Hyperspectral Ratio Feature Selection: Agricultural Product Inspection Example

Songyot Nakariyakul and David Casasent

Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213

ABSTRACT

We describe a fast method for dimensionality reduction and feature selection of ratio features for classification in hyperspectral data. The case study chosen is to discriminate internally damaged almond nuts from normal ones. For this case study, we find that using the ratios of the responses in several wavebands provides better features than a subset of waveband responses. We find that use of the Euclidean Minimum Distance metric gives slightly better results than the more conventional Spectral Angle Mapper distance metric in a nearest neighbor classifier.
Keywords: feature selection, hyperspectral data, product inspection.

1. INTRODUCTION

Hyperspectral (HS) data has been shown to be useful for the detection of objects such as military vehicles [1-3] and mines [3, 4], for land use applications [5], and for many USDA product inspection applications [6-18]. HS data has been shown to be successful in detecting skin tumors on chicken carcasses [6-8], fecal and ingesta contaminants on chicken carcasses [9], aflatoxin levels in corn kernels [10, 11, 13, 14], and internal-damage in almonds [12-14], and in determining wheat grain quality [15, 16], moisture in soybean seeds [17], and sugar content in potatoes [18]. These uses occur, since HS data provides spectral information that uniquely characterizes and identifies the chemical, moisture, and physical properties of the constituent parts of an input object, scene region, or an agricultural product.

HS data is high-dimensional data that contains more than a hundred responses in narrowly spaced spectral bands (λ). High-dimensional data generally does not have enough training samples. It is generally accepted that the number of training samples per class must be at least ten times the number of features (in this case, the number of input λ spectral samples) if one is to accurately predict the class of an unknown test sample [19]. This is known as *the curse of dimensionality*. If the training set is too small (or if too many features are used), poor generalization is expected [20], i.e. classification performance (P_C) for the training set is high, while P_C for the test set is very different. Since a large number of training samples are quite difficult to obtain, we need to reduce the number of features by either feature extraction or feature selection techniques. *Feature extraction* algorithms map (either in a linear or a nonlinear way) *all* of the original features into fewer new features, each of which is a function of *all* original features (we have summarized these methods elsewhere [13]); *feature selection* methods select a small subset of the original features that is most useful for classification (i.e. the response in only several λ bands). For on-line classification, feature selection is more attractive than feature extraction, because it provides faster data acquisition and a less expensive system. Thus, we propose a new HS feature selection technique using ratio features and show that it gives good generalization and better P_C scores than other methods.

An inexpensive and fast inspection system can be fabricated using filters if the response in a maximum of four spectral (λ) bands out of all bands is used. We consider use of λ ratio features (the ratio of the responses at two different λs) and specifically use of our new feature selection algorithm to choose the best two sets of λ band ratios to use. We compare classification results to those obtained using the four best λ bands chosen by our prior feature selection algorithm [13] and the best features chosen by our prior feature extraction algorithm [13] that uses all available λ data. In the ratio feature, we place the lowest numbered waveband in the numerator of the ratio feature; we do not consider both ratios of two wavebands, as they are both expected to have similar information.

Sect. 2 describes the HS almond nut database used. Our fast ratio feature selection algorithm is presented in Sect. 3. We present test results in Sect. 4 and advance conclusions in Sect. 5.

2. DATABASE

As our case study, we consider classification of almond nuts as having concealed internal-damage or not using HS feature selection techniques. Internally damaged almonds are not easily distinguished from normal ones by their external appearances, and, when cooked, they taste bitter. Hyperspectral processing is a non-destructive inspection method that has been shown to be useful for this application [12-14]. The central region of each almond was illuminated by a 100W quartz tungsten halogen lamp. Two different fiber optic transmission spectrometers were used to collect HS spectra; a silicon photodiode array sensor based spectrometer was used to measure the spectrum from λ = 710 – 1000 nm in 0.48 nm intervals, and an InGaAs photodiode array spectrometer was used to obtain the spectrum from λ = 950 – 1390 nm in 3.2 nm intervals. For each almond, ten complete transmission spectra were obtained and the average spectra from each spectrometer was used. Each spectrum was then smoothed by a 19 point Savitzky-Golay second-order filtering operation, sampled at $\Delta\lambda$ = 5 nm increments and combined to produce an HS spectra with 137 spectral samples from λ = 710 – 1390 nm as detailed earlier [12]. For each almond sample, its spectral response was normalized by dividing the spectral response at each λ by the average λ response for that almond sample; this corrects for variations in skin quality, nut thickness, and nut shape [12]. The database contains the spectral responses for 454 almonds with 228 in the training set (173 good and 55 bad nuts) and 226 in the test set (172 good and 54 bad nuts). There are not enough training samples per class for accurate training and classification, since we have only 55 bad nuts for training and this is less than the number of features (137). Thus, this is high-dimensional data for which feature reduction is needed.

3. FAST SPECTRAL RATIO FEATURE SELECTION ALGORITHM

3.1 Prior Work

Feature selection is an N-P complete problem [21], i.e. only an exhaustive search can locate the features that give the best P_C. Using an exhaustive search to find the best single spectral band ratio feature requires (for our 137-feature database) that we evaluate performance of the classifier used for all training set samples for a total of $\binom{137}{2}$ = 9,316 band ratios or sets of ratio features. To select the best two pairs of ratio features requires a search of classification performance on the training set for $\binom{9316}{2}$ = 43,389,270 combinations of two ratio features; this is clearly excessive.

