Title

Evolutionary refinement approaches for band selection of hyperspectral images with applications to automatic monitoring of animal feed quality

Journal

Intelligent Data Analysis, 18(1)

ISSN

1088-467X

Authors

Wilcox, Philip
Horton, Timothy M
Youn, Eunseog
et al.

Publication Date

2014

DOI

10.3233/ida-130626

Peer reviewed

Intelligent Data Analysis 00 (2013) 1–18 1
DOI 10.3233/IDA-130626
IOS Press

Evolutionary refinement approaches for band selection of hyperspectral images with applications to automatic monitoring of animal feed quality

Philip Wilcox[a], Timothy M. Horton[b], Eunseog Youn[a,*], Myong K. Jeong[c], Derrick Tate[b], Timothy Herrman[d] and Christian Nansen[e,f]
[a]*Department of Computer Science, Texas Tech University, Lubbock, TX, USA*
[b]*Department of Mechanical Engineering, Texas Tech University, Lubbock, TX, USA*
[c]*Department of Industrial and Systems Engineering and RUTCOR (Rutgers Center for Operations Research), Rutgers, the State University of New Jersey, Piscataway, NJ, USA*
[d]*Office of the Texas State Chemist, Texas A&M, TX, USA*
[e]*Texas AgriLife Research, Lubbock, TX, USA*
[f]*Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, USA*

Abstract. This paper presents methods for spectral band selection in hyperspectral image (HSI) cubes based on classification of reflectance data acquired from samples of livestock feed materials and ruminant-derived bonemeal. Automated detection of ruminant-derived bonemeal in animal feed is tested as part of an on-going research into development of automated, reliable fast and cost-effective quality control systems. HSI cubes contain spectral reflectance in both spatial dimensions and spectral bands. Support vector machines are used for classification of data in various domains. Selecting a subset of the spectral bands speeds processing and increases accuracy by reducing over-fitting. We developed two methods utilizing divergence values for selecting spectral band sets, 1) evolutionary search method and 2) divergence-based recursive feature elimination approach.

Keywords: Hyperspectral image cubes, animal feed quality monitoring, hyperspectral band selection, reflectance analysis, evolutionary search, divergence, recursive feature elimination

1. Introduction

Monitoring animal feed quality is particularly important because contaminants in feed can, in addition to hindering the development and health of livestock animals, be transferred to milk and meat products and so present health risks to consumers. Specifically, ruminant-derived feed products may be carriers of the Bovine Spongiform Encephalopathy (BSE) prion [19,27]. Potential points of prohibited animal protein introduction include cross-contamination by transporters, protein blenders working with

*Corresponding author: Eunseog Youn, Department of Computer Science, Texas Tech University, Lubbock, TX 79409, USA. Tel.: +1 806 742 3527; Fax: +1 806 742 3519; E-mail: eun.youn@ttu.edu.

multiple sources of animal protein, and feed mills that manufacture protein supplement for cattle and non-ruminant species, with the latter containing prohibited animal protein. The latest case of BSE in the US was confirmed in 2006 from a cow in Alabama [7]. In Europe, there is a ban on use of all rendered animal protein in feedstuffs for food production animals [1,3]. In the US, the Food and Drug Administration (FDA) and state feed control officials evaluate ruminant feed samples for the presence of prohibited animal protein as part of a nationwide BSE detection program.

The current procedure for monitoring of animal feed quality is as follows. Feed samples are collected during inspections of feedlots and feed mills using approved analytical sampling methods and subsequently inspected under laboratory conditions based on standardized microscopy procedures of FDA [2]. The physical preparation of feed samples for inspection takes about 2 hours per sample, and careful microscopy based inspection by specially training technicians takes additional 2 hours per feed sample. If bonemeal fragments or other potentially prohibited ruminant-derived contaminants are found during microscopy, the feed sample is subjected to further analyses using polymerase chain reaction (PCR) technology. The current monitoring system is time consuming and highly subjective to human error, so there is a great need for development of automated quality control procedures for inspection of feed materials.

In a recent study of feed materials and ruminant derived bonemeal, a sequence of linear discriminant analyses (LDA) was deployed to gradually improve the classification accuracy of hyperspectral profiles. Based on independent testing, it was found that the minimum detection level for an automated machine vision based system was about 1% bonemeal contamination (by weight).

In order to overcome the limitations of a manual quality control system, this paper proposes the automatic monitoring approach to animal feed quality based on reflectance data acquired with a hyperspectral imaging system. Using HSI technology has the following advantages over traditional microscopy: it is non-destructive, generally does not require much physical preparation of the target objects, and can provide real-time results. Hyperspectral imaging technology has been evaluated as part of machine vision based quality control of a wide range of food products, including meat [28,37,40,42], fruits and vegetables [8,11,20,21,30,31,35,38,48], grain and flour [6,13,14,16,39,47], and animal feed [5,17,34, 36].

Hyperspectral images contain reflectance information for a large number of narrow spectral bands for each pixel (hyperspectral profile), providing a unique reflectance signature of a given object. This allows for classifying material types based on spectral characteristics at specific wavelengths. However, the large number of spectral bands present in hyperspectral images makes the real-time processing for on-line monitoring of animal feed quality difficult, and there are great risks of over-fitting, so that classification of independent test data becomes associated with low accuracy. Thus, it is crucial to develop an efficient procedure to select a few important spectral bands in which it is possible to detect unique contamination signals in order to reliably distinguish contaminants and feed materials.

Spectral band selection is the process of identifying a spectral band subset that contains a significant amount of information to distinguish contaminants and feed materials. Many different methods have been used to assign importance scores to each spectral band, after which a set of spectral bands is chosen based on those scores. Methods in which spectral bands are considered in sets tend to perform better than those which rank them individually, as nearby spectral bands are frequently highly correlated [12,15,49].

We considered two factors that limit many existing approaches. Deterministic recursive methods add one spectral band to the set at a time, and bands are not removed once added. Bands selected later in the process are thus influenced by bands selected earlier in the process, and so we wanted to develop a method where the bands selected for sets with $n + 1$ bands would be independent of those selected

Table 1
Sensor characteristics

Focal length	35 mm
F/#	F 1.4
Bit depth	12 bit
FOV	7 degrees
Scanner	Push boom
Spectral range	Responsive from 400 to 90 nm $\pm$ 5 nm
Spectral resolution	2.1 nm
Spectral channels	240
Cross-track channels	640
Serial communication	Firewire 800 (IEEE 1394b))
Vendor	Pika II, resonon, Inc.

for sets with n bands. Additionally, wrapper methods for band selection (methods that use the classifier itself in the band selection process) are computationally expensive, so we focused on filter methods that do not involve the classifier.

In this paper, we developed two methods: a divergence recursive feature elimination (DRFE) method (essentially a filter method based on the recursive feature elimination method) and an evolutionary refinement (ER) method. A hyperspectral image camera was used to acquire reflectance data in narrow spectral bands across the visible range from 29 feed material samples, including fish and chicken meal, as well as from commercial samples of bonemeal from ground cattle cadavers. Support Vector Machine (SVM) binary classification was used to distinguish between hyperspectral profiles from feed samples and bonemeal samples on the basis of the hyperspectral reflectance values. Classification accuracy was determined when analyses were based upon different feature selection methods. The organization of the paper is as follows. In Section 2, a hyperspectral imaging system for monitoring animal feed quality is described. Section 3 introduces our proposed spectral band selection methods. Experimental results are presented in Section 4, and the conclusion of the paper is presented in Section 5.

