ESSENTIAL FEATURES AND DATA PATTERN OF NON-DESTRUCTIVE MANGO RIPENESS CLASSIFICATION

Nurnisa Helmee
Embedded, Networks and Advanced Computing Research Cluster (ENAC),
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
Email: nurnisahelmee@gmail.com

Yasmin Mohd Yacob,
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
yasmin.yacob @unimap.edu.my

Salina Mohd Asi,
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
salina@unimap.edu.my

Nik Adilah Hanin Zahri,
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
adilahhanin@unimap.edu.my

Rohani S.M. Farook,
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
rohani@unimap.edu.my

Zulkifli Husin,
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
zulhusin@unimap.edu.my

Rafikha Aliana A.Raof
School of Computer and Communication Engineering
Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
rafikha@unimap.edu.my

ABSTRACT

Data mining is part of knowledge discovery to extract useful information from raw data. It is essential to extract relevant features and data pattern to acquire knowledge and perform data pre-processing tasks to aid classification process. Researches to extract physiological, textural and coloured features via non-destructive manner from mango images have been intensive; nevertheless, extraction of relevant features and data pattern is yet to be performed. This paper focused in extraction of essential features and interval data pattern from non-destructive mango (Mangifera Indica L.) images of ripeness classification. Mango images undergone pre-processing phases to separate it from its background image. Textural features of Angular Second Moment (ASM), Inverse Different Moment (IDM), Contrast and Correlation, each in 0^0, 45^0, 90^0 and 135^0 direction from Gray Level Co-occurrence Matrix (GLCM) were extracted from grayscale image. Coloured features were extracted from three colour channel namely Red-Green-Blue (RGB), Hue-Saturation-Value (HSV) and Lab (Luminosity-a,b-colour opponent dimension). This study employed Colour Moments and Color Coherent Vector (CCV) on each mentioned colour channel as coloured feature descriptors. There were 2 first degree moment features multiplied by 9 colour channel that generated 18 Colour Moment feature descriptors. In CCV, properties of coherent and non-coherent coloured histogram were extracted. Unlike Colour Moment and CCV, there are 20 HSV-Coloured Scale Invariant Feature Transform (CSIFT) image descriptors extraction.Comparison of selected features and its interval data pattern from feature selection methods such Correlation-based Feature Subset Selection (CFS) with Best First, Principal Component with Ranker, Classifier Subset Evaluation with Genetic Search and Multi-interval Discretization (MDLP) are presented and analysis of accuracy and time were performed using Random Sub Space classifier from Weka data mining tool. The results were indeed beneficial for knowledge discovery in mango ripeness classification.

Keywords: data pattern, feature selection, non-destructive, mango ripeness, classification.

1. INTRODUCTION

Mango is one of popular fruits especially in the north state of Malaysia. Unfortunately, sorting and grading of fruits are mainly done by human inspection system, which needs lot of effort and it is also a time consuming process. Manual inspection method of mango exposed to human errors since it is inconsistent and subjective in nature. Human concentration and focus degrade due to the tedious inspection process when dealing with large amount of fruits. Besides, there are chances of damaging the fruit during the fruit sorting process. Thus, non-destructive detection of fruit ripeness is crucial for improving industry production.

A few years ago, there has been a growing interest in this technology from researchers for non-destructive analysis in fruits (Huang et al., 2012). The assessment of fruits ripening such as peach and banana have been studied based on hyperspectral imaging technique (Rajkumar et al., 2012), which revealed a potential use of hyperspectral imaging for the fruits maturity assessment. However, some drawbacks detected such as high costs and difficulties in high-speed data acquisition and processing have limited the use of this technology in a real-time assessment (Kamruzzaman et al., 2013). Apart from that, Fuzzy system is used for mango grading based on maturity and size (Nandi et al., 2014). Feed forward neural network and support vector machine classifiers are used for grading (Khoje et al., 2013). Size and color features based mango grading with fuzzy system is proposed by (Pandey et al., 2014). Area calculation dominant and color method are used in the study. Maturity assessment of mango grading using L*a*b* color space and dominant color is proposed by (Sapan and Bankim, 2014). Weight and area are used for size feature extraction and mean of a and b channel of L*a*b* color space are used as parameters of maturity feature which proposed by (Naik et al., 2015).

