NON-DESTRUCTIVE TESTING OF TEXTURED FOODS BY MACHINE VISION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

PELİN BERİAT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

FEBRUARY 2009

Approval of the Graduate School of Informatics

_____

Prof. Dr. Nazife BAYKAL

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Yasemin YARDIMCI

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Prof. Dr. Yasemin YARDIMCI

Supervisor

Examining Committee Members

Yrd. Doç. Dr. Erhan EREN	(METU, II)	_____
Prof. Dr. Yasemin YARDIMCI	(METU, II)	_____
Dr. Habil KALKAN	(METU, II)	_____
Dr. Hamide ŞENYUVA	(TÜBİTAK)	_____
Yrd. Doç. Alptekin TEMİZEL	(METU, II)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this wok.

Name, Last Name : Pelin BERİAT

Signature : _____

ABSTRACT

NON-DESTRUCTIVE TESTING OF
TEXTURED FOODS
BY MACHINE VISION

Beriat, Pelin
M.Sc., Department of Information Systems
Supervisor: Prof. Dr. Yasemin Yardımcı

February 2009, 87 pages

In this thesis, two different approaches are used to extract the relevant features for classifying the aflatoxin contaminated and uncontaminated scaled chili pepper samples: Statistical approach and Local Discriminant Bases (LDB) approach. In the statistical approach, First Order Statistical (FOS) features and Gray Level Cooccurrence Matrix (GLCM) features are extracted. In the LDB approach, the original LDB algorithm is modified to perform 2D searches to extract the most discriminative features from the hyperspectral images by removing irrelevant features and/or combining the features that do not provide sufficient discriminative information on their own. The classification is performed by using Linear Discriminant Analysis (LDA) classifier. Hyperspectral images of scaled chili peppers purchased from various locations in Turkey are used in this study. Correct classification accuracy about 80% is obtained by using the extracted features.

Keywords: Aflatoxins and Food Safety, Machine Vision, Texture Analysis, Local Discriminant Bases, Hyperspectral Imaging.

ÖZ

MAKİNE GÖRME TEKNİĞİ İLE
DESENSEL GIDALARIN
TAHRİBATSIZ TEST EDİLMESİ

Beriat, Pelin
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Danışmanı: Prof. Dr. Yasemin Yardımcı

Şubat 2009, 87 sayfa

Bu tezde, aflatoksinli ve aflatoksinsiz kırmızı pul biberler örneklerinin sınıflandırılmasında kullanılacak gerekli özniteliklerin çıkarılması için iki farklı yaklaşım kullanılmıştır: İstatistiksel yaklaşım ve Yerel Ayırtaç Tabanları yaklaşımı. İstatistiksel yaklaşımda, Birinci Dereceden İstatistiksel öznitelikler ve Gri Düzey Eş Oluşum Matrisi öznitelikleri çıkarılmıştır. Yerel Ayırtaç Tabanları yaklaşımında, hiperspektral veriden gereksiz öznitelikler atılarak ve/veya tek başlarına nitelikli ayırt edici bilgi sağlamayan öznitelikleri birleştirerek en ayırt edici özniteliklerin elde edilmesi amacıyla, orjinal Yerel Ayırtaç Tabanları algoritması 2B arama yapacak şekilde değiştirilmiştir. Sınıflandırmada Doğrusal Ayırtaç Analizi sınıflayıcısı kullanılmıştır. Bu çalışmada Türkiye'nin farklı şehirlerinden satın alınan pul kırmızı biberlerin hiperspektral görüntüleri kullanılmıştır. Çıkarılan öznitelikler kullanılarak, yaklaşık %80 lik bir sınıflandırma doğruluğu elde edilmiştir.

Anahtar Kelimeler: Aflatoksin ve Gıda Güvenliği, Makine Görmesi, Desen Analizi, Yerel Ayırtaç Tabanları, Hiperspektral Görüntüleme.

This thesis is dedicated to:
My Family and My Fiancé

For their endless support …

ACKNOWLEDGEMENT

I thank my supervisor Prof. Dr. Yasemin Yardımcı for her advice, encouragement and support in supervising this thesis. I would also like to thank Dr. Habil Kalkan for his suggestions and comments throughout this thesis.

I would like to thank Asst. Prof. Dr. Erhan Eren, Asst. Prof. Dr. Alptekin Temizel, and Dr. Hamide Şenyuva for reviewing my work.

In addition, I appreciate the kindness of administrative staff of the Informatics Institute.

Very special thanks go to my fiancé Özkan Bayraktar for his endless support, encouragement, and suggestions since the beginning of this research. He was here for me whenever I needed. I also thank his family for their encouragement during this study.

Finally, I express sincere appreciation to my family for their love and endless support.

This thesis work is supported by The Scientific and Technical Research Council of Turkey (TÜBİTAK) by the project of 106E057: “Food safety with non-invasive methods”.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BGYF	: Bright Greenish Yellow Fluorescence
CCD	: Charge Coupled Device
CFS	: Correlation-based Feature Selection
ELISA	: Enzyme-Linked Immunosorbent Assay
FAO	: Food Agricultural Organization
FDB	: Fisher Distance Based
FOS	: First Order Statistical
FWHM	: Full width at half maximum
GLCM	: Gray Level Cooccurrence Matrix
HDGD	: High Dimensional Generalized Discriminant
HPLC	: High Performance Liquid Chromatography
IARC	: International Agency for Research on Cancer
KNN	: K-Nearest Neighbor
LDA	: Linear Discriminant Analysis
LDB	: Local Discriminant Bases
MRF	: Markov Random Field
PCA	: Principle Component Analysis
ppb	: Parts per billion
SAR	: Synthetic Aperture Radar

SVM : Support Vector Machine

TLC : Thin Layer Chromatography

UV : Ultraviolet

CHAPTER 1

INTRODUCTION

Accessing high quality and safe food is one of the most important public concerns in recent decades. Food safety implies that the food includes safe levels of the substances such as toxins and contaminants injurious to human health. Food quality includes characteristics such as nutritional value and texture for the food to be preferred by the consumers. Since safety and quality of food directly affect human health and quality of life, countries have established strict standards. It is essential for the countries to assure the safety and quality of the imported food to protect their domestic consumers. In addition, they pay attention to exported food safety and quality because of the strong obligations placed by the global market. Diseases arising from the food pose a threat to human health and decrease the economic productivity in respective countries.

Machine vision systems are widely used in automated inspection of food quality (Casasent and Chan, 2004), (Yao, Hruska, Brown and Cleveland, 2006) and provide non-destructive means to evaluate quality and safety of foods.

Multispectral and hyperspectral imaging are broadly used in machine vision systems, especially in food quality inspection applications. By combining information from various spectral bands, more accurate, consistent and faster results can be achieved.

In this thesis, we aim to extract features for detecting contaminated and uncontaminated chili peppers from the hyperspectral data by using non-destructive methods.

1.1 Machine vision

Machine vision is a technology that merges mechanics, optical instrumentation, electromagnetic sensing, and digital input-output devices. The systems for counting objects, barcode reading and defect investigation are highly dependent on machine vision. It is also used in retail automation, robot vision and medical imaging processes such as radiological processing. It has been widely used for examining, monitoring, and controlling a very broad range of applications including food quality inspection which is the focus of the present study.

1.2 Multispectral / Hyperspectral Analysis

Band pass interference filters transmit a particular band while rejecting all other upper and lower frequencies of an electromagnetic spectrum (Shah, 2006). In conventional imaging, single broad band pass filter, which pass broad band of light such as 375 nm to 425 nm, is used. Spectral imaging includes usage of a single waveband and implemented by a narrow band pass interference filter that passes a narrow band of light as little as 1nm. Multispectral imaging and hyperspectral imaging utilize two or more waveband filters. Different features of the target object can be displayed by different spectra.

1.2.1 Multispectral Analysis

A multispectral image covers several narrow bands of wavelengths of the same area. Multispectral imaging system can be implemented by using a single camera and filter wheel or using multiple cameras with a single band pass interference filter (Shah, 2006).

Analysis of multispectral images is becoming an important task because information that is coming from different portions of the spectrum is combined in multispectral images. By analyzing different wavelengths, information that may not otherwise be visible can be gathered. Consequently, more accurate and consistent results can be obtained with multispectral imaging.

In multispectral systems, images are analyzed using a set of selected filters. However, selection of optimal filter set is a hard task. The band pass interference filters that are used in multispectral analysis can be determined by applying band selection or band combination to the hyperspectral data (Shah, 2006).

1.2.2 Hyperspectral Analysis

Hyperspectral imaging collects information from the electromagnetic spectrum and combines spectroscopy and imaging techniques. The hyperspectral data include both spatial and spectral information (Qin, Burks, Kim, Chao and Ritenour, 2008). In hyperspectral imaging full spectral profile from ultraviolet to infrared is coupled with each image pixel which results in large quantities of data. Hyperspectral image covers a set of adjacent bands. In contrast to multispectral analysis, hyperspectral analysis performs continuous data analysis.

1.2.3 Application areas

There are wide range of application areas of multispectral and hyperspectral imaging. One of the major application areas is remote sensing. Multispectral and hyperspectral imaging can be applied for the identification of scene components in remote sensing field. They are used in vegetation and land-use classification (Bachari, Khodja and Belbachir, 2004), and in water resource studies such as flood detection and water quality applications (Ip et al., 2006), (Tilley, Ahmed, Son and

Badrinarayanan, 2003), (Thiemann and Kaufmann, 2002). Another application of hyperspectral analysis in remote sensing is target detection (Gomez, 2002).

Hyperspectral and multispectral imaging technologies can also be used in medical imaging. They have been used in classification of healthy and pathological tissues (Bonnier et al., 2008) and in cancer detection applications (Martin et al., 2006), (Siddiqi et al., 2008).

Hyperspectral can be used as a predecessor to online multispectral machine vision applications for the selection of important wavelengths. Lee, Kang, Delwiche, Kim and Noh (2008) used hyperspectral imaging to select the optimum wavelengths for detecting defects on apples.

Since multispectral and hyperspectral data includes detailed information, it can be used in product inspection applications (Mahesh, Manickavasagan, Jayas, Paliwal and White, 2008), (Ariana and Lu, 2008).

Hyperspectral analysis is also used in aflatoxin detection applications. Casasent et al. (2004) extracted features from hyperspectral data for the inspection of aflatoxin in whole corn kernels. Yao et al. (2006) used hyperspectral imaging to observe Bright Greenish Yellow Fluorescence (BGYF) under UV light to detect aflatoxin contaminated corn kernels. They took images from the spectral range of 450 nm to 900 nm. They found that emission around 500 nm can be a good indicator for differentiation of BGYF positive and BGYF negative corn kernels.

1.3 Thesis Statement

In this thesis, we extract statistical texture features from the hyperspectral images. We also used 2D LDB algorithm for feature extraction. By using the 2D LDB algorithm, both spatial-frequency and spectral axes are searched to get the location of most discriminative

features. The dimensionality of the feature space is reduced by removing the irrelevant features or by merging the ones that do not provide additional information on their own.

The extracted features are used for the classification of contaminated and uncontaminated chili peppers.

1.4 Thesis Overview

Chapter 2 introduces the problem of aflatoxin in foods and particularly in chili peppers. It gives some brief information about methods utilized for the detection of aflatoxin.

Chapter 3 provides a brief overview on the texture feature extraction methods. It also presents Best Bases algorithm and Local Discriminant Bases algorithm for feature extraction. Some feature selection algorithms and classification methods are also reviewed.

Chapter 4 presents the algorithms for feature extraction. It also introduces the data set characteristics and aflatoxin concentrations. This section also describes the hyperspectral imaging system that is utilized in data acquisition step.

Chapter 5 presents the steps and outputs of feature extraction. The classification results are given for statistical approach and LDB approach are given separately.

Chapter 6 presents conclusions and future works.

CHAPTER 2

AFLATOXINS IN CHILI PEPPERS

Turkey is the third fresh chili pepper producer in the world. Nevertheless, it has less than 3% share in trade of processed pepper and does not have noteworthy force in international market. The main reason of its shortfall is the high level of aflatoxin in the chili peppers that is caused by inadequate hygiene conditions during drying, transport and storage in the production process.

Aflatoxins are toxic compounds produced by many species of Aspergillus molds, especially by Aspergillus flavus and Aspergillus parasiticus (Zeringue and Shih, 1998) and aflatoxin contamination in food is an important food safety issue. Since the aflatoxin level of chili pepper produced in Turkey is often above the acceptable limits of developed countries, it cannot be exported to these countries. Hence, it is crucial to prevent the crops from aflatoxin formation by improving the production processes.

Albeit prevention is important, aflatoxin detection has also a great importance both for the protection of consumers and for overcoming the problems experienced in exportation. As aflatoxin formation is uncontrollable to some extent, detecting and distinguishing contaminated part of the crops is necessary. Most of the existing aflatoxin detection methods are destructive methods that use chemical analysis to the food under inspection. Chemical methods are accurate but expensive and time-consuming, and the samples that are analyzed cannot be consumed after the analysis. On the other hand, although

accuracy of non destructive methods is lower than chemical methods, they are fast and cheaper. They also offer the possibility of being integrated into the chili pepper production band. Therefore, in this thesis, a non-destructive visual method is used for detecting aflatoxins in chili peppers.

2.1 Aflatoxins

Molds are microscopic fungi which grow in warm and humid conditions and they may produce some compounds called mycotoxins under specific conditions (Martins, Martins and Bernardo, 2001). Mycotoxins are secondary metabolites of toxigenic fungi and they affect human, livestock health and international trade. They occur in wide variety of agricultural commodities with high carbohydrate and fat content such as corn, wheat, peanuts, pistachio, hazelnuts, and sorghum and can grow in crops at pre-harvest, harvest, storage, and transportation stages or at later points (Jacobsen and Coppock, 1993), (Bothast and Hesseltine, 1975). High temperature, humidity and moisture contents are the most important factors that increase the mold invasion and toxin production (Bhat and Vasanthi, 2003).

Agricultural products that are contaminated with mycotoxins have many economic consequences such as losses arising from destruction or disposal with lower prices of the contaminated low quality crops, refused shipments, losses occurred as a result of decreased human productivity and health care and treatment expenses besides its effects on human and animal health such as kidney and liver damage and cancer (Bhat and Vasanthi, 2003), (Bennett and Klich, 2003).

Although there are more than 300 mycotoxins, 20 of them are more frequently observed in agricultural commodities. Major mycotoxins include aflatoxins, trichothecenes, patulin and fumonisins (Vrabcheva, 2006). Among mycotoxins, aflatoxins are the best known and most

heavily researched because of their wide occurrence and toxicity (Stroka and Anklam, 2002).

There are four major aflatoxins called aflatoxin B1, B2, G1 and G2 based on their exhibition of fluorescence under ultraviolet (UV) light and aflatoxin B1 is the most frequently occurring and most carcinogenic (Hesseltine, Shotwell, Ellis and Stubblefield, 1966). The structures of these four aflatoxins are given in Figure 2.1.

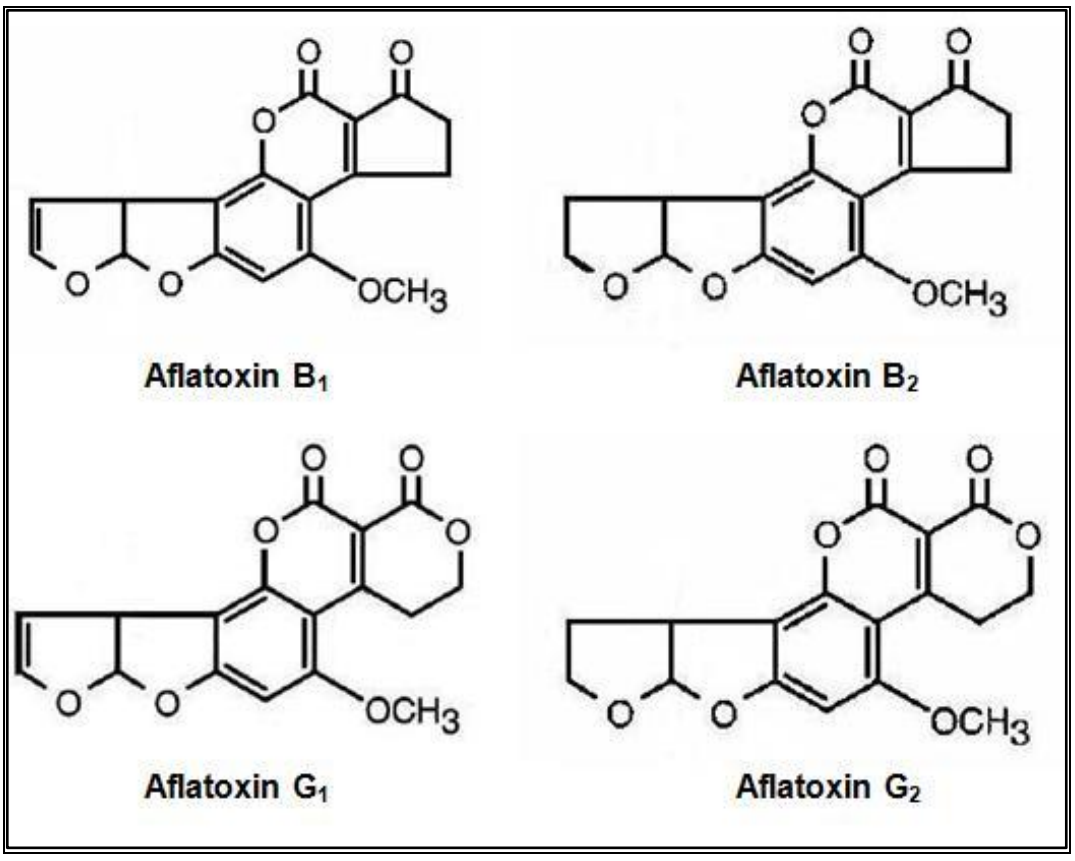

Figure 2.1 – Structures of aflatoxins

2.2 Features That Affect Aflatoxin Production

Aflatoxin contamination can arise as a result of many factors such as drought stress, moisture, temperature, relative humidity, insect damages or other unfavorable weather conditions. Temperature and

moisture content are the most important ones among these factors (Bennett and Klich, 2003).

Besides the storage conditions, drying process is also very important for aflatoxin production for the dried food. Drying the product under the sun takes long and triggers the mold development; therefore, food should be dried as fast as possible in drying machines. Yet world trends favor organic foods and sundrying.

