

SPECTRAL BAND SELECTION FOR OPTICAL SORTING OF PISTACHIO NUT DEFECTS

R. P. Haff, T. Pearson

ABSTRACT. A technique using near-infrared spectroscopy (NIR) was developed for selecting the optimal spectral bands for use in dual-wavelength sorting machines commonly found in food processing plants. A variation of a nearest-neighbor classification scheme selected the two optimal spectral bands given NIR spectra from both sides of an object. The optimal bands were determined for two cases: when both sides contain the defect of interest (AND logic), or when the defect appears on a single side (OR logic). A commercially available sorting machine was used to compare the sorting accuracy using the spectral bands determined with this technique to the accuracy using bands recommended by the manufacturer. The product stream tested was the removal of “small inshell” (small nuts with the shell intact) and shell halves from the stream of nuts with no shells (“kernels”). Results for the selected spectral bands averaged 1.20% false negative (fn) for small inshell and 1.80% fn for half shells with 0.15% false positive (fp) vs. 1.70%, 2.40%, and 0.70%, respectively, using the spectral bands recommended by the manufacturer. Optimal spectral bands were also determined and reported for a variety of other defects and unwanted materials commonly sorted in the pistachio processing plant, including adhering hull, stained, sticks, mold, insect damage and/or webbing, and black spots. Given the success of this technique in pistachio sorting experiments, it is believed that it could be applied to any commodity sorted using commercially available, dual-wavelength, NIR sorting devices.

Keywords. Dual-wavelength sorting, NIR spectroscopy, Pistachio nuts, Spectral bands.

High-speed optical sorters have found extensive use for removing blemished product and foreign materials for many different commodities including tree nuts, peanuts, grain, and vegetables. Some specific examples include removal of kernel bunt in wheat (Dowell et al., 2002) and kernel color class separation (Pasiatkan and Dowell, 2002). Pearson et al. (2001) reported algorithms for use with commercially available monochromatic image-based sorters to detect defects in pistachios. These algorithms were able to distinguish normal nuts from nuts with oily stain, dark stain, and adhering hull defects with an accuracy of 98%. The same algorithm also identified 89.7% of nuts with kernel decay, 93.8% of nuts with *Aspergillus* molds present, and 98.7% of nuts with damage from the naval orange worm (NOW). These systems can inspect and sort objects at a rate of up to 100/second/channel, depending on the size of the object to be sorted. Initially, these machines were monochromatic devices and sorting was based on the measurement of a single band of light, usually in the visible portion of the spectrum. More recently, devices have been developed that measure light in the NIR (from 1000 to 1700 nm), allowing sorting based on chemical rather than

visible differences among products. In pistachio processing plants, visible light sorting devices have largely been replaced with NIR devices, and the commercially available image-based sorters used by Pearson et al. (2001) are in fact no longer in production.

Today, NIR sorting machines are available in a variety of configurations. The simplest is a monochromatic sorter (fig. 1a), which can detect a single band anywhere between 400 and 1700 nm, depending on the choice of detector. Some dual-band sorters measure a single band in the NIR region using an InGaAs sensor with sensitivity between 1100 and 1700 nm and another band in the silicon region between 400 and 1000 nm (fig. 1b). These sorters employ two bandpass filters, one over a silicon sensor and another over an InGaAs sensor. The two sensors are located together at the point where the product flows by the chute. Modular dual-band sorters use a beam-splitting mirror, and then pass each beam through an interference filter over a sensor (fig. 1c). This design allows changing of the mirror, filters, and sensors, so that a silicon-silicon, silicon-InGaAs, or InGaAs-InGaAs arrangement is possible. This design has one limitation in that the beam-splitting mirror requires at least 100 nm difference between the wavelengths of the respective bands.

Dual NIR band sorters can measure specific bands as required for a specific defect and commodity. This is accomplished using a light source that covers roughly the entire NIR spectrum, and selecting the wavelengths of interest using optical bandpass filters. This allows for the measurement of different bands by simply changing the filters. The effectiveness of the sorter depends on the selection of the appropriate filters for the product being sorted. The simplest and most common technique for selecting the two bands to be used is to compare reflectance

Submitted for review in February 2006 as manuscript number FPE 6352; approved for publication by the Food & Process Engineering Institute Division of ASABE in June 2006.

The authors are **Ron P. Haff, ASABE Member Engineer**, Agricultural Engineer, USDA-ARS Western Regional Research Center, Albany, California; and **Tom C. Pearson, ASABE Member Engineer**, Agricultural Engineer, USDA-ARS Grain Marketing and Production Research Center, Manhattan, Kansas. **Corresponding author:** Ron P. Haff, USDA-ARS-WRRC, 800 Buchanan St., Albany, CA 94710; phone: 510-559-5868; fax: 510-559-5684; e-mail: ron@pw.usda.gov.

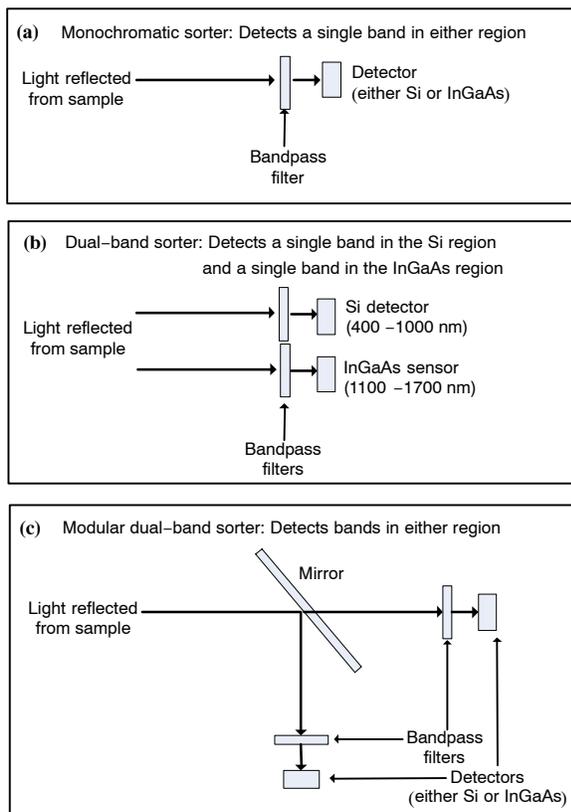


Figure 1. The three most common commercially available NIR sorting devices. The design of the modular dual-band sorter allows easy changing of the mirror, filters, and detectors and is therefore able to detect any bands between 400 and 1700 nm at either detector.

spectra for the good and bad product across the entire NIR range and select the two bands where the differences are greatest. However, this does not take into account variability of absorbance values nor the synergistic effect that two bands may have together, i.e., the ratio of the measurements at two wavelengths can be a more effective basis for sorting than absolute differences at specific wavelengths.

