2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.201800536
Paper Number: 1800536

DETECTION OF SURFACE AND SUBSURFACE DEFECTS OF APPLES USING STRUCTURED-ILLUMINATION REFLECTANCE IMAGING WITH MACHINE LEARNING ALGORITHMS

Yuzhen Lu[a], Renfu Lu[b,*]

[a]Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA; email address: luyuzhen@msu.edu

[b]United States Department of Agriculture Agricultural Research Service (USDA/ARS), Michigan State University, East Lansing, MI 48824, USA

**Corresponding author: renfu.lu@ars.usda.gov*

Written for presentation at the
2018 ASABE Annual International Meeting
Sponsored by ASABE
Detroit, Michigan
July 29-August 1, 2018

ABSTRACT. *Machine vision technology coupled with uniform illumination is now widely used for automatic sorting and grading of apples and other fruits, but it still does not have satisfactory performance for defect detection because there are a large variety of defects, some of which are difficult to detect under uniform illumination. Structured-illumination reflectance imaging (SIRI) offers a new modality for imaging by using sinusoidally-modulated structured illumination, to obtain two sets of independent images, i.e., direct component (DC), which corresponds to conventional uniform illumination, and amplitude component (AC), which is unique to structured illumination. The objective of this study was to develop machine learning classification algorithms by using DC and AC images and their combinations for enhanced detection of surface and subsurface defects of apples. A multispectral SIRI system under illumination of two phase-shifted sinusoidal patterns was used to acquire near-infrared images from 'Delicious' and 'Golden Delicious' apples with various types of surface and subsurface defects. DC and AC images were extracted through demodulation of the acquired images, and were then enhanced using bi-dimensional empirical mode decomposition (BEMD) and subsequent image reconstruction. Defect detection algorithms were developed, by using random forest (RF), support vector machine (SVM) and convolutional neural network (CNN), for DC, AC, and ratio (AC divided by DC) images and their combinations. Results showed that AC images were superior to DC images for detecting subsurface defects and DC images were overall better than AC for detecting surface defects, whereas ratio images were comparable to, or better than, DC and AC images for defect detection. The ensemble of DC, AC and even ratio images resulted in significantly better detection accuracies over using them individually. Among the three classifiers, CNN performed the best with 98% detection accuracies for both varieties of apples, followed by SVM and RF. This research demonstrated that SIRI, coupled with a machine learning algorithm, can be a new, versatile and effective modality for fruit defect detection.*

Keywords. *Apple, defect, empirical mode decomposition, machine learning, structured illumination.*

Introduction

In postharvest handling of apples, defective fruit should be sorted out to ensure high-quality, wholesome products for the consumer. Sorting fruit for defects is traditionally performed manually, which is labor intensive and also prone to inspection inconsistency. Since 1990s, considerable effort has been directed towards the development of computer vision technologies for apple defect sorting (Y. Lu & Lu, 2017c). Early research mainly used monochromatic or black/white imaging (Crowe & Delwiche, 1996; Varghese, 1992; Yang, 1994) and color or RGB (red-green-blue) imaging (Leemans, Magein, & Destain, 1998, 1999), while much of later research was focused on spectral imaging, including hyperspectral (R. Lu, 2003; Mehl, Chen, Kim, & Chan, 2004; Xing & De Baerdemaeker, 2005) and multispectral (Bennedsen, Peterson, & Tabb, 2007; W. Huang, Li, Wang, & Chen, 2015; Kleynen, Leemans, & Destain, 2005; Unay et al., 2011). In addition, other imaging technologies, such as magnetic resonance imaging, thermal imaging, and X-ray imaging, have also been researched for apple defect detection.

Computer imaging with diffuse, uniform illumination is now widely used for automated detection of defects on apples; however, its performance still falls short of the industry's expectations because some surface and subsurface defects are either not visible or can be confused with the normal tissue during image processing and classification. To overcome the shortcomings of conventional computer imaging technology with uniform illumination for fruit defect detection, our lab recently developed a new structured-illumination reflectance imaging (SIRI) system for fruit defect detection (Y. Lu, Li, & Lu, 2016). With using sinusoidally-modulated illumination patterns, SIRI is able to provide two sets of independent images, i.e., direct component (DC) and amplitude component (AC), the former of which correspond to uniform illumination, while the latter are unique for structured illumination. Compared to DC images, AC images have better spatial resolutions and image contrasts, while also possessing depth-resolving features, which are dependent on the spatial frequency of light patterns. Our recent research demonstrated that AC images can effectively detect subsurface bruising in apples at early stage (Li, 2016; Y. Lu & Lu, 2017a), which otherwise cannot be achieved by DC. However, our preliminary research also showed that AC images were inferior to DC images in some cases for detection of surface defects (Li, 2016); the detection results were found to depend on type of defect and vary with variety of apple (Li, 2016). This is because AC images globally enhance fruit surface features, including normal lenticels and other skin textures, which could confound with the appearance of defects, and also diminish the image signal-to-noise ratio (SNR), especially at high spatial frequencies. Since DC and AC provide different features pertaining to defects, it is expected that proper combination of DC and AC images can improve detection accuracy than using them individually, which was the main focus of the current research.

Image segmentation is a critical step for automatic detection of fruit defect, and in many previous studies, it is done by either thresholding (Bennedsen & Peterson, 2005; Y. Lu & Lu, 2017b; Throop, Aneshansley, Anger, & Peterson, 2005; Zhang et al., 2015) or pattern classifiers (Leemans et al., 1999; Unay & Gosselin, 2006, 2007). Image features (i.e., statistical, textural and/or geometric) are then extracted from the segmented blobs, followed by features selection if needed, which are finally fed into a defect detection algorithm. This approach, although advantageous in locating individual defects in the image, needs long computational time and is not effective for segmentation of many types of defects such as russeting and bruise on apples (Leemans & Destain, 2004). A simpler yet still effective alternative is to directly extract discriminative features from the whole image without performing defect segmentation (Kavdir & Guyer, 2004; Mehl, Chao, Kim, & Chen, 2002). In addition, researchers have also explored pixel-based defect detection by training a classifier on a pixel level (Ariana, Guyer, & Shrestha, 2006; Xing, Saeys, & De Baerdemaeker, 2007). This approach, however, overlooks spatial correlations between neighboring pixels, and tends to result in many noisy defect segments.

Development of an appropriate classification algorithm is key to fruit defect detection. Traditionally, linear discriminant analysis, nearest neighbor and decision tree are commonly used classifiers for fruit defect detection, which have been replaced later by more effective algorithms such as support vector machine (SVM) and neural networks (Kavdir & Guyer, 2008; Unay et al., 2011). In recent years, machine learning has emerged rapidly as a new field of artificial intelligence for image processing and data analysis, and a plethora of learning algorithms are currently available (Marsland, 2015). Ensemble learning, which uses an ensemble of multiple learners or classifiers to generate a classification decision using a rule-based approach (e.g., majority voting), has received increasing interest recently. Random forests (RFs) are a method of ensemble learning (Breiman, 2001), which uses decision trees as a base learner to train the data via bootstrap aggregating. The method has achieved remarkable success in many classification tasks (Bosch, Zisserman, & Munoz, 2007; Criminisi & Shotton, 2013). Moreover, convolutional neural networks (CNNs), as a major method of machine learning, have demonstrated superiority over many other state-of-the-art machine learning algorithms (Krizhevsky, Sutskever, & Hinton, 2012). CNN is unique in that it acts as both a feature extractor and a classifier at the same time, thus obviating the need for manually extracting relevant features. CNN has recently found applications in the agricultural and food domain, such as plant disease detection (Mohanty, Hughes, & Salathe, 2016), fruit counting in the orchard (Bargoti & Underwood, 2017; Chen et al., 2017), and quality inspection of fish fillets (Xu & Sun, 2018). These emerging learning algorithms (e.g., RF and CNN) hold great potential to enhance apple defect detection.