2.3 Effects of Aflatoxins

Aflatoxins have been associated with various diseases throughout the world. They are metabolized in liver; hence liver is the most affected organ. Aflatoxins are known to cause acute liver damages induction of tumors and cirrhosis. They may also cause gene mutations, especially on genes involved in the production of liver cancer (Jacobsen and Coppock, 1993). Occurrence of aflatoxic hepatits in humans have been reported in India, Kenya, and Malaysia (Bhat and Vasanthi, 2003).

The risk of cancer development due to aflatoxin poisoning for a person increases if that person is also carrying Hepatits B virus (Murphy, Hendrich, Landgren and Bryant, 2006), (Bhat and Vasanthi, 2003).

In studies carried out in a West African nation Benin, it was found that children exposed to aflatoxin may become stunted, underweight, and more prone to infectious diseases in childhood and later life (Bhat and Vasanthi, 2003).

2.4 Chilli Peppers and Regulations about Aflatoxins

Chilli peppers are one of the most widely consumed spices in the world. By the year 2000, world total pepper production has reached 18.5 million tons. The amount of productions in tons according to the countries is given in Table 2.1 (Duman, Zorlugenç and Evliya, 2002).

Table 2.1 – Amount of pepper production in countries

Country	**Amount of Production (Tons)**
China	7.683.127
Mexico	1.813.252
Turkey	1.390.000
Spain	936.300
Nigeria	715.000
USA	694.950
Equator	600.000
Indonesia	496.908
Italy	316.209
South Korea	307.000
Total	**18.500.622**

In Turkey, Southeastern Anatolia, Mediterranean and Aegean Regions are the major regions where pepper agriculture is made. 76.5% of Turkey's pepper production is based in Southeastern Anatolia Region. Pepper production is not only essential for Turkey domestic market but also has economic importance due to its export potential (Duman et al., 2002). The pepper production process is summarized in Figure 2.2.

Since prevention of aflatoxin is important because of its potential hazards on humans and animals, legal limits have been established to limit its consumption. Since aflatoxin B1 is the most carcinogen of all aflatoxins, maximum levels are set for aflatoxin B1 content and for the content of total aflatoxins in food separately.

European Union determine the maximum permissible level for dried fruits, processed products for direct human consumption, cereal products, groundnuts and nuts as 4 ppb for total aflatoxins. It is 5 ppb and 10 ppb for foods that are used as ingredient in foodstuffs and spices including chili peppers for aflatoxin B1 and total aflatoxins, respectively (Commission Regulation [EC], 2006). For the United States

of America and Turkey, the permissible limit for total aflatoxins is 20 ppb for all foods, and this limit is rather high with respect to the limits of European Union (Karaman and Acar, 2006).

2.5 Aflatoxin Studies about Chili Pepper Produced in Turkey

There are many studies about aflatoxin contamination levels of the chili peppers produced in Turkey. Results of a study where 40 red-scaled pepper samples sold in Van are investigated for aflatoxin B1 and moisture presence show that amount aflatoxin B1 in these samples is between 1.10 – 44.00 ppb and average moisture level is between 12.85 ± 0.72 (Ağaoğlu, 1999).

According to another study applied on 40 red-scaled and 26 red powder samples from Erzurum, 18.2% of red-scaled peppers and 10.7 of the red powder samples are contaminated with aflatoxin between 1.1 and 97.5 ppb. The highest amount of aflatoxin is found in red-scaled pepper (Erdoğan, 2004).

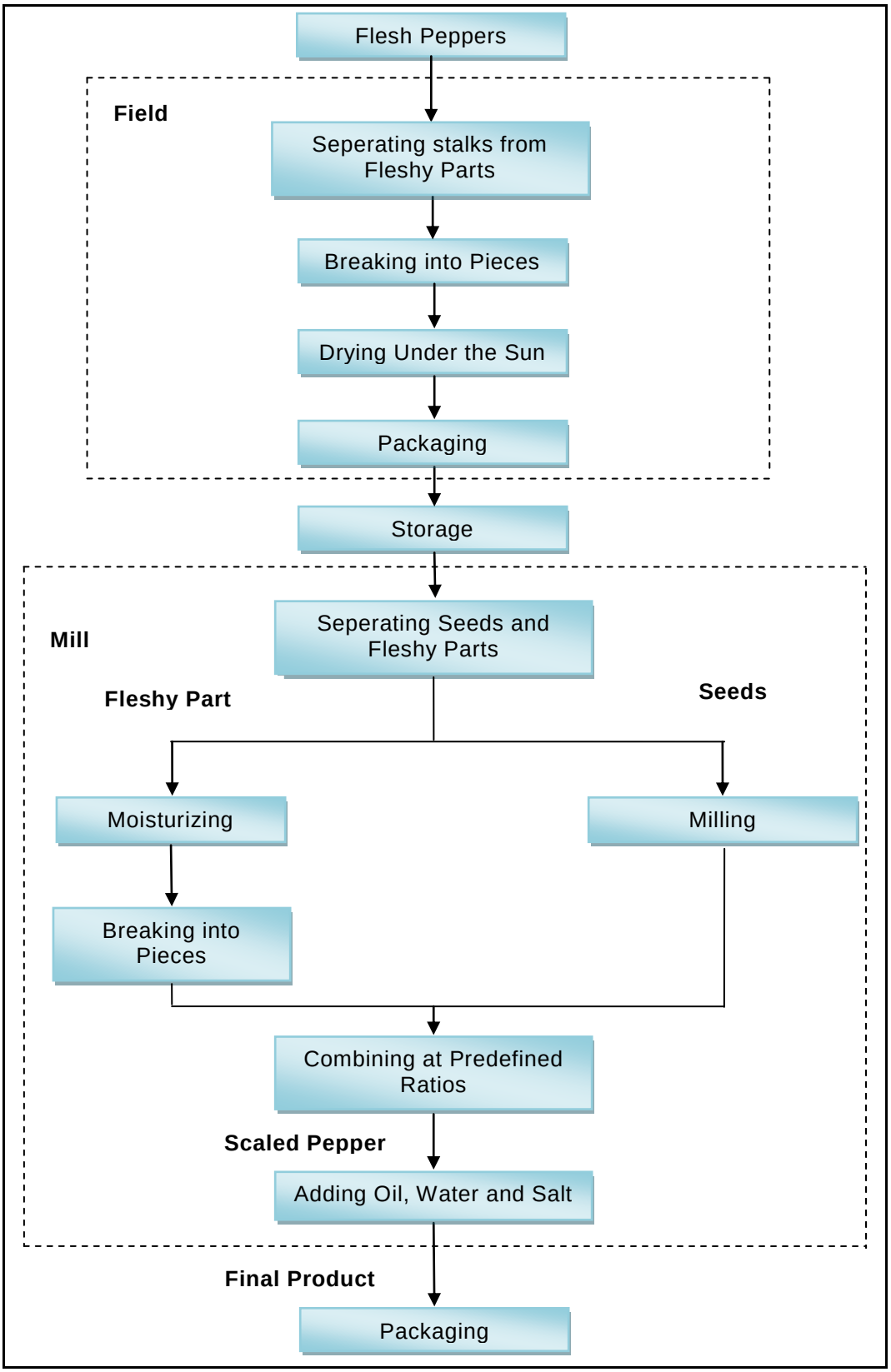

Figure 2.2 – Pepper production process

Another study is carried out on pepper produced in Bursa and Sakarya and results show that 15.4% of the red-scaled peppers produced in Bursa and 50% of the pepper produced in Sakarya are contaminated with aflatoxin between 5.9 – 9.45 ppb and 1.65 – 15.0 ppb, respectively (Üner, Çetin and Ergün, 2000).

In peppers produced in Southeast Anatolian Region, it is found that within 127 samples of various kinds of peppers 90.30% of red-scaled peppers and 72.72% of red powder peppers are contaminated with aflatoxin B1. The average aflatoxin B1 level is 8.429 ppb and 19.975 ppb for red powder pepper and red scaled pepper, respectively (Üner et al., 2000).

2.6 Aflatoxin Detection Methods

There are several methodologies and methods for the detection of aflatoxin. Presently, aflatoxin detection is widely performed by using chemical methods; however we propose to develop vision based method in order to achieve rapid and nondestructive testing.

2.6.1 Chemical Methods

Various chemical methods have been developed for the detection of aflatoxin. Thin Layer Chromatography (TLC), High Performance Liquid Chromatography (HPLC) and Enzyme-Linked Immunosorbent Assay (ELISA) are the most commonly used means for detecting aflatoxin.

TLC was the first method used in aflatoxin detection (Stroka and Anklam, 2002). It can identify the presence or absence of aflatoxin and also quantitatively determine aflatoxin B1, B2, G1 and G2 separately at levels as low as 1 ppb. It is slow and fairly expensive method but it is precise and gives accurate results (Jacobsen and Coppock, 1993).

HPLC is potentially the most sensitive method for aflatoxin testing. It can identify aflatoxin B1, B2, G1 and G2 or their metabolites accurately and quantitatively by either using UV absorbance or fluorescence detectors. It is an expensive method and requires technical expertise and extensive sample clean-up (Jacobsen and Coppock, 1993).

ELISA is an immunochemical method and it is the most widely used test to detect aflatoxins. It is a simple, adaptable and sensitive method and it can identify and measure aflatoxins in food in less than 10 minutes.

All these methods are not only expensive but also destructive. The analyzed substances cannot be consumed after the analysis and test results cannot give information about which parts of the analyzed sample are indeed contaminated. However, visual methods are nondestructive and samples can be consumed after the analyses.

2.6.2 Visual Methods

Bright Greenish Yellow Fluorescence (BGYF) test has been used as a nondestructive method in detection of aflatoxin. It is also known as black light test.

Initially, the association between BGYF and infections of A. flavus was discovered in raw cotton and BGYF is associated with aflatoxins after A. flavus is found out to produce aflatoxins (Bollenbacher and Marsh, 1954) (Ashworth and McMeans, 1966) (Doster and Michailides, 1998). These researches serve as a basis for the development of BGYF test. This test was applied in many crops such as wheat, barley, rice, peanuts, soybeans, white corn and yellow corn (Bothast and Hesseltine, 1975), (Hadavi, 2005), (Steiner, Brunschweiler, Leimbacher

and Schneider, 1992), (Doster and Michailides, 1998) to detect aflatoxin contamination.

A. flavus produces kojic acid that is converted to BGYF compound by peroxidase in the plant. Therefore, BGYF is dependent on the presence of peroxidase (Wicklow, 1999), (Hadavi, 2005). The BGY particles can be detected my machine vision under UV light.

In BGYF test, crops are examined under long-wave UV light and the number of BGY particles is taken as an indication of aflatoxin contamination in a qualitative, but not in a quantitative manner.

BGYF is only a characteristic of living cells. This leads a problem that is referred as aflatoxin false negatives where there are non-BGYF kernels that are highly contaminated with aflatoxins. Since BGYF is dependent on the presence of peroxidase, this can occur when there is no peroxidase activity but the crop is contaminated in post-harvest stages where the conditions for fungi growth are suitable (Hadavi, 2005). Another problem arise from the crops that produce kojic acid but do not produce aflatoxin and exhibit BGYF and referred as aflatoxin false positives. The reason can be that there is not enough time for the toxin production or the fungus has given up the site to other competitive microorganisms such as A. niger which is a more aggressive species later (Hadavi, 2005).

Spectroscopy methods are also used in aflatoxin detection problems. Pearson et al. (2004) used spectral absorbance at 750 and 1200 nm to distinguish the aflatoxin contaminated (>100 ppb) and uncontaminated yellow corn kernels.

CHAPTER 3

LITERATURE REVIEW

3.1 Texture Analysis

Texture is an important characteristic for the analysis of many kinds of images and also has an essential role in many computer vision and image processing applications by including rich source of visual information. It can be defined as the set of local neighborhood properties of the gray levels and a measure of spatial variations of intensities that creates patterns in the image data such as repetitiveness, fineness, coarseness, randomness, directionality and granularity (Fan and Xia, 2003), (Livens, Scheunders, van de Wouwer and Van Dyck, 1997), (Haralick, 1979).

Feature extraction is the first step of texture analysis. Extracted features can be used to solve four major texture analysis problems: texture segmentation, texture classification, texture synthesis and shape from texture (Tuceryan and Jain, 1993).

Texture segmentation:

Texture segmentation is an important topic in image processing and computer vision. It aims to identify the regions in an image and to partition the image into regions with similar characteristics or patterns (Sheshadri and Kandaswamy, 2006).

Two main approaches to texture segmentation are region-based and boundary-based approaches. Region-based approach is concerned with recognition of regions with the same texture characteristics; on the other hand, boundary-based approach is based on the detection of differences in texture in neighboring regions and results in a boundary map of the image (Tuceryan and Jain, 1993).

Texture classification:

The goal of texture classification is to assign an image region into one of the predefined texture classes (Materka and Strzelecki, 1998). Texture classification is a two phased process: learning phase and recognition phase. In the learning phase, texture content of each texture class existing in training data are modeled by using the extracted features that characterize the textural properties of the image. In the recognition phase, a classification algorithm is utilized to compare the features of an unknown sample with training data features and to assign new sample to one of the texture classes (Materka and Strzelecki, 1998).

Texture synthesis:

Texture synthesis deals with building the model of texture and use the model for generating textures that are similar to real ones. It is generally applied in computer graphics, computer vision and image processing applications (Fan and Xia, 2003).

Shape from texture:

The shape from texture problem aims to extract three-dimensional surface orientations by measuring texture

distortions of a two-dimensional image (Tuceryan and Jain, 1993).

3.1.1 Texture Feature Extraction

There are four basic texture feature extraction methods mentioned in the literature. These are statistical, structural, model–based and signal processing methods (Materka and Strzelecki, 1998), (Tuceryan and Jain, 1993).

3.1.1.1 Statistical Methods

Statistical methods try to represent the texture by analyzing the spatial distributions and relationships between the gray levels of an image. Textures are characterized according to statistical measures calculated from the intensity values of pixels. By using statistical operators to those pixels, texture feature descriptors are calculated. Texture feature descriptors can be classified into two groups: first-order texture features and second order texture features (Konak; 2002). First order texture features do not take into account the locations and interactions of pixels, they estimate properties of individual pixels. In contrast, second or higher order texture features consider the relative locations of pixels.

The most widely used statistical features are first-order histogram based features, cooccurrence matrix based features, autocorrelation based features, and features derived from run length matrix.

3.1.1.1.1 First-Order Histogram Based Features

Gray level histogram can represent useful information of an image and it is also simple and fast to compute statistical tool for image analysis. Let image of dimensions N and M is a function $f(x,y)$ with spatial coordinates x and y where $x=0,1,..,N-1$ and $y=0,1,..,M-1$. The value of

f at coordinates *(x,y)* is called the intensity or gray level of image at that point which is noted as *i* where *i=0,1,..,L-1* and *L* is the total number of gray levels in the image. The gray level histogram is a function that shows how many times a particular gray level appears in an image:

$$h(i) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \delta(f(x,y), i) \qquad \text{(Equation 3.1)}$$

where δ is the Kronecker delta function

$$\delta(j,i) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \qquad \text{(Equation 3.2)}$$

A normalized histogram, which gives an estimate of probability of occurrence of a particular gray level can be obtained by dividing *h(i)* by the total number of pixels in the image (Materka and Strzelecki, 1998).

First-order histogram based features include central moments such as mean, variance, skewness and kurtosis and other parameters such as energy and entropy. Equations of these features are given below, where *p(i)= h(i)/NM, i=0,1,..,L-1*.

Mean: $$\mu = \sum_{i=0}^{L-1} ip(i) \qquad \text{(Equation 3.3)}$$

Variance: $$\sigma^2 = \sum_{i=0}^{L-1} (i-\mu)^2 p(i) \qquad \text{(Equation 3.4)}$$

Skewness: $$\mu_3 = \sigma^{-3} \sum_{i=0}^{L-1} (i-\mu)^3 p(i) \qquad \text{(Equation 3.5)}$$

Kurtosis: $\mu_4 = \sigma^{-4} \sum_{i=0}^{L-1} (i - \mu)^4 p(i) - 3$ (Equation 3.6)

Energy: $E = \sum_{i=0}^{L-1} [p(i)]^2$ (Equation 3.7)

Entropy: $H = -\sum_{i=0}^{L-1} p(i) log_2 [p(i)]$ (Equation 3.8)

Mean gives the average intensity level of the image. Variance is a measure of dispersion of the image data around the mean intensity value. Skewness is a measure of degree of asymmetry of the image histogram. If skewness is zero, the histogram is symmetric around the mean; otherwise it is skewed above or below the mean. The kurtosis is an indication of peakedness of the histogram. In order to obtain a Gaussian shaped histogram, kurtosis is normalized to zero by subtracting 3 from the kurtosis equation above (Materka and Strzelecki, 1998). Entropy is a measure of uniformity of the histogram.

3.1.1.1.2 Gray Level Cooccurrence Matrix (GLCM) Based Features

Gray level cooccurrence or in other name spatial gray-tone dependence method is probably the most cited statistical texture analysis method in the literature (Haralick, 1979), (Tuceryan and Jain, 1993), (Arrowsmith, Varley, Picton and Heys, 1999), (Castellano, Bonilha, Li and Cendes, 2004), (Bharati, Liu and MacGregor, 2004), (Smith, Wooster, Powell and Usher, 2002), (Lam, 1996). This approach is based on the spatial dependency among image pixels (Kramer and Aghdasi, 1999). The relationships between neighboring pixels are transformed into a matrix called Gray Level Cooccurrence Matrix (GLCM). GLCMs are constructed to record the relative frequencies of each pixel pair *Pij* with different gray levels *i* and *j* that are separated by a specified

neighborhood distance *d* and direction *θ*. The most commonly used directions are *θ* = 0°, 45°, 90°, 135° (Lam, 1996). GLCM for inverse directions can be calculated by taking the direction parameter as (*θ* + 180) for each *θ* given above. The representation of the four commonly used directions can be seen from the Figure 3.1 below.

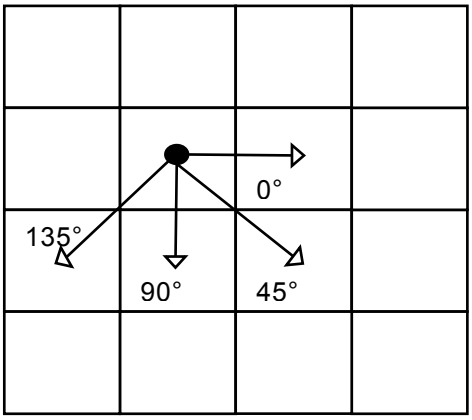

Figure 3.1 – Representation of direction parameter for d = 1

One GLCM is calculated for each direction to represent the spatial dependency of that direction. Resulting GLCM is a square matrix and row and column dimensions are equal to the number of discrete gray levels of the image.

