Thai Journal of Mathematics : 339-349
Special Issue : Annual Meeting in Mathematics 2019

http://thaijmath.in.cmu.ac.th
Online ISSN 1686-0209

Automatic Defect Detection for Mango Fruit Using Non-Extensive Entropy with Gaussian Gain

Kotchakorn Tiemtud, Pornpimon Saprasert, Thanikarn Tormo and Saifon Chaturantabut[1]

Department of Mathematics and Statistics, Thammasat University
emails: kotchakorntiemtud@gmail.com (K. Tiemtud)
pornpimonsaprasert@gmail.com (P. Saprasert)
thanikarn1106@gmail.com (T. Tormo)
saifon@mathstat.sci.tu.ac.th (S. Chaturantabut)

Abstract : Quality inspection process of agricultural products recently becomes a crucial part of food industries. As the amount of these products gets larger, manual quality control based on traditional visual inspection performed by human can be tedious, time-consuming, labor-intensive, and inconsistent. Machine leaning can be used to automate the quality inspection, which will make this process more rapid, consistent, accurate and cost-efficient. This work focuses on identifying defects on mango surfaces. This work applies non-extensive entropy with Gaussian modeling, which is an unsupervised automated texture defect detection technique and therefore it does not require any training image samples in advance. The numerical results from this approach are shown to be effective in detecting various defects, such as cracks, dark spots and bruises from mango image samples. This work also investigates different window sizes used in entropy computation, which will affect the trade-off between computational speed and detection accuracy.

Keywords : unsupervised learning method; non-extensive entropy; Gaussian model; defect detection.
2010 Mathematics Subject Classification : 90C59; 65C50; 94A17.

This work was supported by Faculty of Science and Technology, Thammasat University.

[1]Corresponding author.

1 Introduction

For agricultural product, the visibility of external skin defects is one of the most important factors in the quality and price. The quality check is traditionally done mainly by humans manually. However, this manual inspection can be time-consuming and inefficient, particularly when dealing with large quantity of agricultural product. This work considers an approach that is effective in automated visual inspection for detecting skin defects of fruits, particularly for mango.

Several studies have been conducted in order to detect defects and classify fruits and vegetables [1], such as citrus [2, 3, 4], strawberry [5], kiwi [6] apples [7, 8, 9], tomato [10] and potatoes [11]. Many existing approaches for detecting defects for fruits are based on supervised learning that requires certain set of fruit images for training. This approaches include the support vector machine [10], principal component analysis [9], and neuron network [12]. This work considers unsupervised learning method based on non-extensive entropy with Gaussian gain proposed in [13].

For images with some texture pattern, one of the common measures for uncertainty or ambiguity is entropy. The lower the entropy the higher the receptivity of the pattern and the more regular is the texture. Entropy has been used for texture identification and texture defect detection in previous works, e.g. [14, 15, 16]. The non-extensive entropy with Gaussian gain is recently proposed to use as an entropy for a single feature for texture identification and categorization [13]. It is suitable for textures containing non-additive information content, i.e., the textures containing a set of patterns which repeat over space in a regular manner. For textures containing some degree of regularity, the non-extensive entropy has been shown in [13] to outperform other forms of entropy, such as Shannon, Renyi, Tsallis and Pal entropies. This non-extensive entropy has been applied to applications such as detection of motion disturbance [17] and detection of defect in texture samples [18]. It also has been extended to fuzzy entropy in [19]. However, to our knowledge, it has never been used for detecting defects of any agricultural product. In this work, we therefore investigate the efficiency of this non-extensive entropy on identifying defects on mango surface. This result, of course, can be extended to other kinds of fruits with similar configuration.

To identify defects in our study, after obtaining images of mango skin surface, we construct the gray-level co-occurrence matrix (GLCM) [15] or gray-level spatial dependence matrix for each of many small sub-images, which are basically widows sliding through a given surface image. Note that GLCM is used to represent the textural relationship of pixels. Each probability in the GLCM is then used to compute entropy of each sliding window. The entropy values from all windows can be used to form a Gaussian distribution. The defects can be finally found by considering them as outliers of the distribution using certain statistical criterion.

