

Detection Algorithm for Cracks on the Surface of Tomatoes using Multispectral Vis/NIR Reflectance Imagery

Danhee Jeong¹, Moon S. Kim², Hoonsoo Lee³, Hoyoung Lee², Byoung-Kwan Cho^{3*}

¹Deasang central R&D center, 125-8, Pyogyo-ri, Majang-myeon, Icheon, Gyeonggi-do, 467-813, South Korea

²Environmental Microbiology and Food Safety Laboratory, Agricultural Research Service, U.S.

Department of Agriculture, Powder Mill Road, Bldg 173, BARC-East, Beltsville, MD 20705, U.S.A.

³Department of Biosystems Machinery Engineering, Chungnam National University, Daehak-ro 99, Yuseong-gu, Daejeon, 305-764, South Korea

Received: July 29th, 2013; Revised: August 7th, 2013; Accepted: August 16th, 2013

Abstract

Purpose: Tomatoes, an important agricultural product in fresh-cut markets, are sometimes a source of foodborne illness, mainly *Salmonella* spp. Growth cracks on tomatoes can be a pathway for bacteria, so its detection prior to consumption is important for public health. In this study, multispectral Visible/Near-Infrared (NIR) reflectance imaging techniques were used to determine optimal wavebands for the classification of defect tomatoes. **Methods:** Hyperspectral reflectance images were collected from samples of naturally cracked tomatoes. To classify the resulting images, the selected wavelength bands were subjected to two-band permutations, and a supervised classification method was used. **Results:** The results showed that two optimal wavelengths, 713.8 nm and 718.6 nm, could be used to identify cracked spots on tomato surfaces with a correct classification rate of 91.1%. The result indicates that multispectral reflectance imaging with optimized wavebands from hyperspectral images is an effective technique for the classification of defective tomatoes. **Conclusions:** Although it can be susceptible to specular interference, the multispectral reflectance imaging is an appropriate method for commercial applications because it is faster and much less expensive than Near-Infrared or fluorescence imaging techniques.

Keywords: Food safety, Hyperspectral image, Image processing, Multispectral image, Tomato cracks

Introduction

Food safety has become an extremely important issue throughout countries. Foodborne illnesses in the United States cost approximately \$10–\$83 billion annually. Contamination, bacterial infection, and defective products are common causes threatening food safety. Food poisoning cannot be completely eradicated because of the diverse characteristics of the bacteria involved in the various conditions for growth, life cycle, and incubation period. *Salmonella* spp. is a major food poisoning pathogen worldwide. Over 1.4 million outbreaks of disease associated

with *Salmonella* occur annually in the United States (CDC, 2007). Contaminated foods and water have been known to have a high potential for the transmission of *Salmonella* infection, and those having weak immune systems such as infants and the elderly are especially prone to infection. To prevent *Salmonella* infection, food can be heated by some means before consumption because *Salmonella* cannot withstand high temperatures even if it is present in the food. Even so, *Salmonella* remains a risk due to the increasing market share of fresh-cut products; there is a drift towards the fresh-cut market because of consumer demands for convenience and concern for their health. However, microorganisms are serious contaminants in minimally processed fresh-cut food and this raises problems within every fresh-cut market (Beuchat, 1996). Due to

*Corresponding author: Byoung-Kwan Cho

Tel: +82-42-821-6715; Fax: +82-42-823-6246

E-mail: chobk@cnu.ac.kr

the minimum treatment of these products, bacteria growth is imminent, thus becoming a potential menace for public health (Ahvenainen, 1996; Francis et al., 1999).

Tomato is one of the most critical agricultural products and its industries account for more than \$2 billion annually in the United States. The tomato is also a typical fresh-cut vegetable, and it can cause an outbreak of food poisoning in its uncooked condition. A recent report in the Centers for Disease Control and prevention (CDC) during 1990–2009 shows that outbreaks of food poisoning is associated with tomatoes that infected pathogen such as *Salmonella*, *S. Typhimurium*, *S. Newport*, *S. Saintpaul*, *S. Javiana* (CDC, 2011). The origin of a pathogen infection is the surrounding farm environment. Generally, in vegetable irrigation systems, especially for tomatoes, water is sprayed by a drip irrigation system. At this point, water contaminated with pathogen can touch the surface of vegetable. In addition, pathogen may infiltrate into the tomato through a stem scar where it either survives or dies slowly (Guo et al., 2002). One study reported that once food poisoning bacteria invades a tomato through a crack in the surface of the tomato, it forms a colony (Hedberg et al., 1999).