All sorters have at least two sensor modules so that at least two sides of an object are inspected. A decision to reject an object can be made if the measurement from only one side indicates a defect (called OR logic) or if both sensors indicate that a defect is present (AND logic). OR logic is required when a defect typically appears on a single side of an object. Otherwise, AND logic may be preferred to minimize the rejection of good product. Thus, when selecting spectral bands for dual-band sorters, both logic types should be tested.

Some statistical approaches exist to aid in selecting filters, or a small number of spectral bands, given digitized spectra to identify defects. One approach, used by Pasikitan and Dowell (2002), is to use stepwise discriminant analysis to select a small group of features. However, data within each spectrum is highly correlated, resulting in ill-conditioned matrices during the computation of the F statistic used in stepwise discriminant analysis. This can sometimes lead to selection of bands that are not at all optimal. Bajwa et al. (2004) give a review of spectral band selection for use with hyperspectral images. Two common approaches have been used in recent years: genetic algorithms (GA) (Lestander et al., 2003; Steward et al., 2005; Bajwa et al., 2004), and principle component analysis (PCA). Genetic algorithms select

features using improvements from randomly chosen subsets of features. With PCA, the eigenvectors computed are inspected to ascertain spectral bands having high discrimination power (Bajwa et al., 2004; Lawrence et al., 2003; Mehl et al., 2002). However, the GA and PCA methods do not take into account that more than one spectrum from a different location of each sample might be present and the AND/OR decision logic used by the sorting machines. Additionally, these methods do not exhaustively test all combinations of spectral bands. With the PCA approach, it is assumed that the higher weights of the eigenvectors are due to the different classes of spectra; this may or may not be the case. Given that each spectrum may contain 200 to 500 absorbance values, testing all combinations of two or three bands is not a problem for modern computers.

Pearson et al. (2004) developed a statistical procedure to analyze visible and NIR spectra for the selection of spectral bands for sorting yellow dent corn with mycotoxin contamination in a dual-band sorter. The Mahalanobis distance from absorbance band pairs from each side of a kernel to the contaminated and uncontaminated groups was computed for each possible pair of spectral bands. A kernel was classified as contaminated if the Mahalanobis distance (based on two absorbance values) from either kernel side was closer to the contaminated group than to the uncontaminated group. The pair that resulted in the best classification accuracy was used to test the sorter. However, this method assumes that the data are normally distributed. For some defects, such as shell halves, or where only one side of a kernel is discolored, the spectra from each side of the object are substantially different and the assumption of normal distributions may not be valid. Another method of classification, the *k*-nearest neighbor algorithm, does not need to assume any distribution of the data, as it is trained by a relatively small number of “examples” or instances of data points (Han and Kamber, 2001; Caltepe et al., 2004). However, there is no known *k*-nearest neighbor software that selects features using an exhaustive search of all possible features or that can integrate the AND/OR logic used by sorting machines.

The pistachio industry has rigid quality standards for certain defects. For example, no more than two pieces of shell are generally allowed per ton of shelled kernels. Since this level is not currently attainable using automated sorters alone, product must be manually inspected, which is costly (\$0.20/lb) and often inconsistent. The same is true for other defects and foreign material, such as sticks, which are removed with automated sorters. Improved performance from these sorters would clearly benefit the industry.

The processing stream for pistachio nuts is generally divided in two: those nuts still in their shells (“in shell”), and those with no shells (“kernels”). Both streams make use of automated as well as hand sorting. While most automated sorting is performed with monochromatic sorters at visible wavelengths, the newer dual-band sorting machines with NIR capability are gaining acceptance in the industry. However, it is not clear that the spectral bands currently used as a basis for sorting are optimal. Improved band selection has the potential to increase the accuracy of dual-band sorting devices at minimal cost, thus reducing the need for costly hand sorting.

OBJECTIVE

The objectives of this study were: (1) to develop a technique to determine which spectral bands are optimal for

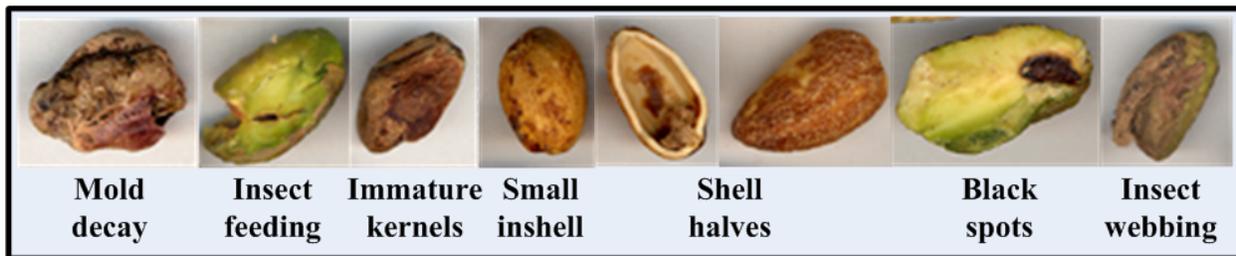


Figure 2. Common kernel defects found in pistachio nut processing streams.

