Computers and Electronics in Agriculture 146 (2018) 12–21

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Pixel based bruise region extraction of apple using Vis-NIR hyperspectral imaging

Wenkai Che, Laijun Sun*, Qian Zhang, Wenyi Tan, Dandan Ye, Dan Zhang, Yangyang Liu

Key Laboratory of Electronics Engineering, College of Heilongjiang Province, Heilongjiang University, Harbin 150080, China

ARTICLE INFO

Keywords:
Bruise of apple
Hyperspectral imaging
Principal component analysis
Pixel based classification
Random Forest

ABSTRACT

Bruises on apples will directly influence its preservation and marketing for they can cause the internal decomposition and flaws of the appearance of apples. Therefore, an effective pixel based bruise region extraction method was proposed in this study to obtain the complete bruise region. Hyperspectral images of 60 apples were obtained via the hyperspectral imaging (HSI) system at 0, 12 and 18 h after the damage experiment. Principal Component Analysis (PCA) was used to compression data size and eliminating redundant data of hyperspectral image cubes. After the selection of the region of interest (ROI) by certain rules, different pixel based apple bruise extraction models were built and compared. The result shows that Random Forest (RF) model have a high and stable classification accuracy, which turns out that RF algorithm is more suitable for classifying bruises on apples than others. The average accuracy of bruise extraction models reached 99.9%. Compared with the most used image processing method in recent literature for extracting bruises of apples, the bruising region predicted by RF model was more consistent with the true bruise region. Additionally, two characteristic wavebands around 675 nm and 960 nm related to the bruise region were singled out for reducing the dimensionality of data by analyzing the feature importance scores of the built RF model. The overall results indicated that the proposed method has a great potential to detect complete bruise region on apples based on hyperspectral imaging for improving the efficiency of apple grading and sorting.

1. Introduction

Apples as one of the most popular fruit in the world are prone to mechanical bruise due to the impact, squeeze and abrasion during picking and transportation. The bruise can not only influence the apple's taste and causes its nutrient loss, but also accelerate its decay velocity and reduce the economic value of the apple (Xing et al., 2007). Hence, bruise on apples is a critical factor in the assessment and grading of apple quality, which can directly influence its preservation and marketing (Li et al., 2016). There are still considerable factories using manual fruit grading method to detect bruises and classify fruits (Baranowski et al., 2012). However, manual fruit grading is a tedious, inefficient work, and the sort result is also less than satisfactory (Pu et al., 2015). For enhancing the efficiency and the accuracy of bruise detection, researchers have explored various green methods to detect the bruise region or the existence of bruise on apples. A setup for bruise detection and grading was developed based on machine vision with the correlation between predicted and measured bruised areas ranging from 0. 63 to 0.84 (Rehkugler and Throop, 1986); Jackson and Harker (2000) detected bruises on apple by measuring the electrical impedance of apple, which achieved the value of R^2 up to 0.71; An X-ray radiographic imaging technique for detecting internal defects in apples was investigated by Schatzki et al. (1997). Unfortunately, the detection accuracy of this technology was lower than 50%, which cannot be used to replace manpower. All methods mentioned above can detect the bruise to varying degrees, which provided various ways for detecting and classifying bruises on apple automatically.

With the development of the HSI technique, more and more studies have been reported on using HSI for the quality and safety inspection of food products (Lorente et al., 2012; López-Maestresalas et al., 2016; Ariana et al., 2006). HSI data is a 3-dimensional cube which contains both spatial (2-dimensional) and spectral (1-dimensional) information of an object. Every pixel in spatial dimension corresponds to a spectrum, meanwhile, all data under a wavelength in spectral dimension can compose an image(Ferrari et al., 2015). HSI technique is frequently used for detecting apple quality for its merits of nondestructive and chemical-free determination. Lu (2003) analyzed hyperspectral images whose wavelength ranged from 900 nm to 1700 nm for predicting the

* Corresponding author at: Room 505, Building A8, Heilongjiang University, No. 74, Xuefu Road, Nangang District, Harbin 150080, China.
E-mail addresses: 2151321@s.hlju.edu.cn (W. Che), 2006036@hlju.edu.cn (L. Sun), 125209217@qq.com (Q. Zhang), 1228348692@qq.com (W. Tan), 850242749@qq.com (D. Ye), 924728661@qq.com (D. Zhang), 1905461824@qq.com (Y. Liu).

https://doi.org/10.1016/j.compag.2018.01.013
Received 26 July 2017; Received in revised form 14 December 2017; Accepted 14 January 2018
Available online 03 February 2018
0168-1699/

time after bruising (1-47d) on two cultivars of apples ('Red Delicious' and 'Golden Delicious') via principal component and minimum noise fraction transforms. The correct detection rate of this approach was from 62% to 88% for red delicious apples and was from 59% to 94% for golden delicious apples. Xing et al. (2005) obtained hyperspectral images of 'Golden Delicious' apples with the spectral region from 400 nm to 1000 nm. For reducing the dimension of HSI data, Principal Component Analysis (PCA) was applied to characteristic waveband selection. A classification algorithm for detecting bruises on apples based on moments thresholding was developed subsequently, and its accuracy reached 86%. In addition, partial least squares method and stepwise discrimination analysis had been studied by ElMasry et al. (2008) for selecting three effective wavelengths (750, 820, 960 nm). Afterwards, the bruised region was obtained by adaptive thresholding method with the averaged image of characteristic images. What's more, Baranowski et al. (2013) detected the bruised region by image thresholding method and the averaged image of images at wavelengths from 680 nm to 960 nm. Various pretreatment methods and classification methods were used and compared for selecting the most accurate model. The research showed that the prediction accuracy was higher than 90% when the best model was built with the combination of second derivative pretreatment method and the linear logistic regression neural networks method. The Bruise susceptibility of apple was studied by Zhu et al. using hyperspectral scattering image ranging from 500 to 1000 nm (Zhu et al., 2016). To get bruised regions, image segmentation method was implemented through Adobe Photoshop CS (San Jose, CA). Keresztes et al. (2017) detected stem, glossy, sound and bruised regions on apples via partial least squares-discriminant analysis (PLS-DA) algorithm. An average bruise detection accuracy reached 94.4% after the post-processing of the predicted binary images.

Image processing method such as thresholding and feature vectors extracting method is one of the most popular methods of most of the existing studies to obtain bruises on apples (Xing et al., 2007; Tian et al., 2014). Nevertheless, considering the influence of the individual diversity and the variety of environment, the use of image processing methods is limited in some circumstances. For instance, bruised region is easily mis-segmented when the surface of apple exists mixed colors or the bruise has a complex shape and a tiny region (Keresztes et al., 2016). Beyond that, it is hard to evaluate the accuracy of the bruise detection with image processing methods since the difficulty in counting the correct selected bruise pixels. From the application point of view, the accurate prediction of bruised regions can be used for estimating the area of the bruise, which is a critical index to grade different levels of apple quality. Hence, an accurate and stable method capable of segmenting bruise on apples is needed.

The overall goal of this research was to find a non-destructive, fast and stable method which can accurately predict the bruised regions on apples. The specific targets were to:

(1) Develop a pixel based bruise extraction model in view of spectral classification for overcoming drawbacks of the traditional bruise segmentation method (image processing method).
(2) Evaluate the accuracy of the bruise extraction method proposed in this study.
(3) Determine several optimal wavebands instead of full bands for reducing dimensions of large-scale HSI image data.

2. Material and method

2.1. Sample preparation

60 fresh apples with a diameter of 7–8 cm, of Fuji variety, were manually selected and purchased in Harbin, Heilongjiang Province, China in September 2016. For obtaining a better apple samples, apples without obvious bruises and different surface colors were given higher priority during visual inspection. After being shipped to the laboratory, apples were numbered and stored at the room temperature of 18 °C and relative humidity of 75% until the experiment was conducted.

