Journal of Food Engineering 126 (2014) 173–177

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

Internal damage inspection of almond nuts using optimal near-infrared waveband selection technique

Songyot Nakariyakul *

Department of Electrical and Computer Engineering, Thammasat University, 99 Moo 18 Phaholyothin Rd., Khlongluang, Pathumthani 12120, Thailand

ARTICLE INFO

Article history:
Received 27 August 2013
Received in revised form 4 November 2013
Accepted 19 November 2013
Available online 24 November 2013

Keywords:
Agricultural product inspection
Almond nuts
Feature selection
Near-infrared spectroscopy
Waveband selection

ABSTRACT

This work presents a statistical method for internal damage inspection of almond nuts based on advanced waveband selection and supervised pattern recognition techniques using near-infrared spectral data. Our proposed method employs an optimal adaptive branch and bound algorithm to select a small set of wavebands for use in a support vector machine classifier. Our case study involves discriminating almond nuts with internal damage from normal ones. Experimental results demonstrate that our method gives significantly higher classification rates than prior algorithms. Our classification model is promising for commercial online processing, since only a few wavebands are used for classification and can thus be recorded by many fast sensor systems.

1. Introduction

The almond, *Prunus dulcis*, is one of the top agricultural exports of many countries including the United States. One problem for the almond industry is the concealed damage in almonds, which is defined as a browning of the kernel interior after cooking or roasting (Reil et al., 1996). Fig. 1 shows images of a normal almond nut and an internally damaged almond nut after cooking (Pearson, 1998). Concealed damage may develop anytime during harvest when rain occurs or after harvest when the kernels are exposed to warm and moist environment (Kadar and Thompson, 1992; Reil et al., 1996). Internally damaged almonds are not easily distinguished from normal ones by their external appearances, since there are no visible defects on the exterior of the kernel before or after cooking. They cause lower consumer consumption because of their reddish-brown internal appearance, bitter flavor, and lower nutrition due to degraded amino acids after roasting and are thus not acceptable by the almond industry. Detection of internally damaged almond nuts with a high accurate rate is thus crucial for quality control.

Near-infrared spectroscopy has been widely investigated for agricultural product inspection, since it provides a noninvasive and accurate inspection system (Huang et al., 2008). It has been successfully applied in many agricultural applications. For example, near-infrared spectroscopy has been used for identifying anomalies and defects in food products such as chicken skin tumor detection (Nakariyakul and Casasent, 2007a), bunch withering disorder in date fruit (Mireei and Sadeghi, 2013), and contaminant detection on poultry carcass (Nakariyakul and Casasent, 2008; Windham et al., 2003). It has also been employed for food composition analysis including determination of soybean composition (Ferreira et al., 2013; Luna et al., 2013), assessment of meat quality in sliced chicken breasts (Grau et al., 2011), pork (Kapper et al., 2012; Liao et al., 2012), and salmon fillets (Kimiya et al., 2013), and measurement of fruit firmness (Jha et al., 2014; Li et al., 2013). These uses occur, since near-infrared spectroscopy accurately measures the transmission and absorption of near infrared light by organic compounds and water.

Prior work (Pearson, 1999) on internal damage detection in almonds showed that, by using the near-infrared spectrum ranging from 700–1400 nm, he could distinguish internally damaged almond nuts from normal ones before roasting at an error rate as low as 12.4%. Near-infrared transmission could detect changes in sugars or oil oxidation levels of almond nuts which in turn affect the concealed damage of almond nuts. However, the near-infrared transmission system that measures the full transmission spectrum of whole almonds is rather slow and cannot achieve an inspection rate of 40 nuts/s required by almond processing plants. Thus, prior related work (Casasent and Chen, 2003; Nakariyakul and Casasent, 2011) considered feature selection techniques to choose only a few significant wavebands to use for classification. Use of a few wavebands is attractive in practice because it leads to savings in measurement cost and speed; i.e., processing is faster and system cost is less. Casasent and Chen (2003) employed a new feature selection algorithm to select the best subset of wavebands for classification. First, they used the Kullback–Leibler distance (KLD) to

* Tel.: +66 2564 3001x3148; fax: +66 2564 3021.
E-mail address: nsongyot@engr.tu.ac.th

0260-8774/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jfoodeng.2013.11.018

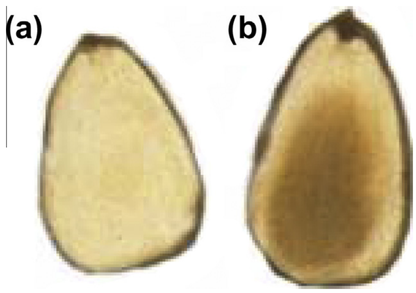

Fig. 1. Images of (a) normal and (b) internally damaged almond nuts after cooking (Pearson, 1998).

reduce the number of original wavebands from 137 to 30; they then applied the modified branch and bound (MBB) feature selection algorithm to select the subset of four and six wavebands from the resultant 30 wavebands. A nearest neighbor classifier was used as the classifier. We refer to this as the KLD–MBB algorithm. Nakariyakul and Casasent (2011) considered the use of ratio features (the ratio of the responses at two different spectral bands), since ratio features are invariant to multiplicative scaling (Keshava, 2004). The fast ratio feature selection algorithm (Nakariyakul and Casasent, 2011) was proposed to select two sets of ratio features to detect internally damaged almonds with a high accuracy rate of 91.2%.