This study was built on, and expanded from, our recent works on using SIRI for defect detection of apples (Li, 2016; Y. Lu et al., 2016; Y. Lu & Lu, 2017a), to develop effective algorithms for detecting both surface and subsurface defects of

apples. The specific objectives of this study were to:

•Acquire DC and AC images for apples with different types of defects using a multispectral SIRI system coupled with a two-phase based image demodulation scheme;

•Enhance the DC and AC images, prior to features extraction and classification, by using a bi-dimensional empirical mode decomposition (BEMD) image preprocessing method; and

•Develop classifiers for DC, AC and ratio (AC divided by DC) images and their combinations, by employing three machine learning algorithms (i.e., random forest, support vector machine, and convolutional neural network), to classify apples into defective and normal classes, and compare their performance for defect detection.

Materials and Methods

Apple samples

Apples of 'Delicious' and 'Golden Delicious' varieties were hand-picked from an orchard of Michigan State University's Horticultural Teaching and Research Center in Holt, Michigan in September 2017. Many of the harvested apples had been inflicted with different types of defect, while they were still on the tree. Each apple was visually inspected and recorded for specific physiological disorders and/or mechanical or natural (i.e., bruising, frost) damage. The list of defects observed for the two varieties of apples are summarized in Table 1. Among the different types of defects, frost ring and bitter pit, as shown in Figure 1, represented the main defects for 'Delicious' and 'Golden Delicious' apples, respectively. Frost ring (a special type of skin russeting) was caused by frost when apples were in bloom, while bitter pit developed beneath the fruit skin but had become externally visible. Mechanical injury, including subsurface bruising, was observed for both varieties, which could have been induced by the fruit-to-fruit impact during transport of apples and stacking loads during storage. The collected apples had been kept in cold storage at 4 °C for about 50 days prior to image acquisition. In total, there were 190 defective and 128 normal apples for 'Delicious' and 156 defective and 94 normal apples for 'Golden Delicious'.

Table 1. Number of apples with different types of defects.

	Bitter pit	Frost ring	Russeting or scab	Lenticel spots or breakdown	Mechanical injury	Insect damage	Rot or decay	Fecal contamination	Mixed defects[a]
'Delicious'		76	37	40	6	7	3	1	20
'Golden Delicious'	64		28		28		1		35

[a] The mixed type indicates a combination of two or more types of defects, usually including mechanical injury.

Figure 1. Photo of a 'Delicious' apple with frost ring (left) and a 'Golden Delicious' apple with bitter pit (right).

Image Acquisition

An in-house assembled multispectral SIRI system, as schematically shown in Figure 2, was used to acquire images from apples, which mainly consisted of a digital light projector (DLP) (DL*i* CEL5500-Fiber, Digital Light Innovations, Austin, TX, USA), a quartz-tungsten-halogen lamp (Newport Corporation, Irvine, CA, USA), a camera (PhotonMax:1024B, air-cooled, Princeton Instruments, Trenton, NJ, USA) with a pixel resolution of 1024×1024, a 35 mm focusing lens and a liquid crystal tunable filter (LTCF) (VariSpec, Cambridge Research and Instrumentation, Inc., Woburn, MA, USA) for waveband selection. Two phase-shifted sinusoidal patterns (with phase offsets of 0 and $2\pi/3$), in 8-bit bmp format, were created in Matlab (The Mathworks, Inc., Natick, MA, USA) and uploaded to the control software of the DLP for sample illumination. Image acquisitions were performed using a custom-developed graphical user interface program in LabVIEW 2016 (National Instruments, Austin, TX, USA), with the exposure time of 300 ms and on-chip 2×2 binning.

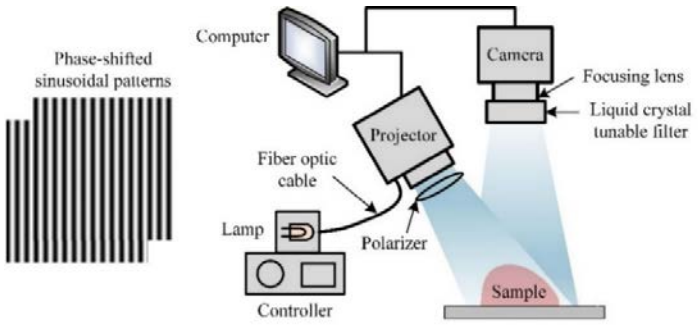

Figure 2. Schematic of a multispectral structured-illumination reflectance imaging system.

Selection of proper spatial frequencies for SI patterns is critical for achieving optimal performance. Our previous studies showed that higher frequency was more conducive to enhancing surface features of fruit, including normal lenticels, but would result in lower SNRs for the acquired images. Based on a preliminary test on the range of spatial frequencies of 0.05-0.3 cycles/mm, the frequency of 0.15 cycles/mm was chosen in this study, which would enhance both surface and subsurface features, while suppressing the lenticel features. Waveband selection is also important for effective defect detection. For the current SIRI system, the spectral images above 850 nm suffered strongly from specular reflectance and were thus not considered. A preliminary study on a comparison of 10 wavebands in the range of 650-830 nm with 20 nm increments, showed that the NIR bands at 710 nm and above, which are insensitive to color variations of fruit skin, had similar performance in recognition of many types of defects. Hence the waveband of 730 nm was selected in this study because it had the highest spectral response, which is also in agreement with the previous studies using 740 nm for apple defect detection (Bennedsen et al., 2007; Throop et al., 2005).

During imaging, the defective apples were placed on a sample holder with an adjustable stage, with the location of defects on the apple facing the camera, while the normal apples were placed with the stem-calyx axis being horizontal (only one side of each apple was imaged). Two images were acquired from each apple under two phase-shifted sinusoidal illumination patterns (i.e., 0 and $2\pi/3$, respectively) as shown in Figure 2.

Image Preprocessing

Image preprocessing for the raw SIRI images consisted of three steps, as shown in Figure 3. First, the acquired images were demodulated into DC and AC images, using spiral phase transform for two phase-shifted pattern images, which is described in detail in Lu et al. (2016a) and Lu & Lu (2017a). Next, object masking was performed on the DC images alone using a unimodal histogram based automatic thresholding technique. The unimodal thresholding is to find a threshold which would maximize the distance to a straight line drawn from the histogram peak to its last non-empty bin position (Y. Lu & Lu, 2017b; Rosin, 2001). Finally, since the derived DC and AC images still suffered from artifacts of noise and vignetting (i.e., an uneven illumination phenomenon characterized by gradual fading out of the image intensity towards its periphery), image enhancement was performed through bi-dimensional empirical mode decomposition (BEMD) and subsequent image reconstruction to remove or minimize these artifacts.

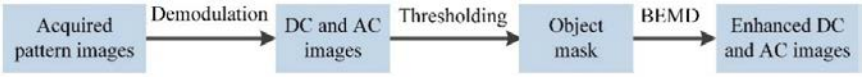

Figure 3. Flow chart of image preprocessing procedures, where DC and AC represent direct and amplitude components, respectively and BEMD represents bi-dimensional empirical mode decomposition.