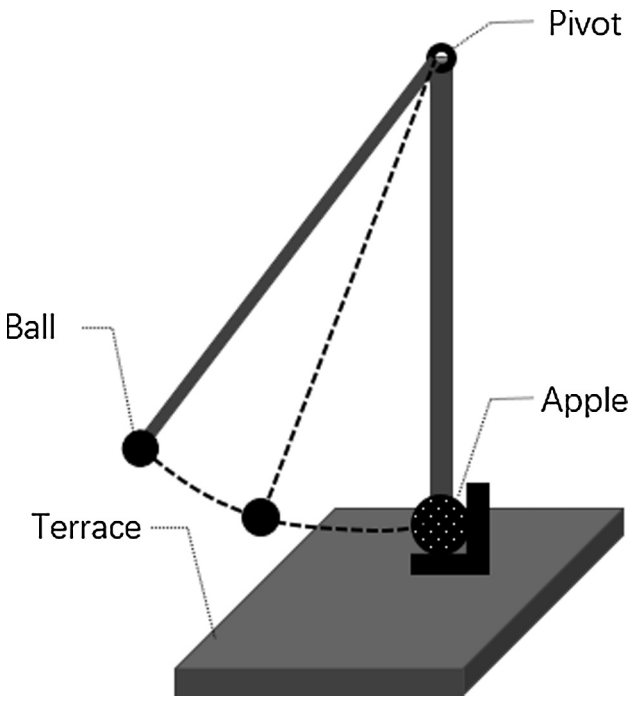

Fig. 1. Apple damage equipment.

2.2. Bruising experiment

Fig. 1 illustrates the design of the equipment for bruising apples. An evenly distributed steel rod is 80 cm in length and 230 g in weight, which connects a pivot to a steel ball whose diameter is 1.6 cm and weight is about 22 g. For controlling the variability during the experiment, the steel rod was set at a fixed angle (57 degrees measured by goniometer) to vertical. Tested apples were fixed at the bottom of the equipment while the steel rod and ball fall onto the apple's equator after toggling the switch, and bruised region with a diameter of approximately 1.4 cm would be caused. Since the angle between the rod and perpendicular line and the height from the ball to the ground was fixed, the force acting on the surface of apple was almost equivalent at each collision, hence, the consistent data was generated in the end.

Apples in this experiment were hit sequentially in ascending order by the sample number. The hyperspectral image was obtained immediately after the collision via the HSI acquisition system. Since the bruise might occur at any time in course of harvest, transportation and storage, hyperspectral images at 12 and 18 h after bruising were also collected for simulating apples bruised at the different time. There were 180 raw hyperspectral images acquired at last.

2.3. Acquisition of hyperspectral images

HSI system which is used to obtain the hyperspectral images consists of: (1) A hyperspectral camera (SOC710, Surface Optics Corp., San Diego, CA, USA); (2) Illumination unit contains four 75 W angle-adjustable tungsten halogen lamps (PHILIPS, USA) around the box; (3) Computer with data acquisition software HyperScanner2.0. As represented in Fig. 2, the hyperspectral images acquisition system keeps the inner environment consistent in each shooting for reducing the interference from outside. The spectrograph covers the spectral range between 400 and 1000 nm (include the visible and near infrared ranges) with a resolution of 4.68 nm, 128 bands. There are 520 × 696 pixels in one band, and the scanning velocity is 30 lines per second. The

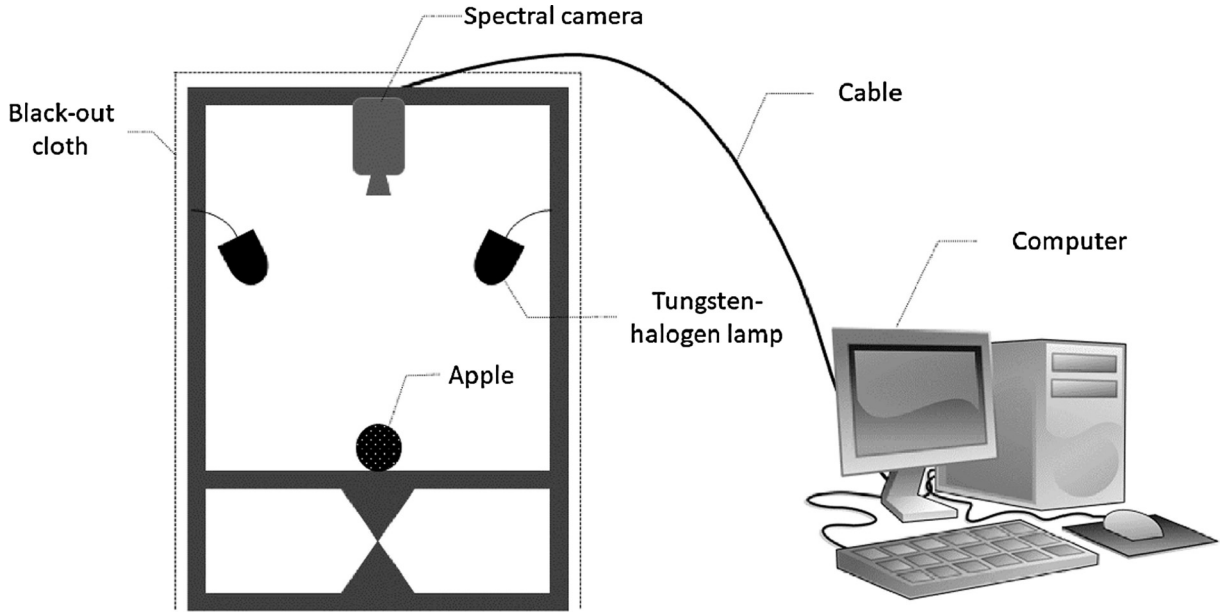

Fig. 2. Hyperspectral imaging acquisition system.

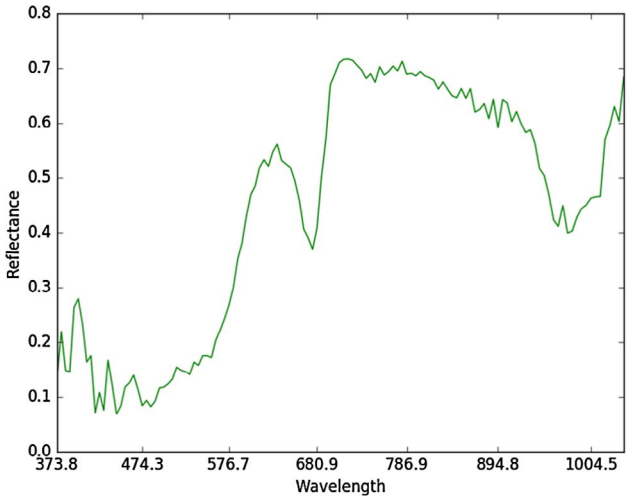

Fig. 3. The calibrated reflectance spectral of apple.

hyperspectral camera supports an internal scan mode which can obtain a complete HSI image automatically without additional hardware like the linear transportation system. Since the dynamic range of data is 12-bits, the value of data ranges from 0 to 4000. The spectral value kept between 20% and 80% (800 to 3200) of the data range via adjusting the exposure time and the angle of four tungsten halogen lamps. The whole HSI system is placed into an aluminum frame whose size is about $120cm \times 120\ \text{cm} \times 160\ \text{cm}$. The frame is covered by a shading cloth which can prevent most of light enter the box for avoiding the influence of external light.

2.4. Analyzing algorithm

All algorithms used in this research for pre-processing data or modeling were based on Python (Python Software Foundation, https://www.python.org/) and ran in PyCharm Community Edition (JetBrains, Prague, Czech Republic). Open source libraries numpy, scikit-learning and matplotlib were used in this study.

