Evaluation of Mangosteen Surface Quality using Discrete Curvelet Transform

Slamet Riyadi
Dept. of Information Technology
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
riyadi@umy.ac.id

Jaenudin
Dept. of Information Technology
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

Laila Ma'rifatul Azizah
Dept. of Information Technology
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
lailamazizah@umy.ac.id

Cahya Damarjati
Dept. of Information Technology
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
cahya.damarjati@umy.ac.id

Tony Khristanto Hariadi
Dept. of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
tonykhariadi@umy.ac.id

***Abstract*—Mangosteen is the major commodity of fruit export from Indonesia. Only free-defect mangosteens are exported which were conventionally classified by human vision. In order to automate the classification between defect and free-defect mangosteen surface and handle high volume of export, machine vision has a great opportunity. The objective of this paper is to classify mangosteen surface images using discrete curvelet transform (DCT). The curvelet transform is a multiscale directional transform, which allows an optimal non-adaptive sparse representation of objects with edges. The methodology of this research involved pre-processing, implementation of DCT, statistical features extraction and classification using linear discriminant analysis. The method has been implemented on a number of 80 mangosteen images and validated using 4-fold cross validation method. The highest accuracy of classification between defect and non-defect surface is 92.5% obtained on second scale of DCT. In conclusion, the proposed method is able to evaluate mangosteen quality surfaces.**

Keywords*—*curvelet, mangosteen, matlab, image processing

I. Introduction

Mangosteen (Garcinia mangostana L.) is one of Indonesia's leading export commodities. The high demand for mangosteen makes mangosteen fruit as the highest contributor of foreign exchange for Indonesia among other fruits. The Central Bureau of Statistics (BPS) notes that throughout 2013, Indonesia's mangosteen exports reached USD 5.73 million or approximately Rp 63 billion (Sindonews 2014). Meanwhile, the Ministry of Commerce also noted that during January-May 2014, martial art exports grew by 153% or USD 13.7 million compared to the same period in 2014.

The mangosteen fruit will be exported must meet the FAO quality standard, ASEAN or importing countries in order to be traded internationally. In this context, appearance measurement techniques must be used to guarantee good external quality of produce that meets the quality standards [1]. Grading Asean standards divide the quality of the mangosteen fruit into three classes, the extra classes, class I and class II.Petaman class in the set Based on the quality of the fruit surface and the extent of damage. Diseases in fruit also cause major problem and cause economic loss, a research written by Ranjit et al. [2]. This surface texture is important in order to extract the exact value for feature extraction. Currently,the process of mangosteen is still done manually by employees at the packing company. The way it feels less effective if done on a large scale because it costs a lot of money, a lot of labor, takes a long time and results are not subjective so that the results are not uniform.

Many methods have been done to handle the problem, one of them by utilizing digital image processing technology. The method has been developed to perform fruit and vegetable quality inspection one part of image processing technology is a multi-scale direction transformation. A paper titled Digital Image Processing for Fruit Grading using Artificial Neural Networks (ANN). This study tried to classify fruit based on some classes by using ANN method [3]. Other research titled Non-destructive method for maturity index determination of Garcinia mangostana L using image processing technology. This reserach shows that Support Vector Machine can be used to classify fruit with accuracy 83.3% [4]. A paper from Arivazhagan et al. tried to solve the issue of texture classification based on curvelet transform features (CSFs) and curvelet co-occurrence features (CCFs). This study shows that these kind of method allows obtaining high degree of success rate in classification [5].

A research was done by Mohana and Prabhakar [6] in paper titled A Novel Technique for Grading of Dates using Shape and Texture Features. This study combined between shape and texture, begin with bilateral filter, then segmentation using threshold to remove the background as noises. Then shape of fruit was extracted using Curvelet Transform and Local Binary Pattern (LBP) from particular region of the fruit. They compared some classifiers, there were k-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). The research found that kNN was the best classifier for this case. Curvelet Transformation also used by Neema and Sasikumar in Image Denoising Method based on Curvelet Transform with Thresholding Function [7]. The problem that they wanted to solve was remove noises in digital images. The study shows that in terms of the PSNR, their proposed denoising technique performs better than other methods. A review titled Color, Size, Volume, Shape, and Texture Feature Extraction Techniques for Fruit [8] performed by Manali and Sumati. This review shows that skin defect detection at higher resolution used wavelet and for low resolution curvelets are best option. Other research was performed by A Majumdar et al. [9] done comparative study between curvelets, wavelets, and contourlets. This study shows that curvelet is the best option for low resolution images for pattern recognition. A paper wrote by Khoje, Bodhe, and Adsul titled Automated Skin Defect Identification System for Fruit Grading based on Discrete Curvelet Transform [10] implemented Curvelet to evaluate Guava and Lemon using Support Vector Machine (SVM) and Probabilistic Neural Network (PNN) as classifiers. The study shows that highest accuracy reach 96%.