Various research has been conducted to extract data pattern and features. However, to date none has attempted to extract essential features and data pattern of non-destructive mango ripeness classification. This paper presents solution for mango industries as well as essential features and data pattern of non-destructive mango ripeness clasification which makes system more efficient as compared to other methods. This paper discusses on literature review, texture feature extraction, colour feature extraction, and feature selection method. Next, this paper reviews on discretization, methodology, results and disscussions and concluded with conclusion section.

2. LITERATURE REVIEW

Image processing techniques have been extensively used in many food process industries (Kishore et al., 2016). Image processing interprates the better details of an image such as shape and color (Bodhe, 2012). Images with different colors can be identified easily but images with same color require texture feature for recognition. Colour histogram, colour moments, fuzzy histogram, mean, variance and range are commonly used colour features (Hussain et al., 2013). Standard quality inspection method needs to be implemented to provide better mango ripeness classification techniques. Non-destructive method of mango ripeness classification not only helps the exporters and farmers but also enhanced the agriculture field.

2.1 TEXTURE FEATURE EXTRACTION

Texture is an important feature of an image that has been widely used in image classification, medical image analysis, automatic visual inspection and remote sensing (Rich and Buhl-brown, 2015). Texture feature extraction is a basic stage to collect features via texture analysis process. There are few techniques to extract texture features such as using statistical, structural, model-based and transform information, in which a popular technique is by using a Gray Level Co-occurrence Matrix (GLCM) (Rayat et al., 2015). GLCM consist of second order statistical information of spatial relationship of pixels of an image. The orientation 0˚, 45˚, 90˚, and 135˚ direction from GLCM were extracted from grayscale image for each feature; to name a few, Angular Second Moment (energy), correlation, contrast, and Inverse Difference Moment (homogeneity) as illustrated in Figure 1 (Mohanaiah et al., 2013).

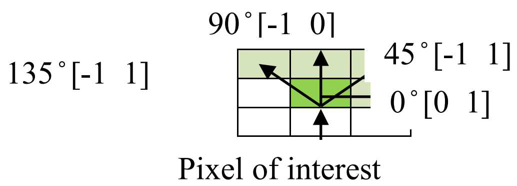

Figure 1: Orientation for the GLCM Matrix

2.2 COLOUR FEATURE EXTRACTION

Colour is the most meaningful feature for image representation of all the visual features. Colour invariant to scaling, translation and rotation of an image (Kumar, 2014). Colour provides extra information that allows dissimilarity between diverse physical causes for colour variation such as assist in distinguish between ripe mango and unripe mango. Colour features can be extracted from a histogram of the image. However, colour histogram is less effective as it merely focus on colour instead of spatial feature. Colour histogram of two distinct objects with the same colour give equal histograms. Three colour channel namely Red-Green-Blue (RGB), Hue-Saturation-Value (HSV) and Lab (Luminosity-a,b-colour opponent dimension) are commonly employed for colour features extraction. There are many colour feature extraction methods employed in food and fruits characterization. Color Coherent Vector (CCV), Colour Moments and HSV-Coloured Scale Invariant Feature Transform (CSIFT) are among them. As an example, L,a,b color space was used in foods detection employed in (Slaughter, 2009) . Table 1 shows summary of color and texture feature used for classification of images.