When 30-60% of the tomato surface turns red or pink, the tomato cracks occur easily. According to the standards for grades of fresh tomato in the United States, if the aggregated growth cracks lengths are within 2 1/2 inch, the tomato may be classified U.S. No. 1 grade (USDA, 1991). However, these cracks may lengthen or deepen during transport, tomato ripening, or conditions in the environment after harvest. Occasionally, they may become contaminated with harmful bacteria. That is, a cracked tomato may provide the proper medium for a bacterial inoculum (Lin and Wei, 1997; Beuchat, 2002). Many solutions such as water rinsing, chlorine, organic acids, ozone, and irradiation have attempted to eliminate bacteria from fresh-cut products. However, these methods cannot eradicate all bacteria and the usage of inappropriate dosage for sanitizers is able to survive some bacteria (Wei, 1995; Beuchat, 2002; Asplund, Williams, 2012 and Nurmi, 1991). Due to the wide distribution of products in the United States, finding the origin of fresh-cut products associated with a foodborne outbreak is very difficult (De Roeber, 1999). Therefore, before the packing and shipping stage, detecting defective tomatoes with cracks on the surface is important. Also those having cracks should be discarded.

Imaging techniques have been developed as an inspection

method for the quality and safety assessment of a variety of agricultural food products. Among these imaging techniques, hyperspectral/multispectral imaging has been developed as a potentially non-destructive method to assess the quality of various agricultural products such as apples, cucumbers, and cantaloupes (Cho et al., 2011; Liu et al. 2006; Vargas et al., 2005). It has been shown as is an effective tool for detecting surface defects and the contamination of agricultural products. For this study, hyperspectral reflectance images were collected from samples of naturally cracked tomatoes. The aim of the present study is to identify the specific wavelength bands that could be applied to detect cracked tomatoes while covering the range of color grades of tomatoes. To assess the process rigorously, the selected wavelength bands were subjected to two-band permutations, and a supervised classification method was used to classify the resulting images.

Material and Methods

Tomatoes

Tomatoes with ($n = 120$) and without ($n = 120$) growth cracks were purchased from a local farm located in Beltsville, Maryland, USA. When collecting the samples, we chose tomatoes with cracked spots of aggregate lengths between 2 1/2 inch and 5 inch because the United States Standard for Grades of Fresh Tomatoes regards tomatoes with aggregate crack lengths less than 2 1/2 inch as the highest grade, and tomatoes with aggregate crack lengths greater than 5 inch are unusable for commercial purposes. Samples consisted of various colors from tomato color classification requirements of the USDA as follows: green, breakers, turning, pink, light red, and red. Prior to obtaining hyperspectral images, samples were cleaned with a paper towel to remove dust and moisture.

Hyperspectral Imaging System

Figure 1 shows a schematic diagram of the hyperspectral Vis/NIR imaging system. It consists of an imaging spectrograph (VNIR Concentric Imaging Spectrograph, Headwall photonics, Fitchburg, MA, USA) and an electron multiplying charge-coupled device (EMCCD: Luca R DL-604M, Andor Technology, South Windor, CT, USA) mated with a C-mount object lens (F1.9 35mm compact lens, Schneider Optics, Hauppauge, NY, USA). The lighting module consists of a halogen lamp,

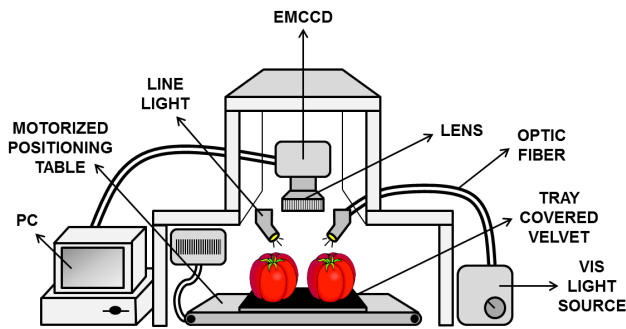


Figure 1. Hyperspectral Imaging System.

power supply (Fiber-Lite DC-950, Dolan-Jenner Industries Inc., MA, USA), and an optical fiber which transfers the light source to the line light reflectors. The sample module has a plate upon which the sample is set and is controlled by a motorized device (Velmex Inc. NY, USA) to move samples line by line. The hyperspectral imaging system is connected to a personal computer to analyze the data from the system.

Hyperspectral Image Acquisition

The interface software, developed using Microsoft Visual Basic 6.0 (Microsoft, WA, USA), controls the system and acquires raw data. Four samples are put into rubber cups for one scan and then the image is divided into four sections, one for each tomato. The rubber cups are covered with black velvet to minimize effect of unwanted reflectance. The exposure and acquisition time were set at 5 ms. References for dark current and white diffuse reflectance were recorded before proceeding with the acquisition of the hyperspectral images of tomatoes. During the scan, raw reflected intensities were recorded line by line.