2. A hyperspectral imaging system for monitoring animal feed quality

Our hyperspectral imager utilizes a push-broom (line-scan) design, with dispersion provided by a diffraction grating. A frame grabber is not required as scanning is performed with a linear stage controlled by a stepper motor. The sensor is then mounted on an aluminum tower-structure at 35 cm above a 5-cm diameter Petri dish holding the feed materials to collect hyperspectral images at a magnification representing a spatial resolution of about 166 hyperspectral profiles (pixels) per mm^2. Sensor specifics are presented in Table 1.

All hyperspectral images were collected in a dark room, and artificial lighting consisted of 2 $\times$ 3 15 W and 12 V light bulbs mounted in two angled rows – one on either side of the lens. As power source for the lighting, we used a voltage stabilizer (Tripp-Lite, PR-7b, www.radioreference.com). A bright pink piece of paper was placed in the bottom of the Petri dish used to hold feed materials, so that hyperspectral profiles from background were easily separated from feed and bonemeal materials. A piece of white Teflon was used for white calibration and, for each spectral band, reflectance profiles were converted into proportion of the reflection from Teflon (denoted relative reflectance). All hyperspectral images were collected at ambient temperature conditions of 21–23°C and 40–50% relative humidity. Prior to imaging, all feed materials were placed in a single layer. Figure 1 shows some of the challenges of the classification tasks in that bonemeal constitutes a heterogeneous mixture of particles and feed samples are very diverse in their composition.

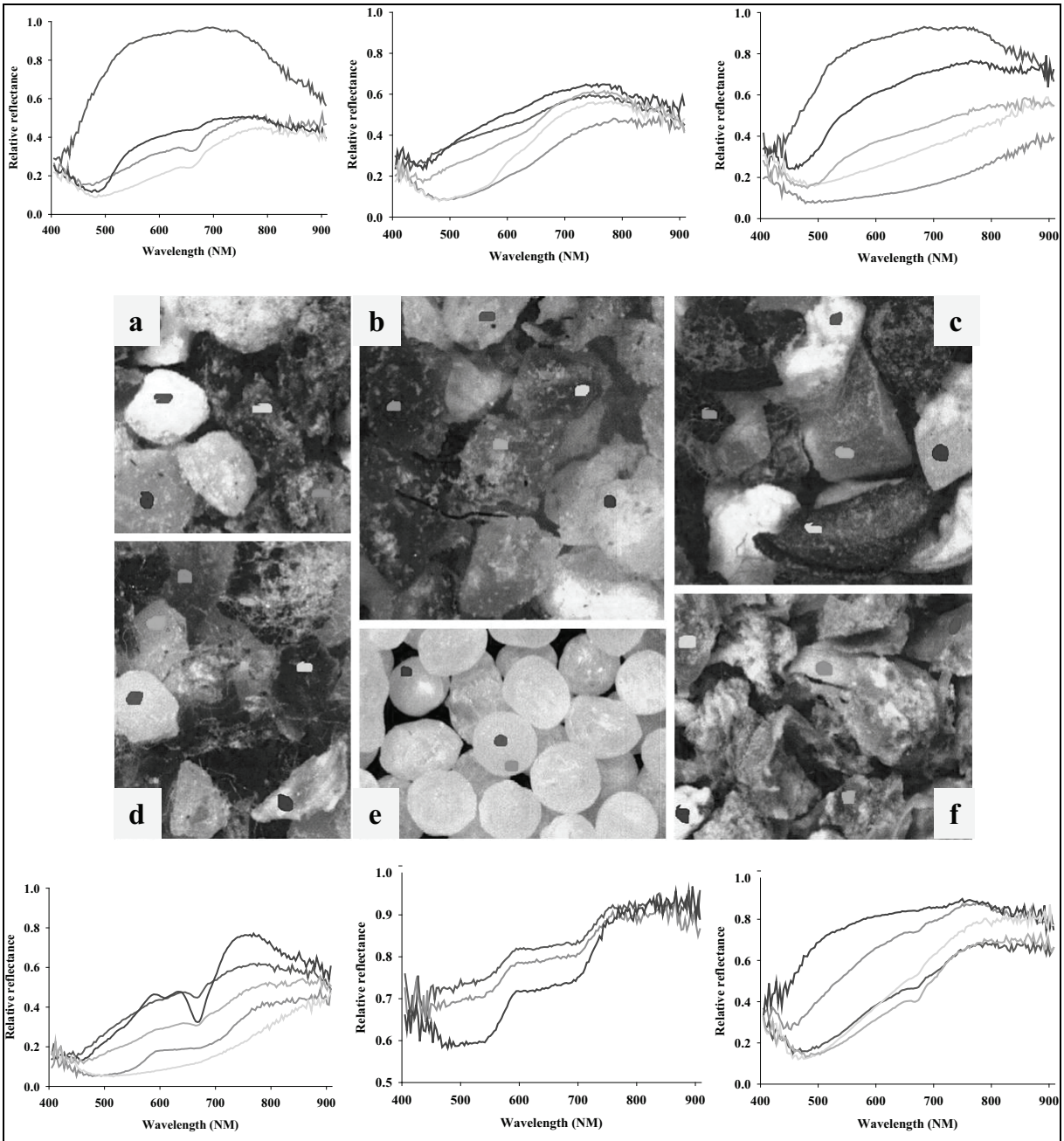

Fig. 1. Reflectance plots on sample particles. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDA-130626)

Our first step after capturing hyperspectral images was to filter out pink hyperspectral profiles that represented the background material on which the feed and bonemeal samples were placed. Next, we separated the data into training and testing sets. Regarding the training data, we selected spectral band sets using both the recursive divergence method and the ER method. The test data was used to evaluate

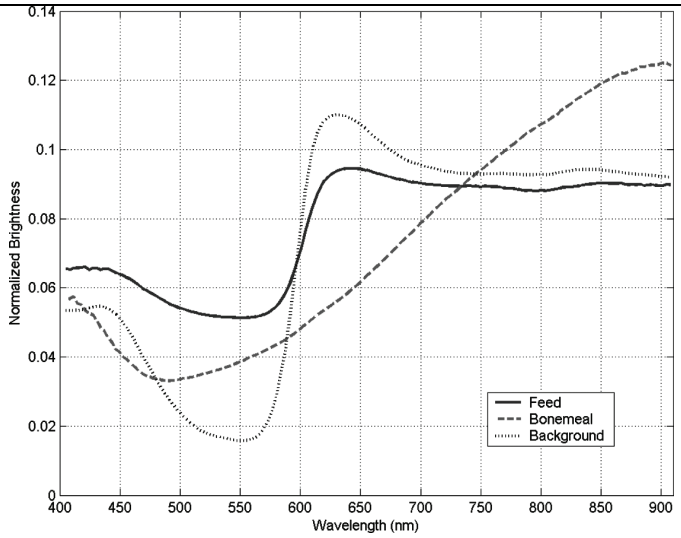

Fig. 2. Example reflectance curves.

the performance of SVM classifiers using those spectral band sets.

In order to automate the filtering of background hyperspectral profiles from our data, we examined the normalized reflectance curves of background hyperspectral profiles and non-background hyperspectral profiles. We identified a distinct curve centered in the visible-light portion of the spectrum (Fig. 2), and we based our filtering method on the values for spectral bands (using 0-based indexing) 19 (465 nm), 44 (544 nm), and 88 (683 nm), which roughly correspond to blue, green, and red light, respectively. We calculated the means and standard deviations of the reflectance values for those three spectral bands in an all-background sample image, after first normalizing the reflectance values for each hyperspectral profile.

To filter background hyperspectral profiles out of our testing and training data, we calculated for each hyperspectral profile the distance (in standard deviations) between the hyperspectral profile's normalized reflectance values for those three spectral bands and the normalized reflectance values from the background sample image. If the sum of those distances was less than a threshold, we identified the hyperspectral profile as background and filtered it out of our training and testing data. We used 10 as the threshold after examining the effects of filtering various test images with various other values. If the sum of the three distances is greater than 10, this means that each band is about more than 3.3 (on average) standard deviations away from the background.