Curvelet transformation is one of the multiscale geometry transformations developed in an attempt to overcome the attachment of limitations of traditional multiscale representations such as wavelets. Conceptually, curvelet transforms are multiscale pyramids with multiple directions and positions on each length scale, and needle-shaped elements on fine scales. Curvelets has a useful geometric feature that distinguishes it from wavelets. Authors of this paper previously published articles on evaluation of mangosteen quality surface using SVM [11] and deep learning [12]. A statistical features extraction of curvelet was also explored [13]. The objective of this paper is to explore the use of curvelet transform in detail, specially the effect of scale variation of curvelet.

II. Theory

A. Discrete Curvelet Transformation

Curvelet transformation is a technology that can separate noise from signal into frequency dimension, dip, azimuth and location. Because of its advantages this curvelet transform is known as a multi-dimensional transformation not shared by other transformation technologies such as Wavelet. Curvelet is widely used for the purposes of the denoising image.

Discrete Curvelet Transform is a development of the previous version, the Continuous Curvelet Transform. In Discrete Curvelet Transform perform the transformation by dividing the image using the middle box on the picture. Curvelet transformation will result in redundant information so as to represent the signal at the edges of the image curve so that the curvelet transform is redesigned later and introduced as Fast Discrete Curvelet Transform (FDCT).The discrete curvelet transform can be represented by (1)

$$C^D(j,l,k) = \sum_{0\le t1,t2<n}^{n} f[t1,t2]\phi^D_{j,k,l}[t1,t2] \quad (1)$$

B. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a generalization of linear discriminant Fisher, a method used in statistics, pattern recognition and machine learning to find linear combinations of features that characterize or separation two or more object classes or events. The resulting combination can be used as a linear classifier, or, more generally, for dimensional reduction before the next classification shown in Fig. 1

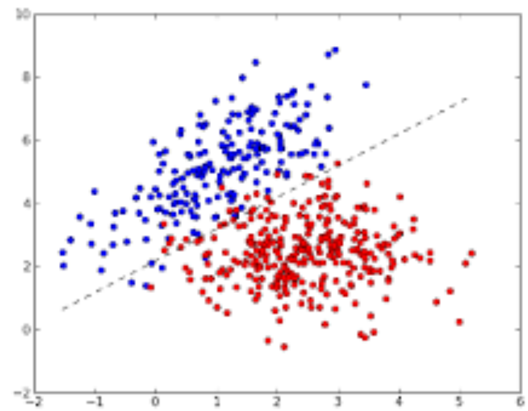

Fig. 1. Scatter plot between two features and separated by a line from linear discriminant analysis

C. K-Fold (Cross Validation)

Cross Validation is one of the methods used to validate the accuracy of a model built on a particular dataset. Model making usually aims to classify new data that has never appeared in the dataset. The data used in the data model development process is called data trainning, while the data to be used for model validation is called data testing. One of the popular cross-validation methods is K-Fold Cross Validation. In this technique the dataset is divided into a number of K-fruit partitions at random shown in Fig. 2

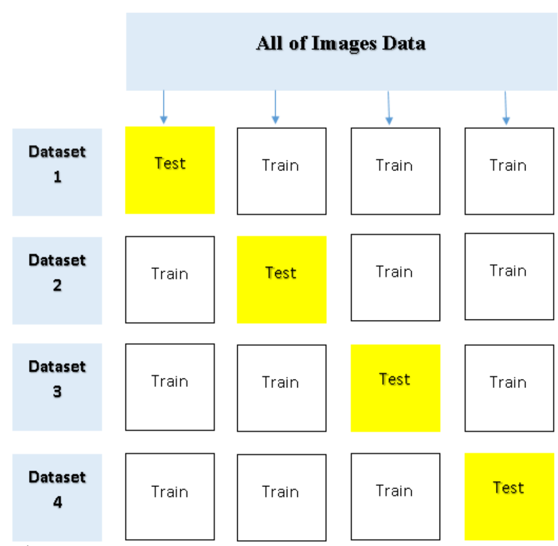

Fig. 2. Illustration of K-Fold cross validation method where data is divided into 4 dataset

