



12-2003

Combination of PCA and DWT features from hyperspectral images for skin tumor detection

John Thomas Fletcher

Follow this and additional works at: https://trace.tennessee.edu/utk_gradthes

Recommended Citation

Fletcher, John Thomas, "Combination of PCA and DWT features from hyperspectral images for skin tumor detection." Master's Thesis, University of Tennessee, 2003.
https://trace.tennessee.edu/utk_gradthes/5223

This Thesis is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Masters Theses by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a thesis written by John Thomas Fletcher entitled "Combination of PCA and DWT features from hyperspectral images for skin tumor detection." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Seong Kong, Major Professor

We have read this thesis and recommend its acceptance:

Accepted for the Council:

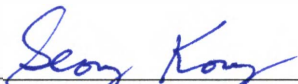
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

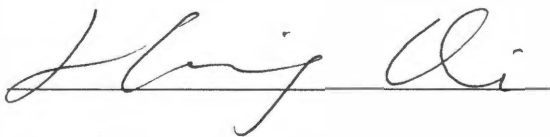
To the Graduate Council:

I am submitting herewith a thesis written by John Thomas Fletcher entitled "Combination of PCA and DWT Features from Hyperspectral Images for Skin Tumor Detection." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

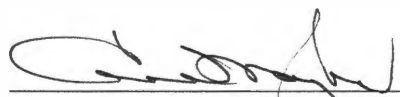

Seong G. Kong, Major Professor

We have read this thesis and
recommend its acceptance:





Acceptance of the Council:


Vice Provost and Dean of
Graduate Studies

Thesis
2003
.F58

Combination of PCA and DWT Features from Hyperspectral Images for Skin Tumor Detection

A Thesis
Presented for the
Masters of Science
Degree
The University of Tennessee, Knoxville

John Thomas Fletcher
December 2003

Acknowledgement

I would like to express my sincerest thanks to my major advisor, Dr. Seong G Kong, for his support, understanding and encouragement during the preparation of this thesis. He has provided excellent technical input and support for the completion of this work. I also thank the other graduate committee members, Dr. J. Wesley Hines and Dr. Hairong Qi, for their comments, assistance and encouragement.

To Dr. Yud-Ren Chen and Dr. Moon S. Kim of the Instrumentation and Sensing Laboratory (ISL), Beltsville Agricultural Research Center, Beltsville, Maryland, for providing their hyperspectral fluorescence images for this research.

Abstract

In this paper, the detection of skin tumors using hyperspectral fluorescence images of poultry carcasses is investigated. Skin tumors are not always visually obvious. The visual region of the spectrum may be too limited to meet all the requirements so that the tumors may be accurately classified so the multiple bands from hyperspectral imaging may be of some use. Each of the hyperspectral fluorescence images will consist of 65 spectral bands ranging from 425 nm to 711 nm. Multiple detection schemes will be utilized to provide adequate classification rates. Principal component analysis (PCA) and discrete wavelet transforms (DWT) are utilized to transform the data from the spectral space to a feature space. A small number of features are selected to provide dimensionality reduction without a significant loss of information. A support vector machine (SVM) classifier is used to determine if a pixel falls in a normal skin or tumorous skin category. To provide additional classifier accuracy, an algorithm based on the average intensity of the pixel signal, is used to combine the two classifiers. The accuracy of the three classifiers was tested using 11 hyperspectral fluorescence images with a combined total of 38 tumors. The PCA-SVM classifier provided a tumor detection rate of 86.8% with 17 false positives and 5 missed tumors. The DWT-SVM classifier provided a tumor detection rate of only 42.1% with 18 false positives and 12 missed tumors. The classifier that selected the best method showed that the PCA/DWT-SVM classifier provided a classification rate of 94.7% with 6 false positives and only one missed tumor.

Table of Contents

1. Introduction.....	1
2. Background.....	6
2.1 Hyperspectral Imaging.....	6
2.2 Previous Work.....	7
3. Feature Extraction.....	11
3.1 Principal Component Analysis.....	11
3.2 Discrete Wavelet Transform.....	16
3.3 Feature Selection.....	22
4. Support Vector Machines for Classification	25
5. Combination of PCA and DWT Feature Extraction Schemes	31
6. Experimental Results	39
6.1 Imaging System	39
6.2 Performance Comparisons.....	44
7. Conclusion	53
7.1 Summary.....	53
7.2 Recommendation for Future Work	54
Bibliography	56
VITA	61

List of Figures

Figure 1: Representation of 3D nature of hyperspectral images.....	6
Figure 2: Eigenvector breakdown of the original signal.....	17
Figure 3: Reconstruction of the signal from the eigenvectors	17
Figure 4: Dyadic DWT filter tree implementation with energy feature extraction	21
Figure 5: Original signal spectrum of a normal skin sample.....	21
Figure 6: Level 1 DWT approximation and detail coefficients of a normal skin sample.	22
Figure 7: Level 2 DWT approximation and detail coefficients of a normal skin sample.	23
Figure 8: Plot of decision surfaces created by support vector machine using the first two principal components as inputs	30
Figure 9: Spectral signature of normal tissue, tumor, and background	31
Figure 10: Block diagram of the proposed multi-channel classifier.....	34
Figure 11: Information content (in percent) of the first 10 principal components	35
Figure 12: Decision boundaries of the PCA-SVM classifier with a radial basis kernel...	35
Figure 13: Decision boundaries of the DWT-SVM classifier with a radial basis kernel..	37
Figure 14: Combination feature to determine classifier selection	38
Figure 15: Diagram of the hyperspectral fluorescence imaging system.....	41
Figure 16: Mosaic of the hyperspectral image of a poultry carcass sample displaying bands 5, 10 thru 60.....	43
Figure 17: Decision boundary of the PCA-SVM classifier	44
Figure 18: Tumors detected by the PCA-SVM classifier on the training image	45
Figure 19: Tumors detected by the PCA-SVM classifier on one of the testing images ...	46

Figure 20: Decision boundary of the DWT using the first two features..... 46

Figure 21: Tumors detected by the DWT-SVM classifier on the training image 47

Figure 22: Tumors detected by the DWT-SVM classifier on the testing image 48

Figure 23: Tumors detected by the PCA/DWT-SVM classifier on the training image 49

Figure 24: Tumors detected by the PCA/DWT-SVM classifier on the testing image 50

Figure 25: Tumor detection rate for the PCA-SVM, DWT-SVM and PCA/DWT-SVM
classifiers..... 55

1. Introduction

Nonmelanoma skin cancer is the most common malignancy worldwide. Unfortunately there are very few options for the medical clinician to diagnosis the skin cancer. The most definitive test has been a biopsy. Since the skin is readily accessible, this procedure is relatively straight forward. Because of the accessibility of the skin, spectroscopy has been used extensively in the desire to develop a non invasive diagnostic procedure. In this paper poultry carcasses are used to provide the tumorous tissue for this investigation.

Hyperspectral imaging reveals useful information for detecting skin tumors on poultry carcasses. Poultry skin tumors are typically ulcerous lesions that are surrounded by a rim of thickened skin and dermis. Skin tumor is not as visually obvious as other pathological diseases like septicemia, airsac, and bruise since its signature appears as shape distortion rather than discoloration [1]. Since the visual region may be too limited to meet the requirements of quality inspection, multiple bands from a hyperspectral image may be needed to extract unwholesome conditions in a material [2]. Fluorescence is a phenomenon where light is absorbed at a given wavelength and then is normally followed by the emission of light at a longer wavelength. There are a number of compounds that emit fluorescence in the visible range when excited with ultraviolet radiation. The altered biochemical and morphological state of the neoplastic tissue is reflected in the spectral characteristics of the measured fluorescence. The theory behind fluorescence enhanced optical imaging involves excitation of fluorophores from within a tissue medium by an

external ultraviolet light source. The excitation results in the emission of fluorescence. Fluorescence techniques are generally regarded as sensitive optical tools and have proven to be useful in a number of scientific areas [3]. Image samples are obtained from a hyperspectral fluorescence imaging system for 65 spectral bands whose wavelength is ranged from 425[nm] to 711[nm].

In the past few years studies have been performed which investigate the use of hyperspectral imaging and fluorescence to detect abnormalities and/or malignancies in skin tissue. In Brancalion et al. [4], in vivo tissue of 18 patients, who were diagnosed with nonmelanoma skin cancers, had the cancerous areas excited by ultraviolet light. Under this excitation, the fluorescence in both basal cell carcinomas (BCC) and squamous cell carcinomas (SCC) showed a distinctly different spectrum. It was also determined that the resultant fluorescent spectrum changed notably as the excitation wavelength was changed. Both the BCC and SCC lesions upon excitation at 295 nm, consistently displayed a larger fluorescence from the lesion compared to the surrounding normal tissue. When excitation was changed to 350nm the fluorescence was reduced in the lesion relative to the normal tissue. In Chang et al. [5], a fiber optic probe was used to direct excitation light to the tissue and collect the fluorescence and diffusely reflected light. The fluorescence measurements consisted of 16 excitation wavelengths, ranging from 330 nm to 480 nm in 10 nm intervals. The reflectance spectra are measured through 4 different source-detector separations. Combinations of the fluorescence and reflectance measurements, consisting of single measurement and groups of two and three, were

investigated using multivariate algorithms to discriminate the pairwise diagnostic classes. PCA was used to reduce the dimensionality of the data set. A Mahalanobis distance based classification algorithm was trained and tested using each of the data sets. It was shown that a combination of fluorescence excitation wavelengths of 340 nm, 350 nm and 410 nm along with a reflectance measurement would provide the best spectral results for classification performance. In Ramanujam et al. [6], optical spectroscopic measurements were made with a fiberoptic probe in contact with each tissue sample. Fluorescence spectra were measured at wavelengths ranging from 300 nm to 460 nm in 20 nm increments. Each emission spectra was measured in 5 nm increments. Diffuse reflectance was measured from 300 nm to 850 nm in increments of 5 nm. Multivariate statistical analysis of the spectral data was performed, which included using PCA, to dimensionally reduce the large number of spectral variables and incorporate the principal component scores into a classification method. It was determined that the optimal excitation wavelengths of 300, 320 and 340 nm obtained a fluorescent spectrum that would provide an effective means of discrimination between malignant and normal tissues.

This paper presents detection of skin tumor on poultry carcasses using hyperspectral fluorescent images. Two methods are investigated to reduce the dimensionality of the hyperspectral fluorescent data. More than one method is used because it was determined through experimentation that a single classifier system was unable to fully discriminate between wholesome and unwholesome areas on the poultry carcass. This was due to the

uneven illumination or fluorescence caused by the curvature and varying elevations of the carcass. The first is to use principal component analysis (PCA) technique to find an efficient representation of the data at reduced dimensional feature space. The PCA technique was chosen because it is a procedure that has been used in many fields with good results. Since PCA transforms spectral signature data into a set of uncorrelated data called principal components. A subset of the Principal Components of the data are found and used to train a Support Vector Machine (SVM) classifier. The SVM classifier was chosen because it produces a decision surface or boundary that maximize the margin or the distance between the boundary and the nearest points in the training set. In the second method a Discrete Wavelet Transform (DWT) is performed on the data set and the approximation and detail coefficients are compacted by taking the RMS value of these coefficients. This compacted data is then used to train a SVM classifier. The combination of these two classifiers, the PCA-SVM and DWT-SVM, provides an increased tumor detection rate. The PCA/DWT-SVM classifier combines the areas of high detection rate of the individual classifiers. This allows the classification rate of the combined classifier to exceed the classification rates achieved by either of the individual classifiers. The combination of the classifiers is performed by an algorithm that measures the average intensity of the pixel spectrum and selects the classifier that should provide the best classification.

The objective of this paper is to develop an approach for the detection of skin tumors. In Chapter 2, a brief discussion of hyperspectral imaging is provided. A review of a few

examples of previous work in the area of automated skin tumor detection is presented. In Chapter 3, the methods used for feature extraction are discussed. Principal component analysis and discrete wavelet transforms are presented with a short discussion of their background and use. The methods used to reduce the dimensionality of the feature space are also presented. In Chapter 4, the support vector machine classifier is introduced with a discussion of its background. In Chapter 5, the feature extraction schemes are presented for the PCA-SVM and DWT-SVM classifiers. The methods by which the reduced feature set used as the inputs to the SVM classifiers are explained. The method to combine the two individual classifiers to increase the detection rate is discussed. In Chapter 6, the imaging system used to capture the hyperspectral fluorescence images is presented. The performance results of the PCA-SVM, DWT-SVM and the PCA/DWT-SVM classifiers are presented and compared. Concluding remarks and recommendations for future work are presented in Chapter 7.

2. Background

2.1 Hyperspectral Imaging

Hyperspectral imaging systems are found useful as no single band image has sufficient information to describe the information of the scene completely. Data produced by hyperspectral imaging systems constitute a three-dimensional (3D) cube of data, stacking multiple 2D single-band images along the spectral axis [7]. A hyperspectral image contains spatial information measured at a sequence of individual wavelength across a sufficiently broad spectral band at high-resolution (10 [nm]) spectrum. Figure 1 demonstrates a conceptualized view of a hyperspectral image data.

