Non-destructive Grading of Pattavia Pineapple using Texture Analysis

Kwankamon Dittakan
College of Computing
Prince of Songkla University
Phuket, Thailand
kwankamon.d@phuket.psu.ac.th

Nawanol Theera-Ampornpunt
College of Computing
Prince of Songkla University
Phuket, Thailand
nawanol.t@phuket.psu.ac.th

Pattaporn Boodliam
College of Computing
Prince of Songkla University
Phuket, Thailand
s5730211039@email.psu.ac.th

***Abstract*—Pineapple is a tropical fruit with may cultivars in Thailand such as Pattavia, Intarachit, Phuket, Nang-Lea and Phu-Lea. Pattavia is a famous pineapple grown in Thailand because of it has a good taste. The fresh Pattavia pineapple can be sub-divided into two grades: (i) $Keaw$ 1 and (ii) $Keaw$ 2 depending on their juiciness. In this paper an automated mechanism to classify the grade of Pattavia pineapple by non-destructive process is proposed. The fundamental idea is to construct the classification model to which is be able to distinguish between $Keaw$ 1 or $Keaw$ 2 images. Local Binary Pattern (LBP) texture analysis is suggested to capture key information of pineapple texture image. The best model produced the excellent results with recorded AUC value of 0.979.**

***Index Terms*—Pineapple Mining, Pineapple Grading, Image Mining, Image Analysis, Agriculture Informatics**

I. Introduction

Agriculture is one of the largest economic sectors in the world. Fruits and vegetables provide a lots of vitamins and minerals which are good source for human health. In term of quality inspection in fruits and vegetables, there are various methods are used to consider for examples: ripeness, firmness, texture and size. The traditional fruit inspection can be done by the human expert, which is a time-consuming process. However quality testing, grading and classification of the fruit are necessary processes for evaluation in agriculture production in order to meet quality standards and increasing marketing value [1].

Pineapple ($Anamas\ Comosus$) is one of the most popular tropical fruits, and the most economically significant plant in the $Bromeliaceae$ family. The origin of pineapple comes from the South America. There are many countries are the main producers of pineapple include: Thailand, Brazil and Philippines which supplying nearly 50 % of the total output in the world. In addition, others important producers include Indonesia, China, India, Costa Rica, Nigeria and Kenya. Pineapples can be eaten or served in fresh, cooked, juiced and can be preserved. Pineapple is a source of vitamin A and B and fairly rich in vitamin C and minerals include calcium, magnesium, potassium and iron. It also contains digesting enzyme, bromelain and also citric acid. Pineapple can be used as supplementary nutritional fruit for good health. The benefits over for health condition include [2]: (i) powerful antioxidant, (ii) immune system support, (iii) cancer prevention,(iv) anti-inflammatory effects, (v) powerful minerals for bone health condition, (vi) blood clotting reduction and (vii) keep teeth healthy.

Pattavia is very popular cultivars to plant and for the canning industry in Thailand. Pattavia is in a group of $Smooth\ Cayenne$ with excellent flavour. The size is about 1.8 to 4.5 kilograms. The shape is in cylindrical form, shallow eyes, orange rind and yellow flesh. Pattavia can be classified into two grades: (i) $Keaw$ 1 and (ii) $Keaw$ 2. $Keaw$ 1 is very juicy and sweet with less acid flavour while $Keaw$ 2 is less juicy and sweet with mildly acid flavour. The price of $Keaw$ 1 is more expensive compares to $Keaw$ 2 in the local market in Thailand. In general people prefers to buy $Keaw$ 1 more than $Keaw$ 2. However because $Keaw$ 1 and $Keaw$ 2 look very alike, so it is very difficult to general public of distinguish between them. In the market, fruit seller could uses their experiences to identify to their customer.

In this paper a mechanism of classifying the grade of Pattavia pineapple is proposed. More specifically a process for the detection the grade of Pattavia pineapple using the texture analysis to pineapple skin imagery is proposed.