Determination of the neighborhood distance parameter *d* is an important decision before generating the GLCM. Classification of coarse textures requires distance parameter to be smaller and fine textures require larger values of distance parameter.

Symmetric or non-symmetric GLCMs can be obtained. In general, if the orientation of objects is important, a non-symmetric matrix can be calculated for the directions of interest. However, if orientation is not important, a symmetric matrix can be obtained by simply adding the matrix calculated for the direction *θ* with its transpose which provides

a means to take into account the inverse direction (θ + 180) (Haralick, 1979), (Konak; 2002).

Haralick, Shanmugam and Dinstein (1973) have proposed some features that can be computed from the GLCM for the purpose of texture discrimination (Materka and Strzelecki, 1998).

Homogeneity: $$\sum_i \sum_j \frac{P(i,j)}{1+|i-j|} \quad \text{(Equation 3.9)}$$

Energy: $$\sum_i \sum_j P(i,j)^2 \quad \text{(Equation 3.10)}$$

Contrast: $$\sum_i \sum_j (i-j)^2 P(i,j) \quad \text{(Equation 3.11)}$$

Entropy: $$-\sum_i \sum_j P(i,j) log P(i,j) \quad \text{(Equation 3.12)}$$

Correlation: $$\sum_{\mathrm{i}} \sum_{\mathrm{j}} \frac{(i-\mu_x)(j-\mu_y)P(i,j)}{\sigma_x \sigma_y} \quad \text{(Equation 3.13)}$$

Inverse Difference: $$\sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j) \quad \text{(Equation 3.14)}$$

Maximum Probability: $$\max_{i,j} P(i,j) \quad \text{(Equation 3.15)}$$

where *P(i,j)* is the *(i,j)*th element of the normalized GLCM matrix and $p(i) = \sum_j P(i,j)$ and $p(j) = \sum_i P(i,j)$. μ_x and μ_y are the means and σ_x and σ_y are the standard deviations of $p(x) = \sum_j P(x,j)$ and $p(y) = \sum_i P(i,y)$, respectively. Homogeneity and inverse difference moment are a measure of uniformity. A diagonal GLCM is obtained when the homogeneity is equal to 1. Energy is a measure of pixel pair repetitions. As energy gets closer to zero, it indicates an increase of

change from one pixel to the next. Correlation is a measure of gray level linear-dependencies.

There are some difficulties encountered while using GLCM method. First problem occurs when the number of distinct gray levels in the image is high. Since the resulting matrix contains L^2 dimensions for L distinct gray levels, texture analysis takes too much time. To overcome this difficulty, number of gray levels can be reduced by applying quantization which also causes some loss of textural information. Another difficulty is about the selection of distance parameter because of the lack of well established methods (Tuceryan and Jain, 1993).

3.1.1.1.3 Autocorrelation Based Features

Autocorrelation is a measure of coarseness of an image and gives the regularity of texture primitives of the image (Tuceryan and Jain, 1993). The autocorrelation function of an image *f(x,y)* with *M,N* dimensions can be defined as follows where *p* and *q* are positional differences:

$$\rho(p,q) = \frac{MN}{(M-p)(N-q)} \frac{\sum_{x=1}^{M-p}\sum_{y=1}^{N-q} f(x,y)f(x+p,y+q)}{\sum_{i=1}^{M}\sum_{j=1}^{N} f^2(x,y)} \quad \text{(Equation 3.16)}$$

If the texture is coarse, the autocorrelation function decreases slowly with increasing distance; otherwise it drops off rapidly. For regular textures function exhibits peaks and valleys (Wan et al., 2004).

3.1.1.1.4 Run length Matrix Based Features

Run length of an image can be defined as the length of consecutively identical gray levels in a specified direction. The most recently used directions that are used are θ = 0°, 45°, 90°, 135°. Four matrices are calculated for these four directions (Haralick, 1979). The rows of run

length matrix indicate the gray level values in the image and the columns indicate the run size for each gray level.

Various texture features can be calculated from the run length matrix. Some of them are short run emphasis, long run emphasis, gray level nonuniformity, run length nonuniformity and run percentage.

Short Run Emphasis: $$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)} \quad \text{(Equation 3.17)}$$

Long Run Emphasis: $$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2 p(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)} \quad \text{(Equation 3.18)}$$

Gray Level Nonuniformity: $$\frac{\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_r}p(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)} \quad \text{(Equation 3.19)}$$

Run Length Nonuniformity: $$\frac{\sum_{j=1}^{N_r}\left(\sum_{i=1}^{N_g}p(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)} \quad \text{(Equation 3.20)}$$

Run Percentage: $$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}jp(i,j)} \quad \text{(Equation 3.21)}$$

where *p(i,j)* is the number of times there is a run of length *j* for gray level *i*, N_g is the number of gray levels and N_r is the number of runs (Tang, 1998). Short and long run emphasis gives the proportion of short runs and long runs in the image, respectively.

3.1.1.2 Structural Methods

Structural methods consider texture as a composition of texture primitives which have nearly repetitive and regular spatial arrangements. They attempt to describe the texture primitives and their placement rules (Tuceryan and Jain, 1993). The first step of

structural methods is the determination of the primitives. A primitive is a connected set that has a list of attributes. For example, the pixel is a primitive with its attribute gray level. Many kinds of primitives such as connected components can be calculated by applying neighborhood operators. The second step is selecting a spatial relationship from the data about texture primitives such as center coordinates and attributes (Haralick, 1979).

Structural methods provide a symbolic description of an image; however their practicality is limited to regular textures. Since most of the natural textures exhibits random behaviour, these methods are more suitable for texture synthesis rather than texture analysis.

3.1.1.3 Model-Based Methods

Model based methods attempt to construct a generative image model that can be used to interpret the image texture and create the observed intensity distribution. The parameters of the model are estimated and these parameters can be used for image analysis or for image synthesis. The model can be considered as a combination of a function that represents the structural information on image and an additive random noise sequence (Tuceryan and Jain, 1993). The major disadvantage of model based methods is the computational complexity of estimating model parameters.

The most widely used model based methods are Markov Random Fields (MRF) and Fractals. Markov Random Field is a probabilistic process which assumes that the intensity of a pixel is determined from the intensities of pixels within its neighborhood (Materka and Strzelecki, 1998). MRF models have been mainly used in texture synthesis and texture segmentation applications (Paget and Longstaff, 1998), (Sinha and Gupta, 2007), (Çeşmeli and Wang, 2001).

A deterministic fractal can be defined by using the concept of self-similarity. *A* bounded set *A* is self-similar when *A* is the union of *N* distinct scaled down and non-overlapping copies of itself. The main difficulty in using fractals is that natural textures show statistical but not deterministic variations as above (Tuceryan and Jain, 1993).

3.1.1.4 Signal Processing Methods

Signal processing methods are used to analyze the frequency content of the image. These methods include spatial domain filters, Fourier, Gabor and Wavelet transforms.

3.1.1.4.1 Spatial Domain Filters

Spatial domain filters are generally utilized to solve edge detection and texture segmentation problems. For edge detection, various masks have been designed such as Robert's or Laplacian operators. Measure of edgeness is computed by applying these masks to the image.

$$\text{Roberts Operators}: \quad M_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad M_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad \text{(Equation 3.22)}$$

$$\text{Laplacian Operator}: \quad L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \qquad \text{(Equation 3.23)}$$

Spatial moments are another set of spatial domain filters. The $(p+q)^{th}$ moments over a region *R* of image *f(x,y)* is given as:

$$m_{pq} = \sum_{x,y \in R} x^p y^q f(x,y) \qquad \text{(Equation 3.24)}$$

which corresponds to filtering the image with a set of spatial masks. Input image is filtered with a set of spatial masks and the resulting images can be used as texture features for texture segmentation (Tuceryan and Jain, 1993).

3.1.1.4.2 Fourier and Gabor Transforms

Transform-based methods are based on Fourier, Gabor, and Wavelet transforms. In texture analysis, the most widely used method is the Wavelet transform.

The Fourier transform deals with global frequency content of an image and provides representation of the image in the frequency domain. Fourier transform based methods perform only well on images that have strong periodicity and directionality. Since Fourier transform do not provide any reference about spatial localization, it may perform poorly in practice. Besides that, these methods have high computational complexity that makes them inefficient for many applications (Konak, 2002).

In order to provide local spatial-frequency analysis, Short Time Fourier Transform that is also known as Windowed Fourier Transform is performed. The image is convolved with a window function that is localized in spatial-frequency domains and the Fourier transform is computed. By sliding the window and repeating the same process, the local frequency contents of the image is obtained (Sinha, Routh, Anno and Castagna, 2003). Gabor transform is a special case of Short Time Fourier Transform where a Gaussian space localizing window is used in order to extract local information from image. To analyze low frequency content, a wide window in space is preferred. In the same way, to analyze high frequency content a narrower window is preferred (Pei and Ding, 2007), (Gabor, 1946), (Ogden, 1997). However, there is usually no single filter resolution that enables the localization of spatial structure in natural texture which limits the applicability of the Gabor Transform. In addition, the features extracted from Gabor transform based methods are highly correlated with each other because Gabor functions are not mutually orthogonal. Also, Gabor transforms are

usually not reversible which makes their applications in texture synthesis limited (Arivazhagan and Ganesan, 2003).

3.1.1.4.3 Wavelet Transform

Wavelet transform is based on multiresolution theory that is concerned with the representation and analysis of signals or images at more than one resolution which lets features that are undetectable in one resolution become detectable in another resolution (Gonzalez and Woods, 2002).

The input image is decomposed into subbands by passband filtering.

A representation of decomposition and reconstruction of a signal is given in Figure 3.2 where h and g are the high and low pass filters, and *H* and *G* are the convolution-subsampling operators using filters *h* and *g,* respectively. If these operators satisfy the orthogonality conditions:

$$HG^* = GH^* \quad and \quad H^*H + G^*\mathrm{G} = I \qquad \text{(Equation 3.25)}$$

where *I* is identity operator, perfect reconstruction is possible (Saito and Coifman, 1995). Reconstruction of the original image is achieved by upsampling, filtering and summing the individual subbands.

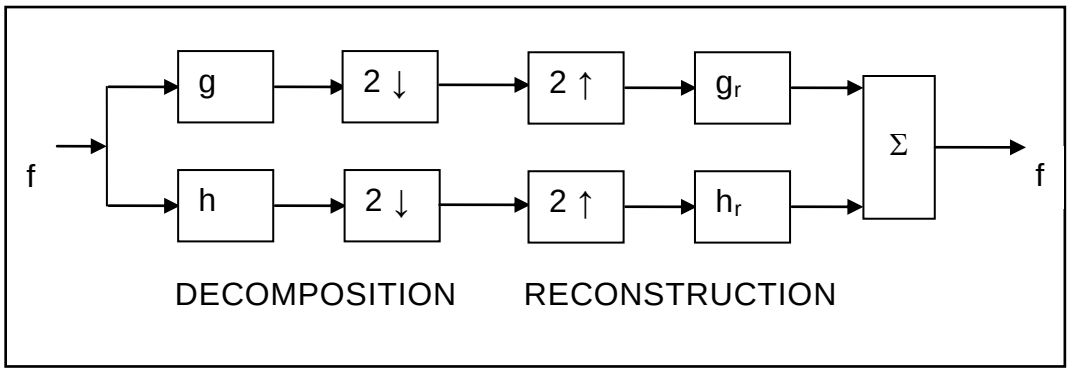

Figure 3.2 – Wavelet based signal decomposition and reconstruction

The wavelet decomposition can be obtained by convolving the image with a set of wavelet basis functions which are usually orthogonal that is, none of the basis functions can be represented as a combination of other basis functions.

Low pass and high pass filters are applied to the rows and columns of the image in order to obtain the wavelet decomposition. By highpass filtering in both directions (HH) gives the diagonal details of the image. HL represents the horizontal details and LH gives the vertical details. LL corresponds to low frequencies.

Choice of wavelet filter has not been considered in this study. Most widely used wavelet types are Haar, Daubechies, Coiflets, Symlets and Butterworth wavelets. All of the wavelets mentioned above are orthogonal. We used Daubechies wavelets in this thesis.

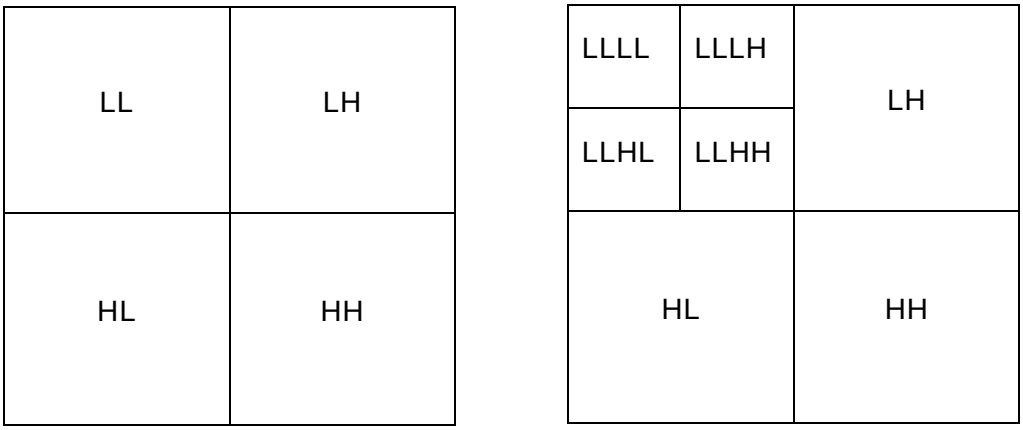

Figure 3.3 – Wavelet decomposition of two dimensional images

In each step of the wavelet decomposition, four subimages are produced and downsampled by a factor of two. However, the size of the decomposition remains the same. For instance, if the input image size is 400x400, four images with size 200x200 are obtained after decomposition.

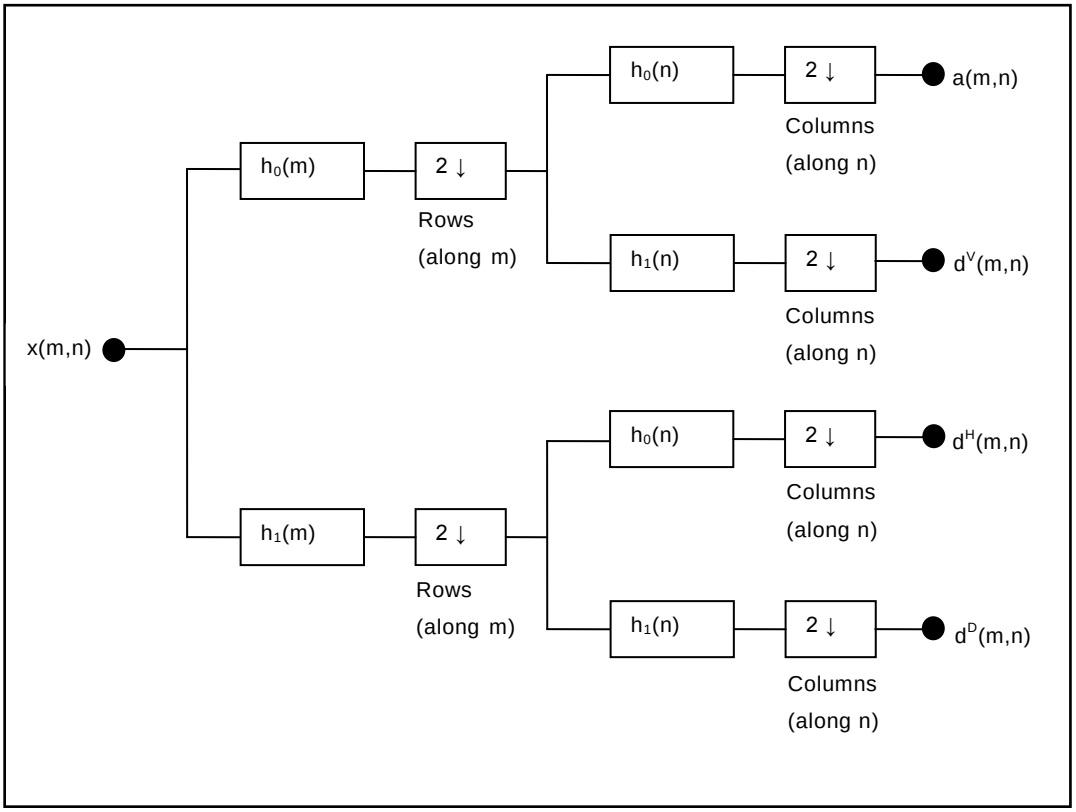

Figure 3.4 – A 2D four band filter bank for subband image coding

As can be seen from Figure 3.4, downsampling operation is applied twice: one before and other after second filtering. The resulting output images subimages contain information of a specific scale and orientation. Output images can be decomposed into smaller subbands.

Texture features can be calculated from each subimage. One of these features is energy that is computed by taking the mean magnitude of wavelet coefficients of the subimage. Variance or cooccurrence matrix features can also be calculated from the subimages. The formulas are given in below equations where *w* is a wavelet coefficient and the subimage is of size *NxN*.

Energy: $$e = \frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N}|w(x,y)| \qquad \text{(Equation 3.26)}$$

Variance: $$\frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N}[w(x,y)-e]^2 \qquad \text{(Equation 3.27)}$$

Wavelet transform has several advantages over other signal processing methods. First of all, as a result of the varying spatial resolution, the most suitable scale that provides a better representation of textures can be obtained. The wide range of choices of wavelets gives the opportunity to select the most suitable wavelet for a specific application. It is also reversible, so it can also be used in texture synthesis applications (Materka and Strzelecki, 1998).

3.1.2 Feature Extraction by Local Discriminant Bases (LDB)

Best bases algorithm was first proposed for signal compression (Coifman and Wickerhauser, 1992). By using orthonormal wavelet or trigonometric bases, the signal is expanded in binary tree structure. Important bases, or nodes of the generated tree, are evaluated by ussing entropy as cost function. Then, the tree is pruned in a bottom-up manner in order to select the best bases for signal compression.

This algorithm includes three main steps:

Step 1: Choose a decomposition method and expand the signal into specified orthonormal bases in binary tree structure and obtain the coefficients.

