the quest to ensure high quality of poultry inspection.

Poultry skin tumors are ulcerous lesions that are surrounded by a rim of thickened skin and dermis [122]. Skin cancer causes skin cells lose the ability to divide and grow normally, and induce abnormal cells to grow out of control to form tumors. Tumorous carcasses often demonstrate swollen or enlarged tissue caused by the uncontrolled growth of new tissue. Tumor is not as visually obvious as other pathological diseases such as septicemia, air sacculitis, and bruise since its spatial signature appears as shape distortion rather than a discoloration. Therefore, conventional vision-based inspection systems operating in the visual spectrum may reveal limitations in detecting skin tumors on poultry carcasses.

Hyperspectral fluorescence imaging offers an instant, non-invasive inspection method for detecting biomedical abnormalities such as defects on poultry carcasses [48][54]. Hyperspectral image data contain spatial information measured at a sequence of individual wavelength across broad spectral bands. Hyperspectral images show a detailed view of the spectral signature of the scene. The spectral signatures are useful for identifying various material compositions due to their unique spectral characteristics at particular wavelengths [5]. Fluorescence techniques are generally regarded as sensitive optical tools, and have proven to be effective in a number of scientific areas [123]. Fluorescence is a phenomenon where light is absorbed at a given wavelength and then is normally followed by the emission of light at a longer wavelength. A number of compounds emit fluorescence in the visible range when excited with ultraviolet radiation. Normal poultry skin often exhibits higher emissions compared to tumorous skin. The altered biochemical and morphological state of the neoplastic tissue is reflected in the spectral characteristics of the measured fluorescence.

3.1.2 Hyperspectral Image System for Chicken Data

Instrumentation and Sensing Laboratory (ISL) at Beltsville Agricultural Research Center, Maryland has developed a laboratory-based line-by-line hyperspectral imaging system

capable of reflectance and fluorescence imaging for uses in food safety and quality research [17][41]. The system employs a pushbroom method in which a line of spatial information with a full spectral range per spatial pixel is captured sequentially to cover a volume of spatial and spectral data. Figure 3.1 shows the ISL hyperspectral imaging system equipped with a CCD camera, a spectrograph, a sample transport mechanism, and two lighting sources for reflectance and fluorescence sensing. Two fluorescent lamp assemblies are used to provide a near uniform UV-A (365 nm) excitation to the sample area for fluorescence measurements. A short-pass filter placed in front of the lamp housing is used to prevent transmittance of radiations greater than approximately 400 nm, and thus eliminate the potential spectral contamination by pseudo-fluorescence. The system acquires the data via line-by-line scans while transporting sample materials via a precision positioning table.

The ISL hyperspectral image system captures 65 spectral bands (K = 65) at the wavelengths from λ_1 (425.4 nm) to λ_{65} (710.7 nm) in visible light spectrum. A hyperspectral image of a poultry sample consists of a spatial dimension of 400×460 pixels where each pixel denotes 1 mm × 1 mm of spatial resolution. Each pixel has a 16-bit gray-scale resolution. The data size of a hyperspectral image sample is approximately 24 mega-bytes (= 460 pixels × 400 pixels × 65 bands × 2 bytes). The speed of the conveyer belt was adjusted based on the predetermined CCD exposure time and data transfer rate.

Spectral signature reveals the characteristics of the different types of tissues. Figure 3.2 shows the relative fluorescence intensity of hyperspectral image data at each spectral band for normal tissues and tumors. Normal tissues have a large peak response at approximately band 22 and a smaller peak at approximately band 45. Tumors show lower fluorescence intensities than normal tissues on average, but have strong response between the bands 40 and 45 relative to the peak near the band 22. Background pixels show low fluorescence intensity and an almost flat response over the entire spectral range due to the carrying tray being covered with a non-fluorescent flat black paint.

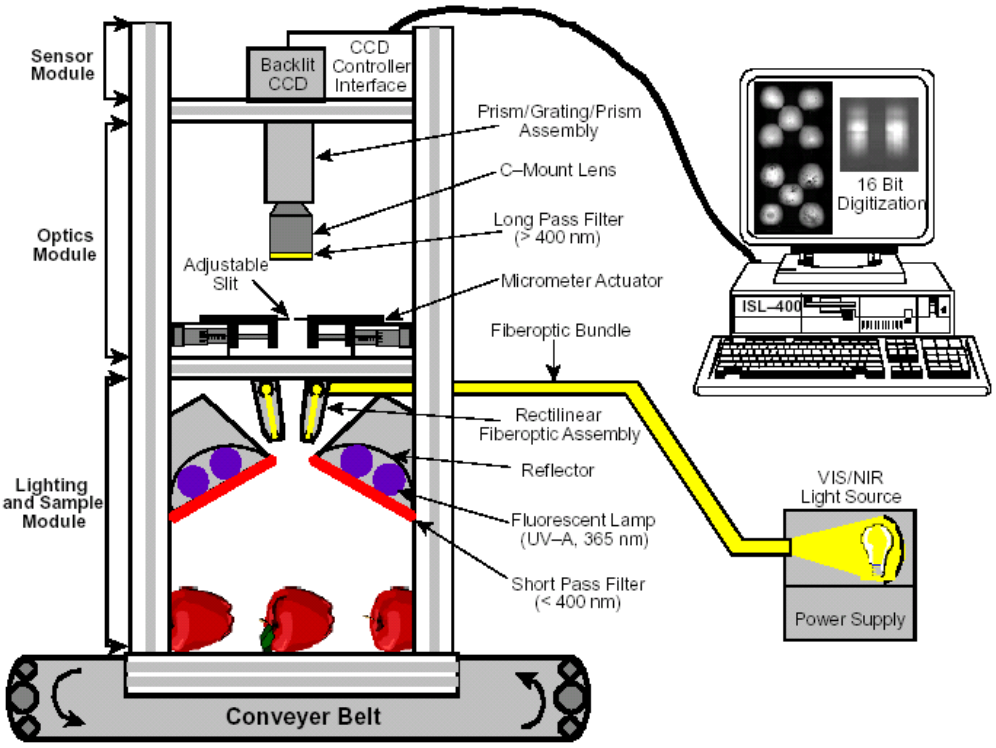


Figure 3.1: Hardware components of the ISL hyperspectral imaging system

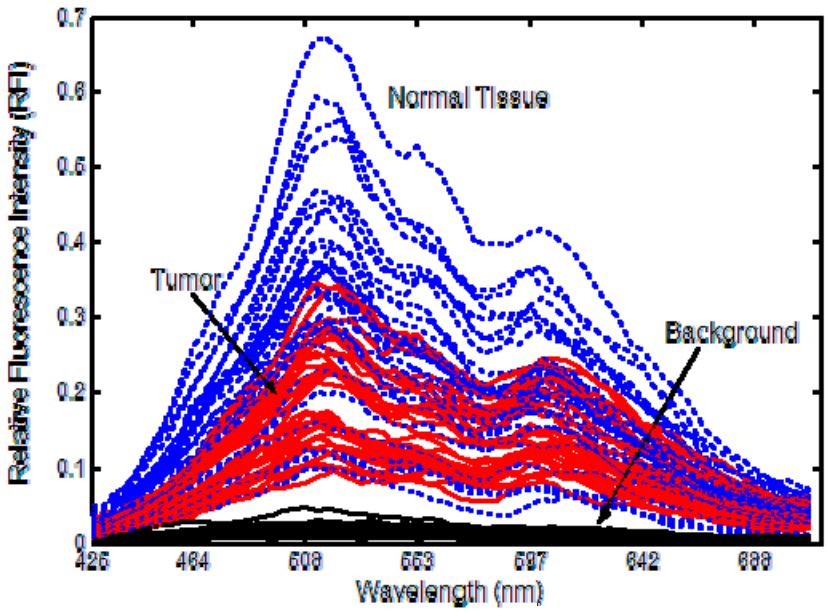


Figure 3.2: Spectral signatures of the tumor and normal tissue measured by relative fluorescence intensity

3.1.3 Hyperspectral Chicken Data Description

Twelve chicken carcasses were collected from a poultry processing plant owned by Allen Family Foods Inc. of Cordova, MD in March and May of 2002. A Food Safety and Inspection Service (FSIS) veterinarian at the plant identifies the condition of the poultry carcasses. Hyperspectral images obtained consist of 460×400 pixels with 65 spectral bands. The spectral band has discrete wavelengths from 425.4 nm (λ_1) to 710.7 nm (λ_{65}). Table 3.1 show the 65 spectral bands used along with the associated wavelength.

The sample poultry carcasses were placed on a tray painted with a non-fluorescent flat black paint to minimize background scattering in a darkened room. The speed of the conveyer belt was adjusted based on the predetermined CCD exposure time and data transfer rate. Figure 3.3 shows 6 spectral images (λ_1, λ_{11},..., λ_{51}) of a hyperspectral image sample obtained by ISL's system.

Table 3.1: Wavelength values of the spectral band used in the image acquisitions

Band #	Wavelength (nm)	Band #	Wavelength (nm)	Band #	Wavelength (nm)	Band #	Wavelength (nm)
1	425.45	2	429.82	3	434.19	4	438.56
5	442.93	6	447.31	7	451.70	8	456.08
9	460.47	10	464.86	11	469.25	12	473.65
13	478.04	14	482.45	15	486.85	16	491.26
17	495.67	18	500.08	19	504.50	20	508.92
21	513.34	22	517.76	23	522.19	24	526.62
25	531.05	26	535.49	27	539.93	28	544.37
29	548.82	30	553.26	31	557.71	32	562.17
33	566.62	34	571.08	35	575.54	36	580.01
37	584.48	38	588.95	39	593.42	40	597.90
41	602.37	42	606.86	43	611.34	44	615.83
45	620.32	46	624.81	47	629.31	48	633.81
49	638.31	50	642.81	51	647.32	52	651.83
53	656.35	54	660.86	55	665.38	56	669.90
57	674.43	58	678.96	59	683.49	60	688.02
61	692.56	62	697.10	63	701.64	64	706.18
65	710.73						

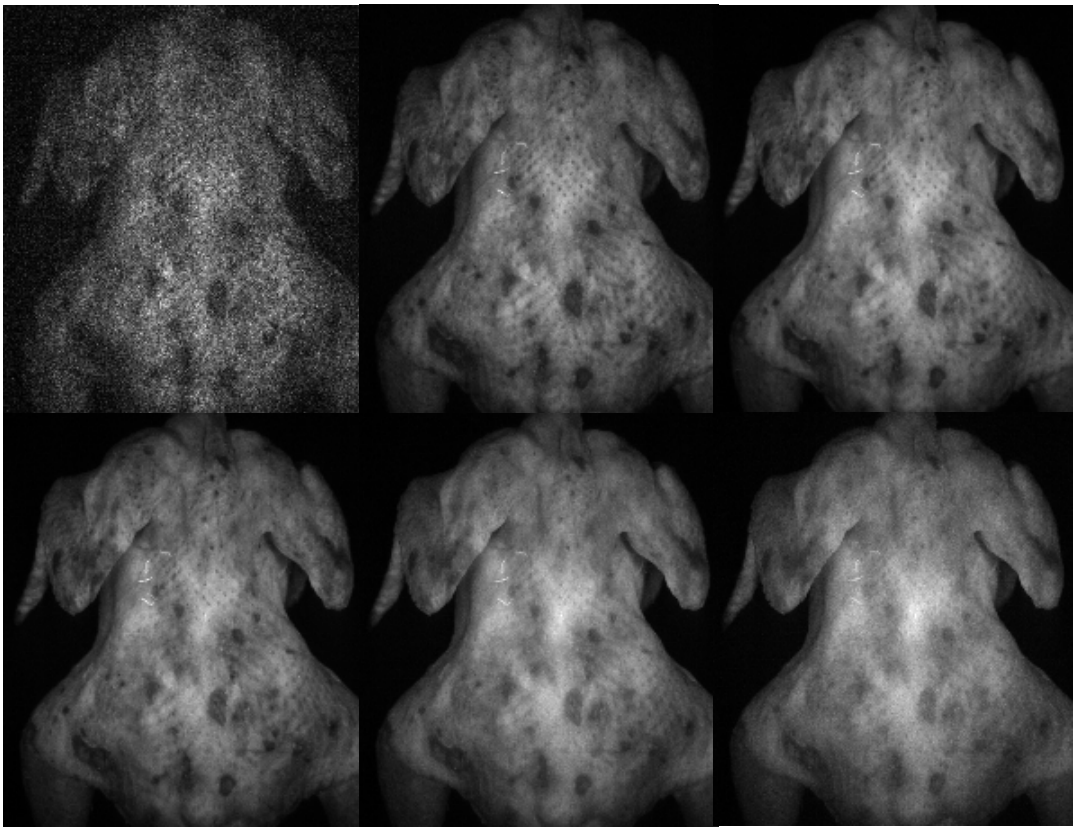


Figure 3.3: Hyperspectral images of a poultry carcass

3.1.4 Data Preprocessing

Image segmentation is performed as preprocessing to remove the poultry carcasses from the background. The background is the tray on which the poultry carcass is placed. Due to the tray painted with non-fluorescent, flat black paint, the fluorescence intensities of these trays are low and almost same for different wavelength. Figure 3.4 displays the histogram of the chicken data. The two peaks in the histogram indicate the intensity of background and poultry carcass. A fixed threshold can easily remove the background. Figure 3.5 shows the image segmentation result.

3.2 Hyperspectral Imaging for Apple Contamination Detection

3.2.1 Background

The other application of using hyperspectral fluorescence image for food safety inspection is automated detection for animal feces contamination on apples. The issue of the detection of apple surface contaminations is very important. For example, contaminated apple juice came to the forefront in recent years with major outbreaks of Escherichia coli infections in people who drank unpasteurized apple juice or cider [40]. The primary source of pathogenic bacteria in unpasteurized juices is animal fecal matter. The government regulations also require no-visual evidence of fecal matter on fruits used to make juices.

In this application, we demonstrate the versatility of the hyperspectral fluorescence imaging techniques for food safety inspection. Here the same hyperspectral image system used in chicken tumor detection is applied to detect apples contaminated with a range of diluted animal feces spanning from visible to invisible to human eye.

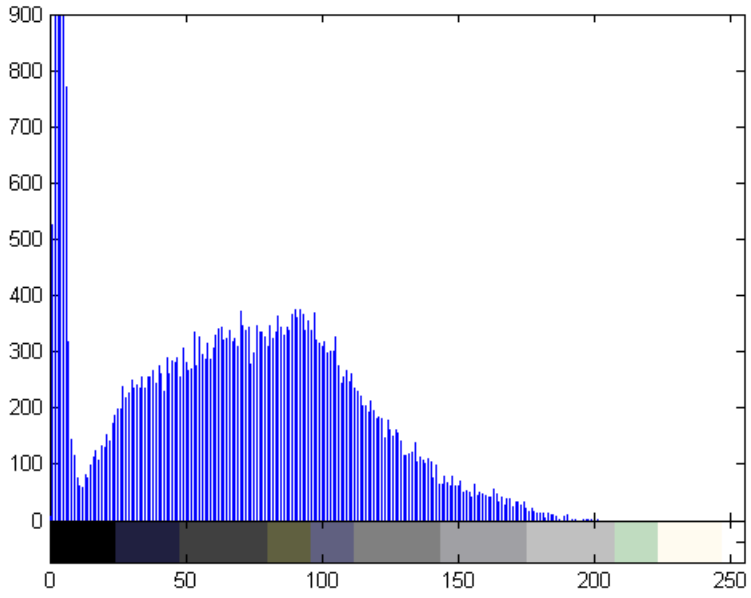
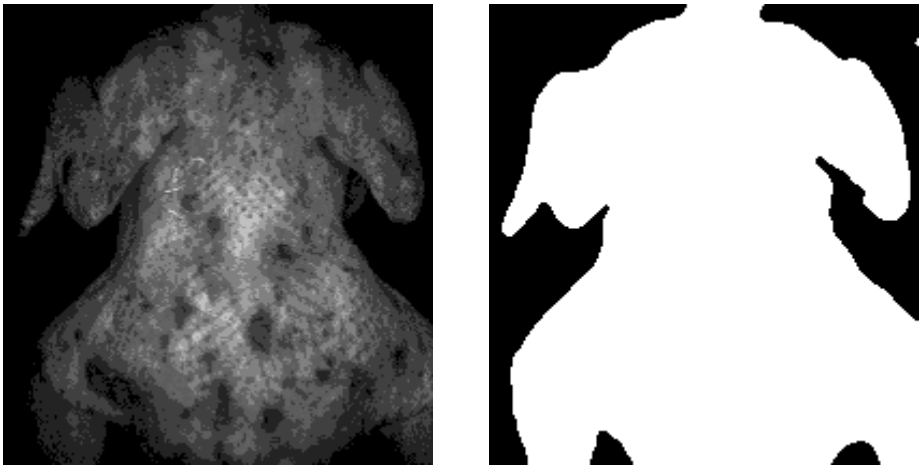


Figure 3.4: Histogram of HS chicken data



(a) Original image (λ_{10}) (b) Segmentation result

Figure 3.5: Segmentation of hyperspectral fluorescence image with a threshold

3.2.2 Hyperspectral Apple Data Description

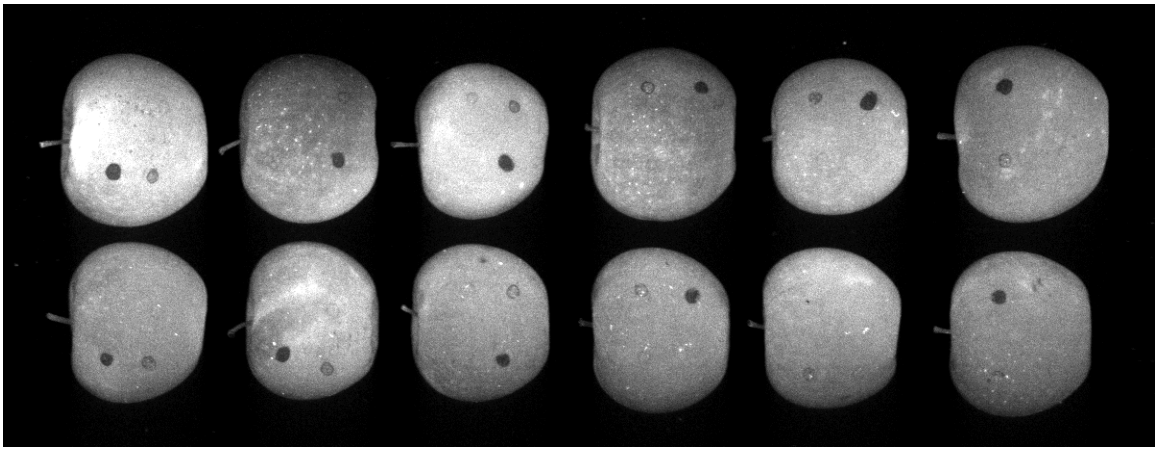
96 Golden Delicious apples were handpicked from crates. Fresh cow feces were collected and diluted 1:2, 1:20, and 1:200 by weight with de-ionized water. Then a drop of each of the three dilutions was immediately applied to each apple. The hyperspectral image system captures 79 spectral bands at wavelengths from 425 to 772nm. Figure 3.6 shows three spectral images (λ_{20}, λ_{40}, λ_{60}) contaminated apples. At the shorter wavelength, like λ_{20} and λ_{40}, the 1:2 feces treatments are clearly visible as darker spots compared to apple surfaces. In contrast, at longer wavelength, the 1:2 feces contamination on apples is not clearly shown. But the fluorescence intensities for feces contaminated spots are generally higher than that of apple surfaces. The spectral signature is shown in Figure 3.7.

3.3 Hyperspectral Imaging for Mouse Skin Cancer Detection

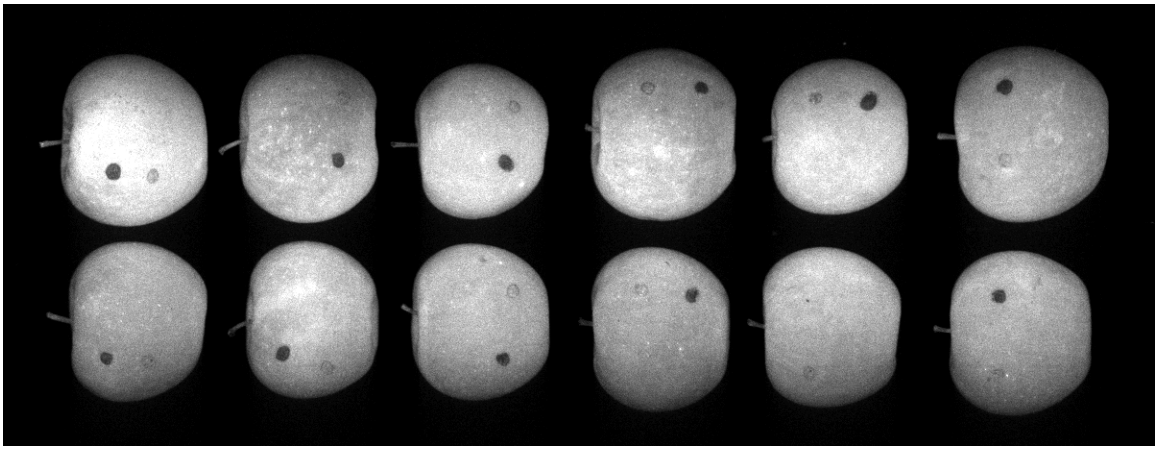
3.3.1 Background

Cancer is the second leading cause of death in the United States, exceeded only by cardiovascular diseases [128]. About one million new cancer cases are expected to be diagnosed and about a half million Americans die of cancer every year. These estimates do not include approximately 1.3 million cases of basal and squamous cell skin cancers that exist in the same time period. Cancers that develop from melanocytes, the pigment-producing cells of the skin, are called melanoma. Melanoma can spread quickly to other parts of the body through the lymph system or through the blood. For most skin cancer patients including melanoma and nonmelanoma skin cancers, early diagnosis and thorough treatment such as complete resection are the keys to gaining a favorable prognosis.

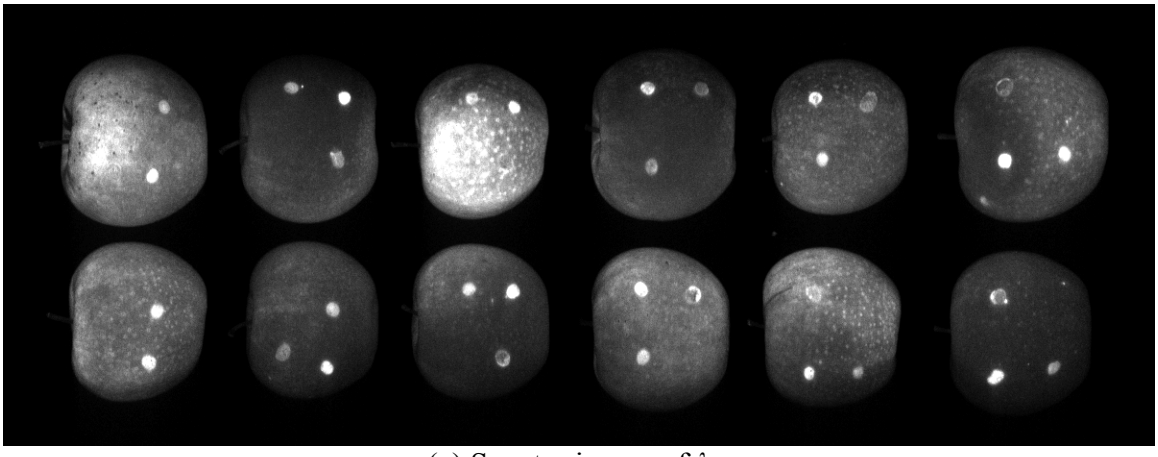
Current diagnostic methods for skin cancers rely on physical examination of the lesion in conjunction with skin biopsy that involves the removal of tissue samples from the body for examination. Biopsy of large lesions often requires substantial tissue removal.



(a) Spectra image of λ_{20}



(b) Spectra image of λ_{40}



(c) Spectra image of λ_{60}

Figure 3.6: Hyperspectral images of contaminated apple

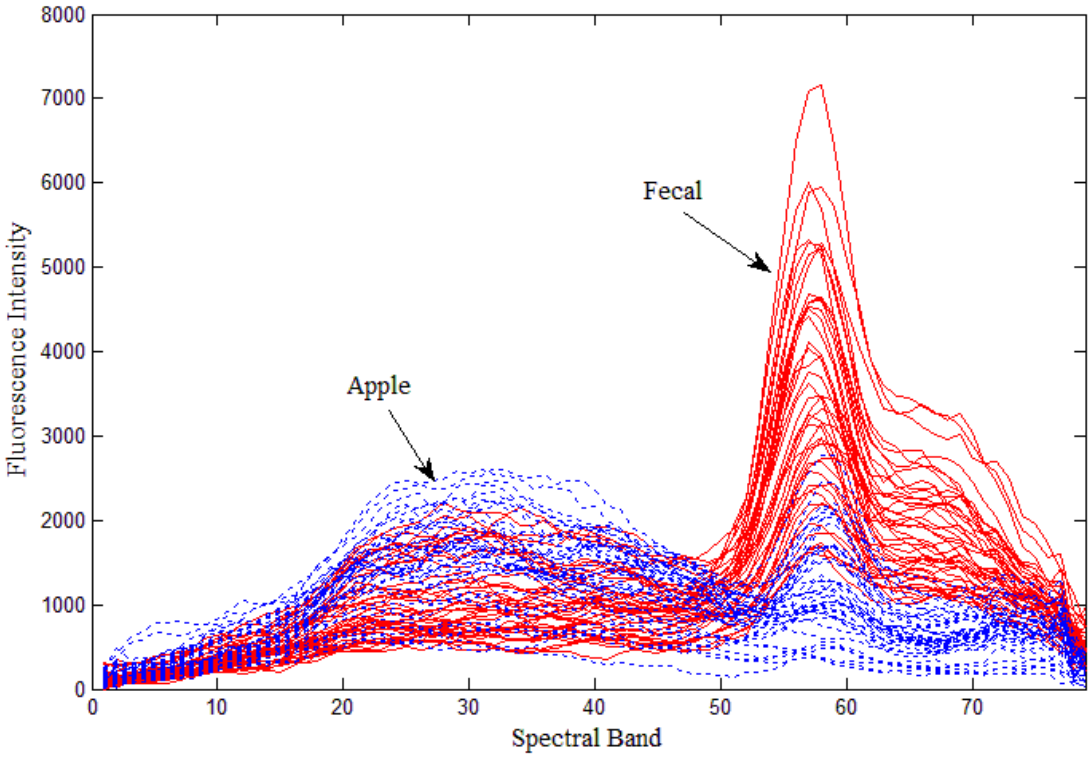


Figure 3.7: Spectral signature of the fecal and apple skin measured by fluorescence intensity

Though this protocol for skin lesion diagnosis has been accepted as the gold standard, it is subjective, invasive, and time-consuming. Since suspicious areas are identified by visual inspection alone, there are a significant number of false positives that undergo biopsy. Conversely, many malignant lesions can also be overlooked. There is an urgent need for objective criteria that would aid the clinician in evaluating whether biopsy is required.

Spectroscopy offers an instant, non-invasive diagnostic tool to detect skin tumor, a typical symptom of skin cancer, accounting for about half of all cancers, based on the spectral properties of tissue [54][48]. Cancer causes the cells grow out of control to form tumors. A hyperspectral image contains spatial information measured at a sequence of individual wavelength across a sufficiently broad spectral range. This enables hyperspectral imaging to reveal useful information for material identification than conventional imaging techniques [6].