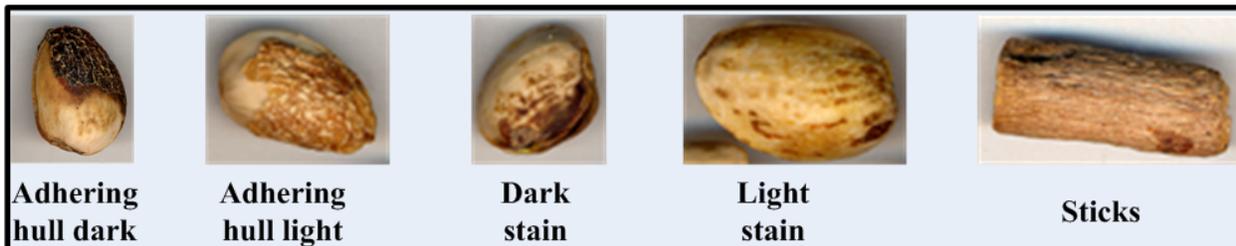


Figure 3. Common defects found in inshell processing streams.

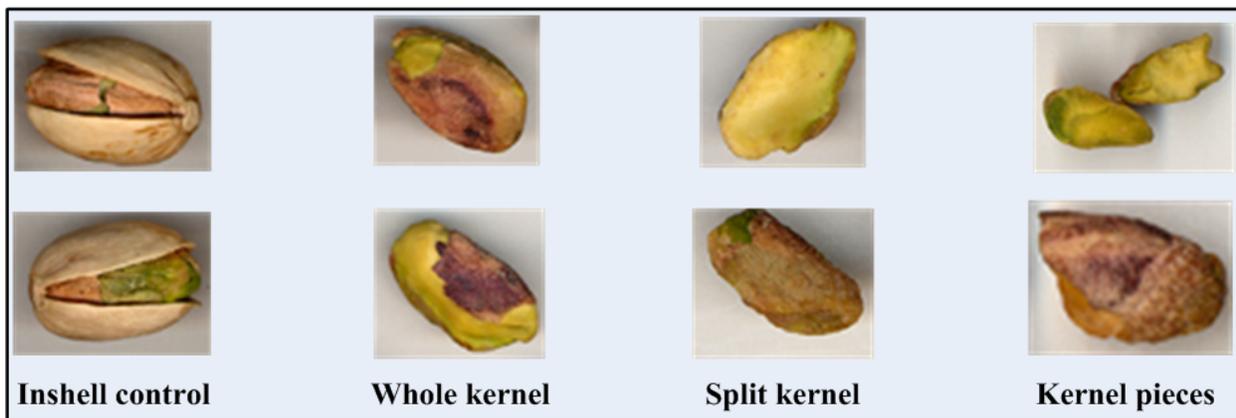


Figure 4. Controls for inshell and kernels.

specific sorting tasks using dual NIR band sorters, (2) to use this technique to determine optimal bands for several common pistachio defects, and (3) to compare sorting results for a few important pistachio defects (shell pieces and small inshell nuts mixed with kernels) using the selected bands to those achieved using manufacturer-recommended filters.

MATERIALS AND METHODS

SAMPLE COLLECTION

Approximately 2 kg samples of several common defects or foreign materials found in the pistachio processing streams were collected at a processing plant after hand sorting by plant personnel. Existing cleaning and sorting machines had processed all samples, and hand sorting was the final step before roasting, salting, and packaging. Figure 2 shows kernel defects collected, figure 3 shows inshell defects, and figure 4 shows controls for inshell and kernel processing streams.

SPECTRA COLLECTION

Single kernel/nut reflectance spectra from 500 to 1700 nm were measured using a diode-array near-infrared spectrome-

ter (DA7000, Perten Instruments, Springfield, Ill.) with a 42 W tungsten halogen lamp light source. The spectrometer measures absorbance using an array of silicon (7 nm resolution, 500 to 1000 nm) and indium-gallium-arsenide sensors (11 nm resolution, 1000 to 1700 nm). Fifteen spectra from each kernel were collected and averaged, and the time required to obtain each spectrum was approximately 33 ms. Spectra of a reference material (Spectralon) were collected after every thirty kernels for calibration purposes. This procedure minimized the effect of fluctuations from the light source. Spectra were obtained from 300 kernels of each defect type and 300 of each control type. Kernels were manually placed on a bifurcated interactance probe attached to the spectrometer and light source (fig. 5). The viewing area was 17 mm in diameter and rested 10 mm above the termination of the illumination and reflectance fibers. The illumination bundle was a 7 mm diameter ring, and the reflectance probe bundle was 2 mm in diameter. Spectra were first collected with each nut oriented on its flattest side, and then a second spectrum was obtained with the kernel axially rotated approximately 90°. This was generally sufficient to ensure that the kernels were exposed through the shell split for in-