Step 2: Evaluate each basis by calculating the information cost function (entropy) from the coefficients. The entropy of a sequence $p=\{p_i\}$ with $\sum_i p_i = 1$ can be defined as:

$$H(p) = -\sum_i p_i log p_i \qquad \text{(Equation 3.28)}$$

Step 3: Compare parent node with its child nodes and prune the binary tree.

While pruning the tree, the nodes in the bottom level are set as children nodes. The children nodes are discarded if the cumulative information of the parent node is greater than its children nodes. Otherwise, the cumulative value of the children nodes is set to parent node. The parent nodes in one level become the child nodes in one higher level. At the end of the pruning, the remaining nodes comprise the best bases.

Entropy is an information measurement and minimization of it leads to efficient signal representation. However, in classification problems, the aim is to differentiate between different classes and the evaluation criterion must be measure of class separability among distributions. Hence, Local Discriminant Bases (LDB) that is based on discrimination power of each subspace was proposed (Saito and Coifman, 1995). In LDB algorithm, the signal is decomposed into orthonormal bases and the discrimination power of each node is calculated by using a dissimilarity measure. Most widely used dissimilarity measures (Rajpoot, 2003) are Kullback-Leiber divergence, Jensen-Shannon divergence, Euclidean distance, Hellinger distance and Fisher distance. The LDB algorithm can be summarized as follows (Saito and Coifman, 1995):

Step 1: Choose a decomposition method and expand the signal into specified orthonormal bases in binary tree structure up to level j.

Step 2: Construct time-frequency energy maps for classes.

Step 3: By using the map, compute the discrimination power of each node by using a dissimilarity measure.

Step 4: Compare parent node with its child nodes and prune the binary tree.

Step 5: Order basis functions by their discrimination power.

Step 6: Use k<n most discriminant basis functions for classification construction.

In pruning step, the nodes in level *j-1* are set as parent nodes and if the cumulative discrimination power of children nodes is greater than the parent node, children nodes are kept and the discrimination value of the parent node is set to this cumulative value. Otherwise, the children nodes are discarded and the parent node is kept.

When applied to images, a quad tree structure is obtained in the first step of the algorithm and other steps remain the same.

3.1.3 Application Areas of Texture Analysis

Texture analysis methods have many application areas, including medical image analysis, remote sensing, and inspection. We will review the literature of remote sensing and inspection as the techniques developed for these domains are likely to work for food inspection as well.

3.1.3.1 Remote Sensing

Texture analysis has been broadly used in remote sensing applications. Ruiz, Fdez-Sarria and Recio (2004) used texture analysis methods for the classification of four different types of landscapes. Gray Level Cooccurrence Matrix, Wavelet, Gabor and Law's energy filters methods were compared according to the accuracy rate of classification. It was stated that better results can be obtained by using cooccurrence features; however, combination of these features with multiscale methods increased the accuracy. Zhang, Xue and Zhang (2005) used wavelet based features to classify seven types of landscapes. Overall accuracy of classification was 84.3%.

Du, Lee and Mango (1992) used wavelet transform features extracted by the examination of subbands for texture segmentation of Synthetic Aperture Radar (SAR) images. They successfully segmented the images into four categories of ridge, multi-year ice, and water.

Haralick et al. (1973) utilized features from GLCM in order to classify remotely sensed images and 82% of the images are correctly classified into eight different types of landscapes.

3.1.3.2 Inspection

Texture analysis applications in the field of inspection include the inspection of defects and quality control.

Most defect detection applications have been in the field of textile inspection. Dewaele et al. (1988), utilized transform based methods for the detection of defects in textile images. Texture features are computed by applying sparse convolution masks to the images. Mitchell, Bowden and Sarhadi (2000) used edge operators in order to detect defects on denim. Latif-Amet et al. (2000) proposed a combination of wavelet features and cooccurrence matrix features for the detection of defects in textile images.

Another application field of inspection is quality control. Siew, Hodgson and Wood (1988) used texture analysis methods for the assessment of carpet wear. They utilized GLCM features along with the first order statistical features. They effectively characterized the carpet wear by using the obtained features.

He, Li and Deng (2007), applied transform based method in the field of food quality inspection. They used wavelet transform in order to classify images of eight varieties of tea that are obtained from near infrared spectroscopy. Principle Component Analysis (PCA) was then

applied to the extracted features for feature reduction. Their model correctly classified the eight classes of tea.

3.2 Feature Selection

Since performance of the classification algorithms are affected by large number of features, feature selection has significance in many areas, especially in hyperspectral imaging. Feature selection is the process of choosing an optimum set of features by eliminating features that are irrelevant, redundant or with little or no predictive information. The aim is to improve comprehensibility and reduce the number of features used in classification while maintaining or increasing classification accuracy (Kim, Street and Menczer, 2003).

Feature selection algorithms require a search procedure through the feature sets. In order to carry out this search, four basic search characteristics must be specified: starting point, search organization, evaluation strategy and stopping criterion.

As a starting point, forward selection that starts with an empty set of features and successively add features to the set can be used. Second option is backward search that starts with the full feature set and successively remove features from the set. Another option can be starting somewhere in the middle and moving from this point by adding or removing features.

As a search organization, exhaustive search strategies that try all possible feature sets and lead to 2^N-1 possible feature subsets for N initial features or heuristic search strategies that focus on a good solution rather than trying all possible paths that can be utilized. Filter or wrapper approaches can be used as evaluation strategy. While determining the stopping criterion feature selector can stop adding or

removing features when the merit of current feature subset is not improved (Hall, 1999).

There are two main categories of features selection methods: filter methods and wrapper methods. Filter methods rely solely on general properties of the features in order to select the best feature set and they are not dependent on any learning algorithm. On the other hand, wrapper methods use a predetermined induction algorithm to evaluate the feature sets (Michalak and Kwasnicka, 2006).

Wrapper approach often leads to better results than filters due to the inclusion of induction algorithm to evaluate alternatives and the relevance of a feature can be determined by estimating the accuracy of the algorithm (Hall, 1999). Their drawback is that they are much slower and complex than filter methods because use exhaustive search strategies and they must repeatedly call the induction algorithm. A filter can provide an initial feature subset for a wrapper to increase the speed (Hall, 1999), (Duch, 2006).

The advantage of using filter approach is that it is faster than the wrapper method. In the case of high dimensional data, they are more practical to use. However, since they filter out the undesirable features before the learning begins, the performance of feature subsets is not assured (Talavera, 2005).

Some filter methods use discretization for feature selection. In these methods, the features that give the same value for all samples when they are discretized are removed from the data (Hall, 1999).

Correlation-based filters select features based on correlations between features and remove the features with high correlation. Pearson linear correlation coefficient or Correlation-based Feature Selection (CFS) algorithm can be utilized for this purpose (Hall, 1999).

Another way of selecting features is based on looking at the differences between probability distributions of classes. A simple measure can be the difference between the joint probability and product probability distributions of each class. Fisher criterion that maximizes the distance between the means of the two classes while minimizing the variance within each class is another measure.

Some filter feature selectors are based on information theory. They use information gain that gives the information gained by adding a new feature to the class as a relevance measure (Duch, 2006).

3.3 Classification

Classification is the procedure of assigning objects to classes based on their features or parameters (He, 2005). There are two main categories of classification: supervised classification and unsupervised classification. In supervised classification, the aim is to assign the samples into pre-determined classes. A set of samples with known classes called training set have also be provided. Unsupervised classification; on the other hand, includes determining the number and location of classes and can be used to define class boundaries precisely with the idea of grouping objects with similar characteristics into same cluster while keeping inter cluster similarities minimum. It implies the use of some clustering algorithms (Lee and Yang, 2009).

There are two phases in constructing a classifier. In the learning phase, the parameters are set in order to separate classes correctly by using the training data. In the testing phase, the classifier with pre-determined parameters are applied to a set of objects whose classes are unknown in order to find what their classes are likely to be.

3.3.1 Classification Methods

Several methods have been proposed to solve classification problems in the literature. Linear classifiers differentiate classes by using linear functions and make classification decision based on the value of linear combination of the features. Nonlinear classifiers produce nonlinear decision boundaries (Lotte, Congedo, Lecuyer, Lamarche and Arnaldi, 2007).

The most widely used classification algorithms are K-Nearest Neighbor (KNN), Neural Network, Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA).

While comparing classification algorithms, accuracy of the classifier is the most important criterion. Other criteria are time to learn, comprehensibility of the results, and computation time and speed of the classifier (Harper, 2005).

3.3.1.1 K-Nearest Neighbor

In this method, an instance is assigned to the class to which majority of its k-nearest neighbors belong. Since the neighbor is close, it is expected to be similar to the instance that is classified and so they are likely to be in the same class. Initially, in the training phase, training samples are mapped to the multidimensional feature space and regions of each class are identified by computing k-nearest neighbors of the training set. In actual classification phase, the test sample with unknown class label is represented in feature space. Afterwards, distances from the test sample to all train samples are calculated and k-nearest samples are chosen. In order to calculate the distances, usually Euclidean distance is used as a distance measure. The test sample is then assigned to the class that is the most dominant one among the k-nearest training samples (Lotte et al., 2007).

The major advantage of this method is that it is easy to implement. It can also lead to good accuracy for the data with instances whose features have different characteristics for different classes (White, 1997).

There are several disadvantages of KNN method. The most serious drawback of this method is its sensitivity to the existence of irrelevant features which have random value for all instances (White, 1997). As the number of training samples increase, the method becomes slower. Another disadvantage occurs when classes with more frequent examples exist. These classes dominate the prediction of new test sample because majority voting is used to assign the new sample to one of the classes. In this method, entire training set represent the object distribution and there is no simplification of objects into a set of comprehensible set of features which is another disadvantage of using KNN method (White, 1997).

3.3.1.2 Neural Networks

Neural network is a mathematical model inspired by biological neural systems (Baesens, Van Gestel, Viaene, Stepanova, Suykens and Vanthienen, 2003). It is a combination of interconnected neurons and it is adaptive which means the structure of the system is changed as external or internal information flow through the network during the learning phase. They are powerful statistical modeling tools for capturing complex input output relationships or for finding patterns in data (He, 2005). A neural network with one layer and one output neuron is called a perceptron which is similar to linear classifiers. Neural networks with multiple layers and output neurons are capable of separating any continuous surface (Dror, 2007). Most widely used network is multilayer perceptron that includes an input layer, one or more hidden layers and an output layer. Each neuron processes the

input and produces an output that is passed to the succeeding neuron. The output of hidden neuron i can be computed as follows:

$$hi = f^{(l)}\left(b_i^{(l)} + \sum_{j=1}^{n} w_{ij}\, x_j\right) \qquad \text{(Equation 3.29)}$$

where $b_i^{(l)}$ is the bias term corresponding to hidden neuron *i* of hidden layer *l*, w_{ij} the weight of the connection between input *j* and hidden neuron *i* and *f* is the transfer function such as sigmoid or hyperbolic tangent. The output of the output layer is calculated as follows:

$$y = f^{(o)}\left(b^o + \sum_{j=1}^{n_h} v_j\, h_j\right) \qquad \text{(Equation 3.30)}$$

where n_h is the number of hidden layers and v_j is the weight of the connection of the hidden layer j to the output neuron (Baesens et al., 2003).

The major advantage of neural network methods is their ability to handle problems with many features and classify instances that have complex distribution in feature space. Another advantage is their arbitrary function approximation mechanism that learns from the training data (White, 1997).

However, neural network methods suffer from several drawbacks. Its black-box nature that means the lack of transparency in hidden steps leads the neural network to perform well only in some situations and makes it impossible checking the system for plausibility and accuracy. They are difficult to set up to produce good results. They are have a tendency to be influenced by noise, and are prone to overfitting. The learning of the neural networks is also susceptible to local minima (He, 2005), (Dror, 2007).

3.3.1.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is one of the well established and simple classification algorithms. It uses hyperplanes to separate the data representing the different classes from the training data. The method finds a direction *w* in the n-dimensional space.

To assign a new sample to one of the classes, the projection of the sample onto *w* and the distance from the new sample to the means of the projections of the training classes are calculated. The aim is to maximize interclass variability while minimizing variability within each class. In order to reach this aim the difference between projected means are maximized and variance of the projected points in each class is minimized. For a two class problem the maximization criteria is computed by using the function given below:

$$j(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2} \qquad \text{(Equation 3.31)}$$

$$m_i = w^T \mu_i \qquad \text{(Equation 3.32)}$$

$$z_i = w^T x_i \qquad \text{(Equation 3.33)}$$

$$s_i = \sum_{z_j \in Z_i} \left(z_j - m_i\right)^2 \qquad \text{(Equation 3.34)}$$

where m_1 and m_2 are the projected means of the points in classes, s_1 and s_2 are the scatter for the projected points and z_i is the projection of any d-dimensional point onto the vector *w*.

The set of projections represent the mapping from d-dimensional space to one dimensional space along *w*. For the two class case, z_1 contains all points in class 1 and z_2 contains all points in class 2.

The decision about a new instance is given calculating the distance from the projection of new instance to the means of the projections of each training data points computed for each class. This means that the probability of an input *x* being in a class *y* is purely a function of this linear combination of the known observations (Dror, 2007).

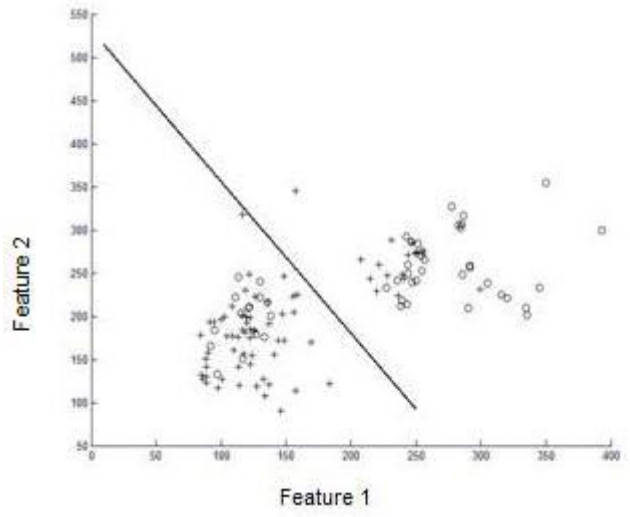

Figure 3.5 – Scatter plot of two features where the solid line is the decision boundary obtained by LDA.

The advantage of LDA method is its simplicity and computational efficiency. The vector *w* can be found quickly using a simple procedure. It also provides a powerful way to reduce the dimensionality of the multivariate data. It has no structural parameters as well (Dror, 2007).

The major disadvantages of LDA are that it performs poor when the means of classes are similar. The number of input variables has to be smaller than the number of samples for the method to be applicable (Burget, 2004).

3.3.1.4 Support Vector Machines

The aim of a Support Vector Machine is to find out an optimal hyperplane that maximizes the margin between two categories of data. They are primarily two-class classifiers but they are extendible for multiclass situation (Sun, Bebis and Miller, 2002). SVM is based on statistical learning theory and structural risk minimization. For the non-linear classification case, a mapping function is utilized to map the input space into a higher dimensional feature space by using a non-linear kernel function $K(x_i,x_j)$ where the data points can be separated linearly (Rajpoot and Rajpoot, 2004).

Optimization criterion is the width of the margin between classes and the classification function can be defined as follows:

$$f(x) = \sum_{i=1}^{N} y_i \alpha_i k(x, x_i) + b \qquad \text{(Equation 3.35)}$$

where x_i is the input data, $y_i \in [-1, +1]$ are the class labels, $k(x, x_i)$ is the kernel function and the sign of $f(x)$ indicated the membership of x. Support vectors of the optimal hyperplane are the data points corresponding to a nonzero α_i.

The advantages of using SVMs for classification are their better performance in higher dimensional spaces, computational efficiency, and robustness to noisy data. Since it is based on margin maximization, it is known to have good generalization properties (Flietstra, 2008), (Lotte et al., 2007).

The main drawback of SVM method is that large numbers of training data are often required for the algorithm to converge but overfitting increases when the number of support vectors increases. As a result of its black box nature, it is hard to infer conclusions about the data. Another disadvantage is that it is hard to find the kernel parameters (Bornstein, Gilmore, Castano and Greenwood, 2007).

CHAPTER 4

METHODOLOGY

4.1 Introduction

Image or texture classification applications generally include the steps such of image acquisition, data pre-processing, feature extraction, feature selection (Casasent and Chan, 2004), (Michalak and Kwasnicka, 2006), (Kalkan, 2008) and classification. Since hyperspectral imaging results in a large number of data to be analyzed, it requires feature extraction and selection prior to classification to reduce the dimension of the feature space. In this thesis, the steps that are followed in order to classify the chili pepper samples are given in Figure 4.1.

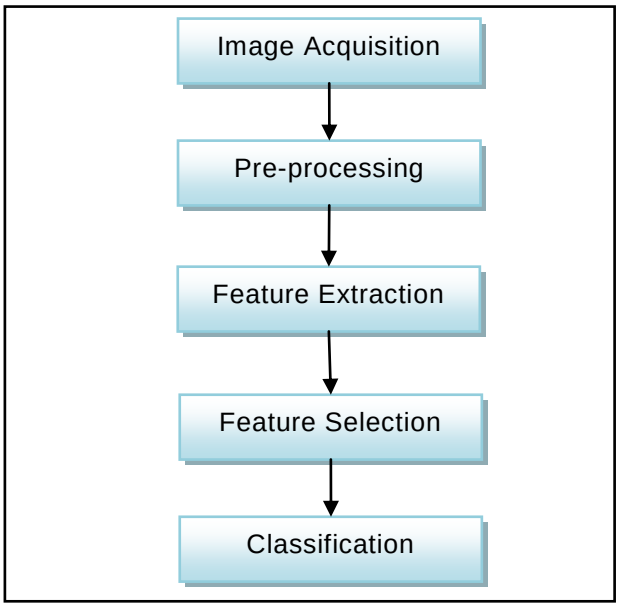

Figure 4.1 – Steps followed for the classification of chili peppers

In this thesis, we try to extract the features to discriminate aflatoxin contaminated and uncontaminated chili peppers using the hyperspectral images that are taken in various wavelengths. We used two approaches for extracting the most discriminative features: Statistical approach and LDB approach. In statistical approach, a combination of First Order Statistical (FOS) features and Gray Level Cooccurrence Matrix (GLCM) features are extracted.