We now address several prior approaches to feature selection with high-dimensional data. For a different HS database, Pearson et al [10] addressed this problem by considering all ratio combinations for only every third spectral response; for his database, this significantly reduced the number of feature sets for which P_C on the Mahalanobis-distance classifier he considered had to be evaluated. We feel that better P_C can be obtained if we do not initially ignore two-thirds of the possible λ band features and perform an exhaustive search on only one-third of the possible λ features. The motivation for considering only every third original feature is that adjacent features (separated by only 5 nm) are expected to contain similar information. This is true for some adjacent spectra. However, in several prior HS case studies [3, 14], we found that the best features often contained adjacent spectral bands. Thus, this approach is not recommended.

In another HS inspection application [16], the best 22 HS feature bands were chosen as ordered by the forward selection (FS) algorithm [22, p. 490], and an exhaustive search of P_C on the training set for all 231 single ratio feature combinations of these 22 spectral bands were used to select the best single ratio of two spectral bands; no reason was given in [16] for selecting only 22 FS bands on which to perform an exhaustive search. The use of forward selection to choose an initial reduced set of N spectral bands is better than simply selecting every third band. However, the best set of two bands does not necessarily include the best single band, the best set of three bands does not necessarily include the best single band or the best two bands, etc. This is known as the nesting problem [23]. Analogous problems occur if backward selection is used to obtain an initial reduced set of features. In addition, these methods produce an ordered set of λ features; *the best ratio features are not necessarily ratios of the best λ features.* This algorithm in [16] is similar to our HDMBB (high-dimensional modified Branch and Bound) feature selection approach [13] in which we use FS and Kullback-Leibler distance (KLD) [24] to select an initial subset of HS features (more than only 22 features), we used this as an ordered subset of features in a new MBB fast and efficient exhaustive search algorithm to select the best three, best four, best five etc set of single λ features to use. In several different HS applications, we showed that it selected the same optimal set of features as a full exhaustive search did. Chen and Casasent [2] applied this algorithm to select ratio

features in a multi-spectral vehicle detection application. Their database contains images taken in 28 spectral bands, each with 4 different polarizations (0°, 45°, 90°, and 135°). Two polarization ratio features at orthogonal polarizations (0 versus 90 degree and 45 versus 135 degrees) were considered for each spectral band for a total of 56 ratio features. The authors used the floating forward selection (FFS) algorithm [25] to select an initial set of 28 ratio features to be used in the MBB search to find the best ratio features from these. Only several polarization ratio features were considered at the same λ band (λ band ratios were not considered).

Another problem with these prior approaches [2, 9, 16] is that we would like to know how much performance P_C was lost by reducing the number of features used; i.e. by using feature selection versus feature extraction methods that employ all original λ features. In prior work [14], a new High-Dimensional Generalized Discriminant (HDGD) feature extraction algorithm was applied to the same database (training and test set) used in this present paper. It was shown [14] to give better classification results than principle component analysis (PCA) and linear discriminant analysis (LDA). The HDGD algorithm is an improvement on LDA; it projects the data onto only some of the eigenvectors $\mathbf{v}_i$ of the within-class covariance matrix $\mathbf{C}_W$. Thus, it produces multiple HDGD output features, rather than just one (as in LDA). The eigenvectors used in the HDGD algorithm were chosen using a new criteria function that considers generalization and discrimination; thus, it has both good discrimination and generalization. In Sect. 4, we compare the P_C performance of our ratio feature selection algorithm to that of the HDGD feature extraction algorithm and the HDMBB feature selection algorithm.

3.2 Overview

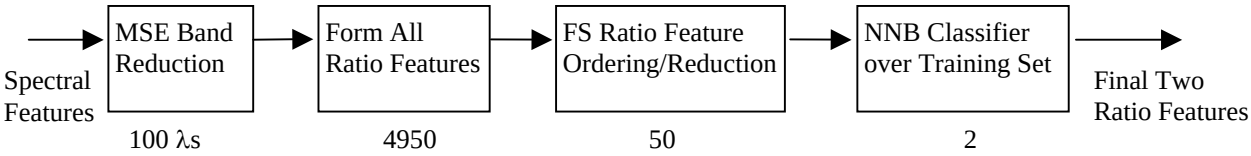

Figure 1. Block diagram of the ratio feature selection algorithm.

Fig. 1 shows a block diagram of our algorithm. We now provide an overview of our new feature selection algorithm; we first reduce the number of feature bands considered by "MSE Band Reduction". This removes bands with similar information. This is allowed because some adjacent bands have very similar responses [7]. However, we do not arbitrarily remove every second band, for example. We use mean square error (MSE) to measure the similarity among all bands and discard those bands with the highest degree of similarity. This is not expected to significantly affect classification performance, since the discarded λ features do not give much new information expected to be useful for classification. We need to discard similar features before computing all possible ratio feature combinations to reduce the search time needed in training. To provide numerical values, we note that for our almond database, we omit 37 spectral bands and consider the response in 100 bands after step 1. After tests on several databases, we can chose MSE limits to determine guidelines for how many λ band features to omit in this first step.

