Postharvest Biology and Technology 123 (2017) 22–32

Contents lists available at ScienceDirect

Postharvest Biology and Technology

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

Selecting optimal wavelengths for detection of insect infested tomatoes based on SIMCA-aided CFS algorithm

Seyed Ahmad Mireei[a,*], Samaneh Amini-Pozveh[a], Majid Nazeri[b]

[a] Department of Biosystems Engineering, College of Agriculture, Isfahan University of Technology, Isfahan 84156-83111, Iran
[b] Department of Laser and Photonics, Faculty of Physics, University of Kashan, Kashan 87317-53153, Iran

ARTICLE INFO

Article history:
Received 22 February 2016
Received in revised form 20 July 2016
Accepted 11 August 2016
Available online 20 August 2016

Keywords:
Wavelength selection
Transmission spectroscopy
Insect infestation
Soft independent modeling of class analogy
Correlation-based feature selection
Machine learning

ABSTRACT

Selection of effective wavelengths is a decisive step for online multispectral imaging systems. In this study, a new approach was utilized to distinguish the most informative wavelengths for detection of insect infestation in tomatoes within 400–1100 nm. Soft independent modeling of class analogy (SIMCA) was first conducted in the entire spectral region after applying different pretreatment procedures. Following satisfactory results obtained from 1st derivative preprocessing (accuracy of 90%), the most effective wavebands for detection of infestation were attained by discrimination power plot of SIMCA analysis. Transmission differences between all possible pairs of wavelengths ($T(\lambda_1)$–$T(\lambda_2)$) in the obtained informative wavebands were then calculated to substitute the 1st derivative spectra. Afterward, correlation-based feature selection (CFS) algorithm was used to find the best pairs of wavelengths. To compare the performance of SIMCA-aided CFS procedure, CFS was also conducted on the raw spectral data. Seven spectral difference features and six wavelength features were found superior by CFS. To classify tomatoes, three different machine learning techniques including Bayesian networks (BNs), artificial neural networks (ANNs), and support vector machines (SVMs) were implemented. The test set validation results of all machine learning techniques revealed that the spectral difference features outperformed the raw spectra features, indicating the superiority of SIMCA-aided CFS procedure for detection of optimal wavelengths. Among different machine learning techniques, the best performance obtained by ANN based on spectral difference features with a classification accuracy of 95.0%. The results of this study can be adapted for developing an online tomato sorting system for detection of infestations.

1. Introduction

Today, the production and quality of tomato (*Lycopersicon esculentum* Mill.) are highly influenced by insect infestations which cause severe damages in both protected and open fields. Infestations destroy the fruit by tunneling the flesh, leaving the galleries, and changing the color of the fruit. *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) is one of the key and problematic pests of tomato which originates from South America, but recently it has propagated to European, North African and Middle Eastern countries, including Iran (EPPO, 2005). The larvae of *Tuta absoluta* can enter through the stem-end, terminal end or other fruit parts that are in contact with contaminated fruits or leaves (USDA, 2011). The galleries bored inside the fruit may be later attacked by the secondary pathogens and causing the fruit rot (EPPO, 2005). Thereby, there is an increasing need to detect tomatoes with insect damage during postharvest handling and before the fruits are shipped to the market. Till now, the research on tomato pests mainly concentrated on management strategies such as the exclusion of insect, application of the chemical pesticide, sanitation, manual inspection, and biological control. No study has been performed so far on the possible methods for identifying the insect infestation in tomatoes.

Since the pest larvae usually produce the conspicuous small holes on the surface of the fruit, the common machine vision technique in the visible range can be a possible means of recognition. Nevertheless, this technique is sensitive to the surface features and may provide information regarding other types of surface damage that leave similar symptoms as holes. Therefore, machine vision has been considered to be unreliable for inspecting the insect infestation in fruits (Jackson and Haff, 2006; Wang et al., 2011; Xing and Guyer, 2008). On the other hand, the infestation can affect both chemical and physical features of fruit. The hemolymph, lipids, and probably chitin content of insect or larvae and

* Corresponding author.
E-mail address: samireei@cc.iut.ac.ir (S.A. Mireei).

http://dx.doi.org/10.1016/j.postharvbio.2016.08.009
0925-5214/

also dehydration of tissue can change the chemical features of the fruit (Rajendran, 2005), while the structural damages such as creating mine and galleries, inside darkening, and possibly fungi contaminations can influence the physical attributes of the fruit (Wang et al., 2011). These phenomena can practically affect the visible and near infrared (Vis/NIR) spectra of fruit and consequently, convince adopting the Vis/NIR spectroscopy as a suitable and nondestructive tool for detecting the infestation.

Meanwhile, the potential of Vis/NIR spectroscopy in nondestructive internal quality evaluation in fruits has been demonstrated by a great deal of research. This technique could successfully detect the internal disorders in apple (Khatiwada et al., 2016; Upchurch et al., 1997), kiwifruit (Clark et al., 2004), pear (Fu et al., 2007), and mangosteen (Teerachaichayut et al., 2011), and internal insect infestations in fig (Burks et al., 2000), cherry (Xing and Guyer, 2008), blubbery (Peshlov et al., 2009), mango (Thanapase et al., 2010), jujube (Wang et al., 2011), and chestnut (Moscetti et al., 2014). Regarding tomato, *Rhizopus Stolonifer* spores (Hahn et al., 2004), and mechanical tissue damage (Wu and Wang, 2014) have been successfully detected by Vis/NIR spectroscopy.