Image Processing Procedure

The obtained hyperspectral images were analyzed using MATLAB software (Mathworks, MA, USA). First, five regions of interest (ROIs) were chosen using MATLAB's drawing mode for every sample (Figure 3) and were labeled as the following: specular, stem, normal, crack, and background. Classification of whether the sample had a cracked spot or not was conducted using dual-band image processing. For this purpose, relative ROIs were selected using the mean value of the maximum differences of reflectance intensity between selected ROIs. Simultaneously, two wavelength bands for each ROI were found. The images based on the selected wavelength bands were subjected to dual-band image processing. Every image was thresholded

using Otsu's method after the dual-band image processing. Several further image processing methods were used to eliminate errors that arose from deep cracks. Tiny holes caused by noise were filled, and then regions were ordered by area. Only the largest was retained. The final images were used to estimate whether the sample was a cracked or normal tomato. Since the cracks are mainly linked to the stem of tomatoes, two feature values (roundness and differential distance value) were calculated from the stem shapes on the images to classify cracked and normal tomatoes. For this assessment, the roundness and different distances for the final image were calculated. Roundness was calculated using the following equation:

$$\text{Roundness} = 4 \times \pi \times \text{area}/\text{perimeter}^2 \quad (1)$$

where the area was obtained by measuring the boundary of the final image and the perimeter was obtained using estimated differences and approximate derivatives in the final image. Overall, the roundness value was obtained by MATLAB's method for identifying round objects. In this case, if the value of roundness is 1, it is regarded as a circle in the final image. If it is far from 1, it is regarded as a cracked spot.

The different distance from the center of the final image that corresponded with the stem of the tomato was also obtained using MATLAB methods for calculating distances between points on a circle. Here, a larger value describes a final image that is not uniform and can indicate a cracked spot.

Both values were used as features of a Linear Discriminant Analysis (LDA) model to prove the distinguishing capacity of the proposed algorithm. The data from each group (i.e., the cracked and normal sample sets) were divided by 70% ($n = 84$) and 30% ($n = 36$) for calibration and validation, respectively.

Results and Discussion

Hyperspectral Reflectance Images

The cracked spots in the tomato images can be recognized by visual observation, as shown in Figure 2, over a broad range of wavelength bands. In accordance with this observation, cracked spots generally appear darker than other areas, such as the normal surface and the specular regions, but they do have a similar brightness to the stem

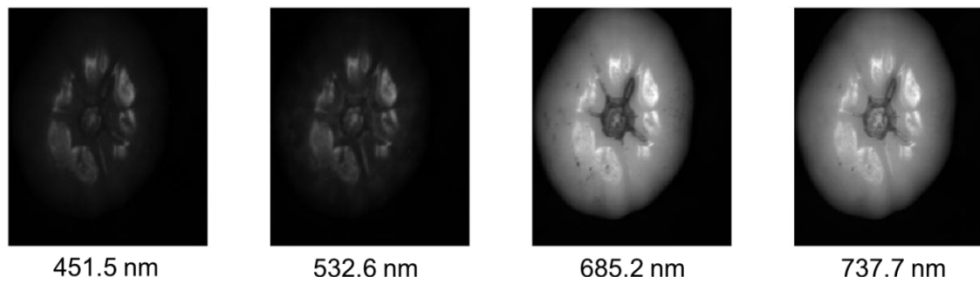


Figure 2. Representative images of cracked tomatoes at several wavelengths.

area. Hence, the algorithm to detect cracked spots was developed to focus on the distinguishing characteristics of cracked spots and stems from other regions of the image.

Regions of Interest

Selected ROIs of cracked tomatoes and a plot of the spectra from each region are shown in Figure 3. The spectral differences between the higher reflectance intensities of the normal surface and spectra of cracked spots were found at wavelengths from 441.9 to 594.6 nm and longer than 952.4 nm. Generally, the spectra of normal regions have higher intensities than the cracked regions. The color variations in tomatoes seem to cause a different response of reflectance intensity because the individual colors of the normal surface generate spectra with different intensities. In addition, the physical length of the cracked spot on the surface will also affect the result. There was some diversity in the features of the cracked spots such as length or depth, and they could sometimes show similar intensities to the spectra of a normal surface. For these reasons, using a single wavelength band for image processing

is limited because there can be unexpected common features. In other words, there can be exceptions. Other research has found that band ratio images can discriminate target spots from other regions better than single band images, even if the surfaces of the samples vary by color (Cho et al., 2011). With this in mind, dual wavelength bands were used for the image processing.

Band Ratio Images

To find optimal wavelength bands, two ROIs combinations, specular/stem and crack/stem were selected. They showed consistent intensity differences even if the surface colors varied. Hence, they were the best choice to use for image processing to minimize unwanted results. The spectra of normal surfaces were not used for image processing because these spectra showed differences related to color. The spectra of the background always showed the lowest reflectance; hence, it did not affect the results and was excluded.