There are now $\binom{100}{2}$ = 4950 combinations of two of those 100 selected bands or 4950 sets of feature ratios. However, since we require two sets of ratio features, we must consider $\binom{4950}{2}$ = 12,248,775 combinations of two sets of feature ratios; thus, this is still too time consuming and prohibitive to use an exhaustive search to search for the two feature ratios with the best performance on the training set. Thus, in step 2, we use "FS Ratio Feature Reduction". We apply forward selection (FS) to the set of 4950 possible ratio features to select the best 50 ratio features (for this present database, we use 50 ratio features as the number of combinations to be searched is reasonable). We chose FS since it considers discrimination. Applying FS to λ ratio feature versus standard features is new. Other ordering methods such as FFS can be used. To select the two best ratio features, we make the assumption that *at least* one of the two best ratio features is one of these 50 FS selected ones; this seems very likely. To select the second ratio feature set (recall, we desire two sets of ratio features), we could choose it from either the remaining first set of 50 FS ordered ratio features or from the other set of 4900 features. To address both choices, we must examine $\binom{50}{2}+\binom{50}{1}\binom{4900}{1}$ = 1225 + 50×4900

= 246,225 combinations of two sets of ratio features. This is a reduction by a factor of approximately 60 compared to the ≅ 12.2 million sets of two ratio features out of 4950. In selecting the final two ratio features, we use P_C score on a nearest neighbor classifier as our performance measure to use to select the best ratio features.

Since our method does not search all possible combinations of ratio features, our result is sub-optimal, i.e. there is no guarantee that our result will yield the best sets of ratio features or a set equivalent to an exhaustive search result. However, this is necessary, and we will note the classification rate P_C we obtain with one and both pairs of ratio features on the training samples using a nearest neighbor (NNB) classifier. To assign an unknown test sample to a class in the NNB classifier, the class of the training set sample that is closest in distance from the test sample determines the class. In practice, one would use a neural network classifier, since its computational load is small. We compare NNB classifier P_C results using two different distance metrics: Euclidean Minimum Distance and the Spectral Angle Mapper as discussed in Sect. 3.3.

3.3 Ratio Feature Selection Algorithm

We now describe our fast ratio feature selection algorithm in detail. We first address step 1, "MSE Band Reduction". When the responses for two wavebands are very similar, we omit one of these original features to reduce the initial number of features considered. This should not lose much information. MSE is used to measure the similarity of two features; it is defined as

$$\mathrm{MSE}(\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b) = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_a(i) - \boldsymbol{\lambda}_b(i)\right)^2 , \tag{1}$$

where $\boldsymbol{\lambda}_a(i)$ is the spectral response in band *a* for the i-th training sample. This measure computes the mean over all N training samples of the square of the difference of two spectral responses. The smaller the MSE, the more similar the two spectra are. For our database, MSE is calculated for all $\binom{137}{2}$ = 9,316 possible pairs of two bands out of all 137 original λ features. We do not search only adjacent bands, although all bands found to be similar were adjacent ones in our initial tests. These MSE values are entered as the elements of a symmetric MSE matrix **M** (note that $\mathrm{MSE}(\lambda_a, \lambda_b) = \mathrm{MSE}(\lambda_b, \lambda_a)$). For our 137-band database, **M** is 137×137. Each **M**(i, j) element is the MSE between the responses of the i-th and j-th wavebands over the training set. We ignore all 137 diagonal elements. To reduce the number of bands used, the minimum element of **M** is determined; this defines the first two bands with the most similarity. Assume bands m and n are most similar, we examine row m and row n of **M** (ignoring element (m,n)), the row with the remaining smallest other element is the band removed (e.g. row m). After we omit row m and column m of **M**, we repeat the procedure until MSE becomes large; for the present database, we used this to remove 37 features. After tests on a number of databases, a guideline for the maximum MSE to use is expected to emerge. Measures besides MSE can of course be used.

In our second step, we consider all ratio feature combinations of the remaining bands. For our present case, these are 4950 combinations of two of the remaining 100 bands. We next use forward selection (FS) to order these 4950 ratio features. We select the best 50 ratio features (ordered by FS). The FS method first selects the best single ratio feature and then adds one ratio feature at a time, which in combination with the first ratio feature maximizes some criterion function J. We use the Mahalanobis distance [26, p. 107] as the criterion function,

$$J = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{C}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \tag{2}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the mean vectors for the good-class and bad-class training samples, respectively, and **C** is the covariance matrix for the training samples. The Mahalanobis distance is large if the mean difference between two classes is large. Various choices for the criteria function are possible; one could use P_C, but computation time may be of concern in the choice. Future work will address this. Others [10] also used Mahalanobis distance. Others used Euclidean [9] distance as the criteria functions, which is a special case of the Mahalanobis distance. To select the best subset of m ratio features out of n original ratio features, the FS algorithm evaluates J for [(2n−m+1)m]/2 subsets of ratio features. For example, to select the best 50 ratio features out of 4950 total ratio features, the FS algorithm requires evaluating J for [(2×4950−50+1)×50]/2 = 246,275 subsets. Since the FS method does not examine all possible subsets,

the resulting subset is not guaranteed to produce the optimal set of ratio values nor the best classification rate P_C. Since the FS method is not optimal and exhibits the *nesting problem*, we retain more than the desired number (2) of ratio features in the FS algorithm; we chose to select 50 ratio features in our initial work. This is an ordered set of features; the best feature, the next best feature used with the first, etc is selected.

As the final step in our ratio feature selection algorithm, we use a nearest neighbor classifier and search the set of ratio features for which pair of ratio features gives the best classification on the training set. We consider selecting one ratio feature from the 50 FS ordered ratio features and one from the other 4900 ratio features or choosing both ratio features from the 50 FS ratio features. To determine the best pair of ratio features, we use the classification rate P_C for an NNB classifier on the training set. We note that this P_C performance measure differs from J in equation (2) used in the prior FS step; thus the best ratio feature chosen by J may not be best using P_C. Future work can address different performance measures for the different steps. As noted earlier, this search requires evaluation of NNB classifier performance on the training set for $\binom{50}{2}+\binom{50}{1}\binom{4900}{1}$ = 246,225 sets of two ratio features. This is our fast ratio feature selection algorithm. To compute P_C for the test set, we use the two ratio features in the NNB classifier. To assign a test sample to a class in the NNB, the shortest distance from that sample to other good-class and bad-class samples in the training set is computed. The NNB classifier assigns the test sample to the same class as its nearest neighbor in the training set. When P_C for the training set is calculated, the training set sample being classified is of course removed from the NNB classifier. Future work should consider a neural net classifier instead of the NNB classifier, since it requires fewer on-line computations.