2.4.1. Reflectance calibration

For lower the influences of camera quantum efficiency and the different configuration of hyperspectral imaging systems, relative corrected images were obtained according to Formula (1) using dark and white reference images.

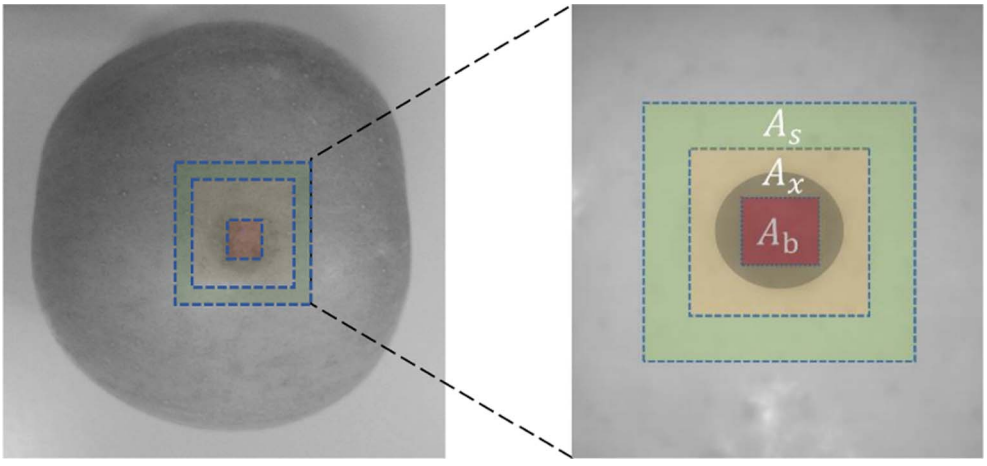

Fig. 4. Select region of interesting.

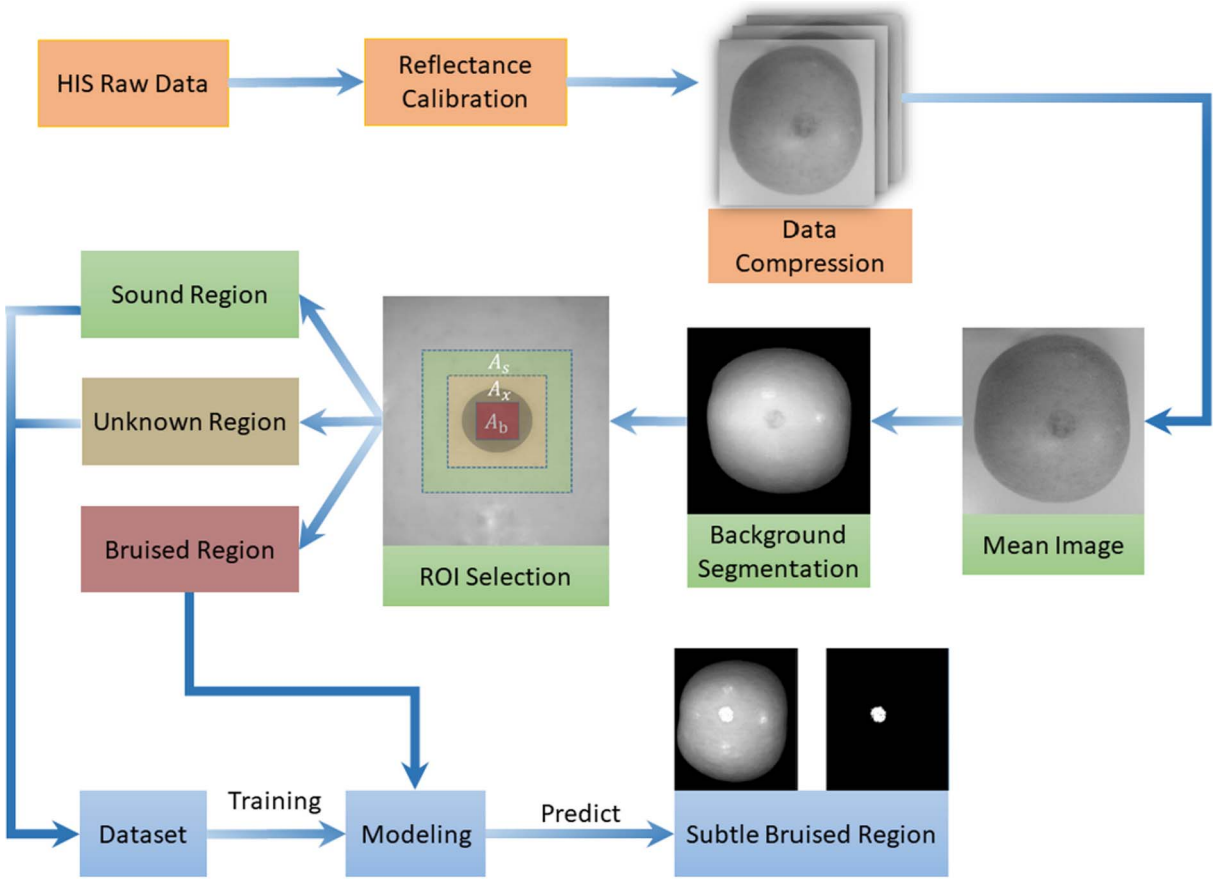

Fig. 5. Flowchart of apple bruise extraction.

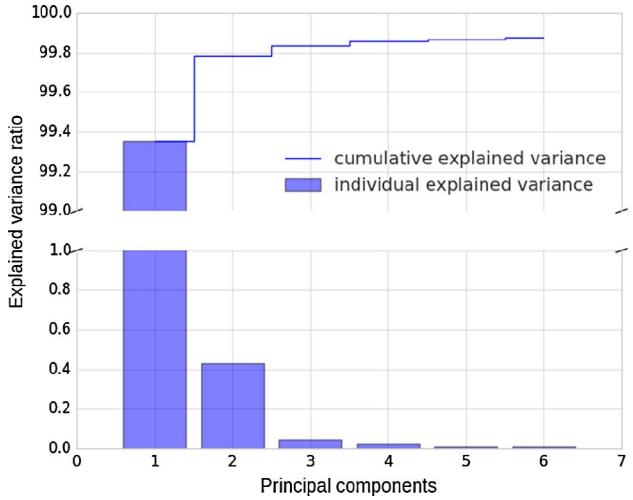

Fig. 6. Explained variance and cumulative explained variance of PC1 to PC6.

$$R(\lambda) = \frac{R_0(\lambda) - B_d}{B_w - B_d} \times 100\% \quad (1)$$

The B_d and B_w in Formula (1) are dark and white reference images respectively. $R_0(\lambda)$ is the raw data at λ wave band, and $R(\lambda)$ is the calibrated hyperspectral data at λ wave band (Pan et al., 2017).

Since the energy of illuminant at each wavelength is not absolute equilibrium, according to the manual of the device, the spectra need a calibration by Formula (2) for obtaining correct reflectance curves.

$$R_\gamma = \frac{L_\gamma}{N_\gamma} * W_\gamma \quad (2)$$

R_γ is the reflectance of measured object at γ wave band, W_γ is the reflectance of reference plate (made from polytetrafluoroethylene) at γ wave band, N_γ and L_γ represent the uncorrected radiance value of measured object and reference plate respectively. The calibrated reflectance spectrum is shown in Fig. 3. The rough regions at both the start and the end of the spectrum may be caused by low quantum efficiency and dark current of the detector (Li et al., 2016). Hyperspectral images with the wavelength range of 450–980 nm were thus reserved for enhancing the signal-to-noise ratio.