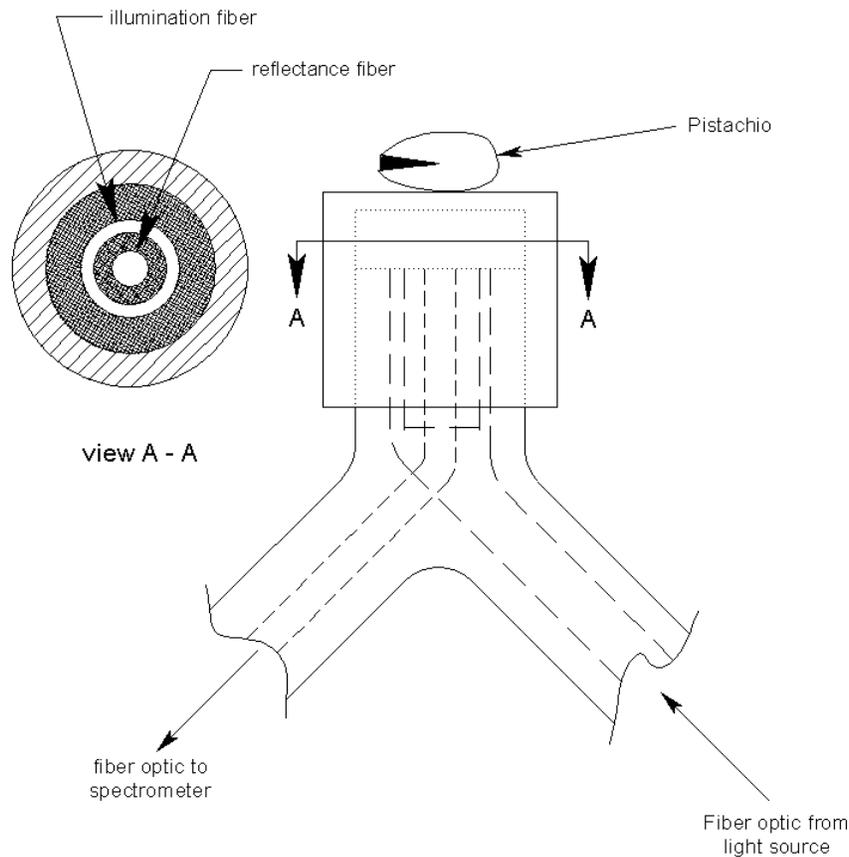


Figure 5. Interactance probe used to collect reflectance spectra of individual kernels.

shell nuts. For split shells and split kernels, each sample was rotated 180°, as they were too thin to acquire spectra at a 90° rotation (side view). All spectra were digitally stored for subsequent analysis.

FILTER SELECTION FROM SPECTRA

All spectra were interpolated (Perten Simplicity software) into 5 nm bins between 500 and 1700 nm, resulting in 241 absorbance values. The spectra were then convolved with Gaussian curves to simulate the effect of different 10 nm

wide bandpass filters. The Gaussian kernel used for the convolutions had a full width at half maximum (FWHM) of 5 nm, matching the width of the bins into which the absorbance bands were interpolated. The resulting spectral bands were used for all further analysis. Figure 6 shows plots of the average spectra for good kernels, small inshell nuts, shell halves (outside), and shell halves (inside). The high absorbance of the inside of the half shell is due to the fact that the inner surface is darker than the outer surface, and the concave inner surface does not reflect much of the incident

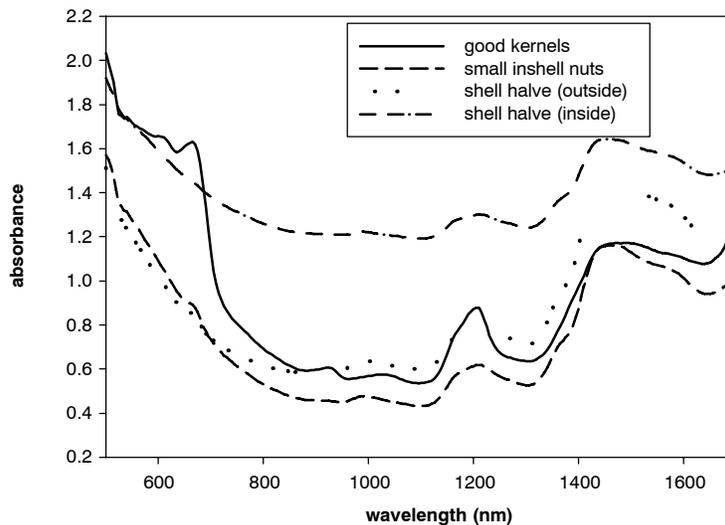


Figure 6. Average spectra for the four streams selected for sorter testing.

light straight back at the sensor. Spectra from small inshell nuts and the outer surface of the half shells are similar, with inshell nuts being slightly darker and having higher absorbance values. Spectra from the kernels exhibit a sharp peak around 1200 nm and a minor peak around 930 nm, where oil absorbs more light.

A k -nearest neighbor classification scheme was used to select the optimal subset of spectral bands for sorting a given “good” product and “defect.” The k -nearest neighbor algorithm was chosen as it can accommodate different distributions of spectra taken from different sides of each nut. In addition, with adaptation, the AND/OR decision logic of sorting machines can be integrated into the feature selection method.

An exhaustive search was performed for the best single spectral band, pair of spectral bands, and triplet of spectral bands. The best single spectral band accommodates monochromatic NIR sorters (fig. 1a). For the case of two spectral bands, optimal band selection was conducted to accommodate the two machine types (figs. 1b and 1c) with one pair constrained so that one band was below 1050 nm (silicon region) and the other was between 1100 and 1700 nm (NIR region). The other pair, intended for the modular design, had no constraints when separated by at least 100 nm, as required due to the beam-splitting mirror. Although sorting machines that detect three NIR spectral bands are not commercially available, the selection of the best triplet of spectral bands was intended to determine whether or not such a machine could improve sorting accuracy compared to dual-band sorters.

For each class of nut (tables 1 through 3) a random sample of 30 good kernels and 30 defect kernels were used as the training set. The remaining 270 nuts for each class served as the validation set. For each of the 241 Gaussian smoothed absorbance values in the spectra, the mean and standard deviation of the training set was computed. Each absorbance value was normalized as shown in equation 1:

$$A_{N\lambda} = \frac{(A_{\lambda} - \bar{A}_{\lambda})}{\sigma_{\lambda}} \quad (1)$$

where $A_{N\lambda}$ is the normalized absorbance centered at wavelength λ , A_{λ} is the Gaussian smoothed absorbance centered at wavelength λ , \bar{A}_{λ} is the training set mean of the Gaussian smoothed absorbance at wavelength λ , and σ_{λ} is the training set standard deviation at wavelength λ .