In our NNB classifier, we consider the two most prominent distance metrics used in HS processing [27]. We consider the Euclidean Minimum Distance (EMD), this measures the distance between two samples **x** and **y**

$$\Delta(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{\sum_{i=1}^{F}(x_i - y_i)^2}\ , \tag{3}$$

where F is the number of features used in the NNB classifier (F = 2 in our case). EMD is monotonic (with the number of features) and rotational-invariant (the minimum distance in all directions is considered). We also consider the Spectral Angle Mapper (SAM) distance metric in our NNB classifier. SAM measures the angle between two samples **x** and **y**

$$\theta(\mathbf{x},\mathbf{y}) = \arccos\left(\frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{x}\|\cdot\|\mathbf{y}\|}\right) = \arccos\left(\frac{\sum_{i=1}^{F} x_i y_i}{\sqrt{\sum_{i=1}^{F} x_i{}^2}\sqrt{\sum_{i=1}^{F} y_i{}^2}}\right) \quad ,0 \le \theta \le \frac{\pi}{2} \tag{4}$$

where $\langle \cdot , \cdot \rangle$ is the dot product operator. The SAM metric is not monotonic but is invariant to multiplicative scaling (this is attractive since it is not changed by differences in the intensity of the HS light source used). The class of the training set sample with the smallest Δ or θ between it and the test input **x** is chosen as the class of the test input in the NNB classifier. Most algorithms for detection and classification in HS data utilize either SAM or EMD as the metric to compare the similarity of two spectra. We compare their performance in Sect. 4. Our almond nut HS database is normalized and the ratio features we consider are invariant to multiplication; thus there may be no advantage to the use of the more common SAM metric in our present application.

4. RESULTS AND DISCUSSION

We first address the similarity among the responses in different wavebands over the training set samples in our HS database. Response plots of the average transmittance spectra of good and bad almond nuts are shown in Fig. 2. Fig. 3 shows the spectral responses for bands 77 and 78 for all training set samples; the first 173 samples are good almond nuts and the other 55 samples are bad almond nuts. The responses for both bands are essentially identical; the MSE between

these two bands is very small (MSE(77, 78) = 7.22×10^{-6}). Thus, band 78 could be removed in step 1 without much loss of information. The 37 bands removed are listed in Table 1 in the order removed. All similar responses found were for adjacent bands.

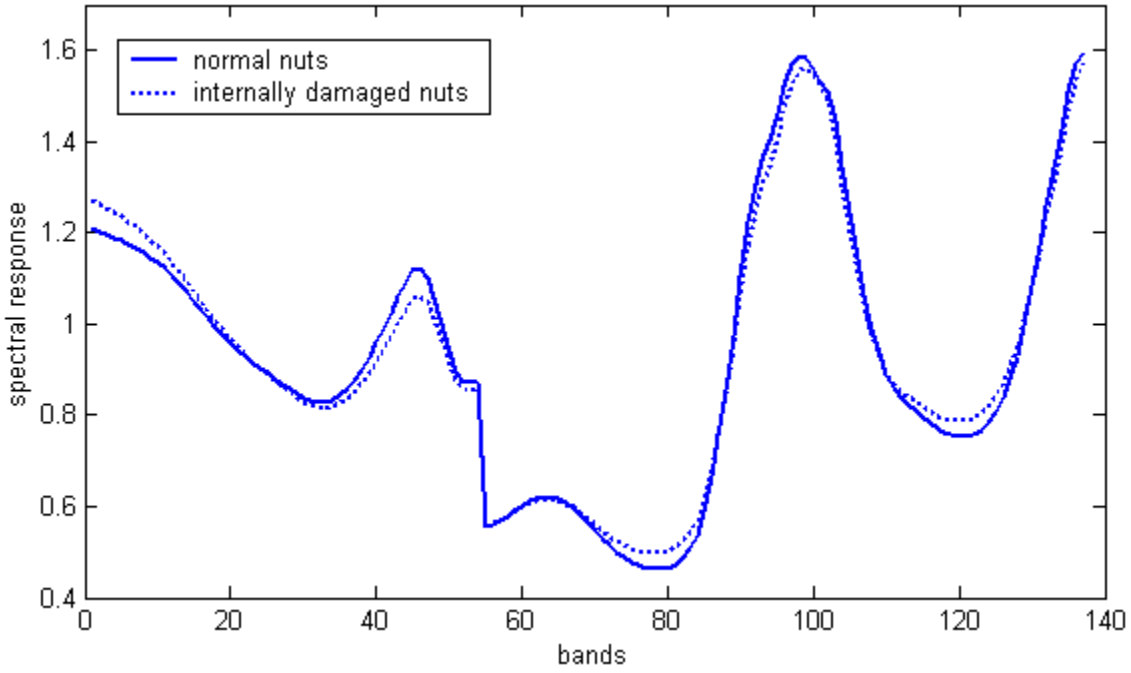

Figure 2. Average normalized spectral responses of normal and internally damaged almond nuts for training set data.

