Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning

Nayeli Vélez Rivera [a], Juan Gómez-Sanchis [b], Jorge Chanona-Pérez [a], Juan José Carrasco [b], Mónica Millán-Giraldo [b,c], Delia Lorente [d], Sergio Cubero [d], José Blasco [d,*]

[a] Instituto Politécnico Nacional – Escuela Nacional de Ciencias Biológicas, Departamento de Ingeniería Bioquímica, Av. Plan de Ayala y Carpio s/n, Col. Santo Tomás, CP 11340, Mexico, D.F., Mexico

[b] Universitat de València – Intelligent Data Analysis Laboratory, Av. Universitat S/N, 46100, Burjassot, Valencia, Spain

[c] Universitat Jaume I – Institute of New Imaging Technologies, Av. Sos Baynat S/N, 12071, Castelló de la Plana, Spain

[d] Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Ctra. Moncada-Náquera km 5, 46113, Moncada, Valencia, Spain

ARTICLE INFO

Article history:
Received 26 September 2013
Received in revised form
19 February 2014
Accepted 27 March 2014
Published online 8 May 2014

Keywords:
Computer vision
Feature selection
Fruit quality
Mango fruits
Non-destructive inspection
Hyperspectral imaging

Mango fruit are sensitive and can easily develop brown spots after suffering mechanical stress during postharvest handling, transport and marketing. The manual inspection of this fruit used today cannot detect the damage in very early stages of maturity and to date no automatic tool capable of such detection has been developed, since current systems based on machine vision only detect very visible damage. The application of hyperspectral imaging to the postharvest quality inspection of fruit is relatively recent and research is still underway to find a method of estimating internal properties or detecting invisible damage. This work describes a new system to evaluate mechanically induced damage in the pericarp of 'Manila' mangos at different stages of ripeness based on the analysis of hyperspectral images. Images of damaged and intact areas of mangos were acquired in the range 650–1100 nm using a hyperspectral computer vision system and then analysed to select the most discriminating wavelengths for distinguishing and classifying the two zones. Eleven feature-selection methods were used and compared to determine the wavelengths, while another five classification methods were used to segment the resulting multispectral images and classify the skin of the mangos as sound or damaged. A 97.9% rate of correct classification of pixels was achieved on the third day after the damage had been caused using k-Nearest Neighbours and the whole spectra and the figure dropped to 91.4% when only the most discriminant bands were used.

* *Corresponding author*. Tel.: +34 96 3424000.
E-mail addresses: jchanona@ipn.mx (J. Chanona-Pérez), blasco_josiva@gva.es (J. Blasco).
http://dx.doi.org/10.1016/j.biosystemseng.2014.03.009
1537-5110/

1. Introduction

The mango (*Mangifera indica* L.) has become important as a product to be exported for several producer countries. The FAO reported that in 2011 Mexico was ranked 5th among top mango-producing countries and was also the world's number one exporter (FAOSTAT, 2011). Mexico produces more than 50 varieties of mango, 'Manila' being one of the most important for both the national and international markets (Ornelas-Paz, Yahiaa, & Gardea, 2008). The 'Manila' mango is a very soft fruit and it is therefore very susceptible to mechanical damage during postharvest processes. Mechanical damage to fruit results from surface loads, but the factors determining the different susceptibility or sensitivity of the fruit to mechanical damage from one species or variety to another is still unknown (Martínez-Romero et al., 2004), and its quantitative evaluation is still a challenge (Li & Thomas, 2014). Hahn (1999) reported on the effect of damage caused by impact on different varieties of mango and 'Manila' was one of the most sensitive, since this cultivar began to develop bruising after being impacted from heights of only 20 cm. One of the main problems in the postharvest period is that the damage caused during postharvest handling may not be visible until several days later, when the fruit has already been marketed. Any process, and consequently the reputation of the producer, is judged by the quality of the final product, which makes quality control a crucial stage of the process. Hence, this is a big problem in the mango industry because this lack of quality cannot be detected by visual inspection during postharvest handling, and damaged fruit with a bad appearance can have an important influence on the consumer's decision to choose a different kind of fruit. However, this situation becomes even worse when the fruit is exported because this is a priority for Mexican fruit and only fruits of excellent quality can be exported.

The detection of damaged fruit in the industry can be achieved by means of automatic systems (Valiente-González, Andreu-García, Potter, & Rodas-Jordá, 2014). Non-destructive methods such as spectrometry or image analysis in the visible and near-infrared range of the spectrum have been widely applied to evaluate aspects of quality and maturity in fruits and vegetables (Cubero, Aleixos, Moltó, Gómez-Sanchis, & Blasco, 2011), including mango fruit (Vélez-Rivera et al., 2014; Wanitchang, Terdwongworakul, Wanitchang, & Nakawajana, 2011). For instance, Hahn (2004) related mechanical damage with loss of firmness in 'Kent' mangos. The author induced the impacts with a pendulum device so as to be able to control the forces involved in causing the damage. In fact, controlled impacts are frequently used in studies for evaluating mechanical damage in olives (Jiménez-Jiménez, Castro-García, Blanco-Roldán, Agüera-Vega, & Gil-Ribes, 2012), oranges (Ortiz, Blasco, Balasch, & Torregrosa, 2011), apples (ElMasry, Wang, Vigneault, Qiao, & ElSayed, 2008), persimmon (Novillo, Salvador, Llorca, Hernando, & Besada, 2014), potatoes or tomatoes (Van Zeebroeck et al., 2003).