This paper is organized as follows. Background materials on GLCM and non-extensive entropy are given in Section 2. Section 3 considers three numerical test cases. The first test applies the non-extensive entropy approach on syntactic images that has different shapes of defects. The second test considers the actual data set

of mango images with and without defects. The last test case considers images of mango surfaces with different colors on them. In the first two numerical tests, we investigate the size of window used in entropy calculation. The last numerical test is performed to confirm that this approach does not give false alarm of defect-free mango surface with varying natural color. Summary of this work and some remarks are finally provided in Section 4.

2 Automatic texture defect detection by entropy modeling

This section will provide details used for specifying defect on an image of a mango surface by using non-extensive entropy with Gaussian gain. First, we will construct the gray-level co-occurrence matrix (GLCM) [15], as described in Section 2.1, which is used to examine the texture by considering the spatial relationship of pixels. In particular, it identifies how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. After creating a GLCM, we can extract statistical measures from this matrix. Each probability in the GLCM will be used to compute the non-extensive entropy with Gaussian gain [13], given in Section 2.2, for each of many sub-regions in the image. Section 2.3 describes how all entropy values of these sub-regions can be used to form a Gaussian distribution and identify defects. In this work, defects can be detected from outliers of the distribution by using the 2 and 3 standard deviation criteria [20].

2.1 Gray-level co-occurrence matrix (GLCM)

We will consider a picture of surface segment as a rectangular image I of size N_x by N_y pixels, where N_x and N_y are the number of pixels in horizontal and vertical directions, respectively. Suppose that the gray level appearing at each pixel is quantized to N_g levels. Let $L_y = \{1, 2, \ldots, N_y\}$ and $L_x = \{1, 2, \ldots, N_x\}$ be the sets of row and column indices of the image. Let $G = \{0, 1, ..., N_g - 1\}$ be the set of N_g quantized gray levels. Then $L_y \times L_x$ is the set of ordered pair (row and column) indices of pixels. The image I can be considered as a function that maps from each pixel in $L_y \times L_x$ to some gray level in G, i.e. $I : L_y \times L_x \to G$. The texture-context information can be determined by the matrix $\mathbf{P} = [P_{ij}]$ where P_{ij} is the *relative frequency* of two neighboring resolution cells separated by distance d in direction θ occurring on the image, one with gray tone $i \in G$ and the other is gray tone $j \in G$. Note that $\mathbf{P}$ is a symmetric matrix of size N_g-by-N_g and it depends on both distance d and the angular direction θ. We therefore also denote P_{ij} as $P(i, j, \theta, d)$. In this work, we consider four directions: $\theta = 0°, 45°, 90°, 135°$ and fixed the value of d. The unnormalized frequencies, denoted by $P(i, j, \theta, d)$, where $\theta = 0°, 45°, 90°, 135°$, are defined as shown in the equations below where # denotes the number of elements in the set.

$$\begin{aligned}
P(i,j,0^\circ,d) &= \#\{(\ (k,l),(m,n)\ ) \in (L_y \times L_x) \times (L_y \times L_x) \,|k-m=0, \\
&\qquad |l-n| = d, I(k,l) = i, I(m,n) = j\} \\
P(i,j,45^\circ,d) &= \#\{(\ (k,l),(m,n)\ ) \in (L_y \times L_x) \times (L_y \times L_x) \,|(k-m=d, l-n=-d) \\
&\qquad or(k-m=-d, l-n=d), I(k,l) = i, I(m,n) = j\} \\
P(i,j,90^\circ,d) &= \#\{(\ (k,l),(m,n)\ ) \in (L_y \times L_x) \times (L_y \times L_x) \,||k-m| = d, \\
&\qquad l-n=0, I(k,l) = i, I(m,n) = j\} \\
P(i,j,135^\circ,d) &= \#\{(\ (k,l),(m,n)\ ) \in (L_y \times L_x) \times (L_y \times L_x) \,|(k-m=d, l-n=d) \\
&\qquad or(k-m=-d, l-n=-d), I(k,l) = i, I(m,n) = j\}
\end{aligned}$$

The corresponding GLCM, denoted by $\bar{\mathbf{P}}_\theta$, is a matrix with (i,j)-th entry $\bar{P}(i,j,\theta,d)$ is a normalized value of $P(i,j,\theta,d)$ so that $\sum_{i\in G}\sum_{j\in G}\bar{P}(i,j,\theta,d) = 1$ for a given distance d and direction θ.

