Band Selection of Hyperspectral Images for Automatic Detection of Poultry Skin Tumors

Zheng Du, Myong K. Jeong, *Member, IEEE*, and Seong G. Kong, *Senior Member, IEEE*

***Abstract*—This paper presents a spectral band selection method for feature dimensionality reduction in hyperspectral image analysis for detecting skin tumors on poultry carcasses. A hyperspectral image contains spatial information measured as a sequence of individual wavelength across broad spectral bands. Despite the useful information for skin tumor detection, real-time processing of hyperspectral images is often a challenging task due to the large amount of data. Band selection finds a subset of significant spectral bands in terms of information content for dimensionality reduction. This paper presents a band selection method of hyperspectral images based on the recursive divergence for the automatic detection of poultry carcasses. For this, we derive a set of recursive equations for the fast calculation of divergence with an additional band to overcome the computational restrictions in real-time processing. A support vector machine is used as a classifier for tumor detection. From our experiments, the proposed band selection method shows high detection accuracy with low false positive rates compared to the canonical analysis at a small number of spectral bands. Also, compared with the enumeration approach of 93.75% detection rate, our proposed recursive divergence approach gives 90.6% detection rate, which is within the industry-accepted accuracy of 90–95%, while achieving the computational saving for real-time processing.**

***Note to Practitioners*—Hyperspectral fluorescence imaging offers an instant, noninvasive inspection method for detecting biomedical abnormalities. However, the huge amount of hyperspectral image data often makes real-time computer processing a challenging task. This paper suggests a band selection method of hyperspectral images based on the recursive divergence for the automatic detection of poultry carcasses. This method avoids transforming the original hyperspectral images to the feature space. Instead, it maximizes the class separability by considering the correlation information of spectral bands. In this paper, we mathematically characterize the use of divergence for band selection. Also, a set of recursive equations for the calculation of divergence with an additional band is derived to overcome the computational restrictions in real-time processing. The method may be extended to detect other biomedical abnormalities as well. In future research, we will incorporate the spatial and spectral information of the data in the development of appropriate band selection techniques for the hyperspectral data processing.**

***Index Terms*—Divergence, hyperspectral imaging, poultry inspection, skin tumor detection, spectral band selection, support vector machine (SVM).**

Manuscript received January 19, 2006; revised June 13, 2006. This paper was recommended for publication by Associate Editor M. Zhang and Editor D. Meldrum upon evaluation of the reviewers' comments. This work was supported by National Science Foundation CMMI-0644830.

Z. Du and S. G. Kong are with the Department of Electrical and Computer Engineering, The University of Tennessee, Knoxville, TN 37996-2100 USA (e-mail: zdu4@utk.edu; skong@utk.edu).

M. K. Jeong is with the Department of Industrial and Information Engineering, The University of Tennessee, Knoxville, TN 37996-0700 USA (e-mail: mjeong@utk.edu).

Color versions of one or more of the figures and tables in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2006.888048

I. Introduction

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

Poultry skin tumors are ulcerous lesions that are surrounded by a rim of thickened skin and dermis [1]. Skin cancer causes skin cells to lose the ability to divide and grow normally, and induces abnormal cells to grow out of control to form tumors. Tumorous carcasses often demonstrate swollen or enlarged tissue caused by the uncontrolled growth of new tissue. Tumor is not as visually obvious as other pathological diseases such as septicemia, air sacculitis, and bruise since its spatial signature appears as shape distortion rather than a discoloration. Therefore, conventional vision-based inspection systems operating in the visual spectrum may reveal limitations in detecting skin tumors on poultry carcasses.

Hyperspectral fluorescence imaging offers an instant, noninvasive inspection method for detecting biomedical abnormalities such as defects on poultry carcasses [2], [3]. Hyperspectral image data contain spatial information measured at a sequence of individual wavelength across broad spectral bands. Hyperspectral images show a detailed view of the spectral signature of the scene. The spectral signatures are useful for identifying various material compositions due to their unique spectral characteristics at particular wavelengths [4]. Fluorescence techniques are generally regarded as sensitive optical tools, and have proven to be effective in a number of scientific areas [5]. Fluorescence is a phenomenon where light is absorbed at a given wavelength and then is normally followed by the emission of light at a longer wavelength. A number of compounds emit fluorescence in the visible range when excited with ultraviolet radiation. Normal poultry skin often exhibits higher emissions compared to tumorous skin. The altered biochemical and morphological state of the neoplastic tissue is reflected in the spectral characteristics of the measured fluorescence.

Hyperspectral sensors collect the electromagnetic spectrum at dozens or hundreds of wavelength ranges in the visible and near infrared spectra. A three-dimensional (3-D) volume of data in spatial and spectral spaces characterizes a hyperspectral image. Such a large amount of hyperspectral image data often makes real-time computer processing a challenging task [6]. In most cases, high spectral correlation of hyperspectral image data enables us to reduce the feature dimensionality without substantial loss of classification accuracy. Among widely used dimensionality reduction methods, the principal component analysis (PCA) rearranges the data in terms of the significance measured by the eigenvalues of the data covariance matrix.

Band selection methods identify a subset of spectral bands significant in terms of information content, and remove the bands of less importance. All the spectral bands do not carry the same amount of information. Many criteria such as distance measures, information-theoretic approaches, and eigenanalysis have been proposed to select the spectral bands in terms of information content.