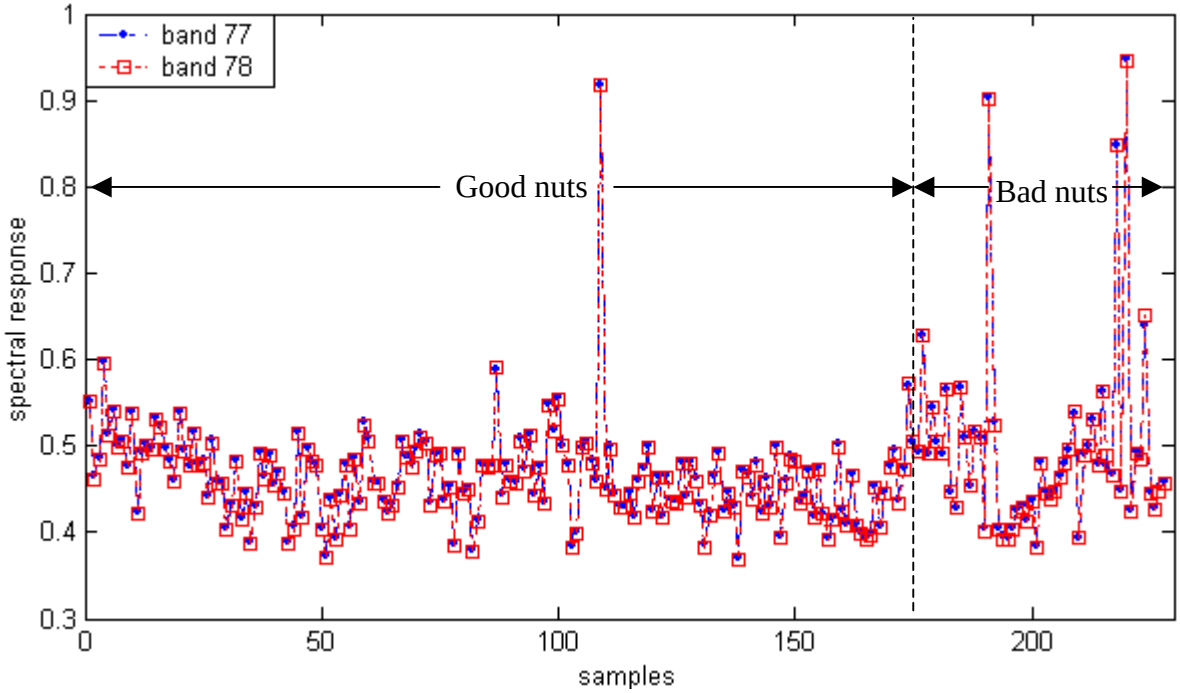

Figure 3. Spectral responses for band 77 and band 78 for all training set samples.

Table 1. The 37 bands removed by MSE band reduction (step 1) in the order removed.

number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
λ feature	78	79	120	121	32	31	53	54	33	77	63	62	64	80	119	122	45	46	52
number	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	
λ feature	65	34	118	76	61	30	117	123	81	3	2	56	55	75	66	4	5	57	

All 4950 combinations of ratio features of two bands out of the reduced set of 100 bands are now considered. Table 2 lists the scores for P_C (percentage of good and bad almonds correctly classified), P_{Good} (percentage of good almonds detected), and P_{FA} (percentage of bad almonds misclassified as good almonds) using one and two ratio features using the EMD metric in our NNB classifier for the training and test set data. The specific waveband numbers used in each ratio feature are also noted. The single best ratio feature to use (ratio feature 115/128) involved a search of all 4950 ratio features for which gave the best P_C classification on the training set. We note that this single best ratio feature (using P_C) was not in the top 50 ratio features (ordered by the criteria function J and the FS algorithm). As we see from Table 2, the P_C scores for the training and test set using one or two ratio features using the EMD metric are comparable; thus, generalization is good. Using two ratio features yields significantly higher P_C scores than using only one ratio feature as we expected. Using P_C, the two best ratio features from the 50 FS ordered set (44/48 and 86/116) are features 1 and 21 ordered by the FS algorithm; note that they are not features 1 and 2 ranked by J in the FS algorithm. The best P_C was obtained when one ratio feature was chosen from the 50 FS ordered set and the other from the remaining 4900 ratio features. The training set P_C for this case is about 2% higher than that using the two best ratio features from the 50 FS ordered set (92% vs 90%). The best ratio feature (using P_C) from the set of 50 FS ordered ratio features was the ratio feature (19/44); it was not the top one ordered by J (it ranked 38 out of the 50 FS ratio features). On the test set, we obtained a noticeably higher test set P_C score of 91.6% (vs 87.6%) when choosing one ratio feature from the 50 FS ordered set and the other from the remaining 4900 ratio features. In these scores, we note that only 7 out of 172 good almond nuts are misclassified as bad nuts, and 12 out of 54 bad almonds are misclassified as good almond nuts. We suspect that some of the test samples in error are not representative of the training set samples. We note that the performance P_{Good} on good nuts is much better than that on bad nuts (described by P_{FA}). We attribute this to differences between the training and test sets and to the small size of the bad training set (54 bad nuts vs 172 good nuts).