Table 1: Summary of color and texture feature used for classification of images

References	Dataset	Feature extraction using		Accuracy
		Color	Textural	
(Naik et al., 2015)	Mango	L*a*b	-	88.89%
(Dayan and Savakar, 2012)	Apple, chikoo, mango, orange, lemon	RGB HSI	-	87%-90%
(Jagadeesh et al., 2013)	Fruits	RGB HSV	GLCM	96.85%(average)
(Arivazhagan et al., 2013)	Plant leafs	HSI	-	94%
(Dah-Jye Lee et al., 2010)	Tomato	-	Direct colour mapping	95%

2.3 FEATURE SELECTION METHOD

A feature is also known as variable or attribute that refers to an aspect of the data (Ladha et al., 2011). Feature are specified or chosen before collecting data. Feature selection facilitating data visualization and data understanding, reducing the measurement and storage requirements, and reducing training period. Feature selection methods include Correlation-based Feature Subset Selection (CFS), correlation, and symmetrical uncertainty. Table 2 depicts a few feature selection methods.

Table 2: Feature selection methods

Contributions	Limitations	Feature selection method
Fast , scalable	Ignores feature dependencies	Information gain
Feature dependencies	Ignores classifier	Correlation-based feature selection (CFS)
Simple, feature dependencies	Risk of over fitting	Sequential forward selection (SFS)

2.4 DISCRETIZATION

Minimum Description Length (MDL) principle is a top-down method, which is based on the optimization of local measure of entropy and include stopping criteria (Fayyad and Irani, 1993).

3. METHODOLOGY

To demonstrate the applicability of data pattern and feature selection extraction in non-destructive mango ripeness classification, 56 images of mangoes 'Kent' were employed in the experiment. The images were taken from a database named as COFILAB: Computers and Optics in Food Inspection (Blasco et al., 2017). The images undergone image pre-processing phase of image segmentation and cropping so that the textural and coloured features were extracted from target region of interest.

In order to generate quality essential features in data pre-processing stage, features were extracted from textural and several coloured extraction methods. This paper extracted Gray Level Co-Occurrence Matrix (GLCM) textured features of Contrast, Correlation, Angular Second Moment (ASM) and Inverse Difference Moment (IDM) from grayscale images; each features from of 0^0,45^0,90^0 and 135^0 of direction. Thus, there are 16 textured features extracted for this study. In addition, coloured features were extracted from Colour Moment, Colour Coherent Vector (CCV) and Hue-Saturation-Value Scale Invariant Feature Transform (HSV-SIFT). Coloured features were extracted from Red-Green-Blue (RGB), Hue-Saturation-Value (HSV) and Lab colour channel incorporating Colour Moment and CCV methods. The colour moment descriptors were 2 first moments named mean and standard deviation from 9 colour channel of 3 colour model mentioned thus, generated 18 Colour Moment-features. There are 27 features from each incoherent and coherent CCV feature; total of 54 features. Unlike Colour Moment and CCV, there are 20 feature descriptors were extracted from SIFT method in HSV colour model.

As illustrated in Figure 2, after the image pre-processing and feature extraction phase, data pre-processing is performed. This experiment demonstrated training process of 56 instances or images in data pre-processing stage. 50 of them are ripe and 6 of them are unripe mango images. The feature selection and data pattern extraction were performed using Weka tool. There were three feature selection methods performed, which were Correlation-based Feature Subset Selection (CFS) with Best First, Principal Component with Ranker, and Classifier Subset Evaluation with Genetic Search. In the case of ranking feature selection method, feature value exceeded the standard deviation were selected. Selection of relevant features were then undergone discretization method namely, Minimum Description Length Principle (MDLP). The relevant features in interval data form were then classified using Random Sub Space classifier available in Weka tool.