3. Evolutionary refinement approaches for spectral band selection

3.1. Related works

Band selection is similar to the feature selection where we select a few important input variables (features) that are most predictive of a given output. Feature selection can identify only a few relevant features and give a better generalization error [10,22,49,50]. Also, based on the success of SVM, several feature selection algorithms in the SVM domain have been proposed including Guyon's SVM-RFE [22], the SVM gradient method [9,24], the M-fold SVM [18] and FGSVM-RFE [44].

Initially, we looked at three individual spectral band ranking methods: separation measure, envelope eccentricity, and signal-to-noise correlation [4]. Let x^+ and x^- be the average value of a feature x taken over the set of all positive (bonemeal) and negative (feed), respectively, data points. Let x^* be the average of x^+ and x^-, and assume that $x^+ \geqslant x^-$ (otherwise, treat it in a symmetric manner). Let the number of positive observations with $x \geqslant x^*$ be n_x^+, and let the number of negative observations with $x < x^*$ be n_x^-. The separation measure is $\sigma_x = n_x^+ n_x^-$, higher values indicate that the values of this feature x can be used to separate more positive/negative pairs of observations. For the envelope eccentricity, let l_x^+ and u_x^+ be the minimum and maximum of the values of feature x for the positive data points; likewise let l_x^- and u_x^- be the same for the negative data points. The overlap index is the ratio $\omega_x = \frac{\min(u_x^-, u_x^+) - \max(l_x^-, l_x^+)}{\max(u_x^-, u_x^+) - \min(l_x^-, l_x^+)}$, and smaller absolute values for this ratio indicate more relevant features. Finally, the signal-to-noise correlation of a feature x with its class is $\tau_x = \frac{x^+ - x^-}{\sigma^+ + \sigma^-}$, where σ^+ and σ^- are the standard deviations of the values of x for the positive and negative samples, respectively.

Recursive feature elimination takes a different approach. Unlike the above methods, features are not individually given a score and then selected based on the highest ranking scores, but are instead evaluated on how important they are when considered alongside the other features [22]. The result is a list of features ranked from most important to least important. We performed the recursive feature elimination (RFE) process using a linear kernel SVM. A brief explanation of SVM-RFE follows: A linear SVM is trained with all features and a weight is identified for each feature. The feature with the weight with the smallest magnitude is removed; this feature (k) is the least important. The linear SVM is retrained without the feature k. The feature with the weight with the smallest magnitude is eliminated and this feature is the next to the least important. Features are recursively eliminated and ranked until all features are exhausted. In this way, the last eliminated feature is the most important.

Another measure that has been used to evaluate feature sets is the between-class divergence [45]. The divergence of a set of features is related to the correlation between those features and can be used to compare the discriminating power of sets of features. In the following, we briefly introduce the general formula for the between-class divergence and then introduce the divergence when data is normally distributed. Let $\mathbf{x}$ be a set of features and $p_i(\mathbf{x})$ ($p_j(\mathbf{x})$) be the probability density function of $\mathbf{x}$ in class i (j). We select class i if $p_i(\mathbf{x}) > p_j(\mathbf{x})$. So the ratio $\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})}$ or equivalently $ln\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})}$ carries information for the discriminatory capabilities regarding $\mathbf{x}$. The mean value over class i for different values of $\mathbf{x}$ is

$$D_{ij} = \int_{-\infty}^{+\infty} p_i(\mathbf{x}) \, ln\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})} d\mathbf{x} \tag{1}$$

Similar arguments hold to define D_{ji} for class j. The sum $J_{ij} = D_{ij} + D_{ji}$ is known as divergence [45].

Each spectral band of the collected data was normally distributed, so the following equation, which treats the data sets as coming from two multivariate normal distributions, can be used (θ_i and $\mathbf{\Sigma}_i$ are the mean vector and covariance matrix, respectively, for class i) [45]:

$$J_{ij}(\mathbf{x}) = \frac{1}{2} tr\left[\left(\sum\nolimits_i^{-1} + \sum\nolimits_j^{-1}\right)(\theta_i - \theta_j)(\theta_i - \theta_j)^t\right] + \frac{1}{2} tr\left[\left(\sum\nolimits_i - \sum\nolimits_j\right)\left(\sum\nolimits_j^{-1} - \sum\nolimits_i^{-1}\right)\right] \tag{2}$$

Except for when both the number of total spectral bands and desired spectral band set sizes are quite small, it is infeasible to calculate the divergence for every possible set of spectral bands since the total number of such subsets grows exponentially (2^n with n being the number of features). The recursive

divergence (RD) feature selection method instead creates a spectral band set incrementally, so that at any step the number of sets to be evaluated is less than or equal to the total number of spectral bands. The first step in the recursive divergence method is to find the single spectral band with the highest divergence. That spectral band is the only element in the initial feature set. At each additional stage, a spectral band is added to the current feature set such that the divergence of the current feature set plus the new spectral band is maximized [15].

3.2. Proposed methods

We propose two new approaches in spectral band selection. One method is a greedy approach and it uses the divergence as an elimination criterion as explained in Section 3.2.1. The other approach is based on evolutionary search as explained in Section 3.2.2. Both approaches attempt to relieve the heavy burden of an exhaustive search in feature subset selection.

3.2.1. Divergence recursive feature elimination spectral band selection

We developed a *divergence recursive feature elimination (DRFE)* method using a similar principle as SVM-RFE. Unlike SVM-RFE, DRFE does not require a particular machine learning method to evaluate relevance of features. DRFE considers subsets of features during the evaluation process, which is an advantage over those approaches implicitly assuming feature independence. Below is the description of the algorithm in detail. Our data set contains 160 features to be considered and therefore there are 2^{160} subsets to be searched. An exhaustive search is not practical. Divergence has been shown to be highly correlated with the success rate [15]. In our DRFE implementation we used a greedy approach (in backward) using divergence values as an elimination criteria. Our DRFE needs O(n^2) divergence computation compared to O(2^n) of the exhaustive search.

Algorithm 1 A divergence recursive feature elimination method (DRFE)

Let $S = \{1, 2, \ldots, n\}$ where n is the total number of features and L be an empty list.

1. While $|S| > 1$:
2. Let $m = |S|$.
3. For each feature $x \in S$, compute d_x, the between-class divergence for $S - \{x\}$.
4. Let $i = \text{argmax}_{j \in S} d_j$. That is, the subset of S without feature i has the maximum divergence value of all subsets of $m-1$ features implies that feature i contributes the least discriminating information of the features in S.
5. $S = S - \{i\}$.
6. Append feature i to the end of L.
7. After the loop terminates, S contains only one feature. Append that feature to the end of L. L now contains all n features, in order from least important to most important.

3.2.2. Evolutionary spectral band selelection method

We also developed a nondeterministic evolutionary search algorithm for spectral band selection, using between-class divergence as a fitness function. We begin with randomly selected sets of spectral bands and, through a number of successive generations, mutate and recombine the sets with the highest divergence values at a given generation in order to find a set with a very high divergence. The procedure followed for a given generation is explained in Algorithms 2–6. Algorithms 2 and 3 are helper algorithms used in Algorithms 4–6. Algorithm 4 describes elites set selection. Descriptions for roulette wheel selection and tournament selection are given in Algorithms 5 and 6, respectively.

Algorithm 2 The BREED helper method – a roulette-wheel method for selecting k unique features from a list of features L where $|L| > k$

Let the set S contain the distinct features from L.
Let F be a frequency list with each entry F_i being the count of the feature S_i in L divided by $|L|$.
Let N be an empty list.
1. While $|N| < k$
2. Choose a feature f from S using the frequencies in F as selection probabilities.
3. If $f \notin N$, append f to N
4. Return N.