When searching for the best single wavelength for separating good kernels from defect kernels, each of the 241 Gaussian smoothed absorbance values (bins) that make up the spectra for an individual kernel can be considered as a point in one-dimensional space. For clarity, we refer to the bins as wavelengths, although in actuality each bin represents a 5 nm wide band of wavelengths. At each wavelength, the spectra from each kernel in the validation set (front and back) were evaluated to determine which k kernels in the training set lie closest in that space. The value for k in this work was chosen as 3, as this reduces the effect of outliers. Distance between points was simply measured as the Euclidean distance. For AND logic, the kernel was classified as defect if absorbance values from both sides of the kernel were closest to defect kernels in the training set, while OR logic required only one side to be closer to a defect kernel. The result was a 241-element array for each kernel in the validation set, with the entries being either good or defect. At

each wavelength, entries for all kernels in the validation set were compared to the known true conditions of the kernels, and the classification accuracy was computed. Here, it would be feasible to select a wavelength whose classification results had desired characteristics, such as the lowest false positive or lowest false negative rate. For this study, the wavelength selected as “best” was that which yielded the lowest overall error rate.

The procedure for selecting the best two wavelengths was exactly as above, except that the Euclidean distances between kernels was in a two-dimensional space, the two dimensions being the selected wavelengths. Rather than 241 possible wavelengths to test in the one-dimensional case, there were 241^2 possible combinations, each of which was tested as described above, and the combination with the highest classification accuracy was selected. Extension to the three-wavelength case is the same, with distances compared in three dimensions.

For the inshell product stream, clean inshell was used as the control, or “good” product, and the defects listed in table 2 were used as the “defect” product. Note that “small inshell” is considered a defect in the kernel processing stream and should not be confused with “clean inshell.” Each defect was analyzed versus clean inshell separately. For kernel processing streams, the “good” product comprised “kernel pieces,” “split kernels,” and “whole kernels,” and the defects are listed in table 1.

SORTER TESTING

A commercial, modular dual-band sorting machine (fig. 1c) was used (Sortex 3000, Sortex, Inc., Stockton, Cal.). The Sortex 3000 has three sensing modules per channel, so a total of six filters and three mirrors were obtained (CVI Laser, Albuquerque, N.M.), allowing the machine to sort at the optimal bands determined as described above for particular defect streams, which will be described below. The accuracy of the machine with the new filters was compared to the accuracy with the manufacturer’s recommended filters.

Sorter testing was performed for shell halves and small inshell product mixed with kernels. The testing was limited to these two classes of defect because replacement of the filters and mirrors is expensive and each defect type would require different filters and mirrors. In this case, it turned out that the optimal filters for these two defect types were the same, so two streams could be tested while purchasing only one set of new mirrors and filters.

Approximately 1 kg of kernels and 1 kg of defects (small inshell or shell halves) were used to train the Sortex 3000 following the manufacturer’s instructions. The training process involved feeding the two streams through the machine, which created a map of the feature space. At the completion of the training, the machine created a decision boundary for separation of the two groups. Subsequent samples were classified based on which side of the decision boundary they were mapped to. While it is possible to manually manipulate the decision boundary, for this experiment the boundary determined by the machine was maintained in every case, as manipulating the boundary would complicate the comparison of accuracy between the new bands and the manufacturer’s recommended bands. The same 1 kg samples of kernels and defects were used for training for both the new filter arrangement and the original.

After training, 1000 small inshell nuts were mixed with 1000 kernels and processed through the sorting machine. Under normal conditions, an air nozzle ejects the bad product. However, it was determined that there was a relatively high amount of misclassification that was the result of failed attempts to eject the product. The testing was therefore done by feeding the samples through one at a time and listening for the sound of the air ejector to determine how the samples were classified. The process was then repeated with 1000 half shells mixed with 1000 kernels. This was done first with the original mirrors and filters, and then with the new ones. The machine was retrained for each stream and each mirror-filter configuration.

RESULTS AND DISCUSSION

FILTER SELECTION

The classification results and wavelength combinations selected using the NIR spectra are listed in table 1 for kernels and table 2 for inshell product. The logic that facilitated the best accuracy is also listed. In all cases, the same logic gave the best accuracy for the different filter selections (one, two, or three bands). The results indicate that dual-band sorters should perform better than monochromatic sorters for most defects mixed with inshell and good kernels. Single-band classification of sticks and inshell nuts had a false negative error rate (defects classified as good) of 12.5%, which fell to 0% for dual-band sorting. Classification errors using dual-band for decayed kernels, immature kernels, shell halves, and small inshell kernels mixed with good kernels fell by at least 50% over single-band discrimination for these defects. Dual-band discrimination using two NIR bands over visible-NIR discrimination improved the accuracy for shell halves mixed with kernels. A hypothetical machine that can sense three different spectral bands may improve sorting of decayed kernels, immature kernels, and kernels with black spots, but none of the defects found in inshell product.

In all cases except immature kernels, the false positive rate (good product rejected as bad) was lower for dual-band classification than for single band. In general, false positive rates were much lower than false negative rates regardless of which combination of filters was used. This occurred because there was generally less variance in the spectra of the good kernels than in most defect streams. As discussed, classification decisions were made based on the distance of each sample from the good group and the defect group. No adjustment was performed to minimize false positive results. In certain cases, it could be desirable to weight the results to decrease the number of false positives, at the cost of higher false negatives. This would depend on the relative risks versus costs associated with the product streams. For example, if a particular defect were associated with a food safety issue, then it would be desirable to weight the results so that false negative results are minimized, at the expense of high false positives. In this study, classification was made to minimize the total error rate with no preference between false positive and false negative results.

It was confirmed, as expected, that OR logic gave the best results for black spots, as these defects usually can only be discriminated from one side. Similarly, shell halves are usually sorted with OR logic using the spectral bands of 1200 and 1450 nm (manufacturer recommended). Figure 7 shows a plot of the normalized Gaussian smoothed absorbance values of 1200 nm versus those of 1450 nm for shell halves (inside surface), shell halves (outside surface), and the two sides of split kernels. Note that the shell inside surface cannot always be separated from the kernels, resulting in reliance on the shell outside to distinguish these products. Consequently, approximately 2% of the shell is classified as kernels using bands at 1200 and 1450 nm. The results indicate that using AND logic and spectral bands at 1190 and 1350 nm resulted in a more accurate separation of shell halves and kernels (fig. 8). At these wavelengths, both halves of the shell separate well from both halves of the split kernels.