Various methods have been extensively used for hyperspectral band selection. Keshava [7] proposed a band selection algorithm based on the spectral angle mapper (SAM) metric, which is the angle between the two spectra. Tu proposes a band selection algorithm coupled with feature extraction for data dimensionality reduction based on canonical analysis (CA) [8]. Using the eigenvalues and eigenvectors generated by CA, a loading factor matrix can be defined, through which a discriminant power (DP) is calculated for each bands. Du [9] used high-order moments for band ranking and divergence for band decorrelation. Du *et al.* [10] used the independent component analysis for the band selection. Ifarraguerri and Parairie [11] presented the band selection algorithm based on the Jefferis–Matusita metric. Guyon *et al.* [12] propose a new method of gene selection utilizing support vector machine (SVM) methods based on recursive feature elimination (RFE). At each step, the coefficients of the weight vector of a linear SVM are used to compute the feature ranking score. The feature with the smallest ranking score is eliminated.

Adjacent spectral bands in hyperspectral images are often highly correlated. The divergence takes into account the correlation that exists among various selected bands, and it is a simple and efficient measurement of statistical class separability used in pattern recognition [13]. We present the band selection method of hyperspectral images based on the maximum divergence for the automatic detection of poultry carcasses. This method avoids transforming the original hyperspectral images to the feature space. Instead, it maximizes the class separability by considering the correlation information of spectral bands. Also, a set of recursive equations for the calculation of divergence with an additional band is derived to overcome the computational restrictions in real-time processing. With a small number of optimal spectral bands selected from hyperspectral image data, we can build a real-time classification system with multispectral image sensors for a specific application [11], [14].

The paper is organized as follows. Section II briefly describes the hyperspectral imaging system. Section III presents a band selection method based on the recursive calculation of divergence with an additional band. In Section IV, the SVM classifier is presented. Experiment results are reported in Section V. Conclusions are drawn in Section VI.

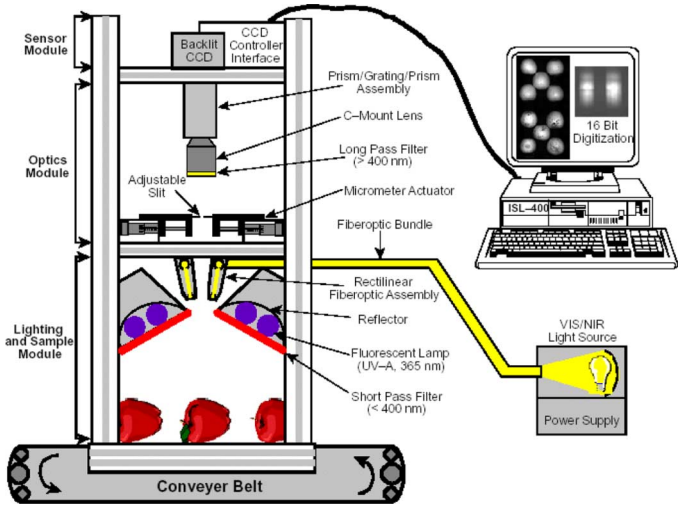


Fig. 1. Hardware components of the ISL hyperspectral imaging system.

II. Hyperspectral Fluorescence Imaging

Instrumentation and Sensing Laboratory (ISL) at Beltsville Agricultural Research Center, Beltsville, MD, has developed a laboratory-based line-by-line hyperspectral imaging system capable of reflectance and fluorescence imaging for uses in food safety and quality research [15], [16]. The system employs a pushbroom method in which a line of spatial information with a full spectral range per spatial pixel is captured sequentially to cover a volume of spatial and spectral data. Fig. 1 shows the ISL hyperspectral imaging system equipped with a CCD camera, a spectrograph, a sample transport mechanism, and two lighting sources for reflectance and fluorescence sensing. Two fluorescent lamp assemblies are used to provide a near uniform UV-A (365 nm) excitation to the sample area for fluorescence measurements. A short-pass filter placed in front of the lamp housing is used to prevent transmittance of radiations greater than approximately 400 nm, and thus eliminate the potential spectral contamination by pseudo-fluorescence. The system acquires the data via line-by-line scans while transporting sample materials via a precision positioning table.

Data produced by hyperspectral imaging systems can be represented by a 3-D cube of images $I(m, n, \lambda_i)$, where (m, n) denotes the spatial coordinate of a pixel in the image of the size $M \times N$ ($m = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$) and λ_i denotes the wavelength of the ith spectral band ($i = 1, 2, \ldots, L$). The value $I(m, n, \lambda_i)$ indicates the fluorescence response of the pixel (m, n) at a wavelength λ_i of the ith spectral band. The ISL hyperspectral image system captures 65 spectral bands ($L = 65$) at the wavelengths from λ_1 (425.4 nm) to λ_{65} (710.7 nm) in visible light spectrum. A hyperspectral image of a poultry sample consists of a spatial dimension of 400×460 pixels where each pixel denotes 1×1 mm of spatial resolution. Each pixel has a 16-bit gray-scale resolution. The data size of a hyperspectral image sample is approximately

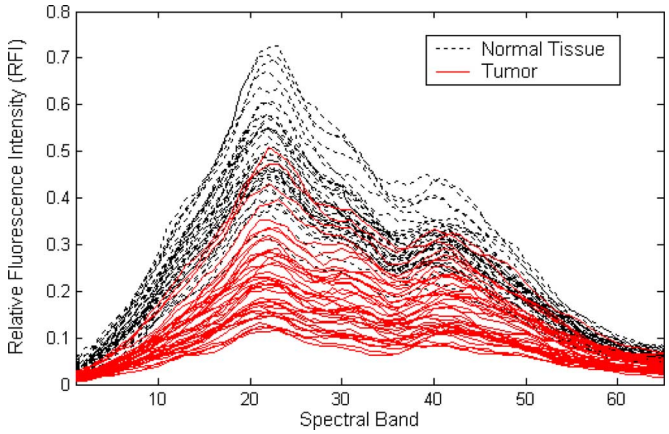


Fig. 2. Spectral signatures of the tumor and normal tissue measured by relative fluorescence intensity.