Algorithm 3 The MUTATE_FEATURES helper method – randomly modify the features in a feature set S

Let i be the number of available spectral bands
Let N be an empty set
1. For each feature $x \in S$
2. Do
3. Let $y = x + r$ where $r \sim N(0, 1)$. Then y is rounded to the nearest integer and rolling the result over so that it is in the range of 0 to $i - 1$, inclusive
4. While $y \in N$
5. Append y to N

Algorithm 4 An evolutionary spectral band selection method (ER) using elites set selection, given a cutoff threshold c and n initial feature sets of size k.

1. Compute the between-class divergence for each of the n sets using the training data.
2. Let m be $n * c$, the number of elite sets to carry into the next generation.
3. Let $S_{selected}$ be the m sets with the highest divergence scores, and let S_{next}, the number of sets to consider in the next generation, initially be $S_{selected}$.
4. Let L be an array containing the pooled distinct spectral band numbers from the sets in $S_{selected}$.
5. While $\binom{|L|}{k} < n$
6. Randomly append a feature number not currently in L to L
7. Let n_{mutate} the number of feature sets to create by mutation, be $\frac{n-m}{2}$. Let $n_{recombine}$, the number of feature sets to create by recombining the pooled features of the m selected sets, be $n - m - n_{mutate}$.
8. Create n_{mutate} new sets by applying MUTATE_FEATURES to randomly selected sets from $S_{selected}$ and append these sets to S_{next}.
9. Create $n_{recombine}$ new sets using roulette-wheel selection by using the BREED method to select sets of k features from L and append these sets to S_{next}.

The while loop in Algorithm 4 is included for cases with small set sizes and large numbers of sets, where otherwise as the feature sets converge the algorithm may end up with too few distinct features to produce enough distinct feature sets for the next generation.

Because this algorithm is nondeterministic, we performed multiple trials of the evolutionary spectral band set search and looked at the mean and standard deviations of the between-class divergence and the balanced success rate from the SVM classifier of the sets selected for each spectral band set size n.

With a crossover operation for two parents, a point in the string is chosen as the crossover point, and all the elements before that point from one parent are concatenated with the elements after that point from the other parent. We are looking for sets of distinct features, so even when keeping them in sorted order, there is a chance that feature k could be found in both parents. For example, feature k could be found both before the crossover point in the left parent and after the crossover point in the right parent.

Roulette-wheel and tournament methods for set selection were tested as well as the elite selection. We tested these methods both with the same recombination and mutation steps from Algorithm 4, where the only difference was that the m sets $S_{selected}$ were selected in a randomized manner using one of the two following methods, instead of with elites selection.

Algorithm 5 A roulette-wheel feature set selection method to select m sets from a set S of n sets

Let F be a frequency array where each entry F_i is the divergence score of set S_i divided by the sum of all the sets' divergence scores.
Let S' be an empty list.
1. While $|S| < m$
2. Let $S1$ be a feature set selected from S using F as the selection probabilities for the sets in S.
3. With probability p_direct, let $S0$ be $S1$, otherwise:
4. Select a feature set $S2$ from S using F.
5. Let $S0$ be the result of BREED $(S1 \cup S2)$.
6. With probability p_mutate, apply MUTATE_FEATURES to $S0$.
7. Append $S0$ to S'
8. Return S'

Algorithm 6 A tournament feature set selection method to select m sets from a set S of n sets

Let S' be an empty list.
1. While $|S''| < m$
2. Let $S1$ be a feature set selected from S by selecting a random subset of sets from S and choosing the set in that subset with the highest divergence score.
3. With probability p_direct, let $S0$ be $S1$, otherwise:
4. Select a feature set $S2$ from S using the same tournament selection as for $S1$.
5. Let $S0$ be the result of BREED $(S1 \cup S2)$.
6. With probability p_mutate, apply MUTATE_FEATURES to $S0$.
7. Append $S0$ to S'
8. Return S'

3.2.3. Parameter selection

Both evolutionary search and SVM training requires careful selection of parameters. Parameter selection for evolutionary search is described in Section 3.3.1 and that for SVM training is described in Section 3.3.2.

3.2.4. Evolutionary search algorithm parameter selection

The parameters to the evolutionary search algorithm are the number of sets to consider at each generation, which is determined by the number of start sets given; the number of generations for which to run; the sigma value to use when mutating candidate sets; and the cutoff value, which controls what fraction of the sets to carry forward at each generation. We tested these for various values in order to determine

Table 2
Divergence values of sets found using different candidate set cutoff thresholds (# sets = 40, # generations = 40, # trials = 10)

	Cutoff = 0.125		Cutoff = 0.250		Cutoff = 0.500	
Set size	Mean div	Std dev div	Mean div	Std dev div	Mean div	Std dev div
2	4.832	0	4.719	0.359	4.489	0.553
4	8.194	0.578	8.216	0.571	8.149	0.343
8	15.504	0.329	15.472	0.219	14.547	0.330
12	19.531	0.407	19.435	0.304	18.358	0.409
16	23.531	0.154	23.028	0.402	21.912	0.460
24	30.160	0.565	29.687	0.493	28.371	0.408
32	35.854	0.362	35.39	0.763	33.775	0.650
40	41.143	0.554	40.162	0.601	38.552	0.636

Table 3
Divergence values of sets found using different numbers of generations (# sets = 40, cutoff = 0.25, # trials = 10)

	20 generations		40 generations		80 generations		160 generations	
Set size	Mean div	Std dev div	Mean div	Std dev div	Mean div	Std dev div	Mean div	Std dev div
2	4.643	0.442	4.719	0.359	4.605	0.479	4.832	0
4	7.758	0.417	8.216	0.571	8.127	0.554	8.168	0.674
8	14.864	0.337	15.472	0.219	15.628	0.242	15.579	0.168
12	18.878	0.339	19.435	0.304	19.734	0.329	19.981	0.250
16	22.232	0.410	23.028	0.402	23.691	0.262	23.665	0.221
24	28.485	0.466	29.687	0.493	30.445	0.430	30.986	0.198
32	33.901	0.647	35.390	0.763	36.540	0.335	37.062	0.212
40	38.900	0.570	40.162	0.601	41.544	0.443	41.964	0.294

Table 4
Divergence values of sets found using different numbers of sets per generation (# generations = 40, cutoff = 0.25, # trials = 10)

	20 sets		40 sets		80 sets		160 sets	
Set size	Mean div	Std dev div	Mean div	Std dev div	Mean div	Std dev div	Mean div	Std dev div
2	4.579	0.534	4.719	0.359	4.719	0.359	4.832	0
4	7.553	0.729	8.216	0.571	8.540	0.402	8.737	0
8	15.111	0.291	15.472	0.219	15.510	0.153	15.658	0.181
12	19.119	0.435	19.435	0.304	19.517	0.274	19.595	0.219
16	22.863	0.370	23.028	0.402	23.300	0.259	23.238	0.176
24	29.559	0.517	29.687	0.493	29.768	0.383	29.839	0.229
32	35.181	0.513	35.390	0.763	35.320	0.313	35.385	0.278
40	39.705	1.118	40.162	0.601	40.104	0.347	40.303	0.262

what effects the various parameters had on the search performance and to select the parameters we would use when comparing this method to other methods (results in Tables 2–4 and 6).

The number of generations most directly affected the magnitude of the divergence values of the larger sets found by the search. The utility of the additional generations appears to be related to the size of the search space (for example, for sets with 32 spectral bands the number of possible sets is $\binom{160}{32} = 4.646 \times 10^{33}$, while for sets with six spectral bands it is $\binom{160}{6} = 2.119 \times 10^{10}$).