Table 1. Classification results for kernel defects and foreign material. Good product comprised “good kernels,” “pieces good,” and “splits good.”

| Kernel Defect | Logic | Best Single Wavelength | | | Best Pair: $\lambda_1 = 500 - 1050$ nm, $\lambda_2 = 1100 - 1700$ nm | | | | Best Pair: 500 - 1700 nm | | | | Best Triplet: 500 - 1700 nm | | | | |
|------------------|-------|------------------------|-------------------|-----------|--|------|-------------|-------------|-----------------------------|------|-------------|-------------|--------------------------------|------|-------------|-------------|-------------|
| | | fp ^[a] | fn ^[b] | λ | fp | fn | λ_1 | λ_2 | fp | fn | λ_1 | λ_2 | fp | fn | λ_1 | λ_2 | λ_3 |
| Decay | AND | 10.1 | 20.1 | 960 | 2.2 | 9.8 | 1045 | 1155 | 2.2 | 9.8 | 1045 | 1155 | 2.5 | 3 | 750 | 1430 | 1615 |
| Feeding | AND | 30.4 | 24.2 | 665 | 29 | 24.2 | 665 | 1205 | 17.9 | 30.7 | 1170 | 1365 | 10.2 | 19.8 | 1380 | 1515 | 1525 |
| Immature kernels | AND | 1.7 | 14.7 | 760 | 3.4 | 3.3 | 905 | 1445 | 2.8 | 3.3 | 1105 | 1445 | 1.1 | 1.6 | 710 | 1060 | 1455 |
| Small inshell | AND | 1.1 | 28.3 | 670 | 0.6 | 17.9 | 670 | 1440 | 0 | 2.7 | 1195 | 1345 | 0.1 | 2.2 | 1200 | 1350 | 1585 |
| Shell halves | AND | 1.2 | 3.6 | 1410 | 0 | 1.2 | 510 | 1425 | 0 | 0 | 1190 | 1350 | 0 | 0 | 505 | 1205 | 1345 |
| Black spots | OR | 22.3 | 40.0 | 1070 | 22.3 | 32.0 | 1050 | 1100 | 10.4 | 28.8 | 1075 | 1110 | 3.5 | 13.8 | 1030 | 1070 | 1095 |
| Insect webbing | AND | 10.6 | 34.8 | 910 | 10.1 | 27.7 | 530 | 1135 | 9.5 | 25.5 | 620 | 1005 | 8.4 | 23.9 | 670 | 915 | 1245 |

[a] False positive (fp): Good product classified as defect.

[b] False negative (fn): Defect product classified as good.

Table 2. Classification results for inshell defects. The good product comprised “clean open.”

| Inshell Defect | Logic | Best Single Wavelength | | | Best Pair: $\lambda_1 = 500 - 1050$ nm, $\lambda_2 = 1100 - 1700$ nm | | | | Best Pair: 500 - 1700 nm | | | | Best Triplet: 500 - 1700 nm | | | | |
|---------------------|-------|------------------------|------|-----------|--|------|-------------|-------------|-----------------------------|------|-------------|-------------|--------------------------------|------|-------------|-------------|-------------|
| | | fp | fn | λ | fp | fn | λ_1 | λ_2 | fp | fn | λ_1 | λ_2 | fp | fn | λ_1 | λ_2 | λ_3 |
| Adhering hull dark | OR | 1.3 | 9.0 | 790 | 1.3 | 7.0 | 855 | 1100 | 1.3 | 7.0 | 855 | 1100 | 1.3 | 7.0 | 835 | 950 | 1115 |
| Adhering hull light | AND | 5.1 | 44.7 | 525 | 0.0 | 31.6 | 995 | 1205 | 0.0 | 30.3 | 1400 | 1650 | 0.0 | 25.5 | 515 | 1460 | 1645 |
| Dark stain | AND | 12.2 | 50.0 | 705 | 2.5 | 31.8 | 785 | 1120 | 2.5 | 31.8 | 785 | 1120 | 1.3 | 21.6 | 750 | 1120 | 1510 |
| Light stain | AND | 5.1 | 60.2 | 1405 | 2.5 | 29.5 | 505 | 1400 | 2.5 | 29.5 | 505 | 1400 | 3.8 | 25.0 | 515 | 1405 | 1685 |
| Sticks | AND | 0.0 | 12.5 | 525 | 0.0 | 0.0 | 710 | 1410 | 0.0 | 0.0 | 710 | 1410 | 0.0 | 0.0 | 525 | 805 | 1400 |

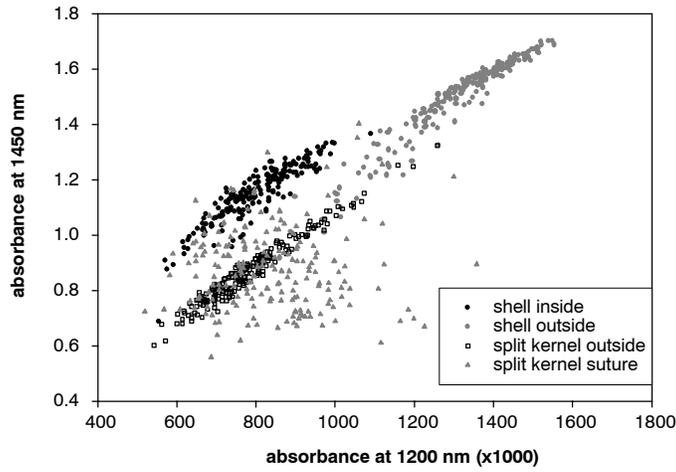


Figure 7. Scatter plot of spectral absorbance at 1200 and 1450 nm for split kernels and shell halves. Note that complete separation of shells and kernels cannot be made at these wavelengths.