2.4.2. Data compression using PCA

The problem of the curse of dimensionality might occur due to the high dimensions of data which contains 696 × 520 × 128 pixels in one cube. Furthermore, not all wavelengths are related to the measured object according to the principle and detection process of NIRS. There is a great amount of redundant information between contiguous wave bands of hyperspectral image. Modeling with full bands data may not only cause the long computation time and consume a huge memory, but also lower the performance for the existence of noise data. Hence, it is necessary to reduce the data size and obtain the useful information before modeling. PCA was used in this research for compression data size, eliminating redundant data, reducing the dimensionality of images and decreasing the influence of noise. The main procedure of PCA including: (1) Calculating eigenvalues and their eigenvectors by covariance matrix of reflectance data. (2) Sorting eigenvalues in descending order, corresponding eigenvectors constitute projection matrix B_k. (3) As shown in Formula (3), the data at original space X can be projected to a new space Y via projection matrix B_k, the k vectors in space Y are corresponding to the k principal components.

$$Y = X * B_k^T \quad (3)$$

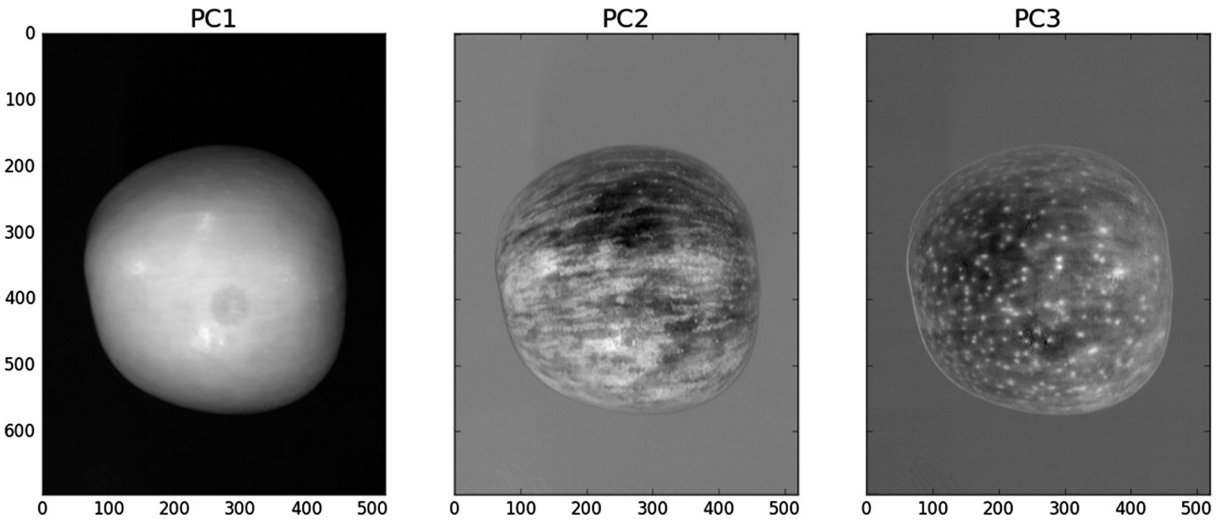

Fig. 7. Principal component image of PC1 to PC3.

Table 1
The comparison between different classification algorithms.

Model	0 h			12 h			18 h		
	Bruise	Sound	Mean	Bruise	Sound	Mean	Bruise	Sound	Mean
SVM-Linear	99.46	92.39	93.99	99.64	100.00	99.91	100.00	100.00	100.00
SVM-RBF	0.00	77.37	59.85	0.00	76.03	57.81	0.00	74.29	55.18
Stochastic Gradient Descent	60.10	98.34	89.69	99.64	99.44	99.49	96.63	100.00	99.13
Decision Tree	95.93	97.74	97.33	99.64	100.00	99.91	99.04	99.34	99.26
Random Forests	99.18	98.53	98.68	98.94	100.00	99.75	99.68	99.89	99.84
Gradient Tree Boosting	96.33	97.74	97.42	99.64	99.89	99.83	99.04	99.34	99.26

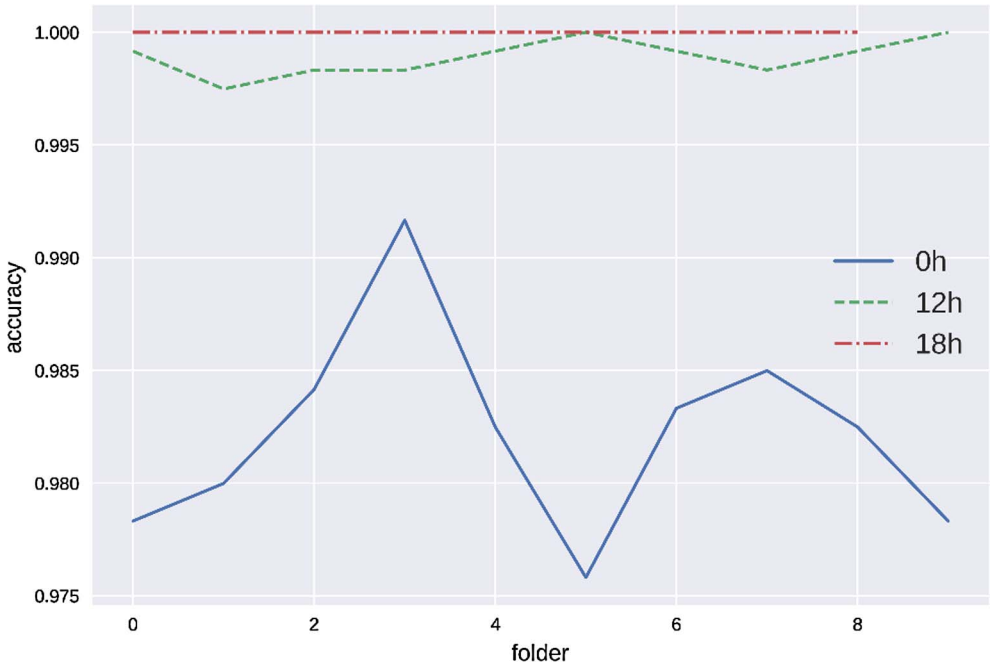

Fig. 8. Results of 10-foler cross validation of the random forest model.

All column vectors in B_k are orthogonal to each other since PCA can find the direction of maximum variance with linear orthogonal transformation. The amount of information in each band decreases with principal components on the basis of the explained variance. The principal component (PC) with little information can be removed for compressing data size (Lü et al., 2011).

2.4.3. Bruise region acquisition algorithm

One of the features of early and slight bruises is that there is little change in its internal chemical constituents. Another important feather is that the lighter the bruise on apples, the fewer differences reflected on spectra between bruised and sound regions. All those features make the bruised region hard to be distinguished from an apple.