Conversely, the number of sets considered was a primary factor in the consistency of the divergence values of the sets found by the search. This was most true for small sets; the effect was less pronounced for larger sets, again presumably because of how rapidly the search space grows as the set size increases. Increasing the number of sets improves the odds that any particular feature will be considered in at least one of the randomly-selected start sets. When independently drawing, without replacement, c sets of

size s from n total features, the probability of any given feature occurring in at least 1 set is:

$$1 - \left(\frac{n-s}{n}\right)^c \tag{3}$$

For instance, when drawing 20 sets of size 10 from 160 total features, the probability of any given feature appearing at least once in those sets is 0.725 (so we would expect roughly 116 distinct features to appear in the 20 sets). When we increase the number of sets to 40, the probability rises to 0.924 (and roughly 148 expected distinct features in the 40 sets).

Changing the cutoff parameter produced results similar to changing the number of generations. Smaller cutoff values produced sets with, on average, higher between-class divergence values. Table 2 shows that cutoff value $= 0.125$ yielded about 6% higher between-class divergence values than that by cutoff value $= 0.5$. The effect of this parameter is closely related to the effects of both of the above parameters: a smaller cutoff means fewer sets are carried forward from the previous generation, meaning more new sets are generated in each generation, similar to the additional sets that are generated and considered when the number of generations is increased; on the other hand, the new sets are generated based on a smaller number of sets, so the search of the scope narrows more quickly, possibly discarding sets that did not perform quite as well but are "close" to sets with higher divergence values than the sets found by focusing on the smaller range of selected sets.

After comparing the results of several different combinations of parameters, we chose the following parameters: generations $= 160$, number of sets $= 80$, cutoff $= 0.125$. These parameters were selected to provide high yet consistent divergence values even for larger spectral band sets. With these parameters, the number of divergence calculations that are performed to find a spectral band set of a given size is $80 + (160 - 1)(80 * (1 - 0.125)) = 11{,}210$. By comparison, the number of divergence calculations required to create the feature ranking of all 160 features using the divergence-based elimination method is $\sum_{i=2}^{160} i = 12{,}879$. For small sets, the 11,210 divergence calculations performed on a small number of spectral bands can be computed much more quickly than the calculations involved in analyzing 160 sets of size 159,159 sets of size $158, \ldots,$ 2 sets of size 1, but as the set size increases the advantage decreases, so these search parameters can also be used to establish an upper limit on the set size for which it is advantageous to use this method. For smaller sets, the number of sets and generations could probably be comfortably decreased to 40 sets and 80 generations to speed things up without having much effect on the performance of the search; the results for different parameters shown in Tables 2 and 3 suggests that this would have little effect for small sets other than a slight decrease in consistency of the divergence values of the resulting sets.

3.2.5. SVM parameter selection

SVM [32,41,43,46,52] has been extensively used on various data domains including hyperspectral data [15,20,51], due to its high generalization performance in various application domains. We have to optimize two parameters of the Gaussian kernel SVM: C and sigma (σ). One of popular approaches is to separate the data set into two parts: training data and testing data and then apply k-fold cross-validation for the training data to select the parameters of SVM. That is, we first divide the training set into k subsets of equal size and each subset is tested sequentially using the classifier trained on the remaining $k-1$ subsets to prevent the overfitting problem [23,25,26,29,33]. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. In this paper, we used 5-fold cross-validation for the parameter selection.

Table 5
Test dataset break down

	Original profiles	Background	Net profiles
Feed (training data)	5,760	59	5,701
Bonemeal (training data)	5,681	160	5,521
Feed (test data)	499,200	24,887	474,313
Bonemeal (test data)	500,000	5,017	494,983

Table 6
Divergence values and balanced success rates (BSR) for sets found with evolutionary divergence-based search (# sets = 80, # generations = 160, cutoff = 0.125, # trials = 27)

Set size	Mean bsr	Sd bsr	Mean div	Sd div
2	0.741	0	4.832	0
3	0.785	0.033	5.862	0.171
4	0.819	0.028	8.334	0.480
5	0.874	0.006	11.977	0.654
6	0.876	0.011	13.313	0.319
7	0.886	0.002	14.499	0.108
8	0.887	0.008	15.730	0.232
9	0.893	0.004	16.945	0.259
10	0.895	0.006	18.049	0.224
11	0.899	0.006	19.033	0.217
12	0.900	0.005	20.092	0.165
14	0.906	0.004	21.994	0.094
16	0.910	0.005	23.880	0.182
18	0.912	0.003	25.829	0.174
20	0.917	0.002	27.556	0.212
24	0.924	0.003	31.041	0.131
32	0.930	0.002	37.081	0.208
40	0.933	0.001	42.247	0.235
48	0.935	0.001	47.163	0.216
64	0.936	0.001	55.946	0.247

4. Experimental results

We evaluated the performance of our proposed methods in two ways: by comparing the divergences of the sets it found to the divergences of the sets selected by the recursive divergence methods, and by comparing the balanced success rate for the SVM classifier when using those sets with the rates from the sets produced by other methods. Because of the nondeterministic nature of the evolutionary divergence search, we ran it multiple times to minimize the effects of variance on the results, with the values in the graphs being the mean values for 20 trials (see Tables 2–4).

Our training data contained 5760 feed hyperspectral profiles and 5681 bonemeal hyperspectral profiles. 59 of the feed hyperspectral profiles were identified as background, as were 160 of the bonemeal hyperspectral profiles, leaving 5701 feed training hyperspectral profiles and 5521 bonemeal training hyperspectral profiles. Our testing data contained 499,200 feed hyperspectral profiles and 500,000 bonemeal hyperspectral profiles. 24,887 and 5,017 feed and bonemeal hyperspectral profiles, respectively, were identified as background, leaving 474,313 feed training hyperspectral profiles and 494,983 bonemeal training hyperspectral profiles. Table 5 summarizes this breakdown of the data.

Figure 3 shows a rough RGB conversion of a sample hyperspectral image and a bitmap mask where the white hyperspectral profiles are the hyperspectral profiles identified as background. The RGB conversion was done by simply selecting the spectral bands corresponding to red, green, and blue light, as discussed in Section 2, and scaling the reflectance values to the 8-bit 0–255 range.

Table 7
Divergence values and balanced success rates (BSR) for sets found with other search methods

	RD		DRFE		RFE	
Set size	BSR	Divergence	BSR	Divergence	BSR	Divergence
2	0.741	4.832	0.717	4.819	0.722	2.234
3	0.759	5.612	0.737	5.492	0.843	4.568
4	0.776	6.390	0.780	8.234	0.845	5.908
5	0.803	7.181	0.874	11.387	0.836	8.321
6	0.865	8.948	0.876	12.295	0.881	11.693
7	0.875	10.901	0.881	13.960	0.886	12.920
8	0.899	13.984	0.883	15.138	0.888	13.234
9	0.900	15.136	0.888	16.567	0.891	13.965
10	0.905	16.288	0.887	17.599	0.891	14.351
11	0.909	17.429	0.890	18.875	0.901	15.286
12	0.907	18.551	0.894	19.994	0.901	16.288
14	0.916	21.189	0.900	22.084	0.905	18.041
16	0.918	23.195	0.905	24.045	0.906	18.666
18	0.921	25.026	0.912	26.036	0.907	20.155
20	0.921	26.912	0.917	27.829	0.909	21.460
24	0.925	30.567	0.923	31.393	0.914	23.443
32	0.930	36.988	0.928	37.429	0.919	26.950
40	0.930	42.650	0.932	42.482	0.923	30.693
48	0.935	47.405	0.935	47.406	0.924	34.087
64	0.935	56.329	0.936	56.448	0.927	41.765

Our tests confirmed that between-class divergence was a fairly good indicator of SVM classifier performance for our data. From Table 7, an average correlation coefficient between SVM success rates and divergence is 0.75 when the set size is greater than 12. The recursive divergence selection method produced sets that performed much better than those created from the individual feature rankings; sets with 8 or more features found using this method even had higher success rates than those found with SVM-RFE. This makes divergence well suited for use as the fitness function for our feature set search.