After selecting the ROIs, we determined two specific wavelength bands by finding the largest difference of reflectance intensity between each pair of ROIs. The largest intensity gap increases contrast between the

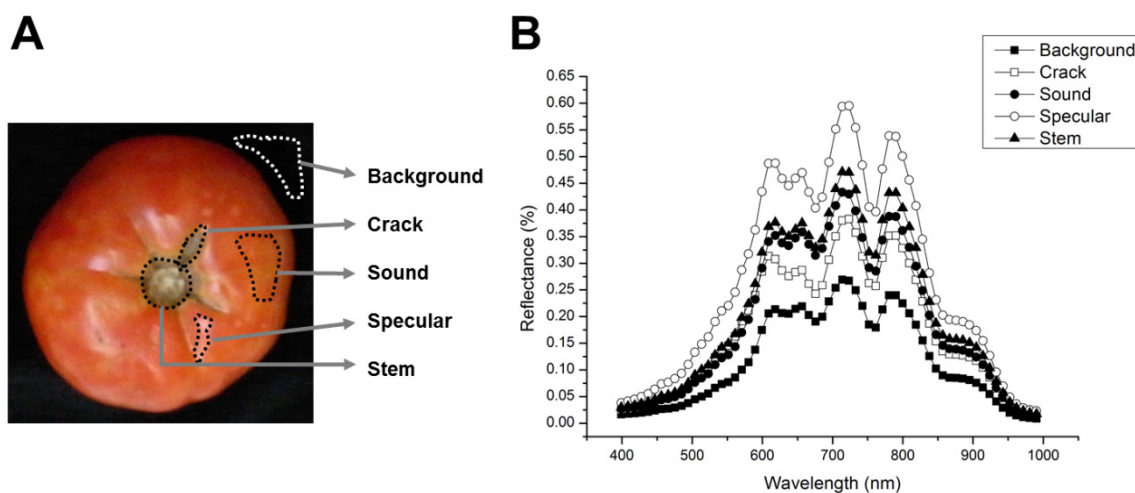


Figure 3. Five selected regions of interest (ROIs) from cracked tomatoes (A) and representative reflectance spectra of the ROIs (B).

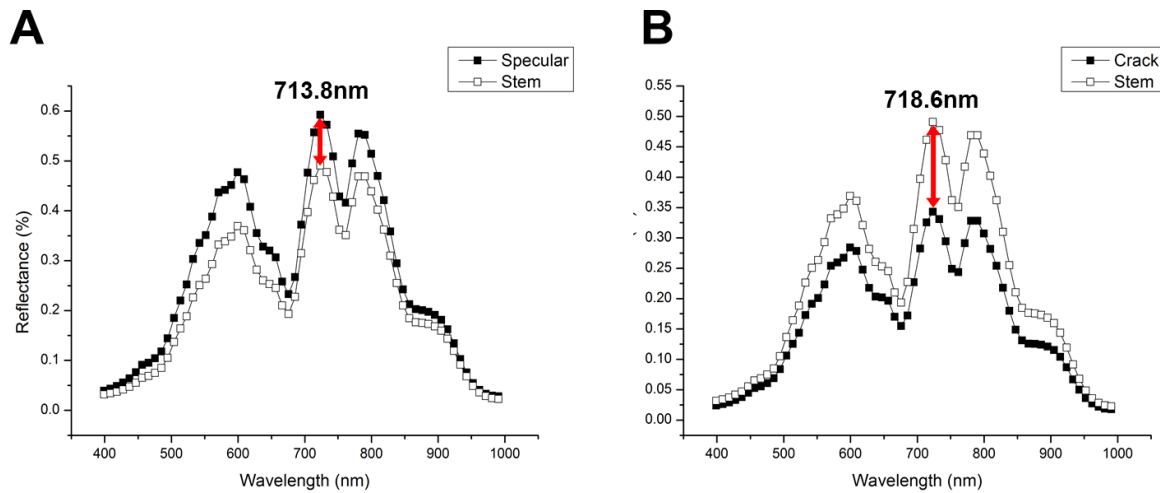


Figure 4. Spectra of specular/stem (A) and crack/stem (B). The wavelength bands marked in red indicate the largest gap between selected ROIs.