A novelty method was provided in this research for recognizing and

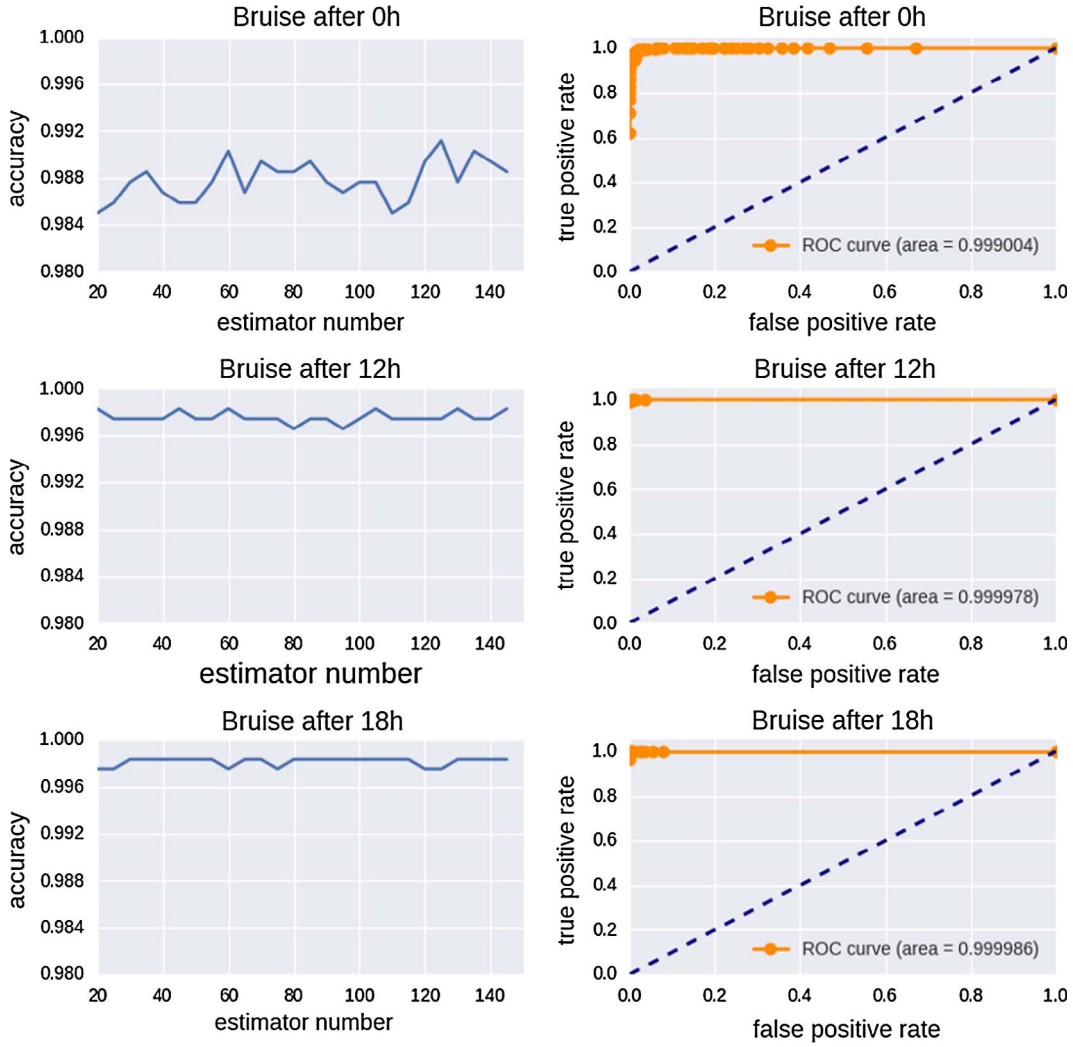

Fig. 9. Classification accuracy curve of 0 h (a), 12 h (c) and 18 h (e), and receiver operating characteristic curve of 0 h (b), 12 h (d) and 18 h (f) of Random Forest model.

extracting the complete bruise region from apples. In this method, the raw data were calibrated at first and PCA was subsequently used to reducing the data size. Then, ROI was manually selected and labeled as bruised region A_b and sound region A_s based on the averaged image of selected principal components images, as shown in Fig. 4. The spectrum of each pixel in bruised and sound regions of interest (ROI) participate in modeling. Then pixels around the boundary of the bruised region would be classified correctly using the fitted model. For avoiding the influence of human labelling errors, the ROIs should be selected under following rules: (a) the labeled bruised region A_b filled with red color in Fig. 4 should be large enough to cover as much as possible bruised pixels but should not exceed the obvious boundary of the bruised region; (b) the labeled sound region A_s filled with green color in Fig. 4 should be small as much as possible but should keep a distance with A_b and not include the visually visible bruised region; (c) the boundary region which is irregular between A_b and A_s should be regarded as unknown region A_x (yellow) for the difficulty of obtaining complete bruised region manually.

The spectrum of each pixel under region A_b and A_s was extracted and unfolded as a dataset DS. For fitting the model as sufficiently as possible, DS was split via hierarchical sample partitioning method into two training sets TR_b, TE_b and two test sets TR_s, TE_s in 5:1 scale. TR_b was combined with TR_s to compose the final training set TR. The same procedure was applied to obtain the final test set TE by combining TE_b and TE_s. To segment the images, five classification methods were tested to determine which one provides the best results. To classify pixels in the boundary of the bruise, five classification algorithms were tested to determine which one provides the best results. Algorithms used for building supervised classification models were listed as follow: Support Vector Machines, Decision Tree (Classification and Regression Trees), Stochastic Gradient Descent and two ensemble methods including Random Forests and Gradient Tree Boosting. All classification algorithms were implemented from Scikit-learn which is a free and open source machine learning library for Python programming language based on Python scientific and numerical libraries Scipy and Numpy. For evaluating the performance of the various models, 10-fold cross validation was used in this study to split the DS into 10 groups. 9 of 10 groups comprises training set, and another group comprises test set. The training set participated in the course of the model building. The test set was used for evaluating the built models by comparing the predicted classes with its actual classes. The main metrics is prediction accuracy which can be calculated by the following formula:

$$\text{Accuracy} = \frac{T_s + T_b}{R_s + R_b} \times 100 \tag{4}$$

Where R_s and R_b represent the number of pixels in the sound region and bruised region respectively. T_s and T_b corresponding to the number of correctly predicted pixels in the sound region and bruised region. The model with the highest performance would be chosen and used for predicting the bruise pixels in boundary region A_x, and the result would be labeled on the image. The overall procedure was depicted as a flowchart in Fig. 5.

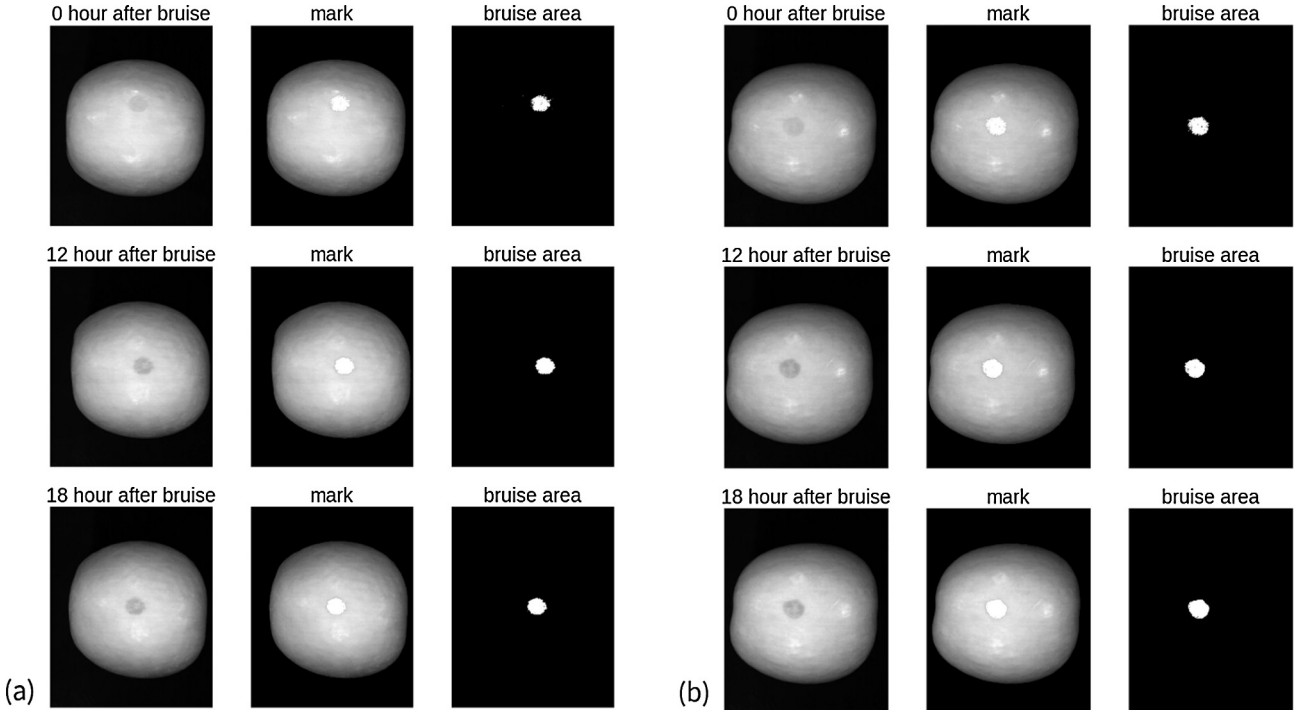

Fig. 10. Predicted bruised region of apples at three stages by RF model.