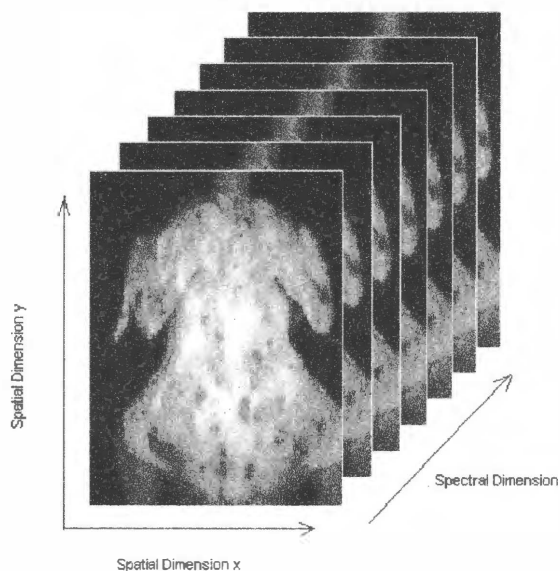


Figure 1: Representation of 3D nature of hyperspectral images

A hyperspectral image can be denoted by $I(u, v, \lambda_i)$, where indices $u = 0, 1, \dots, N-1$, $v = 0, 1, \dots, M-1$ are spatial coordinates and λ_i , $i = 0, 1, \dots, L-1$ indicates the spectral band. For a fixed λ_k , $I(u, v, \lambda_k)$ represents the k^{th} single-band image. If u and v are fixed, then $I(u, v, \lambda)$ stands for the spectral information

Spectral component often involve quantities that are related to the scene phenomenology as reflectance, fluorescence, and emissivity. The resulting spectra can be used in principle to uniquely characterize and identify any given material. To analyze the data some choices need to be made in analyzing the data. How subtle are the discriminations to be made? What is the best spectral dimensionality to be used in the classifier? How is the feature space to be determined? A high dimensional data set is sent through a feature extraction process which will define the possible subspaces based on the descriptions of the specified classes the user desires. Feature extraction algorithms can be used to generate subspace of reduced dimensionality that will suffice for the classes required. The user can then select a subset of the feature space based on the subtly of the classes and the precision required. A classification algorithm is then chosen that can adequately interpret the subset of the feature space with the accuracy required.

2.2 Previous Work

In the paper authored by Chao et al. [8], two black and white cameras take images of a poultry carcass. One of the cameras uses a band-pass filter of 540 nm and the other uses a band pass filter of 700 nm. These wavelengths were found in an earlier paper by Park

and Chen. The two images were reduced to a size of 256 x 240 pixels. The poultry carcass was then segmented from the background using thresholding. The image was divided into 15 horizontal layers with 16 horizontal pixels in each layer. A centroid was calculated for each layer and based on this centroid each layer was divided into several square blocks (16 x 16 pixels). A constant number of blocks (107) were chosen to omit the legs and wings and to delineate the main body of the poultry carcass. The average intensity of each block was used as an input to the neural network classifier. The neural network classifier consisted of 107 input neurons, 10 hidden layer neurons and 2 output neurons. Four models were developed using combinations of the Delta and Norm-cum-delta learning rules and the Tanh and Sigmoid activation functions. In off-line training, validation accuracies were approximately 91% for both the images taken at 540 nm and 700 nm. During the on-line testing the results of the 540 nm image and the 700 nm image were combined using a logic AND function. The wholesome carcasses achieved a 94% classification accuracy and the unwholesome carcasses achieved an 87% classification accuracy.

In the paper by Kong [9], a hyperspectral fluorescence image consisting of 65 bands whose wavelength ranged from 425 nm to 711 nm, of a tumorous chicken carcass was used to develop and test Fuzzy Logic classifier. The hyperspectral image was first reduced in size by means of a 2-dimension Discrete Wavelet Transform utilizing a Daubechies mother wavelet of order 5 (db5). This reduced the size of the image to one-fourth of its original size while maintaining much of the original information. From this

compressed image two prominent features were chosen. Three fuzzy decision rules were developed utilizing the adopted Gaussian membership functions. The use of this small number of rules and the min-max inference scheme could successfully detect poultry skin tumors.

In Park et al. [10], two black and white cameras take images of a poultry carcass. One of the cameras uses a band-pass filter of 540 nm and the other uses a band pass filter of 700 nm. The resultant images were reduced in size to 254 x 240 pixels. The reduction was performed by averaging the pixel intensities using a 3 x 2 template. From the reduced image a 64 x 64 pixel sub-image was extracted for the generation of features. This sub-image was selected by finding the centroid of the chicken body and then using the calculated centroid as the center of the sub-image. Because of limitations on the number of inputs allowed in the neural network software the 64 x 64 pixel sub-image was further reduced in size to 16 x 16 pixels by using a 4 x 4 pixel mask. The resultant compressed sub-image was then used as the input to the neural network classifier. The classifier consisted of 256 input neurons when only one image was utilized and 512 input neurons when both images were utilized. The hidden layer had 32 neurons and the output layer consisted of 2 neurons. In a second test scenario, the Fast Fourier Transform (FFT) was taken of the 64 x 64 pixel sub image. Several test runs were performed using each of the single wavelength images as the input and also using both images as the input. The classifier model using both the pixel intensities of the 540 nm and 700 nm spectral

images as input achieved 93.3% classification accuracy. The classifier model using the FFT of the 700 nm spectral image achieved a classification rate of 90.0%.

In Palmer et al. [11], the fluorescence emission spectra were recorded at excitation wavelengths of 300 nm to 460 nm, in 20 nm increments. At each excitation wavelength the fluorescence was recorded from 10 nm greater than the excitation wavelength up to 600 nm. The diffuse reflectance was measured from 300 nm to 600 nm, in 5 nm increments. PCA was used as a data reduction technique. A Wilcoxon rank-sum test was employed to determine which principal component scores showed statistically significant differences between malignant and non-malignant tissue types. A support vector machine was used as the classification technique. This technique determined that only four of the ten measured were required to maximize classification accuracy. These included 300, 400, 420 and 460 nm.

3. Feature Extraction

The high dimensional hyperspectral fluorescence images are subjected to principal component analysis and discrete wavelet transforms. These two mathematical techniques will create a high dimensional feature space. A small number of significant features are selected from this feature space to provide dimensionality reduction. This reduced dimensional feature space will then be sent to a support vector machine classifier which will differentiate between tumorous and normal skin. There are two distinct classification techniques used. The first is to use a small number of principal components sent to a support vector machine classifier. The second will use a set of coefficients created by a discrete wavelet transform which is sent to a separate support vector machine classifier. To further improve the classification accuracy the outputs of the two classifiers are sent to a selection process. This selection process will choose the technique that will provide the most accurate classification. The results of the three classification outputs are presented for comparison.

3.1 Principal Component Analysis

Pearson [12] introduced a multivariate analysis technique, later independently developed by Hotelling [13], and referred to as Principal Component Analysis (PCA). PCA is a technique used to reduce the dimensionality of a data set which consists of a large number of interrelated variables while retaining as much of the variability or information contained in the data. This can be considered as a data compression technique that is

achieved by transforming the data into a new set of variables, principal components (PCs), that are uncorrelated and ordered such that the first few of the PCs contains a large percentage of the variation that is contained in all of the original data. In a sense, PCA puts redundant or correlated information into separate variables and makes the transformed vectors orthogonal and uncorrelated thus removing any co-linearity problems. The process is simply a linear transformation from the original space to a principal component space which is orthogonal.

The following explanation is based on the derivation provided by Jolloffe [14]. For a given vector \mathbf{x} of n variables, the variances of the variables and the structure of the covariance or correlations between the n variables are assumed to be of some interest. If the value of n is large then the amount of data provided in the n variances and the $\frac{1}{2}n(n-1)$ correlations or covariances would be overwhelming to analyze. A technique to overcome this problem would be to obtain a few variables that will retain most of the original information. For this end we assume a vector of n variables \mathbf{x} , where the data are mean-centered since only the variability of the data is of interest. In matrix form for a matrix of m observations, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]^T$. It is desired that this matrix in n dimensional space is transformed into a space of k dimensions where $k \ll n$. This can be seen in Eq. 1 where, \mathbf{T} is the k dimensional transformed space matrix ($\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_M]$), \mathbf{X} is the original m dimensional data matrix and \mathbf{P} is the linear transformation matrix ($\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M]$).

$$\mathbf{T} = \mathbf{XP} \quad (1)$$

The first principal component t_1 , which has a maximum variance, is found by

$$t_1 = \mathbf{x}\mathbf{p}_1. \quad (2)$$

A linear function $\mathbf{x}\mathbf{p}_1$ is needed that maximizes the variance of t_1 , which can be found by maximizing \mathbf{p}_1 .

$$\text{var}(t_1) = \text{var}(\mathbf{x}\mathbf{p}_1) = \mathbf{p}_1^T \mathbf{x}\mathbf{x}^T \mathbf{p}_1 = \mathbf{p}_1^T \mathbf{\Sigma} \mathbf{p}_1 \quad (3)$$

$$\text{where } \mathbf{\Sigma} = \mathbf{x}\mathbf{x}^T$$

To maximize the variance of t_1 , set the derivative of the variance of t_1 ($\text{var}(t_1)$) to zero and include a Lagrange multiplier λ using the constraint $\mathbf{p}_1^T \mathbf{p}_1 = 1$.

$$\frac{\partial}{\partial \mathbf{p}_1} \left(\text{var}(t_1) - \lambda (\mathbf{p}_1^T \mathbf{p}_1 - 1) \right) = 0 \quad (4)$$

$$\frac{\partial}{\partial \mathbf{p}_1} \left(\mathbf{p}_1^T \mathbf{\Sigma} \mathbf{p}_1 + \lambda \mathbf{I} - \lambda \mathbf{p}_1^T \mathbf{p}_1 \right) = 0 \quad (5)$$

$$(\mathbf{\Sigma} - \lambda \mathbf{I}) \mathbf{p}_1 = 0 \quad (6)$$

This λ is an eigenvalue of $\mathbf{\Sigma}$ and \mathbf{p}_1 is the corresponding eigenvector. \mathbf{I} is the $n \times n$ identity matrix so to choose which of the n eigenvectors gives $\mathbf{x}\mathbf{p}_1$ the maximum variance, λ must be as large as possible since

$$\mathbf{p}_1^T \mathbf{\Sigma} \mathbf{p}_1 = \mathbf{p}_1^T \lambda \mathbf{p}_1 = \lambda \mathbf{p}_1^T \mathbf{p}_1 = \lambda \quad (7)$$

So the eigenvector corresponding to the largest eigenvalue of Σ is \mathbf{p}_1 .

The second PC $\mathbf{x}\mathbf{p}_2$ will maximize $\mathbf{p}_2^T \mathbf{x}\mathbf{p}_2$ subject to being uncorrelated with $\mathbf{x}\mathbf{p}_1$ or the covariance between $\mathbf{x}\mathbf{p}_1$ and $\mathbf{x}\mathbf{p}_2$ is equal to zero.

Given that

$$\text{cov}[\mathbf{x}\mathbf{p}_1, \mathbf{x}\mathbf{p}_2] = \mathbf{p}_1^T \Sigma \mathbf{p}_2 = \mathbf{p}_2^T \Sigma \mathbf{p}_1 = \mathbf{p}_2^T \lambda \mathbf{p}_1 = \lambda \mathbf{p}_2^T \mathbf{p}_1 = \lambda \mathbf{p}_1^T \mathbf{p}_2 \quad (8)$$

any of the terms could be used to indicate the zero correlation between $\mathbf{x}\mathbf{p}_1$ and $\mathbf{x}\mathbf{p}_2$. To maximize the variance of t_2 the derivative of the variance of t_2 is set to zero and using the necessary normalization multiplier

$$\frac{\partial}{\partial \mathbf{p}_2} \left(\text{var}(t_2) - \lambda (\mathbf{p}_2^T \mathbf{p}_2 - 1) - \Phi \mathbf{p}_2^T \mathbf{p}_1 \right) = 0 \quad (9)$$

where λ and Φ are Lagrange multipliers. This result in

$$\frac{\partial}{\partial \mathbf{p}_2} \left(\mathbf{p}_2^T \Sigma \mathbf{p}_2 + \lambda \mathbf{I} - \lambda \mathbf{p}_2^T \mathbf{p}_2 - \Phi \mathbf{p}_2^T \mathbf{p}_1 \right) = 0 \quad (10)$$

$$2\mathbf{p}_1^T \Sigma \mathbf{p}_2 - 2\lambda \mathbf{p}_1^T \mathbf{p}_2 - \Phi \mathbf{p}_1^T \mathbf{p}_1 = 0 \quad (11)$$

Since $\mathbf{p}_1^T \Sigma \mathbf{p}_2 = 0$, $\mathbf{p}_1^T \mathbf{p}_2 = 0$ and $\mathbf{p}_1^T \mathbf{p}_1 = 1$ reduces to $\Phi = 0$. This gives

$$(\Sigma - \lambda \mathbf{I}) \mathbf{p}_2 = 0 \quad (12)$$

So λ is the eigenvalue of Σ and \mathbf{p}_2 is the associated eigenvector. With $\lambda = \mathbf{p}_2^T \Sigma \mathbf{p}_2$ so λ needs to be as large as possible. Using the assumption that Σ does not have repeated

eigenvalues, this implies that λ is the second largest eigenvalue of Σ and \mathbf{p}_2 is its eigenvector.