Nevertheless, various defects like spoilage, microbial contamination, mechanical and freeze damage or fungal infections are very difficult to detect at early stages using standard artificial vision systems (Gómez-Sanchis et al., 2013) and other technologies need to be developed. The advances made in hyperspectral imaging systems, together with the reduction in their price, have recently allowed this technology to be incorporated into postharvest quality control laboratories, which are using it to develop new advanced systems to predict the internal quality and properties of fruit and vegetables (Lorente, Aleixos, et al., 2012), including some types of mechanical damage (Nanyam, Choudhary, Gupta, & Paliwal, 2012). This technology combines the analysis of both spatial and spectral characteristics of a sample (ElMasry, Kamruzzaman, Sun, & Allen, 2012). As a result, it can therefore make use of the spectral information to estimate the internal compounds or also to detect diverse types of damage that cannot be detected by standard computer vision systems because the damage is not fully developed or its appearance is very similar to the sound skin, as is the case of decay lesions in citrus fruits (Blasco, Aleixos, Gómez-Sanchis, & Moltó, 2009) or internal properties in bell peppers (Schmilovitch et al., 2014). However, the complexity of the analysis of hyperspectral images requires a great deal of computing time. In addition, the huge amount of data provided by these systems makes it necessary to have reliable methods to select only the information that can actually optimise the results. Since a hyperspectral image consists of a set of consecutive narrow band monochromatic images, some information contained in close bands is redundant and frequently highly correlated. For this reason, it is very important to select only those bands with the most relevant information, while discarding those that do not contribute in any significant way to improve the results. Under this premise several studies have recently been carried out using different feature selection methods. For instance, Lorente, Aleixos, Gómez-Sanchis, Cubero, and Blasco (2013) and Lorente, Blasco, et al. (2012) compared several methods for selecting important wavelengths to detect decay lesions in citrus fruits at early stages. Serranti, Cesare, and Bonifazi (2013) used interval partial least squares discriminant analysis (iPLS-DA) on hyperspectral images to reduce their spectral dimension and redundancy, and to extract useful image features for differentiating three classes of wheat.

Once the relevant bands have been selected, a multispectral image is obtained, which then has to be segmented in order to achieve the real performance of the system. This segmentation consists of classifying each pixel in the multispectral image as belonging to a region of interest, using the spectral value of the pixels in the different selected bands as inputs of the classifiers (Gómez-Sanchis et al., 2012, 2013). Hence, the main objective of the experiments carried out is to study the possibility of detecting mechanical damage induced intentionally in 'Manila' mangos at early stages of maturity using a hyperspectral vision system, which is achieved by identifying the spectral bands that best categorise whether a mango is damaged or not. The work also aims to obtain the classifier that best fits this problem by segmenting the images using only the selected bands.

2. Material and methods

2.1. Fruit used in the experiments

Ten mangos (*Mangifera indica* L.) cv. 'Manila' were collected from a local market, all of them being similar in terms of stage of maturity (unripe), firmness, colour (dark green), shape and

size, as well as being free of blemishes or any external defects. The average weight of the fruit was 271.82 g ± 2.67. Before the experiments, the fruit was washed with a 1% solution of sodium hypochlorite for 10 min and were then rinsed with distilled water and dried at ambient temperature (Corkidi, Balderas-Ruíz, Taboada, Serrano-Carreón, & Galindo, 2006). Mechanical damage was induced in the fruit in the equatorial zone and each beaten area was marked so as to be able to identify the damaged area at a later stage. The damage to the fruit was created using a pendulum with a length of 242 mm. At the lower end of the pendulum arm there was a spherical aluminium impacting piece weighing 177 g. The impact energy is calculated as the kinetic energy of the pendulum arm just before impact. The energy applied to the fruit was 0.123 J when releasing the pendulum arm from an angle of 45° with respect to the vertical (Van Zeebroeck et al., 2003). Immediately after the fruit was damaged, the samples were stored under conditions of darkness at 25 °C and a relative humidity of 75%. Images of the mangos were acquired 1 h after causing the damage and then every 24 h for seven consecutive days.

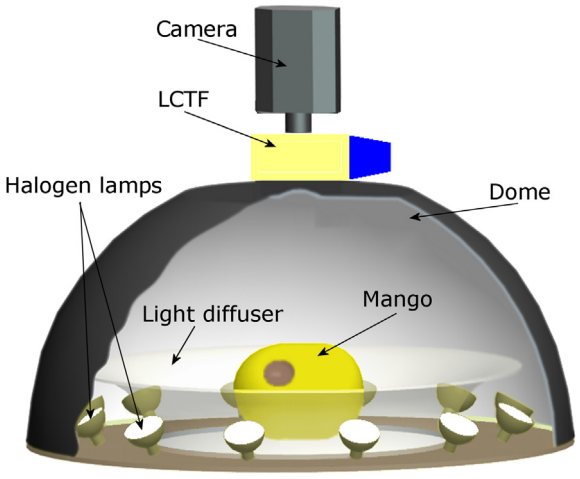

Fig. 1 – Diagram of the computer vision system used to acquire the hyperspectral images.