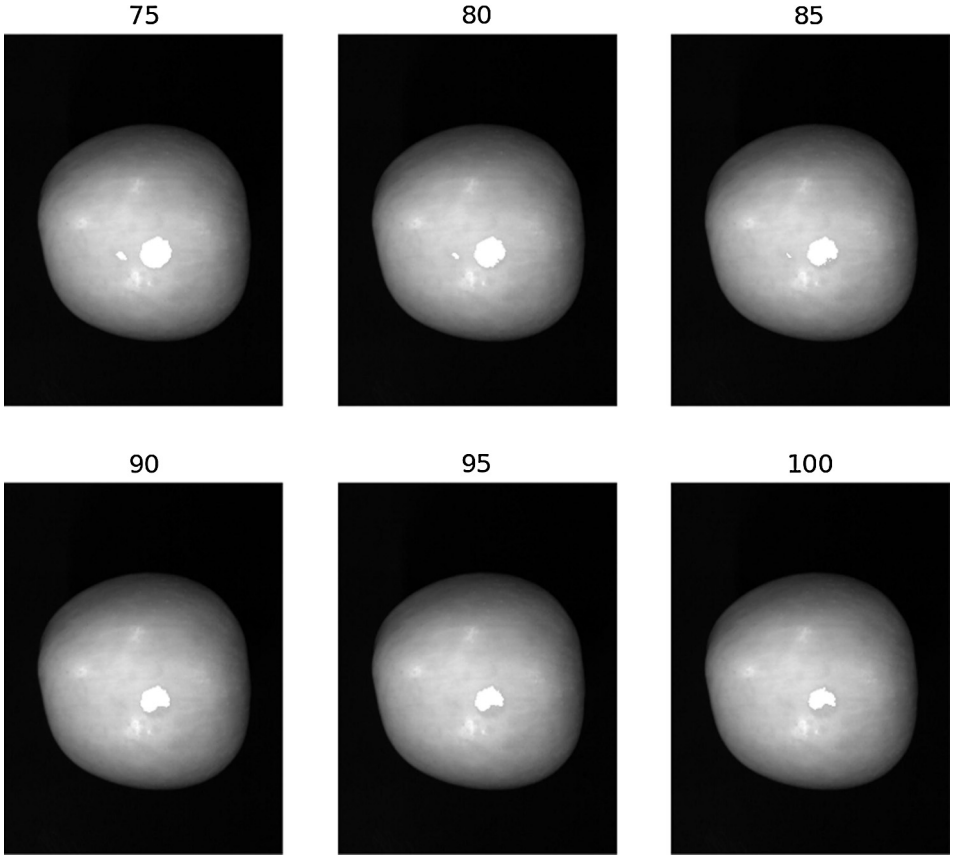

Fig. 11. Bruise of apple segmented by Otsu thresholding method.

3. Results and discussion

3.1. Data compression

The first few principal components (PC) contain most of the information after the PCA. For selecting the appropriate number of PC, the individual and cumulative explained variance of the first six principal components were calculated and illustrated in Fig. 6. The explained variance of the first principal component reached 99.35%, and most of the variances can be explained by the cumulated variance of the

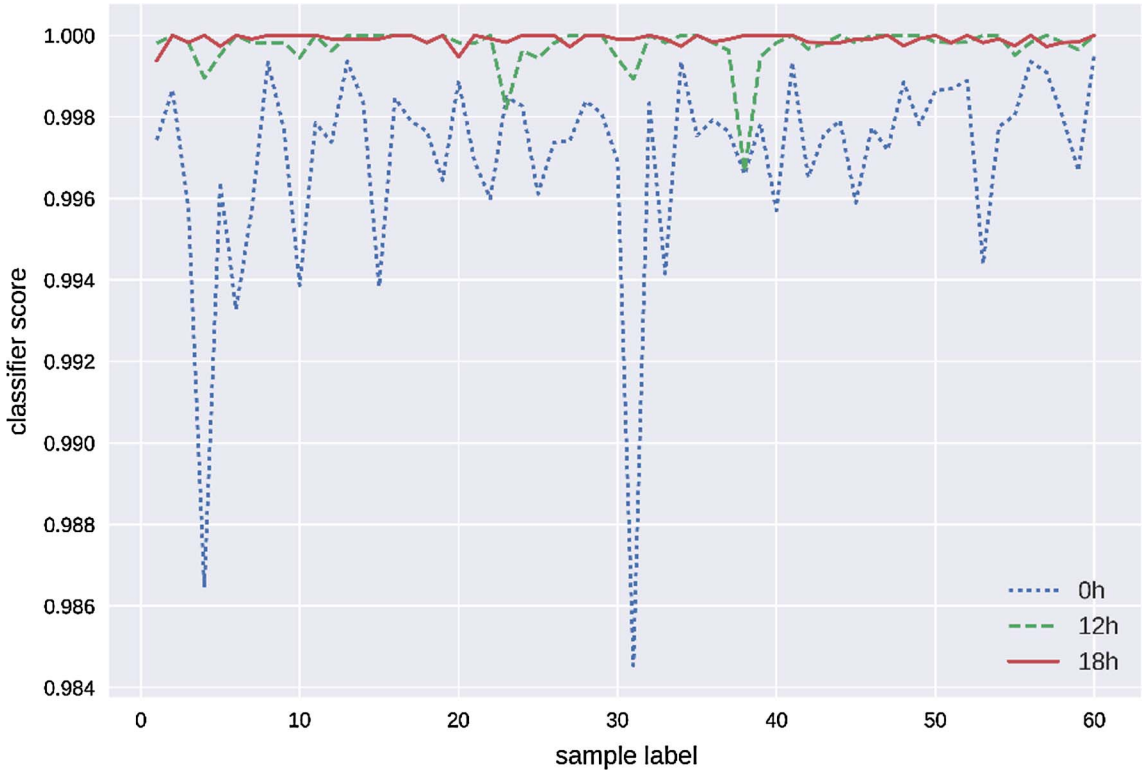

Fig. 12. Performance of all apple bruise extraction models.

first three principal components. The difference between explained variances of the third principal component and the fourth principal component is as low as 0.024%. It is clear that little information has remained after the third principal component. The information contained in first three PCs were reserved because the cumulative explained variance of other PCs is only 0.17%.

Fig. 7 illustrates the first three principal component images obtained after PCA for hyperspectral images. Both of these three images can describe the apple clearly, to be precise, the first principal component reflected the shape and brightness information of apple, the second and the third principal component showed the texture and structure information about the apple.