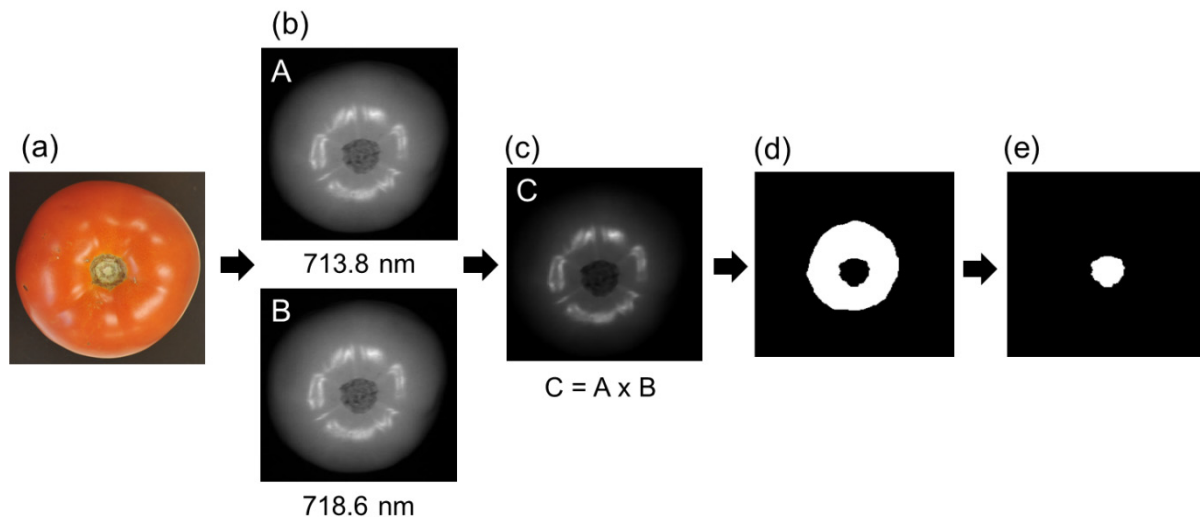


Figure 5. Dual-band image processing on tomatoes from the normal group. Image (a) is an RGB image of a normal tomato, (b) is a hyperspectral images of selected wavelengths (713.8 and 718.6 nm), (c) is the result of dual-band image processing using the two images in (b), (d) is the result after thresholding image (c), and (e) is the final image of the proposed algorithm.

areas in the resulting image. The wavelengths 713.8 nm and 718.6 nm were selected as the most discriminatory wavelength bands between the specular and stem regions and between the crack and stem regions, respectively. The difference in spectra and the selected wavelengths of the two pairs of ROIs are shown in Figure 4.

Figure 5 shows the simple procedure of dual-band image processing on normal tomatoes. Dual-band images were obtained by multiplying images at 713.8 nm (R713.8) by 718.6 nm (R718.6). Otsu’s method was then used to set threshold values and get a binary image from the processed grayscale images. Manually setting the threshold value was not considered because it could not separate the wanted images from the dual-band images.

The image processing procedure on the cracked tomato group was conducted as for the normal tomato group, as shown in Figure 6; however, there were uncontrolled results when the algorithm encountered some of the cracked samples. In particular, it occurred when images prior to the thresholding process had an excessively large cracked spot.

The cracked spots had a variety of features such as length, depth, and color. Among them, the length and depth were the factors that could cause unwanted image results. Specifically, when the crack stretched to the edge of the tomato, open spaces appeared in the images after thresholding. They were the main reason for the failed results because after the image filling process was applied,

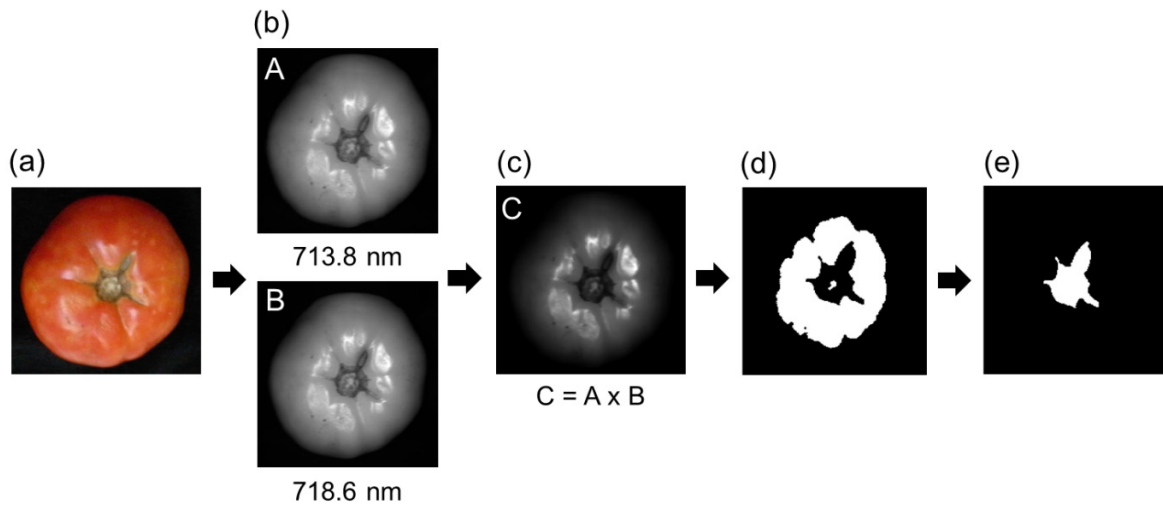


Figure 6. Band ratio image processing on tomatoes from the cracked group. Image (a) is an RGB image of a cracked tomato, (b) is the hyperspectral images of selected wavelengths (713.8 and 718.6 nm), (c) is the result of band ratio image processing using the two images in (b), (d) is the result after thresholding image (c), and (e) is the final image of the proposed algorithm.