24 MB($= 400 \times 460$ pixels $\times$ 65 bands$\times$2 bytes). The speed of the conveyer belt was adjusted based on the predetermined CCD exposure time and data transfer rate.

Spectral signature reveals the characteristics of the different types of tissues. Fig. 2 shows the relative fluorescence intensity of hyperspectral image data at each spectral band for normal tissues and tumors. Normal tissues have a large peak response at approximately band 22 and a smaller peak at approximately band 45. Tumors show lower fluorescence intensities than normal tissues on average, but have strong response between the bands 40 and 45 relative to the peak near the band 22. Background pixels show low fluorescence intensity and an almost flat response over the entire spectral range due to the carrying tray being covered with a nonfluorescent flat black paint.

III. Spectral Band Selection

A. Band Selection Based on the Recursive Divergence

The divergence is one of popular criteria for class separability. Spectral bands in hyperspectral images are highly correlated. The divergence takes into account the correlation that exists among the various selected bands and influences the classification capabilities of the spectral bands that are selected. We use the divergence to determine feature ranking and to evaluate the effectiveness of class discrimination in hyperspectral image data. The divergence is defined as the total average information for discriminating class ω_i from class ω_j, and given by [13]

$$J_{ij}(\boldsymbol{x}) = \int_{-\infty}^{+\infty} [p_i(\boldsymbol{x}) - p_j(\boldsymbol{x})] \ln \frac{p_i(\boldsymbol{x})}{p_j(\boldsymbol{x})} d\boldsymbol{x} \tag{1}$$

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

Suppose that signal classes are characterized by p-dimensional multivariate normal distributions: $N(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\theta}_i$ and $\boldsymbol{\Sigma}_i$ are the mean vector and covariance matrix of class ω_i, respectively. Then, the divergence between these two classes is given by [20]

$$J_{ij}(\boldsymbol{x}) = \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}\right)(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)^t\right] + \frac{1}{2} tr\left[(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)\left(\boldsymbol{\Sigma}_j^{-1} - \boldsymbol{\Sigma}_i^{-1}\right)\right] \tag{2}$$

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

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

$$\begin{aligned} J_{ij}(\boldsymbol{x}) &= tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)^t\right] \\ &= (\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j) \end{aligned} \tag{3}$$

which equals the Mahalanobis generalized distance. The form of (3) is close to that of the Bhattacharyya distance with first and second terms indicating class separability due to mean- and covariance-differences. The advantage of divergence is that both the first and second terms are expressed by the trace of a matrix, while the Bhattacharyya distance is the combination of trace and determinant.

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

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i} \sum_{j=1}^{n} z_{ij}(\boldsymbol{x}_j - \hat{\boldsymbol{\theta}}_i)(\boldsymbol{x}_j - \hat{\boldsymbol{\theta}}_i)^t \tag{4}$$

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

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

$$n_c = \frac{L!}{(L-d)!d!} \tag{5}$$

which can be very large even for moderate values of L and d. For example, selecting the best 6 spectral bands out of 65 bands in our case study of the detection of poultry carcasses means that 82 598 880 band sets must be considered, and evaluating the divergence criterion for every band set in an acceptable time may not be feasible. Thus, we propose the suboptimal band selection method based on the recursive calculation of the divergence.

The basic idea is to build up a set of d spectral bands incrementally, starting with the empty set. That is, the search algorithm constructs the set of spectral bands at the ith stage of the algorithm from that at the $(i-1)$th stage by the addition of a spectral band from the current optimal set. The divergence criterion (2) at stage i can be evaluated by updating its value already calculated for stage $(i-1)$ instead of computing the divergence from their definitions. This results in substantial computational savings.

Let $J_{ij}(\boldsymbol{x}_p^*)$ be the divergence with p selected bands and $J_{ij}(\boldsymbol{x}_p^*, x_{p+1}^*)$ the divergence with the additional band x_{p+1}^*. Suppose the additional band x_{p+1}^* has mean θ_k^*, variance σ_k^2; and the covariance vector between x_{p+1}^* and the elements of

$\boldsymbol{x}_p$, $\boldsymbol{z}_k$ for class k ($= i$ or j). Then the new mean vectors and new covariance matrix are $\boldsymbol{\theta}_k^v = (\boldsymbol{\theta}_{k,p}^*; \theta_k^*)^t$, ($k = i$ or j) and

$$\boldsymbol{\Sigma}_{k,p+1} = \begin{pmatrix} \boldsymbol{\Sigma}_{k,p} & \boldsymbol{z}_k \\ \boldsymbol{z}_k^t & \sigma_k^2 \end{pmatrix}. \tag{6}$$

The divergence with an additional band x_{p+1}^* can be recursively calculated in an efficient way as follows:

$$J_{ij}\left(\boldsymbol{x}_p^*, x_{p+1}^*\right) = J_{ij}\left(\boldsymbol{x}_p^*\right) + \Delta_{ij}\left(x_{p+1}^*\right) \tag{7}$$