BEMD is a two-dimensional extension of the empirical mode decomposition (EMD) method, which was originally proposed by Huang et al. (1998) for nonlinear and non-stationary time series analysis. Different from classic Fourier and wavelet analysis, BEMD requires no a priori basis for data decomposition, and it is empirical, adaptive and fully data-driven (N. E. Huang & Shen, 2014). In this study, BEMD was utilized to decompose the demodulated DC and AC images through a sifting process into a number of intrinsic mode function (IMF) images plus a residual (the number of IMFs was limited to 6 in this study). The resulting IMFs and the residual images provided a representation of the original image at different frequency scales. By selectively discarding the first IMF and the residual, an enhanced image with vignetting and noise removed could be reconstructed, which will be shown later.

Ratio (RT) images, which are obtained by dividing AC by DC, are useful for bruise detection of apples (Y. Lu et al., 2016; Y. Lu & Lu, 2017a), because they would enhance image contrast for bruises and also correct illumination vignetting. In this study, the RT images were derived from the original DC and AC images with their first IMF eliminated. In addition, the ensemble of DC and AC images (denoted as DC-AC), and the ensemble of DC, AC and RT images (i.e., DC-AC-RT) were also exploited as inputs for image classification.

Classification by RF and SVM

Features Extraction

In Image classification with RF and SVM requires the images to be represented by a number of discriminative features. In many studies for fruit defect detection, textural features were used for image classification (Kavdir & Guyer, 2008; Unay et al., 2011). Hence in this study, three types of textural features were extracted, which included 28 Haralick textural features, 59 local binary pattern (LBP) features and 67 Gabor features from each image. In addition, 6 six basic intensity features and 7 Hu-moments features were also extracted, which resulted in a complete set of 167 features for each image.

Haralick features were calculated for the mean and range of 14 different textural measures based on the co-occurrence matrix with a 3-pixel distance over 4 directions defining co-occurrence (Haralick, Shanmugam, & Dinstein, 1979). The LBP features were calculated from 58 uniform LBPs and one non-uniform LBP (Pietikainen, Hadid, Zhao, & Ahonen, 2011), in which the uniformity of LBPs is defined based on the occurrence of at most 2 bitwise transitions when circularly sampled in a neighborhood of 3×3 pixels (Ojala, Pietikainen, & Maenpaa, 2002). For the Gabor features, the image was filtered with a bank of Gabor filters generated with 8 dilations and 8 rotations as a spatial mask of 21×21 pixels, and the average

magnitudes of the filtered images were taken as the Gabor features, in addition to the minimum, maximum and normalized difference of the magnitude (Kumar & Pang, 2002). The basic intensity features included mean, standard deviation, kurtosis, skewness and mean gradient and mean Laplacian (Nixon & aguado, 2012). The Hu-moments features used are invariant shape descriptors (Hu, 1962). A more detailed description of all these features and their computations can be found in Nixon & Aguado (2012).

The extracted features were concatenated in tandem for each image to form a features vector. In the case of two ensembles of input images (i.e. DC-AC and DC-AC-RT), the features extracted from each channel (i.e., DC, AC or RT image) were concatenated as well. For the ensemble of DC and AC images, for instance, the length of the feature vector was 334 (i.e., 2×167). Since each type of features had different scales, feature-wise normalization was done so that all normalized features had a zero mean and a unit variance.

Classification

The image classification was performed for two varieties of apples, separately. For each variety, the dataset was randomly partitioned into training and test sets according to a ratio of 7 to 3, resulting in 223 training and 95 test samples for 'Delicious' and 175 training and 75 test samples for 'Golden Delicious'.

RF is a learning algorithm that uses an ensemble of unpruned decision trees, each of which is built based on a subset (i.e., 2/3) of training data created through bootstrap sampling with replacement using a randomly selected subset of variables or features (Breiman, 2001). Implementation of RF would result in two parameters, i.e., the number of grown trees and the number of randomly selected variables. Our preliminary tests showed that the classification error decreased rapidly and then leveled off without overfitting, when the number of grown trees exceeded a certain value, and it also was insensitive to the number of features used. Hence, fine-tuning of the two parameters was not necessary. In the subsequent analysis, the number of trees was chosen to be 200, and all the features were used in all training cases.

SVM comes in many different versions with or without using kernel tricks to solve linear or nonlinear classification problems. In this study, a linear soft-margin SVM (Cortes & Vapnik, 1995) was chosen for the image classification given that the two-class (defect versus normal) classification could be treated as a linear problem. In the SVM, only one parameter needed to be optimized, i.e., the regularization parameter that controls the tradeoff between maximizing the margin and minimizing the misclassification. Ten-fold cross validations were performed to decide the parameter that was sampled over a range of 0.001 to 1000 based on the minimum classification error.

Given the randomness of data partition, each of the above training cases was repeated 30 times. In each round of training for each apple variety, the sample indices for the training and test sets were shared in training the 5 different image inputs (i.e., three sets of DC, AC and RT images and two combination sets) with the three different classifiers, including the CNN that is presented next, to ensure fair comparison of the results for the different image inputs and the three classifiers.

The above described procedures for image preprocessing, features extraction and model training with RF and SVM were performed in Matlab R2017a (The Mathworks, Inc., Natick, MA, USA).

Classification by CNN

CNN, originally introduced by LuCun et al. (1998), has a specially designed architecture for image classification. A typical CNN architecture embodies three major design characteristics, i.e., local receptive field, shared weights and pooling (or sub-sampling), contributing to the sensitivity to local features and robustness to deformations and shifts, and also reducing learnable parameters as well. As illustrated in Figure 4, a convolution operation is applied to a local receptive field of 3×3 pixels (i.e., the convolution kernel size) in the input image, which is connected to a hidden neuron of the first hidden layer, and sliding the local receptive field across the entire image step by step (by one pixel here) builds up the first hidden layer, which is defined by a number of feature maps. Then maximum pooling operations follow, which downsize the input feature maps by outputting the maximum in a local region of 2×2 pixels. In performing convolutions, all weights (and bias) in the convolution kernel are shared for building each of the feature maps. A CNN usually has more than one convolution-pooling layer, followed by one or more conventionally fully-connected layers before the final output.

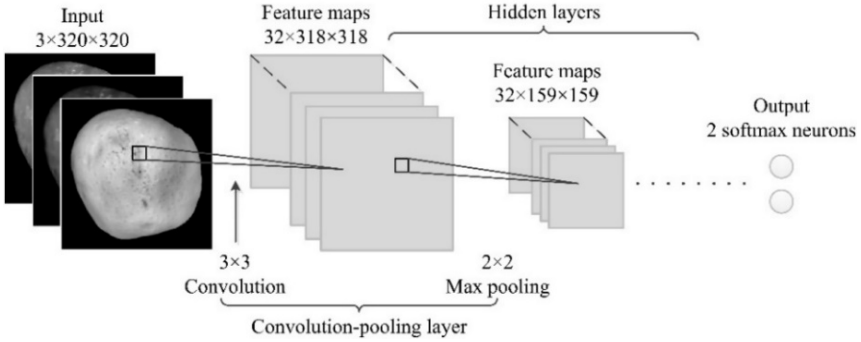

Figure 4. The architecture of a convolutional neural network (CNN).