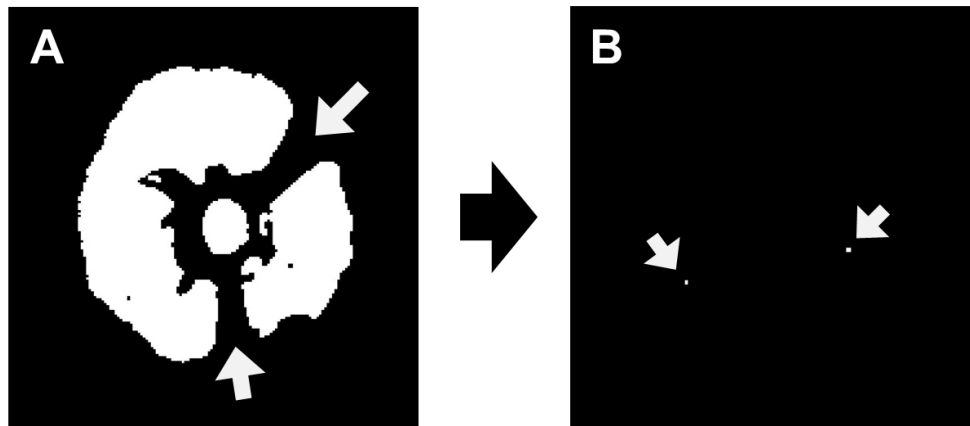


Figure 7. Failed result when the cracked spot is excessively large. Image (A) has been thresholded by Otsu's method and (B) is an image after applying additional image processing such as filling and other methods. Image (A) shows the opened spaces that caused the failed crack detection in (B).

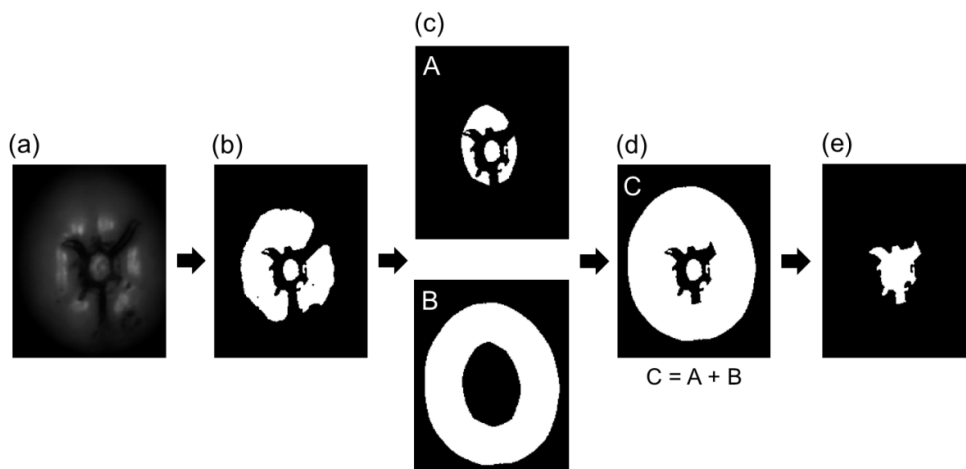


Figure 8. Dual-band image processing with the added mask. Image (a) is the result of dual-band image processing, (b) is the result of thresholding, and (c) shows the added mask. In (c), a mask is added to image (A), and (B) is the filled image within the boundary of the sample except for the filled (A) image. Image (d) is the sum of (A) and (B), and (e) is the final image.

there were either tiny spots or nothing. Open spaces allow the filling-in of the entire image area (i.e., Figure 7). Therefore, a necessary adjustment using a mask layer to close any open spaces was added to the procedure.

Figure 8 shows the adjustment that was used to prevent these exceptions. The edge of the sample boundary was used to devise a mask that closed the open spaces. Based on the edge, a mask was set to be slightly smaller than the original size and then placed on the image. The opened spaces were shut by the mask so that almost all results showed an obvious cracked image. Although the final image does not indicate the full length of the cracked spot, it does not miss existing cracks on the samples.

Classification Ability

Figure 9 shows the roundness values and normal distributions of samples from the final images. Roundness

was defined such that a value of 1 corresponded to a perfectly circular shape. The average values of roundness were 0.83 and 0.41 for 120 normal and 120 cracked tomato images, respectively. There were some outliers (five samples) in the normal tomato group because the stem did not appear as a perfect circle in the final image. In general, however, almost all normal tomato samples show a high value of roundness. Hence, 108 samples (90%) of the normal group had a roundness value between 0.8 and 1. These results support the conclusion that the developed algorithm can reasonably be applied to normal tomatoes. The values of roundness for cracked samples were significantly different to the normal samples. Of the cracked samples, 83 (69.2%) had a roundness value of at most 0.5 and there were no samples with a roundness value greater than 0.85.