The second approach for feature extraction is 2D Local Discriminant Bases (LDB) approach. LDB algorithm is effective for the classification of hyperspectral data because of its speed and simplicity. We use the methodology that (Kalkan, 2008) applied to classification of hyperspectral images of aflatoxin contaminated and uncontaminated hazelnuts. In his study, two feature trees were generated. The first feature tree was on spectral axis where the reflectance energies are located in the nodes of the binary tree. The tree was then pruned by comparing Euclidean distance between corresponding nodes. The second tree was generated in quad tree structure by using wavelet packet decomposition on spatial-frequency axis.

In this chapter, the methodology that is used to distinguish contaminated and uncontaminated chili peppers is introduced. In Subsection 4.2, the data characteristics and image acquisition process is described. In Subsection 4.3, the preprocessing steps that are used in order to remove noise from the images are explained. In Subsections 4.4 and 4.5, our methodology of extracting features is given. In Subsection 4.6 the methodology used for feature selection and classification is described.

4.2 Data Set and Image Acquisition

4.2.1 Data Set

Scaled chili pepper samples that are sold with or without packaging were collected from different cities of Turkey. The number of chili peppers obtained from each city is given in Table 4.1.

Table 4.1 – The cities that the chili pepper samples obtained and number of peppers from each city

City	**Number of chili pepper samples**
Ankara	11
Antalya	6
Diyarbakır	4
Erzincan	2
Hatay	2
İstanbul	2
İzmir	2
Kahramanmaraş	10
Sivas	1
Total	**40**

The samples were sent to chemical analysis in order to determine their aflatoxin levels after the hyperspectral imaging. Aflatoxin levels of each sample can be seen in Figure 4.2. Aflatoxin B1, B2, G1 and G2 levels and total aflatoxin level of each scaled chili pepper is given in Appendix.

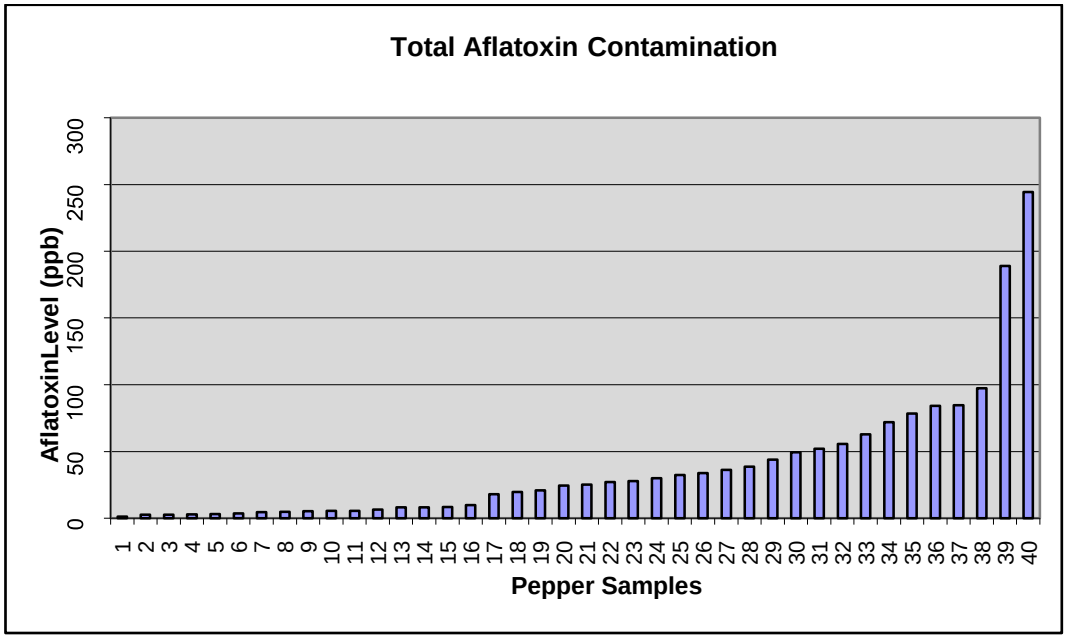

Figure 4.2 – Total aflatoxin levels of scaled chili pepper samples

Initially, 32 chili pepper samples were collected. 23 of them are contaminated with aflatoxin and the remaining 9 are uncontaminated. While extracting the features by statistical approach, this data set, *DS1*, is used. Then, in order to increase the number of uncontaminated samples 8 more pepper samples were bought. The whole data set, *DS2*, is used in the extraction of features by LDB approach.

We set 10 ppb as threshold for aflatoxin contamination. For *DS1*, 9 samples having aflatoxin contamination below 10 ppb are assigned to *Uncontaminated1* class and the remaining 23 samples that have aflatoxin over 10 ppb are assigned to *Contaminated1* class. For *DS2*, 16 samples having aflatoxin contamination below 10 ppb are assigned to *Uncontaminated* class and the remaining 24 samples that have aflatoxin over 10 ppb are assigned to *Contaminated* class. The mean aflatoxin level of the samples in DS1 and DS2 are given in Table 4.2 and Table 4.3, respectively.

Table 4.2 – Number of samples and mean aflatoxin levels (ppb) of the samples for DS1

	Number of Samples	Mean Aflatoxin Level
Afl+	23	59,80
Afl-	9	5,26
Total	32	44,46

Table 4.3 – Number of samples and mean aflatoxin levels (ppb) of the samples for DS2

	Number of Samples	Mean Aflatoxin Level
Afl+	24	60,31
Afl-	16	5,18
Total	40	38,26

4.2.2 Image Acquisition

A hyperspectral imaging system that includes VIS DFK 41AF02 digital Charge Coupled Device (CCD) camera, bandpass filters that are placed on a wheel to position the filters in front of the lens of the camera, a cabin that functions as a dark room, 2 UV-A lamps that have peak intensity at 365 nm and a computer to capture the images is used for image acquisition.

In the image acquisition process, samples are screened by using 14 different filters at 400 nm to 510 nm with 10 nm FWHM, 550 nm and 600 nm with 70 nm and 40 nm FWHM, respectively. The reflected light from the samples are captured by using IC Capture image acquisition tool.

For each sample, three images that are from different parts of the sample are taken. Therefore, the number of total pepper samples to be used in analysis is 120 (72 contaminated and 48 uncontaminated). The exposure time of the camera is set to 2 seconds in order to

ensure achieving sufficient reflectance light. The sample images from each spectral band of a scaled chili pepper with aflatoxin and without aflatoxin are given in Figure 4.3 and Figure 4.4, respectively.

4.3 Pre-processing

The images obtained from the image acquisition process may include impulsive noise arising from the dust on the camera or filters. Median filter, an order statistics filter that replaces the value of the center pixel with the median value of its neighborhood intensity values, is used for removal of impulsive noise (Gonzalez and Woods, 2002). To handle the problem, images are filtered with median filter with size of 7x7.

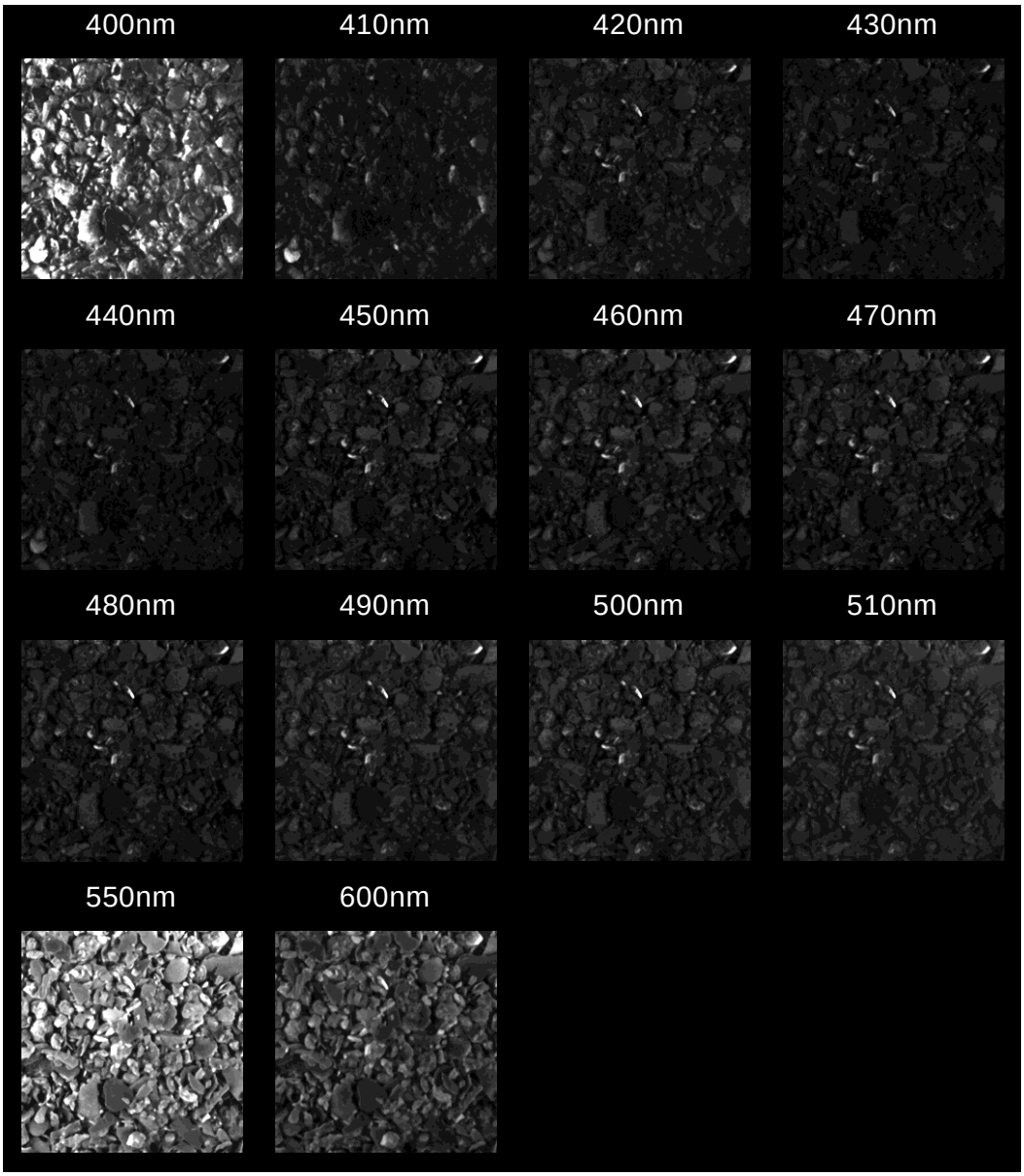

Figure 4.3 – Spectral band images of a contaminated chili pepper sample

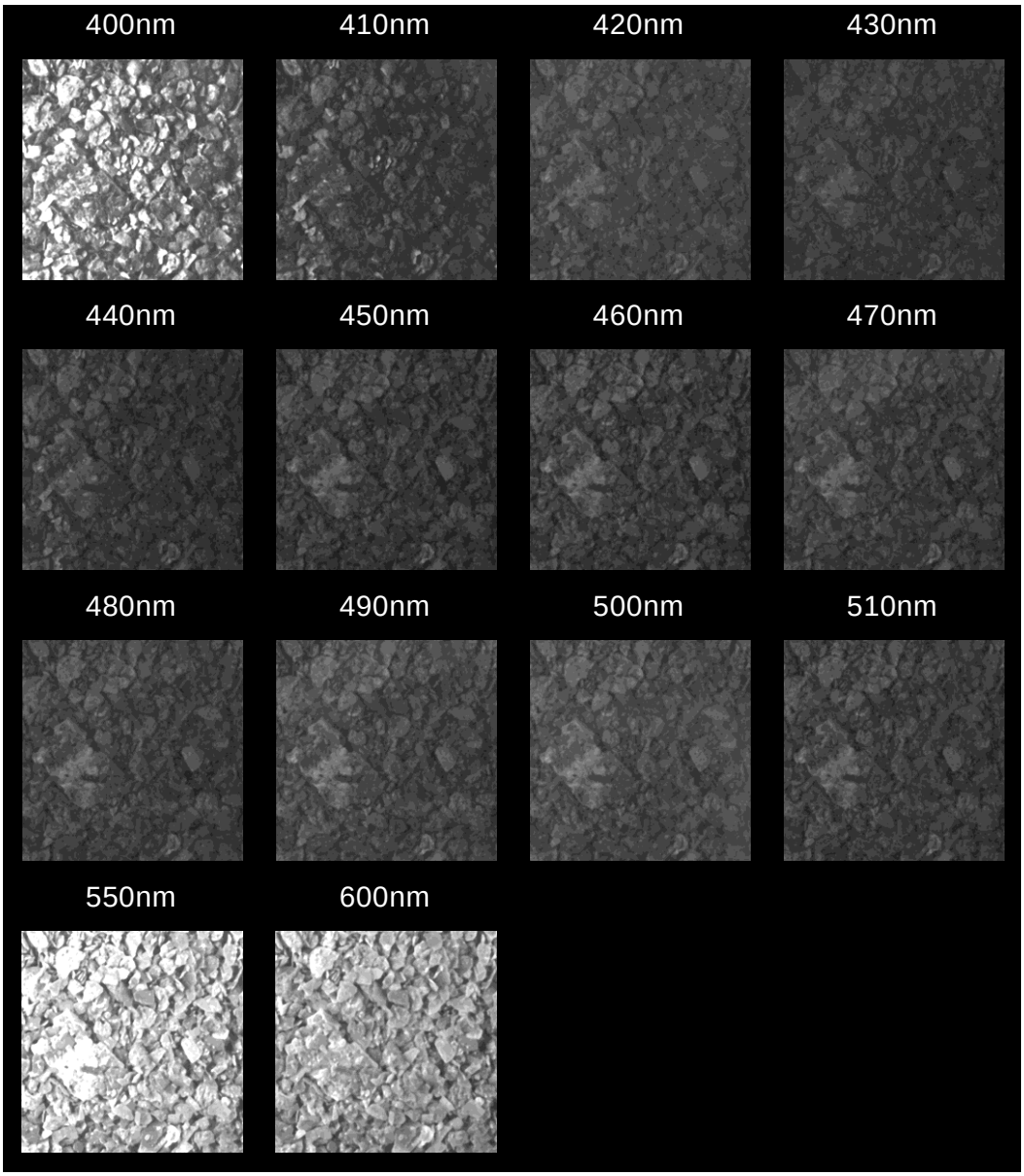

Figure 4.4 – Spectral band images of an uncontaminated chili pepper sample

4.4 Feature Extraction with Statistical Approach

In statistical approach, First Order Statistical (FOS) features and Gray Level Cooccurrence Matrix (GLCM) features are extracted in order to classify the aflatoxin contaminated and uncontaminated chili peppers.

FOS features give information about the distribution of pixels on an image and do not provide any information about their relative positions

and spatial structure. In the first step, first order statistical features; mean, variance, skewness and kurtosis, are calculated for each hyperspectral image.

GLCM gives the distribution of probability of gray level occurrences in an image. GLCM is calculated for a specified neighborhood distance and direction. In the second step, four different GLCM are calculated for four different directions (θ = 0°, 45°, 90°, 135°) by setting the neighborhood distance d to 1. Total 28 features are extracted from each hyperspectral image. The fisher distances between the features of aflatoxin contaminated (*Contaminated1*) and uncontaminated (*Contaminated1*) groups are calculated. It is observed that the discrimination power of the features calculated for different directions are nearly the same and directionality information in chili pepper images has no considerable effect on classification. Therefore, the GLCMs for each four direction are averaged.

The elements of the GLCM are normalized by dividing each entry by the total number of pixel pairs. GLCM features (Haralick, 1979) homogeneity, contrast, correlation, energy, entropy, inverse difference moment and maximum probability are computed which lead to total 7 features for each hyperspectral image. While computing the features; although the number of distinct gray levels is high, quantization is not applied in order not to lose any textural information.

By combining with the FOS features, total 11 statistical features are obtained for each spectral band.

4.5 Feature Extraction with LDB Approach

In order to obtain the most discriminative features from hyperspectral data, 2D structured LDB algorithm is used. The main motivation of using LDB algorithm is that it provides the most discriminative bases in

a computationally efficient manner. It decomposes not only the low frequency bands as in standard wavelet transform but also the high frequency bands that have a potential to carry information about the texture properties.

In original LDB algorithm, Wavelet Packet is used for the decomposition of frequency axis and Local Cosine Packets are used for decomposition of the time axis (Saito and Coifman, 1995). However, in classification of hyperspectral textured images both spectral and spatial-frequency information have importance. Therefore, two feature trees are generated in both spectral and spatial-frequency axis. These feature trees are pruned in spectral and spatial-frequency axes to find the location of most discriminative features.

The LDB algorithm includes three major steps: generation of the feature trees in both spectral and spatial-frequency axis, pruning along the spectral axis and pruning along the spatial-frequency axis.

4.5.1 Feature Tree Generation

Spectral features are important for the classification of hyperspectral images. Adjacent spectral bands may contain highly correlated features and dimensionality reduction has great importance in hyperspectral image classification applications in order to increase the classification performance. One feature tree is generated on the spectral axis and a second tree is generated on spatial-frequency axis in order to capture different spectral and space-frequency localization characteristics.

4.5.1.1 Spectral Feature Tree Generation

A binary tree is generated from the reflectance energies of the hyperspectral images. All of the spectral features are placed at the bottom of the feature tree from left to right. Some of the nodes that do

not correspond to any spectral band in this level are set to null. The feature values of the nodes in the upper levels of the tree are obtained by summing the values of their child nodes. A four level binary tree with 16 spectral bands is given in Figure 4.5 where the bottom nodes of the tree correspond to the feature values extracted from individual spectral bands.

<table>
<tr><td colspan="16"></td></tr>
<tr><td colspan="4">B1+B2+B3+B4</td><td colspan="4">B5+B6+B7+B8</td><td colspan="4">B9+B10+B11+B12</td><td colspan="4">B13+B14+B15+B16</td></tr>
<tr><td colspan="2">B1+B2</td><td colspan="2">B3+B4</td><td colspan="2">B5+B6</td><td colspan="2">B7+B8</td><td colspan="2">B9+B10</td><td colspan="2">B11+B12</td><td colspan="2">B13+B14</td><td colspan="2">B15+B16</td></tr>
<tr><td>B1</td><td>B2</td><td>B3</td><td>B4</td><td>B5</td><td>B6</td><td>B7</td><td>B8</td><td>B9</td><td>B10</td><td>B11</td><td>B12</td><td>B13</td><td>B14</td><td>B15</td><td>B16</td></tr>
</table>

Figure 4.5 – Binary spectral band tree for 4 levels

4.5.1.2 Spatial-Frequency Feature Tree Generation

A feature tree in quad tree structure is generated by decomposing the spectral images into 3-level full wavelet decomposition. From each image in each level, 4 subimages are obtained by filtering the image with high pass or low pass filters through row and column directions. The energy in each subband is computed and used as texture features. 3-level decomposition results in 85 features for each spectral image. The full wavelet decomposition quad tree for 3 levels can be seen in Figure 4.6.