3.2. Bruised region extraction algorithm

For classifying the pixels in boundary region A_x, a pixel based bruise classification model needs to be built with the training set and evaluated on the test set. The performance of this model can also be used to evaluate that whether the predicted bruised region is identical with the true bruised region. In this research, bruise ROI A_b and sound ROI A_s around the bruise were selected from apples at 0 h, 12 h and 18 h after bruising, and unfold pixels in A_b and A_s to two dimensional spectra. Multiple models were built using six different classification algorithms with the spectral data which was a matrix composed of a number of the spectrum. Classification accuracies including bruise accuracy, sound accuracy and mean accuracy of these six models were shown in Table 1. It can be found that classification accuracies of the sound region are generally slightly higher than the accuracies of the bruised region. In a comparison of all six models, SVM model with RBF kernel shows the lowest accuracy (lower than 60%), and performance of other five models stays at a comparable level. Over 70 percent of sound pixels were correctly classified while no bruise pixels were classified correctly by SVM-RBF model. SVM model with linear kernel showed a high accuracy for classifying bruise after 12 h and 18 h of bruising, yet, it performed much worse for classifying bruise just emerged. The result proves that both SVM and SVM-RBF are not that good at dealing with bruise prediction. The mean classification accuracy of Random Forest model at 12 h and 18 h was slightly lower than that of linear SVM model. However, its accuracy at 0 h was much better than linear SVM model. Since Random Forest can combine weak classifiers to get a strong classifier, which has a high classification accuracy and strong anti-noise ability. Coupled with the feature of sorting the feature importance while modeling, Random Forest was more suitable for the detection of bruises on apples and selecting characteristic wave bands.

After a comparison of all models in Table 1, Random Forests was chosen as the bruise classification algorithm. Subsequently, 10-fold cross validation was applied to it, and results are shown in Fig. 8 where the red, green and blue line represents the 10 accuracies of classification models of apples bruising after 0 h, 12 h and 18 h respectively. Same as the conclusion from Table 1, it can be observed that Random Forests algorithm can get a more accurate bruise on apple along with the time after bruising. That also proved that differences between data under bruise and sound region increase as time goes on.

Exhaustive Grid Search method was used in this study to find an optimal number of decision trees with the best classification accuracy. Fig. 9 shows the classification accuracy curve (left plots) and the receiver operating characteristic curve (ROC, right plots) of the RF model with its highest accuracy. The classification accuracy curves of 12 h and 18 h bruise extraction models are apparently higher than that of 0 h bruise extraction model. In view of nutrition ingredient in apple, chemical constituents of apple with early bruise showed little change, which can't cause a remarkable variation on spectral. As time goes by, the content of reducing sugar, vitamin C and total sugar reduced constantly, and the differences between chemical constituents in the bruise and sound region increased, which is the reason why the bruise is easier to be recognized over time. And the ROC curve shows that all the points in ROC space are above the random guess line (blue dash line). Moreover, compared to the points in ROC space of 0 h, points in ROC space of 12 h and 18 h are closer to the top left corner which means high true positive rate and low false positive rate, and their area under the ROC

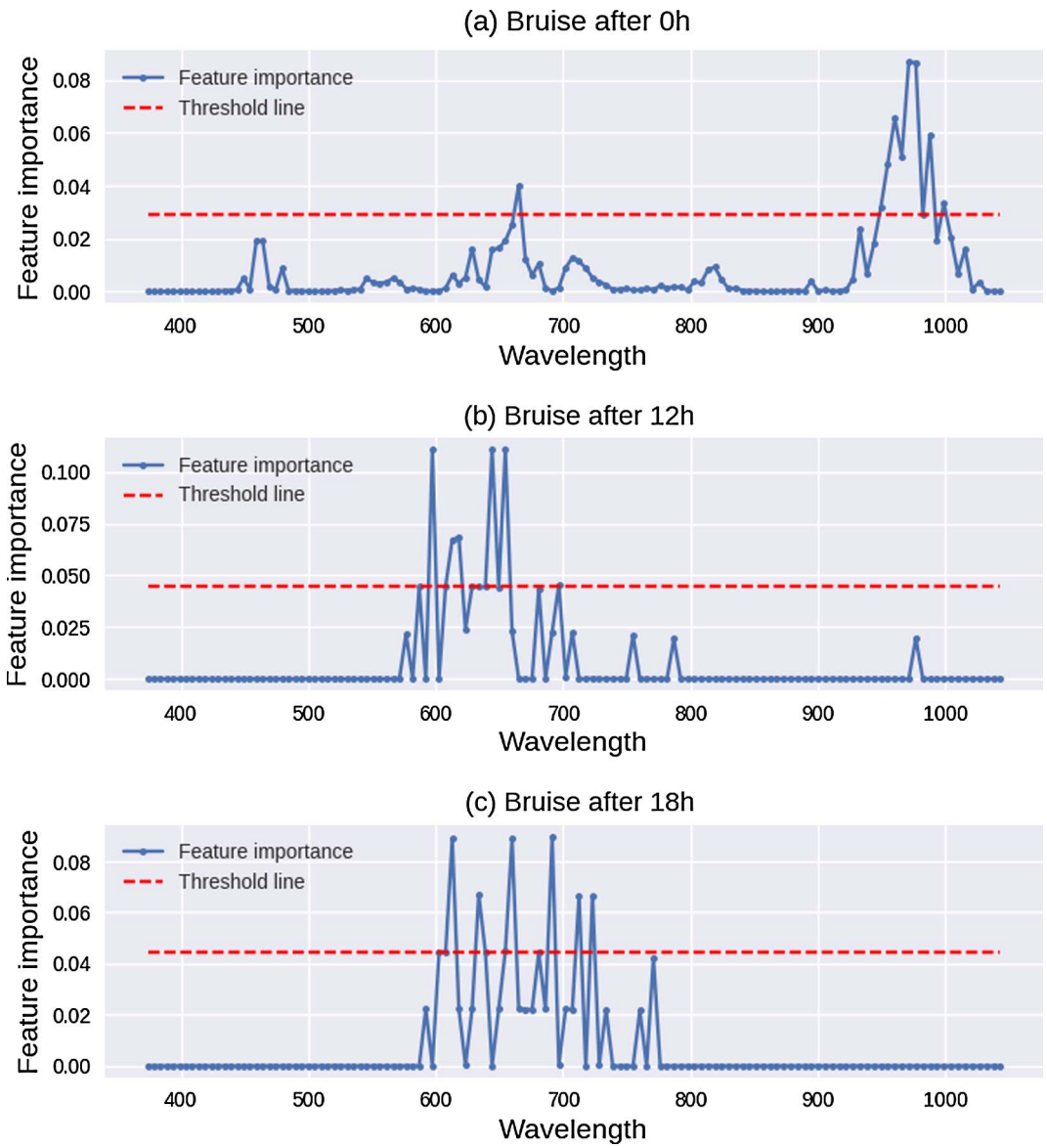

Fig. 13. Feature importance of 0 h, 12 h and 18 h bruise from Random Forest model.

curve (AUC) is closer to 1.

Afterwards, pixels in region A_x were then classified by the built RF model, and a binary mask was created and applied to the mean image of the bruising apple. Fig. 10, (a) and (b) represent the raw bruised apple image, the predicted bruised region on the raw image and the mask image of two random selected tested apples respectively at three stages of 0 h, 12 h and 18 h. Bruised region on raw bruised apple image is roughly circular, but the boundary of the bruised region is irregular. As observed from the plots in the middle column, the predicted bruised region is mostly identical to the bruised region of the raw image. Furthermore, there aren't apparently noisy pixels in the bruise mask image.

As a contrast to the image processing method, the bruised regions segmented by applying Otsu thresholding method to the raw bruised apple image with different threshold values are shown in Fig. 11. The first row of Fig. 11 shows that a misrecognized region always exists while the bruised region is segmented completely. However, the segmented bruise region is incomplete when adjusting the threshold to eliminate the misrecognized region as shown in the second row of Fig. 11. This fact showed the drawback of image thresholding method for segmenting bruise on a complex surface of apples.

60 apples in this experiment were bruised and 180 hyperspectral images were obtained at 0 h, 12 h and 18 h after bruising. Random Forest models of each hyperspectral image were built for accurately predicting the bruised region on apples. The performance of all these models is shown in Fig. 12. The horizontal axis represents the number of sample ranging from 1 to 60, and the vertical axis represents the classification accuracies. The figure illustrates that the holistic accuracies of 0 h models (blue line) with the mean accuracy of 99.75% are generally lower than 12 h and 18 h models (red line and green line) with 99.97% and 99.98% respectively. Overall, all models (0h, 12 h, 18 h) with the mean accuracy of 99.90% and accuracy variance of 0.0014 indicated RF has a high performance and stability on predicting bruised regions of apples.