Figure 10 shows that the results of different distance are less marked than the results of roundness. Different distance is defined as the distance between the centroid of the final image and every point of its edge. We expect the final images of the cracked group have non-uniform boundaries and have a high value of different distance. The results showed that the means of 120 normal and 120 cracked images were 57.6 and 77.2, respectively, supporting our expectation.

Using both features of the final image—roundness and different distance—was expected to give better detection results using an LDA model. LDA is used in machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events.

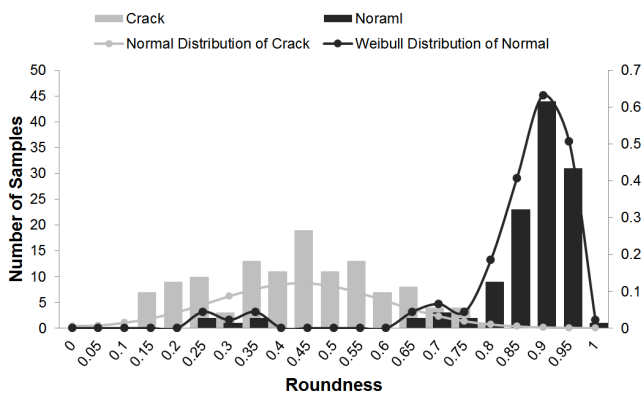


Figure 9. Weibull distribution and normal distribution for roundness of each group of tomatoes. Results are for 120 samples each of normal and cracked tomato groups. Values of roundness close to 1 indicate shapes close to circular.

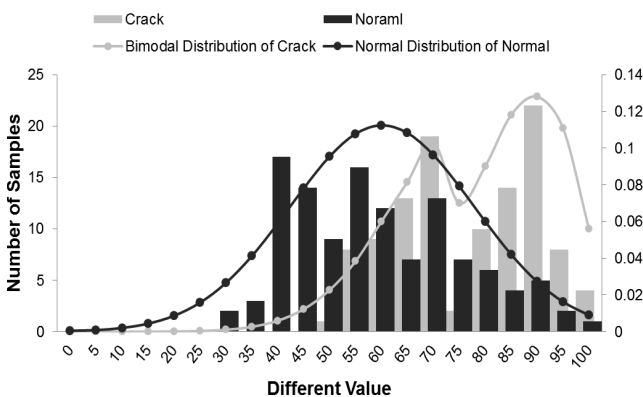


Figure 10. Bimodal distribution and normal distribution for different values of each group of tomatoes. Results are for 120 samples each of normal and cracked tomato groups. When the value of different distance is larger, it has non-uniformity in the final images.

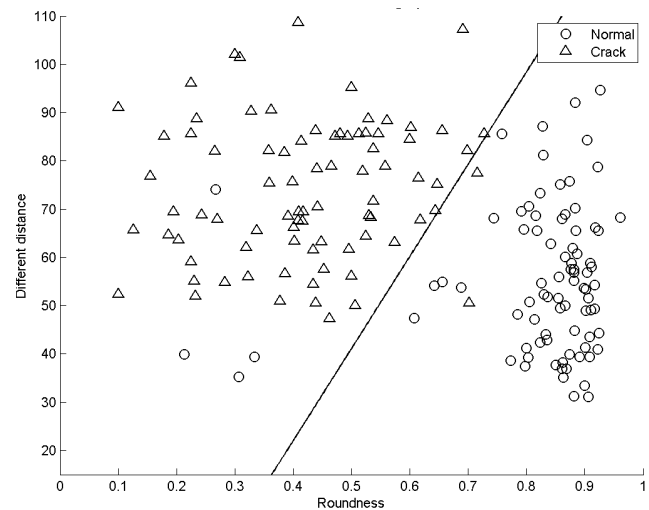


Figure 11. LDA results between roundness and different distance.

Table 1. Classification accuracy of LDA model

Classification Accuracy of Calibration set (%) (n = 84)	Classification Accuracy of Test set (%) (n = 36)
92.03	91.11

The results after applying LDA to both features are shown in Figure 11. The LDA model estimated an appropriate separation border between the normal and cracked groups. First, the calibration set was formed from randomly selected samples of 70% of each group. Based on the calibration result, 30% of the entire sample set was tested randomly. This procedure was repeated twenty times, and the results are shown in Table 1.

The mean of the accuracy was 92.0% for the calibrated results and 91.1% for the random tests. Overall, our proposed simple algorithm can detect cracked spots on the tomato surface with high accuracy without being affected by the diverse colors of tomatoes.