The rest of this paper is organised as follows. Section II presents some previous works with respect to fruit quality testing, grading and classification. The proposed framework for grading the Pattavia pineapple is presented in Section III. The detail of data set used is given in Section IV. Section V described texture analysis adopted in this work. The nature of the feature selection and classifier generation mechanisms used are presented in Section VI. The evaluation of the proposed method is reported in Section VII. Finally, a conclusions and suggestions for future work are discussed in Section VIII

II. Prevous Work

Digital image processing is a mechanism to perform some operations to digital image. The main objectives are: to enhance an image quality or to extract the salient information from the image. In agriculture domain digital image processing has been use for automated quality testing, grading and classification. Many methods for automated quality testing, grading and classification of fruits using their images have been reported in the literature.

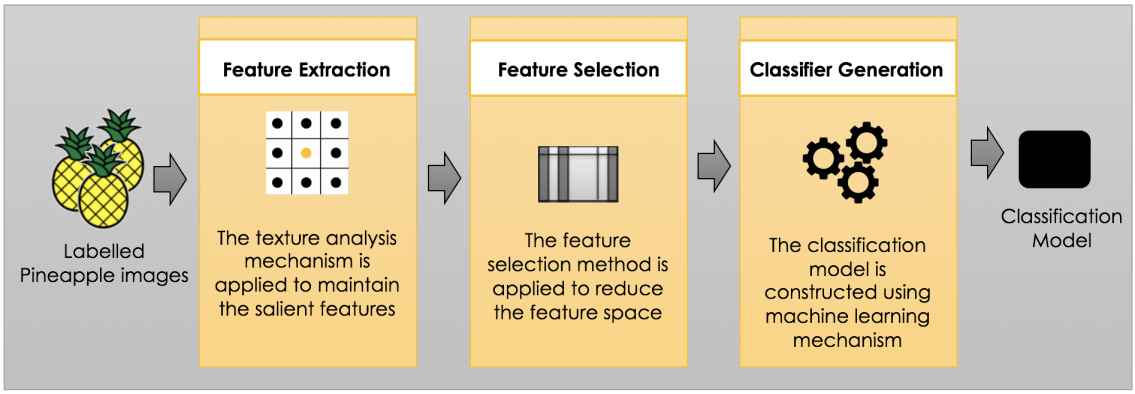

Fig. 1. Proposed non-destructive grading of Pattavia pineapple using texture analysis framework.

The texture, colour and wavelet feature of RGB images were used to test the quality of apple in [3] while in [4] the detected surface defects on apple in near infrared was applied. Blueberry have been classified, in [5] the blueberry with fungal damage was removed using non-expensive computer vision method. But in [6] the early stage of leaf rust was detected using hyper-spectral remote sensing. In [7] a machine vision and neural network system was used to classify three varieties of peach cultivars. Mango is also done the quality control using fuzzy image analysis in [8], in [9] CCD camera was used to capture the mango images and size was calculated for mango size grading system. Colour feature have been used to assess the state of maturation of the pineapple [10].

The common image property can be categorised as being founded on: (i) colour, (ii) shape and (iii) texture. Colour features are typically used in the context of colour image representation methods, colour is the property where the object reflects or emits light producing different sensations on the eye. The methods used to extract colour feature for example colour histogram and colour correlogram.

Shape features are form, outline, external boundary or external surface of an object. The methods used to extract shape properties from the image such as Fourier transform of boundary, Chain code and Boundary approximations.

Texture features provide the spatial arrangement information of colours and/or intensities in digital image. Texture analysis has been wildly used in many applications such as medical image analysis, remote sensing image analysis, industrial inspection and image retrieval. There are four main problems in texture analysis [11]: (i) feature extraction: to describe an image using its texture properties, (ii) image segmentation: to partition an image into sub-regions homogeneous texture, (iii) classification: to categorise the images to pre-defined groups based on its texture, (iv) shape analysis: to establish an image shape from texture information. In order to analyse an image texture, there is a number texture analysis methods such as such as Grey Level Co-occurence Metrics (GLCMs), texture spectrums, wavelet transforms, autocorrelation and power spectrum and lows texture energy measures [12].