2.2. *Imaging system*

The hyperspectral vision system used to acquire the images consisted of a monochrome camera (CoolSNAP ES, Photometrics, USA) configured to acquire images with a size of 551 x 551 pixels and a resolution of 0.27 mm/pixel. The camera was coupled to a liquid crystal tuneable filter (LCTF) (Varispec NIR-07, Cambridge Research & Instrumentation, Inc., USA). The images were acquired in the range between 650 nm and 1080 nm with a resolution of 10 nm. The scene was illuminated by indirect light from twelve halogen lamps (20 W) arranged equidistant from each other inside a hemispherical aluminium diffuser so as to provide spatial uniformity. The inner surface of the dome was painted white with a rough texture to maximise its reflectivity and to light the scene indirectly by means of reflection, which better suits the shape of curved objects like mangos. Figure 1 shows the scheme of the imaging system.

The lamps were switched on for 25 min before each image acquisition to allow the radiation emitted by the lamps to stabilise. The images were acquired by placing the fruit inside the inspection chamber manually with the damaged area facing the camera. This system captured a hyperspectral cube composed of 44 monochrome images for each fruit, which gives a total of 440 monochrome images, each image containing sound and damaged areas. Halogen lamps offer a high degree of directionality, producing spatial variations in the intensity in the plane of the scene that were corrected by preprocessing the images before obtaining the labelled dataset. Moreover, the reflectance that is measured can be altered by the different spectral sensitivity of the elements in the hyperspectral system. To avoid this, a correction was carried out in the images using a white reference with a reflectance value of 99% (CSRT-99-050, Labsphere Inc, USA), in accordance with Equation (1) (Gómez-Sanchis et al., 2008).

$$\rho_{xy}(\lambda) = \rho_{ref}(\lambda)\frac{R_{xy}(\lambda) - R_{dark}(\lambda)}{R_{white}(\lambda) - R_{dark}(\lambda)} \quad (1)$$

where $\rho_{xy}(\lambda)$ is the image after applying this first correction, $\rho_{ref}(\lambda)$ is the reflectance of the white reference, $R_{white}(\lambda)$ is the monochrome image acquired by the hyperspectral vision system corresponding to the white reference, $R_{dark}(\lambda)$ is a dark image acquired by the system in the absence of lighting and $R_{xy}(\lambda)$ is the uncorrected image acquired by the system.

The background was later segmented using the image captured at 700 nm due to the high contrast between the fruit and the background, which makes it easy to separate the two regions using any threshold above a value of 50. Figure 2 shows the RGB (red, green, blue) image of one mango along with the monochromatic images in some examples of wavelengths.

2.3. *Labelled dataset*

A specific application programmed to manually select and store regions of interest (ROI) in hyperspectral images was used to select sound and damaged areas from the captured images. For each day of image acquisition, a total of 44 spectral features were stored for each pixel in the ROIs that were selected, these corresponding to the value of the pixel at each wavelength. In addition, the class the pixel belongs to was also stored as a classification variable. A total of 29,400 pixels were selected from the set of images captured each day, the datasets being divided into a training set (50%) and a test set (50%). The huge number of pixels in the test set was chosen in order to assess the generalisation capability of the statistical models. Figure 3 shows the average reflectance for sound and damaged samples in the labelled dataset.

2.4. *Feature selection methods*

To obtain several sets of important wavelengths with which to discriminate between sound and damaged areas (the damage still not being visible) various methods were studied and compared. These feature selection techniques were: Correlation-based Feature Subset Selection (CFS) (Hall, 2000), Chi Square (ChiS) (Miller & Siegmund, 1982), Fisher score (Duda, Hart, & Stork, 2001), Gini impurity algorithm (GIA)

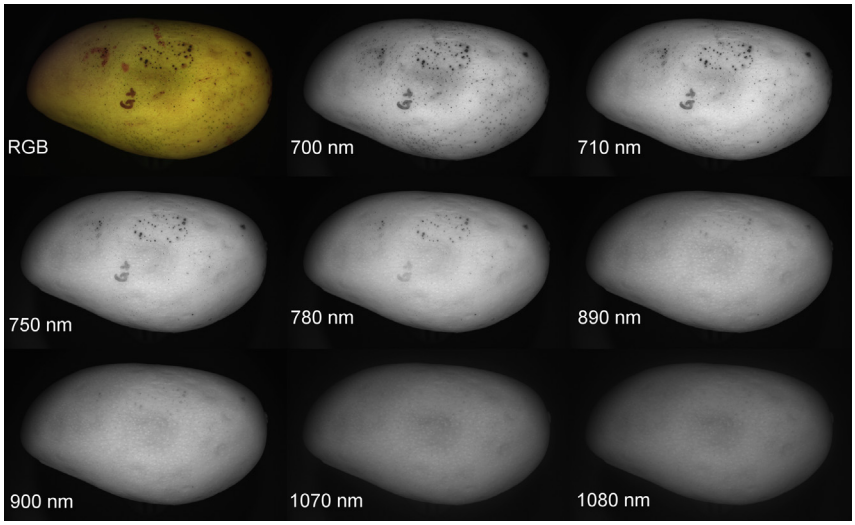

Fig. 2 – Example of RGB and monochromatic images of a mango sample captured at certain wavelengths.