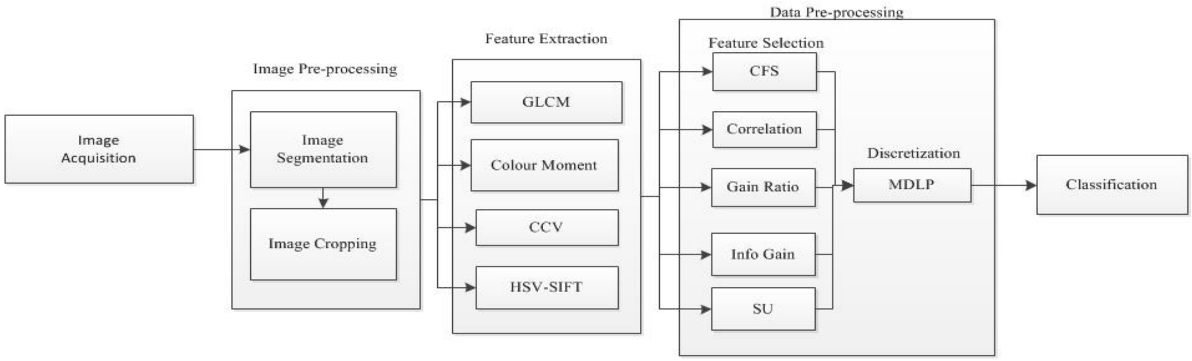

Figure 2: Flowchart of methods in non-destructive mango ripeness classification

4. RESULTS AND DISCUSSIONS

Table 3 shows error rate, correlation score and classification time among three compared feature selection and MDLP discretization method using Random Sub Space classifier. From the experiment, it is observed that Principal Component and Ranker have the best result. Although the error rate of 0.3439 is not the lowest compared to the others, its correlation score is the highest among them. The larger numbers of correlation score indicating the more relevant features. Moreover, the time taken for processing is 0.06 seconds which is good because it takes short time to do processing. Thus, the presence of feature selection and discretization may enhance mango ripeness classification. The results from Table 4 are taken from running of CFS and MDLP, specifically from colour moment feature. From Table 4, it can be said that mango is unripe when mean from Red colour channel feature is until 52.153867 and ripe if the mentioned feature is greater than 52.153867.

Table 3: Comparison of error rate, correlation score and time among compared methods using Random Sub Space classifier

Feature Selection + Search methods / Specifications	Correlation-based Feature Subset Selection (CFS) + Best First	Principal Component + Ranker	Classifier Subset Evaluation + Genetic Search
Error Rate	0.527	0.3439	0.0
Correlation Score	0.0911	0.9894	0.0
Time (seconds)	0.12	0.06	0.05

Table 4: Sample data pattern from CFS feature selection and multi-interval discretization run

Feature extraction	Feature Name	Mango classification	Data Pattern
Colour moment	Mean-Red colour channel	Unripe	(-inf-52.153867]
		Ripe	(52.153867-inf)

5. CONCLUSION

This paper discusses about colour features and texture features extractions followed by feature selection methods and discretization which are essential to perform mango ripeness classification. Since data discretization and feature selection have been performed prior to the learning phase in the classifiers, they extensively reduce the processing effort of the learning algorithm. Apart from that, this proposed method assists to generate knowledges on identifying important features and ripeness range of mango. Essential feature selection and discretization proved that it is proficient in reducing irrelevant attributes and produces good classification accuracy. Finally, the classification results illustrate that essential features and data pattern could achieve better performance. Therefore, this method helps to boost agriculture industrial.

ACKNOWLEDGMENT

The research reported in this paper is supported by Fundamental Research Grant Scheme (FRGS) [Grant number: 9003-00554]. The author would like to express gratitude to Malaysia Ministry of Higher Education (MOHE) and University Malaysia Perlis for the facilities provided.

REFERENCES

Blasco, J, Cubero, S. et al. (2017, January 2) COFILAB: Computers and Optics in Food Inspection. Retrieved from http://www.cofilab.com/.

C. A. Hussain, D. V. Rao, T. Praveen, and B. E. College, (2013). Color Histogram Based Image Retrieval. *Int. J. Adv. Eng. Technol.*, IV, 63–66.

Dah-Jye Lee, James K. Archibald, and Guangming Xiong (2010). Rapid Colour Grading for Fruit Quality Evaluation Using Direct Colour Mapping. IEEE Transactions on Automation Science and Engineering 8(2), 292-302.