III. Methodology

The first step is getting data of images manggosteen and saved as varriable array by system then is resize data of images resulution to 512x512 and converted to greyscale image.The second step is transformation data using curvelet to get the best extraction for detection. The result of curvelet transformation extracted to get the value of extraction on third step. The final step is classification, the values of extraction will be classified using LDA, to classify data of manggosteen defect and non defect.The methodology of the research described in Fig. 3

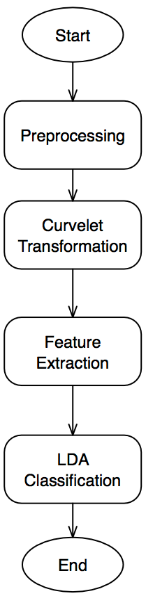

Fig. 3. Methodology which is used this research starting from preprocessing until classification

A. *Pre-processing*

This step is resizing resulution of image data so that the images data which will be prosess have the same resulution. Equation (2) is the formula of resize image

$$f(x,y) = \sum_{i=0}^{1}\sum_{j=0}^{1} a_{ij}x^i y^j = a_{00} + a_{01} + a_{10} + a_{11}xy \quad (2)$$

After date the data of image had resize converted to greyscale. conversion to grayscale image aims to simplify the input image thereby reducing processing time. The following is the formula of conversion to greyscale.

$$P=0.2989R+0.5870G+0.1140B$$

B. *Discrete Curvelet Transformation*

This stap is transformation images data using curvelet. Transformation is performed by fdct function in matlab.The result of is coefficient scale and represented by figure. In the research Analysis will be performed on the scale of each images for getting the best feature extraction.

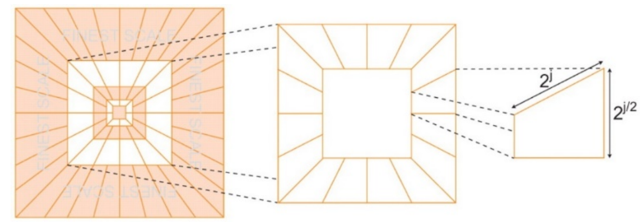

Fig. 4. Discrete curvelet transformation

Fig. 4 shows (a) is coefficient of image, figure (b) is the scale of coefficient and figure c is angle of scale of coefficient. Then the research was conducted on each scale of curvelet extraction results

C. *Feature Extraction*

Feature extraction is the process that raises the unique characteristics of an object in the form of value that will be used for further analysis. Some of previous research that use feature extraction process are in Quality modeling of Malang Oranges [14]. The result of feature extraction is mean, energy, entropy and standard deviation, the illustration is show in Fig. 5.

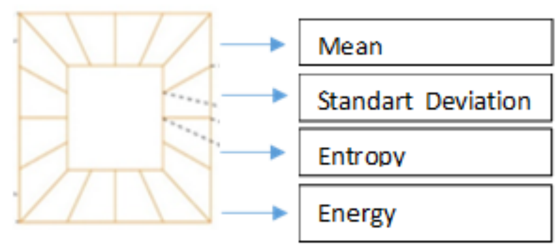

Fig. 5. Features extraction step which computed four statistical values, i.e. mean, standard deviation, entropy and energy

In the research feature performed by equation:

a) Mean

The mean value is derived from the calculation by summing the value of each matrix element from the 1st matrix element to the N matrix element then divided by the many matrix elements present. The equation (3) for calculating the average value is as follows:

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} A_{ij} \quad (3)$$

b) Standard Deviation

The standard deviation value is calculated by subtracting the value of each of the 1st matrix elements to the Nth matrix element with the mean value of each matrix element. The result of this reduction is squared and summed on each 1 to N of pixel. The sum is then quadratically rooted to get the standard deviation value.