3.3. Characteristic wavebands acquisition by Random Forest model

Each decision tree in Random Forest model had split the feature space into different pieces, and grade every feature according to the predicted result. By analyzing the feature importance scores of each feature produced by the fitted model, the characteristic wavelengths which have a great contribution to recognizing bruise on apples then can be found. In this research, feature importance scores of Random

Forest bruising prediction model are sorted in ascending order, then first ten correlated wavelengths were retained. As illustrated in Fig. 13, the blue[1] solid line represents the feature importance score of each wavelength, and the red dash line represents the threshold line who divide first ten and other feature importance scores. The first ten characteristic wavelengths of 0 h after bruising for apples are 971.38, 976.89, 960.38, 987.91, 965.88, 954.89, 665.14, 998.96, 949.4 and 982.4 nm, which mainly distributed in about 960 nm. The first ten characteristic wavelengths of 12 h after bruising are 644.21, 654.66, 597.38, 618.15, 612.95, 696.67, 607.76, 628.56, 587.03 and 638.99 nm respectively, and of 18 h after bruising are 691.4, 612.95, 659.9, 633.77, 712.5, 723.08, 654.66, 607.76, 638.99 and 602.57 nm. It can be seen that 12 h and 18 h after bruising, the characteristic wavelengths are mainly distributed between 600 and 700 nm.

As illustrated in Fig. 13, at the first stage after bruising, peaks of feature importance score were observed around 960 nm which represents the content of sugar and water (ElMasry et al., 2008). As time goes on, peaks of feature importance score transferred to around 675 nm which represents the content of carotenoids and chlorophyll pigments in apple (Xing and De Baerdemaeker, 2005). This reveals that the bruise on apple caused a large consumption of total sugar contents and water at the early stage of bruising, and then caused a consumption of carotenoids and chlorophyll pigments in subsequent stage (WANG et al., 2007).

4. Conclusion

In this research, hyperspectral images of 60 bruised apples at 3 stages (0 h, 12 h, 18 h after bruising) were obtained by the HSI system with a spectral range of 400–1000 nm for the detection of bruises on apples. PCA was applied as a data compression method before classification for compressing the data size and extracting useful information. Both the pixel based classification method and the image processing method were used for making a comparison of obtaining the bruised regions from apples. As a result, the ensemble learning method Random Forest was more precise and stable than other models and image processing method, which get fewer misclassification pixels when extracting apple bruised region, especially in the case that the boundary of the bruise is complex and inconspicuous. The comparison between different bruise detection models proved that Random Forest model was the most effective for pixel based bruised region prediction for its high classification accuracy and generalization ability. Two characteristic wavebands around 675 nm and 960 nm were found according to the built Random Forest model, which concludes that total sugar contents and water were consumed rapidly in the early stage of bruising, then, carotenoids and chlorophyll pigments instead them. The mean accuracy of pixel based Random Forest models reached 99.90%, which indicated that this technique can be used to effectively obtain complete bruised regions on apples and has the potential for being implemented on apple quality sorting and grading system.

References

Ariana, D.P., Lu, R., Guyer, D.E., 2006. Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. Comput. Electron. Agric. 53 (1), 60–70.

Baranowski, P., Mazurek, W., Wozniak, J., Majewska, U., 2012. Detection of early bruises in apples using hyperspectral data and thermal imaging. J. Food Eng. 110 (3), 345–355.

Baranowski, P., Mazurek, W., Pastuszka-Woźniak, J., 2013. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. Postharvest Biol. Technol. 86, 249–258.

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 Sci. Technol. 41 (2), 337–345.

Ferrari, C., Foca, G., Calvini, R., Ulrici, A., 2015. Fast exploration and classification of large hyperspectral image datasets for early bruise detection on apples. Chemometr. Intell. Lab. Syst. 146, 108–119.

Jackson, P.J., Harker, F.R., 2000. Apple bruise detection by electrical impedance measurement. Hort Science 35 (1), 104–107.

Keresztes, J.C., Goodarzi, M., Saeys, W., 2016. Real-time pixel based early apple bruise detection using short wave infrared hyperspectral imaging in combination with calibration and glare correction techniques. Food Control 66, 215–226.

Keresztes, J.C., Diels, E., Goodarzi, M., Nguyen-Do-Trong, N., Goos, P., Nicolai, B., Saeys, W., 2017. Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging. Postharvest Biol. Technol. 130, 103–115.

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 Bioprocess Technol. 5 (4), 1121–1142.

Lu, R., 2003. Detection of bruises on apples using near–infrared hyperspectral imaging. Trans. ASAE 46 (2), 523.

Lü, Q., Tang, M.J., Cai, J.R., Zhao, J.W., ViTTayapaduNg, S., 2011. Vis/NIR hyperspectral imaging for detection of hidden bruises on kiwifruits. Czech J. Food Sci. 29 (6), 595–602.

Li, J., Huang, W., Tian, X., Wang, C., Fan, S., Zhao, C., 2016a. Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral imaging. Comput. Electron. Agric. 127, 582–592.

Li, J., Chen, L., Huang, W., Wang, Q., Zhang, B., Tian, X., Fan, S., Li, B., 2016b. Multispectral detection of skin defects of bi-colored peaches based on vis–NIR hyperspectral imaging. Postharvest Biol. Technol. 112, 121–133.

López-Maestresalas, A., Keresztes, J.C., Goodarzi, M., Arazuri, S., Jarén, C., Saeys, W., 2016. Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. Food Control 70, 229–241.

Pan, L., Sun, Y., Xiao, H., Gu, X., Hu, P., Wei, Y., Tu, K., 2017. Hyperspectral imaging with different illumination patterns for the hollowness classification of white radish. Postharvest Biol. Technol. 126, 40–49.

Pu, Y.Y., Feng, Y.Z., Sun, D.W., 2015. Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: a review. Compr. Rev. Food Sci. Food Saf. 14 (2), 176–188.

Rehkugler, G.E., Throop, J.A., 1986. Apple sorting with machine vision. Trans. ASAE 29 (5), 1388–1397.

Schatzki, T.F., Haff, R.P., Young, R., Can, I., Le, L.C., Toyofuku, N., 1997. Defect detection in apples by means of X-ray imaging. Trans. ASAE 40 (5), 1407–1415.

Tian, Y., Cheng, Y., Wang, X., Li, Q., 2014. Feature vectors determination for pest detection on apples based on hyperspectral imaging. Trans. Chinese Soc. Agricult. Eng. 30 (12), 132–139.

Wang, Y.Y., Hu, W.Z., Pang, K., Zhu, B.W., Fan, S.D., 2007. Effect of mechanical damage on the nutritional compositions in Fuji apple. Food Sci. Technol. 10, 030.

Xing, J., De Baerdemaeker, J., 2005. Bruise detection on 'Jonagold'apples using hyperspectral imaging. Postharvest Biol. Technol. 37 (2), 152–162.

Xing, J., Bravo, C., Jancsók, P.T., Ramon, H., De Baerdemaeker, J., 2005. Detecting bruises on 'Golden Delicious' apples using hyperspectral imaging with multiple wavebands. Biosyst. Eng. 90 (1), 27–36.

Xing, J., Saeys, W., De Baerdemaeker, J., 2007. Combination of chemometric tools and image processing for bruise detection on apples. Comput. Electron. Agric. 56 (1), 1–13.

Zhu, Q., Guan, J., Huang, M., Lu, R., Mendoza, F., 2016. Predicting bruise susceptibility of 'Golden Delicious' apples using hyperspectral scattering technique. Postharvest Biol. Technol. 114, 86–94.

[1] For interpretation of color in Fig. 13, the reader is referred to the web version of this article.