We list in Table 3 the same scoring measures and data using the SAM metric (rather than the EMD one) for our NNB classifier for the training and test set data. The SAM metric cannot be used with only one ratio feature in the NNB classifier; thus, we do not consider it for this case. As before, generalization was good. From Table 3, when we chose two best features from the 50 FS ordered set, we note that one of the ratio features is 17/18; thus, adjacent bands provide useful information for classification. The best P_C was again obtained when one ratio feature was chosen from the 50 FS ordered set and the other from the remaining 4900 ratio features. Using the EMD distance metric (Table 2) in our NNB classifier gives about a 2% higher P_C training score than does use of the SAM metric. The multiplicative scaling advantage of the SAM metric is not of use here, since our HS database was normalized and well recorded. Further tests on other databases are necessary to determine which method is best in general and why. The training set computation time for our fast ratio feature selection algorithm on a 1.7 MHz Pentium IV PC was about 3 hours which is of course much faster than an exhaustive search (10 days).

Table 2. Classification results for our fast ratio feature selection algorithm using the EMD distance matrix in the NNB classifier.

Number of ratio features	Wavebands used in ratio features	Training set			Test set		
		P_C(%)	P_{Good}(%)	P_{FA}(%)	P_C(%)	P_{Good}(%)	P_{FA}(%)
1	115/128	83.8	90.2	36.3	78.8	83.8	37.0
2 ratio features from top 50 FS	44/48 86/116	90.8	94.2	20.0	87.6	91.9	26.0
1 from top 50 FS, 1 from the rest	19/44 39/50	92.5	96.5	20.0	91.6	95.9	22.2

Table 3. Classification results for our fast ratio feature selection algorithm using the SAM distance matrix in the NNB classifier.

Number of ratio features	Wavebands used in ratio features		Training set			Test set		
			P_C(%)	P_{Good}(%)	P_{FA}(%)	P_C(%)	P_{Good}(%)	P_{FA}(%)
2 ratio features from top 50 FS	44/48	17/18	87.7	92.5	27.3	89.4	95.4	30.6
1 from top 50 FS, 1 from the rest	20/49	17/42	90.4	93.1	18.2	88.5	95.4	33.3

Figs. 4 and 5 show plots of the training and test set samples using the two best ratio features for the EMD and SAM distance metrics respectively used in the NNB classifier. From Figs. 4a and 4b, most of the good samples are seen to be separated in distance from bad samples as classification using the EMD distance requires. To obtain a high classification rate using the SAM distance, good and bad samples should fall on different lines through the origin (0, 0) at different angles, and we want the angles between the two lines to be large. This trend is shown in Figs. 5.

In Table 4, we also compare our results using two ratio features (four different wavebands) to those using the best subset of four separate bands chosen from all 137 original wavebands (using the HDMBB feature selection algorithm) and to the best four features (using the HDGD feature extraction algorithm). The HDMBB feature selection algorithm we used to choose the best separate bands was detailed earlier ([13] and Sect 3.1); for the present database, we first used the KLD algorithm to reduce the number of original λ wavebands from 137 to 30; the MBB algorithm was then used to select the optimal subset of wavebands from the reduced 30-dimensional feature space. For feature extraction, we used the HDGD algorithm to combine all of the original features into fewer new features, each of which is a function of all original features. We compare the results of these feature reduction methods in Table 4. Table 4 compares these algorithms using an NNB classifier and the EMD distance metric on the same HS database with the same training and test set.

Table 4. Test set P_C comparison for different feature reduction algorithms.

Algorithm	Features	P_C (Test) %
Feature selection: KLD + MBB	Best 4 λs: 1, 23, 40, 78	83.2
	Best 6 λs: 12, 35, 40, 46, 54, 103	85.9
HDGD feature extracion	Best 4 HDGD features	90.7
Ratio feature selection	Best 2 ratio features: 19/44, 39/50	91.6

In Table 4, we show results using the same number of wavebands (4); this is four separate wavebands or two ratio features. *Our new ratio features give a higher classification rate by 8% compared to the optimal subset of four separate wavebands alone (91.6% vs 83.2%). It also gives a higher classification rate by 1% compared to the HDGD feature extraction algorithm, which uses all λ data.* Feature extraction gives better P_C than standard feature selection, but its P_C is not better than our algorithm for feature selection using ratio features. We next consider the λs selected by the different algorithms. Only one waveband feature (feature 40) of the best 6 λs is in the set of the best 4 λs selected by the KLD/MBB standard feature selection algorithm; This algorithm shows that the 6 best λ features to use contain only one of the best 4 λs (hence algorithms such as FS, that exhibit the nesting problem, cannot select the best λ features). None of the best 4 or 6 separate λs is used in our fast ratio feature selection algorithm. This clearly shows that the best separate wavebands do not necessarily give the best ratio features. Thus, when we consider using ratio features for classification, we need to use a method to order the ratio features, rather than separate waveband feature; our new algorithm does this.