Equation (4) is used for calculating the Deviation Standard value are as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}|A_{i,j}-u_{ij}|^2}{M*N}} \qquad (4)$$

c) Entropy

Entropy functions in showing the size of the disorder of a form. The value of H will be of great value if using an image with evenly distributed gray levels and small values if the image structure is used irregularly. The value of entropy indicates the randomness of the gray-degree distribution of an image. The more random the distribution of the gray degree, the higher the entropy value generated. The equation (5) for calculating the entropy value is as follows:

$$H = -\sum_{i=1}^{G} p(d_i).\,log_2 p(d_i) \qquad (5)$$

d) Energy

Energy is a feature used to measure concentration of intensity pairs on a co-occurrence matrix. The value of energy will produce a great value if the distribution of gray level of the image has a constant or periodic shape. The higher the entropy value the lower the energy value. This is because, the energy value describes the regularity of spreading the gray degree of an image, so it can be said that energy is the inverse of entropy. The equation (6) for calculating energy value is as follows:

$$Energy = \sum_{i,j} P(i,j)^2 \qquad (6)$$

D. *Classification*

Classification will be performed in two step.The first for trainning step and the second for testing step. The step trainning will be performed by fitcdiscr matlab function and testing will be performed by predict matlab function. The dataset will be divided into four folds, each fold consisted of training data and testing data. All fold would be applied with same condition and same formula. The method would classify the images into two class, defect and non-defect class. We can sum the accuracy from those fold and got the final accuracy. The result of classification is the acurration each fold that testing.

IV. Result

Dataset was tested by divided them into four folds. Each fold then was applied same method. The result of each fold will be shown in some tables below. Each table show the scale, class, sample which was consist of two classes (defect and non-defect), and accuracy.

TABLE I. FOLD 1 RESULT

Curvelet	Category	Sample		Total	Accuracy
		Defect	***Non-Defect***		
First Scale	Defect	9	1	10	95%
	Non-defect	0	10	10	
Second Scale	Defect	10	0	10	100%
	Non-defect	0	10	10	
Third Scale	defect	9	1	10	95%
	Non-defect	0	10	10	
Forth Scale	defect	8	2	10	90%
	Non-defect	0	10	15	
Fifth Scale	defect	7	3	10	85%
	Non-defect	0	10	10	
Sixth Scale	defect	8	2	10	90%
	Non-defect	0	10	10	

Table I shows that the highest accuracy is on the scale of and the spatial curvelet with 100% accuracy, while the lowest accuracy is on a scale of 5 with an accuracy of 85%.

TABLE II. FOLD 2 RESULT

Curvelet	Category	Sample		Total	Accuracy
		Defect	***Non-Defect***		
First Scale	defect	5	5	10	75%
	Non-defect	0	10	10	
Second Scale	defect	7	3	10	85%
	Non-defect	0	10	10	
Third Scale	defect	9	1	10	95%
	Non-defect	0	10	10	
Forth Scale	defect	4	6	10	70%
	Non-defect	0	10	15	
Fifth Scale	defect	7	3	10	85%
	Non-defect	0	10	10	
Sixth Scale	defect	6	4	10	80%
	Non-defect	0	10	10	

Table II shows that the highest accuracy is on a scale of 3 and curvelet frequency with an accuracy of 95%, while the lowest accuracy is on a scale of 4 with an accuracy of 70%.

TABLE III. FOLD 3 RESULT

Curvelet	Category	Sample		Total	Accuracy
		Defect	***Non-Defect***		
First Scale	defect	8	2	10	90%
	Non-defect	0	10	10	
Second Scale	defect	9	1	10	95%
	Non-defect	0	10	10	
Third Scale	defect	8	2	10	90%
	Non-defect	0	10	10	
Forth Scale	defect	5	5	10	75%
	Non-defect	0	10	15	
Fifth Scale	defect	7	3	10	85%
	Non-defect	0	10	10	
Sixth Scale	defect	7	3	10	85%
	Non-defect	0	10	10	

Table III shows that the highest accuracy is on a scale of 3 and curvelet frequency with an accuracy of 95%, while the lowest accuracy is on a scale of 4 with an accuracy of 70%.

TABLE IV. FOLD 4 RESULT

Curvelet	Category	Sample		Total	Accuracy
		Defect	*Non-Defect*		
First Scale	defect	8	2	10	90%
	Non-defect	0	10	10	
Second Scale	defect	8	2	10	90%
	Non-defect	0	10	10	
Third Scale	defect	8	2	10	90%
	Non-defect	0	10	10	
Forth Scale	defect	10	0	10	100%
	Non-defect	0	10	15	
Fifth Scale	defect	9	1	10	95%
	Non-defect	0	10	10	
Sixth Scale	Defect	10	0	10	80%
	Non-defect	4	6	10	

Table IV shows that the highest accuracy is on the 2nd scale and the spatial curvelet with an accuracy of 95%, while the lowest accuracy is on a scale of 4 with an accuracy of 75%.