(Breiman, Friedman, Stone, & Olshen, 1984), Information Gain (IG) (Guyon & Elisseeff, 2006), Minimum Redundancy Maximum Relevance (mRMR) (Peng, Long, & Ding, 2005), ReliefF (Kira & Rendell, 1992), Sequential Forward Selection (SFS) (Kohavi & John, 1997), Sparse Logistic Regression (SLR) (Cawley, Talbot, & Girolami, 2007), Stepwise (Draper & Smith, 1998), and Student's *T*-test (Larsen & Marx, 2005). Table 1 summarises the feature selection methods, showing the main properties of each of them.

To determine the best set of bands to detect the damaged areas, a method based on an expert committee (EC) was considered. The EC combines the selection of each individual method to provide a unique set of bands. The first six ranked bands given by each technique were considered as potentially eligible, but only those that were selected by at least three of the techniques were finally chosen. This process was done using the dataset of images corresponding to the third day after producing the damage, since optimum results were obtained this day using all the bands, as explained in the results section.

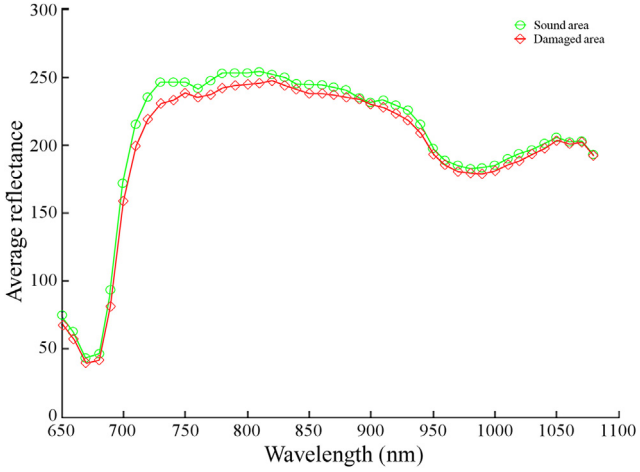

Fig. 3 – Averaged reflectance for the entire spectrum for sound and damaged samples of the labelled dataset: reflectance of the damaged area is represented by symbol ◇, reflectance of the sound area is represented by symbol ○

Table 1 – Summary of the feature selection methods used to decide the most discriminant wavelengths.

Method	Main properties
CFS	Ranks feature variables by giving more relevance to those highly correlated with the corresponding class while uncorrelated with the others
ChiS	Estimates the level of association between the respective class and each input feature in order to rank all features according to their relevance
Fisher score	Gives the correlation coefficient between the features and selects the top-*n* ranked features with larger scores
GIA	Calculates the impurity of features according to the categorization of instances by measuring how often a randomly chosen instance is misclassified
IG	Measures the worth of the feature in relation to the class
mRMR	Selects those features with the highest relevance to the target class and that are also maximally dissimilar to the rest of the classes
ReliefF	Ranks features according to their relevance taking into account how their values can distinguish between instances of the same and different classes that are close to each other
SFS	Selects a subset of features from the original set starting with an empty set, and adding features with the aim of maximising the prediction when they are combined with the previous ones
SLR	The negative log-likelihood function of the binomial distribution is minimised to obtain the Maximum Likelihood Estimation. A constraint is used to shrink the logistic regression model
Stepwise	Iterative method in which all features are evaluated for each iteration in order to sequentially add the most relevant one to the model
Student's *T*-test	Assesses variable significance and determines the probability of two classes being in relation with the feature analysed

Table 2 – Ranking of selected wavelengths according to each feature selection method.

Feature selection methods	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
CFS	710	720	730	740	1070	1080
ChiS	720	730	740	710	780	700
Fisher	730	720	740	780	710	770
GIA	720	730	710	740	700	780
IG	720	730	740	710	780	700
mRMR	730	740	1050	1040	1060	1030
ReliefF	1080	760	650	900	1070	660
T-test	1080	890	900	1070	1060	980
Stepwise	710	700	750	720	670	730
SLR	890	700	980	930	900	750
SFS	730	750	780	820	930	890