Dayan and Savakar (2012). Identification and Classification of Bulk Fruits Images using Artificial Neural Networks International Journal of Engineering and Innovative Technology (IJEIT), 1(3).

D. C. Slaughter. (2009) Nondestructive Maturity Assessment Methods for Mango, 1–18.

Huang, M., Zhu, Q., Wang, B., & Lu, R. (2012). Analysis of hyperspectral scattering images using locally linear embedding algorithm for applemealiness classification. Computers and Electronics in Agriculture, 89, 175–181.

Jagadeesh. D. Pujari, Rajesh. Yakkundimath and A. S. Byadgi (2013). Reduced Color and Texture features based Identification and Classification of Affected and Normal fruits images, International Journal of Agricultural and Food Science, 3(3), 119-127.

Kamruzzaman, M., Sun, D.-W., ElMasry, G., & Allen, P. (2013). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis.Talanta, 103, 130–136.

Khoje, Suchitra, 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 (2013): 1-7.

L.Ladha et al. (2011). Feature Selection Methods and Algorithms. *Int. J. Comput. Sci. Eng.*, 3(5), 1787–1797.

M. Kishore, N. Sengar, N. Minhas, B. Sarkar, and A. Goon. (2016). Image processing based classi fi cation of grapes after pesticide exposure. *LWT - Food Sci. Technol.*, 72, 368–376.

M. Rayat, I. Hossain, I. Ahmed, and H. Kabir, "Automatic Lung Tumor Detection Based on GLCM Features," Springer Int. Publ. Switz. 2015, pp. 109–121, 2015.

Naik, S., Patel, B., & Pandey, R. (2015). Shape, size and maturity features extraction with Fuzzy classifier for non-destructive mango (Mangifera Indica L., cv. Kesar) Grading, (Tiar), 1–7.

Nandi, C. S., B. Tudu, and C. Koley. "Machine Vision Based Techniques for Automatic Mango Fruit Sorting and Grading Based on Maturity Level and Size."Sensing Technology: Current Status and Future Trends II. Springer International Publishing, 2014. 27-46.

Pandey, R.; Gamit, N.; Naik, S., "Non-destructive quality grading of mango (Mangifera Indica L) based on CIELab colour model and size," Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference, pp.1246,1251, 8-10, May2014.

P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar. (2013) Image Texture Feature Extraction Using GLCM Approach, *Int. J. Sci. Res. Publ.*, 3(5), 1–5.

Rajkumar, P., Wang, N., Eimasry, G., Raghavan, G. S. V., & Gariepy, Y. (2012). Studies On banana fruit quality and maturity stages using hyperspectral imaging. Journal Of Food Engineering, 108(1), 194–200.

R. M. Kumar. (2014). A Survey on Image Feature Descriptors, 5(6),7668–7673.

Sapan Naik and Bankim Patel," CIELab based color feature extraction for maturity level Grading of Mango (Mangifera Indica L.)", National journal of system and information technology (0974-3308), 2014

S. Arivazhagan, R. NewlinShebiah*, S. Ananthi, S. Vishnu Varthini. (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. CIGR Journal, 15(1), pp.211.

S. K. Bodhe. (2012). Applications of Image Processing in Agriculture : A Survey, 52(2), 34–40.

S. Naik, B. Patel, and R. Pandey. (2015) Shape , size and maturity features extraction with fuzzy classifier for non-destructive mango (Mangifera Indica L ., cv . Kesar) grading, no. Tiar,1–7.

U.Fayyad K.Irani (1993). Multi interval discretization of continuous valued attributes for classification learning, Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 1022–1027.

W. Rich and D. Buhl-brown. (2015) Texture Analysis of Remote Sensing Imagery with Clustering and Bayesian Inference. *I.J. Image, Graph. Signal Process.*, no. August, 1–10.