Like other deep neural networks, CNN, which contains a large number of learnable parameters, needs a large set of images for model training so as to avoid overtraining. Preliminary attempts showed that the present image data were inadequate for the CNN to produce satisfactory accuracy. To address the problem, data augmentation was performed to increase the number of training images by applying various geometric transformations to the original image, which is a common practice for training CNNs with limited data (Simard, Steinkraus, & Platt, 2003). The transformations included translation, rotation, shear, zooming and flipping, among which the first four were done by uniformly applying a random displacement within a preset range, and flipping (both horizontal and vertical) was randomly performed. Figure 5(a) shows an example of five images generated through these transformations. As a result, the training samples were augmented for 'Delicious' apples by five times to 1,115 images and for 'Golden Delicious' to 875 images. All augmented images were resized to a standard size of 320×320 pixels based on the fruit object size, which were then normalized to have zero mean and unit variance before training

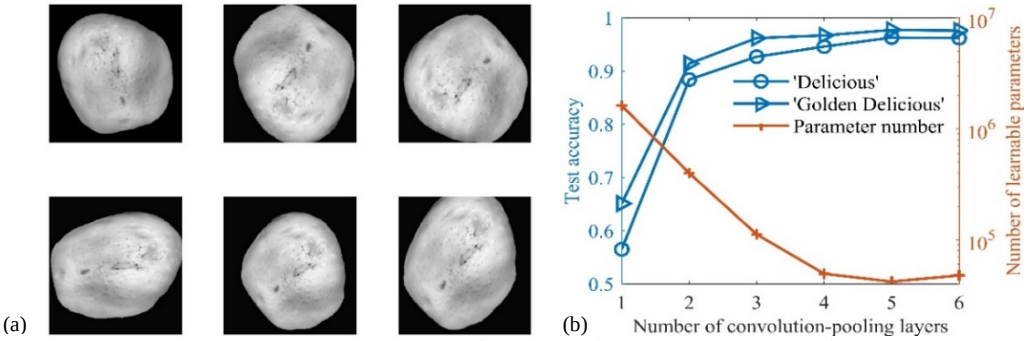

Figure 5. (a) Five augmented images generated by geometrical transformations of an original image (top left) and (b) test accuracy and the number of learnable parameters versus the number of convolution-pooling layers in the convolutional neural networks using the ensemble of direct component, amplitude component and ratio images for two varieties of apple.

Training a CNN model requires much effort and skill to configure numerous hyper-parameters, including learning rate, training epochs, batch size, activation function, number of hidden layers, convolution kernel size, pooling size, dropout, etc. (Bengio, 2012). This study was not intended to optimize all these parameters, but instead, it focused on the number of convolution-pooling layers (relating to the depth of a CNN), which plays a crucial role in model performance and also affects the model complexity in terms of the number of learnable parameters. For each convolution-pooling layer, 32 convolution filters (resulting in 32 feature maps) of size 3×3 pixels with a stride step of 1 pixel were used, followed by 2×2 max-pooling, as shown in Figure 4, and a rectified linear unit was used as the activation function in all the hidden layers and softmax for the output (fully-connected) layer (no other fully-connected layers were added). The optimizer for model training was 'adadelta' (Zeile, 2012) with a starting learning rate of 1, and a dropout trick was used following the last pooling layer and the output layer, with dropout rates of 0.25 and 0.5, respectively. A learning rate updating scheme was implemented, which reduced the learning rate by half when the test accuracy did not improve for 5 epochs. The training process would be terminated if there was no accuracy improvement for 15 epochs. The maximum number of training epochs was set to 50, which was found to be sufficient in this study.

Based on these settings, 1 to 6 convolution-pooling (i.e., hidden layer) layers were explored for the CNN modeling. As shown in Figure 5(b), accuracies of more than 90% were obtained by using 3 or more hidden layers, and the accuracies began to converge or remain stable when the number of hidden layers exceeded 5. It was noted that increasing the hidden layers reduced model parameters, and 5 hidden layers gave the minimum parameters (~ 40k without considering neuron dropout), which were therefore chosen for the CNN modeling for the remaining analysis.

The above described image classification procedures with CNN were performed in Python 2.7 with deep learning packages of Keras (Chollet, 2015) and Theano (Bergstra et al., 2010) in a desktop computer.

Accuracy Evaluation

Image classification results for the test dataset were evaluated using three common performance metrics, i.e., precision (*PRE*) and recall (*REC*) and overall accuracy (*ACC*) (Marsland, 2015):

$$PRE = \frac{\#TP}{\#TP + \#FP} \times 100\% \quad (1)$$

$$REC = \frac{\#TP}{\#TP + \#FN} \times 100\% \quad (2)$$

$$ACC = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN} \times 100\% \quad (3)$$

where # means 'number of', *TP* and *TN* represent true positives (i.e., defective samples that were classified as defective) and true negatives (i.e., normal samples that were classified normal), respectively, and *FP* and *FN* represent false positives

and false negatives, respectively. *PRE* is the percentage of *TP* samples of all the samples classified as positive, where *REC* is the percentage of *TP* samples of all the positive samples (equivalent to the *TP* rate). The three metrics were calculated from the 30 training replicates. Moreover, statistical comparisons were conducted on *ACC*s for the different image inputs and classifiers, using the Fisher's least significance difference (LSD) procedure, coupled with the ANOVA test, at the 5% significance level.

Results and Discussion

Image Enhancement by BEMD

Figure 6 shows an example of the decomposition of a single DC image into six IMF images and a residual image by using BEMD. The first several IMFs captured the high-frequency features of the original image, while the last IMFs and the residual embodied its low-frequency features and basic trend. In particular, the first IMF could be seen as a noise component, carrying no useful information, while the residual basically corresponded to the illumination vignetting. Hence, by eliminating the first IMF image and the residual image, an enhanced DC image could be reconstructed.

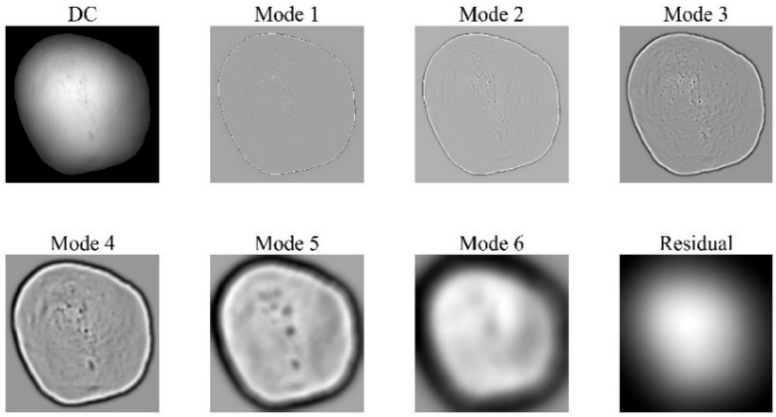

Figure 6. Bi-dimensional empirical mode decomposition of a direct component (DC) image into six mode images and one residual image.

As shown in Figure 7, compared with the original DC and AC images, their reconstructed counterparts possessed greatly improved uniformity and contrast, thus enhancing the clarity of image features like defects. Hence, the BEMD-enhanced DC and AC images were used for further image classification by all three classifiers.

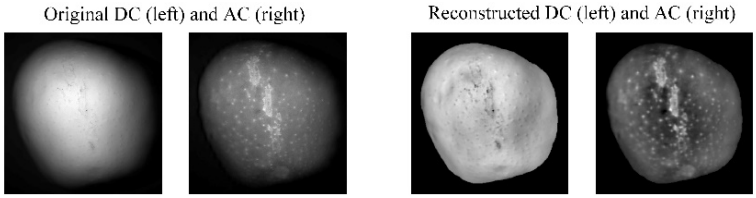

Figure 7. Comparison of the original and reconstructed direct component (DC) and amplitude component (AC) images.

Examples of Defective Apple Images

The Figure 8 shows examples of DC, AC and RT images, after the BEMD enhancement, for several defective 'Delicious' and 'Golden Delicious' apples. The DC and AC images revealed different types of defect features for these apples, while AC and RT images looked similar. Defects such as frost ring, and scab occur on the surface, while bitter pit develops beneath the fruit surface but is externally visible, and these defects could be well revealed in the DC images. On the other hand, the AC images were much more effective in ascertaining subsurface bruising/injury and rot, and certain russeting defect as shown in Figure 7.