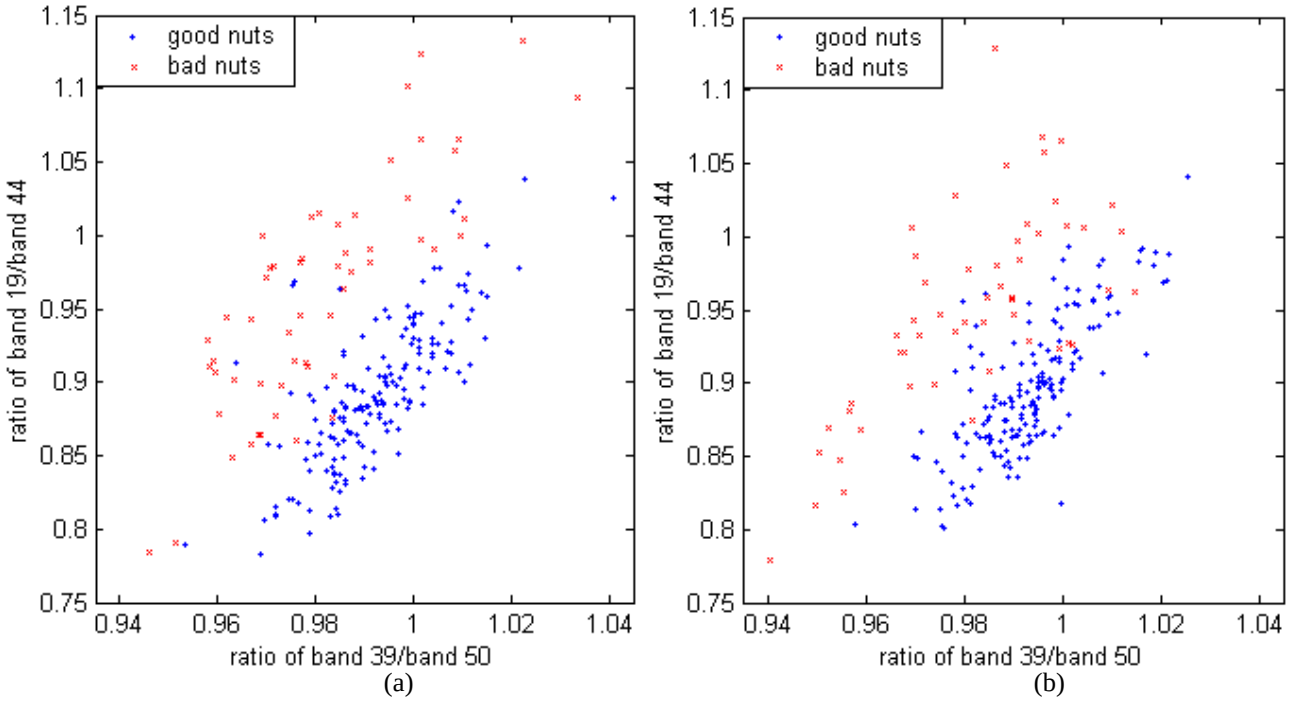

Figure 4. Plot of the training (a) and test (b) set samples for the two best ratio features chosen using the EMD distance metric used in the NNB classifier.

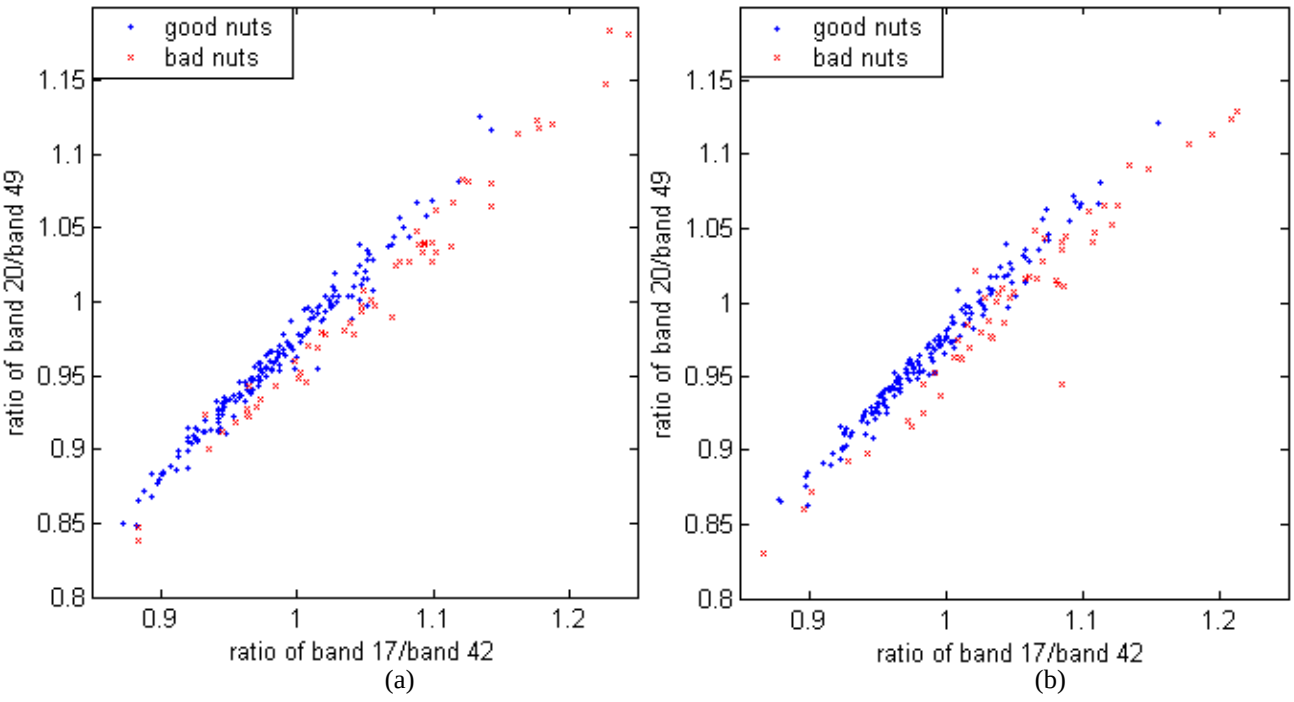

Figure 5. Plot of the training (a) and test (b) set samples for the two best ratio features chosen using the SAM distance metric used in the NNB classifier.