For the general case, to determine the j^{th} principal component the previous constraints hold with the modification that $\mathbf{x}\mathbf{p}_k$ is orthogonal to $\mathbf{x}\mathbf{p}_j$ for $j = 1, 2, \dots, k-1$. To maximize t_k with the given constraints, the derivative of the variance of t_k is set to zero and include a Lagrange multiplier λ_k to ensure $\mathbf{p}_k^T \mathbf{p}_k = 1$ and a $j-1$ Lagrange multiplier Φ_j to ensure $\mathbf{x}\mathbf{p}_k$ is orthogonal to $\mathbf{x}\mathbf{p}_j$ for $j = 1, 2, \dots, k-1$.

$$\frac{\partial}{\partial \mathbf{p}_k} \left(\text{var}(t_k) - \lambda_k (\mathbf{p}_k^T \mathbf{p}_k - 1) - \sum_{j=1}^{k-1} \Phi_j \mathbf{x} (\mathbf{p}_j^T \mathbf{p}_k) \right) = 0 \quad (13)$$

where λ and Φ are Lagrange multipliers. This result in

$$2\Sigma \mathbf{p}_k - 2\lambda_k \mathbf{p}_k - \mathbf{x} \sum_{j=1}^{k-1} \Phi_j \mathbf{p}_j = 0 \quad (14)$$

Left multiply by \mathbf{p}_j^T for $j = 1, 2, \dots, k-1$

$$2\mathbf{p}_j^T \Sigma \mathbf{p}_k - 2\lambda_k \mathbf{p}_j^T \mathbf{p}_k - \mathbf{x} \sum_{j=1}^{k-1} \Phi_j \mathbf{p}_j^T \mathbf{p}_j = 0 \quad (15)$$

So if $j \neq k$ the first two terms are zero which implies that $\Phi_j = 0$ for $j = 1, 2, \dots, k-1$. This results in

$$(\Sigma - \lambda \mathbf{I}) \mathbf{p}_k = 0 \quad (16)$$

So λ_k is the k^{th} eigenvalue and \mathbf{p}_k is its associated eigenvector.

If the values are combined in the \mathbf{P} matrix is formed by the eigenvectors of $\mathbf{X}^T\mathbf{X}$ that correspond to the associated eigenvalues in decreasing order. The \mathbf{T} matrix consists of the scores of the PCs also in decreasing order. The \mathbf{P} matrix can be used to reconstruct the original data with little loss of information by $\mathbf{X} = \mathbf{TP}^T$.

As an example, a set of 100 samples of poultry skin data were used a principal component analysis. The set consisted of 50 samples each of normal skin and tumorous skin. Using the SVD algorithm [15] it was determined that over 99% of the information was contained in the first two principal components. Figure 2(a) displays a hyperspectral fluorescent signal of a sample of normal poultry skin [16]. In Figure 2(b), the first three eigenvectors of the decomposition of the signal are displayed. In Figure 3(a), the reconstruction of the original signal using the eigenvectors of the decomposition is shown. It can be seen that since a large majority of the information was contained in the first two principal components, the signal can be reconstructed with good accuracy using very few of the eigenvectors. Figure 3(b) shows the energy or information content of the first few principal components. It can be seen that the first few principal components contain nearly all of the energy.

3.2 Discrete Wavelet Transform

Wavelet Transform decomposes a signal into multi-resolution components using a wavelet function or mother wavelet as the basis function. Wavelets have been used with

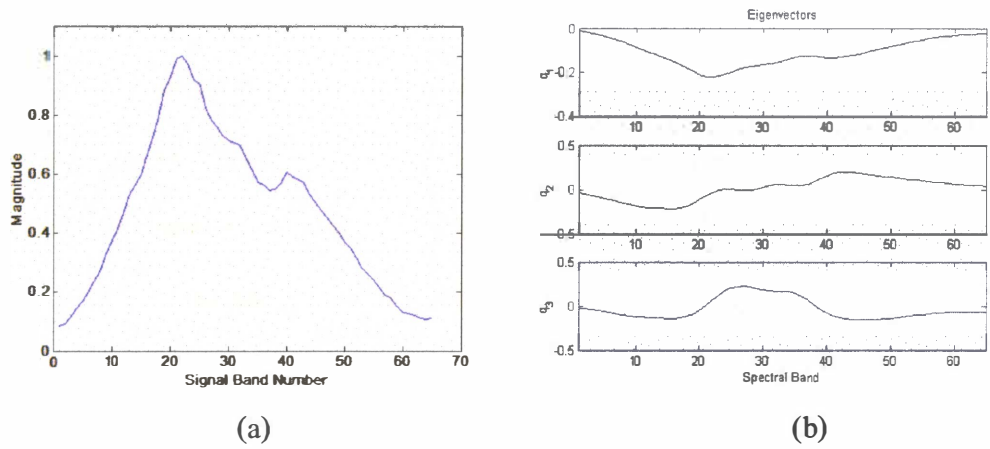


Figure 2: Eigenvector breakdown of the original signal [16]:

(a) Original signal, (b) first three eigenvectors

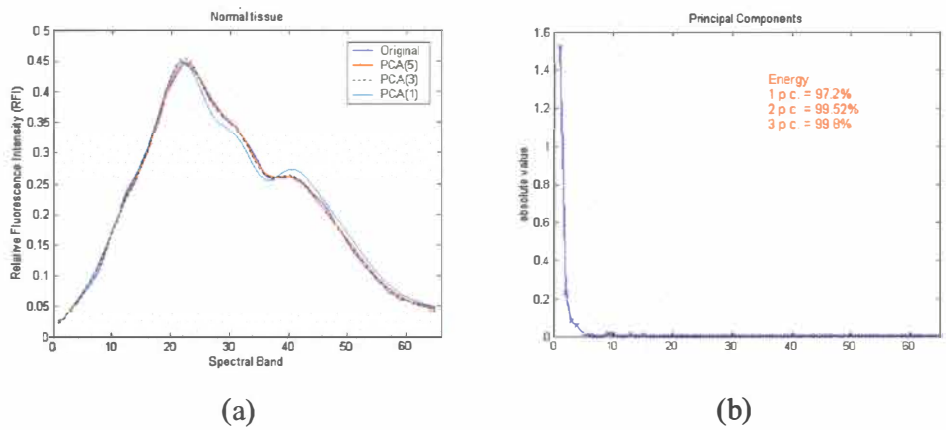


Figure 3: Reconstruction of the signal from the eigenvectors [16]:

(a) Signal reconstruction, (b) Energy content of the first three principal components

success in the classification of hyperspectral imaging. In Bruce et al. [17] the wavelet coefficients are used to create a feature vector which is then reduced in size by Fisher's linear discriminant analysis and then used as the input to a maximum likelihood statistical classifier. Kaewpijit et al. [18] the approximation details were used as the inputs to various statistical classifiers. The support vector machine classifies the feature vector and makes a decision whether each pixel falls in the normal or tumor category.

The wavelet transform cuts up the data into different frequency components and studies each component with a resolution matched to its scale [19]. The DWT can be expressed as an inner product between a signal $f(n)$ and a wavelet basis $\Psi_{j,k}(n)$. Each of the inner products result in a wavelet transform coefficient. The DWT can be expressed as

$$W(f(j,k)) = \sum_{n=0}^{N-1} f(n) \cdot \Psi_{j,k}(n) \quad (17)$$

where $W f(j,k)$ is a DWT coefficient and $f(n)$ is a sequence with length N .

A wavelet basis function $\Psi_{j,k}(n)$ can be generated by shifting and scaling the basic or mother wavelet $\Psi(n)$.

$$\Psi_{j,k}(n) = \frac{1}{\sqrt{j}} \Psi\left(\frac{n-k}{j}\right) \quad (18)$$

Where $j > 0$ and k are real numbers. j is the scaling factor of a basis function and k is the translation variable along the function range. The coefficient is used to normalize the energy of the wavelet. The discretized dyadic wavelet basis is given by

$$\Psi_{j,k}(n) = \frac{1}{\sqrt{2^j}} \Psi\left(\frac{n - 2^j \cdot k}{2^j}\right) \quad (19)$$

or

$$\Psi_{j,k}(n) = 2^{-j/2} \Psi(2^j n - k) \quad (20)$$

The wavelet coefficients are obtained by

$$W_{j,k} = \langle \Psi_{j,k}(n), f(n) \rangle \quad (21)$$

The dyadic DWT can be implemented via the dyadic filter tree algorithm [20]. This algorithm represents the wavelet basis as a set of high and low pass filters. After the filter the signal is decimated by a factor of two. The outputs of the low pass filter are called the approximation coefficients and the outputs of the high pass filter are called the detail coefficients. This decomposition can be iteratively performed until a maximum scale is reached.

For the case presented, the Haar wavelet was chosen as the mother wavelet. The Haar wavelet discontinuous and resembles a step function.

$$\Psi(k) = \begin{cases} 1, & 0 \leq k \leq \frac{1}{2} \\ -1, & \frac{1}{2} \leq k \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

The DWT produces a set of coefficients consisting of one approximation and six detail coefficients. To create a feature vector, the RMS value of each of the coefficients was evaluated by

$$F_i = \sqrt{\frac{1}{K} \sum_{k=1}^K |W_{j,k}|^2} \quad i = 1, 2, \dots, p \quad (23)$$

Where the maximum value of p is determined by $p = \log n$ where n is the number of points in the original signal.

Figure 4 provides a block diagram of the DWT filter tree implementation with the RMS energy extraction for the feature vector. The spectral signal $f(n)$ is the input to the filter tree. Each decomposition level creates an approximation coefficient and detail coefficient from the approximation coefficient on the level above. The wavelet basis is represented as a set of high and low pass filters (HPF and LPF). After the filtering the signal is decimated by a factor of 2. Figure 4, the level 1 approximation coefficients, $W_\phi(1, k)$ are converted into the level 2 approximation, $W_\phi(2, k)$, and detail, $W_\psi(2, k)$, coefficients. The feature vector is created by determining the RMS values of the detail coefficients obtained by the function in Eq. 23. In Figure 5, a sample spectrum of a normal skin sample is shown. There are 64 samples in this spectrum.

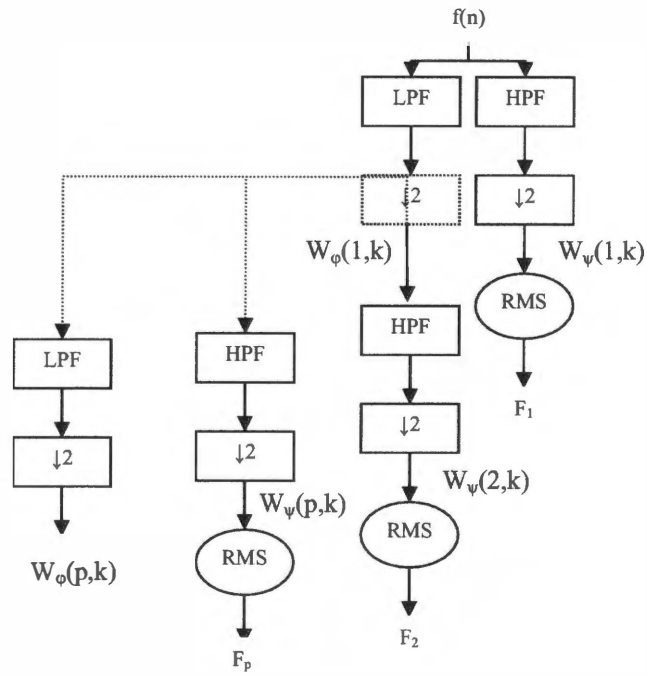


Figure 4: Dyadic DWT filter tree implementation with energy feature extraction

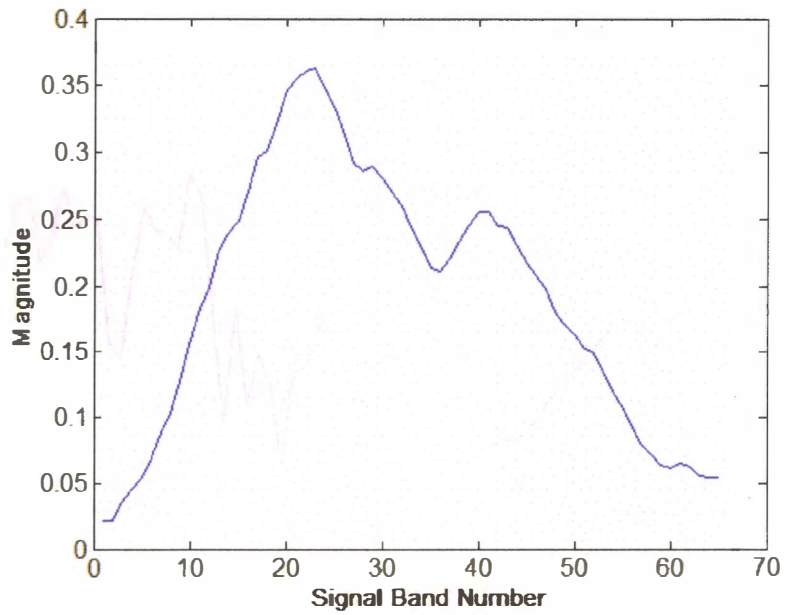


Figure 5: Original signal spectrum of a normal skin sample

In Figure 6(a) the level 1 approximation coefficients are displayed. As one can see the basic shape of the spectrum is retained while the number of samples is reduced by half. The level 1 detail coefficients are displayed in Figure 6(b). It also contains half the samples as the original signal. As indicated in the filter tree implementation. In Figure 7(a) the level 2 approximation coefficients are shown. It has half as many samples as the level 1 approximation coefficient. The level 2 detail coefficients are shown in Figure 7(b).