In the present work, we propose an advanced feature selection technique to select the optimal subset of wavebands based on statistical properties of the training set data for detection of almonds with internal damage. The response of each sample at various numbers of selected wavebands is then fed to a support vector machine (SVM) classifier to identify whether it is internally damaged. We compare the classification results of our SVM-based method to those obtained using the wavebands and extracted features chosen by other feature selection and feature extraction algorithms and show that our proposed method gives excellent detection results with a low false positive error rate.

2. Materials and methods

2.1. Dataset and near-infrared transmission system

The database used in this work was provided by Dr. Tom Pearson from the Agricultural Research Service in Kansas, United States. Mission variety almonds were used because they are more prone to internal damage than other almond varieties. The transmission spectra of 454 almond nuts from 700–1400 nm were measured before cooking. The central region of each almond was illuminated by a 100 W quartz tungsten halogen lamp (Oriel, Stratford, CT, U.S.A.). Two different fiber optic transmission spectrometers were used to collect near-infrared spectra; a silicon photodiode array sensor based spectrometer (Ocean Optics, Dunedin, FL, U.S.A.) was used to measure the spectrum from $\lambda = 700–1000$ nm in 0.48 nm intervals, and an InGaAs photodiode array spectrometer (Control Development, South Bend, IN, U.S.A.) was used to obtain the spectrum from $\lambda = 950–1390$ nm in 3.2 nm intervals. For each nut, ten complete transmission spectra were measured, and the average spectra from each spectrometer were used. Each spectrum was then smoothed by a 19 point Savitzky–Golay second-order filter (Hruschka, 1987), sampled at $\Delta\lambda = 5$ nm increments and combined to produce a spectra with 137 spectral samples from $\lambda = 710–1390$ nm. For each almond sample, its spectral response was normalized by dividing its value by the mean of all values in the sampled spectrum to correct for variations in skin quality, nut thickness, and nut shape (Pearson, 1999). After the normalized spectra were obtained, the nuts were cooked at 135 °C for 90 min in a gravity convection oven (Lab-Line Instrument, Inc., Melros Park, IL, U.S.A.) to induce browning of the kernel interior and were then split at the suture for visual inspection. 109 out of 454 Mission almonds were found to have internal damage. All 454 samples are divided into training, validation, and test sets. Table 1 shows the number of normal and internally damaged samples in each set. This dataset was also used in prior work (Casasent and Chen, 2003; Nakariyakul and Casasent, 2011; Pearson, 1999).

Table 1
Number of normal and internally damaged almond nuts in the training, validation, and test sets.

Type	Training set	Validation set	Test set
Normal	139	34	172
Internally damaged	44	11	54
Total	183	45	226

2.2. Feature selection

Feature selection refers to search techniques that select a good small subset of features (wavebands) from an original larger set, where a statistical criterion function J is used to measure the performance of the selected subset. The branch and bound (BB) algorithm (Narendra and Fukunaga, 1977; Somol et al., 2004) is the only optimal feature selection algorithm that does not exhaustively search all possible combinations of subsets. The BB algorithm employs a bound to eliminate large subsets of possible candidates that are known to give lower bound values in the solution tree. It requires that the criterion function J used satisfy the monotonicity property, i.e., when a new feature is added to a feature set, the J value of the resultant feature set does not decrease. The Mahalanobis distance is used as the criterion function J in this work, since it is monotonic, computationally efficient, and widely used in many prior applications (Pearson, 1999). A subset with a large Mahalanobis distance indicates that the two classes are well separated and can be easily classified. A subset with a larger J value is thus better than one with a smaller J value.

To select the best set of m features out of n original features, the BB algorithm selects the $n–m$ features to be *discarded*. It creates a search tree with $n–m$ levels, where the root represents the set of all n features and the leaves represent all possible subsets of m features. Fig. 2 shows an example of the search tree when $n = 5$ and $m = 2$. As the search traverses down the tree, the J value decreases because more features are omitted. The problem is to find the best leaf (one at level 3 in our Fig. 2 example) with the largest criteria function J value.

The BB algorithm starts by analyzing all nodes at level 1, and the successor nodes (all nodes below a node) of the node with the largest J value are then evaluated further. The search continues until a leaf at the bottom of the tree is reached; this provides an initial bound B for the criteria function J. The algorithm then backtracks to any unexplored nodes at level 2 and, if necessary, those unexplored nodes at level 1. If the J value for a node is less than B, its successor nodes (leaves) at the bottom of the tree must have J values less than the bound B and cannot be the optimal subset. These nodes are omitted or cut off from the tree, and thus J need not be calculated for them. If a new leaf with a J value larger than B is found, the bound B is updated with this new larger J value. The search and backtracking continues until all leaf nodes in the tree are either explored or cut off. The BB algorithm is thus optimal.

Among many versions of the BB algorithms in the literature, the adaptive branch and bound (ABB) algorithm (Nakariyakul and Casasent, 2007b) is shown to be the fastest one for the cases when