4.1. Comparison of the divergences

Figure 4 focuses only on sets with 12 or fewer spectral bands, as that is where both ER and DRFE methods had the biggest advantage. For small sets, both of those methods found sets with significantly higher between-class divergence values than the normal recursive divergence method excepting sets of size 2, in which case the RD method found the same set as the evolutionary search method – for the given search parameters, the ER method found the same set in all 20 trials – and the DRFE method found a set with slightly lower divergence. Also of note is that the divergence values of the sets found by ER tended to have extremely similar divergence values as those found by DRFE. Tables 6 and 7 contain the results of the four tested methods for a range of set sizes between 2 and 64.

The RD method did not start finding sets with higher divergence values than the average set found by the ER method until set size became 40 features. The DRFE method had a slight advantage over both methods starting with sets with 14 features and up, though at set size 40 the RD method slightly outdid it.

The advantage in divergence values for ER compared to the RD method was consistent across the 20 trials. For sets with 4 to 24 spectral bands, the difference between the mean divergence value for the sets found by ER and the divergence value of the set found by the RD method was at least 3 times the standard deviation of the divergence value for the sets found by ER – for sets with 6 and 7 spectral bands

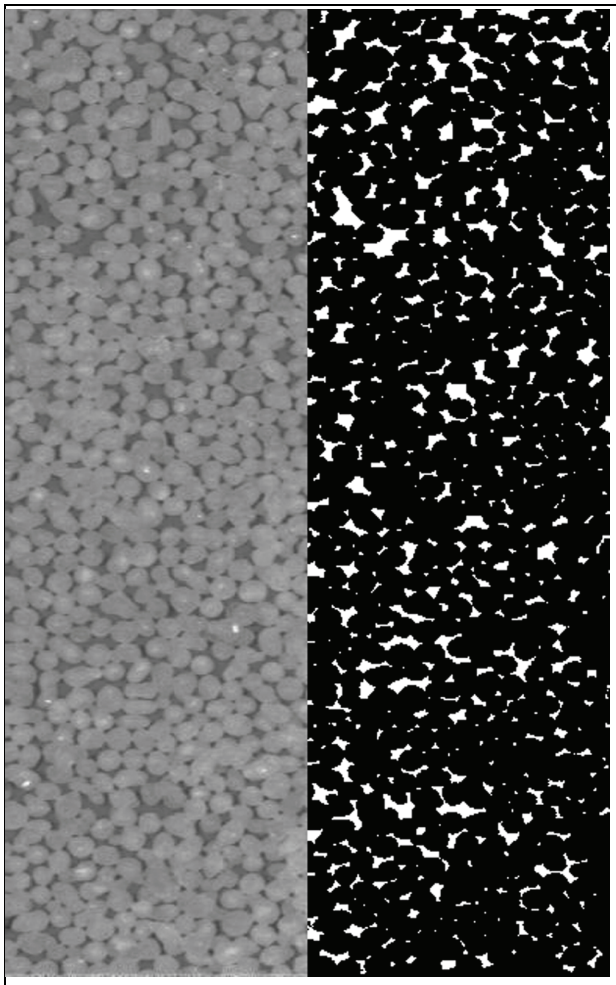

Fig. 3. Sample hyperspectral image (L) and background mask (R).

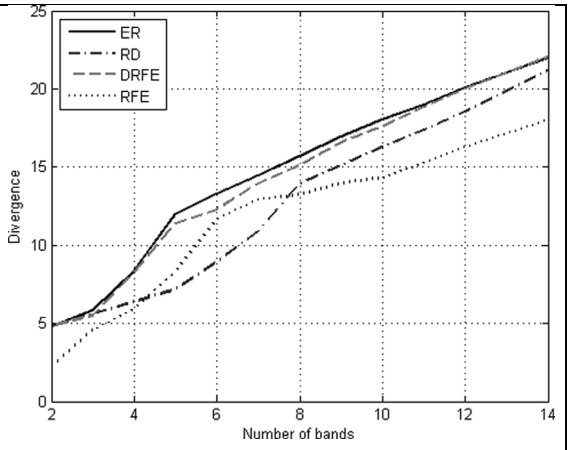

Fig. 4. Divergence vs. spectral band set size for small band sets.

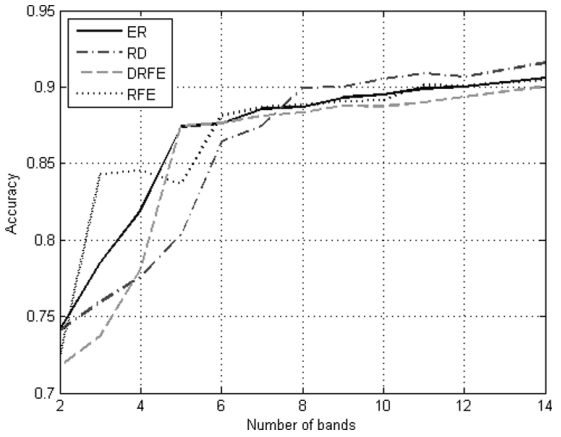

Fig. 5. SVM success rate vs. spectral band set size (small sets).

the difference was over 10 times the standard deviation. For sets with three spectral bands, the difference was 1.5 times the standard deviation.

Also, Table 8 contains results from using roulette-wheel and tournament methods for set selection. Neither of them outperformed the elites selection method.

4.2. Comparison of the balanced success rate

Figure 5 shows that for small sets, the divergence advantages of the ER and DRFE methods also translated into an SVM success rate advantage, although the success rates of the sets found by those two methods were not nearly as similar as the divergence values for sets with just 2 and 3 features. For sets with more than 2 features and fewer than 8 features, the sets selected by the SVM-RFE method also had noticeably higher SVM balanced success rates than those selected by the RD method. For 8–12 features, the sets selected by the RD method outperformed those selected by our methods; however, as the divergence values for those sets were still lower than those for the sets selected by our methods, we

Table 8
Contrast the recombine vs. traditional implementation

	Roulette wheel			Tournament	
n	Mean div	SD div	n	Mean div	SD div
3	5.839	0.318	3	5.962	0.201
4	8.074	0.769	4	8.245	0.553
5	11.71	0.908	5	11.811	0.823
8	15.264	0.204	8	15.751	0.246
12	19.153	0.323	12	20.039	0.179
20	25.529	0.456	20	27.636	0.158
32	33.985	0.501	32	37.009	0.386
48	42.024	0.543	48	46.757	0.395

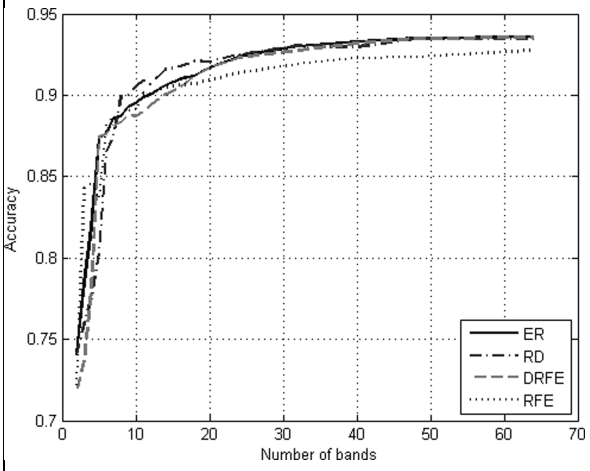

Fig. 6. SVM success rate vs. spectral band set size (all sets).