It should be noted that, unlike fresh bruising which would appear as darkened areas, the bruises encountered in this study had been in existence for a prolonged period of time, thus resulting in brighter appearances due to the desiccation of water in the bruised tissue. Many defective apples had more than one type of defect, some of which were not visible from DC, AC or RT images alone. Hence the ensemble of DC, AC and RT images could have potential to increase discriminative features and thus classification accuracies.

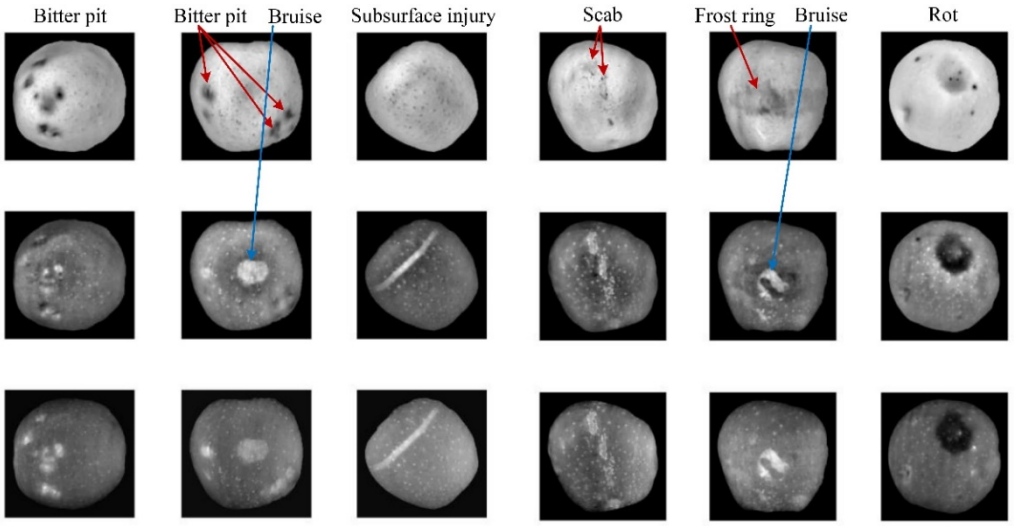

Figure 8. Examples of direct component (DC, top row), amplitude component (AC, middle row) and ratio (RT, bottom row) images of 'Delicious' (the left three columns) and 'Golden Delicious' (the right three columns) apples with different types of defect.

Classification Results

Figure 9 shows the results of PRE, REC and ACC by the three classifiers of RF, SVM and CNN for all five forms of image inputs (i.e., DC, AC, RT, DC-AC and DC-AC-RT) for 'Delicious' and 'Golden Delicious'.

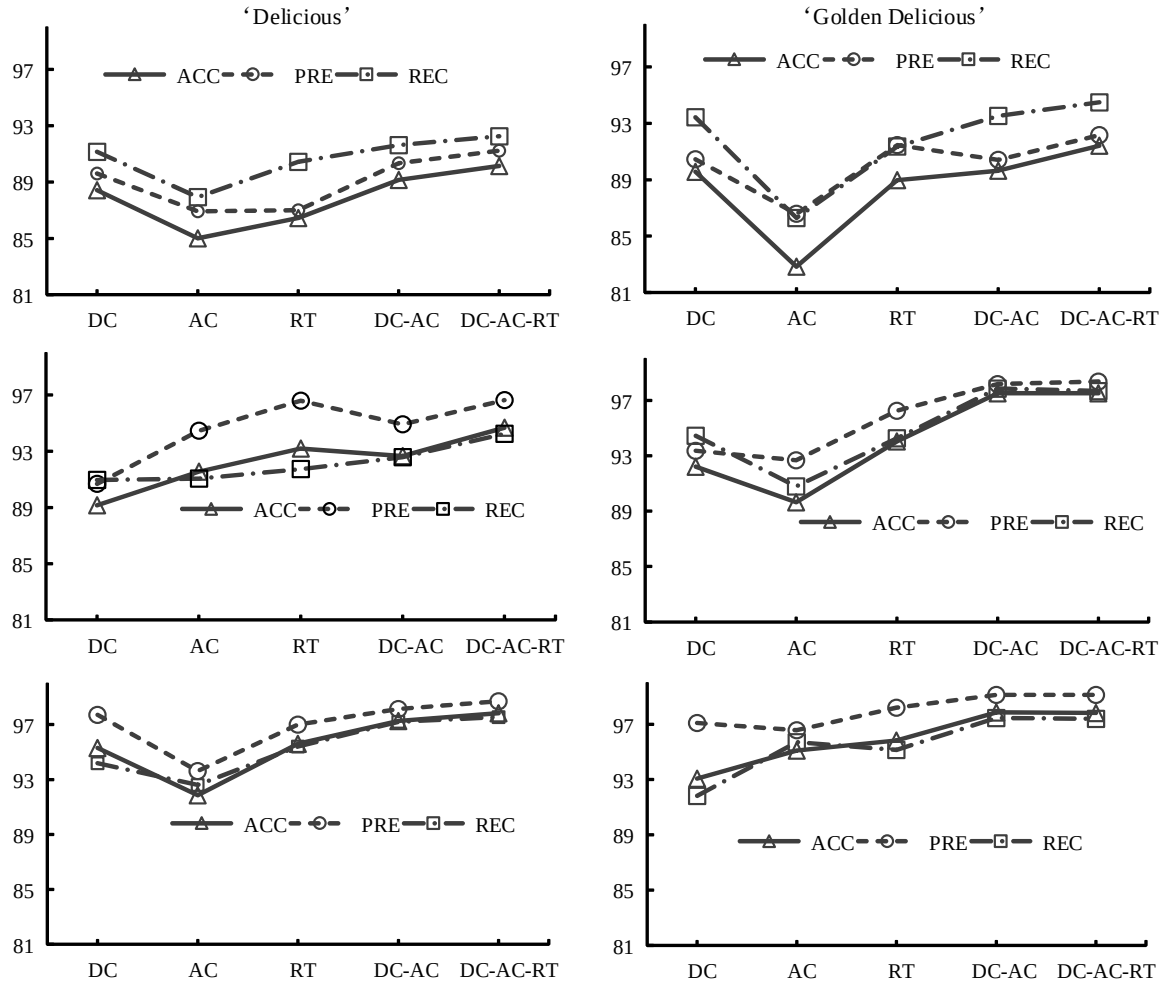

Figure 9. Classification accuracies (in percent) obtained by random forest (top), support vector machine (middle) and convolution neural network (bottom) for the five forms of image input for the test group of 'Delicious' (left pane) and 'Golden Delicious' (right pane) apples, where ACC, PRE and REC denote overall accuracy, precision and recall, respectively; DC, AC and RT denote direct component, amplitude component and ratio images, respectively; and DC-AC and DC-AC-RT represent the two ensembles of two or three types of images.

Detection accuracies, as measured by *PRE*, *REC* and *ACC*, varied with image input and classifier. The three performance

metrics had similar trends of change with the image input. Among the three metrics, *ACC* had the most conservative results, which indicated no accuracy paradox (i.e., good accuracy but bad predictive power with lower REC and PRE) in the classification problem. Hence only *ACC* is used in the following discussion for the model performance comparisons.