3.3.2 Hyperspectral Image System for Mouse Data

The Advanced Biomedical Science and Technology Group at Oak Ridge National Laboratory, Oak Ridge, Tennessee has developed a hyperspectral imaging (HSI) system capable of reflectance and fluorescence imaging [125] [126][127]. For this study the HSI system has been further adapted for skin cancer diagnostics. Figure 3.8 shows a schematic diagram of hardware components of the ORNL hyperspectral imaging system. This system consists of fiber probes for image signal collection, an endoscope, an acousto-optic tunable filter (AOTF) for wavelength selection, a laser excitation source, an endoscopic illuminator (model Olympus CLV-10) equipped with a 300 watt CW Xe arc lamp source, a charge-coupled device (CCD) color camera (model Sony CCD-Iris) for

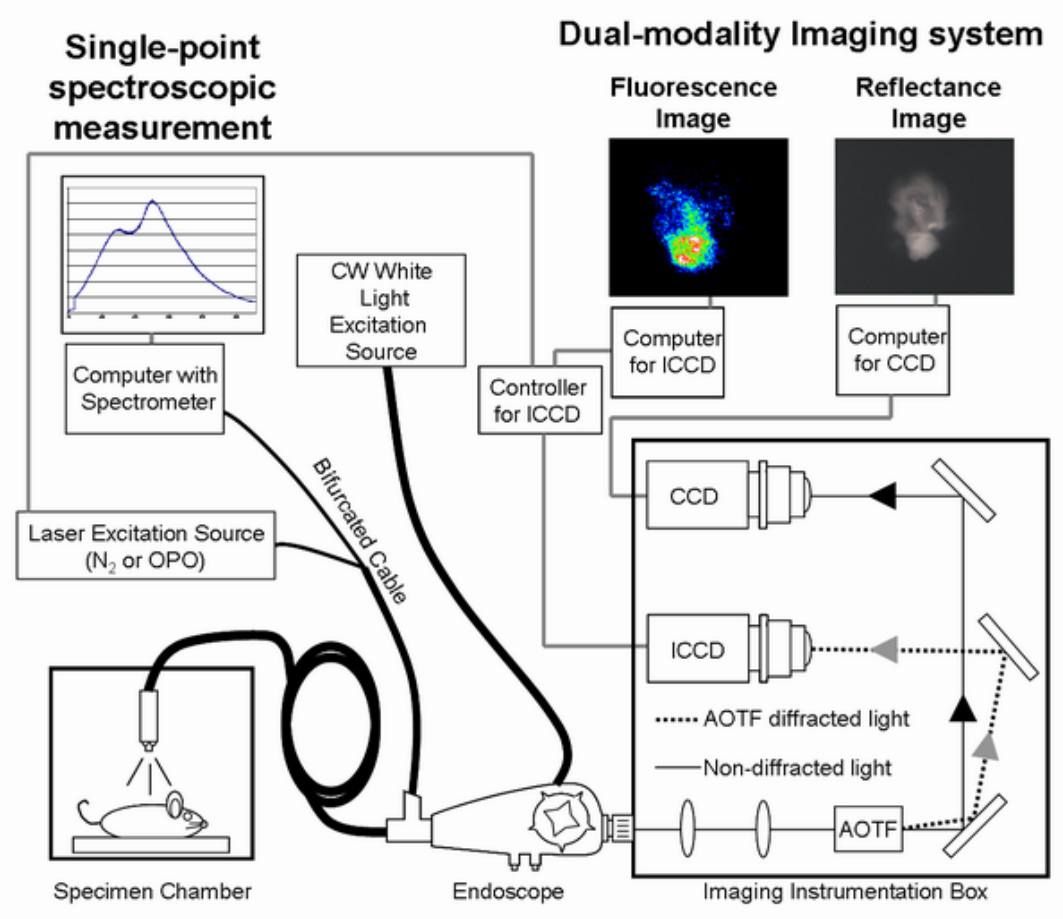


Figure 3.8: Hardware components of the ORNL hyperspectral imaging system

reflection detection, and an intensified charge-coupled device (ICCD) camera (model IMAX-512-T-18 Gen. II) for fluorescence imaging.

The AOTF is an optical bandpass filter whose passband can be electronically tunable using the acousto-optic interaction inside an optical medium whose refractive index is changed by an acoustic wave. The AOTF allows the user to select and transmit a single wavelength from the incoming light. The acoustic wave produces a wavelength-selective single-tone grating in the AOTF transducer that can be varied by simply changing the acoustic frequency. Radio-frequency (RF) signals are used to generate the acoustic waves. The RF amplitude level applied to the transducer controls the filtered light intensity level. Varying the RF frequency corresponding to the wavelength range can do a complete spectral analysis. The AOTF has a dynamic range of 400-650 nm with a 10×10 mm aperture and a spectral resolution of 1-2 nm. The AOTF shows a fast response time (in μs), is accurate, and exhibits a high extinction ratio.

Both fluorescence and reflected lights are collected through the endoscope into the AOTF device via collimating lenses. A mirror placed in front of the AOTF projects the acquired images onto the ICCD camera for fluorescence imaging and onto the CCD camera for reflection measurement. Reflectance images are acquired using an endoscopic illuminator. The reflection source was coupled to a gastrointestinal endoscope (Olympus T120) equipped with an imaging bundle. Fluorescence spectra and images are acquired using a LSI pulsed Nitrogen laser (model VSL-337) with a maximum repetition rate of 20 Hz. For fluorescence imaging, the N_2-pumped laser was coupled to a bifurcated fiber probe (R400-7-VIS/NIR) that was also used to sample point measurements using a miniature fiber optic spectrometer (Ocean Optics USB2000-FLG).

The fluorescent light emitted by the tissues is diffracted by the AOTF (Brimrose TEAF10-0.4-0.65-S) at a 60-degree angle from the undiffracted (zero-order) beam thus separating the reflected image from the fluorescent image. Individual wavelengths by the AOTF are thus sent to the ICCD. A Brimrose AOTF controller (model VFI-160-80-DDS-

A-C2) controls the AOTF. The controller sends an RF signal to the AOTF based on the input provided using Brimrose software. Wavelength selection takes place in microseconds enabling ultra-fast modulation of wavelength output to the ICCD. Wavelength specific images were taken between 440-650 nm every 10 nanometers. In addition to the imaging capability, spectral information from each site was obtained using the Ocean Optics spectrometer coupled to a laptop computer.

Fluorescence images were acquired by gating the intensified ICCD camera. A timing generator incorporated into the ICCD camera's controller (ST-133) allowed the ICCD to operate in the pulsed mode with a wide range of programmable functions. A 500 ns delay between the laser trigger and the detector activation was programmed to synchronize the laser and the detector. The intensifier was gated for 500 ns during which a 5 ns laser pulse was delivered to the tissues. An image was captured 20 times per second, integrated by internal software, and output to a screen once per second. This allows real-time fluorescence detection. Fluorescence images and spectra were acquired and processed with WinView (Roper Scientific) and OOIBase32 (Ocean Optics) software, respectively. Reflection images were captured and processed with SimplePCI image analysis software (Compix, Inc, Cranberry, Township, PA).

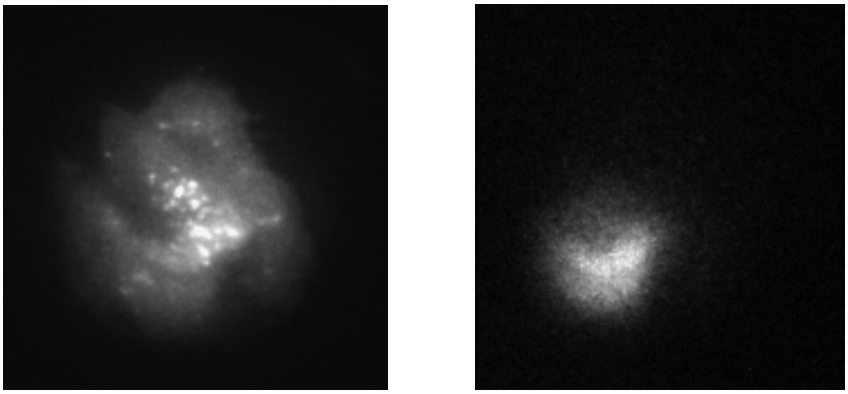
3.3.3 Hyperspectral Mouse Data Description

The hyperspectral image data used in this experiment are fluorescence images data taken from a mouse skin sample by Oak Ridge National Laboratory. The mouse tissue was a sample image taken from a larger study of adult nude mice injected subcutaneously with 100 μL of Fischer rat 344 rat tracheal carcinoma cells (IC-12) to induce tumor formation. The mice were nude to prevent hair from interfering with our measurement and to assist with tumor formation due to having compromised immune systems. The subdermal injection was done as close to the skin surface as possible to allow tumor formation close to the skin surface. This also allowed experiments to be performed *in-vivo* rather than

after tissue extraction. Experiments performed on live, whole animals are a vast improvement over tissue samples due to closer simulation of *in-vivo* human cancerous conditions. After injection, the nude mice were incubated for a period of four days to allow tumor formation to occur. Once a tumor was observed (approximately 5 mm), the mice were anesthetized for approximately 30 minutes to permit data collection. After data collection, the mice were humanely sacrificed to avoid undue suffering. The Oak Ridge National Laboratory committee on the ethical treatment of animals acted as a governing body on all matters concerning animal testing and all work was done under IACUC #A3170-01. Strict protocols advised by the committee were followed when dealing with animals. The hyperspectral images consist of 165×172 pixels with 21 spectral bands from the wavelength λ_1 (440 nm) to λ_{21} (640 nm) with 10 nm spectral resolutions in the spectral region. Such a fine spectral resolution provides sufficient information for precise study of tumor detection. The fluorescence image is enhanced with a Gamma correction with $\gamma = 0.8$. The mapping is weighted toward higher (brighter) output values. Figure 3.9(a) shows a reflectance image of a skin tumor region taken from a mouse sample. Figure 3.9(b) is a fluorescence image of the band λ_6 (490 nm) of the same spot. Images of a small circular area are obtained by an endoscope. Lower part of the fluorescence image (a bright U-shaped area) corresponds to normal tissue, and the upper part is a tumor region.

3.3.4 Registration of Spectral Band Images

Original spectral band images captured by the hyperspectral imaging system are not spatially aligned since the AOTF diffracts the light at different wavelengths. Acousto-optics involves the interaction of sound and light in dielectric material. When sound propagates through a solid or liquid, compressions are created in the material that cause variable refraction of the passing light, pulling color features from it. Filters can be used to pass light with either a single wavelength or multiple wavelengths, depending on the RF signal applied. A piezoelectric oscillator bonded to an acousto-optic medium converts



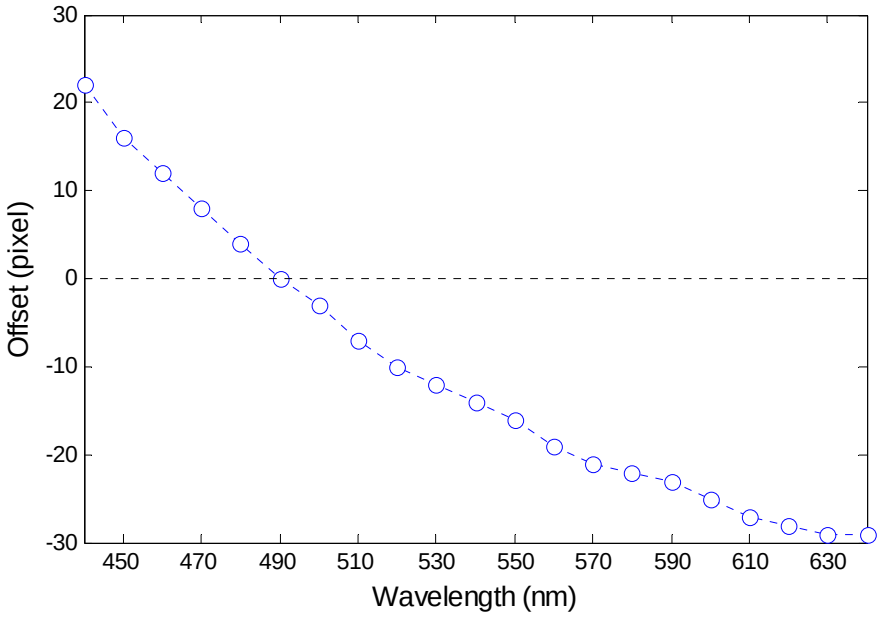
(a) Reflectance image (b) Fluorescence image at wavelength 490 nm

Figure 3.9: Skin tumor region of a mouse skin sample.

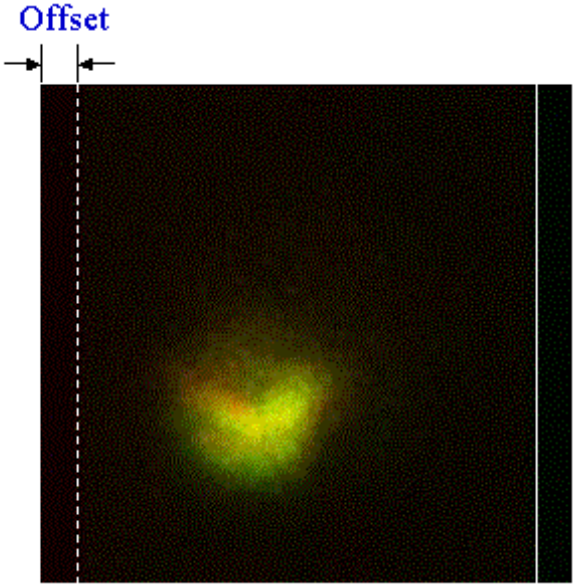
a high-frequency electrical signal into an ultrasonic wave. Figure 3.10(a) shows the amount of offsets between the images measured at different wavelengths that the AOTF generates. The spectral band of the wavelength 490 nm was selected as a reference band with zero offset. Positive offset values indicate the number of pixels of the image shifted to the right with respect to the reference band. Figure 3.10(b) illustrates a pair of band images at the wavelengths of 540 nm and 490 nm spatially co-registered by translating the image at 540 nm for the offset of 14 pixels.

Figure 3.11 shows nine spectral band images from wavelengths 450 nm to 610 nm. All the spectral band images are co-registered with the reference band image (490 nm) according to the procedure described in Section 3.3.4.

Figure 3.12 shows relative fluorescence intensity of a pixel taken from a mouse skin sample in terms of two categories of interest: normal tissue and tumor. Normal tissues have higher fluorescence intensity on average compared to tumors. Normal tissues have a peak response at approximately band 490 nm. Tumors show less distinctive spectral peaks.



(a) The amount of offsets of spectral band images caused by the AOTF with respect to the reference band at 490 nm



(b) Registered band images of 490 nm and 540 nm

Figure 3.10: Offset of image registration

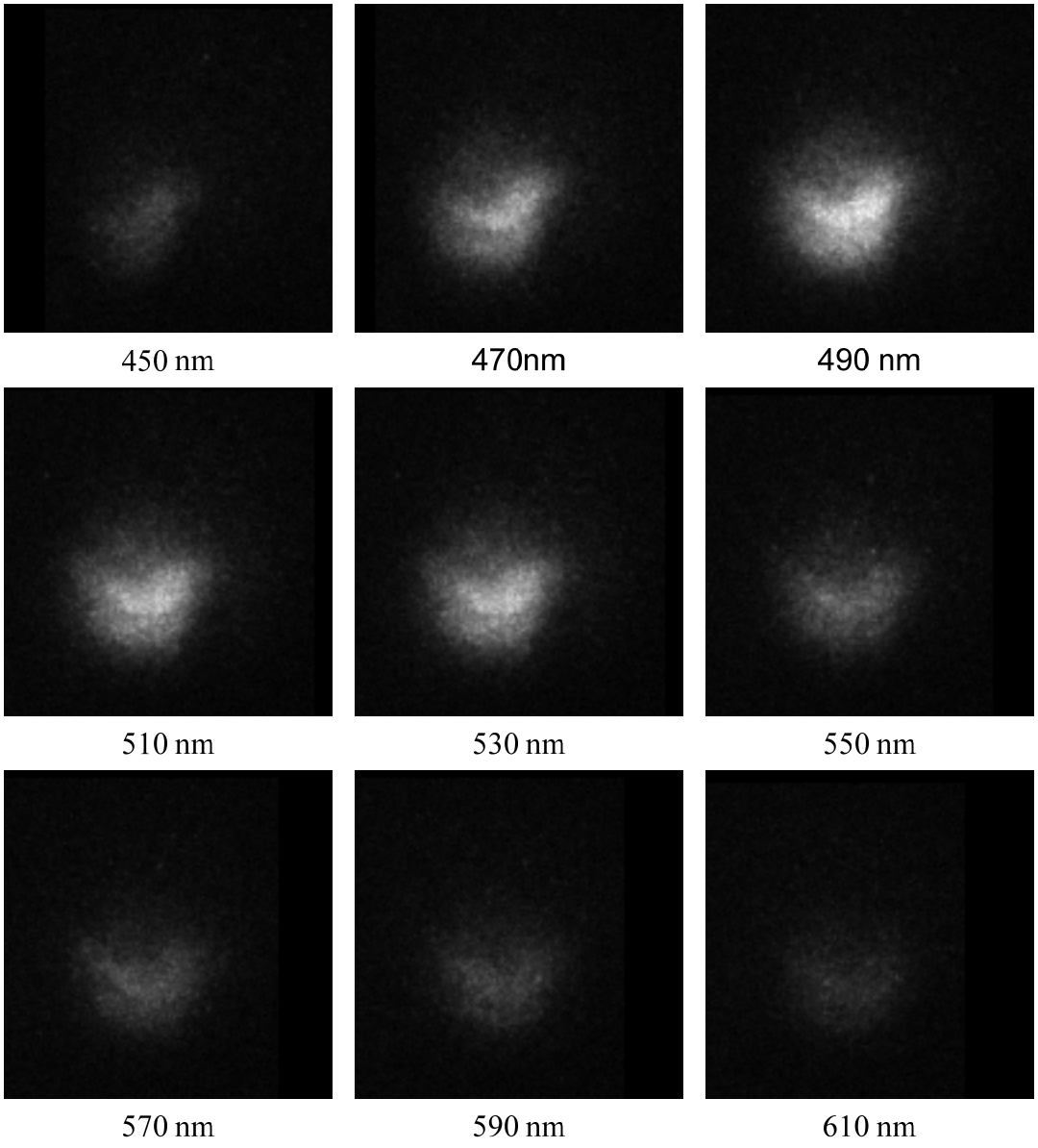


Figure 3.11: Spectral band images from the wavelength 450 nm to 610 nm.

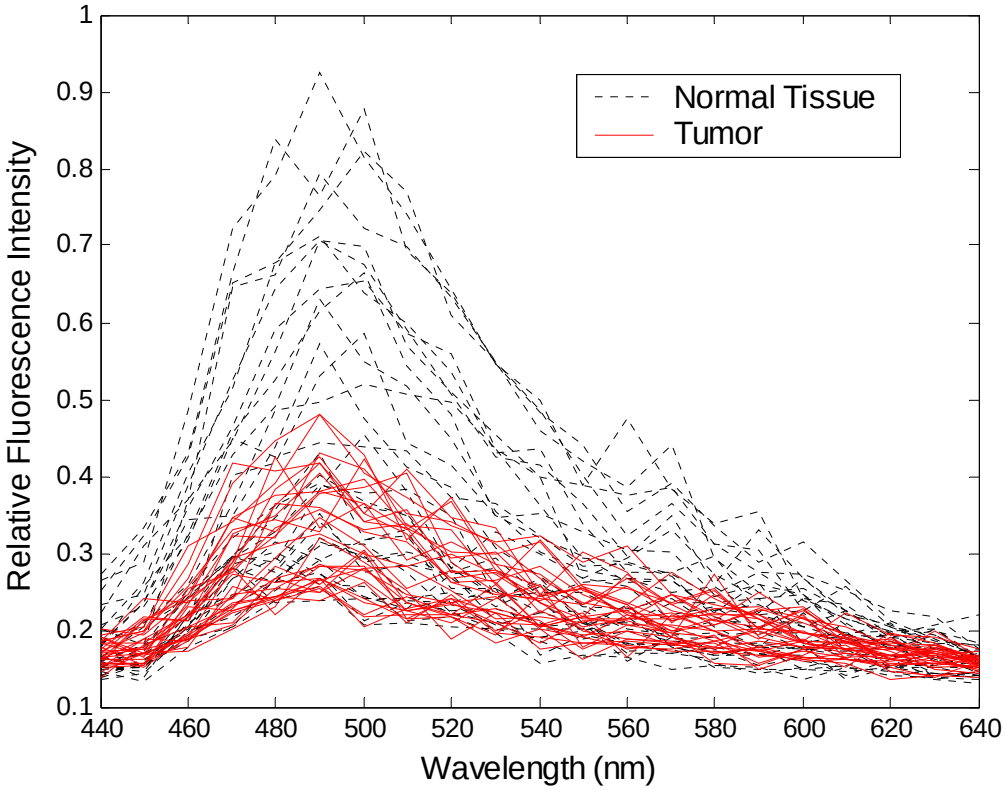
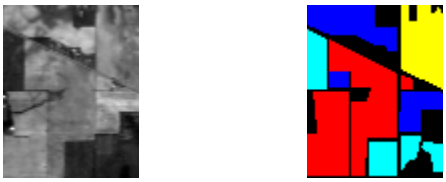


Figure 3.12: Spectral signature of normal tissue and tumor

3.4 Indiana Pine Data

The Indiana Pine data is a well-known publicly available hyperspectral data set, which is used to investigate land use and can be downloaded from ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/. The Indiana Pine data are delivered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), which measured pixel response in 224 bands in the 0.4 to 2.45 μm region of the spectrum with about 10nm intervals, at a spatial resolution of 20 m, and covers an agricultural portion of North West Indiana. The Indian Pines data consists of 145×145 pixels in 220 contiguous spectral bands. This data set includes about two-thirds agriculture, and one-third forest or other natural perennial vegetation. Four of the 224 AVRIS bands do not contain data, leaving 220 bands. Similar to the earlier work on this dataset [103], twenty bands, λ_{104} - λ_{108} (1.36-1.40 μm), λ_{150}-λ_{163} (1.82-1.93μm), and λ_{220} (2.50 μm), where the atmosphere is opaque have been omitted from the data set. Therefore, a total of 200 bands have been used in experiments. The advantage of using this dataset is the availability of the reference image produced from field surveys, which may be used for accuracy assessment purposes.

In this dissertation, we use only a part of the 145 × 145 scene, called the subset scene for a size of 68 × 86. The subset scene contains four classes: Corn-notill, Soybean-notill, Soybean-mintill, and Grass-Trees, and over 75% of this scene are labeled. Figure 3.13 shows the spectral image of subset and the ground truth for subset. Figure 3.14 shows the spectral signature of the four classes.



(a) Spectral image (b) ground truth

Figure 3.13: Spectral image and ground truth of subset

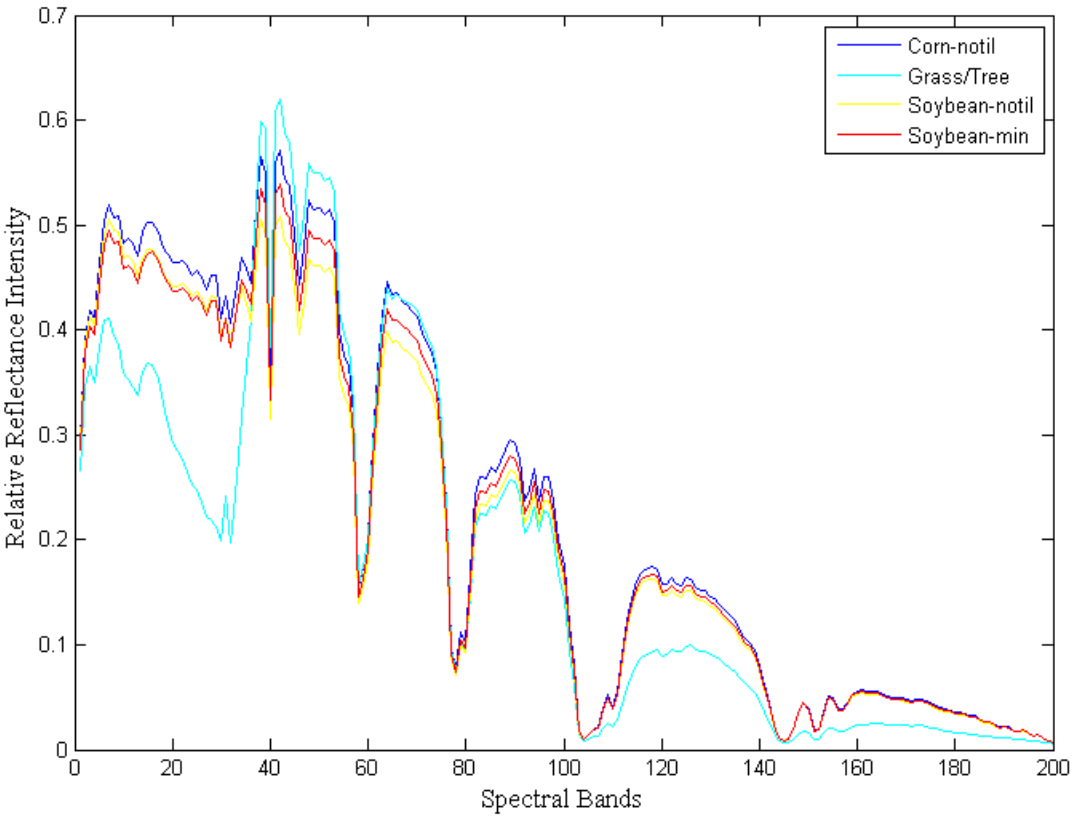


Figure 3.14: Spectral signature of the four classes in subset scene

4 Integration of Spatial and Spectral Information for Hyperspectral Image Band Selection

An important task in hyperspectral data analysis is to reduce the redundancy of the spectral and spatial information without losing any valuable details needed for the subsequent detection, discrimination, and classification processes. This chapter first presents a band selection method of hyperspectral images based on the ***recursive divergence***. A set of ***recursive equations*** for the fast calculation of divergence with an additional band is derived to overcome the computational restrictions in real-time processing.

The major impediment to the combined spatial and spectral approach is that most spatial methods were only developed for single image band. Based on the traditional single-image based local Geary measure, this dissertation successfully ***proposes a Multidimensional Local Spatial Autocorrelation (MLSA) for hyperspectral image data***. This new measure is a fundamental improvement in the description of spatial autocorrelation for hyperspectral imaging. Then based on the proposed spatial measure, this research work ***develops a collaborative band selection strategy that combines both the spectral separability measure (divergence) and spatial homogeneity measure (MLSA) for hyperspectral band selection task***.

The chapter is organized as follows. Section 4.1 introduces the background of band selection. Section 4.2 provides a review of the literature concerning band selection algorithm for hyperspectral image data. Then, the proposed recursive divergence based band selection algorithm is thoroughly described in Section 4.3. In Section 4.4.2, the new Multidimensional Local Spatial Autocorrelation measure is proposed. Section 4.4.3 describes the proposed collaborative band selection method. The experimental results for proposed band selection methods are represented in Section 4.5.

4.1 Introduction

Hyperspectral sensors collect the electromagnetic spectrum at dozens or hundreds of wavelength ranges in the visible and near infrared spectra. A three-dimensional (3-D) volume of data in spatial and spectral spaces characterizes a hyperspectral image [6]. The high spectral resolution is the key characteristic of hyperspectral imaging, which is provided over a large and continuous wavelength region. Spectral resolution is a measure of the narrowest spectral wavelength that can be resolved by sensor. Spectral resolution acts as the pivotal parameter in identifying different materials. It was found that some minerals might have similar features at low-resolution spectrum. However, at high resolution, their spectrum might be quite different. For instance, the AVIRIS hyperspectral sensor can measure the upwelling radiance spectrum from 400 to 2500 nm at 10 nm resolution. Its unique optical sensor delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (also called bands).

The increased number of spectral bands means the higher dimensionality of hyperspectral data. The resulting high-dimensional datasets, on one hand, benefit better discrimination power among similar spectral signatures. On the other hand, such a large amount of hyperspectral image data often involves problems in storage, transmission and processing, which makes real-time computer processing a challenging task.

And in addition, the larger data volumes from such hyperspectral image make traditional processing techniques inefficient for many applications. One of these techniques is supervised classification, which uses labeled samples available for training the classifier and estimating its performance. A rule of thumb used in remote sensing image classification techniques is that the number of training pixels should be at least ten times, but preferably hundreds times, the dimensionality of the data set. However, the size of training samples for such classification techniques is normally limited. Moreover, as the number of available bands increases, such techniques become less applicable, resulting in severe limitations. Thus, beyond a certain point, if the number of training samples per

feature is small, the addition of more dimensions leads to misclassification [81]. This phenomenon is called “Hughes phenomenon” or “Curse of dimensionality”. The Hughes phenomenon states that after adding many features to a classifier which is trained using a fixed size training set, the classification performance starts to degrade.

In order to overcome the curse of dimensionality, dimensionality reduction techniques can be applied to the original hyperspectral bands. There are a number of dimensionality reduction techniques, and according to the adopted reduction strategy they are usually divided into band selection and feature extraction approaches. Feature extraction is often referred to as data transformation from a high order dimension to a low order dimension in order to concentrate the vital information while discarding redundant data. These feature extraction techniques include, but are not limited to, principal component analysis (PCA), fisher’s linear discriminant analysis (FDA), independent component analysis (ICA) [101], projection pursuits [82] and wavelet transform [83][25]. As an alternative, band selection can be defined as the subset of features that provide the highest-class separability. The key difference between band selection and feature extraction is that in the former only a subset of original bands is selected while the latter is based on a functional mapping to generate a completely new feature space in fewer dimensions. Band selection methods are particularly welcome because the selected features retain the original meanings researchers are familiar with.

Band selection is a process that selects a subset of original bands. The optimality of a band subset is measured by an evaluation criterion. Typically, a small number of bands can give sufficient information for the classification. The band selection procedure finds the small subset of bands that are relevant to the target concept. A small subset of relevant bands gives more discriminating power than using more bands. As a result, the band selection gives a better generalization error. Benefits of the band selection include reducing the number of bands, removing irrelevant, redundant, or noisy data, speeding up processing time and improving classification performance.

4.2 Review of Band Selection Techniques

Band selection techniques generally involve both a search algorithm and a criterion function. The search algorithm generates and compares possible subsets of bands by applying the criterion function as a measure for each considered band subset. A criterion function J must be a measure of the overlap or class separability among the distributions. Within-class and between-class scatter matrices are often used to formulate criteria of class separability. In order to formulate criteria for class separability, these matrices need to be converted to a number. After band selection, samples should be close to those in the same class but far from those in different classes. So, this value of J should be larger when the class separability is larger.