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

$$\begin{aligned}\Delta_{ij}\left(x_{p+1}^*\right) = &\frac{1}{2\delta_i}\left[\left(\theta_i^* - \theta_j^*\right) - \left(\boldsymbol{\theta}_{i,p}^* - \boldsymbol{\theta}_{j,p}^*\right)^t \boldsymbol{\gamma}_i\right]^2 \\ &+ \frac{1}{2\delta_j}\left[\left(\theta_i^* - \theta_j^*\right) - \left(\boldsymbol{\theta}_{i,p}^* - \boldsymbol{\theta}_{j,p}^*\right)^t \boldsymbol{\gamma}_j\right]^2 \\ &+ \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p+1} - \boldsymbol{\Sigma}_{j,p+1}\right)\left(\delta_i^{-1}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^t + \delta_j^{-1}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^t\right)\right. \\ &\quad + \left(\boldsymbol{z}_i^t - \boldsymbol{z}_j^t\right)\left(\delta_i^{-1}\boldsymbol{\gamma}_i - \delta_j^{-1}\boldsymbol{\gamma}_j\right) \\ &\quad \left. + \left(\sigma_i^2 - \sigma_j^2\right)\left(\delta_i - \delta_j\right)\right] \end{aligned} \tag{8}$$

where $\boldsymbol{\gamma}_k = \boldsymbol{\Sigma}_{k,p}^{-1}\boldsymbol{z}_k$ and $\delta_k = \sigma_k^2 - \boldsymbol{z}_k^t\boldsymbol{\Sigma}_{k,p}^{-1}\boldsymbol{z}_k$. See Appendix A for the detailed derivation of the incremental divergence.

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

$$\Delta_{ij}\left(x_{p+1}^*\right) = \frac{\left[\left(\theta_i^* - \theta_j^*\right) - \left(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\right)^t \boldsymbol{\Sigma}_p^{-1}\boldsymbol{z}\right]^2}{\sigma^2 - \boldsymbol{z}^t\boldsymbol{\Sigma}_p^{-1}\boldsymbol{z}}. \tag{9}$$

Equation (8) gives an efficient way to calculate the divergence with the additional band. When a new band is to be considered, it is not necessary to compute the divergence of all selected bands; only the incremental divergence is calculated. The procedure for an efficient band selection based on the recursive equation of divergence can be described as follows. The diagram is shown in Fig. 3.

Spectral band selection algorithm with recursive divergence.

Step 1) Set $\boldsymbol{D}$ to the initial band set, $\boldsymbol{S}$ to the empty set. Select a starting band (say λ_i) by exhaustively calculate all bands and find the one with the maximum divergence.

Step 2) Calculate $\Delta_{ij}(x_{p+1}^*)$ according to (8) for all the remaining bands. If V_p represents a set of p spectral bands then, the best band at a given iteration, V_{p+1} is the set for which the incremental effectiveness of additional band has its maximum value.

Step 3) Select the band having the largest incremental effectiveness (say λ_k), and add it to the selected band. If stopping criterion is met, then stop and output selected band set $\boldsymbol{S}$. Otherwise, go to Step 2).

The algorithm will stop when certain detection accuracy achieved. For this task, the industry-accepted accuracy is 90–95%, and using 6 bands can obtain 90.6% accuracy, so the algorithm stops when 6 bands are selected.

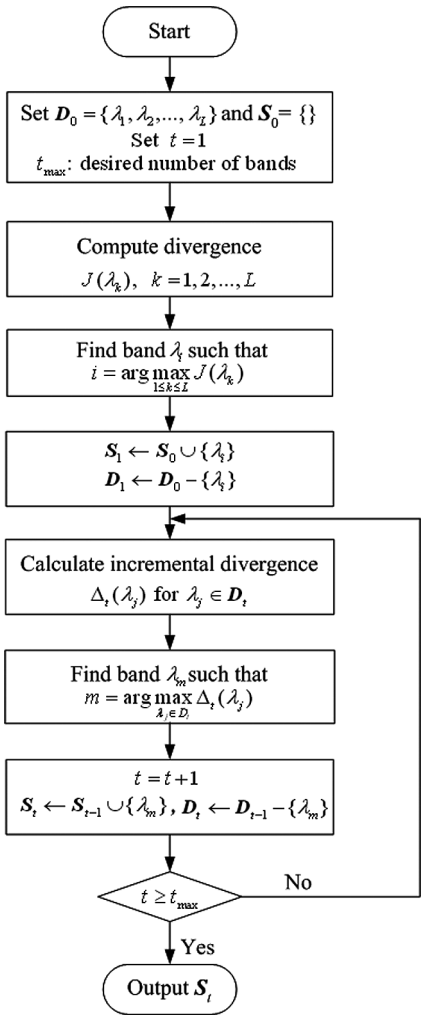


Fig. 3. Block diagram of the proposed band selection method.

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

$$J(\lambda_q) = \sum_{i=1}^{M-1} \sum_{i<j\le M} J_{ij}(\lambda_q). \tag{10}$$

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

$$\Delta\left(x_{p+1}^*\right) = \sum_{i=1}^{M-1} \sum_{i<j\le M} \Delta_{ij}\left(x_{p+1}^*\right). \tag{11}$$

From our experiences, a band selection method based on the CA proposed by Tu [8] showed the best performance compared to other band selection methods, such as PCA-based method, distance based method and information entropy based method. Thus, we compare the performance of our proposed method with CA method in Section V, and the brief introduction of CA approach is given below.