2.2 Non-extensive entropy with Gaussian gain

The non-extensive entropy with Gaussian gain proposed in [13] is defined as follow. Let $X = \{x_1, x_2, ..., x_n\}$ be a variable with the corresponding probabilities $P = \{p_1, p_2, ..., p_n\}$, where $p_i \in [0,1]$, $\sum_{i=1}^n p_i = 1$, and n is the number of probabilistic experiments.

The entropy of X is defined as

$$H(P) = \sum_{i=1}^{n} p_i e^{-p_i^2}. \tag{2.1}$$

The non-extensive entropy with Gaussian gain in (2.1) can be interpreted as the weighted average of single-sided zero mean Gaussian functions with standard deviation of $1/\sqrt{2}$. More details on this entropy can be found in [13].

2.3 Automatic Defect Detection

Suppose a texture image has size $L \times M$. We will calculate the entropy for each window of size $N \times N$, which will be moving 1 pixel each time through the image starting from top-left to bottom-right. Suppose there are total of N_w windows. Let $w \in \{1, 2, ..., N_w\}$. For each window w, the gray-level co-occurrence matrices (GLCM) are compute along four directions $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, and use an offset distance $d = \text{round}(N/4 - 1)$ (for case of asymmetric texture windows N is the lesser dimension) where the function round$(\cdot)$ rounds off the fraction to the next highest integer. Let $\bar{\mathbf{P}}^w_{0^\circ}$, $\bar{\mathbf{P}}^w_{45^\circ}$, $\bar{\mathbf{P}}^w_{90^\circ}$, and $\bar{\mathbf{P}}^w_{135^\circ}$ be these GLCMs corresponding to directions $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. Each of these four matrices are used as the probabilities to compute the entropy in (2.1), denoted by $E^w_{0^\circ}$, $E^w_{45^\circ}$, $E^w_{90^\circ}$, and $E^w_{135^\circ}$, where

$$E^w_\theta = H(\bar{\mathbf{P}}^w_\theta) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} h(\bar{P}(i,j,\theta,d)) \tag{2.2}$$

and the function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(p) = pe^{-p^2}$, which directly follows from (2.1). Then the entropy for each window is then the average of these four entropies:

$$E_w = \frac{E^w_{0^\circ} + E^w_{45^\circ} + E^w_{90^\circ} + E^w_{135^\circ}}{4}, \tag{2.3}$$

where $w = 1, 2, ..., N_w$.

Let $\{E_w\} = \{E_1, E_2, \ldots, E_{N_w}\}$ be the set of entropies computed from N_w windows, where each window has size $N \times N$. We can separate the windows with defect from other widows without defect by using statistical criterion for outliers. In this work, since the data distribution of $\{E_w\}$ seems to be symmetric and mound-shaped, which perfectly fits the normal distribution, then it is reasonable to use the Standard Deviation (SD) criteria as an outlier-labeling methods [20].

The SD method defines outliers based on either 2SD criterion or 3SD criterion. In particular, the outliers are defined as follows.

$$\underline{\textbf{2SD criterion:}} \text{ Outliers are outiside the interval}(\bar{E} - 2\sigma, \bar{E} + 2\sigma) \tag{2.4}$$

$$\underline{\textbf{3SD criterion:}} \text{ Outliers are outiside the interval}(\bar{E} - 3\sigma, \bar{E} + 3\sigma) \tag{2.5}$$

Note that $\bar{E}$ is the sample mean and σ is the sample standard deviation of $\{E_w\}$. Note also that approximately 95% and 99.7% of the data are within 2 and 3 standard deviations of the mean, respectively, for any data set that follows a normal distribution, thus, the observations beyond 2SD or 3SD above and below the mean of the observations may be considered as outliers in the data [20].