Relief is an instance-based feature ranking scheme introduced by Rendell and Kira [84] and later enhanced by Kononenko [85]. The basic idea of Relief is to sample instances at random and then compute their nearest neighbors from the same and opposite class, and adjust a feature weighting vector to give more weight to features that discriminate the instance from neighbors of different classes. The theory is that a useful feature should differentiate between instances from different classes and have the same value for instances from the same class. Relief algorithm has some extensions defined by Kononenko [85]. ReliefF is one of these extensions. It is not limited to two class problems as Relief algorithm is, and it is more robust and can deal with incomplete and noisy data. The key idea of the ReliefF is to estimate the quality of attributes according to how well their values distinguish between instances that are near to each other. In ReliefF algorithm, a set of feature vectors for the training instances and their labels is taken and the vector $\boldsymbol{W}$ which contains the estimations of the qualities of the features is output. Given a randomly selected instance $\boldsymbol{x}_i$ from class L, ReliefF searches for k of its nearest neighbors from the same class called nearest hits H, and also k nearest neighbors from each of the different classes, called nearest misses M. Then, for each feature, it updates the quality estimation $\boldsymbol{W}_j$ based on their values for $\boldsymbol{x}_i$, H, M. If instance $\boldsymbol{x}_i$ and those in H have different values on the j-th feature, then the $\boldsymbol{W}_j$ is decreased. On the other hand, if $\boldsymbol{x}_i$

and those in M have different values on the j-th feature, $\boldsymbol{W}_j$ is increased. This whole process is repeated n times which is set by users.

SVM Recursive Feature Elimination (SVM-RFE) method selects a small subset of genes by using the SVM [88]. SVM-RFE is a recursive feature elimination method based on sensitivity analysis for a cost function

$$J = \frac{1}{2}\|\boldsymbol{w}\|^2 \tag{4.1}$$

where $\boldsymbol{w}$ is the weight vector obtained from SVM. At each step, a feature is discarded according to a criterion related to coefficients of the weight vector of a SVM, $b_j = (w_j)^2$. The feature with the smallest criterion score is eliminated, and the SVM is re-trained at each step.

Several methods are proposed based on the original SVM-RFE algorithm. Unlike the SVM-RFE method, at each step, Duan *et al*. [89] proposed an approach to compute the feature ranking score from a statistical analysis of weight vectors of multiple linear SVMs trained on subsamples of the original training data. In [90], the authors suggested to use ten-fold cross validations to improve the reliability of the top feature subsets selected with SVM-RFE.

Millette [91] designed an expert system, which makes use of the Jeffreys-Matusita distance measure and a library of prototype reflectance data, to make spectral decisions.

$$J_{ij} = 2(1 - e^{-B_{ij}}) \tag{4.2}$$

Where B_{ij} is Bhattacharyya distance

$$B_{ij} = \frac{1}{8}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \left[\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right]^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \frac{1}{2}\ln\left[\frac{\left|\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j\right|/2}{\left|\boldsymbol{\Sigma}_i\right|^{1/2}\left|\boldsymbol{\Sigma}_j\right|^{1/2}}\right] \tag{4.3}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the mean vector and covariance matrix of class ω_i, respectively.

Keshava [92] proposed a band selection algorithm based on the spectral angle mapper (SAM) metric, which is the angle between the two spectra. For two *J*- dimensional spectra, $\boldsymbol{x}$ and $\boldsymbol{y}$, θ is given by the following analytical expression:

$$\theta(\boldsymbol{x}, \boldsymbol{y}) = \arccos\left(\frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}\right) \tag{4.4}$$

Tu *et al*. [93] proposes a band selection algorithm coupled with feature extraction for data dimensionality reduction based on canonical analysis (CA). Using the eigenvalues and eigenvectors generated by CA, a loading factor matrix can be defined, through which a discriminant power (DP) is calculated for each bands. Du [94] used high-order moments for band ranking and divergence for band decorrelation. Du, Qi, and Wang, [95] used the independent component analysis for the band selection. Ifarraguerri and Parairie [96] presented the band selection algorithm based on the Jefferis-Matusita metric. Huang [97] proposes a band selection method based on the pairwise separability criterion and matrix coefficients analysis. Mutual information, which derived from the concept of entropy, measures the statistical dependence between two random variables and therefore can be used to evaluate the relative utility of each band for classification. A mutual information-based band selection method is presented by Guo *et al*. in [98].

4.3 Recursive Divergence Based Band Selection

This approach depends on the concept of a measure of "statistical distance" between the probability densities characterizing the sample classes. Intuitively, this distance or separability measure should has the property that for two feature sets *A* and *B*, if the distance or separability between two class densities were greater for feature set *A* than for feature set *B*, then the error probability obtained for set *A* would be less than for set *B*.

The divergence is one of popular criteria for class separability. Spectral bands in hyperspectral images are highly correlated. The divergence takes into account the

correlation that exists among the various selected bands and influences the classification capabilities of the spectral bands that are selected. We use the divergence to determine feature ranking and to evaluate the effectiveness of class discrimination in hyperspectral image data. The divergence is defined as the total average information for discriminating class ω_i from class ω_j, and given by [99]

$$D_{ij}(\boldsymbol{x}) = \int_{-\infty}^{+\infty} \left[p_i(\boldsymbol{x}) - p_j(\boldsymbol{x}) \right] \ln \frac{p_i(\boldsymbol{x})}{p_j(\boldsymbol{x})} d\boldsymbol{x} \tag{4.5}$$

where $p_i(\boldsymbol{x})$ is the probability density function of $\boldsymbol{x}$ in class ω_i. The divergence is the symmetric version of Kullback-Leibler distance, and it is nonnegative, monotonic, and additive for independent variables.

Suppose that signal classes are characterized by p-dimensional multivariate normal distributions: $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the mean vector and covariance matrix of class ω_i, respectively. Then, the divergence between these two classes is given by [100]

$$D_{ij}(\boldsymbol{x}) = \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1} \right) \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right) \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right)^T \right] + \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j \right) \left(\boldsymbol{\Sigma}_j^{-1} - \boldsymbol{\Sigma}_i^{-1} \right) \right] \tag{4.6}$$

where tr is the notation for the trace of a matrix.

If the covariance matrices of the two normal distributions are equal, that is, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$, then the divergence can be simplified to

$$D_{ij}(\boldsymbol{x}) = tr\left[\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right) \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right)^T \right] = \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right) \tag{4.7}$$

which equals the Mahalanobis generalized distance. The form of Equation (3.6) is close to that of the Bhattacharyya distance with first and second terms indicating class separabilities due to mean- and covariance-differences. The advantage of divergence is that both the first and second terms are expressed by the trace of a matrix, while the Bhattacharyya distance is the combination of trace and determinant.

From the training samples, the sample covariance matrix of class ω_i can be calculated as follows:

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} z_{ij}(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}}_i)^T \tag{4.8}$$

where $z_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{x}_j \in \omega_i \\ 0, & \text{otherwise} \end{cases}$, and $N_i = \sum_{j=1}^{N} z_{ij}$, N is total number of samples, and $\hat{\boldsymbol{\mu}}_i$ is the sample mean vector of class ω_i given by $\hat{\boldsymbol{\mu}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} z_{ij}\boldsymbol{x}_j$.

Given L spectral bands, the number of possible subsets to find the best size d spectral bands is

$$n_c = \frac{L!}{(L-d)!d!}, \tag{4.9}$$

which can be very large even for moderate values of L and d. For example, selecting the best 6 spectral bands out of 65 bands in our case study of the detection of poultry carcasses means that 82598880 band sets must be considered, and evaluating the divergence criterion for every band set in an acceptable time may not be feasible. Thus, we propose the suboptimal band selection method based on the recursive calculation of the divergence.

The basic idea is to build up a set of d spectral bands incrementally, starting with the empty set. That is, the search algorithm constructs the set of spectral bands at the ith stage of the algorithm from that at the (i-1)th stage by the addition of a spectral band from the current optimal set. ***The divergence criterion Equation (4.6) at stage i can be evaluated by updating its value already calculated for stage (i-1) instead of computing the divergence from their definitions***. This results in substantial computational savings.

Let $D_{ij}(\boldsymbol{x}_p^*)$ be the divergence with p selected bands and $D_{ij}(\boldsymbol{x}_p^*, x_{p+1}^*)$ the divergence with the additional band x_{p+1}^*. Suppose the additional band x_{p+1}^* has mean μ_k^*, variance σ_k^2; and the covariance vector between x_{p+1}^* and the elements of $\boldsymbol{x}_p$, $\boldsymbol{z}_k$ for class k (=i or

j). Then the new mean vectors and new covariance matrix are $\boldsymbol{\mu}_k^v = (\boldsymbol{\mu}_{k,p}^*; \mu_k^*)^T$, ($k = i$ *or* j) and

$$\boldsymbol{\Sigma}_{k,p+1} = \begin{pmatrix} \boldsymbol{\Sigma}_{k,p} & \boldsymbol{z}_k \\ \boldsymbol{z}_k^t & \sigma_k^2 \end{pmatrix} \tag{4.10}$$

divergence with the additional of a band x_{p+1}^* can be calculated based on its definition in (4.6) as follows:

$$\begin{aligned} D_{ij}\left(\boldsymbol{x}_p^*, x_{p+1}^*\right) = \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p+1} - \boldsymbol{\Sigma}_{j,p+1}\right)\left(\boldsymbol{\Sigma}_{j,p+1}^{-1} - \boldsymbol{\Sigma}_{i,p+1}^{-1}\right)\right] \\ + \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p+1}^{-1} + \boldsymbol{\Sigma}_{j,p+1}^{-1}\right)\left(\boldsymbol{\mu}_i^v - \boldsymbol{\mu}_j^v\right)\left(\boldsymbol{\mu}_i^v - \boldsymbol{\mu}_j^v\right)^T\right] \end{aligned} \tag{4.11}$$

The inverse of the new covariance matrix with an additional band can be obtained by the following recursive formula

$$\boldsymbol{\Sigma}_{k,p+1}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{k,p}^{-1} + \gamma_k \delta_k^{-1} \gamma_k^t & -\gamma_k \delta_k^{-1} \\ -\delta_k^{-1} \gamma_k^t & \delta_k^{-1} \end{pmatrix} \tag{4.12}$$

where $\gamma_k = \boldsymbol{\Sigma}_{k,p}^{-1} z_k$ and $\delta_k = \sigma_k^2 - z_k^t \boldsymbol{\Sigma}_{k,p}^{-1} z_k$.

Replacing this inverse matrix in the equation of divergence when a new band x_{p+1}^* is to be considered, we can obtain

$$\begin{aligned} D_{ij}\left(\boldsymbol{x}_p^*, x_{p+1}^*\right) &= \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p} - \boldsymbol{\Sigma}_{j,p}\right)\left(\boldsymbol{\Sigma}_{j,p}^{-1} - \boldsymbol{\Sigma}_{i,p}^{-1}\right)\right] \\ &\quad + \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p} + \boldsymbol{\Sigma}_{j,p}\right)\left(\boldsymbol{\mu}_{i,p}^* - \boldsymbol{\mu}_{j,p}^*\right)\left(\boldsymbol{\mu}_{i,p}^* - \boldsymbol{\mu}_{j,p}^*\right)^T\right] + \Delta_{ij}\left(x_{p+1}^*\right) \\ &= D_{ij}\left(\boldsymbol{x}_p^*\right) + \Delta_{ij}\left(x_{p+1}^*\right) \end{aligned} \tag{4.13}$$

where $\Delta_{ij}(x_{p+1}^*)$ is the incremental divergence due to the addition of a band x_{p+1}^*, and can be calculated by the following formulae:

$$\Delta_{ij}(x^*_{p+1}) = \frac{1}{2\delta_i}\left[(\mu^*_i - \mu^*_j) - (\boldsymbol{\mu}^*_{i,p} - \boldsymbol{\mu}^*_{j,p})^T \gamma_i\right]^2 + \frac{1}{2\delta_j}\left[(\mu^*_i - \varpi^*_j) - (\boldsymbol{\mu}^*_{i,p} - \boldsymbol{\mu}^*_{j,p})^t \gamma_j\right]^2$$
$$+ \frac{1}{2} tr\left[(\boldsymbol{\Sigma}_{i,p+1} - \boldsymbol{\Sigma}_{j,p+1})(\delta_i^{-1}\gamma_i\gamma_i^t + \delta_j^{-1}\gamma_j\gamma_j^t)\right] + (\boldsymbol{z}_i^t - \boldsymbol{z}_j^t)(\delta_i^{-1}\gamma_i - \delta_j^{-1}\gamma_j) + (\sigma_i^2 - \sigma_j^2)(\delta_i - \delta_j) \tag{4.14}$$

If the covariance matrices of the two normal distributions are equal, then the incremental divergence due to the addition of a band is given by

$$\Delta_{ij}(x^*_{p+1}) = \frac{\left[(\mu^*_i - \mu^*_j) - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_p^{-1} \boldsymbol{z}\right]^2}{\sigma^2 - \boldsymbol{z}^t \boldsymbol{\Sigma}_p^{-1} \boldsymbol{z}} \tag{4.15}$$

Equation (4.11) gives an efficient way to calculate the divergence with the additional band. When a new band is to be considered, it is not necessary to compute the divergence of all selected bands; only the incremental divergence is calculated.

We can extend the above idea to the band selection of hyperspectral images with multi-classes. Assume that we have L multiple classes in hyperspectral images, and want to select d spectral bands out of K bands. Then, we can define the divergence of a specific band (say λ_q) as the sum of $L(L\text{-}1)/2$ pairwise combinations of $D_{ij}(\boldsymbol{x}_q)$. That is,

$$D(\lambda_q) = \sum_{i=1}^{L-1} \sum_{i<j\le L} D_{ij}(\boldsymbol{x}_q) \tag{4.16}$$

The incremental divergence for multiple classes due to the addition of a band can be defined similarly as follows:

$$\Delta(x^*_{p+1}) = \sum_{i=1}^{L-1} \sum_{i<j\le L} \Delta_{ij}(x^*_{p+1}) \tag{4.17}$$

For multi-class problem, the transformed divergence [104], which gives an exponentially decreasing weight to increasing distances between the classes, can result a better performance. The transformed divergence is defined as

$$TD_{ij}(\boldsymbol{x}) = 2[1 - \exp(-\frac{D_{ij}(\boldsymbol{x})}{8})] \tag{4.18}$$

The procedure for an efficient band selection based on the recursive equation of divergence can be described as follows. The diagram is shown in Figure 4.1.

Algorithm 1: Recursive Divergence Based Band Selection Algorithm

Input: a set of spectral band $\boldsymbol{\Lambda}=[\lambda_1, \lambda_2,..., \lambda_K]$ and training sample set $\boldsymbol{X}$

Output: selected optimal band subset $\boldsymbol{O}$

1. *Set $\boldsymbol{O}$ to the empty set.*

2. *Exhaustively calculate divergence by Equation (4.6) for all bands in $\boldsymbol{\Lambda}$. Find the one with the maximum divergence (say λ_i) as the starting band.*

3. *Calculate divergence incremental $\Delta_{ij}(x^*_{p+1})$ according to Equation (4.12) for all the remaining bands. If $\boldsymbol{O}_p$ represents a set of p spectral bands then, the best band at a given iteration, $\boldsymbol{O}_{p+1}$ is the set which has the maximum divergence value.*

4. *Select the band having the largest divergence incremental value (say λ_k), and add it to the selected band set $\boldsymbol{O}$ and remove it from $\boldsymbol{\Lambda}$.*

5. *If stopping criterion is met, then stop and output selected band set $\boldsymbol{O}$. Otherwise go to Step 2.*

4.4 Collaborative Band Selection

A basic property of spatially located data is that the set of values are likely to be related over space. As Tobler [78] has argued, "the first law of geography: everything is related

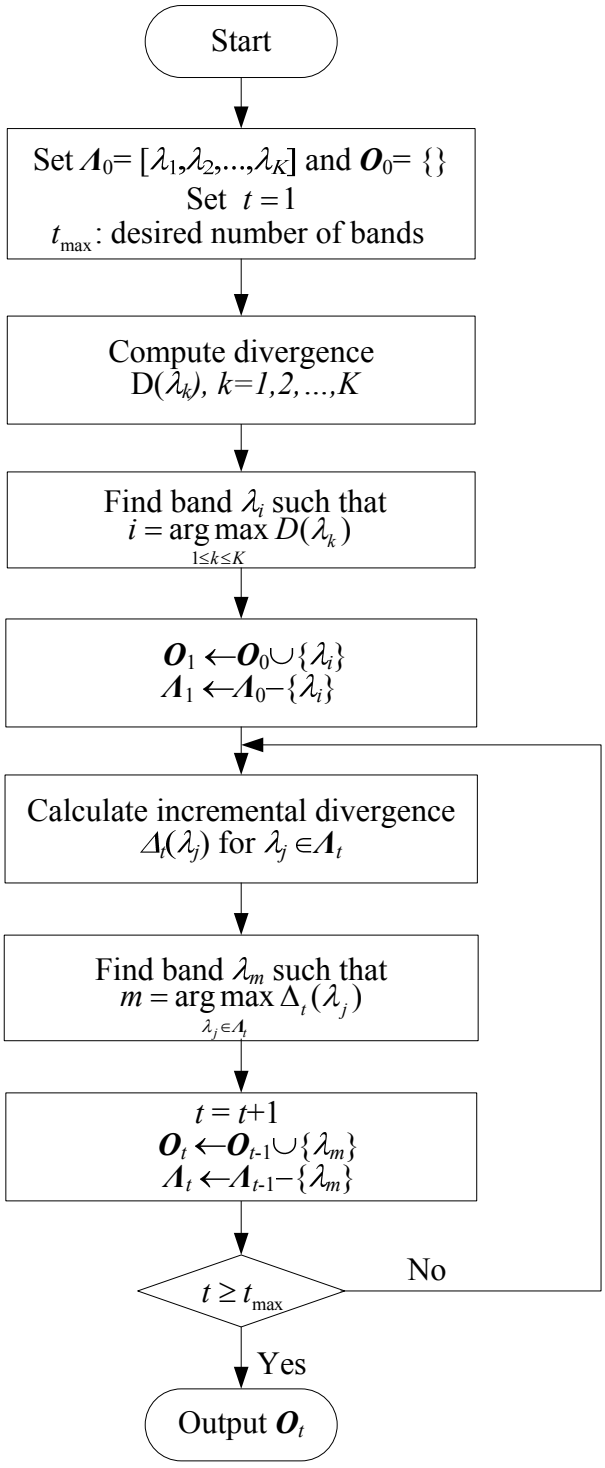


Figure 4.1: Block diagram of the recursive divergence based band selection

to everything else, but near things are more related than distant things." For hyperspectral image data, it is anticipated that there will be some degree of dependency between pixels.

In the spectral information based band selection, a good subset is to maximize the representation of the spectral separability. However, separability maximization does not guarantee a classification process that will produce the best visual result, or the most accurate. It is only when the spectral separability is spatially organized that regional variations become apparent in an image. This implies that using only the spectral separability criterion cannot guarantee obtain the most accurate results. Most hyperspectral image analysis techniques, such as enhancement and classification, are indirectly an attempt to make the spatial heterogeneity clearer. In other words, pixels after process should on average be more similar to neighboring pixels than those pixels that are far away, a characteristic known as spatial autocorrelation. The spatial autocorrelation for the image provides an excellent measure of information in the image. This suggests that to identify band combinations with the highest spatial autocorrelation should not only increase the accuracy of the spectral representation of the objects, but also increase their spatial representation and suppress visually distracting.

4.4.1 Review of Spatial Autocorrelation

Spatial autocorrelation is defined one observation tends to take value (reflectance value of a pixel in hyperspectral image) that is related to those of neighboring observations (surrounding pixels). Spatial autocorrelation is useful since it not only considers the value of the pixel (magnitude of reflectance), but also the relationship between that pixel and its neighboring pixels.

To measure the spatial autocorrelation, some spatial statistics have been developed. These include the Moran *I*, Geary *c* ([67]), *G* statistics ([73]), and LISA ([68]). There are two common aspects for all those spatial analytical techniques. First, they begin from the

randomized distribution assumption of spatial pattern. Second, the spatial pattern is derived from the data only.

There exist two types of measures: global measures, which provide a single value that summarize the level of spatial autocorrelation with respect to the whole region, and local measures, which provide a value for each location with respect to its neighborhood. In the following subsections we briefly review some popular measures of both global and local spatial association.

A. Global spatial autocorrelation

To measure the global measure of spatial autocorrelation, the Moran I is defined by

$$I = \frac{n}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \tag{4.19}$$

where x_i denotes the observed value at location i, $\bar{x}$ is the mean value of the x_i over the n locations. w_{ij} is the (i, j)-th element of a spatial weights matrix. Here we consider symmetric binary weights, with ones if location j is contiguous to location i, and zero otherwise.

As a different approach to measuring spatial association, Geary's c statistic is defined as the following:

$$c = \frac{(n-1)}{2n\sum_{i=1}^{n}(x_i - \bar{x})^2} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \tag{4.20}$$

The Geary statistic is always positive and asymptotically normal. When the Geary value is 1, which means there is no spatial autocorrelation. A low value (between 0 and 1)

indicates a positive spatial autocorrelation while a high value (greater than 1) indicates a negative spatial autocorrelation [64].

B. Local spatial autocorrelation

Global measures of spatial autocorrelation emphasize the average spatial dependence over a study area. These measures have also shown success in quantifying global noise present in image. However, the global measure is useful when spatial dependence is relatively uniform over the study area. If the underlying spatial process is not stationary, global measures may not be representative. In this case, the global measures which generate only a sole average measure of spatial dependency tend to obscure, or suppress, any significant local variation of spatial nonstationarity in the data set. Such a global estimate will be quite uninformative and may be very misleading. Therefore, to measure the spatial dependence at a smaller area is often more appropriate. Local versions of these global measures are gaining attention for their ability to identify the location and spatial dependence within the study area.

To overcome these limitations, local indicators of spatial association (LISA) have been developed [68]. In contrast to existing methods, LISA measures focus on local variations within patterns of spatial dependence. Thus, they have the potential to uncover discrete spatial regimes which might be missed by using existing global techniques. In calculating local spatial association measures on raster data, each cell or pixel receives a value quantifying its spatial dependence to its neighbors, where the neighborhood is determined by the weights matrix.

There have been different proposals for local measures, but two in particular are worth mentioning since they are related to the previous global measures of spatial autocorrelation. Local Moran and local Geary statistics are alternative local indicators.

The local Moran statistic for each observation i is defined as

$$I_i = \frac{(x_i - \bar{x})\sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 / n} \tag{4.21}$$

for any $i = 1, \ldots, n$. Local Moran's I_i measures joined covariation of neighboring localities.

A local Geary statistic for each observation i may be defined as follows

$$c_i = \frac{\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 / n} \tag{4.22}$$

The local Moran allows for the identification of spatial agglomerative patterns, while the local Geary allows for the identification of spatial patterns of similarity or dissimilarity. One advantage of the local Moran and the local Geary is that they can be associated with the global statistics (Moran *I* and Geary *c*) and can be used to estimate the contribution of individual statistics to the corresponding global statistics. In calculating local spatial association measures on image data, each pixel receives a value quantifying its spatial dependence to its neighbors, where the neighborhood is determined by the weights matrix.

4.4.2 Multidimensional Local Spatial Autocorrelation (MLSA)

The measures in previous discusses are all for one-dimensional variable. While the pixels in hyperspectral image are usually multi-dimensional vectors, the classical local Geary statistic cannot be directly applied to hyperspectral image data.

Three methods are proposes here to extend the single-image based local Geary measure to high dimensional data. Let us now consider a hyperspectral image $\boldsymbol{H}$, defined on the *K* dimensional space, where *K* is the number of spectral channels. We use $\boldsymbol{x}_i$ to denote the observed value at location *i*. But now $\boldsymbol{x}_i = (x_{i1}, \cdots x_{ik}, \cdots, x_{iK})^t$, which is represented as a

column vector. k denotes the dimension of the data or the spectral band of the hyperspectral image.

A. Average Approach

The basic idea of this method is to calculate the local Geary's c for each dimension first, and then take an average for the c value among all dimensions. The local Geary's c for K-dimensional data is defined as:

$$c_i^A = \frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2] \tag{4.23}$$

where, the superscript A in c_i^A indicates average, and $m_{2k} = \sum_i (x_{ik}-\bar{x}_k)^2 / n$. We derive the moments for c_i^A under the null hypothesis of no spatial association.

Lemma 1: For hyperspectral image ***H***, if each pixel is *i.i.d*, and spectral image λ_k is independent, then

(i) Expected value for c_i^A is $\frac{2nw_i}{n-1}$,

(ii) Variance for c_i^A is $\frac{1}{K}[\frac{n(w_i^2+w_{i(2)})(3+b_2)}{n-1} - \frac{2nw_i}{(n-1)^2}]$.

Proof of result (i):

$$E(c_i^A) = E\left\{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]\right\}$$

$$= \frac{1}{K}E\left\{\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]\right\}$$

$$= \frac{1}{K}\left\{E[\frac{1}{m_{21}}\sum_j w_{ij}(x_{i1}-x_{j1})^2]+\ldots+E[\frac{1}{m_{2K}}\sum_j w_{ij}(x_{iK}-x_{jK})^2]\right\} \tag{4.24}$$

where $E[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]$ is the expected value for a spectral image. According to [68],

$$E[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]=\frac{2nw_i}{n-1} \tag{4.25}$$

where $w_i=\sum_j w_{ij}$.

Substitute (4.25) to (4.24), we have

$$\begin{aligned} E(c_i^A) &= E\left\{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]\right\} \\ &= \frac{1}{K}\sum_{k=1}^{K}E\left\{[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]\right\} \\ &= \frac{1}{K}\sum_{k=1}^{K}\frac{2nw_i}{n-1}=\frac{2nw_i}{n-1} \end{aligned} \tag{4.26}$$

□

***Proof of result (ii)*:**

$$\begin{aligned} Var(c_i^A) &= Var\left\{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]\right\} \\ &= \frac{1}{K^2}Var\left\{\frac{1}{m_{21}}\sum_j w_{ij}(x_{i1}-x_{j1})^2+\ldots+\frac{1}{m_{2K}}\sum_j w_{ij}(x_{iK}-x_{jK})^2\right\} \end{aligned} \tag{4.27}$$

If all spectral images are independent, we have $\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2$ are independent, then variance can be written as

$$Var(c_i^A)=Var\left\{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik}-x_{jk})^2]\right\} \tag{4.28}$$

$$=\frac{1}{K^2}Var\left\{\frac{1}{m_{21}}\sum_j w_{ij}(x_{i1}-x_{j1})^2+\ldots+\frac{1}{m_{2K}}\sum_j w_{ij}(x_{iK}-x_{jK})^2\right\}$$

$$= \frac{1}{K^2}\left\{ Var\left[\frac{1}{m_{21}} \sum_j w_{ij}(x_{i1} - x_{j1})^2 \right] + \ldots + Var\left[\frac{1}{m_{2K}} \sum_j w_{ij}(x_{iK} - x_{jK})^2 \right] \right\}$$

where $Var\left[\frac{1}{m_{2k}} \sum_j w_{ij}(x_{ik} - x_{jk})^2 \right]$ is the variance for a spectral image. According to [68],

$$Var\left[\frac{1}{m_{2k}} \sum_j w_{ij}(x_{ik} - x_{jk})^2 \right] = \frac{n(w_i^2 + w_{i(2)})(3 + b_2)}{n-1} - \frac{2nw_i}{(n-1)^2} \qquad (4.29)$$

where $w_{i(2)} = \sum_{j \neq i} w_{ij}^2$, $b_2 = m_4 / m_2$, $m_2 = \sum_i (x_i - \bar{x})^2 / n$ and $m_4 = \sum_i (x_i - \bar{x})^4 / n$. After substituting (4.29) to (4.28), we have

$$\begin{aligned} Var(c_i^A) &= Var\{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik} - x_{jk})^2]\} \\ &= \frac{1}{K^2} Var\{\frac{1}{m_{21}}\sum_j w_{ij}(x_{i1} - x_{j1})^2 + \ldots + \frac{1}{m_{2K}}\sum_j w_{ij}(x_{iK} - x_{jK})^2\} \\ &= \frac{1}{K}[\frac{n(w_i^2 + w_{i(2)})(3 + b_2)}{n-1} - \frac{2nw_i}{(n-1)^2}] \end{aligned} \qquad (4.30)$$

Otherwise, if spectral image λ_k are not independent, the variance can be expressed as

$$\begin{aligned} Var(c_i^A) &= Var\{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{m_{2k}}\sum_j w_{ij}(x_{ik} - x_{jk})^2]\} \\ &= \frac{1}{K^2}\sum_{l}^{K}\sum_{n}^{K} Cov(\frac{1}{m_{2l}}\sum_j w_{ij}(x_{il} - x_{jl})^2, \frac{1}{m_{2n}}\sum_j w_{ij}(x_{in} - x_{jn})^2) \end{aligned} \qquad (4.31)$$

□

B. Mean Approach

For this method, we will calculate the mean for the *K*-dimensional data first. This process can convert the *K*-dimensional data to a one-dimensional data. And then the classic local

Geary's c is applied for this one-dimensional data. The local Geary's c for this case is defined as:

$$c_i^M = \frac{1}{\bar{m}_2}\sum_j w_{ij}(\bar{x}_i - \bar{x}_j)^2 \tag{4.32}$$

where the superscript M in c_i^M indicates mean, and $\bar{x}_i = \frac{1}{K}\sum_{k=1}^{K} x_{ik}$, $\bar{x}_j = \frac{1}{K}\sum_{k=1}^{K} x_{jk}$, and $\bar{m}_2 = \frac{1}{n}\sum_i \bar{x}_i^2$ We also derive the moments for c_i^M under the null hypothesis of no spatial association.