3.3 Feature Selection

Mapping the data from one space to another can greatly simplify the task of machine learning. This has allowed the creation of techniques for selecting the best representation of the data. The quantities introduced to describe the data are usually called features.

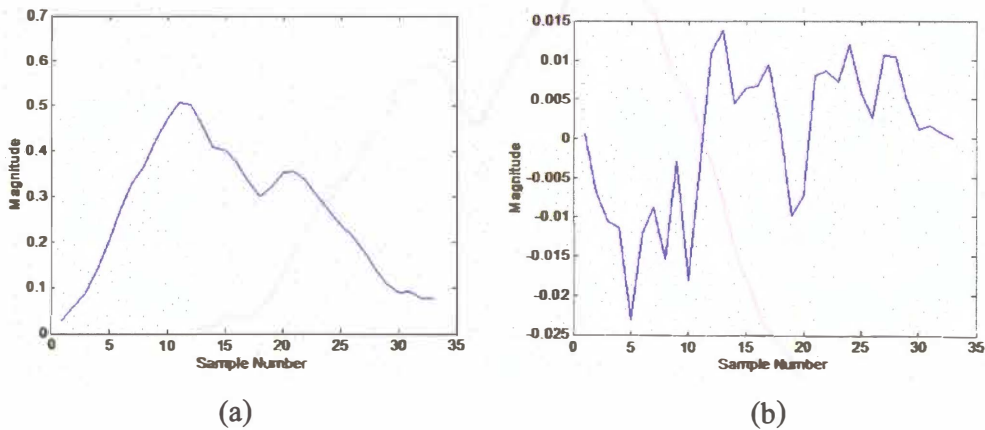


Figure 6: Level 1 DWT approximation and detail coefficients of a normal skin sample:

(a) Approximation coefficients, (b) Detail coefficients

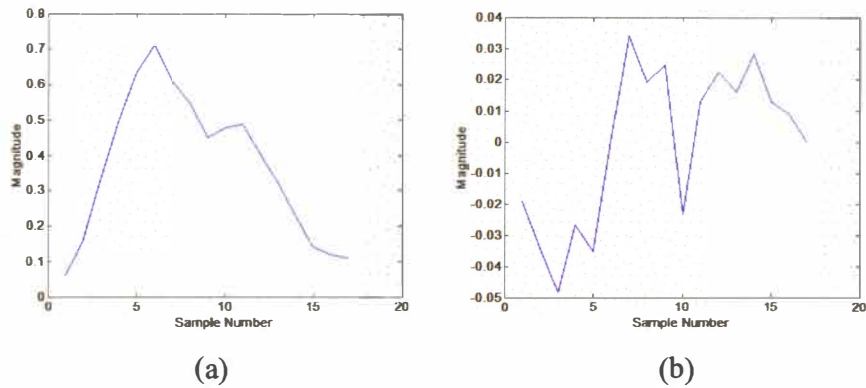


Figure 7: Level 2 DWT approximation and detail coefficients of a normal skin sample:

(a) Approximation coefficients, (b) Detail coefficients

The task of choosing the quantities that most accurately and suitably represent the data is known as feature selection. There are cases where the data in the input space cannot be separated by a linear function, but it can be linearly separated in the feature space. The goal of feature selection is to identify the smallest set of features that still contain the essential information that was included in the original data. This is known as dimensionality reduction. It can be beneficial in the reduction of computational expense and increase generalization performance. It has been seen that computational and generalization performance degrades as the number of features increases. This phenomenon has been referred to as the curse of dimensionality.

As mentioned earlier, PCA is a widely used dimensionality reduction technique in data analysis. It is a linear scheme that reduces the dimensionality of a data set consisting of a high number of interrelated variables, while retaining as much as possible of the variation

present in the data set. The first few principal components retain most of the information or variance present in all of the original data. The principal components are generally linear combinations of information from several of the input data attributes. The remaining principal components contain little structure with the last few being mostly noise. The hyperspectral data will undergo principal component analysis. The resulting principal components are evaluated and the few that contain the large majority of the variance are selected to be the input to the SVM classifier.

Wavelet transforms, using a basis or mother wavelet, decomposes a signal into its multi-resolution components. Each of the iterations transforms the signal into a set of approximation coefficients and a set of detail coefficients. This iterative process can continue until the approximation and detail coefficients consist of a single variable. If the signal contained 64 components, it could undergo a 6-level wavelet decomposition which would result in six detail coefficients and one approximation coefficient. The RMS value of each of the detail coefficients is calculated and the vector containing these values is the inputs to the SVM classifier.

4. Support Vector Machines for Classification

The goal of Support Vector Machines (SVM), when used for classification, is to produce a classifier that will generalize well. For a given set of data there can be many possible separating hyperplanes. There is but one that will separate the data in such a manner that maximizes the margin. The margin is the distance between the classification boundary and the nearest point in each class. The larger the margin, the more separation occurs between the classes and therefore a greater ability to generalize. This linear classifier can be termed the optimal separating hyperplane. It is expected that this classification boundary will generalize as well as or better than other possible boundaries.

The following explanation is based on the derivation provided by Vapnik [21], Hastie et al. [22] and Cristianni et al. [23]. Consider the problem of separating a set of training vectors, which belong to two separate classes. Where \mathbf{x}_m is a column vector of length n and $y_m = \{+1, -1\}$ is the class label for data point \mathbf{x}_m .

$$\mathbf{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}, \mathbf{x} \in \mathfrak{R}^n, y \in \{-1, 1\} \quad (24)$$

The task is to find a classifier with the decision function $f(\mathbf{x})$ such that $y = f(\mathbf{x})$, where y is the class label for the data point \mathbf{x} . The separating hyperplane satisfies the constraints that define the separation of the data samples. The data is said to be linearly separable by the separating hyperplane $\mathbf{w}^T \mathbf{x} - b = 0$ if there exist a vector $\mathbf{w} \in \mathfrak{R}^n$ and a scalar b such that

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i - b &\geq 1 && \text{if } y_i = 1 \\ \mathbf{w}^T \mathbf{x}_i - b &\leq -1 && \text{if } y_i = -1, \text{ where } i = 1, \dots, m. \end{aligned} \quad (25)$$

This corresponds to the decision function

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b) \quad (26)$$

The hyperplane $\mathbf{w}^T \mathbf{x} - b = 0$ is called the separating hyperplane and the two hyperplanes $\mathbf{w}^T \mathbf{x} - b = \pm 1$ are called the supporting hyperplanes. The vectors \mathbf{x}_i that satisfy the equation $\mathbf{w}^T \mathbf{x} - b = \pm 1$ are called the support vectors.

The problem reduces to determining the weight vector \mathbf{w} that maximizes the margin.

Since \mathbf{w} and \mathbf{x} are vectors the supporting hyperplanes can be expressed as $\langle \mathbf{w} \cdot \mathbf{x} \rangle = \pm 1$ or

$$\begin{aligned} \langle \mathbf{w} \cdot \mathbf{x}^+ \rangle + b &= +1, && \text{where } \mathbf{x}^+ \text{ is a positive point} \\ \langle \mathbf{w} \cdot \mathbf{x}^- \rangle + b &= -1, && \text{where } \mathbf{x}^- \text{ is a negative point} \end{aligned} \quad (27)$$

To compute the geometric margin \mathbf{w} must be normalized. The geometric margin γ is the functional margin of the classifier

$$\gamma = \frac{1}{2} \left(\left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|} \cdot \mathbf{x}^+ \right\rangle - \left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|} \cdot \mathbf{x}^- \right\rangle \right) \quad (28)$$

$$\gamma = \frac{1}{2\|\mathbf{w}\|} (\langle \mathbf{w} \cdot \mathbf{x}^+ \rangle - \langle \mathbf{w} \cdot \mathbf{x}^- \rangle) \quad (29)$$

$$\gamma = \frac{1}{\|\mathbf{w}\|} \quad (30)$$

To maximize the margin the norm of \mathbf{w} needs to be minimized or minimize $\langle \mathbf{w} \cdot \mathbf{w} \rangle$ with the limitation $y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1$, for $i = 1, \dots, m$. The Lagrange function is used to determine the optimization problem.

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle - \sum_{i=1}^m \alpha_i [y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) - 1] \quad (31)$$

where $\alpha_i \geq 0$ are the Lagrange multipliers. By differentiating the Lagrange function with respect to \mathbf{w} and with respect to b and setting the result to zero gives

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i = 0 \quad (32)$$

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} = \sum_{i=1}^m \alpha_i y_i = 0 \quad (33)$$

Rearranging gives

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \quad (34)$$

$$0 = \sum_{i=1}^m \alpha_i y_i \quad (35)$$

which is then substituted into the Lagrange function

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle - \sum_{i=1}^m \alpha_i [y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) - 1] \quad (36)$$

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle - \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle + \sum_{i=1}^m \alpha_i \quad (37)$$

$$L(\mathbf{w}, b, \alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle \quad (38)$$

This shows that the function can be realized as a linear combination of the training points. This classifier finds linear boundaries in the input feature space. If the input feature space cannot be separated by a linear boundary the procedure needs to be modified. The substitution of a kernel is equivalent to a mapping of the data into a high dimensional feature space. This means that dataset which may have been inseparable in the input feature space become separable in the modified feature space.

The weight vector \mathbf{w} has the expansion $\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i$, where α is the Lagrange multiplier and y is the class. Substituting into Eq. 26 indicates that the decision function depends on the dot or inner product between patterns. If a set of nonlinear transform functions, $\Phi_j(\mathbf{x}), j = 1, \dots, m$ is defined such that $\Phi(\mathbf{x})$ maps \mathbf{x} into the m -dimensional feature space. Then nonlinear decision boundaries in the input space will map to linear decision boundaries created in the feature space. Once the input space has been mapped to the feature space the linear algorithm only requires the evaluation of inner products of $\mathbf{K}(\mathbf{x}, \mathbf{x}^T) = \langle \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}^T) \rangle$ called the SVM kernel. If the feature space is of a very high dimension $\mathbf{K}(\mathbf{x}, \mathbf{x}^T)$ can be expensive to compute. Luckily, the transform $\Phi(\mathbf{x})$ does not

need to be specified. It is only required that the kernel function $K(\mathbf{x}, \mathbf{x}^T)$ be known.

Using the kernel the Lagrange function is now

$$L(\mathbf{w}, b, \alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (39)$$

and from Eq. (34) the solution becomes

$$f(\mathbf{x}) = \mathbf{w}\Phi(\mathbf{x}) + b \quad (40)$$

$$f(\mathbf{x}) = \sum_{i=1}^m \alpha_i y_i \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}_i) \rangle + b \quad (41)$$

A simple kernel can be substituted which can be computed for much less cost. Frequent choices of kernels include polynomial,

$$K(\mathbf{x}, \mathbf{x}^T) = (1 + \langle \mathbf{x}, \mathbf{x}^T \rangle)^2 \quad (42)$$

radial basis

$$K(\mathbf{x}, \mathbf{x}^T) = \exp\left(-\|\mathbf{x} - \mathbf{x}^T\|^2 / c\right) \quad (43)$$

and sigmoid functions.

$$K(\mathbf{x}, \mathbf{x}^T) = \tanh(\kappa_1 \langle \mathbf{x}, \mathbf{x}^T \rangle + \kappa_2) \quad (44)$$

The SVM algorithm used in this paper was developed by Gunn [24]. It was implemented in MATLAB®. Figure 8 displays two features of the obtained from the samples spectral data. The tumorous skin is in one class and the normal skin is the other class. When an

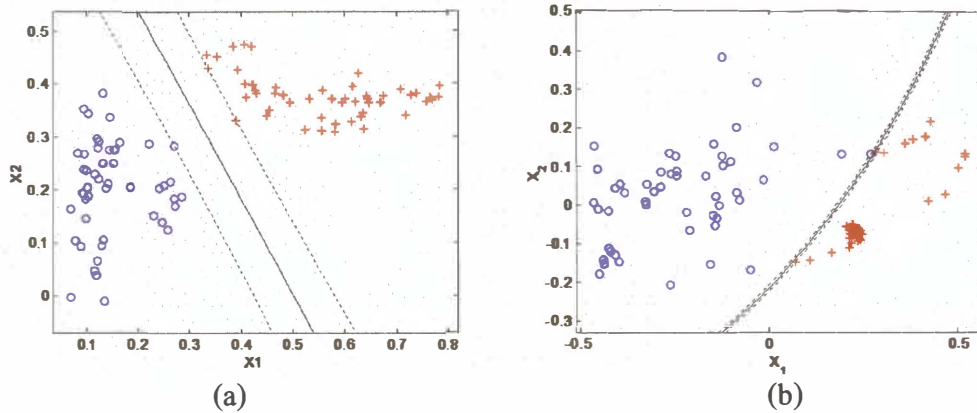


Figure 8: Plot of decision surfaces created by support vector machine using the first two principal components as inputs:

(a) Linearly separable decision surface, (b) Non-linearly separable decision surface

SVM is used to determine the decision boundary it determines not only the decision boundary, solid line, but also the margins, dotted line, which determines the support vectors. The larger the margin, the greater the ability the classifier has to generalize. Figure 8(a) displays a set of features that are linearly separable. The decision surface is represented by the solid line. The margins (dotted lines) are created by the three support vectors. Figure 8(b) displays a second set of features that are separated by a non-linear decision boundary. This example also has two support vectors. In both cases the decision surface was created using a radial basis function as the SVM kernel.

5. Combination of PCA and DWT Feature Extraction

Schemes

Spectral signature reveals the characteristics of the different types of tissues and materials in the scene. Figure 9 shows the relative fluorescence intensity of a pixel as a function of spectral bands for the three categories of interest in this paper: normal tissue, tumorous tissue and background. Pixels obtained from the training hyperspectral from the three nominal categories were used to create the figure. Normal tissues have a large peak response at approximately band 20 and a lesser peak at approximately band 40. Tumors also have a lower fluorescence on average, but show strong response between bands 40 to 45 relative to the peak at approximately band 22. Background pixels show low fluorescence intensity and an almost flat response over the entire spectral range since the carrying tray being covered with a non-fluorescent flat black paint.

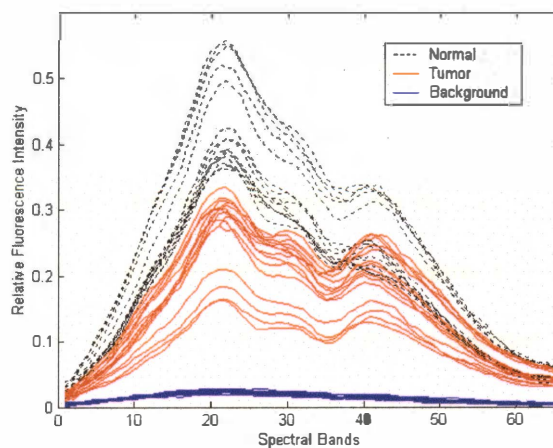


Figure 9: Spectral signature of normal tissue, tumor, and background

This thesis continues the work that was performed in the paper by Fletcher and Kong [25]. In this paper a hyperspectral fluorescence image consisting of 65 bands whose wavelength ranged from 425 nm to 711 nm, of a tumorous chicken carcass was used to test a classifier utilizing Principal Component Analysis (PCA) as a feature reduction technique and a Support Vector Machine (SVM) as the classifier. From the hyperspectral fluorescence image a data set was developed which contained 2000 patterns of each class normal, tumor and background. This data set was divided into training and testing data sets. These two set were further sub-divided into 20 separate sets of 150 patterns consisting of 50 patterns of each class. A training set was chosen and the principal components of that set were calculated. The first two principal components were selected as the input to the SVM classifier. Once the classifier was trained the 20 testing data sets were used to evaluate the performance. The average classification rate over the 20 testing sets was approximately 96%. As a check, a linear regression model was evaluated which produced a 77% classification accuracy.

The continuation of this work used the same training and testing sets of data. 4000 patterns were collected from the hyperspectral image of the training poultry sample. This sample was determined as the training image since it was received from the ISL much in advance of the other eleven samples. It also contained more tumors than images received at a later date which influenced its retention as the training sample. 2000 pixels depicting normal skin and 2000 pixels depicting tumorous skin were selected. The determination of whether a pixel belonged to a tumor or normal skin was based on the results of the

inspection performed by the FSIS Inspector. The set was divided into training and test data sets by selecting every other point. The every other point method insured that each tumor and normal skin area would be represented in both data sets. The background spectral signature is much different from that of the normal and tumor signatures and is very easy to remove from the classification process. This will simplify the calculations and reduce them to a two class problem. Each pattern contained in the training and testing data sets was assigned a target value. The target values were set to 1 for tumor and -1 for normal skin. As each of the two sets contained 2000 patterns each they were further subdivided to 20 separate sets of 100 patterns consisting of 50 patterns of each class. This allowed for a simpler training and testing routine.

Figure 10 provides a graphical representation of the proposed classifier. For this experiment two classifiers were trained. In the first classifier, principal component analysis is used to put the information available into a few features. These features are evaluated to determine the subset that will provide adequate representation of the original signal. This reduced feature set will then be sent to a support vector machine for classification. The second classifier will use discrete wavelet transforms to concentrate the information in a smaller data set. This smaller data set is classified by a SVM. To provide an additional increase in classifier accuracy a simple algorithm is used to provide the direction to use the classifier best suited to classify the signal. The technique chosen was to evaluate the average intensity of the hyperspectral fluorescence.

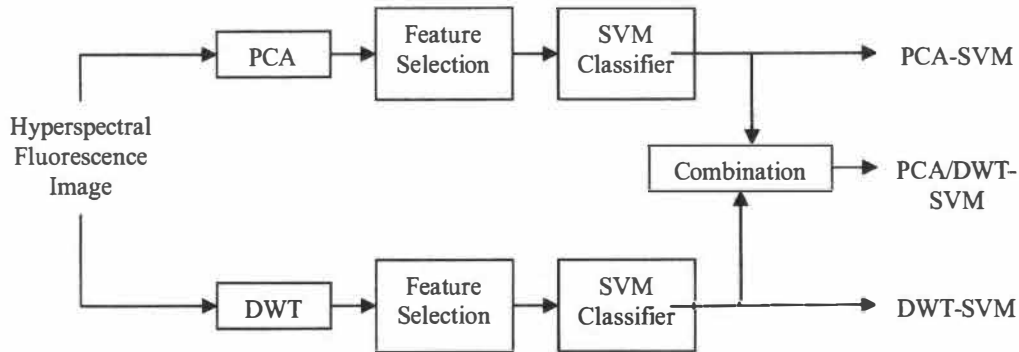


Figure 10: Block diagram of the proposed multi-channel classifier

In the first various combinations of the principal components of the training data set were tested to find a combination that would provide the best results while using a small number of components. The training set was subjected to PCA which resulted in the calculation of the 65 principal components. In Figure 11 the information content of the first 10 principal components are shown. Since the first two principal components contain over 99% of the information it was determined that they would be used as the input to the SVM classifier to provide good results for the fewest number of inputs. The support vector machine for this problem was set up with two neurons in the input layer but only one linear neuron in the output layer. The SVM implementation provided by S. R. Gunn [24] is used to perform the SVM training and classification. The plot of the decision surface for the training set is shown Figure 12. Since the SVM has only one output the decision surface divides the tumor skin, 'o', from the normal skin, shown as '+'.

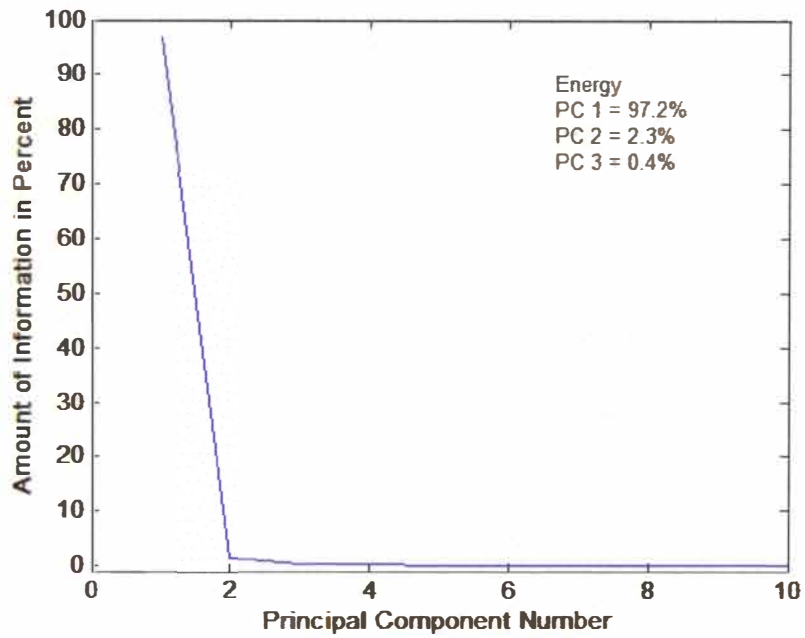


Figure 11: Information content (in percent) of the first 10 principal components

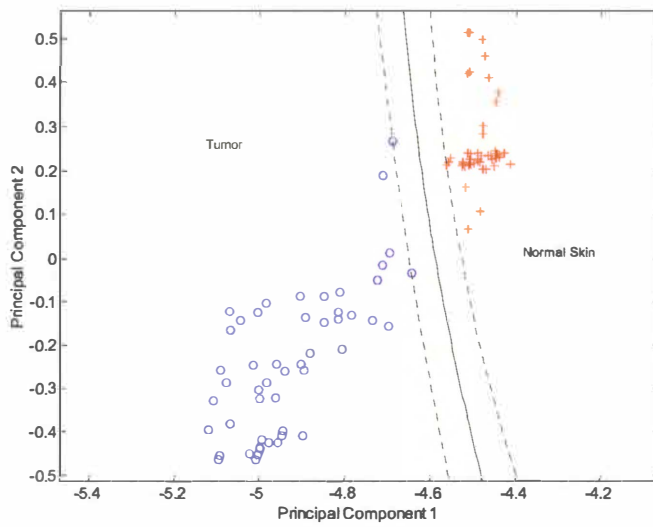


Figure 12: Decision boundaries of the PCA-SVM classifier with a radial basis kernel

This indicates that the SVM classifier was able to differentiate skin tumor of the poultry carcass from the normal tissue and the background. The first and the second principal components provide a reasonably small classification error for the training data set. Additional principal components decreased the error, but the gain was slight. In this paper, first two principal components were used as a feature vector ($m = 2$). The SVM classifier was trained using one of the training data sets. The trained SVM network was then tested using all 20 of the test data sets. The classification rate for each test data set was obtained to check the accuracy of the SVM classifier.

The second classifier used the same training set. But in this instance the training set was truncated to 64 bands by deleting the 65th band. This was done to allow a dyadic discrete implementation. The training set was then subjected to a 6-level DWT using the Haar mother wavelet. The detail coefficients were compressed by taking the RMS values of the coefficients. This reduced the features from the 64 bands to 6 RMS values. The support vector machine for this problem was set up with six neurons in the input layer but only one neuron in the output layer. The plot of the decision surface for the first two wavelet coefficients of the training set is shown Figure 13. Since the SVM has only one output the decision surface divides the tumor skin, 'o', from the normal skin, shown as '+'. This indicates that the SVM classifier was able to differentiate skin tumor of the poultry carcass from the normal tissue. Note that there are six coefficients used in the training of the SVM classifier.

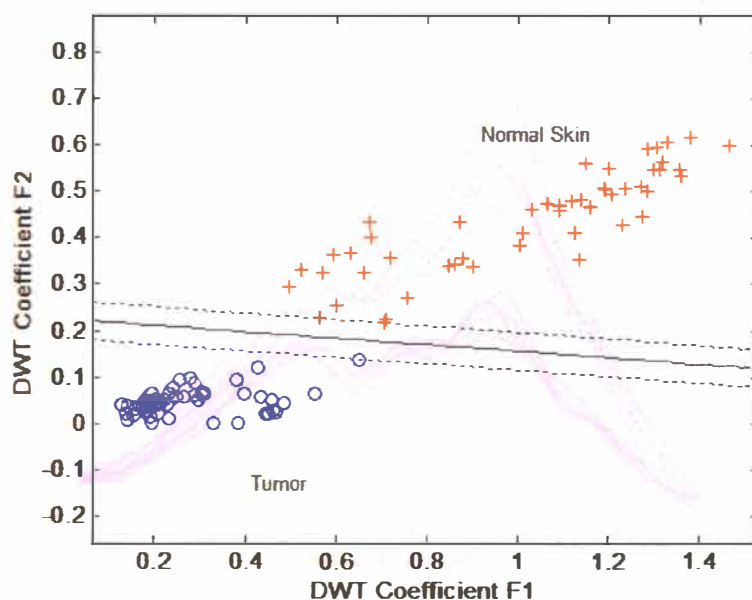


Figure 13: Decision boundaries of the DWT-SVM classifier with a radial basis kernel

In the PCA/SWT-SVM classifier a selector algorithm was used to determine which of the two individual classifiers, PCA-SVM or DWT-SVM, would be used to provide an improved tumor detection rate. From experimental results it could be seen that the PCA-SVM classifier performed with greater accuracy in the brighter areas of the poultry carcass. The DWT-PCA classifier tended to perform better in the darker areas of the carcass. A simple method of using the average intensity of the pixel signal was used to determine which classifier would be used. Figure 14 displays the value of the average pixel intensity chosen that represents the boundary between the light and darks areas. This boundary was determined by evaluating the hyperspectral images and the preliminary classification results. The limit was set to be 85% of the average pixel

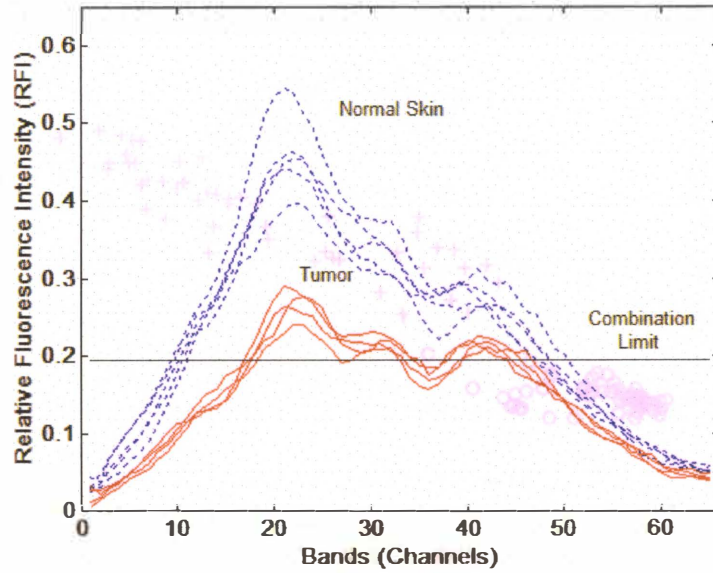


Figure 14: Combination feature to determine classifier selection

intensity of a set of 1000 normal skin samples taken from the center of the carcass body. This boundary limit was determined to be 0.195. The algorithm used to provide the selection in the combination classifier is shown in Eqs. 45 and 46. If the average pixel intensity of the sample to be classified, $f(n)$, is less than or equal to the boundary limit then the DWT-SVM classifier is to be used. Otherwise the PCA-SVM classifier is used.

$$\text{if } \text{avg}(f(n)) \leq \theta, \text{ classifier} = \text{DWT-SVM} \quad \text{Eq. (45)}$$

$$\text{if } \text{avg}(f(n)) > \theta, \text{ classifier} = \text{PCA-SVM} \quad \text{Eq. (46)}$$

6. Experimental Results

Twelve chicken carcasses were collected from a poultry processing plant owned by Allen Family Foods Inc. of Cordova, MD in March and May of 2002. A Food Safety and Inspection Service (FSIS) veterinarian at the plant identifies the condition of the poultry carcasses. They were put into plastic bags to minimize hydration and then placed in a cooler with ice. The images were captured within two hours of their transportation to the Imaging and Sensing Lab.

6.1 *Imaging System*

The Instrumentation and Sensing Laboratory at the Beltsville Agricultural Research Center in Beltsville, MD, developed a hyperspectral imaging system capable of reflectance and fluorescence imaging [2] for use in food safety and quality research. The key components of this system include a CCD camera, a spectrograph, a sample transport mechanism, and appropriate lighting sources. The spectrograph is based on prism-grating-prism optics. The system is equipped with two independent illumination sources for fluorescence and reflectance measurements, respectively. Two fluorescent lamp assemblies provided a near uniform UV-A (365 [nm]) excitation. Short-pass placed in front of the lamp housing were used to prevent transmittance of radiations greater than 400 [nm] and thus eliminating the potential spectral contamination by pseudo-fluorescence. Line-by-line image scans were achieved by transporting sample materials via a precision positioning table.

Hyperspectral images obtained from the Instrumentation and Sensing Laboratory consists of, on average, 400×460 pixels with 65 spectral bands ($N = 460$, $M = 400$, $L = 65$). The spectral band has a discrete value from the wavelength λ_1 (425.4 [nm]) to λ_{65} (710.7 [nm]). The sample poultry carcasses were placed on a tray painted with a non-fluorescent flat black paint to minimize background scattering in a darkened room. The speed of the conveyer belt was adjusted based on the predetermined CCD exposure time and data transfer rate. Figure 15 provides a diagram of the hyperspectral imaging system.

The typical hyperspectral image of a poultry sample consists of a spatial dimension of approximately 400 by 460 pixels where each pixel has 1mm x 1mm of resolution. To complete an entire hyperspectral image requires approximately 200 seconds. In the capture of the hyperspectral images, 65 spectral bands with wavelengths ranging from 425.4nm to 710.7 nm were used. Table 1 show the 65 spectral bands used along with the associated wavelength.

To illustrate the information that the hyperspectral fluorescence image can contain, Figure 16 displays the images of the training image at bands 5, 10, 15 and so on through band 60. One can see that the information or features revealed changes as the wavelengths change.

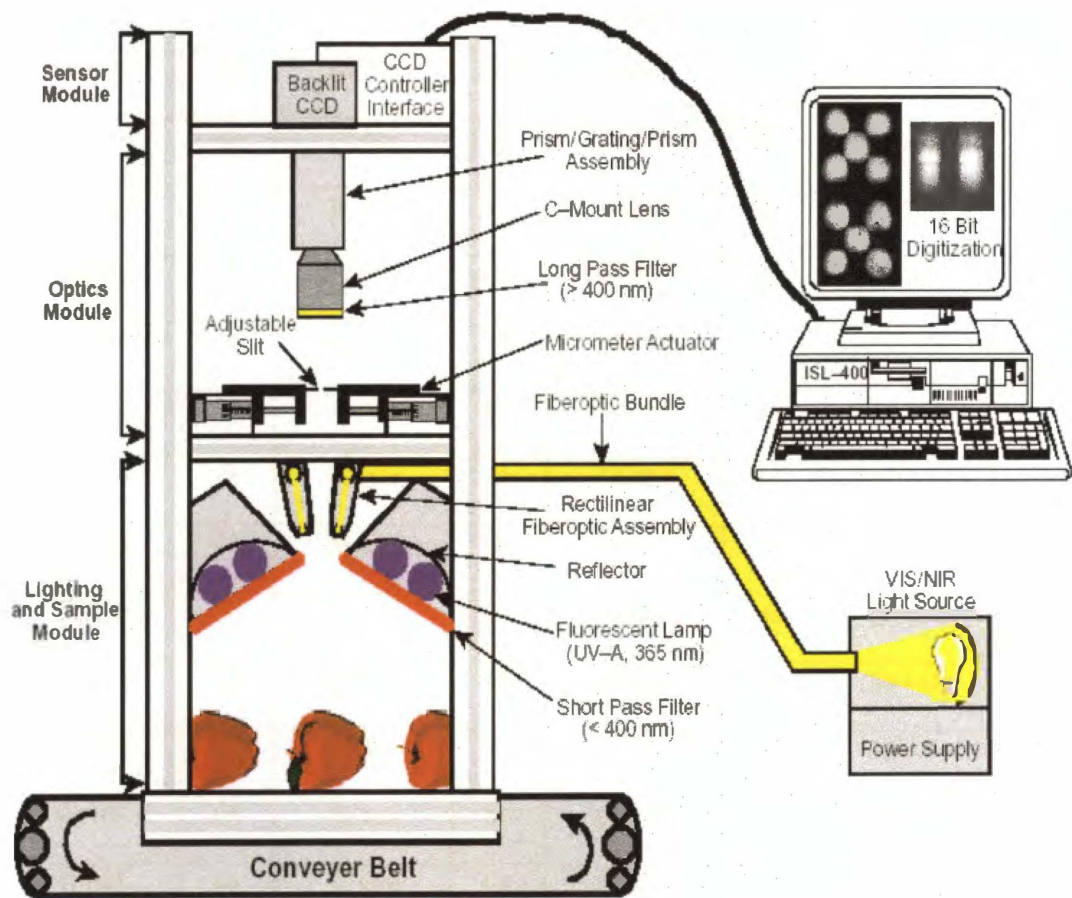


Figure 15: Diagram of the hyperspectral fluorescence imaging system [2]

Table 1: Wavelength values of the spectral band used in the image acquisitions

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

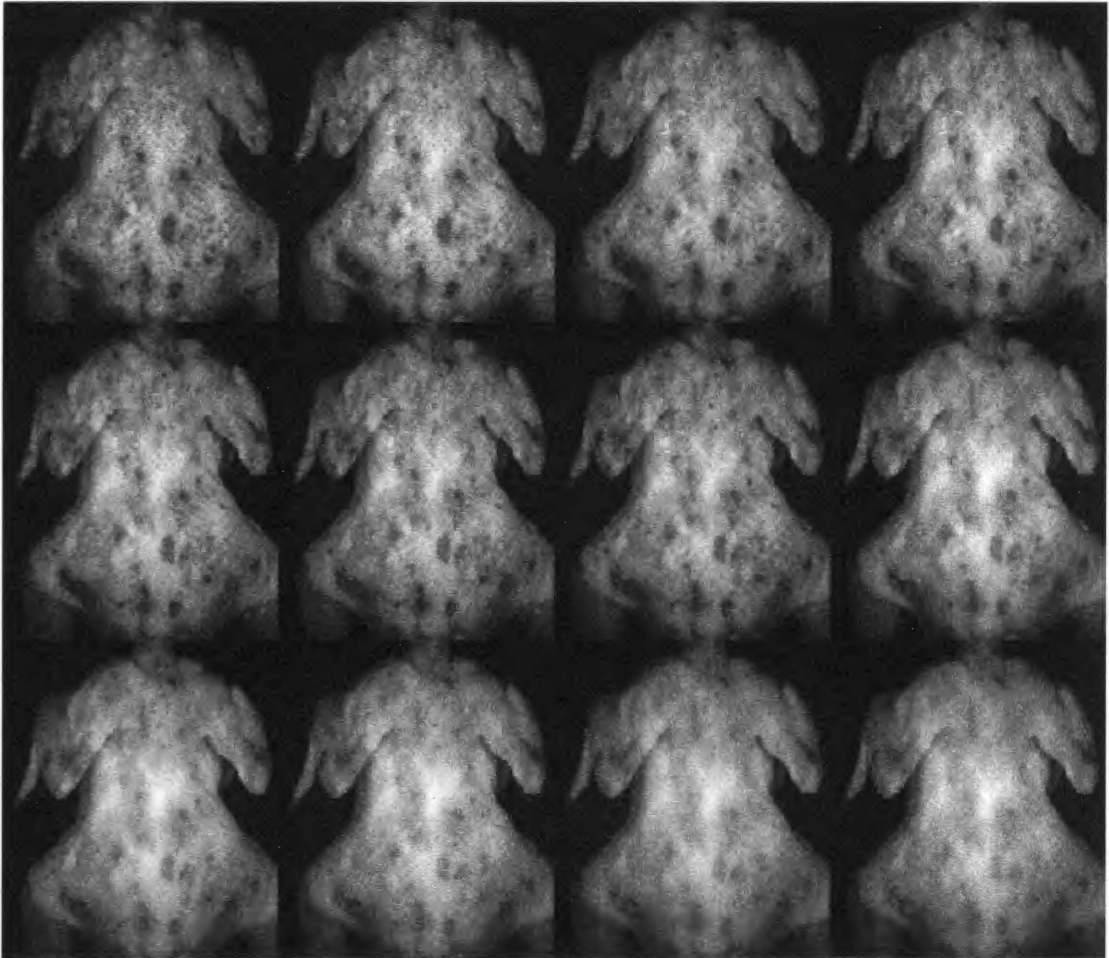


Figure 16: Mosaic of the hyperspectral image of a poultry carcass sample displaying bands 5, 10 thru 60

6.2 Performance Comparisons

6.2.1 PCA-SVM

The trained PCA-SVM classifier was used to evaluate the test data set. Figure 17 shows the decision boundary created by the SVM classifier when trained by the first two principal components of a training set were used as inputs. The trained SVM was then used to classify the testing data set. The PCA-SVM gave 100% classification results.

In Figure 18 the image from which the training information was gathered is shown on the left. On the right is the image with the tumors indicated in white. The PCA-SVM classifier was able to detect most of the tumors. The classifier had problems with the areas that were not well illuminated. Since this image was the source of the training data good results were expected.

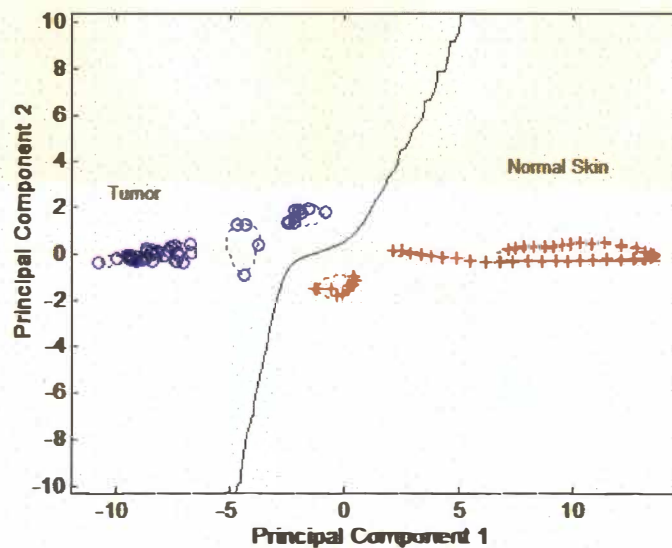
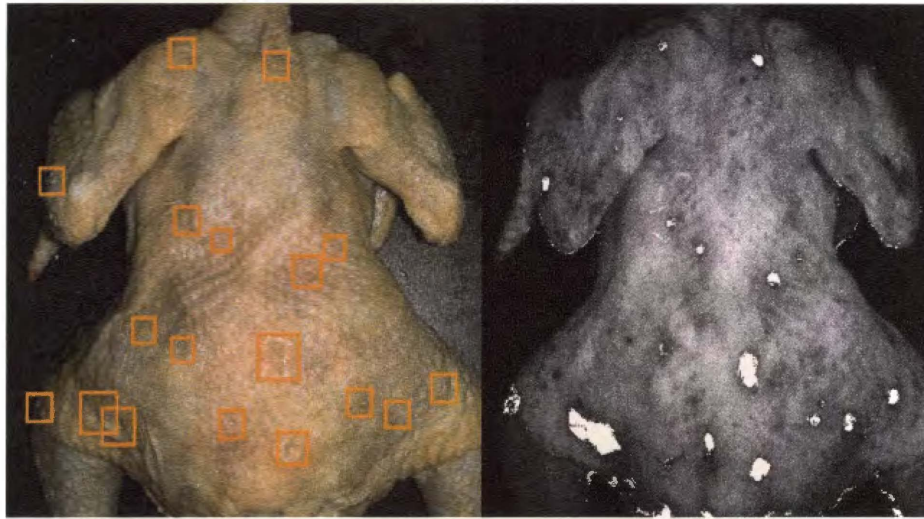


Figure 17: Decision boundary of the PCA-SVM classifier



(a)

(b)

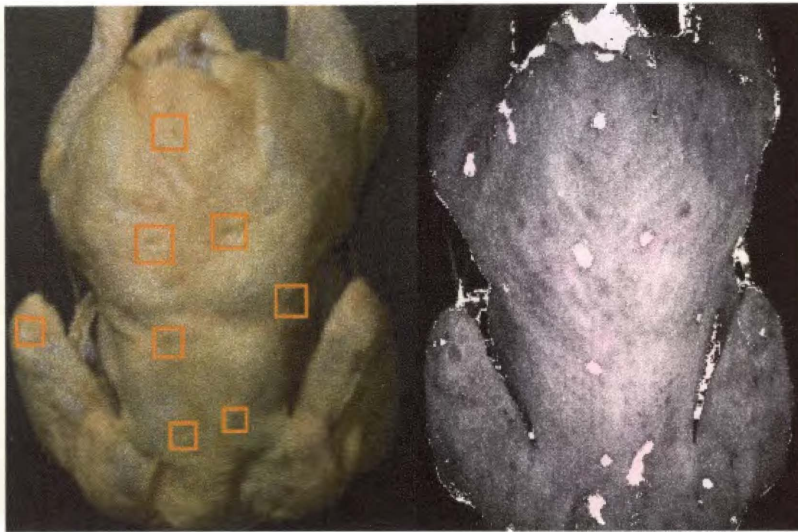
Figure 18: Tumors detected by the PCA-SVM classifier on the training image:

(a) Training image, (b) Tumor detection image

In Figure 19 one of the testing images is shown with a picture of the detected tumors. The PCA-SVM classifier was able to detect most of the tumors especially in the well illuminated areas but had problems in the dimmer areas and the areas that did not have skin attached.

6.2.2 Wavelet

The trained DWT-SVM classifier was used to evaluate the test data set. Figure 20 shows the decision boundary created by the DWT-SVM classifier when trained by the six DWT features used as inputs. The decision boundary created by the first two features is shown in the figure. The trained SVM was then used to classify the testing data set. The DWT-SVM gave 99.8% classification results when used on the 20 testing sets.



(a)

(b)

Figure 19: Tumors detected by the PCA-SVM classifier on one of the testing images:

(a) Testing image, (b) Tumor detection image

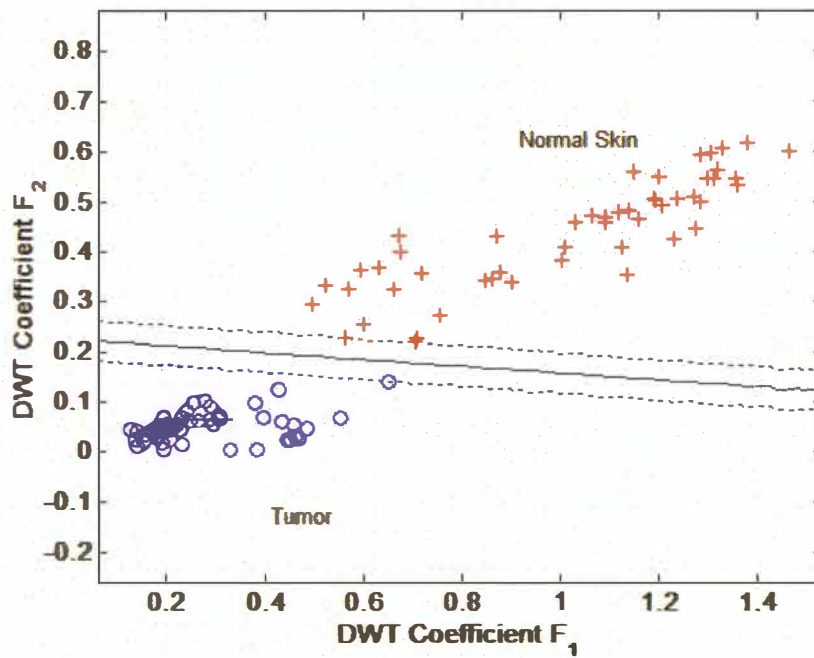


Figure 20: Decision boundary of the DWT using the first two features

In Figure 21(a) the image from which the training information was gathered is shown. In Figure 21(b) the image is displayed with the tumors indicated in white. Again the DWT-SVM classifier was able to detect most of the tumors. The classifier did miss some of the smaller tumors in the main body of the carcass. It seemed to perform well in the darker areas. Since this image was the source of the training data good results were expected.

In Figure 22(a) a test image is shown with the tumors marked. The classified image, Figure 22(b), shows that the tumors in the well illuminated areas were difficult to detect. This classifier performed well in the poorly illuminated areas and in areas that had the skin removed.

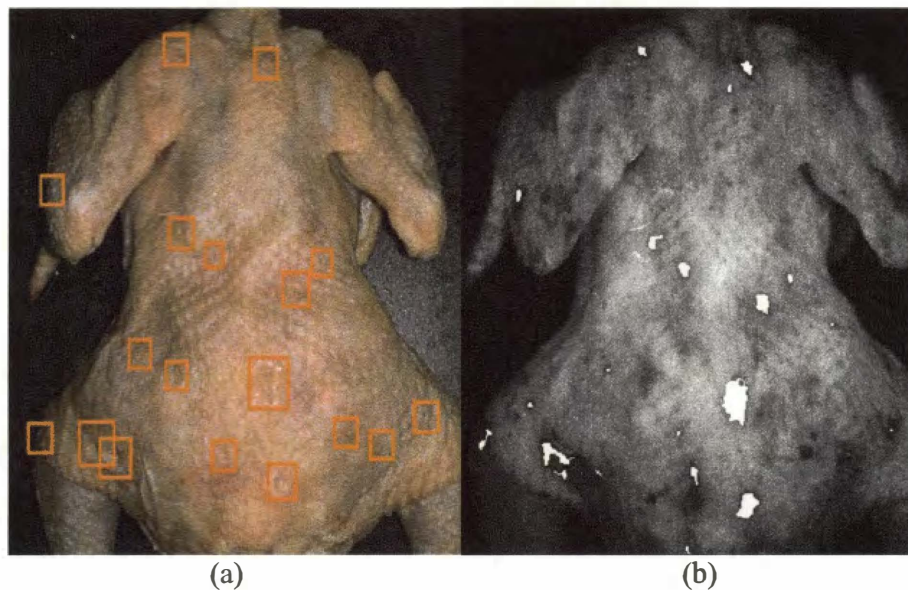


Figure 21: Tumors detected by the DWT-SVM classifier on the training image:

(a) Training image, (b) Tumor detection image

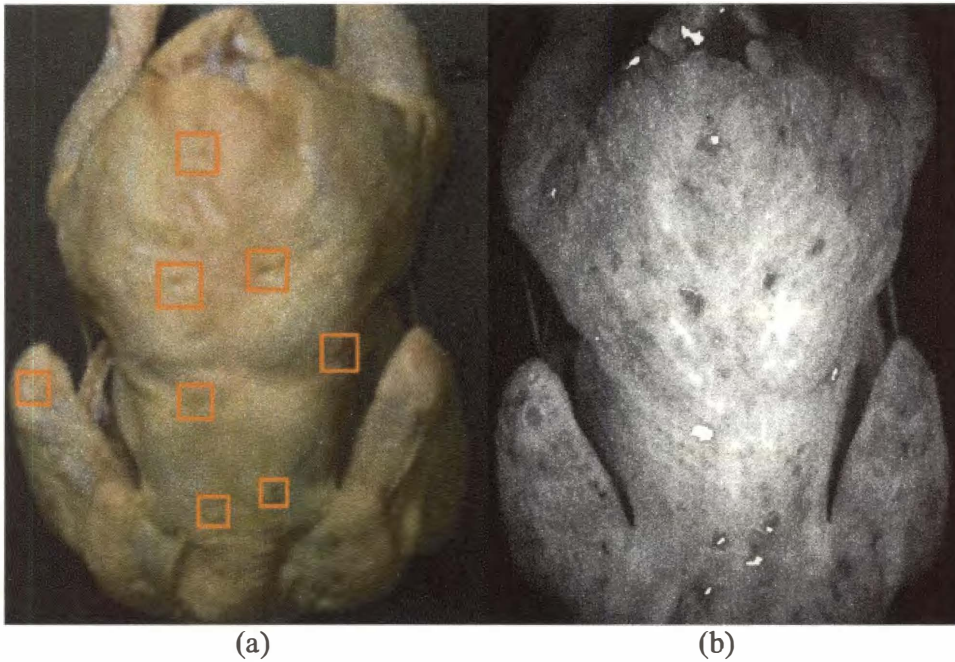


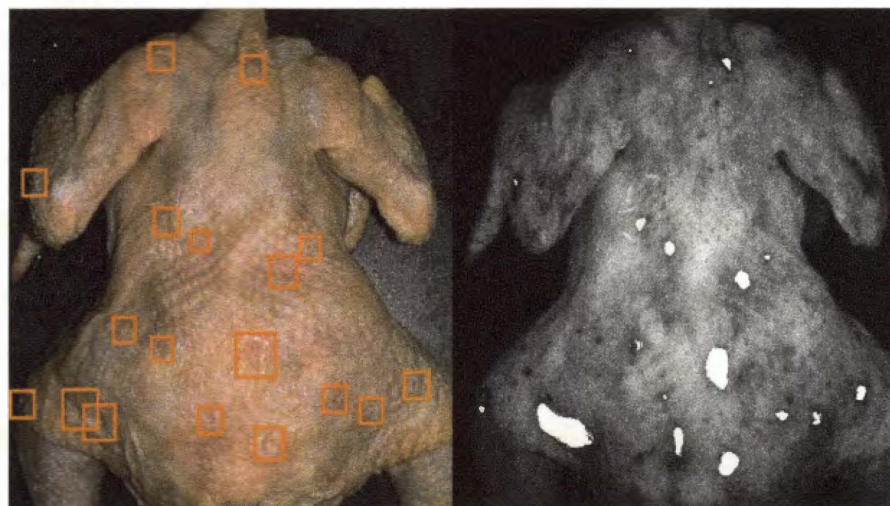
Figure 22: Tumors detected by the DWT-SVM classifier on the testing image:

(a) Testing image, (b) Tumor detection image

6.2.3 Combined Classifier

Even though both of the two approaches to classifier training showed very good results during training when used on images that were not part of the training data the results were mixed. The first classifier that used only principal components 1 and 2 performed well on the brighter parts of the carcass but poorly on the areas that were dimly illuminated. The second classifier that used the DWT reduced data set performed well on the less illuminated areas but tended to miss tumors on the more brightly lit areas of the carcass. By using a selection process that measured the average brightness of the pixel the best classifier was selected which would give the best classification result for that

particular pixel. Figures 23 and 24 show the same two images this time classified with the PCA/DWT-SVM classifier system using a selector based on average pixel brightness. Figure 23(a) displays the training image with tumors marked. Figure 23(b) shows the same image with the detected tumors indicated in white. Figure 24(a) shows one of the testing images with the tumors marked. In Figure 24(b) the tumors detected by the PCA/DWT-SVM classifier are shown in white. In both of the images, the PCA/DWT-SVM classifier was able to detect a very high percentage of the indicated tumors with a very low number of false indications.