2.5. *Classifiers*

To segment the images, five classification methods were tested to determine which one provides the best results. They were used to classify the pixels as belonging to the sound or damaged areas. Firstly, they were used on the entire hyperspectral cube, employing all the spectral bands acquired by the system. Then, they were applied using only the six most important wavelengths selected in the previous step as input. Thus, in addition to obtaining the best combination of methods to select features and classify the fruit, the performance of the selected wavelengths can also be evaluated. The classifiers studied were Linear Discriminant Analysis (LDA) and *k*-Nearest Neighbours (*k*-NN), which are two classifiers widely used in machine vision and hyperspectral computer vision applications (McLachlan, 2004; Polder, van der Heijden, & Young, 2002); Naïve Bayes (NBC), which was chosen as a probabilistic approach based on Bayes' theorem (Becker, Kohavi, & Sommerfield, 2001); and Decision Trees (DT) and Extreme Learning Machine (ELM), because these classifiers have recently been used with good results to detect early stages of decay lesions in citrus using hyperspectral imaging (Lorente, Aleixos, Gómez-Sanchis, Cubero, and Blasco (2013) and Lorente, Blasco, et al. (2012)). In the case of *k*-NN, the *k* value was optimised and automatically computed according to the leave-one-out error on the training set, which gave $k = 5$. For the ELM algorithm, which is an optimised learning algorithm for the Single Layer Feedforward Network (SLFN) (Huang, Zhu, & Siew, 2006), the number of sigmoid hidden nodes was evaluated from 10 to 150 in steps of 10 nodes and the best structure was obtained when this had 100 hidden nodes.

3. Results and discussion

In this section results obtained by the classification models and feature selection techniques are provided only for the test set.

3.1. *Feature selection*

There were no single wavelengths at which it was easy to differentiate the damaged area of the mango from the surrounding sound skin. The results of the feature selection methods are shown in Table 2, where the most important bands are sorted from the most relevant to the least relevant (from left to right). It can be seen that all methods provide similar band sets except mRMR, ReliefF and *T*-test, which contribute with the upper bands of the studied spectrum. Although the match between the bands provided by different feature selection methods is not complete, there are certain areas of the spectrum in which an appreciable accumulation of bands can be detected.

Figure 4 shows the representation of the frequency (histogram) at which each wavelength is selected by the different methods employed, in accordance with Table 1. The eleven bands that have been selected by three or more methods, which are those that have been included in the EC feature selection method, are shown in solid bars. In this figure it can be seen that there is one main interval of bands (700 nm–780 nm) and two other minor intervals (890 nm–900 nm and 1070 nm–1080 nm). Most wavelengths are selected in the first range, which coincides with that observed in Fig. 2, where a major difference between the spectra of sound and damaged areas is observed around the range between 700 nm and 800 nm.

3.2. *Pixel classification*

Table 3 reports the accuracy obtained by the models evaluated for each day of storage in terms of rates of successfully

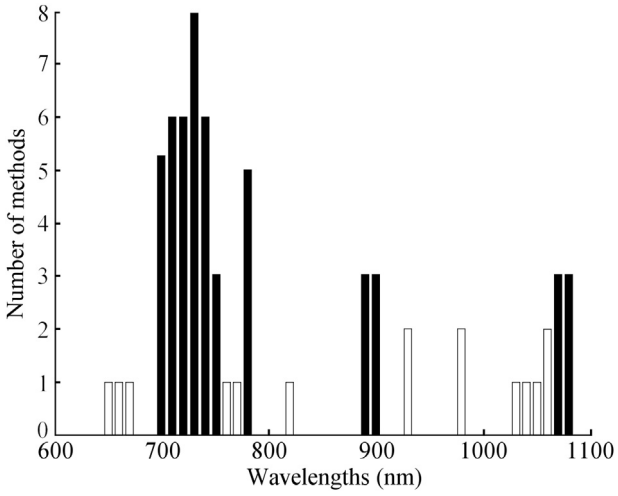

Fig. 4 – Histogram of the wavelengths selected by the feature selection methods. The solid bars are those bands that have been selected by at least three methods and included in the EC.

Table 3 – Accuracy obtained by the classification models using the whole spectrum acquired.

Accuracy (%)	k-NN	NBC	ELM	DT	LDA
Day 1	94.87	67.46	84.63	89.27	89.76
Day 2	95.03	67.54	84.75	89.12	90.59
Day 3	97.95	68.22	90.96	91.92	95.54
Day 4	98.03	71.03	94.91	95.00	95.99
Day 5	98.04	74.91	97.12	96.03	98.00
Day 6	98.13	75.60	97.08	96.20	98.02
Day 7	98.08	75.65	97.13	96.23	98.01

Table 4 – Confusion matrix of validation dataset provided by k-NN classifier using all captured bands and using only the bands chosen by the EC feature selection method.

Prediction/ Reality	All bands		EC bands	
	Damage (%)	No damage (%)	Damage (%)	No damage (%)
Damage	97.72	2.28	91.14	8.86
No damage	1.82	98.18	8.43	91.57

classifying the pixels of the mangos into the two predefined classes when all spectral bands were considered. The accuracy of the models was seen to increase over time, which is logical since the damage evolves. However, from the first day, k-NN achieves an accuracy of 94.9%, which gives an idea of the potential of this technique to detect this damage. While the best accuracy result is obtained with k-NN, the Naïve Bayes classifier provided unexpectedly poor results, which could be partially explained by the fact that this method relies on the assumption that, within each class, the probability distributions for attributes are independent of each other. However, this is not realistic since correlations among features can happen and this can degrade the Naive Bayesian classifier's accuracy (Langley & Sage, 1994).