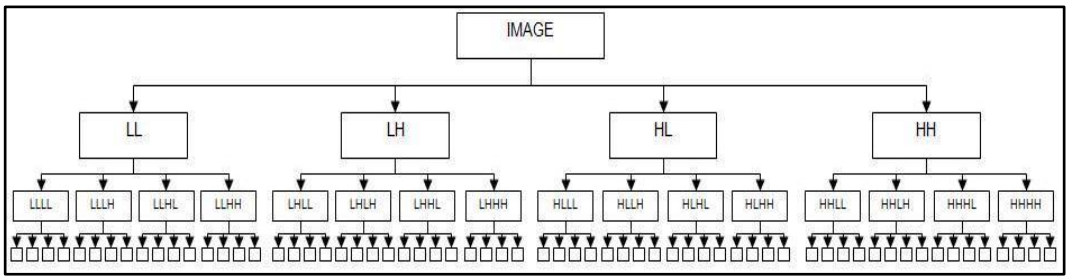

Figure 4.6 – Full wavelet decomposition quad tree for 3 levels

4.5.2 Pruning along the Spectral Axis

The binary spectral tree given in Figure 4.65 is pruned in order to achieve the most discriminative spectral bands in a bottom-up manner by using the pruning algorithm given below where *d* is a distance metric that measures the distance between classes for the feature of a given node.

Spectral axis pruning algorithm:

1. Calculate $max\{d_{child1}, d_{child2}\}$
2. Set $max\{d_{child1}, d_{child2}\}$ as d_{mother} if $d_{mother} < max\{d_{child1}, d_{child2}\}$ else remove children nodes.

The pruning algorithm merges the spectral bands when the discrimination power of the mother node is greater than any of its children nodes according to a distance metric. In this study, Euclidean distance between cumulative probability distribution of nodes is used as the distance metric. Euclidean distance is computed by:

$$Euclidean\ Distance\colon\ D(p,q) = \sum_{i=1}^{n} (p_i - q_i)^2 \qquad \text{(Equation 4.1)}$$

where p_i and q_i are normalized energy distributions of images from *class 1* and *class 2*, respectively. When two spectral bands are merged, the energies of the frequency subbands that correspond to the merged spectral bands are averaged before pruning the spatial-frequency axis.

4.5.3 Pruning along the Spatial-Frequency Axis

The quad tree is pruned in a bottom-up manner by using the algorithm below.

Spatial-frequency Pruning Algorithm:

1. Calculate $max\{d_{child1}, d_{child2}, d_{child3}, d_{child4}\}$
2. Set $max\{d_{child1}, d_{child2}, d_{child3}, d_{child4}\}$ as d_{mother}
   if $d_{mother} < max\{d_{child1}, d_{child2}, d_{child3}, d_{child4}\}$
   else remove children nodes.

If any of the four child nodes have higher discrimination power than the mother node, the child nodes are kept. Otherwise, the mother node is retained and the child nodes removed.

4.6 Feature Selection and Classification

Since hyperspectral imaging leads to a high number of features in the statistical approach or due to a large number of possible bases that can be used to represent the images in the LDB approach, feature selection is needed to find the most discriminative features. Fisher Distance Based (FDB) Feature Selection and Wrapper Based Feature Selection is used to rank the features.

The features that were ranked by FDB Feature Selection were fed into linear classifier one by one in order to find the optimum number of features that gave best classification. In Wrapper Based Feature Selection, all the feature combinations were fed into linear classifier to identify the best feature set giving the maximum classification accuracy.

There are many classifiers that are used in pattern recognition problems such as k-Nearest Neighbor (KNN), Support vector Macine (SVM), Neural Networks. Since the aim of this study is to extract relevant features for classification, a simple classifier, Linear Discriminant Analysis (LDA) is used as classifier.

CHAPTER 5

RESULTS AND DISCUSSIONS

In order to find the best features for the classification for the aflatoxin contaminated and uncontaminated chili pepper samples, two different approaches were used: statistical approach and LDB approach.

DS1 and *DS2* which are explained in Subsection 4.2 were used as data sets in statistical approach and LDB approach, respectively.

5.1 Statistical Approach

In the first step, First Order Statistical (FOS) Features; mean, variance, skewness and kurtosis, were calculated. In the second step, one GLCM was calculated by averaging GLCMs for four different directions (θ = 0°, 45°, 90°, 135°) by setting the neighborhood distance d to 1. Total 7 features; homogeneity, contrast, correlation, energy, entropy, inverse difference moment and maximum probability; were extracted from each hyperspectral image. By adding the FOS features, a total of 154 features were obtained for 14 spectral bands.

In order to extract the most discriminative features among these 154 features, one sample is left out from the data set. For the remaining training set, the best 20 features according to their Fisher distance were calculated. This step is repeated for each training sets that is obtained by leaving out one different sample out. Then, for each feature, a general score was computed by considering the rank and number occurrence of that feature in all training sets. The best 10

features according to their general feature scores were selected for further analysis. The selected features and are given in Table 5.1.

Table 5.1 – The best 10 features and their spectral bands according to their general feature scores

Spectral Band	**Features**		
400	homogeneity	Contrast	energy
480	kurtosis		
490	kurtosis		
500	kurtosis	Energy	
510	mean	Kurtosis	
550	homogeneity		

In order to select the features that give the minimum classification error, two approaches were used: Fisher Distance Based (FDB) and wrapper methods. In FDB, the features were fed to linear classifier one by one according to their fisher discrimination to find the optimal number of features for classification. In wrapper base approach, best combination of features giving the best classification was determined. The classification errors curves of selected features by using Fisher Distance Based (FDB) and wrapper methods are given in Figure 5.1.

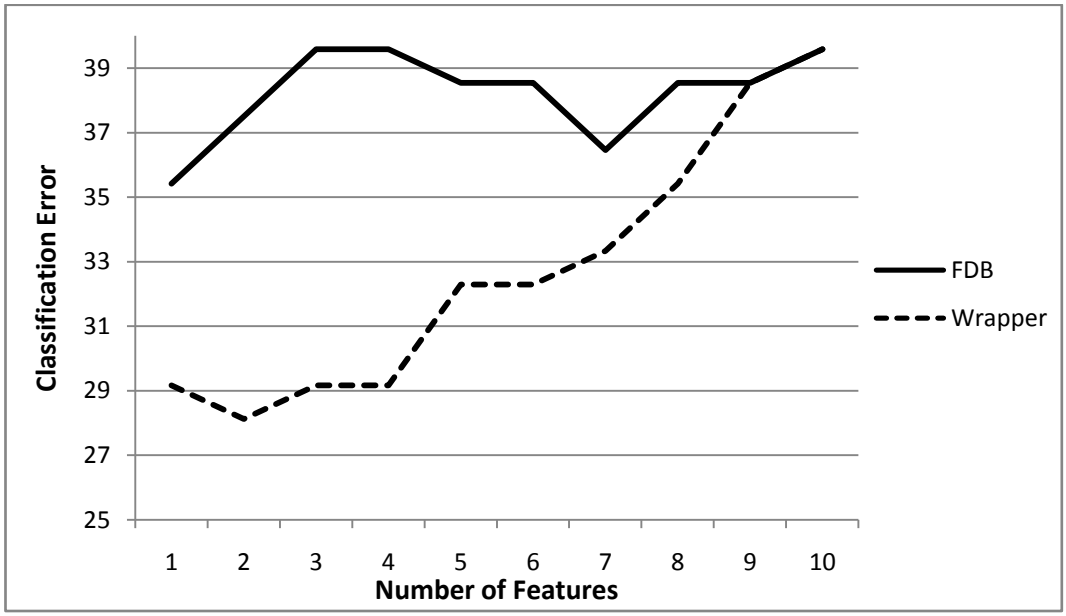

Figure 5.1 – Classification error curves of FDB and wrapper based selected features from extracted features by using statistical approach

The minimum classification error was obtained with 1 feature by using FDB approach and 2 features by using wrapper approach.

By applying leave-one-out cross validation, the total classification error is computed by taking the average classification errors of each test set. The results that are obtained by using the features selected by FDB and wrapper methods are given in Table 5.2.

Table 5.2 – Classification error comparison of FDB and wrapper for statistical approach

	Fisher Distance Based (FDB)	Wrapper Based
Number of features	1	2
Error(%)	35.42	28.12

As can be seen from Table 5.2, the minimum classification error was achieved with two features selected by wrapper approach. It was observed that 20 of the 69 samples in *Contaminated* group and 14 of the 27 samples in the *Uncontaminated* are misclassified by using the

selected feature with FDB approach. By using wrapper approach, 16 samples in *Contaminated* group and 12 samples in *Uncontaminated* group are misclassified.

5.2 LDB Approach

In the first step, a binary spectral tree was generated on the spectral axis. Although the reflectance images of each pepper sample for 14 different bands were collected in the image acquisition step, images of 550 nm and 600 nm were not put into the tree generated because of their wide FWHM. Therefore, for 12 spectral bands from 400-510 nm, reflectance energies were placed at the bottom of the binary tree. Remaining 4 nodes at the bottom of the binary tree were set to null and not used in the pruning. Energies of the merged bands are averaged.

As the second step, 3-level full wavelet decomposition was obtained by using Daubechies 8 tap filter that result in 85 subband images for each spectral band. Energies of subband images were used as features. A total of 1190 features are achieved for 14 spectral bands. These features were placed on the spatial-frequency quad tree. Then, the spatial-frequency tree was pruned.

At the end of pruning of spectral axis and spatial-frequency axis, a feature map of best discriminative features was obtained.

In order to select best discriminative features, at each time one of the samples is left out for testing and a feature map is obtained by using LDB algorithm from the remaining training samples. As a result of that, one feature map was constructed for each training set. The most common feature map is given in Figure 5.2. As can be seen from the Figure 5.2, spectral bands 400-410 and 440-470 are merged and the

low spatial-frequency subbands of spectral bands 420, 430, 550 and 600 are pruned.

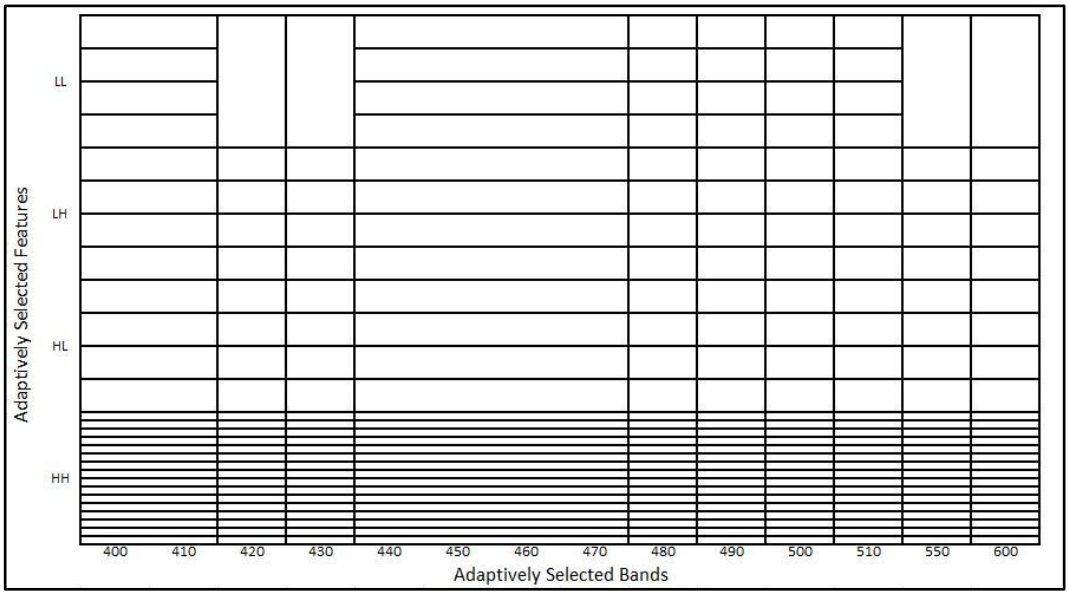

Figure 5.2 – General feature map

Since there were some different features in the feature maps obtained from different training sets, the most common and discriminative features to be used in classification were selected by computing a general feature score for each feature in the feature maps. While calculating the feature score, a score value is assigned to each feature according to their discrimination power rank. The features in all feature maps were then combined and all the scores assigned to each feature were summed and became the general score of that feature. The features were sorted according to their scores. After this analysis, the most discriminative 11 features are selected. The most discriminative features are shown in Figure 5.3, where the darkness of the area shows the discrimination power of the feature.

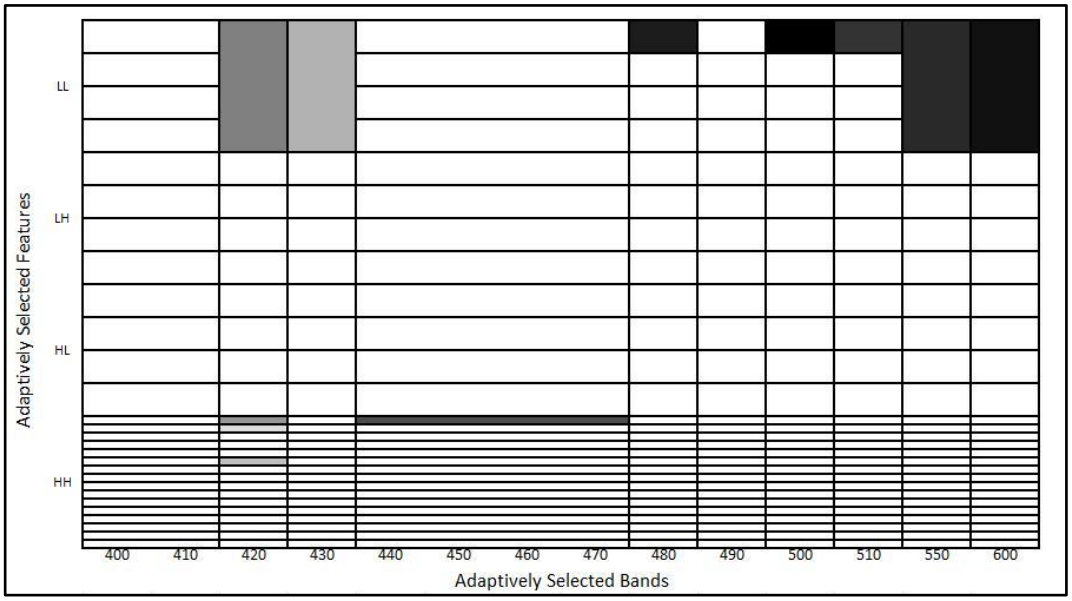

Figure 5.3 – Most discriminative features selected by fisher distance

The extracted features were fed into a linear classifier incrementally by using two approaches. In the first case, the features were ranked according to their Fisher distance and added one by one in order to determine the optimal number of features for classification. In the second case, wrapper approach was used in order to obtain the best combination of features that gives the best classification. The classification errors curves of selected features by using Fisher Distance Based (FDB) and wrapper method are given in Figure 5.4.

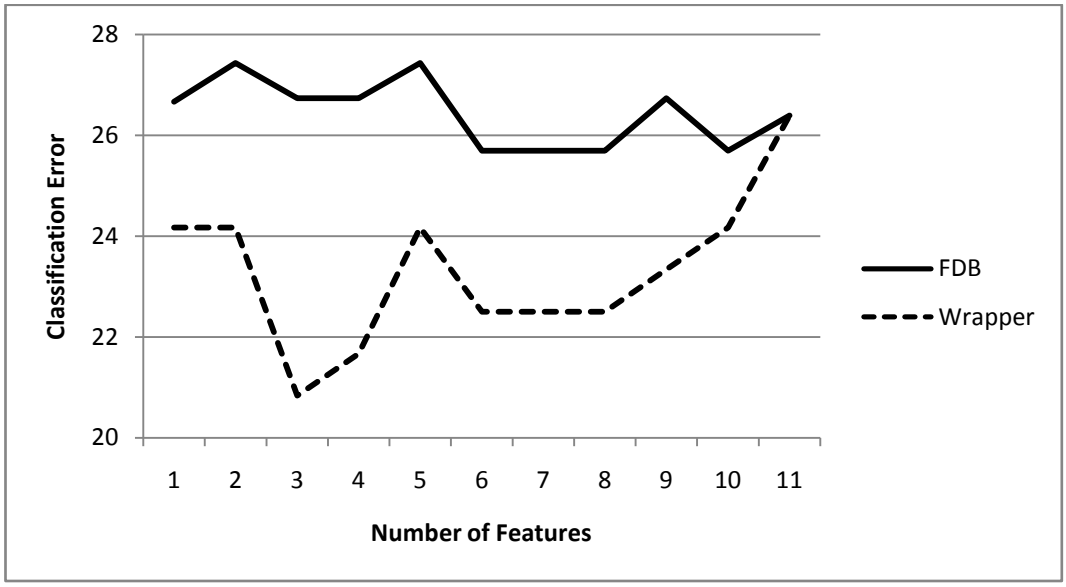

Figure 5.4 – Classification error curves of FDB and wrapper based selected features from extracted features by using LDB approach

It is preferable to use minimum number of features that give the lowest error. The minimum classification error was obtained with 6 features by using FDB approach and 3 features by using wrapper approach.

By applying leave-one-out cross validation, the total classification error was computed by taking the average classification errors of each train set. The results that are obtained by using the features selected by FDB and wrapper methods are given in Table 5.3.

Table 5.3 – Classification error comparison of FDB and wrapper for LDB approach

	Fisher Distance Based (FDB)	Wrapper Based
Number of features	6	3
Error(%)	25.69	20.83

As can be seen from Table 5.3, the minimum classification error was achieved with three features selected by wrapper approach. It was observed that 13 of the 72 samples in *Contaminated* group and 16 of the 48 samples in the *Uncontaminated* are misclassified by using the

selected 6 features with FDB approach. By using wrapper approach, 15 samples in *Contaminated* group and 10 samples in *Uncontaminated* group are misclassified. The mean aflatoxin level of the original data set is 38.26 ppb. When the peppers that are classified as contaminated are discarded, the mean aflatoxin level becomes 15.93 ppb.