Figure 10 shows statistical comparisons of the classification results (in terms of *ACC*) by the three classifiers for the five forms of image inputs. DC and AC images did not consistently outperform each other, although DC had better results in 4 out of the six cases. These results are in general agreement with the study of Li (2016) that AC did not offer a distinct advantage over DC for surface defect detection of apples. RT images, however, compared favorably to both DC and AC images, except for one case when using RF for classifying 'Delicious' apples. Further, combining DC and AC images, i.e., concatenating the relevant features for the three classifiers of RF, SVM and CNN, improved the classification over using either DC or AC alone, statistically significantly in 3 out of the six cases. The overall improvements ranged between 2%-5%, which confirmed our hypothesis that the combination of AC and DC images provides additional discriminative features for enhanced defect detection. The combination of DC and AC images also compared favorably to RT images, resulting in significant improvements in 4 of the six cases. Moreover, the ensemble of DC, AC and RT (i.e., DC-AC-RT) images for the three classifiers achieved significant improvements in all cases for both varieties, when compared with AC, DC and RT alone, and it also significantly outperformed the ensemble of DC and AC (i.e., DC-AC) when using SVM for classifying 'Delicious' apples.

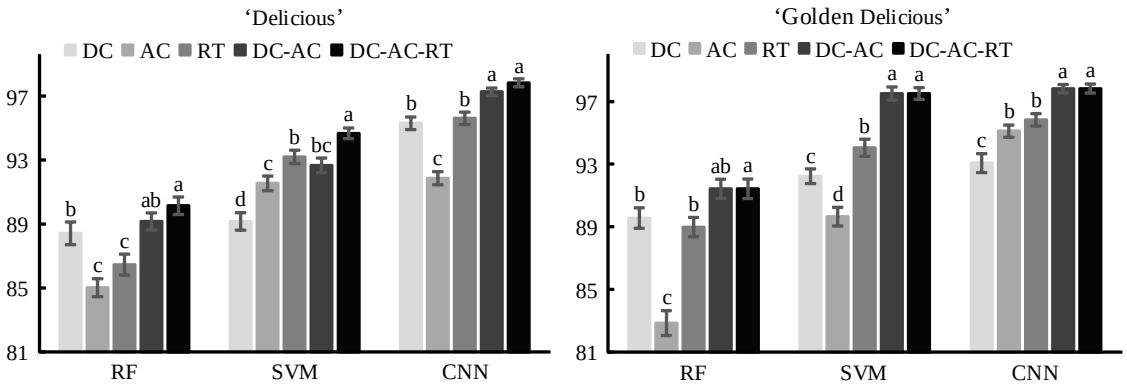

Figure 10. Comparisons of overall accuracies (in percent) by the three classifiers of random forest (RF), support vector machine (SVM) and convolutional neural network (CNN) for the five forms of image inputs for 'Delicious' and 'Golden Delicious' apples. DC, AC and RT denote direct component, amplitude component and ratio images, respectively, and DC-AC and DC-AC-RT represent two ensembles of DC and AC images and DC, AC, and RT images, respectively. Classification results for each classifier in the same group of five input forms with different letters are different at the 5% significance level.

Further comparisons for the three classifiers are shown in Figure 11. Overall, CNN was the best classifier; it was significantly better than RF for all 10 groups of comparisons, and also significantly better than SVM for 6 out of the 10 groups. When the ensemble of DC, AC, and RT images was used, CNN achieved overall accuracies of 98% for both varieties of apple. SVM performed significantly better than RF in all cases, except for the DC of 'Delicious' apples. SVM achieved the highest accuracies of 94.7% and 97.5% for 'Delicious' and 'Golden Delicious' apples, respectively, when using DC-AC-RT. The plausible reasons for the poorer performance of RF could be due to the relatively intuitively linear two-class classification and a smaller training dataset used in the study, which worked better for a linear SVM than for an RF that is intrinsically suited for multiclass problems involving a large number of training instances. It was also noted that RF gave relatively larger standard errors (i.e., longer error bars in the figure) than those obtained by SVM and CNN, which could be due to the limited training data for implementing the randomization in the RF algorithm.

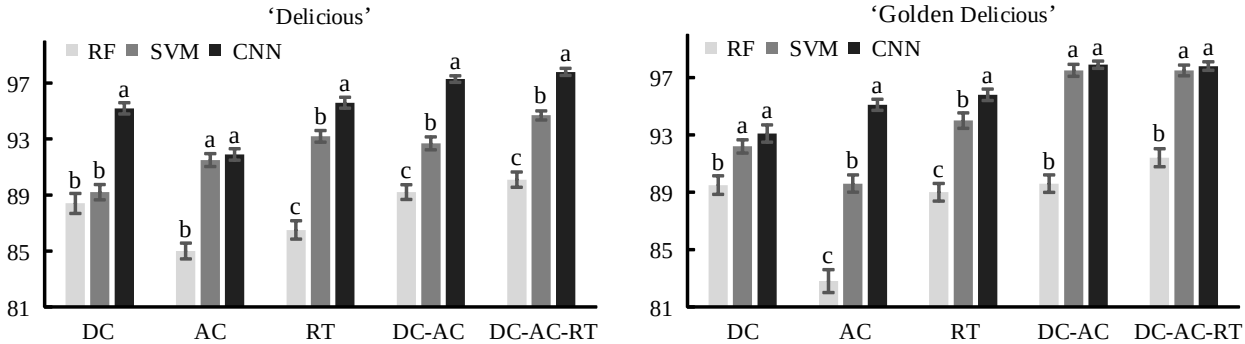

Figure 11. Comparisons of random forest (RF), support vector machine (SVM) and convolutional neural network (CNN) in terms of overall test accuracies (in percent) for the five different image inputs for 'Delicious' and 'Golden Delicious' apples. DC, AC and RT denote direct component, amplitude component and ratio images, respectively, and DC-AC and DC-AC-RT represent the two ensembles of DC and AC and DC, AC, and RT, respectively. Results in the same groups with different letters are different at the 5% significance level.

With the ensemble of DC-AC-RT images, CNN gave *REC*s of 97.6% and 97.4% for 'Delicious' and 'Golden Delicious' respectively, indicating less than 3% of (i.e., at most 2 on overage) defective apples misclassified as normal. Figure 12 shows DC, AC and RT images for the two most frequently misclassified apples for each variety, in the 30 training replications with random data partitioning. These apples contained only slight defects (i.e., old bruise, russeting and bitter pit) and had weak image contrast for the DC, AC and RT images, thus resulting in their misclassifications. It should be noted that the wavelength used in the study (i.e., 730 nm) was not optimized for detecting the types of defects observed in the apple samples, and hence it may not represent the optimum wavelength for enhanced detection.

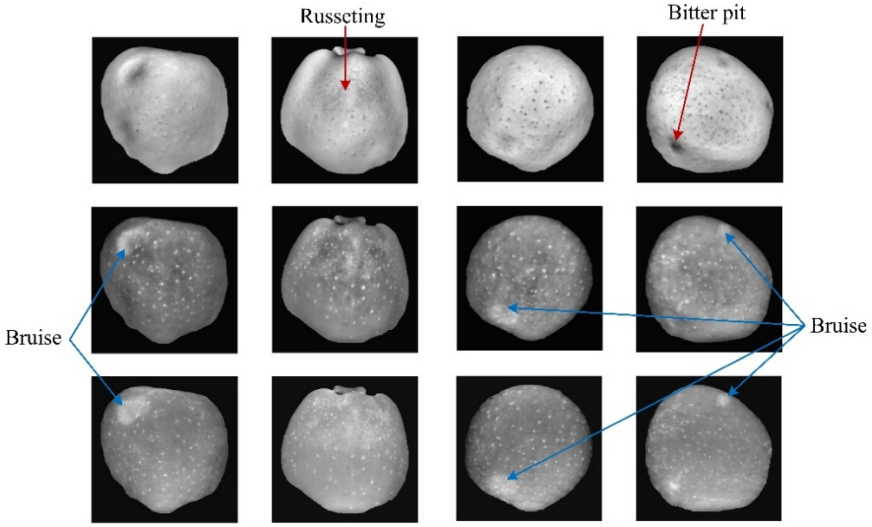

Figure 12. The direct component (DC, top row), amplitude component (AC, middle row) and ratio (RT, bottom row) images for the two defective 'Delicious' (the left two columns) and 'Golden Delicious' (the right two columns) apples which are misclassified by CNN using the ensemble of DC-AC-RT image input.