As shown in Table 3, all the classification models except NBC achieved a relatively good level of accuracy from the first day, k-NN offering the best performance, which becomes stable from day 3 (97.95%) onwards. Moreover, the classification accuracy of this model did not increase significantly after this day. Hence, since one objective of this study was to detect the damage as early as possible, the feature selection was carried out using the data obtained on the third day of the experiments. Table 4 shows the classification results obtained using k-NN on the third day of the experiments, since this is the only method that performs well from the first day on. The results given were obtained using all bands acquired by the hyperspectral computer vision system and also only those chosen by the EC method. With average success of 97.9% and 91.4% respectively, the second approach provides poorer results than when using the whole range acquired, which is logical since the bands that are not considered may still contain information that is independent although not very relevant. For this reason, more effort is needed to improve the results while using even fewer bands in order to create a system that can be effectively used by the industry. Therefore, the present study contributes to a better selection of 'Manila' variety mango fruit by enabling the early detection of damage using a smart system, which is especially important for the export market. Figure 5 shows the segmented image in comparison to the image of the mango obtained at 900 nm in Fig. 2. Most errors correspond to isolated pixels that could be corrected using a median filter. On the other hand, some errors are found in the borders of the fruit due to the curvature of the surface of the mangos. These could probably be resolved by removing the borders of the mangos before applying the classifier, that is, by not considering the borders in the problem.

These results can be considered acceptable, since no currently available systems are capable of detecting this damage. Hence, this work lays the foundations for future implementation of an automatic system capable of detecting early damage caused by mechanical stress. In addition, these results are consistent with other recent works carried out using this technology for damage detection in other fruits. For instance, Nagata, Tallada, and Kobayashi (2006) identified up to five wavelengths to detect damage in strawberry in 87% of cases using stepwise linear discriminant analysis, while Liu, Chen, Wang, Chan and Kim (2006) obtained a success rate of 90% in detecting damage in cucumbers. Qin, Burks, Zhao, Niphadkar, and Ritenour (2011) achieved success rates of 95% in detecting canker in citrus fruits and Huang, Chen, Li, and Zhang (2013) obtained success rates of 96% in detecting mechanical damage in apples in both cases by selecting two wavelengths using PCA. In any case, the damage caused by impacts and the detection capability of non-destructive systems depends on the state of maturity of the fruit or the impact energy thresholds, as stated by Kitthawee, Pathaveerat, Srirungruang, and Slaughter (2011) in a study to correlate bruise occurrence with impact or compression in young coconuts. But, in these works, the damage was detected after it became clearly visible, while in our case, damage was not visible at the time of the experiments. However, to extend

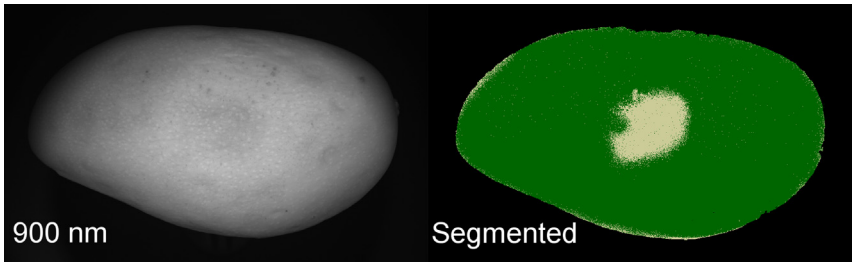

Fig. 5 – Segmented image in comparison to the one obtained at 900 nm.

the usefulness of the proposed system, it is necessary to conduct more studies with fruit at different stages of maturity and storage temperatures.

4. Conclusions

An NIR hyperspectral imaging system was used to detect mechanical damage induced in 'Manila' mangos by means of image analysis. Images were captured for seven days after the damage was produced in order to estimate the moment in which the damage could be effectively detected in the images. Using all the bands, it was concluded that after the third day the results were good for almost all classifiers, with *k*-NN achieving the best results.

In order to reduce the dimensionality of the data, eleven feature selection methods were used and the information combined in the so-called expert committee method, which only selected those bands that were repeatedly chosen by at least three different methods of selection. Results revealed three regions in the spectrum that was studied where there was an accumulation of selected bands: 700 nm–780 nm, 890 nm–900 nm, and 1070 nm–1080 nm.

In addition, five classifiers were evaluated to segment the images of the mangos into two classes: damaged and non-damaged. Among these classifiers, Naïve Bayes yielded very poor scores while the others – *k*-NN, ELM, DT and LDA – achieved scores above 90% for correct classification three days after the damage was produced. This percentage even rose to 95% for all these classifiers after the fourth day. The high performance achieved by *k*-NN on the third day (97.95%) leads us to use this model to segment the multispectral images consisting of only the selected features, with scores above 91% being achieved in the correct classification of sound and damaged areas.