Conclusion

The proposed simple algorithm for detecting cracked spots on the surface of tomatoes achieves its purpose and can be applied to every color of tomato. However, reflectance multispectral images have a minor drawback that can be solved by applying a mask during the image processing procedure. The masked images give a better result although the entire cracked spot does not appear in the final image. The essential aim of this study is to separate tomato samples into cracked and normal groups. In accordance with algorithm that we developed in this study, we can detect cracked samples with a rate of success of greater than 90%.

We suggest further studies that should investigate other methods; e.g. using physical filters in the light source to minimize the specular. It may also help to develop simple image processing methods. Adding extra image processing increases the handling time and decreases productivity in commercial applications. Using other techniques such as NIR or fluorescence imaging could also be considered, but the image acquisition time can be longer than for hyperspectral imaging. Even if the present study indicates limited results that do not show perfectly detected cracks in the image, it does show reasonable classification ability and it is likely that the developed algorithm can be applied practically for commercial applications.

Conflict of Interest

The authors have no conflicting financial or other interests.

Acknowledgement

This work was partially supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ009078), Rural Development Administration, Republic of Korea. It was also partially supported by Technology Development Program for Agriculture and Forestry, Ministry for Food, Agriculture, Forestry and Fisheries, Republic of Korea.

References

- Ahvenainen, R. 1996. New approaches in improving the shelf life of minimally processed fruit and vegetables. *Trends in Food Sci. & Technol.* 7(6):179-187.
- Asplund, K. and E. Nurmi. 1991. The growth of *Salmonellae* in tomatoes. *Internatl. J. Food Microbiol.* 13(2):177-181.
- Beuchat, L. R. 1996. Pathogenic microorganisms associated with fresh produce. *J. Food Protec.* 59(2):204-216.
- Beuchat, L. R. 2002. Ecological factors influencing survival and growth of human pathogens on raw fruits and vegetables. *Microbes & Infection* 4(4):413-423.
- Centers for Disease Control and prevention (CDC). 2007. Multistate outbreaks of *Salmonella* infections associated with raw tomatoes eaten in restaurants - United States, 2005-2006. *J. American Med. Assoc.* 298(15):1753-1755.
- Cho, B., I. Baek, N. Lee and C. Mo. 2011. Study on bruise detection of Fuji apple using hyperspectral reflectance imagery. *J. Biosystems Engr.* 36(6):484-490.
- De Roever, C. 1999. Microbiological safety evaluations and recommendations on fresh produce. *Food Control* 10(2):117-143.
- Francis, G. A., C. Thomas and D. O'Beirne. 1999. The microbiological safety of minimally processed vegetables. *Intl. J. Food Sci. & Technol.* 34(1):1-22.
- Guo, X., J. Chen, R. E. Brackett and L. R. Beuchat. 2002. Survival of *Salmonella* on tomatoes stored at high relative humidity, in soil, and on tomatoes in contact with soil. *J. Food Prot.* 65(2):274-279.
- Hedberg, C. W., F. J. Anquilo, K. E. White, C. W. Langkop, W. L. Schell, M. G. Stobierski, A. Schuchat, J. M. Besser, S.

- Dietrich, L. Hesel, P. M. Griffin, J. W. McFarland and M. T. Osterholm. 1999. Outbreaks of *salmonellosis* associated with eating uncooked tomatoes: implications for public health. *Epidemiol. & Infection*. 122(3):385-393.
- Lin, C. Mand C. I. Wei. 1999. Transfer of *Salmonella montevideo* onto the interior surfaces of tomatoes by cutting. *J. Food Prot.* 60(7):858-862.
- Liu, Y., Y. Chen, C. Wang, D. Chan and M. S. Kim. 2006. Development of hyperspectral imaging technique for the detection of chilling injury in cucumbers; Spectral and image analysis. *Applied Engr. in Agric.* 22(1):101-111.
- Painter, J. A., R. M. Hoekstra, T. Ayers, R. V. Tauxe, C. R. Braden, F. J. Angulo and P. M. Griffin. 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998-2008. *Emerging infectious diseases* 19(3):407.
- United States Department of Agriculture (USDA). 1991. Standards for grades of fresh tomatoes. pp. 1-9.
- Vargas, A. M., M. S. Kim, Y. Tao, Lefcourt, A., Chen, Y., Luo, Y., Song, Y and Buchanan, R. 2005. Detection of fecal contamination on cantaloupes using hyperspectral fluorescence imagery. *J. Food Sci.* 70(8):E471-E476.
- Wei, C. I., T. S. Huang, J. M. Kim, W. F. Lin, M. L. Tamplin and J. A. Bartz. 1995. Growth and survival of *Salmonella montevideo* on tomatoes and disinfection with chlorinated water. *Journal of Food Protection* 58(8):829-836.
- CDC (Centers for Disease Control and Prevention). 2011. Foodborne Outbreak Online Database (FOOD). Data retrieved August 1, 2013 from <http://wwwn.cdc.gov/foodborneoutbreaks/Default.aspx>.