The following steps summarize the procedure for the automated defect detection discussed in this section.

Algorithm 1 : Automated defect detection algorithm

INPUTS:

1) Images of size $N_x \times N_y$ pixels;
2) N for window size $N \times N$;
3) L, M for the scaled gray-level image $L \times M$;
4) N_g (Note: This work uses $N_g = 256$, since gray-scale images in digital photography are generally composed of 256 shades of gray, ranging from black to white.)

OUTPUT: Defective windows of the input image (if any).

1: Transform an input image to the gray-level texture image of size $L \times M$
2: For a fixed window w of size $N \times N$, perform the following steps.
   (i) Construct the four $N_g \times N_g$ gray-level co-occurrence probability matrices $\bar{\mathbf{P}}_\theta$ corresponding to the angles $\theta = 0°, 45°, 90°, 135°$ at an offset distance of $d = \text{round}(N/4 - 1)$
   (ii) Compute entropy E_θ^w for each $\theta = 0°, 45°, 90°, 135°$ using (2.2).
   (iii) Find the average entropy E_w for each window w by using (2.3)
3: Slide the image window across the entire image starting from the top-left corner and ending at the bottom-right corner, and go back to (i)-(iii) in step 2. Let the set of entropies computed across all windows be $\{E_w\}$.
4: Fit $\{E_w\}$ with the Gaussian distribution.
5: The defective windows would be outliers that satisfy the 2SD criterion in (2.4) or 3SD criterion in (2.5).

From Algorithm 1, the main steps for identifying whether a given mango image has some defect are computing entropies for all windows (steps 2-3) and fitting these entropies to the normal distribution (step 4). If there is any outliner based on the criterion in (2.4) or (2.5), we conclude that the mango in the input image has some defect on its surface.

The effectiveness of the above procedure is illustrated in the next section by applying it on some synthetic and actual image data sets.

3 Numerical Results

All procedures explained in the previous section are implemented in MATLAB program to generate the numerical results here. This section considers 3 numerical test cases. The first one applies the non-extensive entropy approach on 33 syntactic images that has different shapes of defects. The second test considers the actual data set of 40 mango images with and without defects. The last test case considers 30 images of mango surfaces with different shades of colors on them. Some sample images of these 3 cases are shown in Figure 1. The accuracy of detecting the defects is measured by the following ratio as done in [18]:

$$R = \frac{\text{Ncorr}}{\text{Nfalse}}, \tag{3.1}$$

where Ncorr is the number of images correctly classified as defective and Nfalse is the number of images falsely classified as defective. Large value of R implies accurate defect detection. In the first two numerical tests, we investigate the size of window used in entropy calculation, as well as the criteria used for specifying outliers. The last numerical test is performed to confirm that this approach does not give false alarm of defect-free mango surface with varying natural color.

In the following numerical tests, the defective windows are high- lighted by white and the non-defective windows by black in the result image which is of the same size as that of the original image.

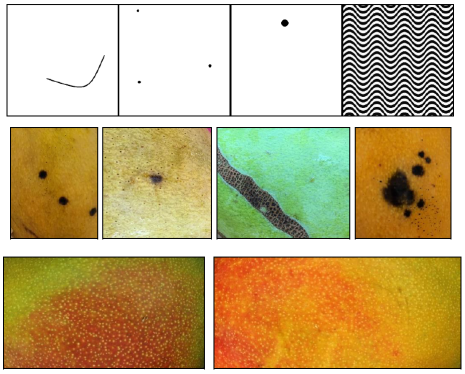

Figure 1: Examples of tested images: synthetic images and real mango images

3.1 Numerical Test 1

In the first numerical test, different types of defects are synthesized on both plain and textured backgrounds. The defects are specified based on both 2SD and 3SD criteria

given in (2.4) and (2.5), respectively. The window size, N, is varied from 3 to 9 pixels. The accuracy of each case is shown through the ratio R define in (3.1).

Two examples of defect detection results are shown in Figures 2 and 3 when non-extensive entropy with Gaussian gain method is used with 2SD and 3SD criterion; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels. Notice from Figure 3 that using the 2SD criteria gives more accurate result, since the 3SD criterion cannot detect any defect when window sizes $N = 5, 7, 9$ are used. Similarly, Table 1 provides the ratios R defined in (3.1), which reflect that the 2SD criterion can classify overall images with defects more accurately than the 3SD criterion .

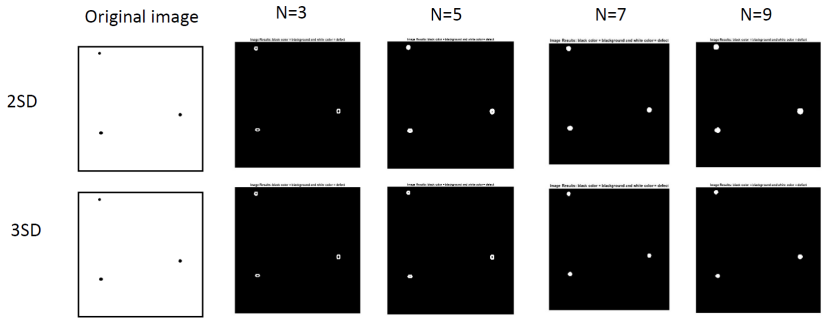

Figure 2: Defect detection output for a synthetic image (numerical test 1) using non-extensive entropy with Gaussian gain with the 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

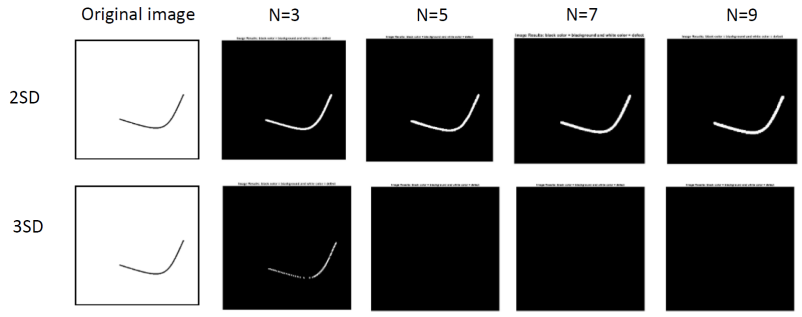

Figure 3: Defect detection output for a synthetic image (numerical test 1) using non-extensive entropy with Gaussian gain with the 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

R = Ncorr/Nfalse (Outlier Criterion)	$N = 3$	$N = 5$	$N = 7$	$N = 9$
R (2SD)	4.5	4.5	4.5	5.8
R (3SD)	3.7143	3.125	3.7143	3.7143

Table 1: The ratios R = Ncorr/Nfalse from testing 33 synthetic images (numerical test 1) using non-extensive entropy with Gaussian gain with the 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

3.2 Numerical Test 2

This numerical test considers the real mango surface images. Two example of classification outputs are shown in Figures 4 and 5 when non-extensive entropy with Gaussian gain method is used with 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels. Notice that in Figure 4, 2SD criterion can clearly identify the defect spot better than 3SD criterion for all widow sizes. When the defect scatters on the surface image as shown in Figure 5, it becomes harder to identify the defect. In this case, 2SD criterion cannot detect the defect for all window sizes, and 3SD criterion can detect the defect clearly only when the window size is 3. The performance on the overall images in the data set of real mango surface are demonstrated through Table 2. Notice that, similar to the previous results, the 2SD criterion gives larger ratios R (more accurate detection) than the 3SD criterion for all window sizes.

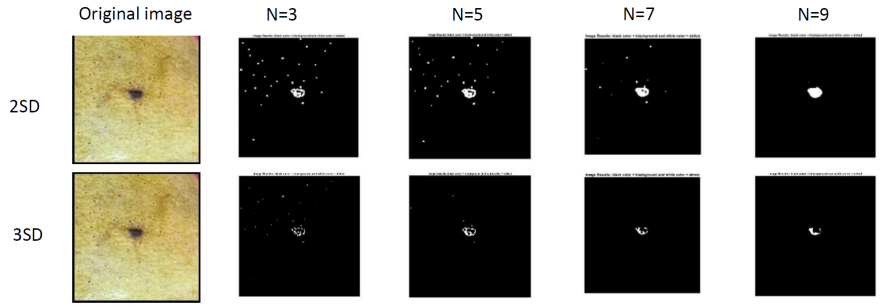

Figure 4: Defect detection output for a mango surface image (numerical test 2) using non-extensive entropy with Gaussian gain with 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

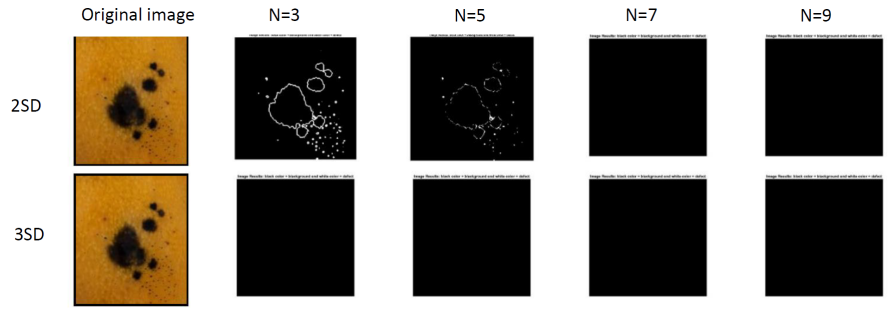

Figure 5: Defect detection output for a mango surface image (numerical test 2) using non-extensive entropy with Gaussian gain with 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

3.3 Numerical Test 3

This last numerical test is performed to confirm that this approach does not give false alarm of defect-free mango surfaces with varying natural color. This test considers 30 images of mangoes' surfaces with different varying colors on them that has no defect. Figure 6 illustrates two of the test images, as well as the prediction outputs, which correctly

R = Ncorr/Nfalse (Outlier Criterion)	$N = 3$	$N = 5$	$N = 7$	$N = 9$
R (2SD)	2.6364	2.6364	2.077	3
R (3SD)	0.4815	0.4286	0.4815	0.4815

Table 2: The ratios R = Ncorr/Nfalse from testing 40 mango images (numerical test 2) using non-extensive entropy with Gaussian gain with 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

classify that the surface images have no defect on it. This result shows that the method we used in this work is stable and not sensitive to varying natural color of mangoes's surface.

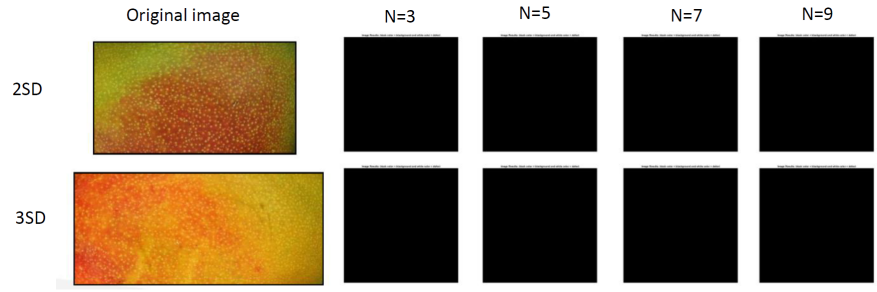

Figure 6: Defect detection output for a mango surface image (numerical test 3) using non-extensive entropy with Gaussian gain with 2SD and 3SD criteria; window size $N \times N$, where $N = 3, 5, 7, 9$ pixels.

4 Conclusions

In the present work, we introduce an efficient automated defect identification for mango skin, which is based on using gray-level co-occurrence matrix with the non-extensive entropy with Gaussian gain. The defects are considered as outliers from the distribution of entropies, which are identified by using 2 and 3 standard deviation (SD) criteria. For natural texture of mango skin surface considered in this work, the 2SD criterion is shown to provide better classification results. When defects scatter over mango surface, small window size trends to give more accurate detection than large window size (see Numerical Test 2). The numerical test also demonstrates that this approach is not sensitive to varying shades of colors on mango skin. The results presented in this work can be extended to other kinds of fruits with certain visible textural outer skin.

Acknowledgement: The authors would like to thank the reviewers for the useful comments to improve this paper. The authors gratefully acknowledge the financial support provided by Faculty of Science and Technology (Thammasat University), Contract No. 10/2563.

References

[1] S. Chopde, M. Patil, A. Shaikh, B. Chavhan, M. Deshmukh, Developments in computer vision system, focusing on its applications in quality inspection of fruits and vegetables-a review, Agricultural Reviews 38 (2) (2017).

[2] J. Blasco, N. Aleixos, J. Gómez,E. Moltó, Citrus sorting by identification of the most common defects using multispectral computer vision, Journal of Food Engineering 83 (3) (2007) 384–393.

[3] E. Cerruto, S. Failla, G. Schillaci, Identification of blemishes on oranges, In International conference on agricultural engineering, AgEng 96 (1996).

[4] R. Thendral, A. Suhasini, Automated skin defect identification system for orange fruit grading based on genetic algorithm, Current Sci. 112 (8) (2017) 1704–1711.

[5] G. ElMasry, N. Wang, A. ElSayed, M. Ngadi, Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry, Journal of Food Engineering 81 (1) (2007) 98–107.

[6] Q. Lü, M. Tang, Detection of hidden bruise on kiwi fruit using hyperspectral imaging and parallelepiped classification, Procedia Environmental Sciences 12 (2012) 1172–1179.

[7] G. ElMasry, N. Wang, C. Vigneault, J. Qiao, A. ElSayed, Early detection of apple bruises on different background colors using hyperspectral imaging, LWT-Food Science and Technology 41 (2) (2008) 337–345.

[8] D. Unay, B. Gosselin, Stem and calyx recognition on 'Jonagold'apples by pattern recognition, Journal of Food Engineering, 78 (2) (2007) 597–605.

[9] B. Zhu, L. Jiang, Y. Luo, Y. Tao, Gabor feature-based apple quality inspection using kernel principal component analysis, Journal of Food Engineering 81 (4) (2007) 741–749.

[10] N.A. Semary, A. Tharwat, E. Elhariri, A.E. Hassanien, Fruit-based tomato grading system using features fusion and support vector machine, In Intelligent Systems' 2014, Springer (2015) 401–410.

[11] D.C. McRae, A review of developments in potato handling and grading, Journal of Agricultural Engineering Research 31 (2) (1985) 115–138.

[12] M.A. Shahin, E.W. Tollner, R.D. Gitaitis, D.R. Sumner, B.W. Maw, Classification of sweet onions based on internal defects using image processing and neural network techniques, Transactions of the ASAE 45 (5) (2002) 1613.

[13] S. Susan, M. Hanmandlu, A non-extensive entropy feature and its application to texture classification, Neurocomputing, 120 (2013) 214–225.

[14] M.C. Padma, P.A. Vijaya, Entropy based texture features useful for automatic script identi

cation, International Journal on Computer Science and Engineering 2 (02) (2010) 115–120.

[15] R.M. Haralick, K Shanmugam, I. Dinstein, Textural Features for Image Classification, 1973.

[16] D-M Tsai, B-T Lin, Defect detection of gold-plated surfaces on PCBs using entropy measures, The International Journal of Advanced Manufacturing Technology 20 (6) (2002) 420–428.

[17] S. Susan, M. Hanmandlu, Unsupervised detection of nonlinearity in motion using weighted average of non-extensive entropies, Signal, Image and Video Processing 9 (3) (2015) 511–525.

[18] S. Susan, M. Sharm, Automatic texture defect detection using Gaussian mixture entropy modeling, Neurocomputing, 239 (2017) 232–237.

[19] S. Susan, M. Hanmandlu, A novel Fuzzy Entropy based on the Non-Extensive entropy and its application for feature selection, In 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2013) 1–8.

[20] S. Seo, A review and comparison of methods for detecting outliers in univariate data sets, PhD thesis, University of Pittsburgh, 2006.

(Received 17 June 2019)
(Accepted 24 December 2019)