Acknowledgements

This work has been partially funded by the INIA through project RTA2012-00062-C04-01, and by projects 20110627 and 20121001 at the IPN (SIPIPN-Mexico), 133102 (CONACyT) and Catedra CocaCola para jóvenes investigadores 2011 (Coca-Cola-CONACyT). Nayeli Vélez Rivera thanks CONACyT for the scholarship. Jorge Chanona-Pérez thanks CONACyT and the Secretaría Académica of the IPN for financial support for the sabbatical stay.

REFERENCES

Becker, B., Kohavi, R., & Sommerfield, D. (2001). Visualizing the Simple Bayesian Classifier. In U. Fayyad, G. Grinstein, & A. Wierse (Eds.), *Information visualization in data mining and knowledge discovery* (pp. 237–249). San Francisco (USA): Morgan Kaufmann Publishers.

Blasco, J., Aleixos, N., Gómez-Sanchis, J., & Moltó, E. (2009). Recognition and classification of external skin damages in citrus fruits using multispectral data and morphological features. *Biosystems Engineering, 103*, 137–145.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton (USA): Chapman Hall/CRC.

Cawley, G. C., Talbot, N. L. C., & Girolami, M. (2007). Sparse multinomial logistic regression via Bayesian L1 regularisation. *Advances in Neural Information Processing Systems, 19*, 209–216.

Corkidi, G., Balderas-Ruíz, K. A., Taboada, B., Serrano-Carreón, L., & Galindo, E. (2006). Assessing mango anthracnose using a new three-dimensional image-analysis technique to quantify lesions on fruit. *Plant Pathology, 55*, 250–257.

Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food and Bioprocess Technology, 4*, 487–504.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. New York, (USA): John Wiley Sons.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York, (USA): John Wiley Sons.

ElMasry, G., Kamruzzaman, M., Sun, D.-W., & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical Reviews in Food Science and Nutrition, 52*, 999–1023.

ElMasry, G., Wang, N., Vigneault, C., Qiao, J., & ElSayed, A. (2008). Early detection of apple bruises on different background colors using hyperspectral imaging. *LWT – Food Science and Technology, 41*, 337–345.

FAOSTAT. (2011). *Production quantities by country*. Retrieved March 21, 2014, from http://faostat3.fao.org/home/index.html#VISUALIZE.

Gómez-Sanchis, J., Blasco, J., Soria-Olivas, E., Lorente, D., Escandell-Montero, P., Martínez-Martínez, J. M., et al. (2013). Hyperspectral LCTF-based system for classification of decay in mandarins caused by *Penicillium digitatum* and *Penicillium italicum* using the most relevant bands and non-linear classifiers. *Postharvest Biology and Technology, 82*, 76–86.

Gómez-Sanchis, J., Gómez-Chova, L., Aleixos, N., Camps-Valls, G., Montesinos-Herrero, C., Moltó, E., et al. (2008). Hyperspectral system for early detection of rottenness caused by *Penicillium digitatum* in mandarins. *Journal of Food Engineering, 89*, 80–86.

Gómez-Sanchis, J., Martín-Guerrero, J. D., Soria-Olivas, E., Martínez-Sober, M., Magdalena-Benedito, R., & Blasco, J. (2012). Detecting rottenness caused by *Penicillium* in citrus fruits using machine learning techniques. *Expert Systems with Applications, 39*, 780–785.

Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. In I. Guyon, S. Gunn, M. Nikravesh, & L. Zadeh (Eds.), *Feature extraction foundations and applications* (pp. 1–28). Berlin (Germany): Springer.

Hahn, F. (1999). *Detección de la firmeza en mango usando un acelerómetro. [Mango firmness detection using an accelerometer]*. Technical Report CIAD/DUC/RT/17/99. Culiacán, Sinaloa, México.

Hahn, F. (2004). Mango firmness sorter. *Biosystems Engineering, 89*, 309–319.

Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)* (pp. 359–366).

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: a new learning scheme of feedforward networks. *Neurocomputing, 70*, 489–501.

Huang, W., Chen, L., Li, J., & Zhang, C. (2013). Effective wavelengths determination for detection of slight bruises on apples based on hyperspectral imaging. *Transactions of the Chinese Society of Agricultural Engineering, 29*, 272–277.

Jiménez-Jiménez, F., Castro-García, S., Blanco-Roldán, G. L., Agüera-Vega, J., & Gil-Ribes, J. S. (2012). Non-destructive determination of impact bruising on table olives using Vis–NIR spectroscopy. *Biosystems Engineering, 113*, 371–378.