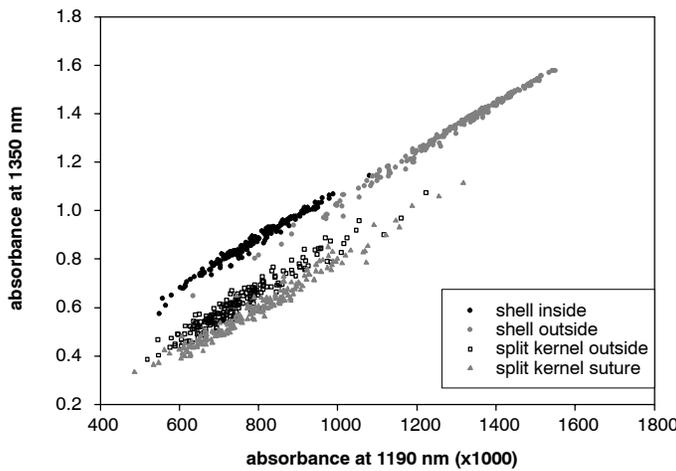


Figure 8. Scatter plot of spectral absorbance at 1190 and 1350 nm for split kernels and shell halves. Note the difference in absorbance values for the inside and outside of shells and the suture side and outside of split kernels.

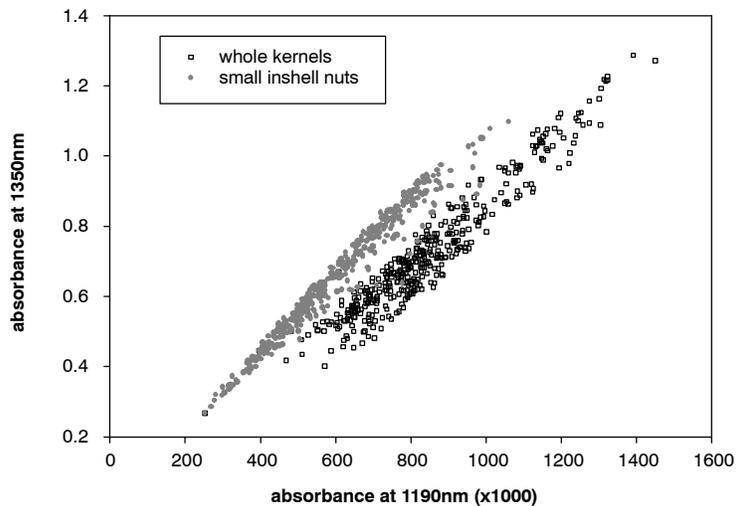


Figure 9. Scatter plot of small inshell nuts vs. whole kernels at the selected wavelengths of 1190 and 1350 nm.

For sorting small inshell mixed with kernels, nearly the same filters were chosen as for shell halves. Again, AND logic was more accurate than OR logic. Since small inshell nuts are all intact, the inner side of the shell is not exposed;

thus, AND logic is able to identify more small inshell with fewer false positive errors.

Figures 9 and 10 display scatter plots at the selected and manufacturer-recommended spectral bands for small inshell

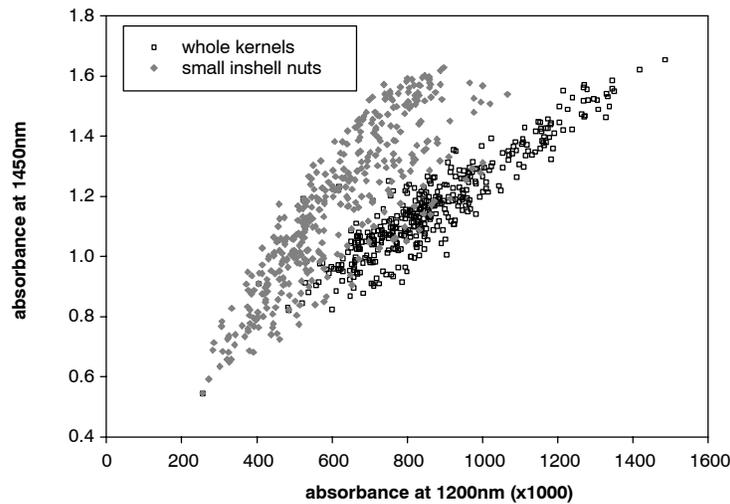


Figure 10. Scatter plot at the manufacturer-recommended spectral band of 1200 and 1450 nm for small inshell nuts and whole kernels. Note that many small inshell nuts are inseparable from kernels at these two bands.

and whole kernels, respectively. The accuracy at the selected bands of 1190 and 1350 nm was 100% for kernels and over 97% for small inshell. However, to achieve 100% accuracy for kernels at the manufacturer-recommended bands requires that over 10% of the small inshell nuts be classified as kernels. This inaccuracy can be seen on the scatter plots.

For a few defect types, results can be compared with those reported by Pearson et al. (2001) using imaging algorithms developed for high-speed image-based sorters. Pearson et al. (2001) reported 98% accuracy in detecting adhering hull defect, oily stain, and dark stain (individual results were not reported), with 1.4% fp and 2.3% fn. Clearly, these are better than results obtained with NIR in this study, probably since the imaging methods were able to sense the texture of some defects, whereas VIS-NIR sorters have no spatial resolution. The best results for adhering hull (dark) were 1.3% fp and 7.0% fn for one, two, or three wavelengths. For adhering hull (light), the best results were 0% fp with 25.5% fn. The best results for detection of dark stain were also disappointing, with 1.3% fp and 21.6% fn.