With respect to the work reported in this paper, texture analysis mechanism was used to extract the feature from the Pattavia pineapple skin images. More specifically the Local Binary Pattern (LBP) [13] was adopted. LBP offer advantages on change invariant and ease to compute. LBP is discussed in further detail in Section V

III. Proposed Framework

The framework of non-destructive grading of Pattavia pineapple using texture analysis is presented in this section. The schematic of the proposed framework is illustrated in Figure 1. The framework consists of three processes: (i) feature extraction, (ii) feature selection and (iii) classifier generation.

With respect to the work in this research, a collection of known grades ($Keaw$ 1 and (ii) $Keaw$ 2) pineapple images (refers to known grade of pineapple labelled by domain experts) is used as training data set. The suitable feature extraction process is then applied to the training set which is the a collection of image translation is conducted, typically Local Binary Pattern (LBP) is adopted for this work. When the input data is translated into appropriate format representation so that feature vectors can be generated. Once the feature vectors have been generated, the feature selection mechanism is applied for feature space reduction but still maintain a good discrimination of data. The classifier generation is then proceeded so that the desired classifier is generated. Once the classifier model construction process is completed the produced classifier could be applied to predict the grade of unseen pineapple skin image in the future.

IV. Data Set.

The information of a collection of data set used in this research is presented in this section. A collection of 83 images was obtained from the skin of Pattavia pineapple brought in the fruit market in Suratthani Province in the south of Thailand during March to August in 2017. The images were separated into two groups: (i) 40 images of $Keaw$ 1 and (ii) 43 images of $Keaw$ 2 by domain experts. $Keaw$ 1 is the top grade of the Pattavia pineapple with very sweet and juicy, in another word $Keaw$ 2 is the lower grade from $Keaw$ 1 with less sweet and

juicy. Figure 2 shows the example of pineapple images: Figure 2(a) presents skin and internal of Pattavia pineapple $Keaw$ 1 pineapple images, and Figure 2(b) presents skin and internal of Pattavia pineapple $Keaw$ 2 pineapple images.

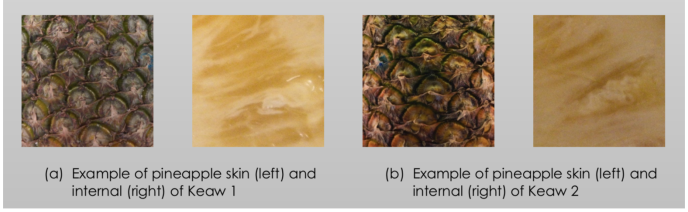

Fig. 2. Example of Pineapple grade $Keaw$ 1 and grade $Keaw$ 2 images

From the image, it can be seen that the images from Fig. Ref fig: Kaew (a) and Figure ref fig: Kaew (b) are difficult to distinguish between them.

V. Local Binary Pattern Based Representation.

Local Binary Pattern (LBP) is a texture descriptor which has been applied in many image processing and computer vision applications. LBP was introduced by Ojala et al. [13] for texture analysis. LBP offers the main advantages on simple and powerful descriptor. LBP has been applied in many applications, the example include: medical image analysis [14] [15], Face image analysis [16] [17], and remote sensing image analysis [18] [19]. The fundamental idea is to define LBP pattern or LBP code which encode the local texture content for every isolate pixels in an image surrounding with its eight neighbours. To generate a set of LBPs from pineapple texture image, the image is first converted into the greyscale colour space. A 3×3 pixel window is then used and moved over the image, the centre pixel is compared with its eight neighbours. The process of LBP is illustrated in Figure 3. If the neighbouring intensity value is greater than the centre pixel intensity value (or positive different value) a 1 is encoded, otherwise a 0 is encoded. An eight digit binary number is defined by concatenating all these binary codes according to its 3×3 pixel neighbourhood in a clockwise direction. Thus its corresponding decimal value is calculated from eight digit binary number. The resulting number is referred to Local Binary Patterns or LBP codes for labelling the centre pixel. The potential LBP code with respect to an 8-digits of 256 (2^8) different patterns can be defined.

An example image of LBP process is presented in Figure Example. Figure Example shows the original image of pineapple skin, the colour image was transformed into greyscale colour as presented in Figure Example(b). The processed image by LBP is illustrated in Figure 4(C).

Variations of LBP can be defined using difference number of sampling points (p) and the radius from the centre point (r). The notation $LBP_{p,r}$ is used to describe the variation of LBP method. With respect to the work in this research three different variations of LBP were considered: (i) $\text{LBP}_{8,1}$, 8 sampling points with a radius of 1 , (ii) $\text{LBP}_{8,2}$, 8 sampling points with a radius of 2, and (iii) $\text{LBP}_{8,3}$, 8 sampling points with a radius of 3. The examples of LBP variations is presented in Figure 5.

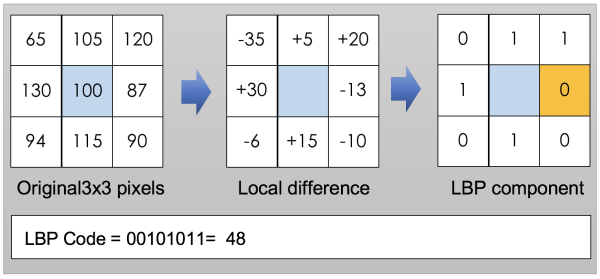

Fig. 3. Local Binary Pattern Process

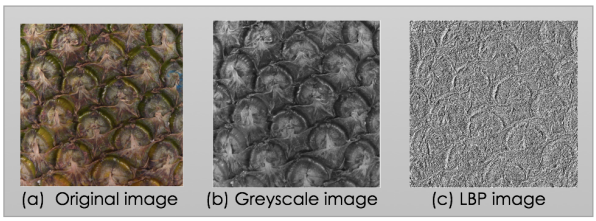

Fig. 4. Example of LBP process applied to pineapple image

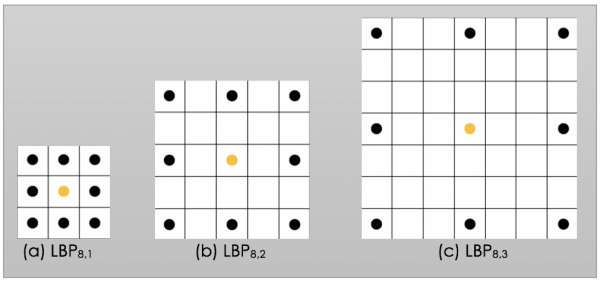

Fig. 5. LBP Variations

VI. Feature Selection and Classifier Generation

The fundamental idea of feature selection is presented in this section. When a set of LBPs code have been obtained (as described in previous section) the next step in the framework (see Figure 1) is feature selection. Prior to feature selection data discretisation was applied in order to reduce the number of values so that the numeric attributes were converted into a set of ranged attributes. With respect to the work presented in this paper 10 was the number of ranges to be considered according to this work.

The 256 features ware used to describe each pineapple image. However before classifier generation could be commenced, so as to reduce the number of dimensions in the feature vector feature selection was then applied. Feature selection mechanism offers advantages on shorter training

times, enhanced classifier generalisation process by reducing overfitting problem, and remove redundant and irrelevant attributes but still maintain a good discrimination between data classes. Three feature selection strategies were considered: (i) Gain ratio, (ii) Information gain and (iii) Relief-F. Note that, the common for feature selection techniques is to identify a parameter K which is the maximum number of "best" features to be selected.

Once the feature selection was completed and the reduced feature vector was obtained, the classifier generation (the right box in Figure 1) is applied. With respect to the evaluation presented later in this paper eight different learning methods were considered: (i) Decision Tree, (ii) Binary Decision Tree, (iii) Random Forest, (iv) Nave Bayes, (v) Bayesian Network, (vi) Logistic Regression, (vii) Sequential Minimal Optimisation (SMO) and (viii) Neural Network.

VII. Evaluation

The evaluation of the proposed approach for detection of pineapple grade is presented in this section. Extensive evaluation was conducted with respect to proposed approach. However only the most significant results obtained is reported in this section (there is insufficient space to allow for the presentation of all results obtained). The evaluation was conducted by considering a specific as study directed at digital external skin image taken from 83 pineapples (detail was mentioned in Section IV. The labeled pineapple data was separated into two classes: (i) $Keaw$ 1 and (ii) $Keaw$ 2. As noticed in Section VI, before classification could be commenced the input was first discretised thus the numberis feature was converted into 10 ranges feature. For classifier generation purposes a number of classification learning methods ware used as implemented in the Waikato Environment for Knowledge Analysis (WEKA) machine learning workbench ([20]). For the evaluation purposes Ten-fold Cross Validation (TCV) was applied throughout and the performance of the generated classifiers recorded in term of : (i) accuracy (AC), (ii) area under the ROC curve (AUC), (iii) sensitivity (SN), (iv) specificity (SP) and (v) precision (PR). The overall aim of the evaluation was to provide evidence directed to weather the grade of pattavia pineapple can be detected using the proposed approach or not. To end this four set of experiments were conducted with the following objectives:

1) To identify the most appropriate LBP representation with respect to three variations.
2) To determined the most appropriate feature selection methods.
3) To identify the most appropriate number of top K values with respect to desired feature selection mechanism.
4) To indicate the most appropriate classification learning algorithms.

Each set of experiments is discussed in further detail in Sub-section VII-A to Sub-section VII-D, respectively.

A. LBP Variations

The variations of the LBP representation can be defined in term of the values p and r, where p is the number of sampling points and r is the neighbourhood radius; the notation $LBP_{p,r}$ was used to indicate particular variations (noted in Section V)

This sub-section reports on the evaluation conducted to compare the operation of a range of LBP variations with increasing values of r: (i) $\text{LBP}_{8,1}$, (ii) $\text{LBP}_{8,2}$ and (iii) $\text{LBP}_{8,3}$. For the experiments Information gain feature selection was used with K = 60 (information gain feature selection and K = 60 were used because the experiments report in Sub-section VII-B and Sub-section VII-C, had revealed that this were appropriate feature selection algorithms and value of K, respectively) and a Neural network learning method as these had been found to work well in context of grade detection for pineapple (see Sub-section VII-D. The outcomes are presented in Table I (best results indicated in bold font with respect to AUC values). From the table it can be observed that the best result was produced using $\text{LBP}_{8,1}$; a recorded AUC value of 0.979 and sensitivity of 0.940. Hence we conclude that in the context of the experiments the $\text{LBP}_{8,1}$ was the most appropriate variation to adopt.

TABLE I
Results using variations of LBP representations

LBP variations	AC	AUC	SN	SP	PR
$\text{LBP}_{8,1}$	**0.940**	**0.979**	**0.940**	**0.940**	**0.940**
$\text{LBP}_{8,2}$	0.855	0.902	0.855	0.853	0.856
$\text{LBP}_{8,3}$	0.819	0.871	0.819	0.820	0.820

B. Feature Selection Methods

This sub-section reports on the evaluation conducted to identify a best mechanism for dimension reduction of the feature vector. Three techniques of feature selection were considered: (i) Gain ratio, (ii) Information gain and (iii) Relief-F. For the experiments used to compare these three techniques the $\text{LBP}_{8,1}$ was used as this had been found to produce the best result was established in the previous sub-section. Again K=60 was adopted together with Neural network learning method for the same reasons as before (because they had been found to produce the best outcomes). The results from the experiments are presented in Table II. From the table it can be clearly seen that the Information gain feature selection techniques produced the best outcome with respect to all evaluation metrics considered (a best AUC value of 0.979 and sensitivity of 0.940) and it can be concluded that Information gain is the most appropriate measure in the context of the grade detection application considered in this paper.

C. Number of Features

This sub-section reports on the evaluation conducted to investigate the most appropriate value for K with respect to Information Gain feature selection technique (because it had been found to product the best result as presented in previous

TABLE II
RESULTS USING THREE FEATURE SELECTION TECHNIQUES

Feature selection	AC	AUC	SN	SP	PR
Gain Ratio	0.867	0.947	0.867	0.868	0.868
Information Gain	**0.940**	**0.979**	**0.940**	**0.940**	**0.940**
Relief-F	0.735	0.843	0.735	0.736	0.736

sub-section). For the evaluation a sequence of experiments was conducted using a range of K values from 10 to 100 incrementing in steps of 10. For the evaluation the $\text{LBP}_{8,1}$ variation together with Neural network classification were again used, because they had already been shown to produce the best in classification performance. The obtained results are presented in Table III. From the table it can be observed that K = 60 produced the best performance with respect to all the evaluation metrics considered (AUC = 0.979 and sensitivity = 0.940).

TABLE III
RESULTS USING A RANGE OF DIFFERENT VALUES FOR K WITH RESPECT TO INFORMATION GAIN FEATURE SELECTION

Top K values	AC	AUC	SN	SP	PR
$K = 10$	0.843	0.927	0.843	0.844	0.844
$K = 20$	0.867	0.943	0.867	0.868	0.868
$K = 30$	0.843	0.940	0.843	0.846	0.846
$K = 40$	0.855	0.962	0.855	0.855	0.855
$K = 50$	0.916	0.961	0.916	0.918	0.918
$K = 60$	**0.940**	**0.979**	**0.940**	**0.940**	**0.940**
$K = 70$	0.916	0.970	0.916	0.918	0.918
$K = 80$	0.880	0.966	0.880	0.879	0.880
$K = 90$	0.867	0.943	0.867	0.868	0.868
$K = 100$	0.867	0.931	0.867	0.866	0.868

D. Classification Learning Methods

This sub-section reports on the evaluation performed to determine the most appropriate classification learning methods. Eight different learning methods were considered: (i) Decision Tree, (ii) Binary Decision Tree, (iii) Random Forest, (iv) Nave Bayes, (v) Bayesian Network, (vi) Logistic Regression, (vii) Sequential Minimal Optimisation (SMO) and (viii) Neural Network. For the experiment, in each case the $\text{LBP}_{8,1}$ variation dataset was used together with the K=60 with respect to information gain feature selection technique were used. The results are presented in Table IV. From the table it can be clearly be seen that neural network learning outperformed all the other classification learning algorithms considered with AUC of 0.979 and sensitivity of 0.940. The decision tree learning method produced the worst performance.

VIII. CONCLUSIONS

A framework for Non-destructive grading of Pattavia pineapple using texture analysis has been proposed. The main findings evidenced by the reported evaluation were:

- The proposed approach produced the great results with AUC value of 0.979 and sensitivity value of 0.940.

TABLE IV
RESULTS USING TEN DIFFERENT CLASSIFICATION LEARNING METHODS

Learning methods	AC	AUC	SN	SP	PR
Decision Tree	0.602	0.586	0.602	0.599	0.602
Binary Decision Tree	0.687	0.646	0.687	0.686	0.687
Random Forest	0.651	0.774	0.651	0.649	0.650
Nave Bayes	0.723	0.826	0.723	0.732	0.742
Bayesian Network	0.723	0.823	0.723	0.732	0.742
Logistic Regression	0.867	0.876	0.867	0.868	0.868
SMO	0.880	0.834	0.880	0.877	0.880
Neural Network	**0.940**	**0.979**	**0.940**	**0.940**	**0.940**

- The most appropriate LBP variations with respect to LBP texture descriptor was$\text{LBP}_{8,1}$.
- The most appropriate feature selection mechanisms was Information Gain.
- The most appropriate K value to be used with respect to information gain feature selection was found to be K = 60 (top 60 features were selected).
- The most appropriate classification learning methods was neural network.

Therefore, in conclusion, the proposed framework produced the good results indicating the Pattavia pineapple could be classified the grade without destroy the product. However for the future work more feature extraction may be carried out. Also the larger training data sets will be conducted.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Prince of Songkla University (Contract Number ENG610416M).

REFERENCES

[1] Mandeep Kaur and Reecha Sharma. Quality detection of fruits by using ann technique. *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, 10(Volume 10):35–41, July 2015.

[2] Farid Hossain, Shaheen Akhtar, and Mustafa Anwar. Nutritional value and medicinal benefits of pineapple. *International Journal of Nutrition and Food Sciences*, 4:84–88, January 2015.

[3] D. Unay and B. Gosselin. Artificial neural network-based segmentation and apple grading by machine vision. In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–630, 2005.

[4] B. S. Bennedsen, D. L. Peterson, and Amy Tabb. Identifying defects in images of rotating apples. *Comput. Electron. Agric.*, 48(2):92–102, August 2005.

[5] Gabriel Leiva-Valenzuela, German Mondragon, Domingo Mery, and J. Aguilera. The automatic sorting using image processing improves postharvest blueberries storage quality. In *In Proceedings of International Congress on Engineering and Food (ICEF-2011)*, pages 279–280, 01 2011.

[6] V. Ahlawat, O. Jhorar, L. Kumar, and D. Backhouse. Using hyperspectral remote sensing as a tool for early detection of leaf rust in blueberries, 2013.

[7] A. Alipasandi, H. Ghaffari, and S.Z. Alibeyglu. Classification of three varieties of peach fruit using artificial neural network assisted with image processing techniques. *International Journal of Agronomy and Plant Production*, 4(9):2179–2186, 2013.

[8] Tajul Rosli Razak, Mohd Nazari A Bakar, Mohd Nazari bin Abu Bakar, Khairul Adilah bt Ahmad, and Ab Razak Mansor. Mango grading by using fuzzy image analysis. In *International Conference on Agricultural, Environment and Biological Sciences (ICAEBS'2012)*, pages 18–22, May 2012.

[9] Suchitra Khoje and Shrikant Bodhe. Comparative performance evaluation of size metrics and classifiers in computer vision based automatic mango grading. *International Journal of Computer Applications*, 61(9):1–7, January 2013.

[10] L. Angel, S. Lizcano, and J. Viola. Assessing the state of maturation of the pineapple in its perolera variety using computer vision techniques. In *20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*, page 1, 2015.

[11] Z. Guo, L. Zhang, D. Zhang, and S. Zhang. Rotation invariant texture classification using adaptive lbp with directional statistical features. In *2010 IEEE International Conference on Image Processing*, pages 285–288, Sept 2010.

[12] M. Sharma and S. Singh. Evaluation of texture methods for image analysis. In *The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001*, pages 117–121, Nov 2001.

[13] Timo Ojala, Matti Pietikinen, and David Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 582–585 vol.1, Oct 1994.

[14] Kwankamon Dittakan and Frans Coenen. Early detection of osteoarthritis using local binary patterns: A study directed at human joint imagery. In *PRICAI 2016: Trends in Artificial Intelligence - 14th Pacific Rim International Conference on Artificial Intelligence, Phuket, Thailand, August 22-26, 2016, Proceedings*, pages 93–105, 2016.

[15] Arnau Oliver and Jordi Freixenet. False positive reduction in mammographic mass detection using local binary patterns. In *Intl Conf on Medical Image Computing and Computer Assisted Intervention*, 2007.

[16] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In Tomás Pajdla and Jiří Matas, editors, *Computer Vision - ECCV 2004*, pages 469–481, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[17] Abdenour Hadid and Senior Member. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2006.

[18] S. Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer. A 3d teacher for car detection in aerial images. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007.

[19] Kwankamon Dittakan, Frans Coenen, Rob Christley, and Maya Wardeh. A comparative study of three image representations for population estimation mining using remote sensing imagery. In *Advanced Data Mining and Applications, 9th International Conference, ADMA 2013, Hangzhou, China, December 14-16, 2013, Proceedings, Part I*, pages 253–264, 2013.

[20] I. H. Witten, E. Frank, and M. A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Amsterdam, Third edition, 2011.