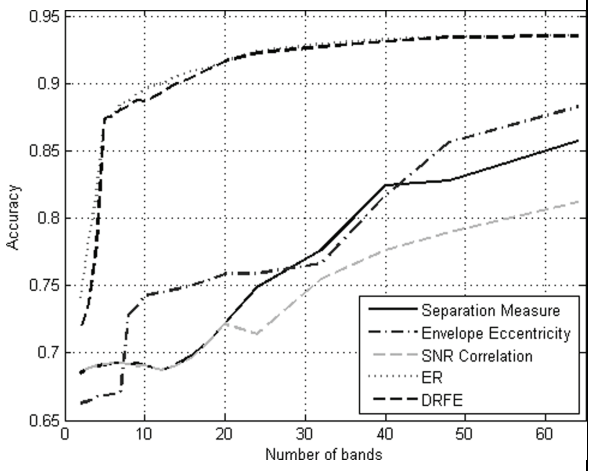

Fig. 7. SVM success rate vs. spectral band set size.

believe this to be mainly the result of chance and not representative of an inherent advantage of the RD method.

As set size increased, the success rates converged – particularly for the divergence-based methods, as shown in Fig. 6. The sets from the recursive divergence (RD), divergence-based elimination (DRFE), and evolutionary divergence (ER) search methods resulting in very close success rates for sets with 20 or more features, with declining gains from adding more features. For instance, for the evolutionary search method going from 2 to 5 features caused a 0.133 increase in balanced success rate, while going from 5 to 64 features only caused a 0.062 increase, and from 12 to 64 only caused a 0.036 increase. The sets found by the RD and DRFE methods followed almost the same curve; the sets found by the SVM-RFE method also followed a similar curve, but with success rates that were consistently about 0.01 lower for sets with 20 or more spectral bands. Additionally, the standard deviation of the balanced success rates for the sets found by the ER method declined consistently as larger spectral band sets were considered, while the standard deviation of the divergence values for those sets did not – there is a correlation between divergence and SVM success rate, and set size itself appears to have more influence on the consistency of the success rates than the consistency of the divergence values of the sets, as seen in Table 7.

Also note that for 48 features, the subsets selected by the RD and DRFE methods outperformed the set of all features, and for 64 features, the subsets selected by those two methods as well as the one selected by the ER method did so as well.

Figure 7 shows the success rates of the ER and DRFE methods compared to the separation measure, envelope eccentricity, and signal-to-noise correlation rankings. As mentioned before, nearby spectral are often highly correlated, so methods that look at combinations of spectral bands instead of individual spectral bands have a significant advantage; this is reflected in the success rates of the sets found by the individual spectral band feature rankings.

5. Conclusion

There is an increasing demand for reliable and sensitive machine vision systems to be used in quality control of feed and food products due to growing public concerns about food safety. Because feed materials are highly diverse in composition, they can be considered a challenging model system for detection of contaminants. This paper presents methods for spectral band selection in hyperspectral image (HSI) cubes, containing spectral reflectance in both spatial dimensions and spectral bands. A hyperspectral image camera was used to acquire reflectance data in narrow spectral bands across the visible range from 29 feed material samples. Support Vector Machine binary classification was used to distinguish between hyperspectral profiles from feed samples and bonemeal samples on the basis of the hyperspectral reflectance values. Classification accuracy was determined when analyses were based upon different feature selection methods. Two new methods were developed utilizing divergence values for selecting spectral band sets, 1) divergence-based recursive feature elimination (DRFE) approach, and 2) evolutionary search (ER) method. The ER algorithm and DRFE methods provide a significant advantage for selecting small spectral band sets over the recursive divergence (RD) method other than for two spectral bands, in which case ER never found a set with higher divergence than the one selected by the RD method, and the DRFE method found one with slightly lower divergence. The advantage of the evolutionary spectral band set search decreases as the set size increases, with the recursive divergence method reclaiming the advantage for most cases when 40 spectral bands are considered, likely due to the increasing size of the search space. Future work will explore ways to determine how many features would be best based on various feature rankings.

References

[1] Council decision of 4 December 2000 concerning certain protection measures with regard to transmissible spongiform encephalopathies and the feeding of animal protein, *Official Journal of the European Communities* **L306** (2000), 32–33.
[2] Food and drug administration, fourth draft: framework of the fda animal feed safety system, available from http://www.fda.gov/AnimalVeterinary/SafetyHealth/AnimalFeedSafetySystemAFSS/ucm196795.htm, (8 January 2010), (accessed 16 March 2011).
[3] Regulation (EC) No 1774/2002 of the European parliament and of the council of 3 october 2002 laying down health rules concerning animal by-products not intended for human consumption, *Official Journal of the European Communities* **L273** (2002), 1–95.
[4] G. Alexe, S. Alexe, P. Hammer and B. Vizvari, Pattern-based feature selection in genomics and proteomics, *Annals of Operations Research* **148** (2006), 189–201.
[5] V. Baeten, A.M. Renier, G. Sinnaeve and P. Dardenne, Analyses of feeding stuffs by near-infra red microscopy (NIRM): Detection and quantification of meat and bonemeal meal (MBM), in: *Proceed Sixth Food Authenticity and Safety Inter Symp (FASIS)*, Nantes, France, (2001), 1–11.
[6] T. Baye and H. Becker, Analyzing seed weight, fatty acid composition, oil, and protein contents in Vernonia galamensis germplasm by near-infrared reflectance spectroscopy, *Journal of the American Oil Chemists' Society* **81** (2004), 641–45.
[7] Centers for disease control and prevention, BSE in an Alabama cow, available from http://www.cdc.gov/ncidod/dvrd/bse/news/alabama_cow_031506.htm, (26 August 2010), (accessed 14 March 2011).