B. Band Selection Based on Canonical Analysis

Canonical analysis [8] computes the transformation that maximizes the between-class scatter and minimizes the within-class scatter. Let $\boldsymbol{x}_{ki}$ be the ith sample and $\boldsymbol{\mu}_k$ be the mean of the class $k(k = 1, 2, \ldots, K)$, respectively. Let N_k be the number of samples in class k. Then the within-class scatter matrix is defined as

$$\boldsymbol{S}_w = \sum_{k=1}^{K} \sum_{l=1}^{N_k} (\boldsymbol{x}_{kl} - \boldsymbol{\mu}_k)(\boldsymbol{x}_{kl} - \boldsymbol{\mu}_k)^t. \tag{12}$$

Between-class scatter matrix is defined as the sum of outer products of the centered means of each class

$$\boldsymbol{S}_b = \sum_{k=1}^{K} (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^t \tag{13}$$

where $\boldsymbol{\mu}$ indicates the mean of all the data. A linear transformation is given by a matrix whose columns are the eigenvectors of the matrix $\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$

$$\boldsymbol{S}_w^{-1}\boldsymbol{S}_b\boldsymbol{e}_j = \boldsymbol{e}_j d_j \tag{14}$$

where d_j are eigenvalues arranged in descending order $(d_1 \geq d_2 \geq \ldots \geq d_L)$. The eigenvector $\boldsymbol{e}_j$ corresponds to the eigenvalue d_j. The term $r_{ji} = \sqrt{d_j}(e_{ji}/\|\boldsymbol{e}_j\|)$ denotes the loading factor of canonical component j at the ith band, and e_{ji} is the ith element for eigenvector $\boldsymbol{e}_j$. The discriminating power of the ith band can be measured by the CA score as

$$J_i^{CA} = \sum_{j=1}^{K-1} r_{ji}^2 = \sum_{j=1}^{K-1} \frac{d_j e_{ji}^2}{\|\boldsymbol{e}_j\|^2}. \tag{15}$$

The bands are ranked in terms of the CA score. The CA selects the spectral bands corresponding to the first p largest CA scores.

IV. Classification of Spectral Bands With SVM

A support vector machines is used as a classifier for tumor detection. After band selection, all spectral characteristics were used as features to train a SVM classifier. The SVM [17], [18] finds the optimal separating hyperplane that maximizes the margin between the classes. Consider the case of classifying a set of linearly separating data. Assume a set of training vectors $\boldsymbol{u}_i$ that belong to two classes with the class label $y_i = \{+1, -1\}(i = 1, \ldots, n)$. The data set is called linearly separable by a hyperplane $\boldsymbol{w}^t\boldsymbol{u} + b = 0$ if a vector $\boldsymbol{w}$ and a scalar b exists such that

$$\begin{aligned} \boldsymbol{w}^t\boldsymbol{u}_i + b \geq +1 \quad &\text{if } y_i = +1 \\ \boldsymbol{w}^t\boldsymbol{u}_i + b \leq -1 \quad &\text{if } y_i = -1 \end{aligned} \tag{16}$$

which can be combined into an inequality

$$y_i(\boldsymbol{w}^t\boldsymbol{u}_i + b) - 1 \geq 0, \quad i = 1, 2, \cdots, n. \tag{17}$$

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

$$J(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{n} \alpha_i \left[1 - y_i(\boldsymbol{w}^t\boldsymbol{u}_i + b)\right], \quad \alpha_i \geq 0. \tag{18}$$

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

$$\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{u}_i \tag{19}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \tag{20}$$

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

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\boldsymbol{u}_i^t \boldsymbol{u}_j\right) \tag{21}$$

subject to the constraints

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

and

$$\alpha_i \geq 0, \quad i = 1, 2, \cdots, n. \tag{22}$$

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

$$\alpha_i \left[1 - y_i(\boldsymbol{w}^t\boldsymbol{u}_i + b)\right] = 0, \quad i = 1, 2, \cdots, n. \tag{23}$$

Training samples $\boldsymbol{u}_i$ corresponding to nonzero Lagrange multipliers (α_i) are called support vectors. Support vectors lie on the class boundaries at the distance $1/\|\boldsymbol{w}\|$ from the hyperplane. All remaining samples in the training set but support vectors do not play a role in finding optimal decision boundaries. The discriminant function corresponding to the optimal hyperplane depends both on the Lagrange multipliers and on the support vectors, i.e.,

$$f(\boldsymbol{u}) = \sum_{i \in S} \alpha_i y_i \left(\boldsymbol{u}_i^t \boldsymbol{u}\right) + b^* \tag{24}$$

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

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

$$f(\boldsymbol{u}) = \operatorname{sgn}\left(\sum_{j \in S} y_j \alpha_j K(\boldsymbol{u}, \boldsymbol{u}_j) + b\right). \tag{25}$$

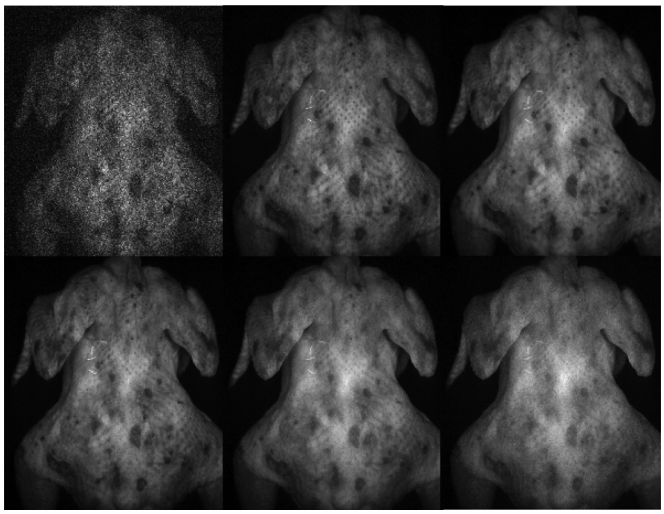
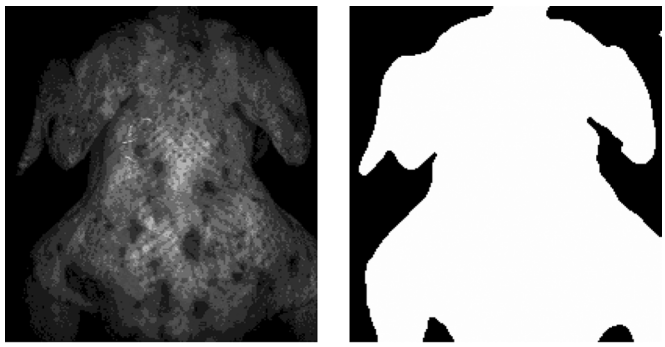


Fig. 4. Hyperspectral images of a poultry carcass.