Lemma 2: For hyperspectral image ***H***, if each pixel is *i.i.d*, then

(i) Expected value for c_i^M is $\frac{2nw_i}{n-1}$,

(ii) Variance for c_i^M is $\frac{n(w_i^2 + w_{i(2)})(3+b_2)}{n-1} - \frac{2nw_i}{(n-1)^2}$

***Proof of result (i)*:**

Since $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ are independent, $\bar{x}_i = \frac{1}{K}\sum_{k=1}^{K} x_{ik}$ and $\bar{x}_j = \frac{1}{K}\sum_{k=1}^{K} x_{jk}$ are also independent. According to [68],

$$E(c_i^M) = E\left\{\frac{1}{\bar{m}_2}\sum_j w_{ij}(\bar{x}_i - \bar{x}_j)^2\right\} = \frac{2nw_i}{n-1} \tag{4.33}$$

□

***Proof of result (ii)*:**

Similar to (i), according to [68],

$$Var(c_i^M) = Var\left\{\frac{1}{\bar{m}_2}\sum_j w_{ij}(\bar{x}_i - \bar{x}_j)^2\right\} = \frac{n(w_i^2 + w_{i(2)})(3+b_2)}{n-1} - \frac{2nw_i}{(n-1)^2} \tag{4.34}$$

□

C. Vector Approach

The local Geary statistic for observation *i* is defined as [68]

$$c_i = \frac{1}{m_2}\sum_j w_{ij}(x_i - x_j)^2 \tag{4.35}$$

This definition can be re-written as

$$c_i = \frac{1}{m_2}\sum_j w_{ij}(x_i - x_j)^2 = \sum_j w_{ij}\frac{(x_i - x_j)^2}{m_2} = \sum_j w_{ij}(x_i - x_j)m_2^{-1}(x_i - x_j) \tag{4.36}$$

Now, consider in multi-dimensional space, suppose $\boldsymbol{x}_i, \boldsymbol{x}_j \in N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,and are *i.i.d*, then

$$c_i^V = \sum_j w_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j) = \sum_j w_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j) \tag{4.37}$$

where the superscript V in c_i^V indicates vector, $\boldsymbol{\Sigma}$ is the covariance matrix for $\boldsymbol{x}$. The observed values $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ have expectations:

$$E(\boldsymbol{x}_i\boldsymbol{x}_i^T) = \mathrm{cov}(\boldsymbol{x}_i) + \boldsymbol{\mu}\boldsymbol{\mu}^T = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \tag{4.38}$$

$$\begin{aligned} E[(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T] &= E[\boldsymbol{x}_i\boldsymbol{x}_i^T - \boldsymbol{x}_i\boldsymbol{x}_j^T - \boldsymbol{x}_j\boldsymbol{x}_i^T - \boldsymbol{x}_j\boldsymbol{x}_j^T] \\ &= 2E[\boldsymbol{x}_i\boldsymbol{x}_i^T] - 2E(\boldsymbol{x}_i)E(\boldsymbol{x}_i^T) \\ &= 2\,\mathrm{cov}(\boldsymbol{x}_i) + 2\boldsymbol{\mu}\boldsymbol{\mu}^T - 2E(\boldsymbol{x}_i)E(\boldsymbol{x}_i^T) \\ &= 2\boldsymbol{\Sigma} \end{aligned} \tag{4.39}$$

The moments for c_i^V under the null hypothesis of no spatial association are derived.

Lemma 3: For hyperspectral image $\boldsymbol{H}$, if $\boldsymbol{x}_i, \boldsymbol{x}_j \in N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,and are *i.i.d*, then,

(i) Expected value for c_i^V is $2w_iK$,

(ii) Variance for c_i^V is $w_{i(2)}(4K^2 + 8K) + (w_i^2 - w_{i(2)})(4K^2 + 2K) - (2w_iK)^2$

Proof of result (i):

$$
\begin{aligned}
E(c_i^V) &= E\left[\sum_j w_{ij}(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)\right] \\
&= \sum_j w_{ij}E\left[(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)\right] \\
&= w_iE\left[(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)\right] \\
&= w_iE\left\{tr[(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)]\right\} \\
&= w_iE\left\{tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\right]\right\} \\
&= w_itr\left\{E\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\right]\right\} \\
&= w_itr\left\{\boldsymbol{\Sigma}^{-1}E\left[(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\right]\right\} \\
&= w_itr\left\{\boldsymbol{\Sigma}^{-1}(2\boldsymbol{\Sigma})\right\} \\
&= 2w_iK
\end{aligned}
\tag{4.40}
$$

where K is the dimension of the features.

□

Proof of result (ii):

The variance for c_i^V can be denoted as

$$
\mathrm{var}\left(c_i^V\right) = E\left[\left(c_i^V\right)^2\right] - E^2\left(c_i^V\right) \tag{4.41}
$$

while

$$
\begin{aligned}
E\left[\left(c_i^V\right)^2\right] = E\Bigg[&\sum_{j\neq i} w_{ij}^2(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j) \\
&+\sum_{k\neq l\neq i}\sum w_{ik}^2(\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k)(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\Bigg] \\
= w_{i(2)}&\mathrm{E}\left[(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)\right] \\
&+\left(w_i^2-w_{i(2)}\right)E\left[(\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k)(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\right]
\end{aligned}
\tag{4.42}
$$

The first term in right hand side of Equation (3.42) can be represented as [74]

$$\begin{aligned}
&\mathrm{E}\left[(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)\right] \\
&\quad = tr(\boldsymbol{\Sigma}^{-1}2\boldsymbol{\Sigma})tr(\boldsymbol{\Sigma}^{-1}2\boldsymbol{\Sigma})+2tr(\boldsymbol{\Sigma}^{-1}2\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}2\boldsymbol{\Sigma}) \\
&\quad = (2K)(2K)+2(4K)=4K^2+8K
\end{aligned} \tag{4.43}$$

For the second term in Equation (3.42),

$$\begin{aligned}
&E\left[(\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k)(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\right] \\
&\quad = \mathrm{cov}\left((\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k),(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\right) \\
&\quad + E\left[(\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k)\right]E\left[(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\right]
\end{aligned} \tag{4.44}$$

Using the following results [74],

$$\mathrm{cov}((\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k),(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l))=2K \tag{4.45}$$

and

$$\begin{aligned}
&E\left[(\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k)\right]E\left[(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\right] \\
&\quad = tr(\boldsymbol{\Sigma}^{-1}2\boldsymbol{\Sigma})tr(\boldsymbol{\Sigma}^{-1}2\boldsymbol{\Sigma})=4K^2
\end{aligned} \tag{4.46}$$

We obtain

$$E\left[(\boldsymbol{x}_i-\boldsymbol{x}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_k)(\boldsymbol{x}_i-\boldsymbol{x}_l)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_l)\right]=2K+4K^2 \tag{4.47}$$

The variance value turns out to be

$$\begin{aligned}
&\mathrm{var}\left(c_i^V\right)=E\left[\left(c_i^V\right)^2\right]-E^2\left(c_i^V\right) \\
&\quad = w_{i(2)}(4K^2+8K)+(w_i^2-w_{i(2)})(4K^2+2K)-(2w_iK)^2
\end{aligned} \tag{4.48}$$

□

The first two approaches are not recommended because there is a possibility for loss or corruption of information of the image. In this dissertation, we only use the vector

approach to calculate the local spatial autocorrelation for hyperspectral image data. This measure is named as Multidimensional Local Spatial Autocorrelation (MLSA) measure.

4.4.3 Collaborative Band Selection

In the spectral information based band selection, a good subset is to maximize the representation of the spectral separability. However, separability maximization does not guarantee a classification process that will produce the best visual result, or the most accurate. It is only when the spectral separability is spatially organized that regional variations become apparent in an image. In other words, pixels should on average be more similar to neighboring pixels than those pixels that are far away, a characteristic known as spatial autocorrelation. This suggests that using only the divergence criterion cannot guarantee obtain the most accurate results. As an alternative, the spatial information based method determines the band subset with only spatial information to make the spatial heterogeneity clearer. While in classification stage, the most available classifiers are spectral classifiers, they assign each pixel to particular class depend on the spectral similarity. This suggests that for hyperspectral classification procedure, spectral information is still in the dominant position. The spatial information is a useful supplement to increase the classification accuracy.

In this section, we developed a collaborative band selection method that combines both spectral separability measure and spatial homogeneity measure of hyperspectral band selection. The collaborative band selection algorithm consists of three major steps of computation. At the first step, divergence is used to measure class separability of data samples for each subset combination. We sort and rank the divergence value in a descendent order for the all combinations. Then the collaborative band selection algorithm selects several subsets which have the largest divergence value. Since the classifier makes decision based on the spectral similarity, hence the large divergence value usually indicates higher classification accuracy. The algorithm chooses several

band combinations with the largest divergence values can guarantee the selected subset has good classification performance.

At the second step, the algorithm attempts to improve the prediction accuracy by integrating spatial information. At this step, the multidimensional local spatial autocorrelation measure for those selected subsets in the previous step is calculated. By adopting the proposed spatial measures, it is possible to efficiently calculate the spatial autocorrelation for hyperspectral image. From the training samples, the average multidimensional local spatial autocorrelation measure for class ω_l of specific band subset (say $\boldsymbol{U}$, $\boldsymbol{U}=[\lambda_1, ..., \lambda_p]$) be calculated as:

$$C_l(\boldsymbol{U}) = \frac{1}{N_l}\sum_{i=1}^{N_l} c_i^V(\boldsymbol{U}) \tag{4.49}$$

The average multidimensional local spatial autocorrelation measure value for all training samples is

$$C(\boldsymbol{U}) = \sum_{l=1}^{L} C_l(\boldsymbol{U}) \tag{4.50}$$

The third step is to combine the spectral information (divergence value) with the spatial information (average multidimensional local spatial autocorrelation measure value). The ration between divergence and average multidimensional local spatial autocorrelation measure value is used to combine the divergence and spatial autocorrelation in this step:

$$DC(\boldsymbol{U}) = D(\boldsymbol{U}) / C(\boldsymbol{U}) \tag{4.51}$$

where $\boldsymbol{U}$ is a specific band subset. For the divergence measure, the larger value indicates more separate between two classes. While for the multidimensional local spatial autocorrelation measure, the smaller value means stronger spatial similarity. For Equation (4.51), the band subset has large class separabiliy and strong spatial similarity will yield a bigger output value. The rule to find the optimal subset at this step is defined as:

$$\boldsymbol{U}^* = \max_{\boldsymbol{U}} DC(\boldsymbol{U}) = \max_{\boldsymbol{U}}[D(\boldsymbol{U}) / C(\boldsymbol{U})] \tag{4.52}$$

The procedure for the collaborative band selection method is described as follows. The diagram is shown in Figure 4.2.

Algorithm 2. Collaborative Band Selection Algorithm

Input: a set of spectral band $\boldsymbol{\Lambda} = [\lambda_1, \lambda_2, ..., \lambda_K]$, training sample set $\boldsymbol{X}$
User defined value p

Output: selected optimal band subset $\boldsymbol{O}$

1. *Set $\boldsymbol{O}$ to the empty set.*

2. *Exhaustively calculate divergence $D(\lambda_k)$ by Equation (4.17) for all bands in $\boldsymbol{\Lambda}$. Sort and rand the $D(\lambda_k)$ in a descendent order. Find largest p divergence value and corresponding subsets.*

3. *Calculate the average multidimensional local spatial autocorrelation measure value for these p subsets.*

4. *Calculate the combination value according to Equation (4.51)*

5. *Select the band having the largest combination value. Add it to the selected band set $\boldsymbol{O}$ and remove it from $\boldsymbol{\Lambda}$.*

6. *If stopping criterion is met, then stop and output selected band set $\boldsymbol{O}$. Otherwise go to Step 2.*

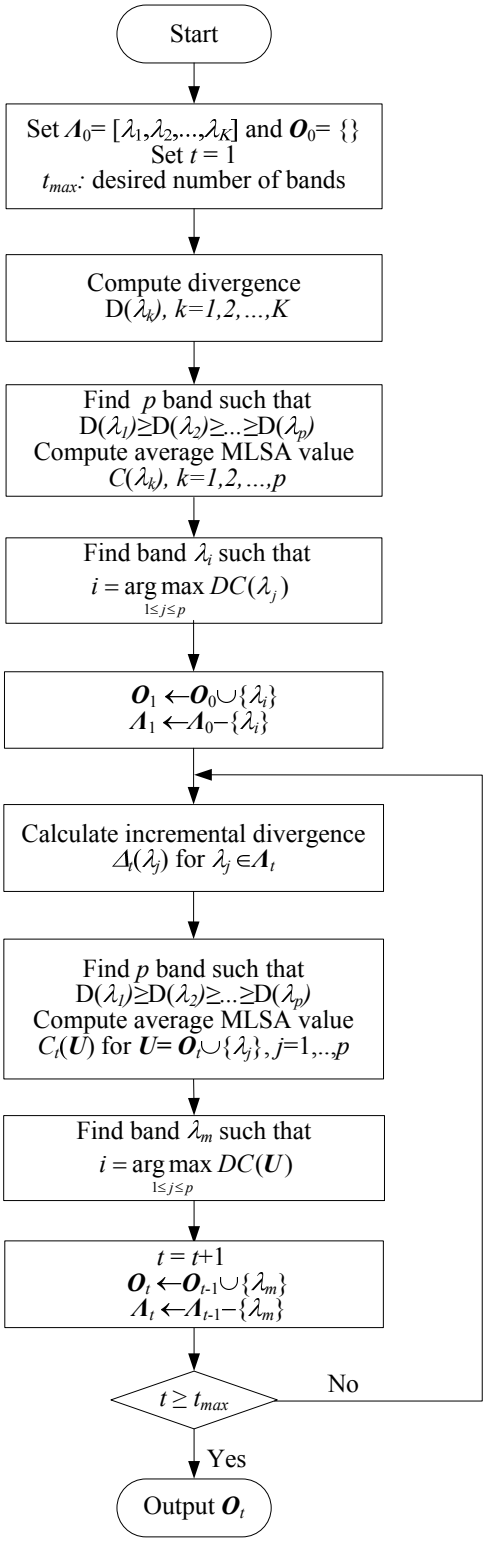


Figure 4.2: Block diagram of the collaborative band selection algorithm

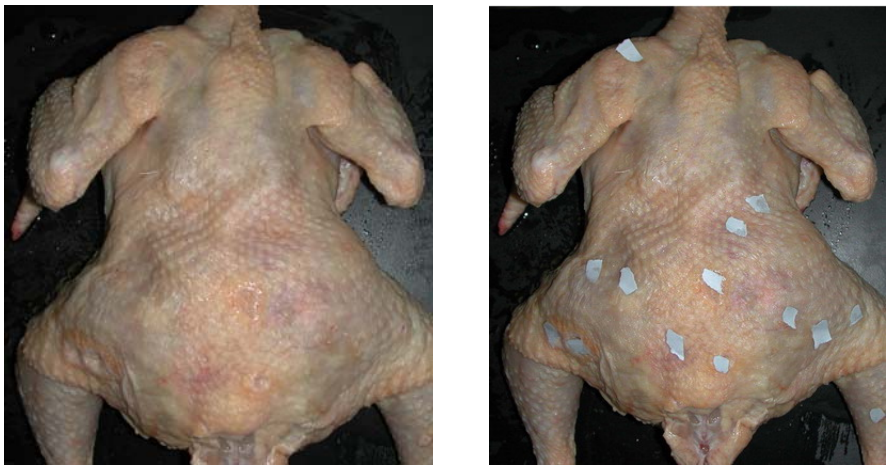
4.5 Experimental Results for Band Selection

In this section, a serious of experiments is presented to illustrate the performance of the proposed recursive divergence band selection (RD) and collaborative band selection (CBS) method. Here, the classification accuracy is got from the pixel classifier. The MLC classifier is used to get the pixel classification results. We do not choose SVM as classifier for this experiment is because the performance of SVM depends on its parameter setting. While MLC do not have this problem, so its output is fairer to evaluate the performance of band selection methods.

4.5.1 Hyperspectral Poultry Data

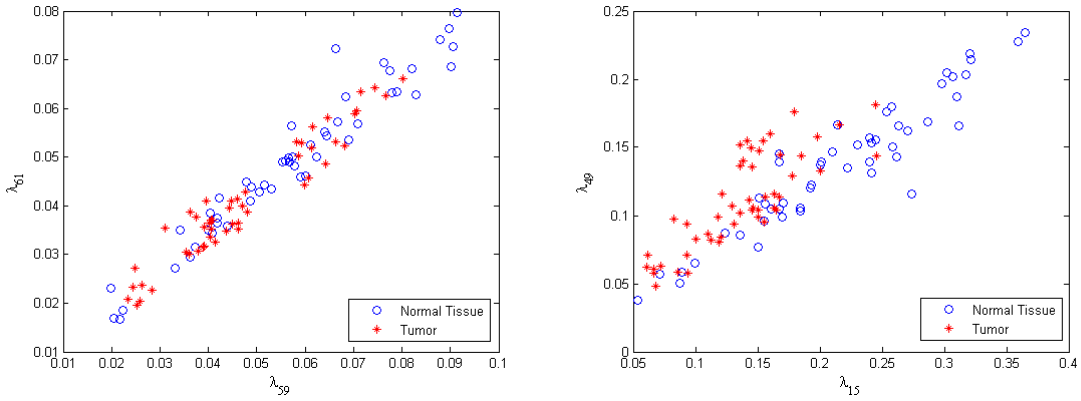
Since we have twelve HS chicken data cubes. we use one cube for training purpose. The other eleven HS cubes are used for testing. The tumors on the poultry carcass are verified and labeled by a Food Safety and Inspection Service (FSIS) veterinarian (shown in Figure 4.3). The size of this data cube is 460×400 pixels. Tumor and normal tissue pixels are collected from this data cube and all these pixels with they labels are used for band selection and classifier training purpose. We also examine the detection result for the training set to determine the other parameters, such as window size for calculate local geary value, and parameters for band selection. From the training data cube, we extract 1500 pixels from different tumor area and 5000 pixels from different normal tissue area. Twenty percent of these samples (300 tumor pxiels and 1000 normal pxiels) are used in training, and the other are used as the test set.

We first discuss the RD method. Figure 4.4 shows the band combinations with minimum and maximum divergence values for two bands case. This result is obtained from exhaustively searching all combinations with two bands. There are total 65×64/2=2080 combinations. Bands λ_{59} and λ_{61} gives minimum divergence value 0.8262 among all 2080 combinations. As shown in Figure 4.4 (a), the normal tissue and tumor pixels are highly overlapped. That means if we use these two bands for classification, we must get plenty



(a) Without labeled (b) labeled tumors

Figure 4.3: Labeled tumors



(a) Minimum divergence value: 0.8262 (b) Maximum divergence value: 10.8444

Figure 4.4: Minimum and maximum divergence value for 2-band case

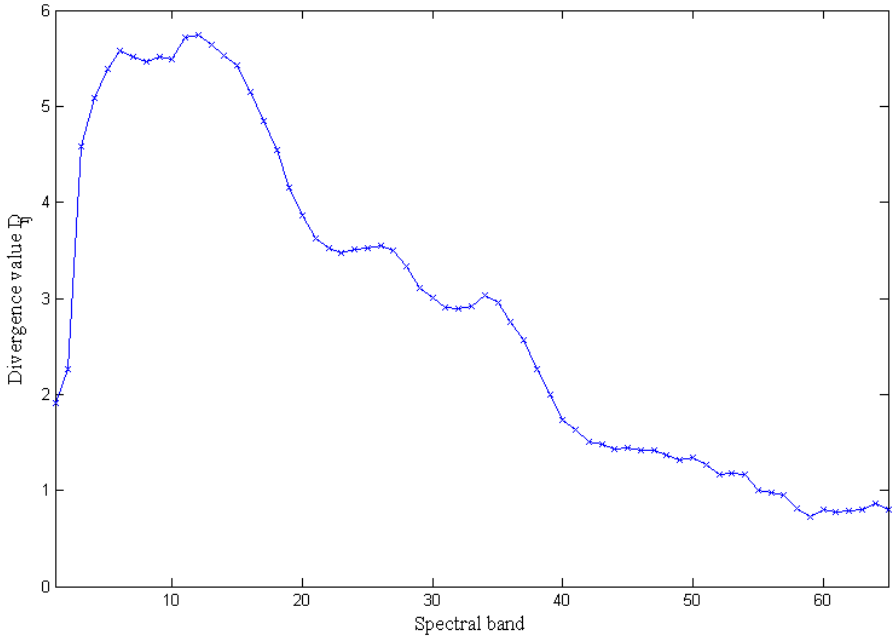
of misclassified pixels. While for bands λ_{15} and λ_{49}, which gives the largest divergence value 10.8444, the tumor and normal tissue are more separate than the previous one. Figure 4.4 illustrates the relationship between divergence value and class separability. The larger divergence value indicates more separate between two classes.

Now we study band selection for 1-band case, which is to choose the best one band from total 65 spectral bands. Figure 4.5 (a) shows the divergence value for normal tissue and tumor classes on all 65 spectral bands. The divergence has a large peak value at approximately band λ_{12} with divergence value 5.7454. If we use only the divergence value as the criterion, band λ_{12} will be chosen since it has the largest divergence value among all 65 bands. Can this band yield the best classification accuracy? Figure 4.5 (b) plots the classification accuracy corresponding with different divergence values. The maximum value of accuracy happens at band λ_9, which has divergence value 5.5088. These plots suggest that, in general, the larger divergence value usually result higher classification accuracy. Unfortunately, the largest divergence value cannot always guarantee the highest classification accuracy. This is why we need spatial information as a useful supplement to increase the classification accuracy.

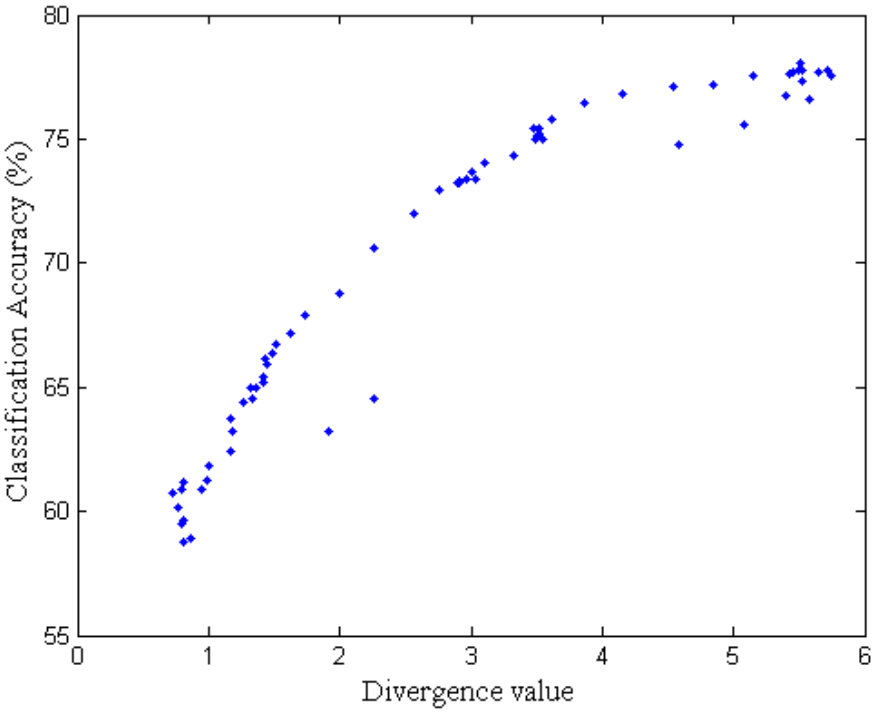
Table 4.1 lists the divergence value, average local Geary value and accuracy for ten bands. These ten bands have biggest divergence values among all 65 bands. The average local Geary values are calculated by Equation (4.50) with 3×3 window. If we apply the CBS method, and assume the parameter is set to 10, which means we consider the spatial information for the top 10 bands. Then the optimal band is λ_{15}.

Table 4.1: Divergence (D), average local Geary (C) and accuracy (Acc) for 1-band case

Band	12	11	13	6	14	7	9	10	8	15
D	5.7454	5.7120	5.6459	5.5760	5.5281	5.5165	5.5088	5.4895	5.4581	5.4327
C	0.1079	0.1130	0.1022	0.1539	0.1000	0.1411	0.1132	0.1180	0.1318	0.0965
D/C	53.247	50.548	55.243	36.231	55.281	39.096	48.664	46.521	41.412	56.297
Acc	0.7755	0.7775	0.7767	0.7663	0.7759	0.7734	0.7803	0.7778	0.7773	0.7778



(a) Divergence value for 1-band case



(b) divergence value vs. classification accuracy for testing data

Figure 4.5: Divergence value and accuracy for 1-band case

The RD method chooses λ_{12} as the optimal band for one-band case, and λ_{12} yields 77.55% classification accuracy. For CBS method, the λ_{15} is chosen as the optimal band, which has 77.78%. The highest accuracy 78.03% is obtained by λ_9, which in only 0.25% higher than that obtained by CBS. Hence for one band case, the CBS method can obtain better performance than RD method.

Next we discuss the 2-band case. In this step, two spectral bands will be selected from 65 bands. There are 65×64/2=2080 combinations in total. Figure 4.6 plots the classification accuracy corresponding with different divergence values. This figure illustrates similar phenomenon that the largest divergence value cannot always guarantee the highest classification accuracy.

Table 4.2 lists several results for 2-band case. The subset [λ_{14}, λ_{65}] gives the highest accuracy rate of 83.31%. The subset [λ_{15}, λ_{49}] has the largest divergence value among 2080 possible combinations. This subset yield accuracy of 81.69%, which is not the highest accuracy rate. The RD method chooses subset [λ_{12}, λ_{46}], which has 80.90%. The CBS method selects subset [λ_{15}, λ_{47}], and this subset can produce 81.96% accuracy. The RD and CBS method adopt sequential forward search strategy, which keep the selected bands from the previous stage, and add a new band to make the subset optimal criterion. The sequential forward search technique can only give the near-optimal solution, this results RD method get poor performance. While for CBS method, because it combines the spatial information with the spectral information, even it also adopts the sequential forward search technique, the final performance is much better than using the divergence measure only. The accuracy yielded by CBS method is only 1.35% less than the optimal accuracy. We can conclude that for two band case, the CBS method also obtains better performance than RD method.

Table 4.3 lists results for 3-band case. Similar conclusions can be found. The highest accuracy 84.27% is yielded by subset [λ_{14}, λ_{58}, λ_{65}] among all combinations of three bands, which is 65×64×63/6=43680 combinations. The RD method chooses subset [λ_{12},

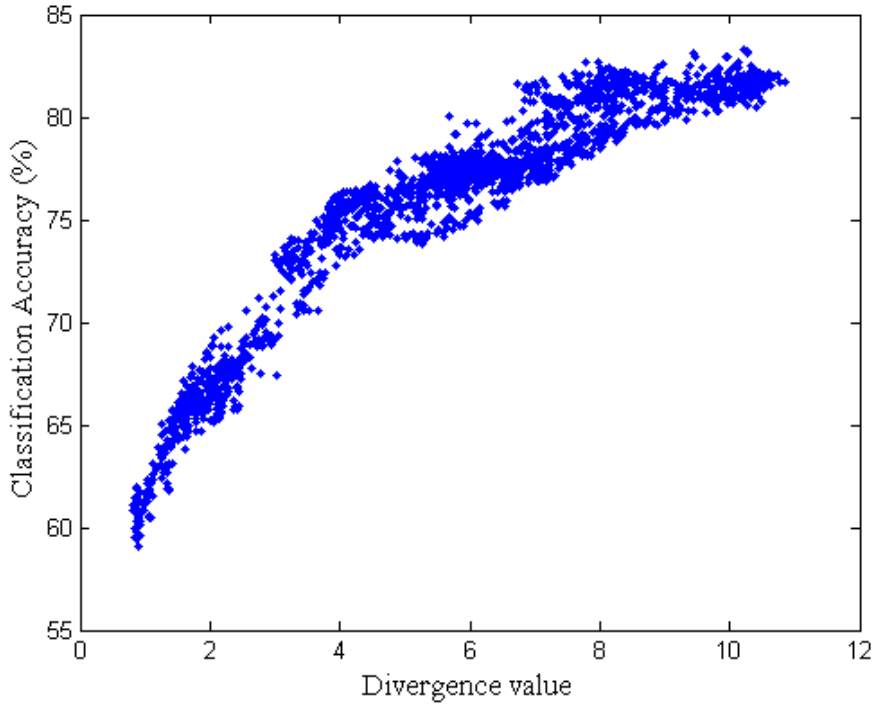


Figure 4.6: divergence value vs. classification accuracy for 2-band

Table 4.2: Divergence (*D*), average local geary (*C*) and accuracy (Acc) for 2-band case

Band		*D*	*C*	Acc
14	65	10.2135	0.3150	0.8331
15	49	10.8444	0.3338	0.8169
12	46	9.8367	0.3112	0.8090
15	47	10.5840	0.3214	0.8196

Table 4.3: Divergence (*D*), average local Geary (*C*) and accuracy (Acc) for 3-band case

Band			*D*	*C*	Acc
14	58	65	11.4371	0.7567	0.8427
12	46	2	10.3414	0.7331	0.8120
15	47	26	11.3698	0.4434	0.8329

λ_{46}, λ_2], which has 81.20% accuracy. The CBS method selects subset [λ_{15}, λ_{47}, λ_{26}], and this subset can produce 83.29% accuracy. The accuracy yielded by CBA method is very close to the optimal one. So for three band case, the CBS method still obtains better performance than RD method.