[8] Y.R. Chen, S.R. Delwiche and W.R. Hruschka, Classification of hard red wheat by feedforward backpropagation neural networks, *Cereal Chemistry* **72** (1995), 317–19.
[9] H. Cho, S. Baek, E. Youn, M. Jeong and A. Taylor, A two stage classification procedure for near-infrared spectra data based on multi-scale vertical energy wavelet thresholding and SVM-based gradient-RFE feature elimination, *Journal of the Operational Research Society* **60** (2009), 1107–1115.
[10] H. Cho, S. Kim, M. Jeong, Y. Park, T. Ziegler and D. Jones, Genetic algorithm-based feature selection in high-resolution NMR spectra, *Expert Systems with Applications* **35** (2008), 967–975.
[11] S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis and J. Blasco, Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables, *Food and Bioprocess Technology* (2010), doi: 10.1007/s11947-010-0411-8.
[12] S. De Backer, P. Kempeneers, W. Debruyn and P. Scheunders, A band selection technique for spectral classification, *IEEE Geoscience and Remote Sensing Letters* **2** (2005), 319–323.
[13] S.R. Delwiche, S.K. Moon and Y. Dong, Damage and quality assessment in wheat by NIR hyperspectral imaging, in: *Proc SPIE 7676, 767607* (2010), doi:10.1117/12.851150.
[14] S.R. Delwiche and W.R. Hruschka, Protein content of bulk wheat from near-infrared reflectance of individual kernels, *Cereal Chemistry* **77** (2000), 86–88.
[15] Z. Du, M.K. Jeong and S.G. Kong, Band selection of hyperspectral images for automatic detection of poultry skin tumors, *IEEE Transactions on Automation Science and Engineering* **4** (2007), 332–339.
[16] O.A. El-Sebai, R. Sanderson, M.P. Bleiweiss and N. Schmidt, Detection of Sitotroga cerealella (Olivier) infestation of wheat kernels using hyperspectral reflectance, *Journal of Entomological Science* **41** (2006), 155–64.
[17] F.A.F. Pierna, V. Baeten, A.M. Renier, R.P. Cogdill and P. Dardenne, Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bonemeal meal (MBM) in compound feeds, *Journal of Chemometrics* **18** (2004), 341–349.
[18] L. Fu and E. Youn, Improving reliability of gene selection from microarray functional genomics data, *IEEE Transactions on Information Technology in Biomedicine* **7** (2003), 191–196.
[19] O. Fumière, P. Veys, A. Boix, C. von Holst, V. Baeten and B. Berben, Methods of detection, species identification and quantification of processed animal proteins in feedingstuff, *Biotechnology, Agronomy, Society and Environment* **13** (2009), 59–70.
[20] J. Gomez, J. Blasco, E. Molto and G. Camps-Vails, Hyperspectral detection of citrus damage with Mahalanobis kernel classifier, *Electronics Letters* **43** (2007), 1082–1084.
[21] J. Gómez-Sanchis, L. Gómez-Chova, N. Aleixos, G. Camps-Valls, C. Montesino-Herrero, E. Moltó and J. Blasco, Hyperspectral system for early detection of rottenness caused by penicillium digitatum in mandarins, *Journal of Food Engineering* **89**(1) (2008), 80–86.
[22] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine Learning* **46** (2002), 389–422.
[23] T. Hastie, S. Rosset, R. Tibshirani and J. Zhu, The entire regularization path for the support vector machine, *Journal of Machine Learning Research* **5** (2004), 1391–1415.
[24] L. Hermes and J. Buhmann, Feature selection for support vector machines, in: *Proceedings of the International Conference on Pattern Recognition (ICPR'00)* **2** (2000), 712–715.
[25] C.W. Hsu, C.C. Chang and C.J. Lin, A practical guide to support vector classification, technical report department of computer science and information engineering, National Taiwan University, 2003. http://www.csie.ntu.edu.tw/_cjlin/libsvm/.
[26] C.L. Huang and C.J. Wang, A GA-based feature selection and parameters optimization for support vector machines, *Expert Systems with Applications* **31** (2006), 231–240.
[27] C.M. Kahn and S. Line, The Merck Veterinary Manual, 9th ed., Whitehouse Station, NJ: Merck, 2005.
[28] S.G. Kong, Y. Chen, I. Kim and M.S. Kim, Analysis of hyperspectral-fluorescence images for poultry skin tumor inspection, *Applied Optics* **43** (2007), 824–833.
[29] B.C. Kuo and K.Y. Chang, Feature extractions for small sample size classification problem, *IEEE Transactions on Geoscience and Remote Sensing* **45** (2007), 756–764.
[30] D.J. Lee and J.K. Archibald, Color image processing for date quality evaluation, in: *Proc SPIE 7539*, 75390V (2010), doi:10.1117/12.841135.
[31] A.M. Lefcout and M.S. Kim, Technique for normalizing intensity histograms of images when the approximate size of the target is known: Detection of feces on apples using fluorescence imaging, *Computers and Electronics in Agriculture* **50** (2006), 135–147.
[32] K. Lin and M. Chen, On the Design and Analysis of the Privacy-Preserving SVM Classifier, *IEEE Transactions on Knowledge and Data Engineering* (2010), p. 99 doi: 10.1109/TKDE.2010.193.
[33] S.W. Lin, Z.J. Lee, S.C. Chen and T.Y. Tseng, Parameter determination of support vector machine and feature selection using simulated annealing approach, *Applied Soft Computing* **8** (2008), 1505–1512.
[34] Y. Liu, K. Chao, M.S. Kim, D. Tuschel, O. Olkhovyk and R.J. Priore, Potential of raman spectroscopy and imaging

methods for rapid and routine screening of the presence of melamine in animal feed and foods, *Applied Spectroscopy* **63** (2009), 477–480.

[35] P.M. Mehl, Y. Chen, M.S. Kim and D.E. Chan, Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations, *Journal of Food Engineering* **61** (2004), 67–81.

[36] C. Nansen, T. Herrman and R. Swanson, Machine vision detection of bonemeal in animal feed samples, *Applied Spectroscopy* **64**(6) (2010).

[37] B. Park, S.-C. Yoon, W.R. Windham, K.C. Lawrence, G.W. Heitschmidt, M.S. Kim and K. Chao, Line-scan hyperspectral imaging for real-time poultry fecal detection, in: *Proc SPIE 7676* 7676I, (2010), doi: 10.1117/12.850258.

[38] Y. Peng and R. Lu, Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content, *Postharvest Biology and Technology* **48** (2008), 52–62.

[39] J. Perez-Mendoza, J.E. Throne, F.E. Dowell and J.E. Baker, Detection of insect fragments in wheat flour by near-infrared spectroscopy, *Journal of Stored Products Research* **39** (2003), 305–312.

[40] J. Qiao, M.O. Ngadi, N. Wang, C. Gariépy and S.O. Prasher, Pork quality and marbling level assessment using a hyperspectral imaging system, *Journal of Food Engineering* **83** (2007), 10–16.

[41] M. Rojas, I. Dopido, A. Plaza and P. Gamba, Comparison of support vector machine-based processing chains for hyperspectral image classification, in: *Proc SPIE 7810* 78100B (2010), doi:10.1117/12.860413.

[42] C.B. Singh, D.S. Jayas, J. Paliwal and N.D.G. White, Near-infrared hyperspectral imaging for quality analysis of agricultural and food products, in: *Proc SPIE 7676*, 767603 (2010), doi:10.1117/12.850371.

[43] Z. Sun, H. Guo, X. Li, L. Lu and X. Du, Estimating urban impervious surfaces from Landsat-5 TM imagery using multilayer perceptron neural network and support vector machine, *J Appl Remote Sens* **5** (2011), doi:10.1117/1.3557816.

[44] Y. Tang, Y. Zhang, Z. Huang, X. Hu and Y. Zhao, Recursive fuzzy granulation for gene subsets extraction and cancer classification, *IEEE Transactions on Information Technology in Biomedicine* **12** (2008), 723–730.

[45] S. Theodoridis and K. Koutroumbas, Pattern Recognition, Third Edition. San Diego: Academic Press, 2006.

[46] V.N. Vapnik, Statistical learning theory, New York: Wiley-Interscience, 1998.

[47] D. Wang, F.E. Dowell and D.S. Chung, Assessment of heat-damaged wheat kernels using near-infrared spectroscopy, *Cereal Chemistry* **78** (2001), 625–628.

[48] S.C. Yoon, B. Park, K.C. Lawrence, W.R. Windham and G.W. Heitschmidt, Development of real-time line-scan hyperspectral imaging system for online agricultural and food product inspection, in: *Proc SPIE 7676*, 76760J (2010), doi:10.1117/12.850460.

[49] E. Youn and M. Jeong, Class dependent feature scaling method using naïve Bayes classifier for text datamining, *Pattern Recognition Letters* **30** (2009), 477–485.

[50] E. Youn, L. Koenig, M. Jeong and S. Baek, Support vector based feature selection using Fisher's linear discriminant and support vector machine, *Expert Systems with Applications* **37** (2010), 6148–6156.

[51] R. Zhang and J. Ma, Feature selection for hyperspectral data based on recursive support vector machines, *International Journal of Remote Sensing* **30** (2009), 3669–3677.

[52] M. Zhou, J. Shu and Z. Chen, Classification of hyperspectral remote sensing image based on genetic algorithm and SVM, in: *Proc SPIE 7809*, 78090A (2010), doi:10.1117/12.860153.