(a) (b)

Fig. 5. Segmentation of hyperspectral fluorescence image with a threshold. (a) Original image (λ_{10}). (b) Segmentation result.

Frequent choices of kernels include polynomial, radial basis, and sigmoid function. We have used radial basis kernel $K(\boldsymbol{u}, \boldsymbol{u}_j) = e^{-\|u-u_j\|^2/2p^2}$ with parameter $p = 2$ in our experiments. These values have been selected after series of numerical experiments on the training and testing data to get the best generalization ability. Generally, they should be chosen in such a way that at smallest possible number of support vectors the best performance of the classifier on the testing data be observed. The SVM implementation provided by Gunn [19] is used to perform the SVM training and classification.

V. Experiment Results

Twelve chicken carcasses were collected from a poultry processing plant owned by Allen Family Foods, Inc., Cordova, MD, in March and May 2002. A Food Safety and Inspection Service (FSIS) veterinarian at the plant identifies the condition of the poultry carcasses. Hyperspectral images obtained consist of 460 × 400 pixels with 65 spectral bands. The spectral band has discrete wavelengths from 425.4 nm (λ_1) to 710.7 nm (λ_{65}). The sample poultry carcasses were placed on a tray painted with a nonfluorescent 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. Fig. 4 shows six spectral images $(\lambda_1, \lambda_{11}, \ldots, \lambda_{51})$ of a hyperspectral image sample obtained by ISL's system.

Image segmentation is performed as preprocessing to remove the poultry carcasses from the background. The background is the tray on which the poultry carcass is placed. Due to the tray painted with nonfluorescent, flat black paint, the fluorescence intensities of these trays are almost same. A fixed threshold can easily remove the background. Fig. 5 shows the image segmentation result.

In order to train SVM, 400 pixels were picked from the training data (as shown in Fig. 4), where 200 pixels are from tumor tissue and 200 pixels are from normal tissue. These 400 pixels are used to train a SVM classifier. Then we use this SVM classifier to detect tumors on the 11 testing chicken data.

TABLE I
Band Selection Result

Methods	Number of selected bands					
	1	2	3	4	5	6
RD	4	4,65	4,65,13	4,65,13,54	4,65,13,54,22	4,65,13,54,22,32
ES	4	20,52	20,52,64	17,20,52,64	8,13,19,52,64	8,17,30,34,52,64
CA	56	56,51	56,51,43	56,51,43,61	56,51,43,61,48	56,51,43,61,48,26

TABLE II
Computation Time for Selecting Six Bands of the Different Methods

Method	RD	ES	CA
Mean Total Time (s)	0.5	71829	0.063

Table I shows the selected spectral bands for the given number of bands based on the recursive divergence (RD), exhaustive search (ES) and CA on the poultry carcasses data. The computational time for selecting the best six spectral bands out of 65 bands using three different methods are summarized in Table II. We used a computer with Pentium IV 2.6-GHz processor to conduct the experiments. The programming language used is MATLAB 6.5. For the ES method, 82598880 band sets must be considered; therefore, it requires the longest computational time (71829 s); while for the RD method, 375 band sets need to be considered because of the recursive property, so it took only 0.5 s. The CA method just ranks the bands without any search strategy and it took 0.063 s.

Fig. 6 shows the divergence values of the selected bands from different methods. The divergences with one band for three methods are around 20, and the divergence increases to around 40 when two spectral bands are selected. With more bands, the divergence will increase because of its monotonic property, but the incremental decreases. Fig. 6 shows that the bands selected by the RD method have a larger divergence value than those with the CA method. Consider the 3-band case as an example: RD method selects bands λ_4, λ_{65}, and λ_{13}, while CA λ_{56}, λ_{51}, and λ_{43}. In case of CA, λ_{56} and λ_{51} are adjacent bands, which are usually highly correlated. Since the divergence takes into account the correlation that exists among the selected bands, this results in a smaller divergence value for CA method. In general, it can be observed that larger divergence means more

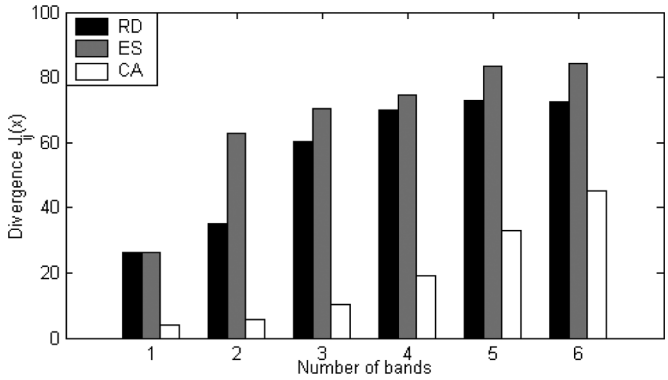


Fig. 6. Divergence of selected bands.