Selection of most informative wavelengths in the acquired spectral region is one of the main problems in the analysis of spectroscopic data. An appropriate selection of wavelengths can reduce the dimensionality and complexity of data, and simultaneously enhance the predictive ability of the model (Luo et al., 2012). Additionally, by reducing the analysis time, adopting the effective wavelengths can be also profitable for the model to be implemented in an online quality control system. Different wavelength selection procedures have been reported in the literature and they resulted in different prediction accuracies. For example, principal components analysis (PCA) (Xing et al., 2005), partial least squares (PLS) and stepwise discrimination analysis (ElMasry et al., 2008), correlation analysis between the wavelength ratio (Lee et al., 2008), and receiver operating characteristic (ROC) analysis between the wavelength difference (Luo et al., 2012) of reflectance were used to extract effective features for bruise detection in apple. Moreover, model-fitting procedure for detection of insect infested jujubes (Wang et al., 2011), analysis of area under the ROC curve for detection of flaws in hazelnut (Moscetti et al., 2013), and genetic algorithm for detection of insect infestation in tart cherry (Xing et al., 2008) and chestnuts (Moscetti et al., 2014) were successfully implemented for extracting the informative features from Vis/NIR spectrum.

In this paper, soft independent modeling of class analogy (SIMCA) was first utilized to investigate the feasibility of Vis/NIR technique for detect of insect infestation in tomato. Then, the wavelength variables with strong discriminatory powers were retained, while the weaker wavelengths were discarded. After calculating the spectral difference of wavelengths in the obtained wavebands, correlation-based feature selection (CFS) algorithm was performed to find the best pairs of wavelengths. The ability of the selected features for discrimination of tomatoes was then examined by different machine learning techniques. This approach is quite dissimilar from all of the aforementioned methods.

The objective of this study were (1) to evaluate the feasibility of SIMCA analysis in detecting intact and insect infested tomatoes from Vis/NIR spectra, and the effectiveness of SIMCA in identifying the informative spectral regions, and (2) to determine and compare the potential of CFS algorithm in extracting the effective wavelengths from the obtained informative spectral regions as well as raw spectral data using different machine learning techniques, including Bayesian networks (BNs), artificial neural networks (ANNs), and support vector machines (SVMs).

2. Materials and methods

2.1. Tomato samples

Tomatoes (cv. 'Halil') with different stages of ripening were collected from an open field located in the southeast of Isfahan, Iran between July and September in 2015. At the beginning, the intact and infested samples were distinguished by the visual inspection. However, since the most infested samples belonged to the middle and final stages of ripening, the intact samples were selected at the same levels of ripening to ensure that the color variation or ripening stage of the tomato does not influence the developed procedure. All the tests were conducted within 24 h of harvesting. Before spectral recording, each sample was first labeled and its morphological properties were measured. Following spectroscopic measurement on a whole tomato, the samples were cut into halves for correctly assigning each spectrum into intact and infested classes. Based on visual inspection, the samples with internal insect or larvae, mines and galleries, and discolored flesh related to infestation were regarded as infested. Finally, a total of 73 intact tomatoes and a further 92 samples with different insect infestation levels were selected for the experiments.

2.2. Spectral measurements

Fig. 1 illustrates the experimental setup for acquiring Vis/SWNIR spectra of tomato. The whole setup was covered with a polyurethane made tent to prevent ambient light and reflections being recorded. A photo-diode array (PDA) spectrometer (model: Spec-700, OPTC Inc., Iran) was used to acquire the spectral characteristics of tomatoes in half-transmittance mode. The spectrometer could collect the spectra of the samples in the range of 400–1100 nm with the wavelength resolution and dispersion of 2 nm and 0.5 nm/pixel, respectively. Moreover, the standard deviation (SD) of the absorbance spectra obtained from 50 repeated scans from the white reference was less than 0.05, indicating the high repeatability and consistency of the spectrometer in the obtained wavelength region. Because of high water content of fruits including tomato, it is difficult to pass the light through the entire fruit in the long wavelength NIR region (1100–2500). Thus, short-wavelength near infrared (SWNIR) spectroscopy in the transmission mode is suggested to be suitable for revealing the internal attributes of the fruit (Xing and Guyer, 2008). However, in this study, considering the intensity of radiation and the signal-to-noise ratio especially in the SWNIR region (800–1100 nm), half-transmittance measuring system was employed. For this purpose, six 50 W tungsten/halogen lamps were arranged around the equator of the sample and a single 200 μm fiber optic probe was positioned underneath the sample to acquire and pass the transmitted light to the spectrometer (Fig. 1). The light intensity of the lamps could be adjusted via a dimmer circuit. Before collecting the main spectra, the light intensity required to give the spectrum with acceptable signal-to-noise ratio was first determined using different samples. Then, the respective voltage value was recorded and used in the subsequent experiments. In order to calculate the relative transmission spectra of the samples and remove the optical interferences, the reference and dark spectra were recorded at the beginning each of 20 spectral measurements. During sample spectrum acquisition, the fruit was placed centrally on a fruit holder (containing the optical fiber) with a stem-calyx axis oriented horizontally. Two spectra were collected for each tomato by 180° rotating the sample around the stem-calyx axis and the average spectrum was used for further computations.