SORTING EXPERIMENTS

As stated above, the optimal spectral bands for use in the modular dual-band sorter for the separation of both half shells and small inshell nuts from kernels were 1350 and 1190 nm. Table 3 shows the sorting results for 1000 samples of each defect stream mixed with 1000 kernels. In separating shell halves from kernels, false negatives fell from 2.4% with the original filters to 1.8% with the new filters, while false positives were reduced from 0.7% to 0.1%. Note that the results are consistent with the roughly 2% false negatives predicted from the band selection process described earlier. For the small inshell / kernel stream, false negatives decreased from 1.7% with the original filters to 1.2% with the new filter setup, while false positives fell from 0.7% to 0.1%. The substantial decrease in false positives for both streams is particularly significant as this reduces the amount of higher-value product diverted to a lower-value stream. Results were analyzed using a χ^2 significance test; it was found that the classification results using the computed bands were significantly different from those using the manufacturer-recommended bands, with p-values of 0.01 for small inshell and 0.02 for half shells.

Table 3. Results of sorting shell halves and small inshell from kernels with a Sortex 3000 modular dual-band sorting machine. Each stream consisted of 1000 defect samples and 1000 kernels.

| | Original Bands (1200 and 1450 nm) | | Computed Bands (1190 and 1350 nm) | |
|---------------|--------------------------------------|------|--------------------------------------|------|
| | fp % | fn % | fp % | fn % |
| Shell halves | 0.7 | 2.4 | 0.2 | 1.8 |
| Small inshell | 0.7 | 1.7 | 0.1 | 1.2 |

It is important to note that the results reported here were obtained using the default mapping that the machine produces during the training process. This was done so that results could be compared without introducing the complication of manually manipulating the decision boundary. Since manual manipulation of the boundary generally improves the sorting results, the results reported here do not necessarily represent the best performance that can be obtained. Rather, these results represent a comparison between the two wavelength pairs under identical training conditions.

CONCLUSION

A method was developed to select the optimal spectral bands for sorting using the three most common types of commercially available NIR sorting machines. While the method was applied to the sorting of pistachio streams in this report, it should apply equally well to any product sorted using these machines. The method is particularly suited for use in modular dual-band sorting devices, as the mirrors and filters used to select the sorting wavelengths can be easily changed. The mirrors and filters required are relatively expensive, depending on the wavelengths involved. The mirrors obtained for this study cost in the neighborhood of \$1,000 US, but the price drops substantially as the quantity increases, since the majority of the expense involves setup costs for the fabrication of the mirrors. The filters cost approximately \$200 each. Therefore, for a two-channel Sortex 3000 with three modules per channel as used in this study, the cost of converting to the new wavelengths was approximately \$8,400 US. However, as stated above, conversion of multiple machines would reduce the cost per machine. This is a relatively small cost, compared to the price of a new

machine, which can range from \$40,000 US to around \$100,000 US depending on the model. Furthermore, improved sorting, especially the reduction in false positives as described earlier, should make conversion of these machines cost-effective over a relatively short period of time.

While only two processing streams (split shell / kernels and small inshell / kernels) were tested in this study due to cost constraints, optimal bands were determined and reported for most defects and foreign material commonly sorted in pistachio processing plants. This was done for the three most common types of commercially available NIR sorting machines (monochromatic, dual-band NIR / silicon, and modular dual-band NIR). It is anticipated that the improvement in sorting accuracy demonstrated for the two processing streams tested here would be repeated for other processing streams by simply changing the machine optics to those reported here.

Results indicate that addition of a third NIR band would improve sorting accuracy for most processing streams. Such machines are not commercially available at this time.

Finally, comparing results with those reported by Pearson et al. (2001) for detection of defects using visible light image-based sorters indicates that dual-band sorting with no spatial resolution is not as effective for detecting certain defect classes, including adhering hull and dark stain.

REFERENCES

Bajwa, S. G., P. Bajcsy, P. Groves, and L. F. Tian. 2004. Hyperspectral image data mining for band selection in agricultural applications. *Trans. ASAE* 47(3): 895-907.

- Caltepe, Z., T. C. Pearson, and A. E. Cetin. 2004. Identification of insect-damaged wheat kernels using transmittance images. *Electronic Letters* 41(5): 23-24.
- Dowell, F. E., T. N. Boratynski, R. E. Ykema, A. K. Dowdy, and R. T. Staten. 2002. Use of optical sorting to detect wheat kernels infected with *Tilletia indica*. *Plant Disease* 86(9): 1011-1013.
- Han, J., and M. Kamber. 2001. *Data Mining: Concepts and Techniques*. San Francisco, Cal.: Morgan Kaufmann.
- Lawrence, K. C., W. R. Windham, B. Park, and R. J. Buhr. 2003. A hyperspectral imaging system for identification of faecal and ingesta contamination on poultry carcasses. *J. Near Infrared Spec.* 11(4): 269-281.
- Lestander, T. A., R. Leardi, and P. Geladi. 2003. Selection of near-infrared wavelengths using genetic algorithms for the determination of seed moisture content. *J. Near Infrared Spec.* 11(4): 433-446.
- Mehl, P. M., K. Chao, M. Kim, and Y. R. Chen. 2002. Detection of defects on selected apple cultivars using hyperspectral and multispectral image analysis. *Applied Eng. in Agric.* 18(2): 219-226.
- Pasikatan, M. C., and F. E. Dowell. 2002. Evaluation of high-speed color sorter for segregation of red and white wheat. *Applied Eng. in Agric.* 19(1): 71-76.
- Pearson, T. C., M. A. Doster, and T. J. Michailides. 2001. Automated detection of pistachio defects by machine vision. *Applied Eng. in Agric.* 17(5): 729-732.
- Pearson, T. C., D. T. Wicklow, and M. C. Pasikatan, 2004. Reduction of aflatoxin and fumonisin contamination in yellow corn by high-speed dual-wavelength sorting. *Cereal Chem.* 81(4): 490-498.
- Steward, B. L., A. L. Kaleita, R. P. Ewing, and D. A. Ashlock. 2005. Genetic algorithms for hyperspectral range operator selection. ASAE Paper No. 053063. St. Joseph, Mich.: ASAE.