Figure 4.7 compares the classification accuracy for bands selected by RD and CBS method. For the CBS method, different p values are used. The 3×3 window is adopted to calculate local Geary measure. When p =10, the result shows the best performance. Meanwhile, the incremental of classification accuracy becomes very small after using four bandns. Hence, the first four selected bands are used. Table 4.4 lists the four optimal bands selected by RD, CBS and the exhaustive search method for the highest classification accuracy among all combinations.

Figure 4.8 displays the classification accuracy for testing set with band selection results. The overall accuracy for selected bands is shown in Figure 4.8 (a). The CBS method generates higher classification accuracy than the RD method as we expected. Exhaustive search methods provide the best performance. For the tumor accuracy (shown in Figure 4.8 (b)), the RD and CBS method all generate high accuracies. This means the bands selected by RD and CBS method are all provided good discriminant ability for tumor pixels. But for normal tissue accuracy (shown in Figure 4.8 (c)), the CBS method produces the higher accuracy than RD. This suggests that the band selected by the CBS method will generate less false positive error than RD.

4.5.2 Hyperspectral Apple Data

For this dataset, the goal is to find one or two optimal spectral bands from 79 bands for fecal detection. Each hyperspectral cube contains twelve apples as shown in Figure 4.9. The four apples on the left side are picked as training data; the other six apples are used as testing data. In the training data, the pixels from fecal and apple surface are randomly picked. Total 2000 fecal pixels and 4000 apple surface pixels are gathered for band

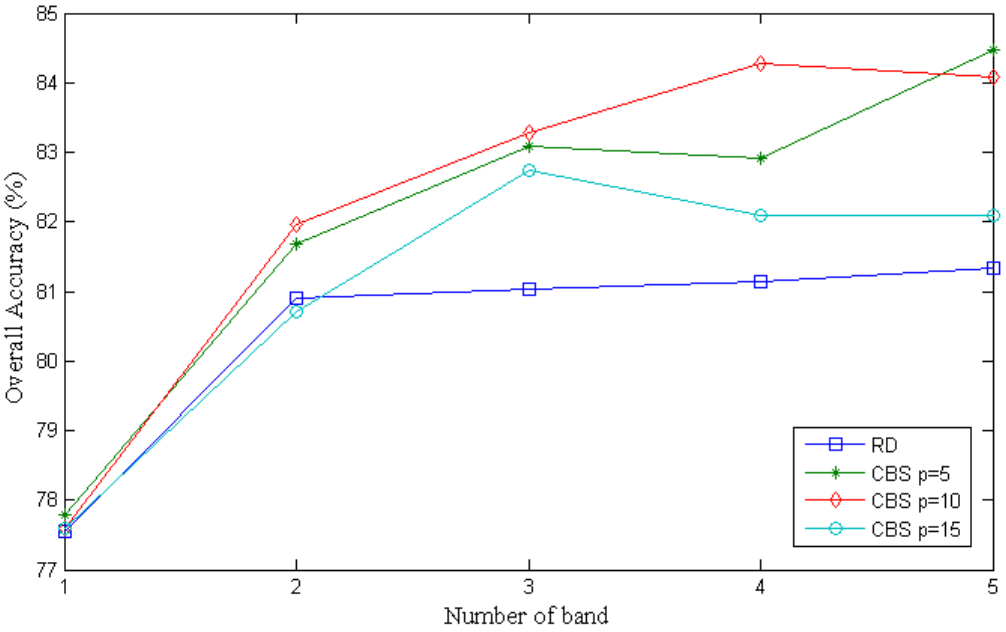
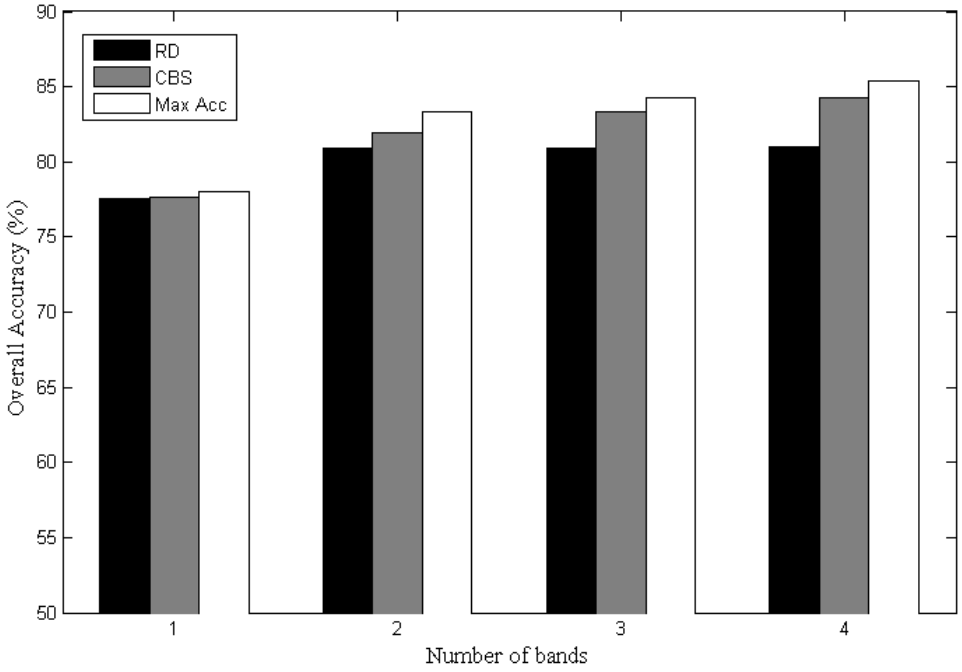


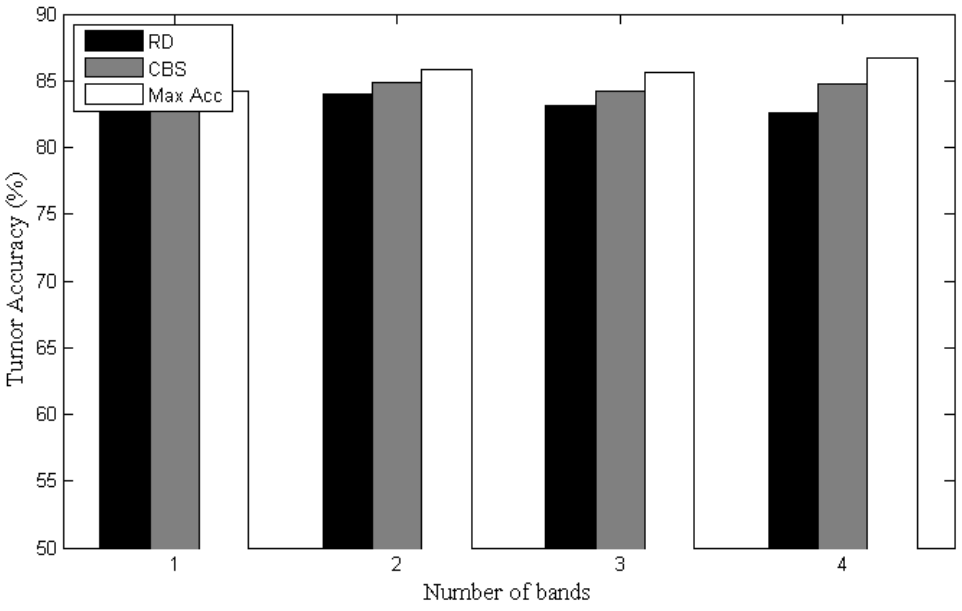
Figure 4.7: Accuracy for RD and CBS results

Table 4.4: Band selection result for chicken data

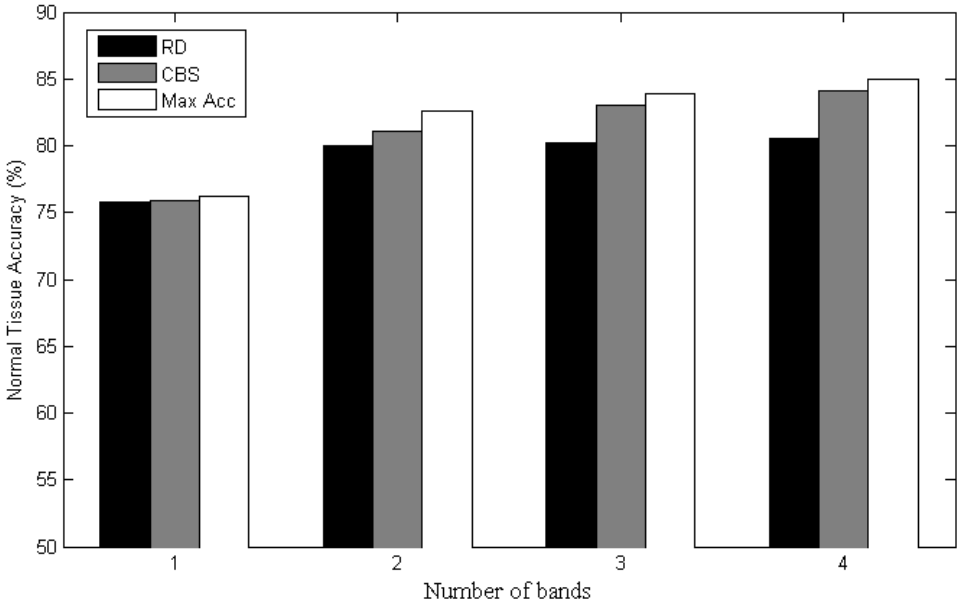
Methods	Number of selected bands			
	1	2	3	4
RD	12	12,46	12,46,2	12,46,2,1
CBS	15	15,47	15,47,26	15,47,26, 40
Max Accuracy	9	14,65	14,58,65	14,25,40,65



(a) Overall accuracy



(b) Accuracy for Tumor



(c) Accuracy for Normal Tissue

Figure 4.8: Classification accuracy on band selection results for chicken data

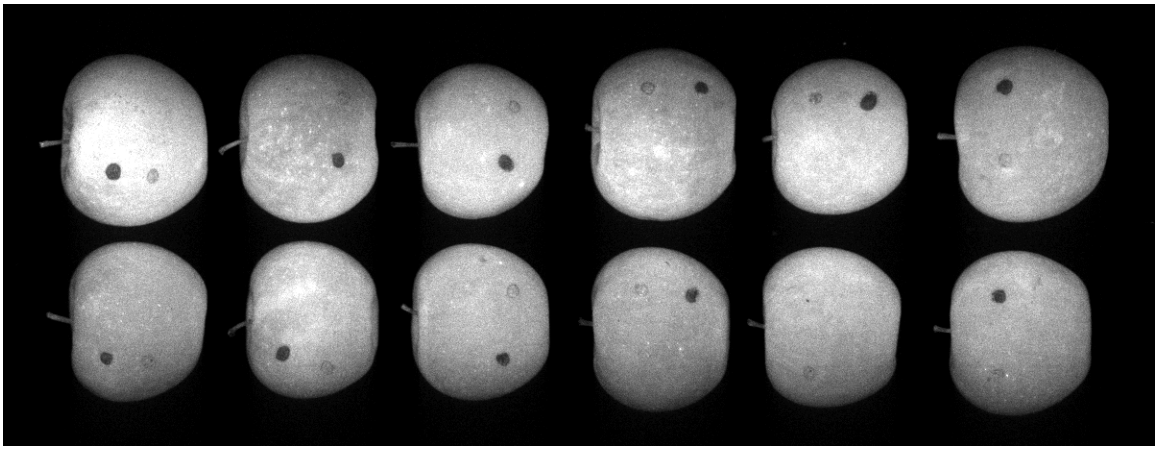


Figure 4.9: One hyperspectral cube of apple data

Table 4.5: Band selection results for apple data

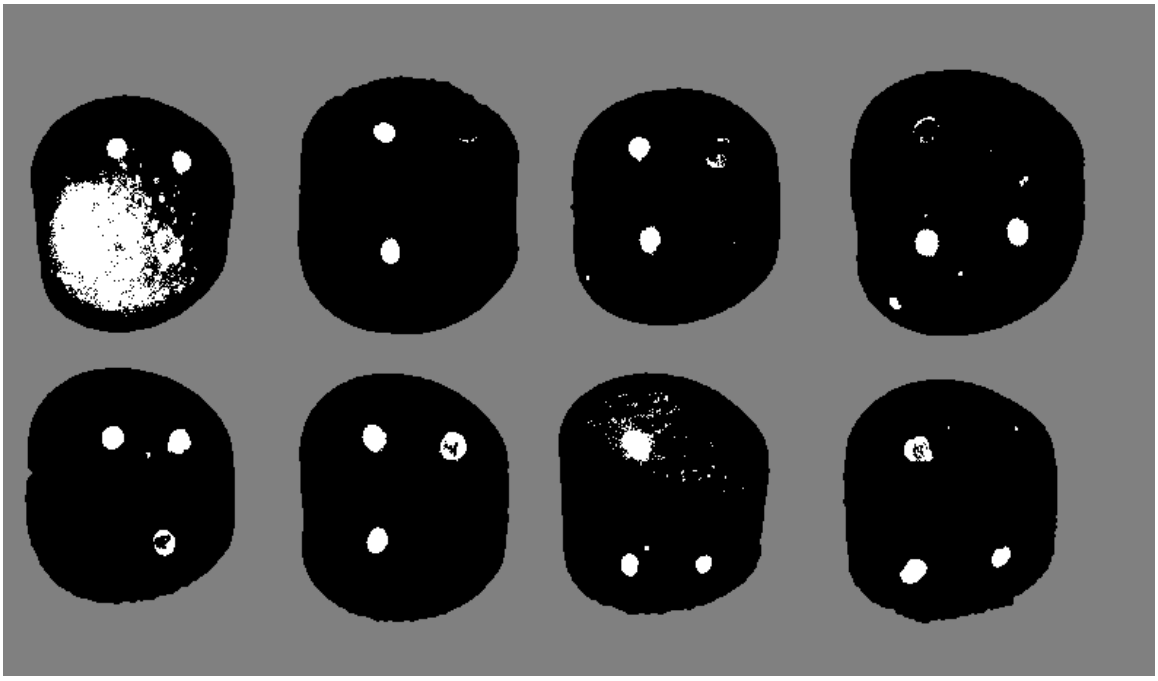
Methods	Number of selected bands	
	1	2
RD	55	55,36
CBS	54	54,38

selection and MLC classifier training. The testing data contain 4250 pixels from fecal and 160656 pixel from apple surface. The band selection results are listed in Table 4.5. For the CBS method, we use 3×3 window with p=10.

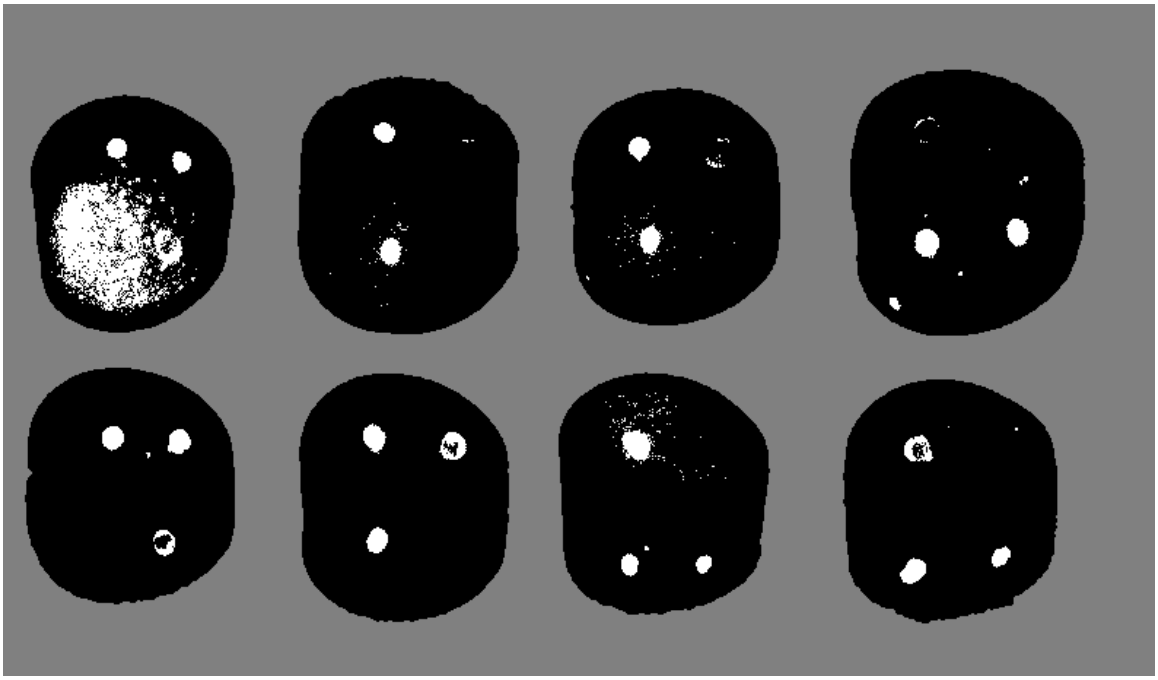
The classification results with 1 band selected by RD and CBS method are displayed in Figure 4.10 (a) and (b), respectively. The white spots indicate fecal, black area is apple surface and grey color denotes background. From Table 4.5 we known, for 1-band case, the RD and CBS method obtain very similar results. RD method selects λ_{55}, while CBS chooses λ_{54}. From the classification results shown in Figure 4.10, it is clear that with only spectral band, the detection accuracy is not satisfying. Several feces are missed and in addition, lots of apple surface are misclassified as fecal.

Using two spectral bands can greatly improve the performance. As shown in Figure 4.11, all feces are correctly detected. Although there are some apple surface pixels are misclassified as fecal, this kind of error can be easily removed through spatial filtering. Table 4.6 lists the pixel classification results. Both methods can correctly detect most of the fecal pixels. But the bands selected by CBS method generates less false positive error, i.e., apple surface pixels are misclassified as fetal.

As a comparison, Figure 4.12 and Figure 4.13 shows the ratio images of two selected band and the binary images of automated histogram-threshold method proposed in [43]. The fecal is indicated with white spots. In this case, both RD and CBS method correctly identified 100% of feces. Compared with the band selection methods proposed in [43]

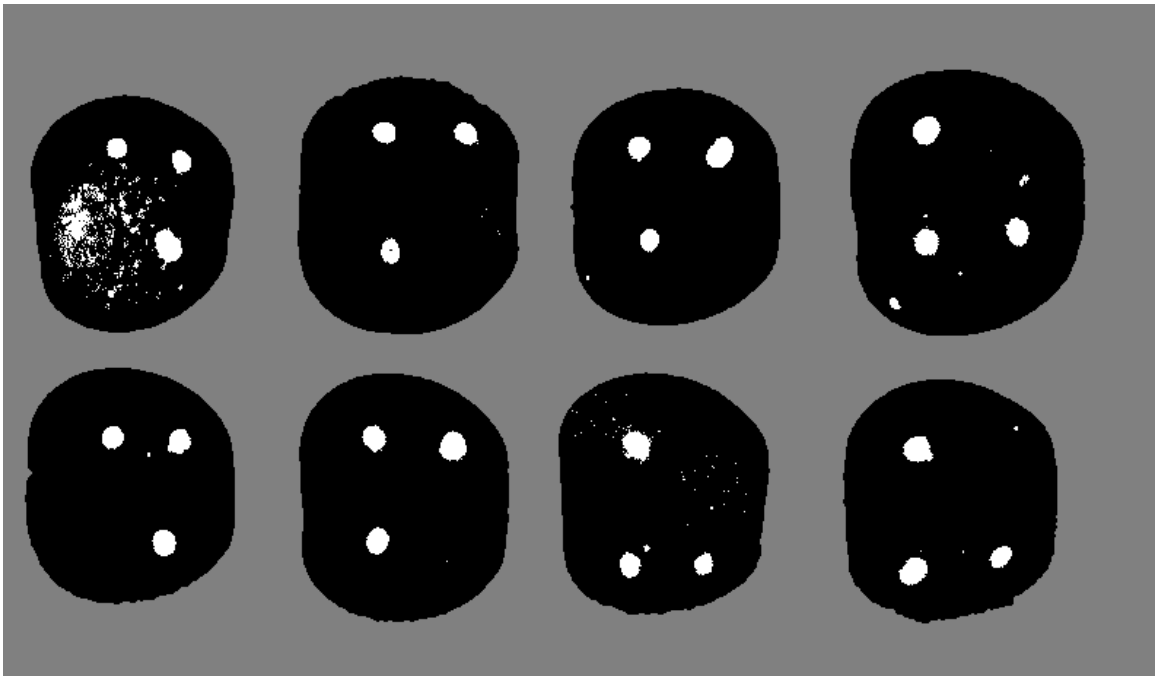


(a) Classification result for λ_{55}

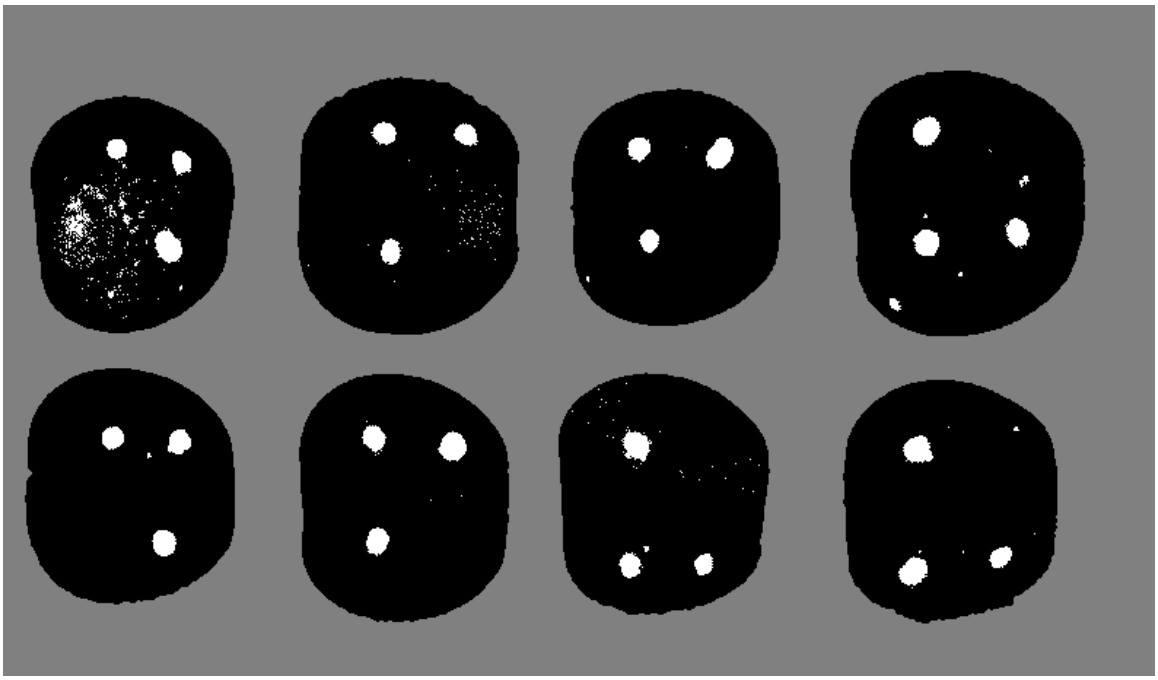


(b) Classification result for λ_{54}

Figure 4.10: Classification results with 1 band on apple data

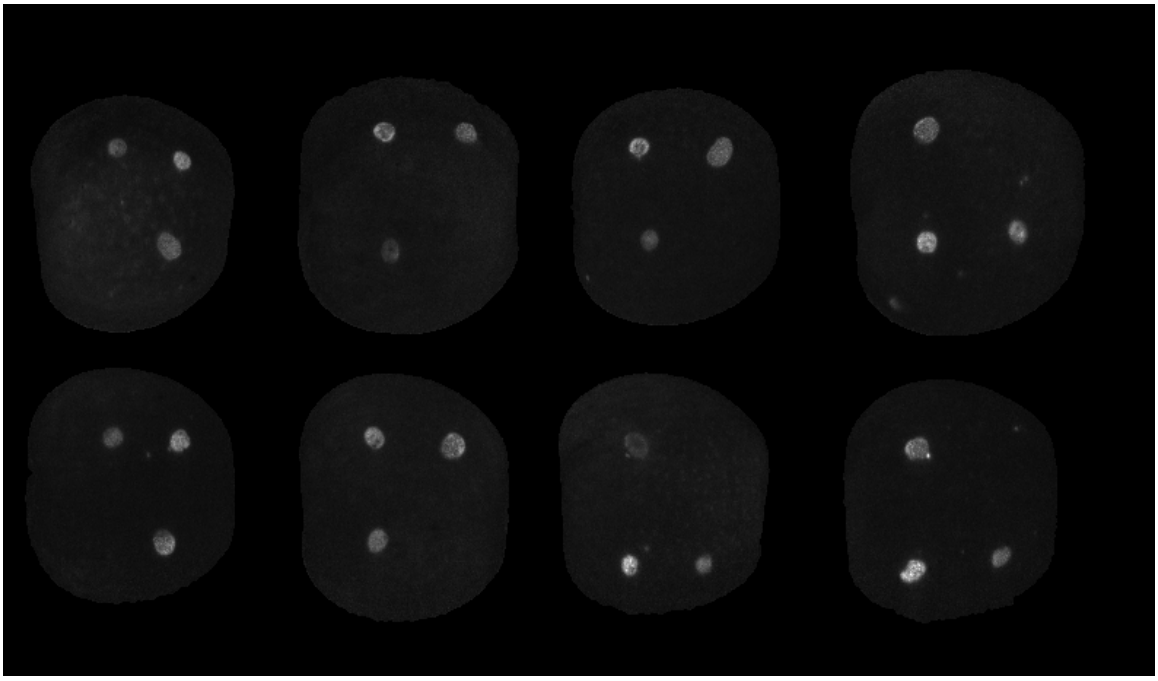


(a) Classification result for [λ_{55}, λ_{36}]

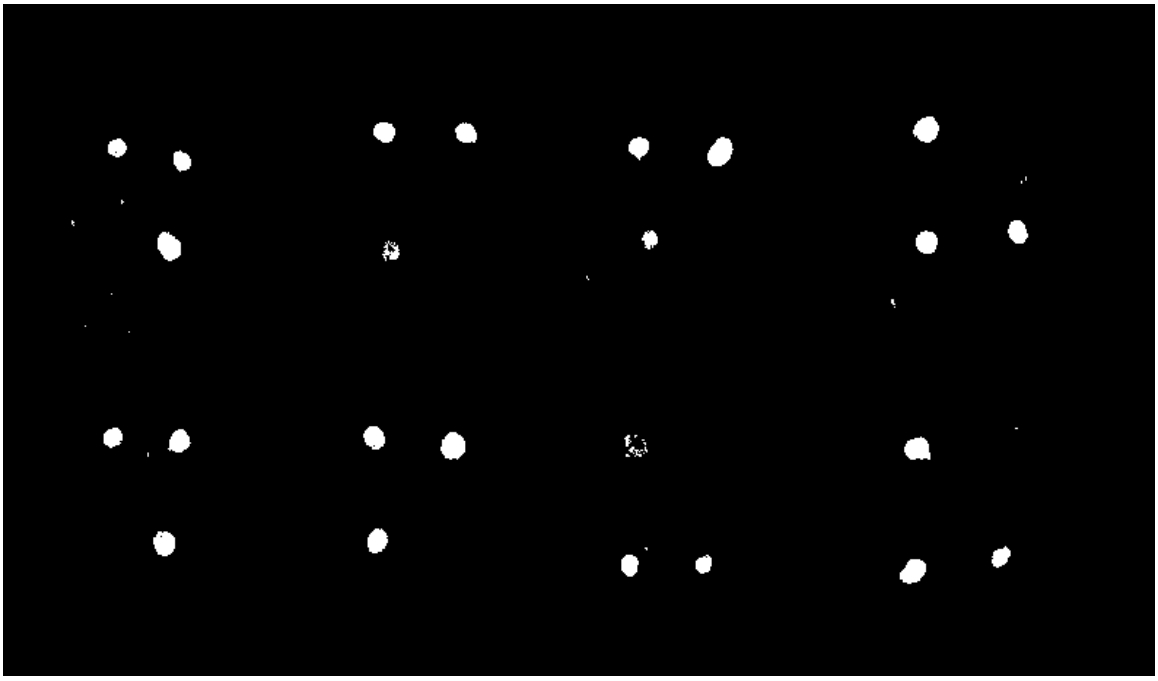


(b): Classification result for [λ_{54}, λ_{38}]

Figure 4.11: Classification results with 2 bands on apple data

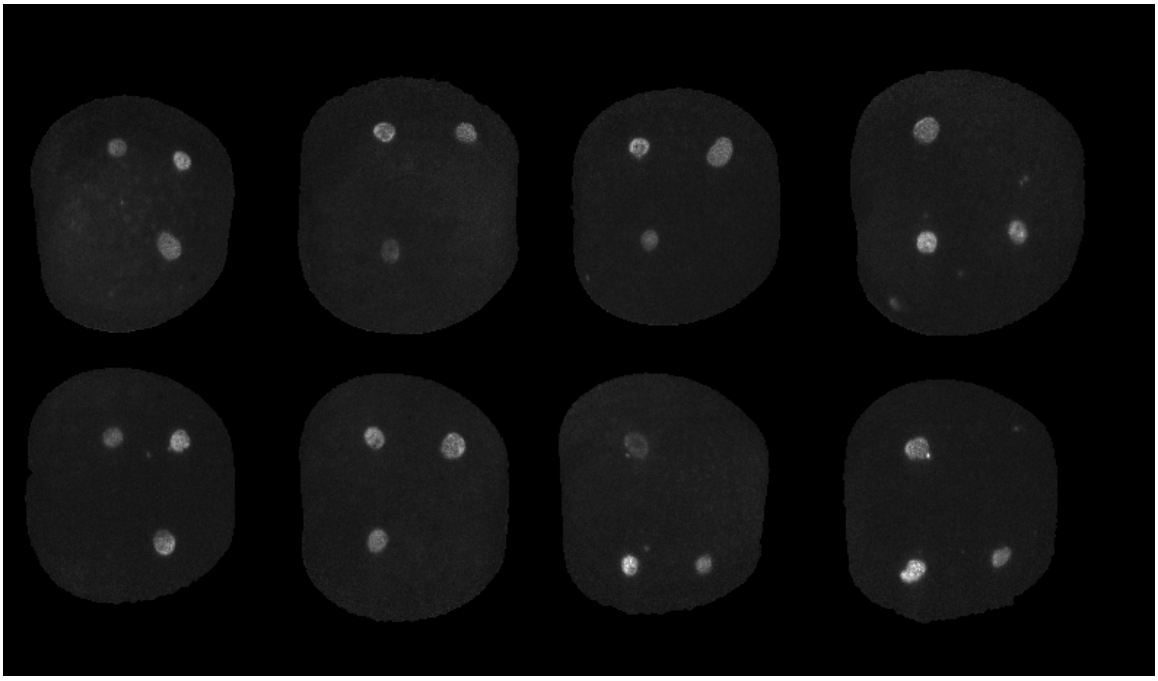


(a) $\lambda_{55}/\lambda_{36}$ image

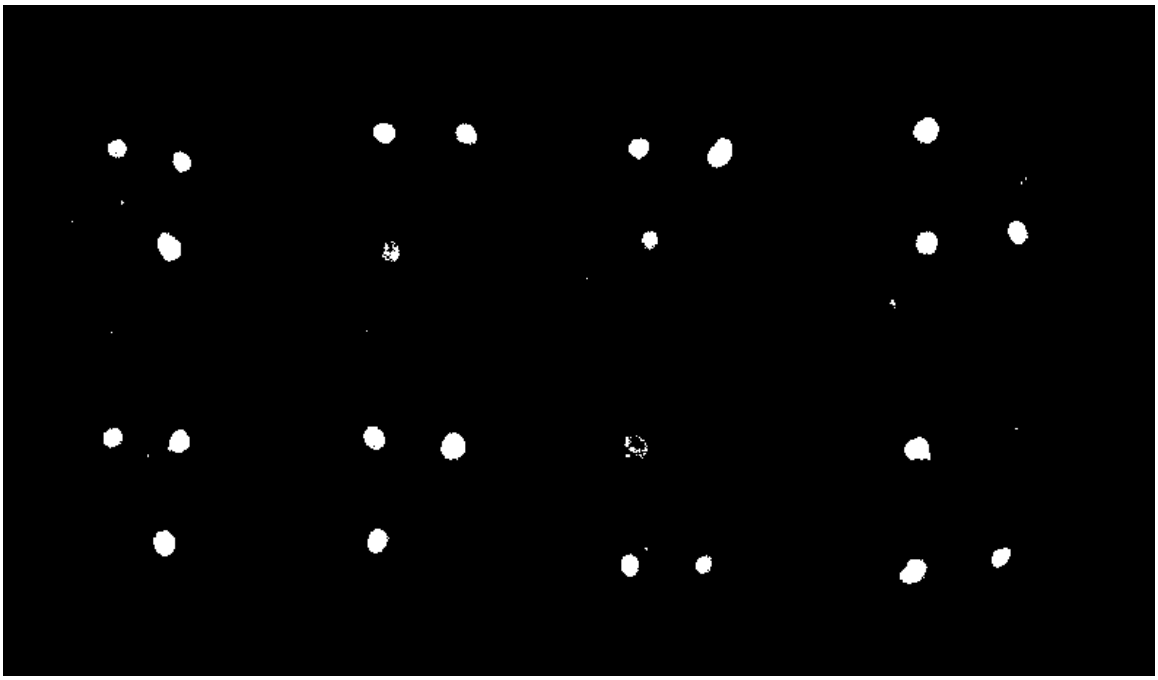


(b): Threshold result for [λ_{55}, λ_{36}]

Figure 4.12: Threshold results for [λ_{55}, λ_{36}] on apple data



(a) $\lambda_{54}/\lambda_{38}$ image



(b): Threshold result for [λ_{54}, λ_{38}]

Figure 4.13: Threshold results for [λ_{54}, λ_{38}] on apple data

Table 4.6: Pixel classification results on apple data

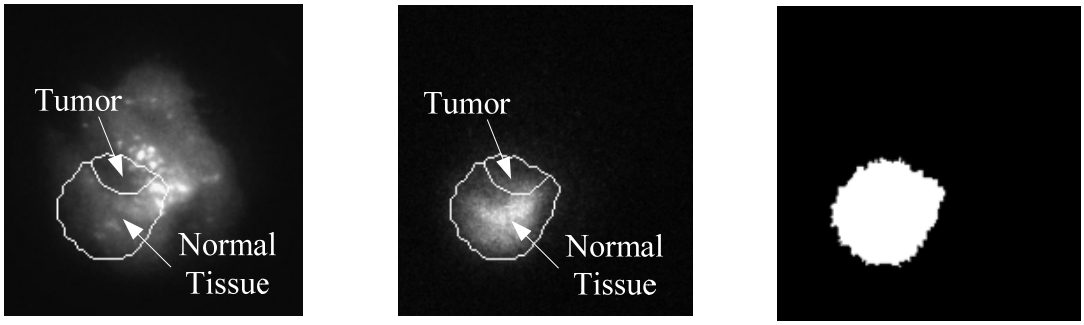
	Fecal	Apple surface
$[\lambda_{55}, \lambda_{36}]$(RD)	4244/4250	158371/160656
$[\lambda_{54}, \lambda_{38}]$(CBS)	4245/4250	158689/160656

[44], our proposed methods are more computational efficient and has a stronger mathematic background.

Compared to the chicken tumor detection, to detect fecal on apple surface is relatively easier. So the RD and CBS yield very similar results. Both of the methods can generate high accuracy. If compared in pixel level, CBS is a litter better than RD for less false positive errors.

4.5.3 Hyperspectral Mouse Data

The hyperspectral data used in the section is a hyperspectral fluorescence dataset taken on mouse skin by the Oak Ridge National Laboratory. This hyperspectral dataset consists of, on average, 165×172 pixels with 21 spectral bands. The spectral band has a discrete value from the wavelength λ_1 (440 nm) to λ_{21} (640 nm) with 10 nm intervals in the spectral region. The hyperspectral mouse data contains two classes of tumor and normal tissue. Fluorescence intensity of the background pixels is almost same. Before performing the classification, a k-mean clustering algorithm is applied on the image to remove the background and to acquire the region of interest (ROI) shown in Figure 4.14 (c). Also, depending on the reflection image (Figure 4.14(a)), the ground truth can be mapped, which is displayed in Figure 4.14(b), where the black area in the Figure 4.14 (b) indicates the background. We can see the lower left part is normal tissue (U-shaped, bright area in the fluorescence image), and the upper part is the tumor. In order to perform band selection, training samples are randomly selected for the two classes, other than the background. The number of training and testing pixels used are tabulated in Table 4.7.



(a) Reflection Image (b) 490nm Fluorescence Image (c) ROI

Figure 4.14: Reference images

Table 4.7: Mouse data used for training and testing

Class Name	Training	Testing
Normal Tissue	200	2036
Tumor	200	517

To evaluate the performance of the proposed band selection methods, we still conduct the exhaustive search method to find the optimal band subset for one to three-band cases. Then we compare the classification accuracy for different selected bands.

Table 4.8 lists the one spectral band selected by RD, CBS and exhaustive search method. The band λ_{11} gives the largest divergence value among all 21 bands with accuracy of 62.55%. But the highest classification accuracy 62.59% is yielded by band λ_8, which is also the band selected by the CBS method. For one band case, the advantage of CBS method is not significant. Next, we discuss the choosing two bands case.

Table 4.9 lists the band selection result for 2-band case by RD, CBS and exhaustive search method, respectively. Through exhaustively searching all 2-band combinations, which is 21×20/2=210, the subset [λ_6, λ_{16}] yields the highest accuracy of 82.82%. The subset [λ_8, λ_{13}] has the largest divergence value and accuracy of 82.67%. This subset is also the output of the CBS method. This accuracy is very close to the optimal one. For RD method, because of the limitation of the sequential forward searching, it selects a

Table 4.8: Divergence (D), average local geary (C) and accuracy (Acc) for 1-band case

Band	*D*	*C*	Acc
11	7.5626	0.0189	0.6255
8	6.8867	0.0148	0.6259

Table 4.9: Divergence (*D*), average local geary (*C*) and accuracy (Acc) for 2-band case

Band		*D*	*C*	Acc
6	16	16.6747	0.0552	0.8282
8	13	22.5978	0.0483	0.8267
11	1	19.5453	0.0539	0.7560

subset [λ_1, λ_{11}] with 75.60% accuracy. For two-band case, the CBS method shows greater improvement than the RD method.

The results for 3-band case are listed in
Table 4.10. The optimal accuracy is 90.76% by subset [λ_1, λ_8, λ_{19}]. The RD output [λ_1, λ_6, λ_{11}] results 88.42% accuracy. While CBS method selects subset [λ_1, λ_8, λ_{13}], which yields accuracy of 88.81%. Still, the CBS method is better.

Figure 4.15 displays the overall accuracy for testing set with band selection results. The CBS method generates higher classification accuracy than the RD method as we expected. The accuracy has a big jump from 1 band to 3 bands, the incremental is about 50%. Then the accuracy still increases but not that dramatically.

4.5.4 Indiana Pine Data

The previous two hyperspectral datasets contain only two classes. In this experiment, we will test the proposed band selection for multiple-class situation. The Indiana pine data used in this experiment contains four classes. From the subset scene, a random sample of 20% of the pixels was chosen from the known ground truth of the four classes: Corn-notill, Soybean-notill, Soybean-min, Grass-Trees. The remaining 80% of the known

Table 4.10: Divergence (D), average local geary (C) and accuracy (Acc) for 3-band case

Band			D	C	Acc
1	8	19	26.0890	0.0723	0.9076
11	1	6	29.5938	0.0792	0.8842
8	13	1	30.5100	0.0795	0.8881

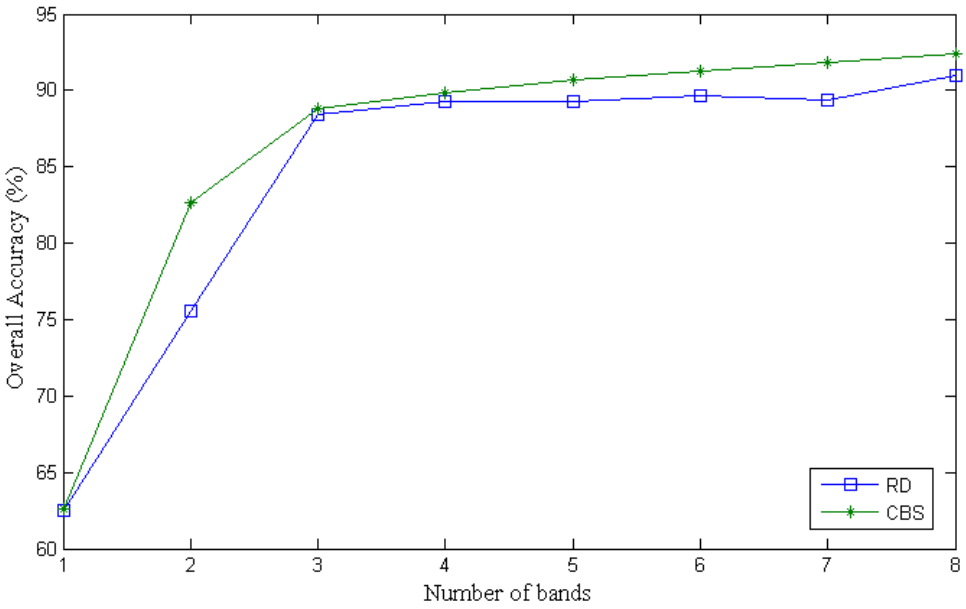


Figure 4.15: Classification accuracy on band selection results for mouse data

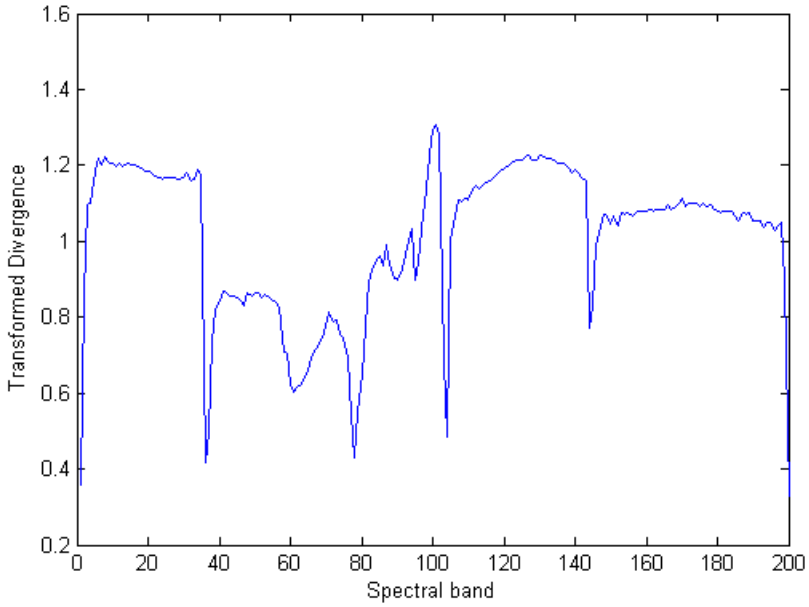
ground pixels in the scene are acted as testing data. Table 4.11 lists the pixel numbers used as training and testing of each class.

Since this is a multi-class problem, we adopt transformed divergence (TD) instead of classic divergence as the criterion. For multi-class problem, the transformed divergence gives an exponentially decreasing weight to increasing distances between the classes, therefore can result a better performance. Figure 4.16 shows the transformed divergence for one band case. Figure 4.16 (a) plots the TD value for each band. The peak value appears at band λ_{101}. The Figure 4.16 (b) plots the TD value vs. pixel classification accuracy on testing data. A similar observation can be found as that in the previous experiments, is that the largest divergence value cannot guarantee the highest classification accuracy.

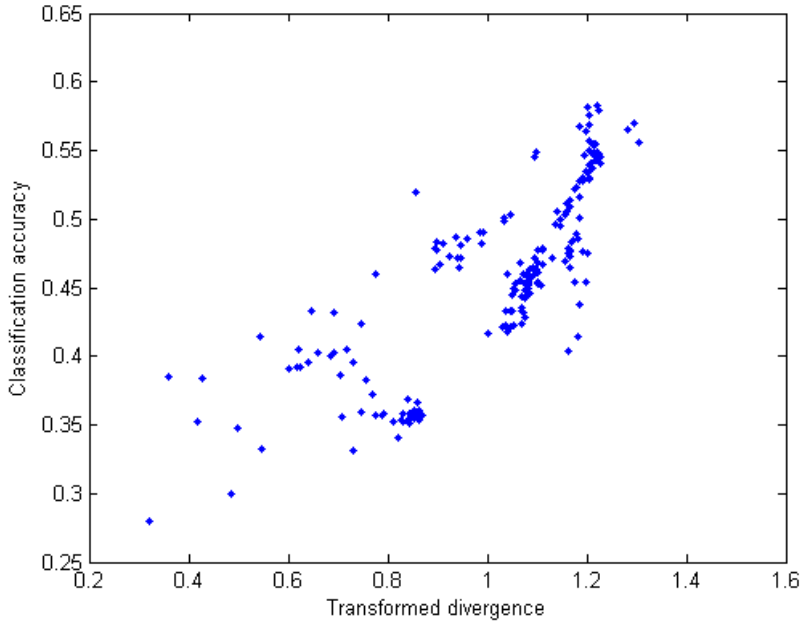
Experiments are carried out to assess the performance of our band selection methods and to compare with two classical subset selection algorithms (ReliefF [85] and CA [93]). Figure 4.17 presents the comparison of classification accuracy of four band selection methods. For all methods, the total accuracy improves as more spectral bands are added into the band subsets. Two proposed methods RD and CBS achieve better performance than the comparing methods. The CBS method yields the highest accuracy among all methods. The RD method produces the second best accuracy. They both outperform the competitive methods.

Table 4.11: Indiana pine data used for training and testing

	Corn-notil	Grass/Tree	Soybean-notil	Soybean-min
# of training	202	146	145	385
# of testing	806	586	582	1541



(a) Transformed divergence value for each band



(b) TD vs. testing accuracy

Figure 4.16: Transformed divergence for 1 band case

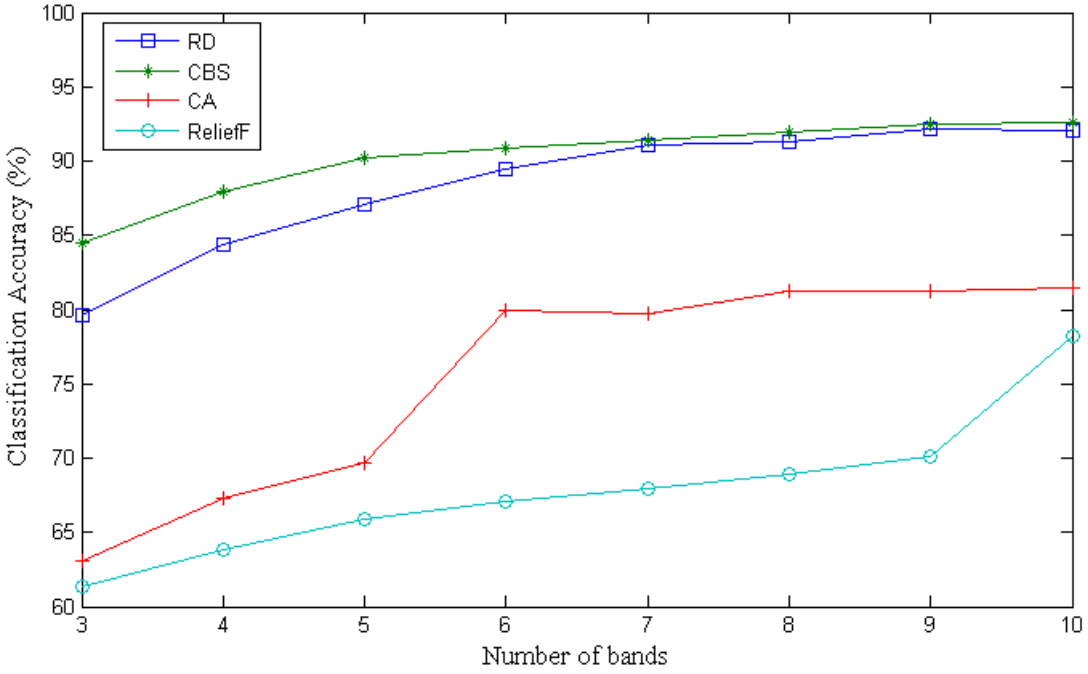


Figure 4.17: Comparison of band selection method performance

5 Integration of Spatial and Spectral Information for Hyperspectral Image Classification

In this chapter, we propose a collaborative classification method for hyperspectral image classification. By using the proposed the collaborative classification method, ***the spectral and spatial information of image can be combined simultaneously***. This method fully utilizes the spatial-spectral relationships inherent in the data, and thus improves performance in hyperspectral classification task.

The chapter is organized as follows. Section 5.1 explains the procedure of hyperspectral classification. Section 5.2 provides a review of previous work for hyperspectral classification. Then, the proposed collaborative classification method is described in Section 5.3. The classification results are presented in Section 5.4.

5.1 Introduction

The main issue on hyperspectral imaging is concentrated on classification. Classification is usually a name given when one is dealing with grouping a large number of pixels into multiple classes. Classification of a hyperspectral image means to identify each pixel into multiple classes in the scene. Once all pixels are classified into one of several classes, the data may be used to produce thematic maps. The thematic map may provide much valuable information, such as to produce summary statistics of the objects in a scene. Therefore, to obtain an accurate classification result or thematic map is a very important issue.

There are two major hyperspectral image classification methods: Supervised classification and unsupervised classification. In unsupervised classification techniques,

no prior information is required. The algorithms aggregate pixels into various classes according to the clusters found in the spectral space. While a typical supervised classification procedure for hyperspectral image usually has three steps:

1) a set of training data of each class is first derived from where both ground truth and spectral information are available,

2) a classification criterion (e.g. discriminant function) is developed and used to classify pixels to each class,

3) the pixel is then assigned to the class depends on the output of the discriminant function.

The supervised spectral classification methods produce satisfying results for many cases, but a main limitation of such methods is that they assign a pixel to a class only depend on the spectral similarity, without any consideration for the spatial locations of that pixel. When the objects in the hyperspectral image do not have unique spectral signatures, the classification results they generated often display noisy or unrealistic features, such as isolated pixels assigned to a particular class. In this situation, additional information is required to distinguish them. Information captured in neighboring pixels or information about pattern surrounding the pixel of interest may provide useful supplementary information. This type of information is referred to as spatial information. Therefore, integration of information from spectral and spatial domain for classification presents the potential for increased classification performance for hyperspectral image classification.

5.2 Review of Classification Techniques

In the previous studies, the methodologies for taking spatial context have been categorized into four different groups [106]. The first approach, which is also the most widely used technique, is to perform postprocess after the image has been classified by a pixel-wise classifier. An example of the postprocessing is using a majority filter. The majority filtering process assigns a pixel's label according to its neighbors. If the local

neighborhood is dominated by certain class, the label of the targeted pixel is assigned to that class to reflect the majority. The postprocessing process can remove outliers in a homogeneous area.

The opposite approach to postprocessing is preprocessing of images. The methods in this category are based on image segmentation. A given image is divided into many homogeneous regions according to their spatial and spectral similarities by using an appropriately chosen criterion. Each homogeneous area is classified by comparing the similarities of the means of each segment to pre-labeled samples.

The well-known ECHO (Extraction and Classification of Homogeneous Objects) algorithm [107][108] is in this category. ECHO separates image pixels into fields of spectrally similar before the pixels are assigned to categories. Classification is then conducted for each field, rather than individual pixels. First, ECHO divides the image into small groups which is consisting of four pixels. For each group, the pixels are tested for homogeneity by a distance to the average value of the group. If the distance is in the tail of the Gaussian density, the groups are not homogenous and are rejected. Then, each individual group is compared to its adjacent 'field', which can be a group of one or more connected groups that have previously been merged. If the two appear statistically similar by some appropriate criterion, then they are merged. Otherwise this group is compared to another adjacent field or becomes a new field. For classification process, a maximum likelihood (ML) classifier was applied for each field resulted from the segmentation process.

Karakahya *et al* [109][110] proposed a two-stage process to the classification of remote sensing images. First, a spatial filter is used to achieve more homogeneous regions, which can improve spectral separability. Then, a maximum likelihood classifier is employed to classify the land covers. De Jong *et al* [111] proposed a SSC (spatial and spectral classifier) method for hyperspectral image classification. SSC method starts by dividing an image into homogeneous and heterogeneous regions based on spectral variation. Then

a conventional per-pixel classification method is applied to classify the homogeneous parts. The heterogeneous parts are classified based on a combination of spectral and contextual information.

The third category is to add new components to the original spectral vectors. This method can also be called as stacked vector method. The new components are features that can carry spatial information. The simplest way is to add mean and variances of neighboring pixel values to the original vectors. Alternatively, the additional components can be derived from texture descriptors such as Fourier coefficients or coocurrent matrices [112]. Camps-Valls et al.[113] proposed a composite kernel machine to enhance the classification accuracy of hyperspectral images. Through exploiting the properties of Mercer's kernels, that method constructs a family of composite kernels which can combine spatial and spectral information. The experiment results prove that this method can efficiently combine contextual and spectral information for hyperspectral image classification.

The final category tries to combines the spectral and spatial (contextual) information and classifies pixels using both sources at the same time through modeling of the scene. The spatial (contextual) information is used by setting up a probabilistic model, which is then used for decision making. The widely used Markov Random Fields (MRF) [114][115] method is in this category.

The MRF is used to construct an a priori model in the Bayesian sense so as to accomplish the Maximum a Posteriori (MAP) estimate. Such a MAP solution often provides more satisfactory results than a Maximum Likelihood classifier (MLC) [105]. In general, the optimization of MRF function is difficult, so there have been many approaches proposed for approximately solving the optimization problem. The Iterative conditional mode (ICM) [116] is the most commonly used one. The ICM method iteratively minimizes the functional with respect to single pixel in much the same manner as used by iterative coordinate descent. The ICM algorithm has widely been applied since it is quick and

produces reasonably accurate categories. The drawback is that it may arrive at a local optimum, hence emphasizing the choice of initial means.

5.3 Collaborative Classification

As we mentioned before, for the spectral classification method, even if the parameter estimation for discriminant function is accurate, the spectral distances does not consider the pixel neighborhood, as an isolated pixel in a uniformly labeled region is more likely to have the same label even if its spectral vector makes it belonging to another class. Therefore, integration of information from spectral and spatial domain for classification presents the potential for increased classification performance for hyperspectral image classification.

From the previous research work, although various approaches have been proposed, they have this or that kind of drawbacks. A comment problem for the postprocess approach is that its performance heavily relies on the initial pixel-wise classification accuracy. If the initial classification accuracy is not good, the postprocessing procedure can even lead to a degraded performance. In addition, the resulting classified thematic maps are often blocky and do not properly identify class boundaries and loose details unnecessarily. For the second approach, although the classification result is usually more reliable than that of individual pixels, the classification accuracies of these algorithms are very sensitive to the initial segmentation results. These segmentation results are critical to achieving a good thematic map. The third approach has a common problem of excessive dimensionality of augmented vectors, slow processing speed and poor performance at the object boundaries since the texture measures are based on a certain size of neighborhood. And for the MRF method, the key is to model the contextual information. For the MRF model, the least square fit method [117] and coding method [118][119] are the two most widely known technique for estimating the model parameters. If the relevant parameters of the model are not accurately defined, this model cannot show its full effectiveness. The

success of the model parameters estimation relies on the complete and correct understanding of image neighborhood configurations. In practical sense, the neighborhood configurations are difficult to acquire. The parameter within MRF model is therefore trivial and restricts the model capabilities.

Here, we propose a cost function based collaborative classification method inspired by [120]. This method can efficiently integrate spectral and spatial information for classification process. The cost function consists of two parts:

- A spectral similarity term to measure how similar a given sample to a particular class.
- A spatial similarity term to measure how similar a pixel to its neighborhood.

The cost function F is defined as follows:

$$F(\boldsymbol{x}_i, \omega_l) = S(\boldsymbol{x}_i, \omega_l) + \alpha B(\boldsymbol{x}_i) \tag{5.1}$$

The first term $S(\boldsymbol{x}_i, \omega_l)$ in the right side of equation (5.1) represents the spectral distance between a sample $\boldsymbol{x}_i$ to a particular class ω_l, which is used to measure how similar a given sample to a particular class in spectral space. This spectral distance measure can be distance-based measure, such as Euclidean distance, Mahalanobia distance, Bhattacharyya distance, or projection-based method, such as spectral angle mapper (SAM), or decision value from discriminant function, such as decision value from support vector machine (SVM) classifier, maximum likelihood classifier (MLC) and maximum a posteriori (MAP) classifier.