Discussion

The classification results for two varieties of apples, obtained by the three machine learning classifiers, demonstrated that using the ensembles of DC-AC and DC-AC-RT resulted in significant improvements in the overall detection accuracy, compared to using them separately. While AC images have been shown to be much more effective in detecting subsurface defects such as bruising, which are externally invisible (Y. Lu et al., 2016; Y. Lu & Lu, 2017a), DC images are generally better in detecting surface defects of apples. This further confirmed that DC and AC images provide different image features and hence are complementary and synergistic for defect detection. The classification results from this research also compare favorably with those reported in previous studies using conventional color or NIR imaging techniques. Unay et al. (2011) reported an overall classification accuracy of 93.5% for the two-class classification of 'Jonagold' apples using SVM combined with features selection. In classifying 'Fuji' apples, Zhang et al. (2015) obtained an accuracy of 95.6% using weighted relevance vector machine with features selection. Recently, Moallem et al. (2017) reported an accuracy of 92.5% for 'Golden Delicious' apples using SVM. It should be noted that in this study features selection was not implemented in using RF and SVM, which could further boost the accuracy. Hence, SIRI can be advantageous for fruit defect detection over conventional imaging technique under uniform illumination.

CNN provides a potent tool for image classification given its superior performance and unique advantage of integrating features extraction and classification. However, training a CNN model is by no means an easy task, because of its network architecture and many hyper-parameters that need considerable tuning effort and long training time. In this study, simplifications were made to focus on the number of convolution-pooling layers so as to speed up computations and thus ease the training task. Unlike SVM (and/or RF), CNN generally requires much more training samples per class, which may be problematic in preparing datasets since it can be difficult to collect sufficient samples for all types of defects. This was also the main reason that only two-class classification was performed in this research. Although this issue could be alleviated by artificial data augmentation, as done in this study, it is highly desirable and also beneficial to collect more real samples in order to realize the full potential of CNNs.

As aforementioned, this study only dealt with a two-class classification problem. In practical applications, it would be more desirable and valuable to perform multi-class (e.g., multiple levels of defect severity), even defect-specific classification and segmentation, which should be the subject of future research. In utilizing DC, AC and RT images, there are also other ways for integrating them for enhanced detection. For example, one may build models using them separately, and make classifications based on certain voting rules as done in ensemble learning. In the current SIRI system, images were taken from samples when they were in stationary status, and the image processing and analysis were performed off-line. Further research is needed to improve the software and hardware, so that SIRI can acquire and process patterned images from the moving samples in real time.

Conclusions

In this research, a multispectral SIRI system was used to obtain the DC and AC images of 'Delicious' and 'Golden Delicious' apples with different types of surface and subsurface defects. The proposed BEMD preprocessing method was found effective in removing vignetting and noise artifacts in both DC and AC images, thus enhancing subsequent image features extraction. While DC tended to be more useful for detection of surface defects and AC images were more capable of detecting subsurface defects, they had lower defect classification accuracies when used separately. However, by combining the DC and AC images, the three classifiers (i.e., RF, SVM and CNN) produced improved classification accuracies, and using the three sets of images, including the ratio of AC to DC, resulted in further improvements in the classification accuracy for both varieties of apples. Among the three machine learning classifiers, the CNN, trained with artificially augmented images, yielded the best classification results, with overall accuracies of 98% for both varieties of apples, followed by SVM and RF that were trained using extracted image features. Since SIRI enables acquiring both DC and AC images, which provide different, complementary features, the technique, coupled with an appropriate machine learning algorithm, has potential to be a new, effective, and versatile modality for defect detection of fruit and other food products.

Disclaimer

Mention of commercial products is solely for providing factual information, and it does not imply the endorsement of the products by USDA over those not mentioned.

Acknowledgements

The authors would thank Dr. Zhao Zhang for helping prepare apple samples.

References

Ariana, D., Guyer, D. E., & Shrestha, B. (2006). Integrating multispectral reflectance and fluorescence imaging for defect detection on apples. *Computers and Electronics in Agriculture, 50*, 148-161. http://doi.org/10.1016/j.compag.2005.10.002

Bargoti, S., & Underwood, J. P. (2017). Image segmentation for fruit detection and yeild estimation in apple orchards. *Journal of Field Robotics, 34*(6), 1039-1060. http://doi.org/10.1002/rob.21699

Bengio, Y. (2012). *Practical recommendations for gradient-based training of deep architectures*. Berlin: Springer.

Bennedsen, B. S., & Peterson, D. L. (2005). Performance of a system for apple surface defect identification in near-infrared images. *Biosystems Engineering, 90*(4), 419-431. http://doi.org/10.1016/j.biosystemseng.2004.12.005

Bennedsen, B. S., Peterson, D. L., & Tabb, A. (2007). Identifying apple surface defects using principle component analysis and artificial neural networks. *Transcation of the ASABE, 50*(6), 2257-2265. http://doi.org/10.13031/2013.24078

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., . . . Bengio, Y. (2010). *Theano: a cpu and gpu math complier in python*. Paper presented at the Proceedings of the Python for Scientific Computing Conference (SciPy). Retrieved from http://conference.scipy.org/proceedings/scipy2010/pdfs/bergstra.pdf

Bosch, A., Zisserman, A., & Munoz, X. (2007). *Image Classification using Random Forests and Ferns*. Paper presented at the ICCV 2007. IEEE 11th International Conference on Rio de Janeiro, Brazil.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32. http://doi.org/10.1023/A:1010933404324

Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., . . . Kumar, V. (2017). Counting apples and oranges with deep learning: a data-driven approach. *IEEE Robotics and Automation Letters, 2*(2), 781-788. http://doi.org/10.1109/LRA.2017.2651944

Chollet, F. (2015). Keras. https://github.com/keras-team/keras

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297. http://doi.org/10.1007/BF00994018

Criminisi, A., & Shotton, J. (2013). *Decision forests for computer vision and medical image analysis*. London: Springer.

Crowe, T. G., & Delwiche, M. J. (1996). Real-time defect detection in fruit .1. design concepts and development of prototype hardware. *Transcation of the ASAE, 39*(6), 2299-2308. http://doi.org/10.13031/2013.27740

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1979). Textural features for image classification. *IEEE Transactions on*

Systems, Man and Cybernetics SMC-3(6), 610-621. http://doi.org/10.1109/TSMC.1973.4309314

Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory, 8*(2), 179-187. http://doi.org/10.1109/TIT.1962.1057692

Huang, N. E., & Shen, S. S. P. (2014). *Hilbert-Huang transform and its applications*. Singapore: World Scientific.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zhang, Q., . . . Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. of Royal Society of London A, 454*, 909-995.