5.3 LDB Approach on Intensity Based Divided Data Set

Since most of the discriminative features are generally from the low subbands, it is an indication that the average intensities of the hyperspectral images are different in *Contaminated* and *Uncontaminated* group. It was observed that images from *Contaminated* group are generally darker than the images of *Uncontaminated* group. In Figure 5.6, histogram of the total intensity values of the images is given. An intensity based division was performed according to the total of mean intensity values of each spectral band for each pepper sample before classification. Threshold on the histogram was selected as 920 while dividing the *Contaminated* and *Uncontaminated* groups are into two sets: *ContaminatedDark*, *ContaminatedLight*, *UncontaminatedDark*, and *UncontaminatedLight.*

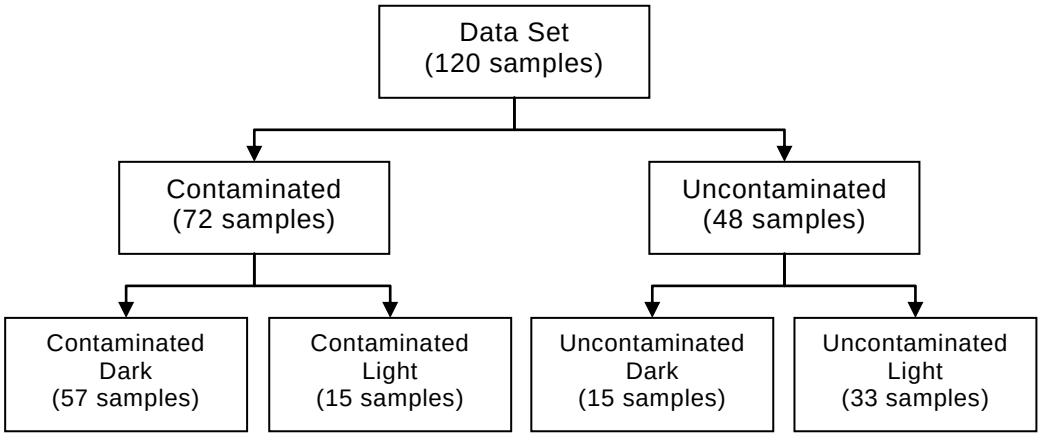

Figure 5.5 – Intensity based sets and their sample sizes

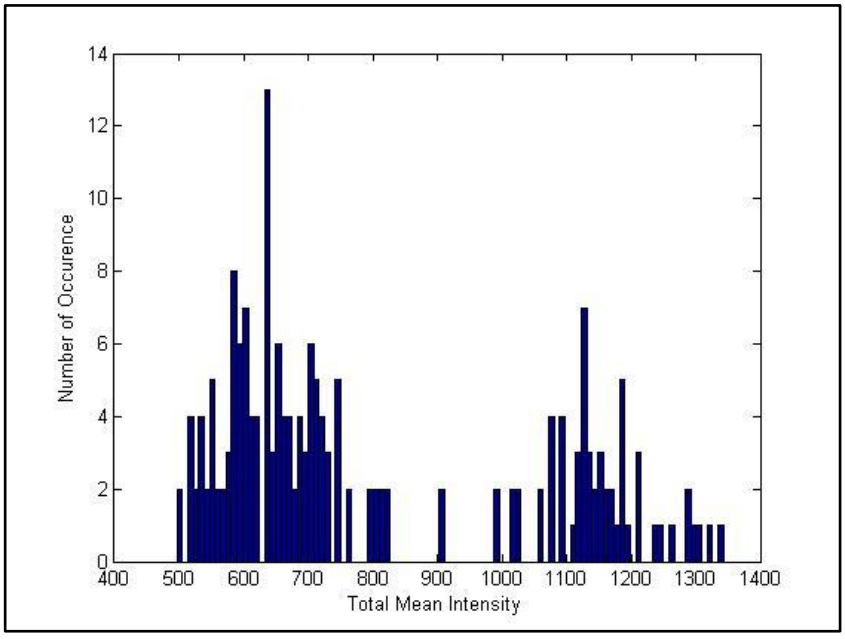

Figure 5.6 – Histogram of total intensities of hyperspectral images for all spectral bands

The feature maps for *Dark* and *Light* groups were obtained separately by using the same way that was defined above. The feature map of *Dark* group is given in Figure 5.7 and *Light* group is given in Figure 5.8 where the darkness of the area shows the discrimination power of the feature. For these two groups, the best discriminative 12 features considering their general scores were selected and displayed in these feature maps.

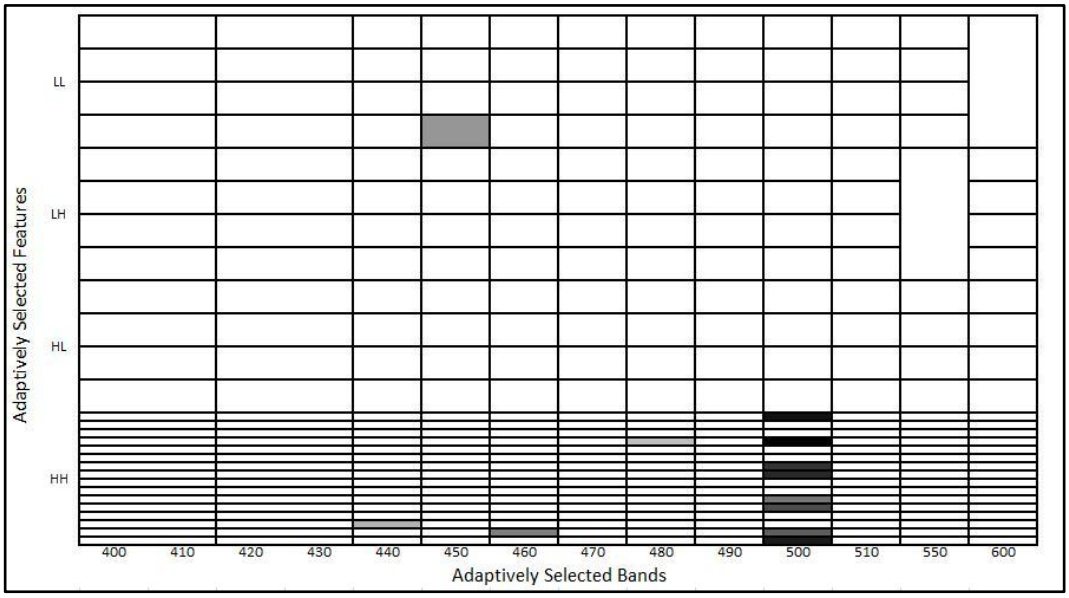

Figure 5.7 – Most discriminative features of dark group according to Fisher distance

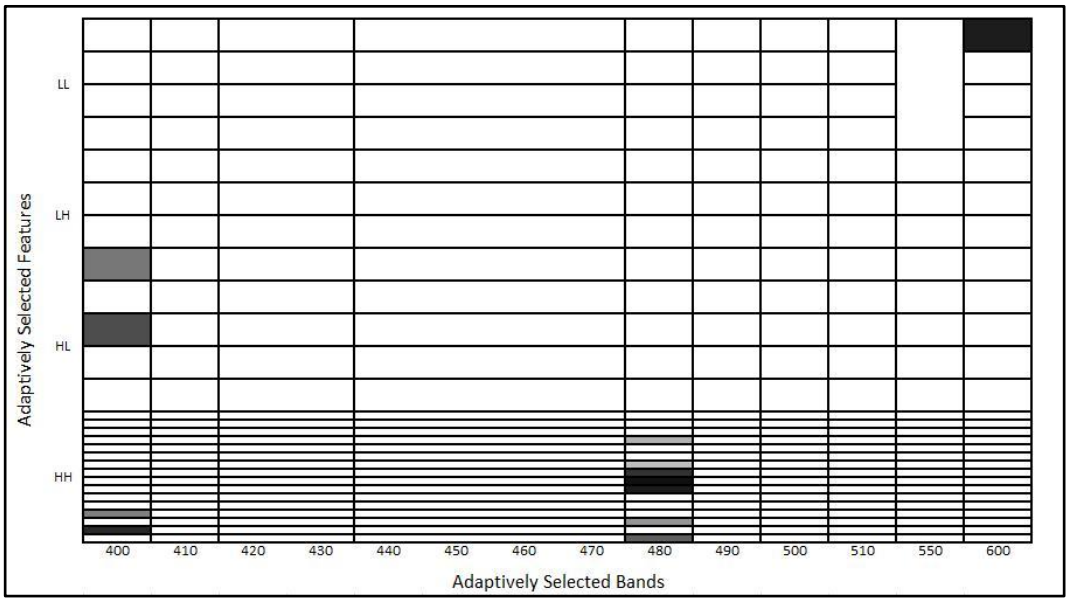

Figure 5.8 – Most discriminative features of light group according to Fisher distance

Classification was also performed separately for *Dark* and *Light* groups by using FDB and wrapper approaches. The best discriminative features are fed into linear classifier. The total

classification error was computed by taking the average of the results obtained from *Dark* and *Light* groups. The minimum classification error and the number of features used in order to achieve the minimum classification error with FDB and wrapper methods for *Dark* and *Light* groups are given in Table 5.4.

Table 5.4 – Classification error comparison of FDB and wrapper for dark and light groups

	Fisher Distance Based(FDB)		Wrapper Based	
	Dark	Light	Dark	Light
# of Features	8	2	4	2
Error(%)	32.77	39.72	26.39	14.58
Avg Error(%)	36.25		20.83	

Dividing the *Contaminated* and *Uncontaminated* sets into *Dark* (*UncontaminatedDark and ContaminatedDark*) and *Light (UncontaminatedLight and ContaminatedLight)* groups and applying LDB approach separately for each group has no effect on the classification performance. 17 of the 72 *Contaminated* peppers and 8 of the 48 *Uncontaminated* peppers were misclassified by using the wrapper approach. The number of misclassified samples for each group is given in Table 5.5.

Table 5.5 – The number of misclassified samples in each group

	Contaminated	Uncontaminated	Total
Light	1	5	6
Dark	16	3	19
	17	8	25

The mean aflatoxin level of the original data set is 38.26 ppb. When the peppers that are classified as contaminated are discarded, the mean aflatoxin level becomes 29.24 ppb.

The classification errors obtained by dividing the set into *Dark* and *Light* groups and not dividing *Contaminated* and *Uncontaminated* data sets are the same. However, the number of misclassified samples of Contaminated and Uncontaminated groups differs. By dividing the data in Dark and Light groups, the number of misclassified contaminated pepper samples is increased by 2. Since classifying a contaminated pepper as uncontaminated is not preferred, it can be said that extracting features from the whole dataset without dividing by using LDB and selecting the most discriminative features by using wrapper approach is more preferable.

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this thesis, two different approaches; statistical approach and LDB approach; are used to extract the features to classify the aflatoxin contaminated and uncontaminated scaled chili pepper samples. Two different feature selection algorithms are used to select the best features among the extracted features. A standard LDA classifier is used for evaluating the classification accuracy of the selected features.

In the statistical approach, statistical texture features; First Order Statistical (FOS) features and Gray Level Cooccurrence Matrix (GLCM) features are extracted. By leaving one sample out at each step, Fisher distances between features are calculated for the remaining samples and the best 20 features are selected. For each feature a general feature score is computed according to the rank and number of occurrence in all training sets and the best ten features that have the highest scores are selected. Thus, the feature set with 154 features for 14 spectral bands is reduced 10 features. It is desired to achieve the best classification accuracy with fewer features. Two different feature selection algorithms; FDB and wrapper; are used to reduce the feature space dimension and the selected features are fed into LDA. The classification accuracy of 64.58% is obtained by FDB features. Better accuracy of 71.88% is achieved by using wrapper based selected features. The statistical approach does not produce good results in classification of chili peppers.

In LDB approach, the original LDB algorithm is modified to perform 2D searches to extract the most discriminative features from the hyperspectral images. The feature extraction step, reduces the dimensionality of the feature space by removing irrelevant features and/or merging the features that do not have sufficient discriminative information on their own. Two feature trees are generated on spectral and spatial-frequency axis and these feature trees are pruned in order to select the most discriminative features. By leaving out one sample out at each step, one feature map of best 20 features according to their Fisher distance ranks was constructed for each training set and general score of features are calculated with the same way in statistical approach. The feature set of 1190 features is reduced to 11 features after feature extraction and selection. FDB and wrapper based feature selection is applied in order to find the features that gives the best classification. A classification accuracy of 74.31% is achieved by using FDB and a better accuracy of 79.17% is reached by using wrapper based ranked features. The most discriminant features are generally from low subbands of the spectral bands 420, 430, and 600 nm. When the peppers that are classified as contaminated are discarded, the mean aflatoxin level of the final chili pepper lot becomes 15.93 ppb compared to the original set's 38.26 ppb.

Since the best features are selected from the low bands, the data set is examined for further analysis; it is observed that most of the contaminated peppers have lower mean intensity value than uncontaminated peppers. Therefore, the data set is divided into two groups including dark and light peppers and LDB approach is applied to these two groups separately. The most discriminant features are generally from the spectral band of 500 nm for dark peppers and 480 and 600 nm for light peppers. The classification accuracy that is obtained by wrapper based features is 73.61% for dark peppers and 85.42% for light peppers. And average classification accuracy is

79.17%. However, the number of misclassified contaminated pepper samples is increased which is not preferable. In this case, discarding the peppers that are classified as contaminated results in a mean aflatoxin level of 29.24 ppb.

The previous studies (Bollenbacher et al., 1954) (Ashworth et al., 1975) (Doster et al., 1998) show that there is a relationship between aflatoxin contamination and BGYF component of the crops when the reflectance of samples examined under UV illumination. However, Fersaie, McClure and Monroe (1978) found that there is no direct relationship between aflatoxin and BGYF. They observed some contaminated samples do not exhibit BGYF whereas some uncontaminated samples do exhibit BGYF. Our results support the findings of their study. BGYF results from the reaction of host plant peroxidase with the fungi metabolite, kojic acid. Contaminated, yet non-BGYF peppers can occur due to insufficient amount of peroxidase in the plant. Similarly, BGYF can be produced by other fungi which results in uncontaminated-BGYF peppers.

Another possible reason for the mismatch could be the deliberate addition of foreign materials such as aluminum and silicon into peppers to affect the results of chemical analysis. Such fraudulent measures are well-known within the farming community to guarantee passing the safety limits. This results in aflatoxin contaminated peppers to be found as uncontaminated according to chemical analysis. In this case, the ground truth value which we compare our results would be incorrect.

6.1 Future Work

The number of pepper samples used in this study may not be enough for determining the most discriminative features. The same

methodology can be applied to chili peppers by increasing the size of the data set.

We used filters that have fixed FWHM and filters are skipped by using a filter wheel. An electronically tunable filter with tunable FWHM can be used in order to take images from any pass band region. We recently acquired such a device for future research.

In this thesis, the pepper samples are illuminated by UV. They can be illuminated by other light sources covering IR, NIR or VIS region to increase the classification accuracy. We used hyperspectral reflectance imaging in this thesis. As an alternative to reflectance, transmittance of samples can be used. However, pepper samples were non-uniform in their quantity and their scale. Optimal sample thickness should be determined and used throughout the whole pepper set if transmittance images will be used.

While extracting the features, the correlation between features is not taken into consideration. The feature extraction algorithm can be modified to get most discriminative and independent features by considering the correlation between them.

Different texture features such as Markov Random Fields (MRF) or combination of GLCM and wavelet features can be extracted from the hyperspectral data which may lead to better classification results.

The wrapper based feature selection cannot be performed on the whole set of extracted features because of its computational complexity. Parallel processing can be used to try out all the possibilities.

In order to perform classification, LDA is used. Other non-linear classifiers such as Neural Network, SVM can be used in order to achieve better classification accuracy.

.

REFERENCES

Ağaoğlu, S. (1999). "Van İlinde Açıkta Satılan Kırmızı Pul Biberlerde Aflatoksin B1 Varlığının Araştırılması." Van Tıp Dergisi 6(4): 28-30.

Ariana, D. P. and R. Lu (2008). "Quality evaluation of pickling cucumbers using hyperspectral reflectance and transmittance imaging: Part 1. Development of a prototype." Sensing and Instrumentation for Food Quality and Safety 2(3): 144-151.

Arivazhagan, S. and L. Ganesan (2003). "Texture classification using wavelet transform." Pattern Recognition Letters 24(9-10): 1513-1521.

Arrowsmith, M. J., M. R. Varley, et al. (1999). Hybrid Neural Network System for Texture Analysis. Image Processing and Its Applications, Manchester, UK.

Ashworth, L. J., and McMeans, J. L. (1966). "Association of Aspergillus flavus and aflatoxins with a greenish yellow fluorescence of cotton seed. " Phytopathology 56:1104-1105.

Bachari, N. e. i., S. Khodja, et al. (2004). Multispectral Analysis of Satellite Images. ISPRS - International Society for Photogrammetry and Remote Sensing, Istanbul, Turkey.

Baesens, B., T. Van Gestel, et al. (2003). "Benchmarking state-of-the-art classification algorithms for credit scoring." Journal of the Operational Research Society 54(6): 627-635.

Bennett, J. W. and M. Klich (2003). "Mycotoxins." Clinical Microbiology Reviews 16(3): 497-516.

Bharati, M. H., J. J. Liu, et al. (2004). "Image Texture Analysis: Methods and Comparisons." Chemometrics and Intelligent Laboratory Systems 72(1): 57-71.

Bhat, R. V. and S. Vasanthi (2003). "Mycotoxin Food Safety Risk in Developing Countries." Food Safety in Food Security and Food Trade (2020 Focus Briefs)

Bollenbacher, K., and Marsh, P. B. (1954). "A preliminary note on a fluorescent-fiber condition in raw cotton. " Plant Dis. Rep. 38: 375-379.

Bonnier, F., D. Bertrand, et al. (2008). "Detection of pathological aortic tissues by infrared multispectral imaging and chemometrics." The Analyst 133(6): 784-790.

Bornstein, B., M. Gilmore, et al. (2007). Automated Mineral Detection in Visible / Near-Infrared Spectra for Focus-of-Attention. 38th Lunar and Planetary Science Conference, (Lunar and Planetary Science XXXVIII), Texas, USA.

Bothast, R. J. and C. W. Hesseltine (1975). "Bright Greenish-Yellow Fluorescence and Aflatoxin in Agricultural Commodities." Applied Microbiology 30(2): 337-338.

Burget, L. (2004). Complementarity of Speech Recognition Systems and System Combination. Computer Graphics and Multimedia. Brno, Czech Republic, Brno University of Technology. Unpublished Ph.D. Dissertation: 145.

Casasent, D. and X. W. Chen (2004). Aflatoxin detection in whole corn kernels using hyperspectral methods. Proceedings of SPIE, Providence, RI, USA, International Society for Optical Engineering.

Castellano, G., L. Bonilha, et al. (2004). "Texture Analysis of Medical Images." Clinical Radiology 59(12): 1061-1069.

Coifman, R. R. and M. V. Wickerhauser (1992). "Entropy-Based Algorithms for Best Basis Selection." IEEE Transactions on Information Theory 38(2): 713-718.

Çeşmeli, E. and D. Wang (2001). "Texture Segmentation Using Gaussian-Markov Random Fields and Neural Oscillator Networks." IEEE Transactions on Neural Networks 12(2): 394-404.

Dewaele, P., L. Van Gool, et al. (1988). Texture inspection with self-adaptive convolution filters. Proceedings 9th International Conference on Pattern Recognition, Rome, Italy.

Doster, M. A. and T. J. Michailides (1998). "Production of Bright Greenish Yellow Fluorescence in Figs Infected by Aspergillus Species in California Orchards." Plant Disease 82(6): 669-673.

Dror, G. (2007). Analysis of Gene Expression Data, Tel Aviv University.

Du, L. J., J. S. Lee, et al. (1992). Texture Segmentation of Sar Images Using the Wavelet Transform. IEEE International Geoscience and Remote Sensing Symposium (IGARSS '92), Houston, Texas, USA.

Duch, W. (2006). "Filter Methods." Studies in Fuzziness and Soft Computing 207: 89-118.

Duman, A. D., B. Zorlugenç, et al. (2002). "Kahramanmaraş'ta Kırmızı Biberin Önemi ve Sorunları." KSÜ Fen ve Mühendislik Dergisi / KSU J. Science and Engineering 5(1).

EC (2006). European Commission Regulation (EC) No 1881/2006 of 19 December 2006: Setting maximum levels for certain contaminants in foodstuffs, Official Journal of the European Union.

Erdoğan, A. (2004). "The Aflatoxin Contamination of Some Pepper Types Sold in Turkey." Chemosphere 56(4): 321-325.

Fan, G. and X. G. Xia (2003). "Wavelet-based texture analysis and synthesis using hidden Markov models." IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 50(1): 106-120.

Fersaie, A., McClure, W.F. and Monroe, R.J. (1978). "Development of Indices for Sorting Iranian Pistachio Nuts According to Fluorescence", Journal of Food Science 43(3): 1550-1552.

Flietstra, B. C. (2008). A Data Mining Approach for Acoustic Diagnosis of Cardiopulmonary Disease. Operations Research. Massachusetts, USA, Massachusetts Institute of Technology. Unpublished M.S. Thesis: 111.

Gabor, D. (1946). "Theory of communication." Journal of the Institute of Electrical Engineers 93(3): 429-457.

Gomez, R. B. (2002). "Hyperspectral imaging: a useful technology for transportation analysis." Optical Engineering 41(9): 2137-2143.

Gonzalez, R. C. and R. E. Woods (2002). Digital Image Processing (Second Edition), Prentice Hall.

Hadavi, E. (2005). "Several Physical Properties of Aflatoxin-Contaminated Pistachio Nuts: Application of BGY Fluorescence for Separation of Aflatoxin-Contaminated Nuts." Food Additives and Contaminants 22(11): 1144-1153.

Hall, M. A. (1999). Correlation-based Feature Selection for Machine Learning. Philosophy. Hamilton, New Zealand, The University of Waikato. Unpublished Ph.D. Dissertation: 198.

Haralick, R. M. (1979). "Statistical and structural approaches to texture." Proceedings of the IEEE 67(5): 786-804.

Haralick, R. M., K. Shanmugam, et al. (1973). "Textural Features for Image Classification." IEEE Transactions on Systems, Man and Cybernetics 3(6): 610-621.

Harper, P. R. (2005). "A review and comparison of classification algorithms for medical decision making." Health Policy 71(3): 315-331.

He, P. (2005). Classification methods and applications to mass spectral data. Philosophy. Kowloon Tong, Hong Kong, Hong Kong Baptist University. Unpublished Ph.D. Dissertation: 128.

He, Y., X. Li, et al. (2007). "Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model." Journal of Food Engineering 79(4): 1238-1242.

Hesseltine, C. W., O. L. Shotwell, et al. (1966). "Aflatoxin Formation by Aspergillus flavus." Bacteriological Reviews 30(4): 795-805.

Ip, F., J. M. Dohm, et al. (2006). "Flood Detection and Monitoring with the Autonomous Sciencecraft Experiment Onboard EO-1." Remote Sensing of Environment 101(4): 463-481.

Jacobsen, B. J. and R. W. Coppock (1993). "Mycotoxins and Mycotoxicoses."

Kalkan, H. (2008). Feature Extraction from Acoustic and Hyperspectral Data by 2D Local Discriminant Bases Search. Information Systems. Ankara, Turkey, Middle East Technical University. Unpublished Ph.D. Dissertation: 90.

Karaman, S. and B. Acar (2006). "Uluslararası Gıda Ürünleri Ticareti ve Aflataoksin Yasal Düzenlemeleri / International Food Trade and Legislation for Aflatoxins." Doğuş Üniversitesi Dergisi 7(2): 190-197.

Kim, Y., W. N. Street, et al. (2003). "Feature Selection in Data Mining." Data Mining: Opportunities and Challenges.

Konak, E. S. (2002). A Content-Based Image Retrieval System for Texture and Color Queries. Computer Engineering. Ankara, Turkey, Bilkent University. Unpublished M.S. Thesis: 80.

Kramer, D. and F. Aghdasi (1999). Texture analysis techniques for the classification of microcalcifications in digitized mammograms. Proceedings of the IEEE AFRICON '99 Conference, Cape Town, South Africa.

Kumar, S., J. Ghosh, et al. (2001). "Best-bases feature extraction algorithms for classification of hyperspectral data." IEEE Transactions on Geoscience and Remote Sensing 39(7).

Lam, S. W. C. (1996). Texture feature extraction using gray level gradient based co-occurrence matrices. IEEE International Conference on Systems, Man, and Cybernetics, Beijing, China.

Latif-Amet, A., A. Ertüzün, et al. (2000). "An efficient method for texture defect detection: sub-band domain co-occurrence matrices." Image and Vision Computing 18(6-7): 543-553.

Lee, C. H. and H. C. Yang (2009). "Construction of supervised and unsupervised learning systems for multilingual text categorization." Expert Systems With Applications 36(2): 2400-2410.

Lee, K., S. Kang, et al. (2008). "Correlation analysis of hyperspectral imagery for multispectral wavelength selection for detection of defects on apples." Sensing and Instrumentation for Food Quality and Safety 2(2): 90-96.

Livens, S., P. Scheunders, et al. (1997). Wavelets for texture analysis, an overview. International Conference on Image Processing and Its Applications, Dublin, Ireland.

Lotte, F., M. Congedo, et al. (2007). "A review of classification algorithms for EEG-based Brain-Computer Interfaces." Journal of Neural Engineering 40(4).

Mahesh, S., A. Manickavasagan, et al. (2008). "Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes." Biosystems Engineering 101(1): 50-57.

Martin, M. E., M. B. Wabuyele, et al. (2006). "Development of an Advanced Hyperspectral Imaging (HSI) System with Applications for Cancer Detection." Annals of Biomedical Engineering 34(6): 1061-1068.

Martins, M. L., H. M. Martins, et al. (2001). "Aflatoxins in Spices Marketed in Portugal." Food Additives and Contaminants 18(4): 315-319.

Materka, A. and M. Strzelecki (1998). "Texture analysis methods - A review." COST B11 Report.

Merry, R. J. E. (2005). Wavelet Theory and Applications - A literature study. Eindhoven, Netherlands: 49.

Michalak, K. and H. Kwasnicka (2006). "Correlation-based feature selection strategy in classification problems." International Journal of Applied Mathematics and Computer Science 16(4): 503-511.

Mitchell, T. A., R. Bowden, et al. (2000). "Efficient texture analysis for industrial inspection." International Journal of Production Research 38(4): 967-984.

Murphy, P. A., S. Hendrich, et al. (2006). "Food Mycotoxins: An Update." Journal of Food Science 71(5): 51-65.

Ogden, R. T. (1997). Essential Wavelets for Statistical Applications and Data Analysis, Birkhauser.

Paget, R. and I. D. Longstaff (1998). "Texture Synthesis Via a Noncausal Nonparametric Multiscale Markov Random Field." IEEE Transactions on Image Processing 7(6): 925-931.

Pearson, T. C., Wicklow, D.T., et al. (2004). "Reduction of Aflatoxin and Fumonisin Contamination in Yellow Corn by High-speed Bi-chromatic Sorting". Cereal Chemistry. 81(4): 490-498.

Pei, S. C. and J. J. Ding (2007). "Relations Between Gabor Transforms and Fractional Fourier Transforms and Their Applications for Signal Processing." IEEE Transactions on Signal Processing 55(10): 4839-4850.

Qin, J., T. F. Burks, et al. (2008). "Citrus Canker Detection Using Hyperspectral Reflectance Imaging and PCA-based Image Classification Method." Sensing and Instrumentation for Food Quality and Safety 2(3): 168-177.

Rajpoot, K. M. and N. M. Rajpoot (2004). Wavelets and support vector machines for texture classification. Proceedings 8th IEEE International Multitopic Conference (INMIC'2004), Lahore, Pakistan.

Rajpoot, N. (2003). Local Discriminant Wavelet Packet Basis for Texture Classification. Proceedings SPIE Wavelets X (Wavelets X '2003), San Diego, California, USA, International Society for Optical Engineering.

Ruiz, L. A., A. Fdez-Sarria, et al. (2004). Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. ISPRS - International Society for Photogrammetry and Remote Sensing, Istanbul, Turkey.

Saito, N. and R. R. Coifman (1995). "Local discriminant bases and their applications." Journal of Mathematical Imaging and Vision 5(4): 337-358.

Shah, C. P. (2006). Detection of Pecan Weevil Larvae In Pecan Nutmeat Using Multispectral Imaging. Oklahoma, USA, Oklahoma State University. Unpublished M.S. Thesis: 101.

Sheng, S. (2004). Best Bases for Classification. The Pennsylvania State University CiteSeer Archives, The Pennsylvania State University.

Sheshadri, H. S. and A. Kandaswamy (2006). "Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms." Medical Imaging and Information Sciences 23(3): 105-107.

Siddiqi, A. M., H. Li, et al. (2008). "Use of Hyperspectral Imaging to Distinguish Normal, Precancerous, and Cancerous Cells." CANCER (Cancer Cytopathology) 114(1): 13-21.

Siew, L. H., R. M. Hodgson, et al. (1988). "Texture measures for carpet wear assessment." IEEE Transactions on Pattern Analysis and Machine Intelligence 10(1): 92-105.

Sinha, A. and S. Gupta (2007). Approximation of Conditional Density of Markov Random Field and Its Application to Texture Synthesis. IEEE International Conference on Image Processing (ICIP'2007), San Antonio, Texas, USA.

Sinha, S. K., P. S. Routh, et al. (2003). "Time-Frequency Attribute of Seismic Data using Continuous Wavelet Transform." 73rd Annual International Meeting (The Society of Exploration): 1481-1484.

Smith, A. M. S., M. J. Wooster, et al. (2002). "Texture Based Feature Extraction: Application to Burn Scar Detection in Earth Observation Satellite Sensor Imagery." International Journal of Remote Sensing 23(8): 1733-1739.

Steiner, W. E., K. Brunschweiler, et al. (1992). "Aflatoxins and Fluorescence in Brazil Nuts and Pistachio Nuts." Journal of Agricultural and Food Chemistry 40(12): 2453-2457.

Stroka, J. and E. Anklam (2002). "New Strategies for the Screening and Determination of Aflatoxins and the Detection of Aflatoxin-Producing Moulds in Food and Feed." Trends in Analytical Chemistry 21(2): 90-95.

Sun, Z., G. Bebis, et al. (2002). Quantized wavelet features and support vector machines for on-road vehicle detection. Seventh International Conference on Control, Automation, Robotics and Vision (ICARCV '2002), Singapore.

Talavera, L. (2005). An evaluation of filter and wrapper methods for feature selection in categorical clustering. Lecture Notes in Computer Science. 3646: 440-451.

Tang, X. (1998). "Texture Information in Run-Length Matrices." IEEE Transactions on Image Processing 7(11): 1602-1609.

Thiemann, S. and H. Kaufmann (2002). "Lake Water Quality Monitoring Using Hyperspectral Airborne Data - a Semiempirical Multisensor and Multitemporal Approach for the Mecklenburg Lake District, Germany." Remote Sensing of Environment 81(2): 228-237.

Tilley, D. R., M. Ahmed, et al. (2003). "Hyperspectral Reflectance of Emergent Macrophytes as an Indicator of Water Column Ammonia in an Oligohaline, Subtropical Marsh." Ecological Engineering 21(2-3): 153-163.

Tuceryan, M. and A. K. Jain (1993). Texture Analysis. Handbook of Pattern Recognition & Computer Vision (2nd Edition). Singapore, World Scientific Publishing Co.: 207-248.

Üner, Y., Ö. Çetin, et al. (2000). "Baharatta Görülen Mikrobiyolojik Kontaminsayonlar / Microbiological Contaminations Observed in Spices." İstanbul Üniversitesi Veteriner Fakültesi Fakülte Dergisi 26(1).

Vrabcheva, T. (2006). Activities on Food Safety and Quality with a Focus on Mycotoxins. Scientific & Technical Workshop on Food Safety and Quality in JRC Information Day and S&T Workshops, Sofia, Bulgaria.

Wan, Y. Y., J. X. Du, et al. (2004). Bark texture feature extraction based on statistical texture analysis. Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP 2004), Hong Kong, China.

White, R. L. (1997). "Object classification in astronomical images." Statistical Challenges in Modern Astronomy II: 135-148.

Wicklow, D. T. (1999). "Influence of Aspergillus Flavus Strains on Aflatoxin and Bright Greenish Yellow Fluorescence of Corn Kernels." Plant Disease 83(12): 1146-1148.

Yao, H., Z. Hruska, et al. (2006). Hyperspectral bright greenish-yellow fluorescence (BGYF) imaging of aflatoxin contaminated corn kernels. Proceedings of SPIE, Boston, MA, USA, International Society for Optical Engineering.

Zeringue, H. J. and B. Y. Shih (1998). "Extraction and Separation of the Bright-Greenish-Yellow Fluorescent Material from Aflatoxigenic Aspergillus spp. Infected Cotton Lint by HPLC-UV/FL." Journal of Agricultural and Food Chemistry 46: 1071-1075.

Zhang, S., X. Xue, et al. (2005). Feature Extraction and Classification with Wavelet Transform and Support Vector Machines. IEEE International Geoscience and Remote Sensing Symposium (IGARSS '05), Seoul, Korea.

APPENDIX

Aflatoxin B1, B2, G1 & G2 Levels and Total Aflatoxin Levels of Scaled Chili Pepper Samples

Aflatoxin B1, B2, G1 and G2 levels and total aflatoxin levels of scaled chili pepper samples are given according to their total aflatoxin levels in ascending order in the table below.

Sample Number	City	B1(ppb)	B2(ppb)	G1(ppb)	G2(ppb)	Total (ppb)
1	Ankara	1,31	0,00	0,00	0,00	**1,31**
2	Kahramanmaraş	2,56	0,12	0,35	0,00	**2,56**
3	Ankara	2,65	0,10	0,00	0,00	**2,65**
4	Diyarbakır	2,93	0,88	0,11	0,00	**2,93**
5	Erzincan	3,12	0,15	0,49	0,00	**3,12**
6	Ankara	3,63	0,11	0,27	0,00	**3,63**
7	Ankara	4,46	0,23	0,00	0,00	**4,46**
8	Erzincan	4,88	0,11	0,49	0,00	**4,88**
9	Ankara	5,16	0,45	0,00	0,00	**5,16**
10	Kahramanmaraş	5,49	0,23	0,24	0,00	**5,49**
11	Kahramanmaraş	5,49	0,13	0,51	0,06	**5,49**
12	Kahramanmaraş	6,46	0,28	0,36	0,00	**6,46**
13	Hatay	8,10	0,42	0,40	0,00	**8,10**
14	Diyarbakır	8,27	0,42	0,48	0,00	**8,27**
15	Ankara	8,48	0,45	0,14	0,00	**8,48**
16	Ankara	9,90	0,53	0,00	0,00	**9,90**
17	İzmir	15,83	1,76	0,31	0,00	**17,90**
18	Ankara	17,67	0,69	1,31	0,00	**19,67**
19	Kahramanmaraş	11,12	0,36	9,25	0,18	**20,91**
20	Kahramanmaraş	21,70	1,48	1,19	0,00	**24,37**
21	Antalya	21,19	3,32	0,67	0,07	**25,25**
22	İzmir	23,03	2,71	1,34	0,13	**27,21**
23	Ankara	26,53	0,74	0,45	0,00	**27,72**
24	Kahramanmaraş	25,95	2,06	1,89	0,00	**29,90**
25	Kahramanmaraş	23,58	1,72	7,02	0,00	**32,32**
26	İstanbul	30,77	1,82	1,25	0,04	**33,88**
27	Kahramanmaraş	32,54	2,67	0,96	0,00	**36,17**

Sample Number	City	B1(ppb)	B2(ppb)	G1(ppb)	G2(ppb)	Total (ppb)
28	Antalya	34,37	2,17	1,88	0,08	**38,50**
29	Diyarbakır	39,22	3,72	1,02	0,00	**43,96**
30	İstanbul	44,81	3,20	1,29	0,10	**49,40**
31	Antalya	44,32	2,94	4,67	0,17	**52,10**
32	Sivas	53,30	2,14	0,24	0,07	**55,75**
33	Hatay	58,72	3,69	0,40	0,00	**62,81**
34	Ankara	67,10	4,54	0,19	0,00	**71,83**
35	Antalya	65,92	3,62	8,41	0,47	**78,42**
36	Antalya	69,93	4,66	9,02	0,46	**84,07**
37	Ankara	73,28	8,13	2,60	0,56	**84,57**
38	Antalya	85,95	9,76	1,48	0,11	**97,30**
39	Kahramanmaraş	170,48	13,86	4,27	0,29	**188,90**
40	Diyarbakır	196,90	11,13	34,81	1,57	**244,41**