Meanwhile, the second term $B(\boldsymbol{x}_i)$ in the right side of the equation represents a measurement for spatial similarity of local i with its neighborhood and is defined as

$$B(\boldsymbol{x}_i) = \sum_j w_{ij} V(\boldsymbol{x}_i, \boldsymbol{x}_j), \; j \in N_i \tag{5.2}$$

where N_i represents the neighborhood of the pixel at location i. N_i can be n-th order neighborhood system. Figure 5.1 illustrates an example of 5-th order neighborhood

5	4	3	4	5
4	2	1	2	4
3	1	×	1	3
4	2	1	2	4
5	4	3	4	5

Figure 5.1: 5-th order neighborhood system

system. w_{ij} is spatial weight for each neighboring pixel. This spatial weight can be defined by either adjacency or distance criteria.

- Adjacency criteria: w_{ij} is 1 if location i is adjacent to location j, and 0 otherwise. According to the distance.
- Distance criteria: w_{ij} is 1 if the distance between location i and j is within a given distance, and 0 otherwise.

$V(\boldsymbol{x}_i, \boldsymbol{x}_j)$ measures the spatial similarity between two pixels at location i and j respectively. How to define this term is the key to measure the spatial similarity among $\boldsymbol{x}_i$'s neighborhood. For example, in MRF method, it determines the spatial relationship based on the class label of $\boldsymbol{x}_i$ and its neighbors. A indication function T(Γ) is defined for this purpose. If the neighborhood pixel $\boldsymbol{x}_j$ has the same label as $\boldsymbol{x}_i$, the value of indication function is 1, otherwise it is 0. In [120], the author proposed a more simple way to measure the spatial similarity for class labels. If the neighborhood pixel $\boldsymbol{x}_j$ has the same label as $\boldsymbol{x}_i$, the spatial similarity measure decreased by certain value, otherwise it increases by certain value.

However a major drawback of earlier methods is they use merely the class label to estimate the spatial information. As we known, the class label of a pixel is determined by its spectral characteristic. When a classifier assigns a pixel to a particular class, it only considers the spectral similarity of the pixel, without any spatial information involved. To

overcome this drawback, in this section, we propose a new formula to estimate the spatial information. The new measure $V(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is defined as

$$V(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} -\dfrac{1}{c_j} & \text{if } \omega_i = \omega_j \text{ and } j \in N_i \\ \dfrac{1}{c_j} & \text{if } \omega_i \neq \omega_j \text{ and } j \in N_i \end{cases} \tag{5.3}$$

where c_j is the multidimensional local spatial autocorrelation value for location j. If the neighborhood pixel $\boldsymbol{x}_j$ has the same label as $\boldsymbol{x}_i$, the function decreased by c_j, otherwise it increases by c_j. The c_j value provides the spatial autocorrelation for location j, and can be calculated by Equation (4.37). The smaller value of c_j means more similar of $\boldsymbol{x}_j$ with its neighboring pixels, then the classification result for $\boldsymbol{x}_j$ should be more reliable. α is a positive scalar to control the influence of spatial term. If α equals to 0 means no spatial information is considered.

In our approach, we use the SVM as the classifier, so the cost function in our approach is defined as follows:

$$F(\boldsymbol{x}_i, \omega_l) = S_{SVM}(\boldsymbol{x}_i, \omega_l) + \alpha \sum_j w_{ij} V(\boldsymbol{x}_i, \boldsymbol{x}_j),\ j \in N_i \tag{5.4}$$

$S_{SVM}(\boldsymbol{x}_i, \omega_l)$ is the distance between a sample $\boldsymbol{x}_i$ and a given SVM hyperplane ω_l.

$$S_{SVM}(\boldsymbol{x}_i, \omega_l) = -(\sum y_{SV} \alpha_{SV} K(\boldsymbol{x}_i, \boldsymbol{x}_{SV}) + b) \tag{5.5}$$

Where $\boldsymbol{x}_{SV}$ is support vectors, y_{SV} is the label of x_{SV}, and α_{SV}, b are the SVM parameters obtained by training a one-against-all SVM classifier.

The decision rule for equation (5.5) is

$$x_i \in \omega_l \text{ if } \arg\min_{l=1,2,\ldots,L} F(x_i, \omega_l) \tag{5.6}$$

The algorithm can be summarized as

Algorithm 3: Collaborative Hyperspectral Classification Algorithm

1. *Calculate* $S_{SVM}(\boldsymbol{x}_i, \omega_l)$ *for* $\boldsymbol{x}_i$ *and obtain the initial label for* $\boldsymbol{x}_i$.

2. *Calculate* $F(\boldsymbol{x}_i, \omega_l)$ *for all* $\boldsymbol{x}_i$ *and re-labeling them according Equation (5.6)*

3. *Go to step 2 until no change is detected.*

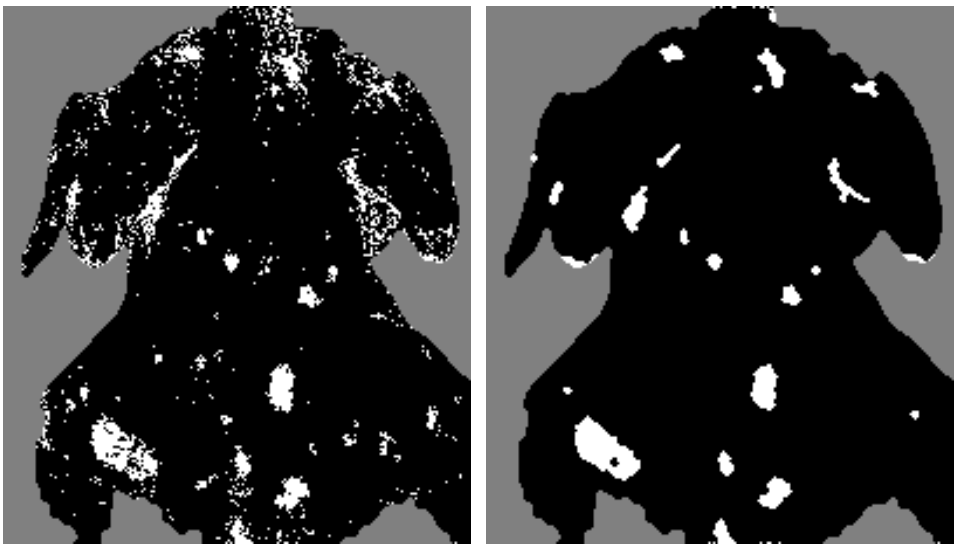
5.4 Experimental Results

In this section, the classification results with cost function based collaborative classification method are presented. In these experiments, the SVM classifier is adopted instead of MLC since SVM can provide better classification performance than MLC. For the chicken data and mouse data, *2*-th order neighborhood system is adopted.

5.4.1 Hyperspectral Chicken Data

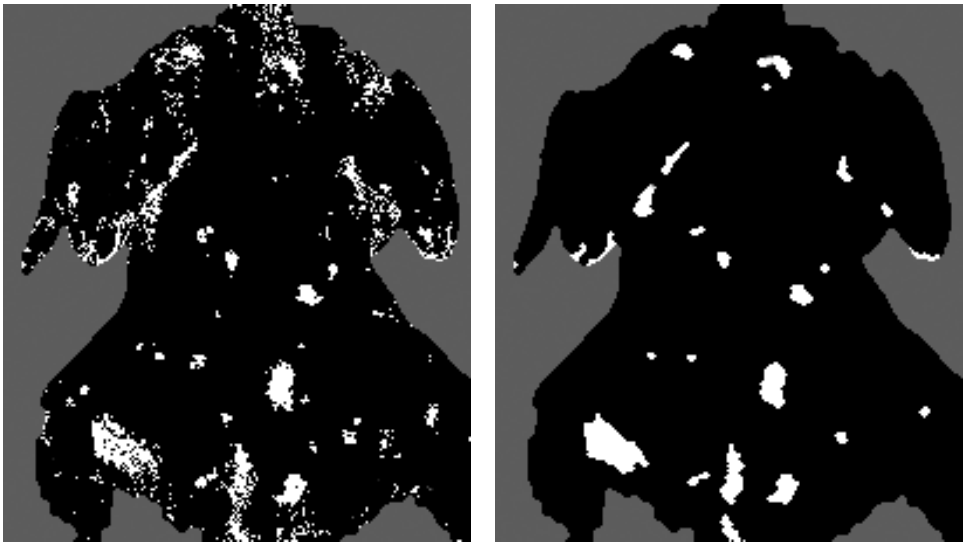
From Figure 4.7, the incremental of classification accuracy become very small after four bands. Hence, the first four selected bands are used to training classifier and classify the test data.

Figure 5.2 and Figure 5.3 shows the classification results for four bands selected by RD method and CBS method, respectively. Figure 5.2 (a) displays the classification result from SVM classifier. The SVM classifier uses only the spectral information as the criterion for decision making. As shown in the figure, the classified pixels have a lot of false positive errors, which is many normal pixels are misclassified as tumor. Also, we can find plenty of isolated pixels, which are misclassified as tumors. The SVM classification result for CBS results is shown in Figure 5.3 (a). It also has lots of false positive errors, but compared to the result in Figure 5.2 (a), the bands selected by CBS method generate less false positive errors than those chosen by RD method. This



(a) Classified by SVM (b) Classified by collaborative method

Figure 5.2: Classification result with RD result for training data



(a) Classified by SVM (b) Classified by collaborative method

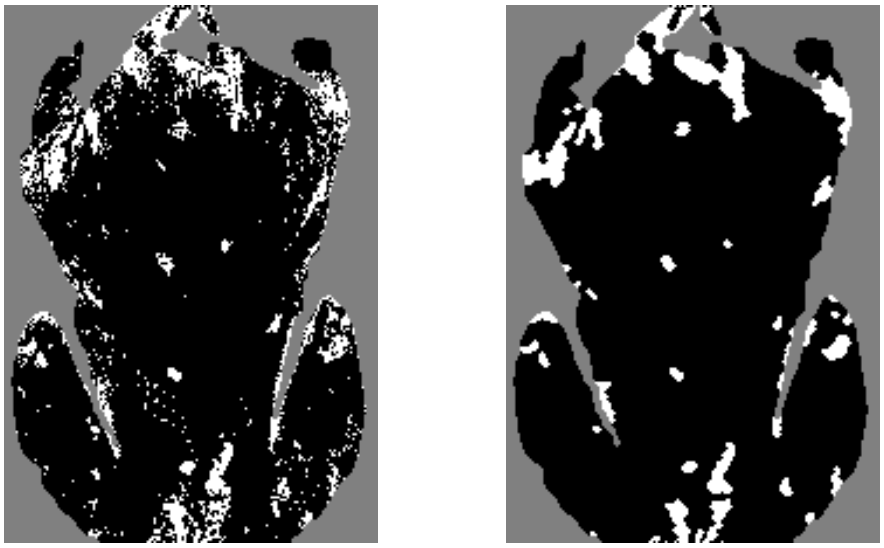
Figure 5.3: Classification result with CBS result for training data

observation is consistent with the one we got from Figure 4.8 (c). The classification results by using cost function based collaborative classification method are displayed in Figure 5.2 (b) and Figure 5.3 (b) for bands selected by RD method and by CBS method, respectively. Since the collaborative classification method integrating spectral and spatial information together during the decision-making process, the performance is much better than using spectral information only. Using same selected bands, the collaborative classification method can remove most false positive errors and generate more accuracy classification result than SVM classifier.

Figure 5.4 and Figure 5.5 show the classification result for one of the testing chicken data. Similar conclusion can be drawn from these results, the CBS method can generate less false positive error than the divergence method and by integrating the spectral and spatial information, the classification accuracy for hyperspectral image can be improved.

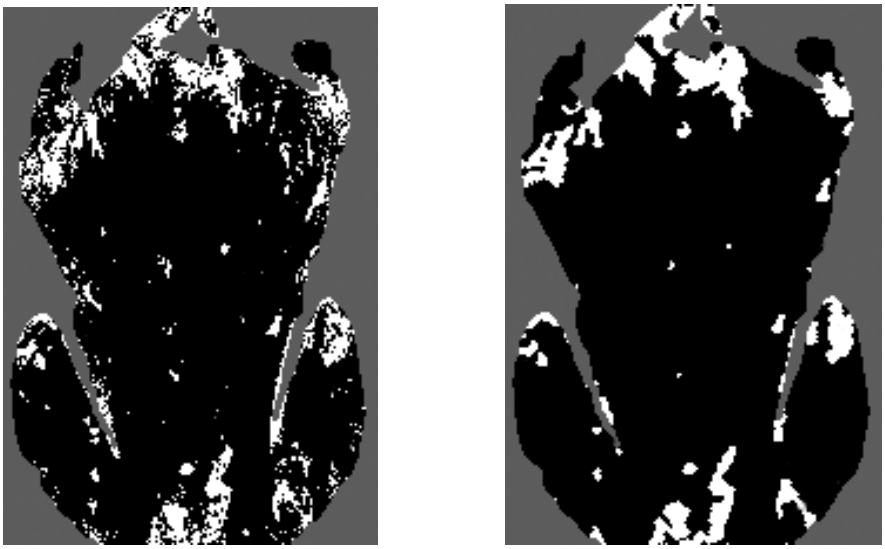
The output of the classifier shows the locations of potential tumors. A potential tumor is a region that consists of pixels identified as a tumor in classification. Because some normal tissues are spectrally very similar to tumors, the classifier usually yields more tumor spots than actually exist. The post processing is to further remove the false positives based on the location, shape and size of potential tumors. Most tumors have a round shape and size of the tumor cannot be too big or too small. Also because of the illumination reason, false positive errors are more likely happened at the edge area of the image. Employing these rules during post processing, the final detection results are shown in Figure 5.6. White spots indicate the tumors correctly detected and then white areas enclosed by a rectangle indicate false positives.

Table 5.1 summarizes the tumor detection results on 11 poultry samples with 4 spectral bands selected using the RD, Hybrid, and ES. The average detection rates were 90.6% for RD and CBS and 93.75% for ES. Band selection with the RD has 3 missed tumors and 19



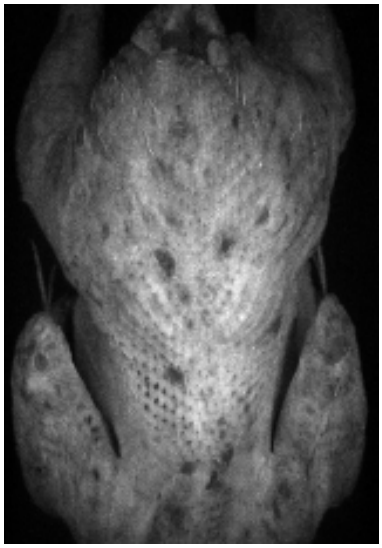
(a) Classified by SVM (b) Classified by collaborative method

Figure 5.4: Classification result with RD result for testing data

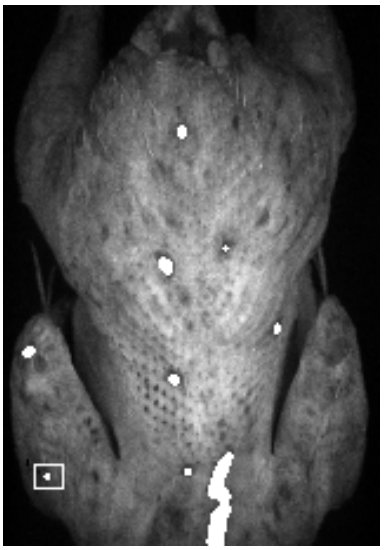


(a) Classified by SVM (b) Classified by collaborative method

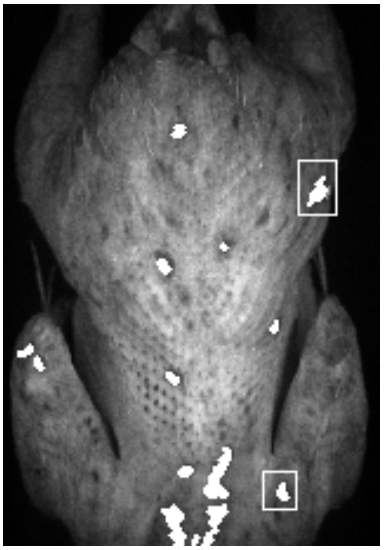
Figure 5.5: Classification result with CBS result for testing data



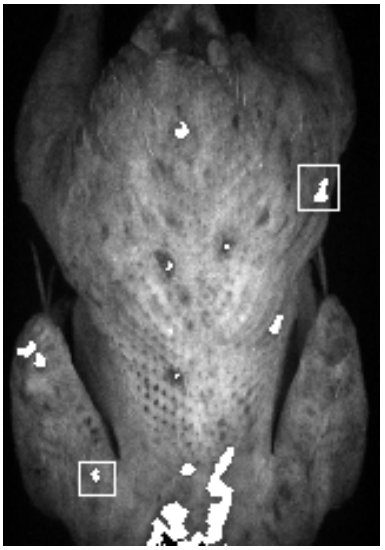
(a) Original image (λ_{10})



(b) Exhaustive search



(c) RD



(d) CBS

Figure 5.6: Tumor detection results with four spectral bands selected

Table 5.1: Tumor detection performance of the RD, CBS, and Exhaustive search with 4 bands

Image No.	Number of Tumors	RD		CBS		ES	
		# Found	FPs	# Found	FPs	# Found	FPs
1	8	8	2	8	2	8	1
2	2	2	1	2	1	2	1
3	0	0	2	0	1	0	1
4	3	3	1	3	1	3	1
5	2	2	1	2	1	2	1
6	2	2	2	2	1	2	2
7	2	2	2	2	1	2	1
8	0	0	0	0	0	0	0
9	4	3	2	3	1	3	1
10	7	5	2	5	2	6	2
11	2	2	1	2	1	2	1
Total	32	29 (90.6%)	16	29 (90.6%)	12	30(93.75%)	12

false positives (FPs) on average while the CBS shows 3 missed tumors and 12 false positives. The CBS method can efficiently reduce the number of false positive errors.

5.4.2 Hyperspectral Mouse Data

For this data set, we will analyze the performance of the collaborative classification method in a more quantitive way. First, we use the SVM to classify testing data with all the spectral information, i.e., all 21 bands. Figure 5.7 shows the classification results with SVM and collaborative classification method. We conduct classification of the pixels in a small circular region of interest generated by an endoscope. Normal tissue is indicated as the dark area and white region corresponds to tumor. The gray region outside the region of interest is not considered. It is clear that the result generated by collaborative classification method has less misclassified pixels. Table 5.2 compares the performance of SVM and the collaborative classification method from overall accuracy, tumor accuracy and normal tissue accuracy. In all three terms, the collaborative classification method shows better results. The Overall accuracy increases from 92.23% to 93.85%. The improvement for tumor accuracy is 1.71%, for normal tissue is 1.6%.

Next we analyze the classification result on the bands selected by different band selection methods. The band selection results for mouse data are presented in Section 5.2. Figure 5.8 displays the classification results with band λ_{11}. This band is the first band selected by RD method. The classification accuracy with this band is only 61% (listed in Table 5.3). This result points out that with only spectral band, we cannot precisely detect tumor. We need more spectral information. This maybe is the motivation of investigating hyperspectral imaging.

The classification results with three selected bands [λ_{11}, λ_1, λ_6] are displayed in Figure 5.9 and Table 5.4. The accuracy jumps to 85% with three bands. The collaborative classification method yields higher accuracy in overall accuracy, tumor accuracy and

(a) SVM (b) collaborative method

Figure 5.7: Classification of skin tumor with all the 21 spectral bands

Table 5.2: Classification accuracy of skin tumor with all the 21 spectral bands

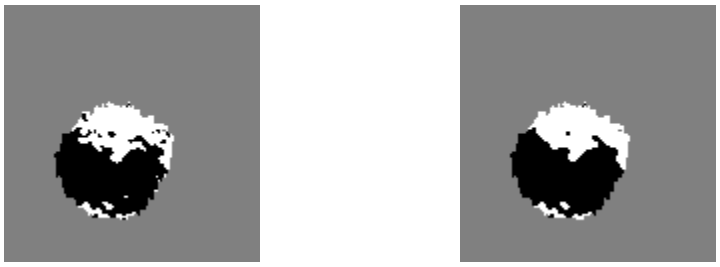
	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	92.23	95.17	91.22
Collaborative method	93.85	96.88	92.82

(a) SVM (b) Cost function

Figure 5.8: Classification of skin tumor with band λ_{11}

Table 5.3: Classification accuracy of skin tumor with band λ_{11}

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	61.40	87.38	52.53
Collaborative method	61.13	86.92	52.32



(a) SVM (b) Collaborative method

Figure 5.9: Classification of skin tumor with band [λ_{11}, λ_1, λ_6]

Table 5.4: Classification accuracy of skin tumor with band [λ_{11}, λ_1, λ_6]

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	84.97	91.59	82.70
Cost function	85.88	94.70	82.86

normal tissue accuracy than the SVM method. Figure 5.10 and Table 5.5 show the result with five bands. And Figure 5.11, Table 5.6 shows the results with ten bands. For both cases, the collaborative classification method outperforms the SVM method. With 5 spectral bands, the collaborative classification method yields 88.66% overall accuracy, while SVM generates only 86.12%. For ten-band case, SVM can correctly classify 90.48% pixels; while collaborative classification method can make 93.10% pixels correctly classified.

For the bands selected by the CBS method, Figure 5.12 and Table 5.7 show the results with band λ_8. The classified result is poor. It only give 59% accuracy, this result is just better than the random guess. But the accuracy appears a big promotion with three bands. As shown in Figure 5.13 and Table 5.8, the accuracy reaches 90% for the collaborative classification method and 88% for SVM. Recall that for RD method, three bands can only get 85% accuracy. This proves the efficiency of integrating spatial information once more. The classification results with five spectral bands are listed in Figure 5.14 and Table 5.9. The accuracy further increases to 91.75% and 91.75 for the collaborative classification method and for SVM, respectively. With ten bands, the accuracy reaches 92.94% and 90.96% for the collaborative classification method and for SVM, respectively. These results are shown in Figure 5.15 and Table 5.10.

(a) SVM (b) Collaborative method

Figure 5.10: Classification of skin tumor with band [λ_{11}, λ_1, λ_6, λ_{14}, λ_{13}]

Table 5.5: Classification accuracy of skin tumor with band [λ_{11}, λ_1, λ_6, λ_{14}, λ_{13}]

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	86.12	92.06	84.09
Collaborative method	88.66	94.39	86.70

(a) SVM (b) Collaborative method

Figure 5.11: Classification of skin tumor with 10 bands

Table 5.6: Classification accuracy of skin tumor with 10 bands

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	90.48	94.08	89.25
Collaborative method	93.10	95.79	92.18

(a) SVM (b) Collaborative method

Figure 5.12: Classification of skin tumor with band [λ_8]

Table 5.7: Classification accuracy of skin tumor with band [λ_8]

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	59.90	91.12	49.23
Collaborative method	59.18	90.65	48.43

(a) SVM (b) Collaborative method

Figure 5.13: Classification of skin tumor with band [λ_8, λ_{13}, λ_1]

Table 5.8: Classification accuracy of skin tumor with band [λ_8, λ_{13}, λ_1]

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	88.38	90.81	87.55
Collaborative method	90.60	92.06	90.10

(a) SVM (b) Collaborative method

Figure 5.14: Classification of skin tumor with band [λ_8, λ_{13}, λ_1, λ_6, λ_{11}]

Table 5.9: Classification accuracy of skin tumor with band [λ_8, λ_{13}, λ_1, λ_6, λ_{11}]

	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	89.25	93.46	87.81
Collaborative method	91.75	95.64	90.42

(a) SVM (b) Collaborative method

Figure 5.15: Classification of skin tumor with 10 bands

Table 5.10: Classification accuracy of skin tumor with 10 bands

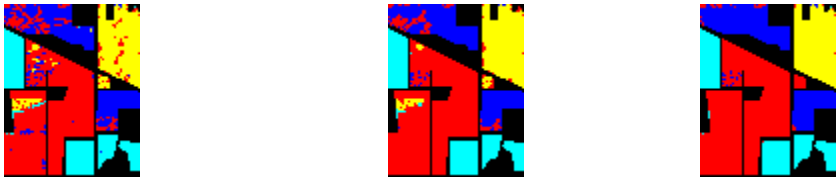
	Overall Accuracy (%)	Tumor Accuracy (%)	Normal Tissue Accuracy (%)
SVM	90.96	94.70	89.68
Collaborative method	92.94	96.88	91.59

5.4.3 Indiana Pine Data

In this part, we will test the performance of the proposed classification method for multi-class problem. From the subset scene, there are four classes: Corn-notill, Soybean-notill, Soybean-min, Grass-Trees. The training data are used to train four binary SVM classifiers. The results with two different neighborhood systems: *2*-th order and 5-th order are presented and compared.

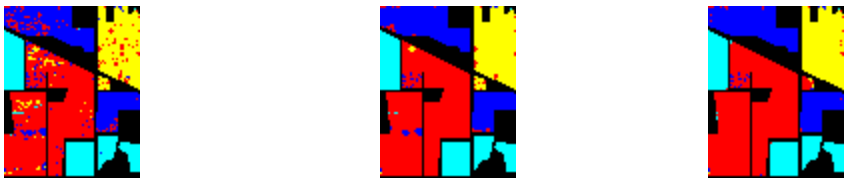
First, we conduct classification on RD results. Figure 5.16 - Figure 5.18 show the classification results for 3, 4 and 5 bands, respectively. The Figure 5.16 (a) shows the classification result obtained by SVM. It is obviously that there are lots of misclassified pixels. And the accuracy for it is 85.43%. While by using the proposed classification method, the accuracy can be greatly enhanced. When using the *2*-th order neighborhood system (3 by 3 window), the final accuracy is 90.75%. If using *5*-th order neighborhood system (5 by 5 window), the accuracy can reach 96.70%. For 4-band case, the initial accuracy with SVM is 87.25%, when applied proposed method, the optimal accuracy can increase to 96.70%. For 5 bands, the initial accuracy is 89.05%, the proposed method yields 96.76% accuracy. From these results, we find that for this data set, when using the *2*-th order neighborhood system, the optimal accuracy can be obtained with four spectral bands. If using the *5*-th order neighborhood system, then three spectral bands are enough to get a good result.

The classification results for CBS method are displayed in Figure 5.19 - Figure 5.21, respectively. For 3-band case, the SVM accuracy for CBS method is 86.57%, which is higher than that of RD. This proves that the CBS method can generate better band selection result. If using proposed classification method, the final accuracy can increase to 91.61% and 96.73% for *2*-th order and 5-th order neighborhood system, respectively. For 4-band case, the initial SVM accuracy is 89.36%, also higher than that of RD. Same conclusion can be drawn for 5-band case. When using the *2*-th order neighborhood



(a) Accuracy: 85.43% (b) Accuracy: 90.75% (c) Accuracy: 96.70%

Figure 5.16: Classification results for 3 bands selected by RD

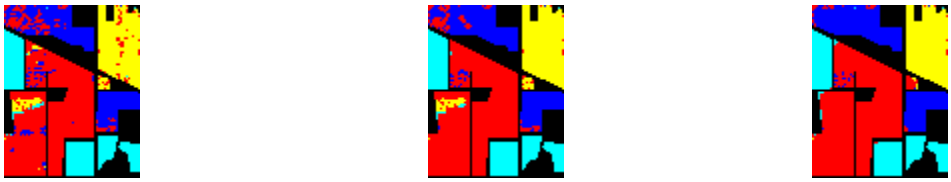


(a) Accuracy: 87.25 % (b) Accuracy: 94.82% (c) Accuracy: 96.70%

Figure 5.17: Classification results for 4 bands selected by RD

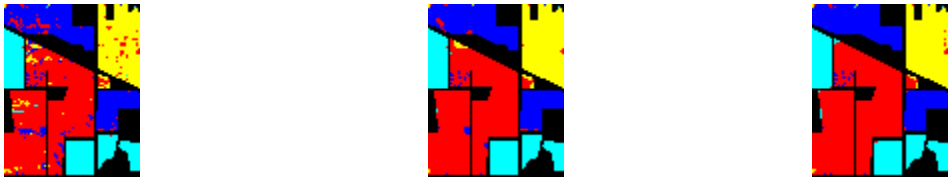
(a) Accuracy: 89.05% (b) Accuracy: 94.84% (c) Accuracy: 96.76%

Figure 5.18: Classification results for 5 bands selected by RD



(a) Accuracy: 86.57% (b) Accuracy: 91.61% (c) Accuracy: 96.73%

Figure 5.19: Classification results for 3 bands selected by CBS



(a) Accuracy: 89.36% (b) Accuracy: 95.39% (c) Accuracy: 96.70%

Figure 5.20: Classification results for 4 bands selected by CBS

system, the optimal accuracy 95.39% can be obtained with four spectral bands. And for the 5-th order neighborhood system, using three spectral bands can yield 96.73% accuracy.

Table 5.11 shows classification results of different methods for same dataset. The Euclidean classifier [129] and SVM classifier [130] only use the spectral information. The BLOOC+DAFE+ECHO [129] uses contextual and spectral information to classify homogeneous objects. It is important to stress that all these methods are using 200 spectral bands for classification, while in our experiment, we use only 3 or 4 spectral bands and we can obtain better classification performance than all of them. It can be concluded that the proposed integration method is a great improvement for hyperspectral image classification.