Huang, W., Li, J., Wang, Q., & Chen, L. (2015). Development of a multispectral imaging system for online detection of bruises on apples. . *Journal of Food Engineering, 146*, 62-71. http://doi.org/10.1016/j.jfoodeng.2014.09.002

Kavdir, I., & Guyer, D. E. (2004). Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. *Biosystems Engineering, 89*(3), 331-344. http://doi.org/10.1016/j.biosystemseng.2004.08.008

Kavdir, I., & Guyer, D. E. (2008). Evaluation of different pattern recognition techniques for apple sorting. *Biosystems Engineering, 99*, 211-219. http://doi.org/10.1016/j.biosystemseng.2007.09.019

Kleynen, O., Leemans, V., & Destain, M. F. (2005). Development of a multi-spectral vision system for the detection of defects on apples. *Journal of Food Engineering, 69*(1), 41-49. http://doi.org/10.1016/j.jfoodeng.2004.07.008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Paper presented at the Advances in Neural Information Processing Systems 25.

Kumar, A., & Pang, G. K. H. (2002). Defect detection in textured materials using Gabor filters. *IEEE Transactions on Industry Applications, 38*(2), 425-440. http://doi.org/10.1109/28.993164

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. of The IEEE, 86*(11), 2278-2324. http://doi.org/10.1109/5.726791

Leemans, V., & Destain, M. F. (2004). A real-time grading method of apples based on feature extracted from defects. *Journal of Food Engineering, 61*, 83-89. http://doi.org/10.1016/S0260-8774(03)00189-4

Leemans, V., Magein, H., & Destain, M. F. (1998). Defects segmentation on 'Golden Delicious' apples by using colour machine vision. *Computers and Electronics in Agriculture, 20*, 117-130. http://doi.org/10.1016/S0168-1699(98)00012-X

Leemans, V., Magein, H., & Destain, M. F. (1999). Defect segmentation on 'Jonagold' apples using colour vision and a Bayesian classification method. *Computers and Electronics in Agriculture, 23*, 43-53. http://doi.org/10.1016/S0168-1699(99)00006-X

Li, R. (2016). *Development of a structured illumination reflectance imaging system for enhanced detection subsurface and surface defects in apple fruit.* (MS), East Lansing, Michigan, Michigan State University, Department of Biosystems and Agricultural Engineering

Lu, R. (2003). Detection of bruises on apples using near-infrared hyperspectral imaging. *Transcation of the ASAE, 46*(2), 523-530. http://doi.org/10.13031/2013.12941

Lu, Y., Li, R., & Lu, R. (2016). Structured-illumination reflectance imaging (SIRI) for enhanced detection of fresh bruises in apples. *Postharvest Biology and Technology, 117*, 89-93. http://doi.org/10.1016/j.postharvbio.2016.02.005

Lu, Y., & Lu, R. (2017a). Development of a multispectral structured-illumination reflectance imaging (SIRI) system and its application to bruise detection of apples. *Transcation of the ASABE, 60*(4), 1379-1389. http://doi.org/10.13031/trans.12158

Lu, Y., & Lu, R. (2017b). Histogram-based automatic thresholding for bruise detection of apples by structured-illumination reflectance imaging. *Biosystems Engineering, 160*, 30-41. http://doi.org/10.1016/j.biosystemseng.2017.05.005

Lu, Y., & Lu, R. (2017c). Non-destructive defect detection of apples by spectroscopic and imaging technologies:a review. *Transactions of the Asabe, 60*(5), 1765-1790. http://doi.org/10.13031/trans.12431

Marsland, S. (2015). *Machine learning: an algorithmic perspective*. Boca Raton, FL, USA: CRC Press.

Mehl, P. M., Chao, K., Kim, M., & Chen, Y. R. (2002). Detection of defects on selected apple cultivars using hyperspectral and multispectral image analysis. *Applied Engineering in Agriculture, 18*(2), 219-226. http://doi.org/10.13031/2013.7790

Mehl, P. M., Chen, Y. R., Kim, M. S., & Chan, D. E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering, 61*, 67-81. http://doi.org/10.1016/S0260-8774(03)00188-2

Moallem, P., Serajoddin, A., & Pourghassem, H. (2017). Computer vision-based apple grading of golden delicious apples based on surface features. *Information Processing in Agriculture, 4*(1), 33-40. http://doi.org/10.1016/j.inpa.2016.10.003

Mohanty, S., Hughes, D. P., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Frontier in Plant Science, 7*, 1419. http://doi.org/10.3389/fpls.2016.01419

Nixon, M. S., & aguado, A. S. (2012). *Feature extraction & image processing for computer visiion, third edition*. Singapore: Elsevier Ltd.

Ojala, T., Pietikainen, M. P., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transcations on Pattern Analysis and Machine Intelligence, 24*(7), 971-987. http://doi.org/10.1109/TPAMI.2002.1017623

Pietikainen, M., Hadid, A., Zhao, G., & Ahonen, T. (2011). *Computer vision using local binary patterns*. London: Springer.

Rosin, P. L. (2001). Unimodal thresholding. *Pattern Recognition, 34*(11), 2083-2096. http://doi.org/10.1016/S0031-3203(00)00136-9

Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. *ICDAR, 3*, 958-962.

Throop, J. A., Aneshansley, D. J., Anger, W. C., & Peterson, D. L. (2005). Quality evaluation of apples based on surface defects: development of an automated inspection systems. *Postharvest Biology and Technology, 36*, 281-290. http://doi.org/10.1016/j.postharvbio.2005.01.004

Unay, D., & Gosselin, B. (2006). Automatic defect segmentation of 'Jonagold' apples on multispectral images: a comparative study'. *Postharvest Biology and Technology, 42*, 271-279. http://doi.org/10.1016/j.postharvbio.2006.06.010

Unay, D., & Gosselin, B. (2007). Stem and calyx recognition on 'Jonagold' apples by pattern recognition. *Journal of Food Engineering, 78*, 597-605. doi: 10.1016/j.jfoodeng.2005.10.038

Unay, D., Gosselin, B., Kleynen, O., Leemans, V., Destain, M. F., & Debeir, O. (2011). Automatic grading of Bi-colored apples by multispectral machine vision. *Computers and Electronics in Agriculture, 75*(1), 204-212. http://doi.org/10.1016/j.compag.2010.11.006

Varghese, A. Z. (1992). *Feasibility of using machine vision for automated inspection of apples.* (PhD), The Pennsylvania State University.

Xing, J., & De Baerdemaeker, J. (2005). Bruise detection on 'Jonagold' apples using hyperspectral imaging. *Postharvest Biology and Technology, 37*(2), 152-162. http://doi.org/10.1016/j.postharvbio.2005.02.015

Xing, J., Saeys, W., & De Baerdemaeker, J. (2007). Combination of chemometric tools and image processing for bruise detection on apples. *Computers and Electronics in Agriculture, 56*, 1-13. http://doi.org/10.1016/j.compag.2006.12.002

Xu, J. L., & Sun, D. W. (2018). Computer vision detection of salmon muscle gaping using convolutional neural network features. *Food Analytical Methods, 11*(1), 34-47. http://doi.org/10.1007/s12161

Yang, Q. (1994). An approach to apple surface feature detection by machine vision. *Computers and Electronics in Agriculture, 11*(2-3), 249-264. doi: 10.1016/0168-1699(94)90012-4

Zeile, M. Z. (2012). Adadelta: an adaptive learningrate method. Retrieved from https://arxiv.org/abs/1212.5701

Zhang, B., Huang, W., Gong, L., Li, J., Zhao, C., Liu, C., & Huang, D. (2015). Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. *Journal of Food Engineering, 146*, 143-151. http://doi.org/10.1016/j.jfoodeng.2014.08.024