5. CONCLUSIONS

Our fast ratio feature selection algorithm for high-dimensional data was described. It gave good generalization and required much less calculations than does an exhaustive search. Thus, our proposed method is very useful and practical for product inspection using HS data. Initial results show that using only two ratio features (four different wavelengths) for classification gives a very high P_C test set score (91%). We found that using ratio features gave a higher classification rate than did use of the best feature selection subset of separate wavebands or than did use of a feature extraction algorithm using all λ responses. None of the best 4 or 6 separate λs is used by our fast ratio feature selection algorithm; thus an algorithm such as ours is necessary. Furthermore, use of the EMD distance metric in the NNB classifier was shown to give slightly higher classification rates (by 2%) than did use of the more standard SAM distance metric for our database. Further tests on more databases are needed.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Tom Pearson (ARS, Manhattan, Kansas) for providing the HS database used in this paper.

REFERENCES

[1] B. Thai, and G. Healey, "Invariant subpixel target identification in hyperspectral imagery," *Proc. SPIE*, vol. 3717, pp. 14-24, 1999.

[2] X.-W. Chen, and D. Casasent, "Feature selection from high-dimensional hyperspectral and polarimetric data for target detection," *Proc. SPIE*, vol. 5437, pp. 171-178, 2004.

[3] D. Casasent, and X.-W. Chen, "Feature reduction and morphological processing for hyperspectral image data," *Applied Optics*, vol. 43 (2), pp. 227-236, 2004.

[4] J. Goutsias, and A. Banerji, "A morphological approach to automatic mine detection problems," *IEEE Trans. Aerospace and Electronic Systems*, vol. 34(4), pp.1085-1096, 1998

[5] M.J. Muasher, and D.A. Landgrebe, "The K-L expansion as an effective feature ordering techniques for limited training sample size," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-21, pp. 438-441, Oct 1983.

[6] K. Chao, P.M. Mehl, M. Kim, and Y.-R. Chen, "Detection of chicken skin tumors by multispectral imaging," *Proc. SPIE*, vol. 4206, pp. 214-223, 2001.

[7] S.G. Kong, Y.-R. Chen, I. Kim, and M. Kim, "Analysis of hyperspectral fluorescence images for poultry skin tumor inspection," *Applied Optics*, vol. 43(4), pp. 824-833, 2004.

[8] S. Nakariyakul, and D. Casasent, "Hyperspectral feature selection and fusion for detection of chicken skin tumors," *Proc. SPIE*, vol. 5271, pp. 128-139, March 2004.

[9] B. Park, W.R. Windham, K.C. Lawrence, and D.P. Smith, "Classification of hyperspectral imagery for identifying fecal and ingesta contaminants," *Proc. SPIE*, vol. 5271, pp. 118-127, 2004

[10] T.C. Pearson, D.T. Wicklow, E.B. Maghirang, F. Xie, and F.E. Dowell, "Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy," *Trans. of the ASAE*, vol. 44(5), pp.1247-1254, 2001.

[11] D. Casasent, X.-W. Chen, and S. Nakariyakul, "Hyperspectral methods to detect aflatoxin in whole kernel corn," *Mycopathologia*, vol. 157(4), pp. 422, May 2004.

[12] T. Pearson, "Spectral properties and effect of drying temperature on almonds with conceal damage," *Lebensm.-Wiss. U.-Technol*, vol. 32, pp. 67-72, 1999.

[13] D. Casasent, and X.-W. Chen, "Feature reduction methods for hyperspectral data," *Proc. SPIE*, vol. 4572, pp. 1-11, 2001.

[14] D. Casasent, and X.-W. Chen, "Waveband selection for hyperspectral data: optimal feature selection," *Proc. SPIE*, vol. 5106, pp. 259-270, April 2003.

[15] F. Dowell, M. Ram, and L. Seitz, "Predicting scab, vomitoxin, and ergosterol in single wheat kernels using near-infrared spectroscopy," *Cereal Chem.*, vol. 76(4), pp.573-576, 1999.

[16] S. Delwiche, R. Stephen, and M. Kim, "Hyperspectral imaging for detection of scab in wheat," *Proc. SPIE*, vol. 4203, pp. 13-20, 2000.

[17] D. Lamb, and C. Hurburgh, "Moisture determination in single soybean seeds by near-infrared transmittance," *Trans. of the ASAE*, vol. 34(5), pp. 2123-2129, 1991.

[18] M. Mehrubeoglu, and G. Cote, “Determination of total reducing sugars in potato samples using near-infrared spectroscopy,” *Cereal Foods World*, vol. 42(5), pp. 409-413, 1997.
[19] A. Jain, R. Ruin, and J. Mao, “Statistical pattern recognition: a review,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22(1), pp. 4-37, 2000.
[20] E. Baum, and D. Haussler, “What size net gives valid generalization?” *Neural Computation*, vol. 1, pp. 151-160, 1989.
[21] T. Cover, and J. Campenhout, “On the possible orderings in the measurement selection problem,” *IEEE Trans. Systems, Man, and Cybernetics*, SMC-7(9), pp. 657-661, 1977.
[22] K. Fukunaka, *Introduction to Statistical Pattern Recognition*, 2nd Ed., Academic Press Inc., New York, 1992.
[23] T.M. Cover, “The best two independent measurements are not the two best,” *IEEE Trans. Systems, Man, and Cybernetics*, vol. 4, pp. 116-117, 1974.
[24] T.M. Cover and J. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, NY, 1991.
[25] P. Pudil, J. Novovicova, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, 1994.
[26] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd Ed., A Wiley-Interscience Publication, New York, pp. 107, 2001.
[27] N. Keshava, “Angle-based band selection for material identification in hyperspectral processing,” *Proc. SPIE*, vol. 5093, pp. 440-451, 2003.