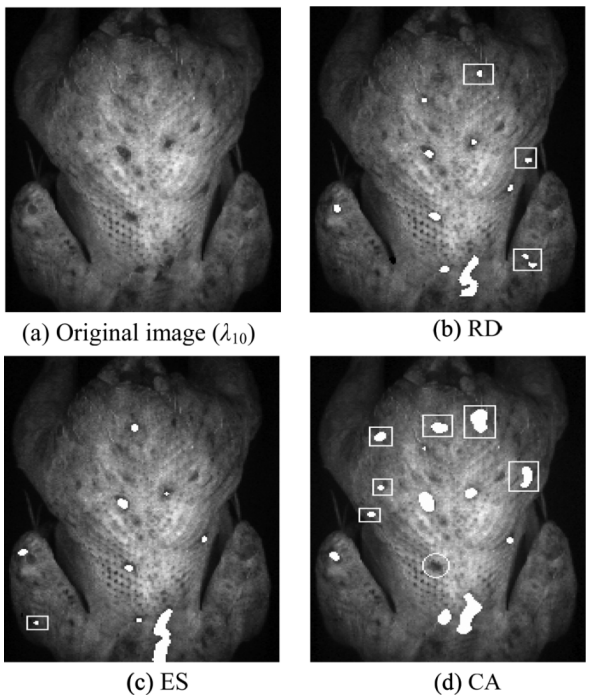


Fig. 7. Tumor detection results with the four spectral bands selected. (a) Original image. Tumor detection results with (b) RD, (c) ES, and (d) CA.

separability, and tumor detection results in Fig. 7 validate this statement.

Fig. 7 shows the original images with tumors detected by the RD, ES, and CA methods. White spots indicate the tumors correctly detected and the white areas enclosed by a rectangle indicate false positives. Circled areas are the tumors not detected by the algorithms.

Table III summarizes the tumor detection results on 11 poultry samples with six spectral bands selected using the RD, ES, and CA. The average detection rates were 90.6% for RD and 93.75% for ES and 81.25% for CA. Band selection with the RD has three missed tumors and 19 false positives (FPs) on average while the CA shows six missed tumors and 28 false positives.

In the chicken industry, the acceptable human inspection accuracy is 90–95%. Our accuracy defines the ratio of the numbers of tumor spots detected and identified by human experts.

TABLE III
TUMOR DETECTION PERFORMANCE OF THE RD, ES, AND CA WITH SIX BANDS

Image No.	Number of Tumors	RD		ES		CA	
		# Found	FPs	# Found	FPs	# Found	FPs
1	8	8	3	8	1	7	6
2	2	2	2	2	2	2	1
3	0	0	2	0	2	0	2
4	3	3	1	3	1	2	3
5	2	2	2	2	1	2	3
6	2	2	2	2	2	2	3
7	2	2	2	2	1	2	1
8	0	0	0	0	1	0	1
9	4	3	2	3	1	2	3
10	7	5	2	6	2	5	2
11	2	2	1	2	1	2	3
Total	32	29 (90.6%)	19	30(93.75%)	15	26 (81.25%)	28

Though its accuracy is around 90%, the proposed recursive divergence method correctly recognized all 11 testing chicken samples with or without tumors. This level of accuracy is well acceptable in industry.

VI. CONCLUSIONS AND FUTURE WORK

Hyperspectral imaging offers an instant, noninvasive diagnostic procedure based on the analysis of the spectral properties of the tissue. This paper presented a band selection method in hyperspectral imaging based on the maximum divergence for the detection of skin tumors on poultry carcasses. The divergence takes into account the correlation that exists among the various selected bands and influences the classification capabilities of the spectral bands that are selected. Also, a set of recursive equations for the calculation of incremental divergence with an additional band is derived to overcome the computational restrictions in real-time processing.

With a small number of optimal spectral bands selected from hyperspectral image data, we could build a real-time classification system with multispectral image sensors for the detection of poultry skin tumors. Our proposed band selection method reduces the computational complexity for real-time processing of hyperspectral images. A support vector machine classifier with radial basis function kernel finds an optimal decision boundary in a reduced feature space for detecting skin tumors. The tumor detection accuracy averaged over 11 different testing hyperspectral images was 90.6% for recursive divergence, while the CA produced only 81.25% accuracy.

One of the distinguishing properties of hyperspectral image data is the high dimensional spectral information coupled with a two-dimensional pictorial representation amenable to image interpretation. In the future work, we will incorporate the spatial and spectral information of the data in the development of appropriate band selection techniques for the hyperspectral data processing.

APPENDIX
DERIVATION OF INCREMENTAL DIVERGENCE

The divergence with the additional of a band x^*_{p+1} can be calculated based on its definition in (2) as follows:

$$J_{ij}\left(\boldsymbol{x}^*_p, x^*_{p+1}\right) = \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p+1} - \boldsymbol{\Sigma}_{j,p+1}\right)\left(\boldsymbol{\Sigma}^{-1}_{j,p+1} - \boldsymbol{\Sigma}^{-1}_{i,p+1}\right)\right] + \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}^{-1}_{i,p+1} + \boldsymbol{\Sigma}^{-1}_{j,p+1}\right)\left(\boldsymbol{\theta}^{\nu}_i - \boldsymbol{\theta}^{\nu}_j\right)\left(\boldsymbol{\theta}^{\nu}_i - \boldsymbol{\theta}^{\nu}_j\right)^t\right].$$

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

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

where $\boldsymbol{\gamma}_k = \boldsymbol{\Sigma}_{k,p}^{-1} z_k$ and $\delta_k = \sigma_k^2 - z_k^t \boldsymbol{\Sigma}_{k,p}^{-1} z_k$ [20].