(a)

(b)

Figure 23: Tumors detected by the PCA/DWT-SVM classifier on the training image:

(a) Training image, (b) Tumor detection image

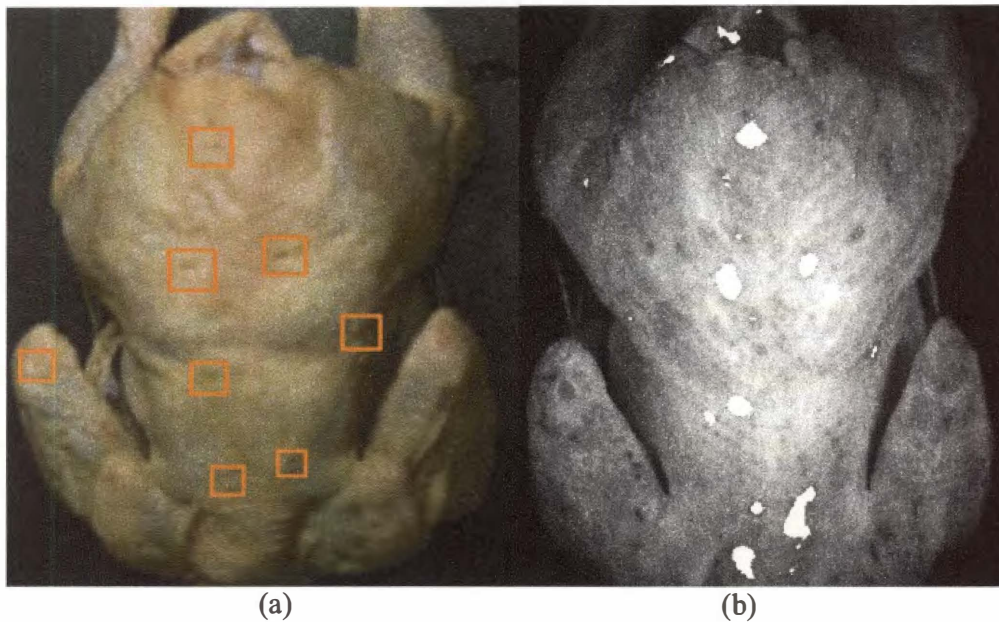


Figure 24: Tumors detected by the PCA/DWT-SVM classifier on the testing image:

(a) Testing image, (b) Tumor detection image

The three classification schemes were then used on the testing images. Table 2 displays the results of the PCA-SVM classifier when used to detect tumors on the 11 testing images. The results indicate that the PCA-SVM classifier was good at detecting the tumor since there were 38 tumors detected. However, the results also show that there were 17 false positives and 5 missed tumors. Table 2 also displays the results of the DWT-SVM classifier. The DWT-SVM classifier had different results than the PCA-SVM classifier. There were 28 tumors detected, which is lower than the PCA-SVM classifier, with 18 false positives and 12 missed tumors. In Table 3 the results of the PCA/DWT-SVM classifier were displayed. In this case 37 tumors were detected with only one tumor missed. There were only 6 false positives.

Table 2: Classification results of the PCA-SVM classifier and DWT-SVM classifier

Image No.	# Tumors	PCA-SVM classifier			DWT-SVM classifier		
		# indicated	False positive	# missed	# indicated	False positive	# missed
1	8	8	2	1	5	3	3
2	4	4	2	0	3	3	1
3	0	0	2	0	0	3	0
4	5	5	2	1	4	1	2
5	2	2	2	0	1	2	1
6	2	2	1	0	2	2	0
7	2	2	0	1	1	0	1
8	0	0	0	0	0	0	0
9	5	5	1	1	4	2	2
10	7	7	2	1	5	1	2
11	3	3	2	0	3	1	0
Total	38	38 (100%)	17	5 (13%)	28 (74%)	18	12 (31%)

Table 3: Classification results of the combined PCA/DWT-SVM classifier.

Image No.	# Tumors	PCA/DWT-SVM classifier		
		# indicated	False positive	# missed
1	8	7	3	1
2	4	4	1	0
3	0	0	0	0
4	5	5	1	0
5	2	2	0	0
6	2	2	0	0
7	2	2	0	0
8	0	0	0	0
9	5	5	0	0
10	7	7	1	0
11	3	3	0	0
Total	38	37 (97.4%)	6	1 (2.6%)

7. Conclusion

7.1 Summary

This paper presents detection of skin tumor on poultry carcasses using hyperspectral fluorescence images. Skin tumor is not as visually obvious as other pathological diseases since its signature appears as shape distortion rather than discoloration. This makes it difficult to detect patterns from the poultry images based on reflections. The hyperspectral imaging system captures the fluorescence data from the poultry images. Hyperspectral images have proved useful in providing information for detecting poultry skin tumors. The features obtained have shown to contain a higher level of contrast and thus provide a greater possibility of detecting tumors. Since hyperspectral images contain a very large amount of information it is necessary to provide some means of data compression in order to reduce the amount of calculations. The PCA has the ability to transform the spectral information into a small number of principal components without losing significant information. By using the discrete wavelet transform (DWT) and performing an additional compression by placing the detail coefficients into a single number it has been shown that the large amount of information can be placed into a smaller feature space. This smaller dimensional feature space reduces computation for on-line implementation. In the PCA-SVM classifier, since the first two principal components contained over 99% of the information, it was determined that only these two principal components were needed as input to the SVM classifier. The DWT-SVM

classifier used the six RMS value of the detail coefficients as inputs to the SVM classifier. To further improve the classification rate, a selector was used to choose the classifier, PCA-SVM or DWT-SVM, which would most accurately classify the pixel. During the examination of the individual classifier results it was determined that the PCA-SVM classifier has the best results on the brighter areas of the carcass. The DWT-SVM classifier has better results on the darker areas and tended to misclassify tumors on the brighter areas. Because of this need a selection process was implemented based on the average intensity of a pixel. The pixels that exhibited a higher intensity were sent to the PCA-SVM classifier. The pixels with lower average pixel intensity used the DWT-SVM classifier. The combination of the two provided an improvement in the classification rate. Figure 25 shows the classification results for the three classification schemes. The PCA-SVM classifier has a tumor detection rate of 86.6% with 17 false positives and 5 missed tumors. The DWT-SVM classifier had a tumor detection rate of 42.1% with 18 false positives and 12 missed tumors. The PCA/DWT-SVM classifier has a tumor detection rate of 94.7% with only 6 false positives and 1 missed tumor. The addition of the selection scheme increased the tumor detection rate and decreased the occurrence of false positives.

7.2 Recommendation for Future Work

Some future work could be performed in order to further the detection accuracy using hyperspectral fluorescence imaging.

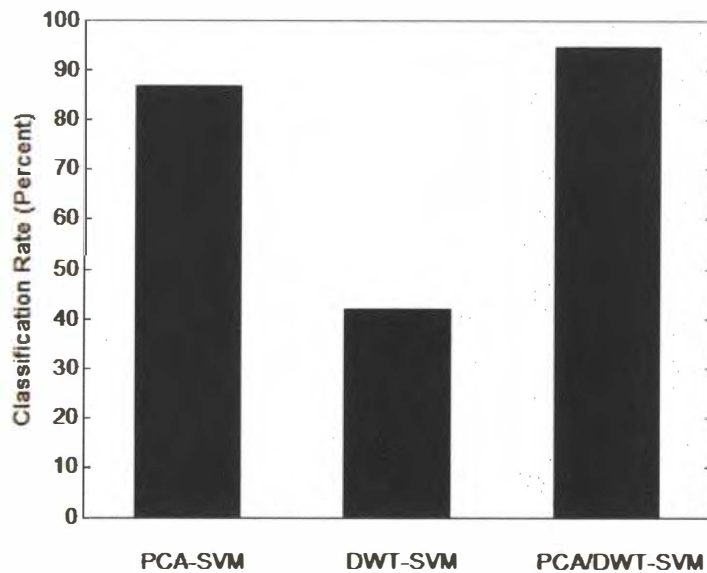


Figure 25: Tumor detection rate for the PCA-SVM, DWT-SVM and PCA/DWT-SVM classifiers

- a) Incorporate the use of Independent Component Analysis (ICA) as a method of dimensionality reduction. ICA is a statistical and computational technique to determine the latent variables of a data set. The data set is assumed to be a linear or non-linear combination of these latent variables or independent components. They are assumed to be non-gaussian and mutually independent. ICA is a method to find these components.
- b) Use pixel unmixing to determine the abundance of each of the end-member classes contained in each pixel. Since the hyperspectral fluorescence images are scanned one line at a time it is possible that there is mixing of signals in the pixels contained in a row.



Bibliography

Bibliography

- [1] B. W. Calnek, H. J. Barnes, C. W. Beard, W. M. Reid, and H. W. Yoder, *Diseases of Poultry*, Ch. 16, Iowa State University Press, Ames, Iowa, 1991.
- [2] M. S. Kim, Y. R. Chen, and P. M. Mehl, "Hyperspectral Reflectance and Fluorescence Imaging System for Food Quality and Safety," *Transactions of the American Society of Agricultural Engineers*, Vol. 44, No. 3, pp.721-729, 2001.
- [3] B. Albers, J. DiBenedetto, S. Lutz and C. Purdy, "More Efficient Environmental Monitoring with Laser-induced fluorescence Imaging," *Biophotonics Int.*, pp. 42-54, Nov. 1995
- [4] L. Brancalion, A. J. Durkin, J. H. Tu, G. Menaker, J. D. Fallon and N. Kollias, "Invivo Fluorescence Spectroscopy of Nonmelanoma Skin Cancer," *Photochemistry and Photobiology*, Vol. 73, No. 2, pp.178-183, 2001.
- [5] S. K. Chang, Y. N. Mirabal, E. N. Atkinson, A. Malpica, M. Follen, R. Richards-Kortum, "Combination of Fluorescence and Reflectance Spectroscopy for In Vivo Detection of Cervical Pre-Cancers," *Proceedings of the Second Joint EMBS/BMES Conference*, October 23-26, pp.2265-2266, 2002.
- [6] N. Ramanujam, G. Palmer, T. Breslin, K. Gichrist, "Diagnosis of Breast Cancer Using Auto Fluorescence and Diffuse Reflectance Spectroscopy," *Proceedings of the Second Joint EMBS/BMES Conference*, October 23-26, pp.2279-2280, 2002.

- [7] D. A. Landgrebe, "Hyperspectral Image Data Analysis as a High Dimensional Signal Processing Problem," *IEEE Signal Processing Magazine*, Vol. 19, No. 1, pp.17-28 2002.
- [8] K. Chao, Y. R. Chen, W. R. Hruschka, and F. B. Gwozd, "On-line inspection of poultry carcasses by a dual-camera system," *Journal of Food Engineering*, Vol. 51, pp.185-192, 2002.
- [9] S. G. Kong, "Inspection of Poultry Tumor using Hyperspectral Fluorescence Imaging," *Proc. Int. Conf. Quality Control by Artificial Vision (QCAV'2003)*, Gatlinburg, TN, pp.455-463, May 2003.
- [10] B. Park, Y. R. Chen, and M. Nguyen, "Multi-spectral Image Analysis using Neural Network Algorithm for Inspection of Poultry Carcasses," *J. of Agriculture Engineering Research*, Vol. 69, pp.351-363, 1998.
- [11] G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist and N. Ramanujam, "Comparison of Multiexcitation Fluorescence and Diffuse Reflectance Spectroscopy for the Diagnosis of Breast Cancer," *IEEE Transactions on Biomedical Engineering*, Vol. 50, No. 11, pp.1233-1242, Nov. 2003
- [12] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *Phil. Magazine*, Vol. 2, No.6, pp.559-572, 1901.
- [13] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *J. Educ. Psychol.*, Vol. 24, pp.417-441,498-520, 1933.
- [14] I.T. Jolliffe, *Principal Component Analysis 2nd Edition*, Springer-Verlag, New York, NY, 2002

- [15] MatLab, The Mathworks, 3 Apple Hill Drive, Natick, MA 01760.
- [16] S. G. Kong, Y. R. Chen, I. Kim and M. S. Kim, "Analysis of Hyperspectral Fluorescence Images for Poultry Tumor Inspection," *in print, Applied Optics*, 2003.
- [17] L. M. Bruce, C. H. Kroger and J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE Transactions of Geoscience and Remote Sensing*, Vol. 40, No. 10, pp. 2331-2338, 2002.
- [18] S. Kaewpijit, J. Le Moigne and T. El-Ghazawi, "Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis," *IEEE Transactions of Geoscience and Remote Sensing*, Vol. 41, No. 4, pp. 863-871, 2003.
- [19] B.B. Hubbard, *The World According to Wavelets 2nd Edition*, A. K. Peters, Ltd., 1998.
- [20] Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA 2002
- [21] V.N. Vapnik, *The Nature of Statistical Learning Theory 2nd Edition*, Springer-Verlag, New York, NY, 2000
- [22] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, New York, NY, 2001
- [23] N. Cristianini, J. Showe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, Cambridge, UK, 2000
- [24] S. R. Gunn, "Support Vector Machines for Classification and Regression," *Technical Report, ISIS, Department of Electronics and Computer Science*, University of Southampton, 1998.

- [25] J. T. Fletcher and S. G. Kong, "Principal Component Analysis for Poultry Tumor Inspection using Hyperspectral Fluorescence Imaging," *Proc. Of the 2003 International Joint Conference on Neural Networks*, Portland, OR, pp.149-153, July 2003.

VITA

John Thomas Fletcher was born in Clarksville, Tennessee. He graduated from Clarksville High School in 1971. After spending six years in the US Navy in the Navy Nuclear Power Program he entered Tennessee Technological University and in 1982 received a Bachelor of Science degree in Electrical Engineering.

Upon graduation from Tennessee Technological University, he began work as an electrical engineer with the Equipment Group of Texas Instruments in Dallas, TX. In 1985 he transferred to the Industrial Automation Group located in Johnson City, TN which became Siemens Industrial Automation. In 2001, he entered the University of Tennessee Graduate School to pursue a graduate degree in Electrical Engineering.