(a) Accuracy: 90.01% (b) Accuracy: 95.39% (c) Accuracy: 96.76%

Figure 5.21: Classification results for 5 bands selected by CBS

Table 5.11: Classification accuracy for other works

Method	Euclidean classifier	BLOOC+DAFE+ECHO	SVM
Accuracy	67.43%	93.50%	95.90%

6 Conclusions and Future Works

This dissertation addresses the problem of integrating spectral and spatial information for hyperspectral image classification. The huge amount of hyperspectral image data often makes real-time computer processing a challenging task. To overcome this problem, this dissertation proposes two band selection algorithms in Chapter 4. First, this research work proposes a band selection method for hyperspectral images based on recursive divergence. This method avoids transforming the original hyperspectral images to the feature space. A set of recursive equations for the calculation of divergence with an additional band is derived to overcome the computational restrictions in real-time processing. In the recursive divergence based band selection, a good subset is to maximize the representation of the spectral separability. However, separability maximization does not guarantee a classification process that will produce the best visual result, or the most accurate. After carefully investigating the existing spatial autocorrelation measure, this dissertation proposes a Multidimensional Local Spatial Autocorrelation (MLSA) for hyperspectral image data. Based on the new local spatial autocorrelation measure, a collaborative band selection strategy is developed that combines both the spectral separability measure and spatial homogeneity measure of hyperspectral band selection.

In Chapter 5, this dissertation discusses the classification for hyperspectral image. A collaborative classification method is proposed in this chapter to integrate the spectral and spatial information for the classification process. The proposed collaborative classification method consists of a spectral similarity term to measure the similarity of a given sample to a particular class and a spatial similarity term to measure how similar a pixel to its neighboring pixels. It can fully utilize the spatial-spectral relationships inherent in the data and enhance the classification performance.

The proposed band selection and classification method was applied for three exclusive applications, including chicken tumor detection, apple surface contamination detection, skin cancer detection and a public dataset, Indiana pine data. The performances of the proposed methods clearly show the necessity and efficiency of integrating spatial information for hyperspectral image processing. They outperformed the other compared method which using spectral information only. And it can be concluded that the proposed integration method is a great improvement for hyperspectral image classification.

Although this dissertation makes a deep and extensive research on integrating the spectral and spatial information for hyperspectral classification, there are still many problems worth further consideration. In order to combine the spatial information with spectral information, first we need to have a criterion to measure the spatial information in the hyperspectral image. Although there exist many spatial statistic measures, most of them have only been developed for the single image band. In this dissertation, we propose a multidimensional local spatial autocorrelation measure to assess the spatial information for hyperspectral data. It is possible to propose more spatial measures for hyperspectral image. How to develop other spatial method for hyperspectral image is an important and challenging work in future. Additionally, in this dissertation, the sequential forward search strategy is used in band selection procedure. The advantage of this search strategy is computational efficient. But this strategy can only find a near-optimal solution. To further improve the band selection performance, it is important to find a better search strategy.

Bibliography

[1] J. Taranik, D. Mouat, and C. Elvidge, "Hyperspectral technology for geologic applications," *Proceedings of International Geoscience and Remote Sensing Symposium*, pp.917-920, 1993.

[2] R. Birk and T. McCord, "Airborne hyperspectral sensor systems," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 9, No.10, pp.26-33, 1994.

[3] E. Perry, H. Foote, G. Petrie, K. Steinmaus, D. Irwin, and A. Stephan, "Exploitation of hyperspectral imagery and lidar for landuse classification," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vol. 2, pp.1013-1015, 1998.

[4] T. Wilson and R. Felt, "Hyperspectral remote sensing technology (HRST) program," *Proceedings of IEEE Aerospace Conference*, Vol. 5, pp.193-200, 1998.

[5] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Magazine*, Vol. 19, No.1, pp.12-16, 2002.

[6] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Processing Magazine*, Vol. 19, No. 1, pp.17-28, 2002.

[7] W. Mauser, "The airborne visible/infrared imaging spectrometer AVIS-2 - multiangular and hyperspectral data for environmental analysis**,"** *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vol. 3, pp.2020-2022, 2003.

[8] A. Bannari, M. Chevrier, K. Staenz, and H. McNairn, "Senescent vegetation and crop residue mapping in agricultural lands using artificial neutral networks and hyperspectral remote sensing**,"** *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vol. 7, pp.4292-4294, 2003.

[9] C. Yang, J. Everitt, and J. Bradford, "Airborne hyperspectral imaging and yield monitoring of grain sorghum yield variability," *American Society of Agricultural Engineers Meeting Paper*, No. 02-1079, 2002.

[10] M. Eismann, "Strategies for hyperspectral target detection in complex background environments**,"** *IEEE Aerospace Conference*, pp.10-20, 2006.

[11] T. Nichols, J. Thomas, W. Kober, and V. Velten, "Interference-invariant target

detection in hyperspectral images," *Proceedings of SPIE*, Vol. 3372, pp. 176-87, 1998.

[12] B. Thai and G. Healey, "Invariant subpixel target identification in hyperspectral imagery," *Proceedings of SPIE*, Vol. 3717, pp. 14-24, 1999.

[13] J. Goutsias and A. Banerji, "A morphological approach to automatic mine detection problems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 34, No. 4, pp. 1085-1096, 1998.

[14] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Illumination-invariant face recognition in hyperspectral images," *Proceeding of SPIE*, Vol. 5093, pp. 275-282, 2003.

[15] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No.12, pp. 1552-1560, 2003.

[16] K. Rajpoot, N. Rajpoot, and M. Turner, "Hyperspectral Colon Tissue Cell Classification," *SPIE Medical Imaging, San Diego*, Vol. 5370, 2004.

[17] S. Kong, Y. Chen, I. Kim, and M. Kim, "Analysis of Hyperspectral Fluorescence Images for Poultry Skin Tumor Inspection," *Applied Optics*, Vol. 43, No. 4, pp.824-833, February 2004.

[18] S. Kong, Z. Du, M. Martin, and T. Vo-Dinh, "Hyperspectral Fluorescence Image Analysis for Use in Medical Analysis," *Proceedings of SPIE Conference on Biomedical Optics*, San Jose, CA, January 2005.

[19] S. Kong, M. Martin, and T. Vo-Dinh, "Hyperspectral Fluorescence Imaging for Mouse Skin Tumor Detection," *ETRI Journal*, Vol. 28, No. 6, pp.770-776, December 2006.

[20] D. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing,* New York, Wiley, 2003.

[21] A. Dell, P. Gamba, A. Ferrari, J. Palmason, J. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geoscience and Remote Sensing Letters,* Vol. 1, No.4, pp.322-326, 2004.

[22] R. Sundberg, R. Kennett, J. Gruninger, and R. Haren, "Extraction of spatial and spectral scene statistics for hyperspectral scene simulation," *Proceedings. Of IEEE International Geoscience and Remote Sensing Symposium,* pp.3237-32405, 2004.

[23] J. Richards and X. Jia, *Remote sensing digital image analysis: an introduction*, Springer, 1999.

[24] A. Cheriyadat and L. Bruce, "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data," *Proceedings of IEEE Geoscience and Remote Sensing Symposium*, Vol. 6, pp.3420-3422, 2003.

[25] L. Bruce, C. Koger, J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 10, pp.2331-2338, 2002.

[26] L. Jimenez, D. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 6, pp. 2653-2667, 1999

[27] G. Zavattini, S. Vecchi, R. Leahy, D. Smith, and S. Cherry, "A hyperspectral fluorescence imaging system for biological applications," *IEEE Nuclear Science Symposium Conference Record*, Vol.2, pp. 942-946, 2003.

[28] J. Lakowicz, *Topics in fluorescence spectroscopy, Volume 1 Techniques*, New York: Plenum Press, 1991.

[29] V. Vapnik, *Statistical Learning Theory*, New York: John Wiley & Sons, 1998.

[30] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, New York: Cambridge University Press, 2000.

[31] J. Zhang, Y. Zhang, and T. Zhou, "Classification of hyperspectral data using support vector machine," *IEEE Proceedings of International Conference on Image Processing*, Vol. 1, pp.882-885, 2001.

[32] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vol. 1, pp.288-290, 2001.

[33] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing

images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 8, pp.1778-1790, 2004.

[34] N. Venkateswarlu, and V. Raju, "Three Stage ML Classifier," *Pattern Recognition*, Vol. 24, No. 11, pp.1113-16, 1991.

[35] C. Lee, and D. Landgrebe, "Fast Likelihood Classification," *IEEE Transactions on Geoscience Remote Sensing*, Vol. 29, No. 4, pp.509-517, 1991.

[36] X. Jia and J. Richards, "Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets," *IEEE Transactions on Geoscience Remote Sensing*, Vol. 32, No. 2, pp. 274-81, 1994.

[37] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, *2nd edition*, Wiley-Interscience, 2000.

[38] F. Bryan "Diseases transmitted by foods," Atlanta: Centers for Disease Control, 1982.

[39] X. Cheng, *Hyperspectral imaging and pattern recognition technologies for real-time fruit safety and quality inspection*, Ph.D. Dissertation, University of Maryland, 2004.

[40] Y. Chen, and M. Kim "Machine vision sees the food contaminants we can't see," *Agricultural research,* pp. 4-8, 2002.

[41] M. Kim, Y. Chen, and P. Mehl, "Hyperspectral Reflectance and Fluorescence Imaging System for Food Quality and Safety," *Transactions of the American Society of Agricultural Engineers*, Vol. 44, No.3, pp.721-729, 2001.

[42] R. Lu, "Detection of bruises on apples using near-infrared hyperspectral imaging," *Transactions of the American Society of Agricultural Engineers*, Vol. 46 No.2, pp.523-530, 2003.

[43] M. Kim, A. Lefcourt, Y. Chen and S. Kang, "Uses of Hyperspectral and Multispectral Laser Induced Fluorescence Imaging Techniques for Food Safety Inspection," *Key Engineering Materials*, Vol. 270-273, pp.1055-1063, 2004.

[44] A. Lefcourt, M. Kim, Y. Chen and S. Kang, 'Systematic Approach for Using Hyperspectral Imaging Data to Develop Multispectral Imaging Systems: Detection of Feces on Apples," *Computers and Electronics in Agriculture*, Vol. 54, pp.22-35,

2006.

[45] M. Kim, A. Lefcourt, Y. Chen, I. Kim, D. Chan, and K. Chao, "Multispectral Detection of Fecal Contamination on Apples Based on Hyperspectral Imagery: Part I. Applications of Visible and Near-Infrared Reflectance Imaging," *Transactions of the American Society of Agricultural Engineers*, Vol. 45, No.6, pp.2027-2037, 2002.

[46] M. Kim, A. Lefcourt, Y. Chen, I. Kim, D. Chan, and K. Chao, "Multispectral Detection of Fecal Contamination on Apples Based on Hyperspectral Imagery: Part II. Applications of Hyperspectral Fluorescence Imaging," *Transactions of the American Society of Agricultural Engineers*, Vol. 45, No.6, pp.2039-2047, 2002.

[47] I. Kim, M. Kim, Y. Chen, and S. Kong, "Detection of skin tumors on chicken carcasses using hyperspectral fluorescence imaging," *Transactions of the American Society of Agricultural Engineers*, Vol. 47, No.5, pp.1785-1792, 2004.

[48] K. Chao, P. Mehl, and Y. Chen, "Use of hyper- and multi-spectral imaging for detection of chicken skin tumors," *Applied Engineering in Agriculture*, Vol. 18, No. 1, pp. 113-119, 2002.

[49] S. Nakariyakul and D. Casasent, "Fusion algorithm for poultry skin tumor detection using hyperspectral data," *Applied Optics*, Vol. 46, No. 3, pp. 357-364, 2007.

[50] J. Fletcher and S. Kong, "Principal component analysis for poultry tumor inspection using hyperspectral fluorescence imaging," *Proceedings of the International Joint Conference,* Vol.1, pp.149-153, 2003.

[51] Z. Du, M. Jeong, and S. Kong, "Band Selection of Hyperspectral Images for Automatic Detection of Poultry Skin Tumors," IEEE Transactions on Automation Science and Engineering, Vol. 4, No. 3, pp.332-339, 2007.

[52] *Cancer Facts & Figures 2005*. *American Cancer Society Inc*, Atlanta, GA, 2005.

[53] A. Geller, D. Miller, G. Annas, M. Demierre, B. Gilchrest, and H. Koh, "Melanoma Incidence and Mortality Among US Whites, 1969-1999," *Journal of the American Medical Association*, Vol. 288, No.14, pp.1719-1720, 2002.

[54] J. Zhang, C. Chang, S. Miller, and K. Kang, "Optical biopsy of skin tumors," *Proceedings of Conference on Engineering in Medicine and Biology*, Vol. 2, pp.13-

16, 1999.

[55] Z. Du, S. G. Kong, M. Martin, and T. Vo-Dinh, “Hyperspectral Fluorescence Signature Analysis for Skin Cancer Detection,” *World Congress on Medical Physics and Biomedical Engineering*, Seoul, 2006.

[56] D. Ferris, R. Lawhead, E. Dickman, N. Hotzapple, J. Miller, S. Grogan, S. Bambot, A. Agrawal, and M. Faupel, "Multimodal Hyperspectral Imaging for the Noninvasive Diagnosis of Cervical Neoplasia," *Multimodal Hyperspectral Imaging for the Noninvasive Diagnosis of Cervical Neoplasia*, Vol. 5, No.2, pp.65-72, 2001.

[57] K. Rajpoot and N. Rajpoot, "Wavelet based segmentation of hyperspectral colon tissue imagery," *7th International Multi Topic Conference*, pp.38-43, 2003.

[58] K. Rajpoot and N. Rajpoot, "SVM Optimization for Hyperspectral Colon Tissue Cell Classification," *Proceedings 7th International Conference on Medical Image Computing and Computer Assisted Intervention*, 2004.

[59] B. Brisco, R. Hirose, T. McNairn, and K. Staenz, "Precision Agriculture and the Role of Remote Sensing," *Canadian Journal of Remote Sensing*, Vol. 24, No.3, pp.315-327, 1998.

[60] P. Goel, S. Prasher, J. Landry, R. Patel, and A. Viau, "Hyperspectral Image Classification to Detect Weed Infestations and Nitrogen Status in Corn," *Transactions of the American Society of Agricultural Engineers*, Vol. 46, No.2, pp.539-550, 2003.

[61] C. Yang, J. Everitt, and C. Mao, "A hyperspectral imaging system for agricultural applications," *American Society of Agricultural Engineers Meeting Paper No.01-1107*, 2001.

[62] C. Yang, J. Everitt, and J. Bradford, "Airborne hyperspectral imaging and yield monitoring of grain sorghum yield variability," *Proceedings of American Society of Agricultural Engineers Annual Meeting*, 2002.

[63] K. Rajpoot, N. Rajpoot, and M. Turner, "Hyperspectral Colon Tissue Cell Classification," *SPIE Medical Imaging, San Diego*, 2004.

[64] A. Stein, F. van der Meer, and B. Gorte, *Spatial Statistics for Remote Sensing*. Kluwer Academic Publishers, Dordrecht, 1999.

[65] M. Shaban, and O. Dikshit, “Improvement of classification in urban areas by the use of textural features: The case study of Lucknow city, Uttar Pradesh,” International *Journal of Remote Sensing*, Vol. 22, No. 4, pp. 565-593, 2001.

[66] P. Gong, D. Marceau, and P. Howarth, “Comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data,” *Remote Sensing of Environment*, Vol. 40, No. 2, pp.137-151, 1992.

[67] A. Cliff and J. Ord, *Spatial Processes*, *Models and Applications*, London: Pion, 1981.

[68] L. Anselin, “Local indicators of spatial association - LISA,” *Geographical Analysis*, No. 27, pp. 93-115, 1995.

[69] G. Jacquez and D. Greiling, "Local clustering in breast, lung and colorectal cancer in Long Island, New York," *International Journal of Health Geography*, Vol. 2, No.3, 2003.

[70] D. McGrath and C. Zhang, "Spatial distribution of soil organic carbon concentrations in grassland of Ireland," *Applied Geochemistry*, Vol. 18, pp.1629-1639, 2003.

[71] E. LeDrew, H. Holden, M. Wulder, C. Derksen, and C. Newman, "A spatial statistical operator applied to multidate satellite imagery for identification of coral reef stress," *Remote Sensing of the Environment*, Vol. 91, pp.271-279, 2004.

[72] P. Goovaerts, G. Jacquez, and A. Marcus, "Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies," *Remote Sensing of Environment*, Vol. 95, No.3, pp.351-367, 2005.

[73] A. Getis and J. Ord, “The analysis of spatial association by use of distance statistics,” *Geographical Analysis*, Vol. 24, pp.189-206, 1992.

[74] J. Schott, *Matrix Analysis for Statistics*, 2nd ed., New Jersey: John Wiley & Sons, 2005.

[75] T.Warner and M. Shank, “Spatial Autocorrelation Analysis of Hyperspectral Imagery for Feature Selection,” *Remote Sensing of Environment*, Vol.60, No.1, pp.58-70, 1997.

[76] T.Warner, K. Steinmaus and H. Foote, “An evaluation of spatial autocorrelation

feature selection," *International Journal of Remote Sensing*, Vol.20, No.8, pp.1601-1616, 1999.

[77] J. Read and N. Lam, "Spatial methods for characterising land cover and detecting land-cover changes for the tropics," *International Journal of Remote Sensing*, Vol.23, No.12, pp.2457-2474, 2002.

[78] W. Tobler, "A computer movie simulating urban growth in the Detroit Region," *Economic Geography Supplement*, Vol. 46, pp.234-240, 1970.

[79] M. Wulder and B. Boots, "Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic," *International Journal of Remote Sensing*, Vol. 19, No.11, pp.2223-2231, 1998.

[80] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Processing Magazine*, Vol. 19, No. 1, pp.17-28, 2002.

[81] M. Zortea and V. Haertel, "Experiments on feature extraction in remotely sensed hyperspectral image data," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Vol. 2, pp.964-9672, 2004.

[82] H. Lin and L. Bruce, "Parametric projection pursuit for dimensionality reduction of hyperspectral data," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Vol. 6, pp.3483-34856, 2003.

[83] L. Jiang, B. Mann, and A. Mathur, "Wavelet transform for dimensionality reduction in hyperspectral linear unmixing," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Vol. 6, pp.3513-35156, 2002.

[84] L. Rendell and K. Kira, "A Practical Approach to Feature Selection," *Proceedings of Ninth International Conference on machine learning*, pp.249-256, 1992.

[85] I. Kononenko, "Estimating Attributes: Analysis and Extensions of RELIEF," *Proceedings of seventh European Conference on Machine Learning*, pp.171-182, 1994.

[86] J. Quinlan, *C4.5: Progeams for machine learning*. Morgan Kaufmann, San Mateo, Calif., 1993.

[87] Y. Yang and J. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," *Proceedings of the 14th International Conference on Machine Learning*, pp.412-420, 1997.

[88] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning*, Vol. 46, No. 1, pp.389-422, 2002.

[89] K. Duan, J. Rajapakse, H. Wang, and F. Azuaje, "Multiple SVM-RFE for gene selection in cancer classification with expression data," *IEEE Transactions on NanoBioscience*, Vol. 4 No.3, pp.228-234, 2005.

[90] L. Fu and S. Eun, "Improving reliability of gene selection from microarray functional genomics data," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 7 No.3, pp.191-196, 2003.

[91] T. Millette, "An Expert System Approach to Spectral Band Selection for Remote Sensing Analysis," *10th Annual International Geoscience and Remote Sensing Symposium,* pp.1285-1288, 1990.

[92] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42 No.7, pp.1552-1565, 2004.

[93] T. Tu, C. Chen, J. Wu, and C. Chang, "A fast two-stage classification method for high-dimensional remote sensing data," *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 36, No. 1, pp.182-191, 1998.

[94] Q. Du, "Band selection and its impact on target detection and classification in hyperspectral image analysis," *Advances in Techniques for Analysis of Remotely Sensed Data*, pp.374-377, 2003.

[95] H. Du, H. Qi, and X. Wang, "Band selection using independent component analysis for hyperspectral image processing," *Proc. Applied Imagery Pattern Recognition Workshop*, pp.93-98, 2003.

[96] A. Ifarraguerri and M. W. Prairie, "Visual method for spectral band selection," *IEEE Geoscience and Remote Sensing Letters*, Vol. 1, No. 2, pp.101-106, 2004.

[97] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, Vol. 2 No.2, pp.156-159, 2005.

[98] B. Guo, S. Gunn, B. Damper, and J. Nelson, "Adaptive Band Selection for Hyperspectral Image Fusion Using Mutual Information," *7th International Conference on Information Fusion*, 2005.

[99] P. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*, New York: McGraw-Hill, 1978.

[100] K. Fukunaga, *Statistical Pattern Recognition* (2nd edition), Morgan Kaufman, 1990.

[101] M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy, "Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Vol. 6, pp.2893-28956, 2001.

[102] A. Getis, and J. Ord, "The Analysis of Spatial Association by the Use of Distance Statistics," *Geographical Analysis*, Vol. 24, pp.189-206, 1992.

[103] B. Milenova and M. Campos, "Mining high-dimensional data for information fusion: a database-centric approach," *8th International Conference on Information Fusion*, Vol. 1, pp.7-14, 2005.

[104] J. Jensen, *Introductory digital image processing: A remote sensing perspective*, Prentice-Hall, Englewood Cliffs, NJ, 1996.

[105] Q. Zhang, *Design of an Adaptive Classification Procedure for the Analysis of High-Dimensional Data with Limited Training Samples*, PhD dissertation, Purdue University, 2001.

[106] J. Kittler and J. Foglein, "Contextual Classification of Multispectral Pixel. Data," *Image and Vision Computing*, Vol.2, No.1, pp.13-28, 1984

[107] D. Landgrebe, " The Development of a Spectral-Spatial Classifier for Earth Observational Data," *Pattern Recognition*, Vol. 12, pp. 165-175

[108] R. Kettig and D. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogenous Objects," *IEEE Transactions on*

Geoscience and Remote Sensing, Vol. GE-14, pp. 19-26, 1976

[109] H. Karakahya and B. Yazgan, "A Spatial Filtering Approach to the Classification of Multispectral," *Proceedings of International Conference on Recent Advances in Space Technologies*, pp.359-364, 2003

[110] H. Karakahya, B. Yazgan, and O. Ersoy, "A Spectral-Spatial Classification Algorithm for Multispectral Remote Sensing Data," *Lecture Notes in Computer Science*, pp.1011-1017, 2003.

[111] S. De Jong, T. Hornstra and H. Maas, "An Integrated Spatial and Spectral Approach to the Classification of Mediterranean Land Cover Types: the SSC Method," *Journal of Applied Geosciences*, Vol. 3, No. 2, pp.176-183, 2001.

[112] R. Haralick, K. Shanmugam, and I. Dinstein "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 6, pp. 610-621, 1973.

[113] 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 No.1, pp.93-97, 2006.

[114] S. Li, *Markov Random Field modeling in computer vision*, Springer-Verlag, 1995.

[115] T. Kasetkasem and P. Varshney, "An Image Change Detection Algorithm Based on Markov Random Field Models." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 8, pp. 1815-1823, 2002.

[116] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, Vol. 68, pp. 259–302, 1986.

[117] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, Vol. 36, pp.192-236, 1974.

[118] H. Elliott, H. Derin, R. Cristi, and D. Geman, "Application of the Gibbs distribution to image segmentation," *IEEE Proceedings of the International Conference on Acoustic, Speech and Signal Proceeding*, Vol. 9, pp. 678-681, 1984.

[119] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No.1, pp.39-55, 1987

[120] H. Kwon, S. Der, and N. Nasrabadi, "An adaptive segmentation algorithm using iterative local feature extraction for hyperspectral imagery," *Proceedings of International Conference on Image Processing*, Vol. 1, pp.74-771, 2001.

[121] P. Goovaerts, "Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data," *Journal of Geographical Systems*, Vol. 4, No. 1, pp. 99-111, 2002.

[122] B. Calnek, H. J. Barnes, C. Beard, W. Reid, and H. Yoder, *Diseases of Poultry*, Iowa State University Press, 1991.

[123] B. Albers, J. DiBenedetto, S. Lutz, and C. Purdy, "More efficient environmental monitoring with laser-induced fluorescence imaging," *Biophotonics International Magazine*, Vol. 2, No. 6, pp.42-54, 1995.

[124] C. Shah, P. Watanachaturaporn, P. Varshney, and M. Arora, "Some recent results on hyperspectral image classification," *Proceedings of 2003 IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp.346-353, 2003.

[125] F. Moreau, S. Moreau, D. M. Hueber, and T. Vo-Dinh, "Fiberoptic Remote Multisensor System Based on AOTF," *Applied Spectroscopy*, Vol. 50, No. 10, 1996.

[126] T. Vo-Dinh, D. L. Stokes, M. Wabuyele, M.E. Martin, J. M. Song, R. Jagannathan, E. Michaud, R. J. Lee, and X. Pan "Hyperspectral Imaging System for In Vivo Optical Diagnostics", *IEEE Engineering in Medicine and Biology*, Vol. 23, pp.40-59, 2004.

[127] M. Martin, M. Wabuyele, M. Panjehpour, B. F. Overholt, R. Denovo, S. Kennel, G. Cunningham, and T. Vo-Dinh "An AOTF-Based Dual-modality Hyperspectral Imaging System Capable of Simultaneous Fluorescence and Reflectance Imaging," *Medical Engineering & Physics*, Vol. 2, No. 2, pp.149-155,2005.

[128] *Cancer Facts and Figures*, American Cancer Society, Vol. 19, 2002.

[129] S. Tadjudin, *Classification of high dimensional data with limited training samples*, Ph.D. dissertation, Purdue University, 1998.

[130] J. Gualtieri, S. Chettri, R. Cromp, and L. Johnson, "Support vector machine classifiers as applied to AVIRIS data," *Proceedings of Eighth JPL Airborne*

Geoscience Workshop, 1999.

Appendix

Appendix: Hyperspectral Image Toolbox

Hyperspectral imaging is a technique that combines conventional imaging and spectroscopy to acquire both spatial and spectral information from an object. Hyperspectral imaging produces three-dimensional images. The third dimension contains spectral (or wavelength) information for each pixel on the hyperspectral image cube. Because of this combined feature of imaging and spectroscopy, hyperspectral imaging can enhance and expand our capability of detecting some chemical constituents in an object as well as their spatial distributions. Hyperspectral imaging has been used in a wide range of scientific and industrial fields including space exploration; remote sensing for environmental mapping, geological search or mineral mapping, atmospheric composition analysis and monitoring, military target detection or recognition.

To help researchers using hyperspectral imagery, a MATLAB toolbox for analyzing the hyperspectral image is very necessary. The Hyperspectral Image Toolbox incorporates with both standard algorithms for hyperspectral image analysis and also includes original work in hyperspectral band selection and classification. Figure 1 shows the schematic diagram of the Hyperspectral Image Toolbox.

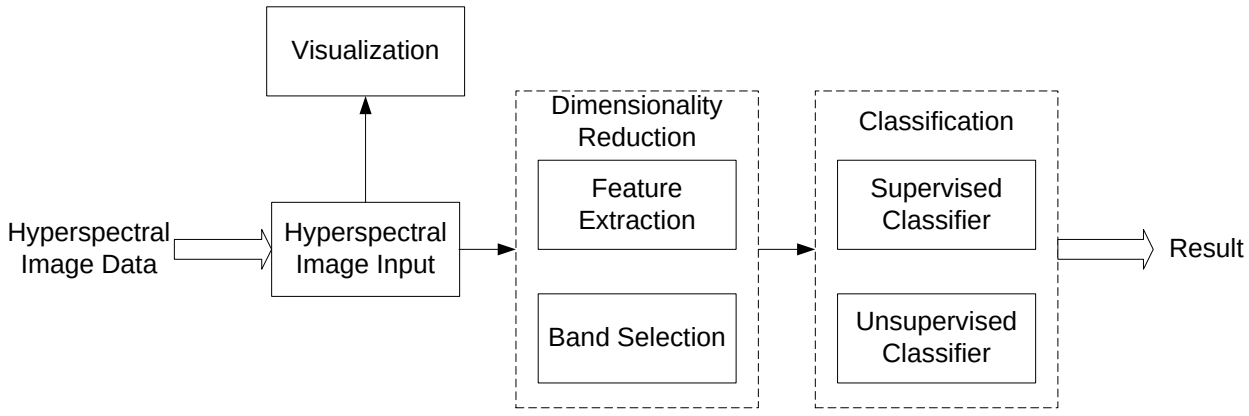


Figure 1: Schematic diagram

VITA

Zheng Du was born in Wuhan, China, on April 24, 1978. After finishing high school in 1996, he entered the Huazhong University of Science and Technology, where he received the Bachelor and Master degree in Engineering in June 2000 and June 2003.

Upon earning his Master's degree, Zheng Du entered the doctoral program in computer engineering at the University of Tennessee, Knoxville. His research interests include hyperspectral image analysis, pattern recognition, statistical learning, machine learning and support vector machines with applications in hyperspectral image. Zheng Du received his Doctor of Philosophy degree with a major in Computer Engineering in August 2008.