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

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

where

$$\begin{aligned} \Delta_{ij}\left(x_{p+1}^*\right) = &\frac{1}{2\delta_i}\left[\left(\theta_{i,p+1}^* - \theta_{j,p+1}^*\right) - \left(\boldsymbol{\theta}_{i,p}^* - \boldsymbol{\theta}_{j,p}^*\right)^t \boldsymbol{\gamma}_i\right]^2 \\ &+ \frac{1}{2\delta_j}\left[\left(\theta_{i,p+1}^* - \theta_{j,p+1}^*\right) - \left(\boldsymbol{\theta}_{i,p}^* - \boldsymbol{\theta}_{j,p}^*\right)^t \boldsymbol{\gamma}_j\right]^2 \\ &+ \frac{1}{2} tr\left[\left(\boldsymbol{\Sigma}_{i,p+1} - \boldsymbol{\Sigma}_{j,p+1}\right)\left(\delta_i^{-1}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^t + \delta_j^{-1}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^t\right)\right. \\ &\qquad + \left(z_i^t - z_j^t\right)\left(\delta_i^{-1}\boldsymbol{\gamma}_i - \delta_j^{-1}\boldsymbol{\gamma}_j\right) \\ &\qquad \left. + \left(\sigma_i^2 - \sigma_j^2\right)\left(\delta_i - \delta_j\right)\right]. \end{aligned}$$

ACKNOWLEDGMENT

The authors are grateful to Dr. Y. R. Chen and Dr. M. S. Kim at Instrumentation and Sensing Laboratory, Beltsville Agricultural Research Center, Beltsville, MD, for providing the hyperspectral fluorescence image data for this research.

REFERENCES

[1] B. W. Calnek, H. J. Barnes, C. W. Beard, W. M. Reid, and H. W. Yoder, *Diseases of Poultry*. Ames, IA: Iowa State Univ. Press, 1991, ch. 16, pp. 386–484.

[2] J. Zhang, C. Chang, S. Miller, and K. Kang, "Optical biopsy of skin tumors," in *Proc. Conf. Engineering in Medicine and Biology*, 1999, vol. 2, pp. 13–16.

[3] K. Chao, P. Mehl, and Y. R. Chen, "Use of hyper- and multi-spectral imaging for detection of chicken skin tumors," *Appl. Eng. Agricult.*, vol. 18, no. 1, pp. 113–119, 2002.

[4] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.

[5] B. Albers, J. DiBenedetto, S. Lutz, and C. Purdy, "More efficient environmental monitoring with laser-induced fluorescence imaging," *Biophoton. Int. Mag.*, vol. 2, no. 6, pp. 42–54, 1995.

[6] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[7] N. Keshava, "Best bands selection for detection in hyperspectral processing," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2001, vol. 5, pp. 3149–3152.

[8] T. Tu, C. Chen, J. Wu, and C. Chang, "A fast two-stage classification method for high-dimensional remote sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 1, pp. 182–191, Jan. 1998.

[9] Q. Du, "Band selection and its impact on target detection and classification in hyperspectral image analysis," *Adv. Tech. Anal. Remotely Sensed Data*, pp. 374–377, 2003.

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

[11] A. Ifarraguerri and M. W. Prairie, "Visual method for spectral band selection," *IEEE Geosci. Remote Sensing Lett.*, vol. 1, no. 2, pp. 101–106, Apr. 2004.

[12] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, no. 1, pp. 389–422, 2002.

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

[14] B. Park, Y. R. Chen, M. Nguyen, and H. Hwang, "Characterizing multispectral images of tumorous, bruised, skin-torn, and wholesome poultry carcasses," *Trans. Amer. Soc. Agricult. Eng.*, vol. 39, no. 5, pp. 1933–1941, 1996.

[15] M. Kim, Y. Chen, and P. Mehl, "Hyperspectral reflectance and fluorescence imaging system for food quality and safety," *Trans. Amer. Soc. Agricult. Eng.*, vol. 44, no. 3, pp. 721–729, 2001.

[16] S. G. Kong, Y. Chen, I. Kim, and M. Kim, "Analysis of hyperspectral fluorescence images for poultry skin tumor inspection," *Appl. Opt.*, vol. 43, no. 4, pp. 824–833, 2004.

[17] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[18] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[19] S. Gunn, "Support vector machines for classification and regression," Image Speech and Intelligent Systems Group, Dept. Electron. Comput. Sci., Univ. Southhampton, Southampton, U.K., Tech. Rep. MP-TR-98-05, 1998.

[20] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

Zheng Du received the B.S. and M.S. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2000 and 2003, respectively. He is currently pursuing the Ph.D. degree at the University of Tennessee, Knoxville.

His research interests include hyperspectral image analysis, pattern recognition, statistical learning, machine learning, and support vector machines with applications in hyperspectral image.

Myong K. Jeong (M'04) received the Ph.D. degree in industrial and systems engineering from the Georgia Institute of Technology, Atlanta, in 2004.

He is currently an Assistant Professor in the Department of Industrial and Information Engineering, University of Tennessee, Knoxville. His research interests include data mining, pattern recognition, electronics manufacturing, semiconductor manufacturing processes, and sensor data analysis.

Dr. Jyong is a member of the Institute for Operations Research and the Management Sciences (INFORMS), the Institute of Industrial Engineers (IIE), and the Society for Mechanical Engineers (SME).

Seong G. Kong (SM'03) received the B.S. and the M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1982 and 1987, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1991.

He was an Associate Professor of electrical engineering at Soongsil University, Seoul, from 1992 to 2000. He served as Department Chair from 1998 to 2000. During the years 2000–2001, he was with School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, as Visiting Scholar. Since 2002, he has been an Associate Professor in the Department of Electrical and Computer Engineering at the University of Tennessee, Knoxville. His research interests include pattern recognition, image processing, and intelligent systems.

Dr. Kong received an honorable mention paper award from the American Society of Agricultural and Biological Engineers in 2004 and the Most Cited Paper Award for *Computer Vision and Image Understanding* in 2007. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS.