Lebensm.-Wiss. u.-Technol., 32, 73–78 (1999)

Use of Near Infrared Transmittance to Automatically Detect Almonds with Concealed Damage

Tom C. Pearson

USDA-ARS-WRRC, 800 Buchanan Street, Albany, CA 94710 (U.S.A.)
(Received March 6, 1998; accepted September 1, 1998)

A near infrared inspection system was built to automatically detect whole almond kernels with concealed damage at an inspection rate of 40 nuts/s. The inspection device detects transmitted light through whole almonds from six different near infrared LEDs using a sine wave modulation–demodulation scheme. Multiple linear regression and discriminant analysis to classify nuts as concealed damaged or undamaged was performed. A classification error rate of 14.3% on the validation set was obtained with discriminant analysis, and a classification error rate of 20.0% was obtained with regression analysis. Most of the incorrectly classified nuts were those where it was difficult to objectively determine if they were damaged or undamaged.

Introduction

This study investigates the feasibility of an automated near infrared (NIR) detection system for almonds with concealed damage or internal browning, after drying but before roasting. For the reasons discussed in (1), almonds with concealed damage would preferably be detected after drying but before roasting. However, no internal browning occurs before roasting. Previous work (1) showed that after drying, nuts that absorb less light in the oil absorption band near 930 nm and absorb more light near 700 nm will develop concealed damage after cooking. This study examines the feasibility of using NIR light emitting diodes to detect these spectral features.

Materials and Methods

Prototype detection system design

The prototype detection system had to be able to inspect nuts at a rate of 40 nuts/s which is comparable to the rates of automated color sorters at almond processing plants. To achieve this rate with a single channel system, almonds need to travel in a single file stream at a speed of approximately 1.0 m/s. To maintain a sorting rate of 40 nuts/s, a maximum of 25 ms can be used to acquire the necessary information, process it, and activate an air nozzle to divert the nut from the stream if concealed damage is detected. A parallel study (1) showed that, by using the absorbance spectrum, first derivative spectrum, and second derivative spectrum from 700–1400 nm, almonds with concealed damage can be distinguished from undamaged almonds at an error rate as low as 12.4%. However, the technique used for this study required the use of the whole spectrum from 700–1400 nm. Low cost photodiode array spectrometers cannot currently acquire a full transmission spectrum of whole almonds in 25 ms. The prototype system measures transmitted light from six different light emitting diodes (LED), as shown in **Figure 1**. **Table 1** lists the part number and manufacturer of the six different LEDs used while **Figure 2** displays the actual normalized emission spectra from each of the LEDs measured by a spectrometer (Ocean Optics, #PC1000, Dunedin, FL, U.S.A.). The LEDs used for the prototype had peak emission wavelengths of 600, 830, 890, 940, and 950 nm. These six LEDs were chosen as they comprised all of the readily available LEDs in this wavelength range that had sufficient power to transmit light through a whole almond kernel. It should be noted that other LEDs are available that have different peak emission wavelengths but require bulk orders or very long lead times. The light intensity of each LED was sine wave modulated at different frequencies. A 50 mm diameter plano-convex lens (Edmund Scientific Co., Barrington, NJ, U.S.A., #E32,970) focused the light emitted from all the LEDs onto a single, 3 mm diameter, fiber optic cable, (Edmund Scientific Co., Barrington, NJ, U.S.A., #P38659) to diffuse the light and deliver it to the nut.

0023-6438/99/020073 + 06 $30.00/0
Article No. fstl.1998.0489

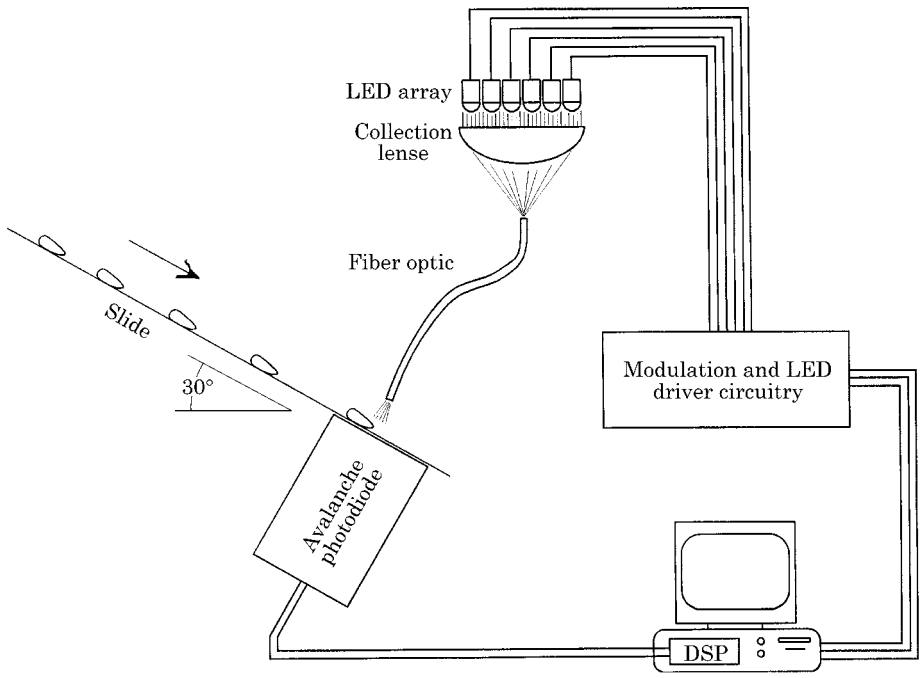

Fig. 1 Schematic of sorting machine

Table 1 Source of LEDs used for the prototype and modulation frequencies

LED peak wavelength (nm)	Modulation frequency (KHz)	Manufacturer part number	Manufacturer
660	9	SL660WCT3	UDT Sensors, Inc., Hawthorne, CA
830	12	L3989-01	Hamamatsu, Hamamatsu City, Japan
880	21	L2791-02	Hamamatsu, Hamamatsu City, Japan
890	24	L2690-02	Hamamatsu, Hamamatsu City, Japan
940	15	L2388-01	Hamamatsu, Hamamatsu City, Japan
950	18	LN54	Panasonic, Osaka, Japan

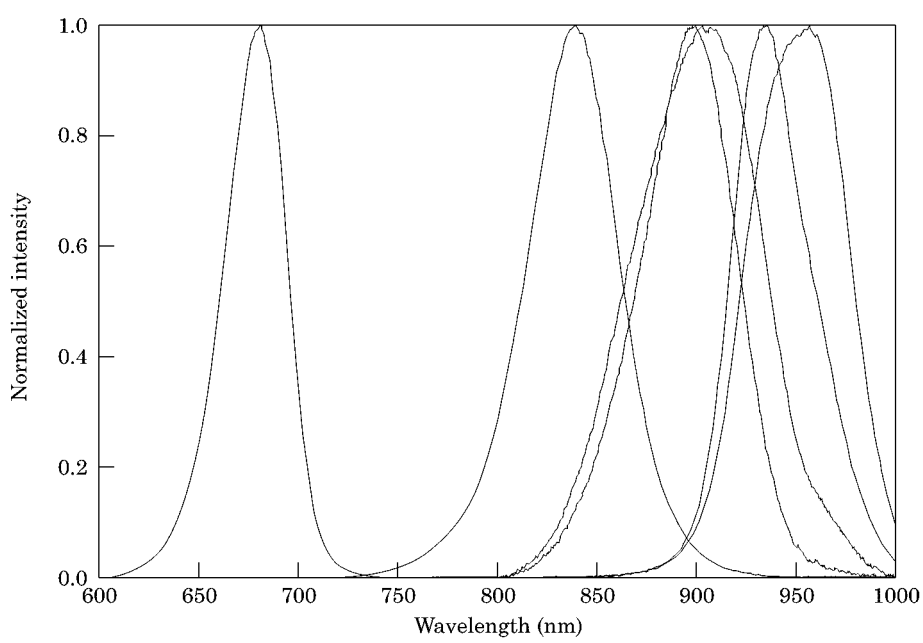

Fig. 2 Normalized emission spectra from the six LEDs used on the sorting machine. Spectra obtained after LEDs were burned in for 140 h

The transmitted light through the nut was detected by an avalanche photodiode module (Hamamatsu, Hamamatsu City, Japan, #C5460). The signal from the avalanche photodiode was input to a digital signal processing board (DSP) (Dalanco Spry, Rochester, NY, U.S.A., #310). The DSP board was equipped with a TMS320C31 digital signal processor, 14 bit analog to digital converter (A/D), and mounts into an ISA slot of a personal computer (60 MHz Pentium, Micron Electronics, Inc., Nampa, ID, U.S.A.). The DSP board performed discrete Fourier transforms to demodulate the input signal from the avalanche photodiode. The demodulated response from all LEDs was transferred to the PC, where the response from each LED was used to classify the nut as undamaged or concealed damaged.

The modulating sine waves to drive the LEDs were generated with precision, serially programable, sine wave generators (Micro Linear, San Jose, CA, ML2035). The sine wave generators output a sine wave at a frequency determined by an input clock frequency and a 16 bit digital word serially programed into the sine wave generator. A 3.6864 MHz TTL input clock (SaRonix, Palo Alto, CA, U.S.A., #S1500) was used, as this gave the sine wave generators a resolution of 0.44 Hz while allowing their maximum output frequency of 25 KHz. The 16 bit word to set the output frequency of a sine wave generator was supplied by a PC digital interface card (Keithly Metrabyte, Taunton, MA, U.S.A., PIO-24). Sine wave modulating frequencies of 9, 12, 15, 18, 21, and 24 KHz were used. These frequencies were chosen because their periods are all integer multiples of the sampling period, 1.67 ms. The DSP board was programed to sample 300 points at 180 KHz which requires 1.67 ms. Furthermore, the resolution of the discrete Fourier transform (DFT) with these sampling parameters is 600 Hz. Therefore, each LED modulation frequency coincides with a computed point in the DFT.

The magnitude of the sine wave driving the LEDs was from 2.5–7.5 volts. The minimum of a 2.5 volt differential across the LEDs was required to prevent to the LED from turning completely off and losing the sine wave shape of the emitted light intensity. The combined light from all LEDs transmitted through almonds caused the avalanche photodiode output signal to range from approximately 0.5–3.0 volts which corresponds to an A/D output of 1640–8200.

To minimize analysis time, the DFT of the avalanche photodiode output signal was computed for only the six modulating frequencies of the LEDs. This method allowed the computation of the six DFT points concurrently with the data acquisition. Equation [1] (2) shows the trigonometric form for computing the DFT of a signal

$$\tilde{F}\left(\frac{n}{NT_s}\right) = \sum_{k=0}^{N-1} [f(kT_s)w(kT_s) * \Delta'(kT_s)] * [\cos(2\pi nk/N) - J\sin(2\pi nk/N)] \quad \text{Eqn [1]}$$

where $\tilde{F}(n/NT_s)$ is the discrete Fourier transform at frequency n/NT_s, N is the total number of samples (300), T_s is the time interval between samples, n is an integer set to determine the frequency of the specific DFT point, $f(t)$ is the signal from the avalanche photodiode, $w(t)$ is the window function $\Delta'(t)$ is the sampling function, and k is an integer from 0 to $N-1$ defining the sample number in sequential order. The real and imaginary components of Eqn [1] can be computed each time a sample is acquired and a running sum of these quantities can be stored. Thus, upon acquiring N number of samples, the value of Eqn [1] only needs to be computed for $k = N-1$ and added to the running sum to obtain the DFT at frequency n/NT_s. The Hanning window value for each k, the $\cos(2\pi nk/N)$, and $\sin(2\pi nK/N)$ values for each k at each of the six modulation frequencies specified by n were loaded to the DSP memory and called by the DSP when needed. A Hanning window, rather than a rectangular window, was found to give a superior signal to noise ratio. The computation of a DFT began when the photodiode signal drops below a set threshold, indicating the presence of a nut. Twelve sequential DFTs were computed as the nut passed by the photodiode. This required approximately 20 ms, which is within the time frame allowed to inspect 40 nuts/s.

Prototype testing

The prototype was evaluated using a sample of 324 Mission almond kernels from the 1997 harvest. To induce concealed damage, nuts were exposed to the long moisture treatment as discussed in (1). Half of these nuts were then dried to their original bulk mass in an air convection dryer at 55 °C, and the other half dried at 110 °C as discussed in (1). These nuts and 81 control nuts were individually inspected by the prototype device. Nuts were inspected in separate batches of 81 nuts. Before and after inspecting a batch, LED light emission standards were measured by placing a 0.1% transmission neutral density filter (Ealing, Holliston, MA, U.S.A., #35-5941) between the photodiode and the fiber optic, and the photodiode signal was sampled as if a nut were present. The average of the two LED emission standards was used to normalize the absorbance for each nut in the batch. The LED emission intensities measured before and after each batch never deviated more than 1% from each other. The 12 DFTs obtained while the nuts were sliding by the photodiode were stored for analysis.

After obtaining the LED light transmission for all nuts, they were cooked, split at the suture, imaged, and their mean gray level computed as discussed in (1). As in (1), nuts with a mean gray level below 160 were considered to have concealed damage.

In addition to testing the prototype with almond samples, the consistency of the 12 DFTs were tested by sliding a small rectangular piece of Teflon through the device. When no nut is present between the photodiode and fiber optic, the photodiode is saturated by the unobstructed light from the LEDs. The step response

of the photodiode to go from a saturated state to accurately measuring the modulated LED signal when a nut suddenly blocks the light incident on the photodiode needs to be determined. As a nut is passed by the photodiode, 12 DFTs are taken. This, in effect, samples each LED 12 times as the nut passes by the photodiode. In an ideal case, the measured intensity of the 12 LED samples for each individual LED should be the same when the Teflon piece is inspected by the prototype. However, due to the frequency response limitations of the photodiode, this may not be the case. The main objective of this test is to determine whether the first few DFTs are equivalent to the remainder. The Teflon piece (McMaster-Carr, Los Angeles, CA., U.S.A., #873K13) was 6.35 mm thick, 38.1 mm long, and 12.5 mm wide. The piece of Teflon was passed through the prototype 60 times and the 12 DFTs acquired for each pass were saved in sequential order. This resulted in a data set containing 60 sets of 12 DFTs numbered, one to 12, in the order taken. Tukey's Studentized Range Test (3, 4) with $\alpha = 0.05$, was used to test the equivalence of the mean DFT value for each of the twelve sequential samples for each individual LED.

Data treatment

Each LED absorbance value was normalized by the mean absorbance values of all six LEDs. All normalized LED absorbance values, all possible ratios of normalized absorbance values, and all possible differences of two normalized absorbance values were computed. Redundant ratios and differences were not used, leaving a total of 15 each. Three sets of principle components were computed, one for the normalized absorbance values, a second set for ratios, and a third for the differences.

Prediction of concealed damage with discriminant analysis

A discriminant function was developed to classify nuts into one of two categories, concealed damaged or undamaged using principle components of the LED absorbance values, ratios of absorbance values, and differences in absorbance values. The discriminant analysis was performed with SAS proc discrim and proc princomp (3, 4). Principle components were selected for classifying nuts as concealed damaged or undamaged with stepwise discriminant analysis using a significance for entry and elimination from the model of 0.05. Covariance matrices were tested for equivalance using the pool = test option in SAS. The stepwise selection was trained using odd numbered samples only. The even numbered samples were used as a validation set. After performing discriminant analysis with all variables selected by the stepwise procedure, the least significant principle component, determined from the stepwise procedure, was eliminated and discriminant analysis was performed again. This was repeated until the error rate of the validation set reached a minimum.

Prediction of concealed damage with regression analysis

A prediction equation was developed with the mean image gray level of the cooked almond kernels as the dependent variable, while the independent variables were the principle components of the LED absorbance values, ratios of absorbance values, and differences in absorbance values. Principle components were selected based on their ability to predict the mean gray level by comparing the adjusted R^2 of all possible combinations of principle components. The odd numbered samples were used for calibration while the even numbered samples were used as a validation set. After model selection by the adjusted R^2 procedure on the calibration set, the last principle component, determined from the stepwise procedure on the validation set, was eliminated and regression was performed again. This was repeated until the standard error of prediction of the validation set reached a minimum. For classification, nuts were considered concealed damaged if their predicted mean gray level was less than 160. The range of mean gray levels for all nuts was from 10–250, with low gray levels indicating very severe browning.

For comparison purposes, regression analysis was also performed with the mean gray level as the dependent variable and the six absorbance, 15 ratio and 15 difference values as independent variables without transforming them into principle components. Regression models were selected by comparing the model adjusted R^2 of all possible combinations of ten or less variables. To speed the computation of adjusted R^2 for the models, the effect of multicollinearity was reduced by eliminating highly correlated ($|r| > 0.9$) independent variables. A correlation matrix for all independent variables was computed. If two independent variables were highly correlated ($|r| > 0.9$), then the variable with the lower correlation with the mean gray level was eliminated. When no independent variables remained with $|r| > 0.9$, then adjusted R^2 model selection was performed on the remaining variable set.

Results and Discussion

Repeatability of DFT values

From the tests of sliding the rectangular Teflon piece through the prototype, it was determined by Tukey's Studentized Range Test (3) that the mean intensity of the first two DFTs was significantly different at the 0.05 level than the mean intensity of the last ten DFTs. This was true for all six LEDs. The mean intensity of the last ten DFTs were determined equivalent at the 0.05 level. Due to these results, data from the first two DFTs were not used for the prediction of concealed damage. The average of the last ten DFT values was used for the prediction of concealed damage.

Classification based on discriminant analysis

Classification from a discriminant function is based on the posterior probability exceeding a set threshold. The false positives can be reduced by raising the posterior probability threshold for classifying a nut as concealed

Table 2 Error rates at different posterior probability thresholds for classifying nuts as undamaged or concealed damaged

Posterior probability of concealed damage threshold	False positive error rate (%)	False negative error rate (%)	Total error rate (%)
0.0	100.0	0.0	73.5
0.1	54.9	1.9	40.8
0.2	42.4	1.9	31.6
0.3	33.3	7.7	26.5
0.4	28.5	9.6	23.5
0.5	22.2	15.4	20.4
0.6	17.4	19.2	17.9
0.7	9.0	28.8	14.3
0.8	7.6	32.7	14.3
0.9	2.8	42.3	13.3
1.0	0.0	100.0	26.5

damaged. **Table 2** shows the false positive, false negative and total error rates of the validation set for a range of posterior probability threshold levels. The total error rate is the total number of incorrect classifications, divided by the total number of nuts in the validation set. Using a posterior probability threshold of 0.7, there are 9.0% false positives and 28.8% false negatives, giving a total error rate of 14.3%. Most, 76%, of the incorrectly classified nuts are on the border between actually being considered concealed damaged or undamaged as they have mean gray level values between 140 and 180. Recall that the division between classifying a nut as concealed damaged or undamaged was a mean gray level of 160. Almonds that show no indication of browning whatsoever have mean gray levels above 210, while almonds that are clearly concealed damaged have mean gray levels below 140.

Classification based on regression analysis

Three regression models were developed to predict the mean gray level of almonds after cooking. One regression model used principle components of the absorbance, ratios and differences of LED light absorbance values. Another regression model used the raw LED absorbance, ratio and difference values after removing highly correlated values. A third regression model comprised only the six normalized LED absorbance values. For each of these three models, the adjusted R^2 of the validation set and classification error rates on the validation set are listed in **Table 3**.

For the regression procedure using raw LED absorbance, ratio and difference values after removing highly correlated independent variables, only twelve of the original 36 independent variables were left after removing all the highly correlated independent variables. The remaining variables were: absorbance values from the 660, 830, 880, 890 and 940 nm LEDs, the ratio between 830 and 890 nm LEDs, and the differences between two LEDs having the following peak emission wavelengths: 660–880, 830–940, 830–950, 940–950, 950–880, 950–890. The model with the highest adjusted R^2, 0.35, contained the following eight variables: absorbance from the 660 nm LED, ratio between 830 and 890 nm LEDs, and differences between two LEDs having the following peak emission wavelengths: 660–880, 830–950, 830–950, 940–950.

The regression model comprising only the six LED absorbance values had an adjusted R^2 of 0.40. Dropping any one of the LED absorbance values resulted in large drop of the model adjusted R^2. The highest adjusted R^2 using only five LED absorbance values was 0.29.

Classifying nuts as concealed damaged or undamaged, based on the predicted mean gray level by principle component regression yields similar classification results to those obtained with the discriminant analysis procedure. The maximum adjusted R^2 for the regression model was 0.46 on the validation set. This was obtained using five principle components, two from the LED absorbance set and three from the ratio set. The total classification error rate was 20.0%. Only 9.1% of the undamaged nuts were classified as concealed damaged. Most of the errors arise from classifying concealed damaged nuts as undamaged, 39.4% of the concealed damaged nuts were incorrectly classified as undamaged. As with the discriminant analysis procedure, most of the classification errors involve nuts with mean gray levels near 160, the division between undamaged and concealed damage. Of all of the incorrectly classified nuts, 80% have a mean gray level between 140 and 180.

Since almonds with concealed damage are detected before roasting, the false positive error rates obtained through this study may be acceptable to almond processors. Nuts classified as having concealed damage can still be sold as a product that is not intended to be roasted. This would minimize the economic loss for the processor.

Conclusion

A prototype to automatically detect almonds with concealed damage that could operate at commercial speeds,

Table 3 Comparison of the three regression models used to classify nuts as concealed damage or undamaged

Model	Adjusted R^2	Total classification error (%)	False positive error rate (%)	False negative error rate (%)
Principle components	0.46	20.0	9.1	39.4
LED absorbance, ratio and difference values	0.35	22.9	11.2	43.8
LED absorbance values	0.40	22.1	10.7	42.6

about 40 nuts/s, was developed. The automated inspection device detected transmitted light through whole almonds from six different near-infrared LEDs. Each LED was modulated at a different frequency so that the light from all LEDs was transmitted through the almond at the same time. The demodulation of the signal from the photodiode was performed by a digital signal processor in real time, by computing a discrete Fourier transform at the six modulating frequencies. Multiple linear regression and discriminant analysis to classify nuts as concealed damaged or undamaged was performed on principle components of all normalized LED absorbance values, all possible ratios of two normalized absorbance values, and all possible differences in two normalized absorbance values. A classification error rate of 14.3% on the validation set was obtained with three principle components selected in a stepwise discriminant analysis procedure. A classification error rate of 20.0% was obtained with seven principle components selected in a stepwise regression analysis procedure. Most of the incorrectly classified nuts were on the border between actually being considered concealed damaged or undamaged. Furthermore, detection before roasting enables other uses for the almonds classified as concealed damaged, lowering the economic loss caused by the false positive errors.

Acknowledgements

The author would like to thank the Almond Board of California for partial funding of this project.
Mention of a specific product or vendor does not imply endorsement by USDA over other products which might be equally applicable.

References

1 PEARSON, T. C. Spectral properties and effect of drying temperature. Almonds with concealed damage *Lebensmittel-Wissenschaft und-Technologie*, **32**, 62–72 (1999)
2 BRIGHAM, E. O. *The Fast Fourier Transform and its Applications*. Englewood Cliffs, NJ: Prentice Hall (1988)
3 SAS INSTITUTE INC., *SAS / STAT User's Guide, Version 6, Fourth Edition, Volume 1*, SAS Institute Inc., Cary, NC. (1989)
4 SAS INSTITUTE INC., *SAS / STAT User's Guide, Version 6, Fourth Edition, Volume 1*, SAS Institute Inc., Cary, NC. (1989)