Kira, K., & Rendell, L. (1992). The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 129–134).
Kitthawee, U., Pathaveerat, S., Srirungruang, T., & Slaughter, D. (2011). Mechanical bruising of young coconut. *Biosystems Engineering, 109*, 211–219.
Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.
Langley, P., & Sage, S. (1994). Induction of selective bayesian classifiers. In *Proceedings of the Tenth international conference on uncertainty in artificial intelligence* (pp. 399–406).
Larsen, R. J., & Marx, M. L. (2005). *Introduction to mathematical statistics and its application*. Upper Sadle River (USA): Prentice Hall.
Li, Z., & Thomas, C. (2014). Quantitative evaluation of mechanical damage to fresh fruits. *Trends in Food Science & Technology, 35*, 138–150.
Liu, Y., Chen, Y.-R., Wang, C. Y., Chan, D. E., & Kim, M. S. (2006). Development of hyperspectral imaging technique for the detection of chilling injury in cucumbers; spectral and image analysis. *Applied Engineering in Agriculture, 22*, 101–111.
Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., & Blasco, J. (2013). Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks. *Food and Bioprocess Technology, 6*, 530–541.
Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O. L., & Blasco, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology, 5*, 1121–1142.
Lorente, D., Blasco, J., Serrano, A. J., Soria-Olivas, E., Aleixos, N., & Gómez-Sanchis, J. (2013). Comparison of ROC feature selection method for the detection of decay in citrus fruit using hyperspectral images. *Food and Bioprocess Technology, 6*(12), 3613–3619.
Martínez-Romero, D., Serrano, M., Carbonell, A., Castillo, S., Riquelme, F., & Valero, D. (2004). Mechanical damage during fruit post-harvest handling: Technical and physiological implications. In P. Dris, & M. S. Jain (Eds.), *Quality handling and evaluation: Vol. 3. Quality handling and evaluation* (pp. 233–252). Dordretch, (The Netherlands): Kluwer Academic Publishers.
McLachlan, G. J. (2004). *Discriminant analysis and statistical pattern recognition*. New York, (USA): John Wiley Sons.
Miller, R., & Siegmund, D. (1982). Maximally selected chi square statistics. *Biometrics, International Biometric Society, 38*, 1011–1016.
Nagata, M., Tallada, J. G., & Kobayashi, T. (2006). Bruise detection using NIR hyperspectral imaging for strawberry (*Fragaria* × *ananassa* Duch.). *Environmental Control in Biology, 44*, 133–142.
Nanyam, Y., Choudhary, R., Gupta, L., & Paliwal, J. (2012). A decision-fusion strategy for fruit quality inspection using hyperspectral imaging. *Biosystems Engineering, 111*, 118–125.
Novillo, P., Salvador, A., Llorca, E., Hernando, I., & Besada, C. (2014). Effect of CO^2 deastringency treatment on flesh disorders induced by mechanical damage in persimmon. Biochemical and microstructural studies. *Food Chemistry, 145*, 454–463.
Ornelas-Paz, J., De, J., Yahiaa, E. M., & Gardea, A. A. (2008). Changes in external and internal color during postharvest ripening of 'Manila' and 'Ataulfo' mango fruit and relationship with carotenoid content determined by liquid chromatography–APcI+-time-of-flight mass spectrometry. *Postharvest Biology and Technology, 50*, 145–152.
Ortiz, C., Blasco, J., Balasch, S., & Torregrosa, A. (2011). Shock absorbing surfaces for collecting fruit during the mechanical harvesting of citrus. *Biosystems Engineering, 110*, 2–9.
Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Analysis and Machine Intelligence, 27*, 1226–1238.
Polder, G., van der Heijden, G. W. A. M., & Young, I. T. (2002). Spectral image analysis for measuring ripeness of tomatoes. *Transactions of ASAE, 45*, 1155–1161.
Qin, J., Burks, T. F., Zhao, X., Niphadkar, N., & Ritenour, M. A. (2011). Multispectral detection of citrus canker using hyperspectral band selection. *Transactions of the ASABE, 54*, 2331–2341.
Serranti, S., Cesare, D., & Bonifazi, G. (2013). The development of a hyperspectral imaging method for the detection of Fusarium-damaged, yellow berry and vitreous Italian durum wheat kernels. *Biosystems Engineering, 115*, 20–30.
Schmilovitch, Z., Ignat, T., Alchanatis, V., Gatker, J., Ostrovsky, V., & Felföldi, J. (2014). Hyperspectral imaging of intact bell peppers. *Biosystems Engineering. Special Issue: Image Analysis in Agriculture, 117*, 83–93.
Valiente-González, J. M., Andreu-García, G., Potter, P., & Rodas-Jordá, A. (2014). Automatic corn (*Zea mays*) kernel inspection system using novelty detection based on principal component analysis. *Biosystems Engineering. Special Issue: Image Analysis in Agriculture, 117*, 94–103.
Van Zeebroeck, M., Tijskens, E., Van Liedekerke, P., Deli, V., De Baerdemaeker, J., & Ramon, H. (2003). Determination of the dynamical behaviour of biological materials during impact using a pendulum device. *Journal of Sound and Vibration, 266*, 465–480.
Vélez-Rivera, N., Blasco, J., Chanona-Pérez, J. J., Calderón-Domínguez, G., Perea-Flores, M. J., Arzate-Vázquez, I., et al. (2014). Computer vision system applied to classification of 'Manila' mangoes during ripening process. *Food and Bioprocess Technology, 7*, 1183–1194.
Wanitchang, P., Terdwongworakul, A., Wanitchang, J., & Nakawajana, N. (2011). Non-destructive maturity classification of mango based on physical, mechanical and optical properties. *Journal of Food Engineering, 105*, 477–484.