Testing result for the fourth fold is then done the calculation of the average accuracy to determine the accuracy of each result of feature extraction. The complete percentage of overall accuracy is shown in the table V.

TABLE V. ACCURACY RESULT

Curvelet	Accuracy (%)				Average %
	Fold 1	*Fold 2*	*Fold 3*	*Fold 4*	
First Scale	95	70	90	90	86. 25
Second Scale	100	95	85	90	92. 5
Third Scale	90	95	90	90	91. 5
Forth Scale	90	70	75	100	83. 75
Fifth Scale	85	85	85	95	87. 5
Sixth Scale	90	80	85	80	83. 75

Table V shows the highest level of accuracy is on the scale of 2 and the frequency, which for accuracy on the scale of 2 is 92. 5% and for the frequency 93. 75%, while the lowest accuracy is on the scale of 4 and the scale of 6 that is 83. 75%.

V. CONCLUSION

The method developed in this study can detect mangosteen defects with the highest degree of accuracy of 92.5% for the second scale and the lowest is 83.75% for the sixth scale image decomposition using curvelet transformation. Suggestions for further research related to this research that are classify by other methods such as SVM and increase the image of trained mangosteen handicapped with low disability level to the high disability level.

ACKNOWLEDGMENT

Authors would like to acknowledge Directorate of Research and Community Services, Directorate General of Research and Development, Ministry of Research, Technology and Higher Education, Republic of Indonesia for the PTUPT research grant and Universitas Muhammadiyah Yogyakarta.

REFERENCES

[1] B. J. Cubero, N. Aleixos, E. Molt, J. Gmez-Sanchis, “Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables,” in Food Bioprocess Technol, 2010, pp. 1–18.

[2] K. Ranjit, H. Chethan, and C. Naveena, “Identification and classification of fruit diseases,” Int. J. Eng. Res. Appl., vol. 6, no. 7, 2016, pp. 11–14.

[3] Y. P. Wiharja and A. Harjoko, “Digital image processing for fruit quality classificaation using artificial neural network,” IJEIS, vol. 4, no. 1, 2014, pp. 57–68.

[4] I. Prabasari, N. A. Utama, N. A. U. Hasanah, S. Riyadi, and T. K. Hariadi, “Non destructive method for maturity index determination of Garcinia mangostana L using image processing technology,” Int. Soc. Hortic. Sci., 2016.

[5] S. Arivazhagan, L. Ganesan, and T. G. S. Kumar, “Texture classification using curvelet statistical and co-occurrence features,” in 18th International Conference on Pattern Recognition (ICPR’06), 2006.

[6] S. Mohana and C. Prabhakar, “A novel technique for grading of dates using shape and texture features,” Mach. Learn. Appl. An Int. J., vol. 1, no. 2, 2014, pp. 15–29.

[7] N. Neema and M. Sasikumar, “Image denoising method based on curvelet transform,” Int. J. Sci. Technol. Eng., vol. 2, no. 12, 2016, pp. 387–402.

[8] M. R. Satpute and S. MJagdale, “Color, size, volume, shape and texture feature extraction techniques for fruits: a review,” Int. Res. J. Eng. Technol., no. 2010, 2016, pp. 2395–56.

[9] A. Majumdar and B. Arusharka, “A comparative study in wavelets, curvelets and contourlets as feature sets for pattern recognition,” Int. Arab J. Inf. Technol., vol. 1, no. 1, 2009, pp. 47–51.

[10] A. Khoje, S.K. Dr, Dr Bodhe, “Automated skin defect identification system for fruit grading based on discrete curvelet transform,” Int. J. Eng. Technol., 2013, pp. 3251–3256.

[11] Slamet Riyadi, Afrizal Zuhri, Tony K. Hariadi, Indira Prabasari, Nafi Ananda Utama, "Optimized estimation of mangosteen maturity stage using SVM and color features combination approach," International Journal of Applied Engineering Research Vol. 12 No. 24, 2017, pp. 15034-15038

[12] Laila M. Azizah, Sitti Fadillah Umayah, Slamet Riyadi, Cahya Damarjati, "Deep learning implementation using convolutional neural network in mangosteen surface defect detection," Proceeding 7th IEEE International Conference on Control System, Computing and Engineering 2017, pp. 244-248

[13] Cahya Damarjati, Wahyu Indah Triyani, Slamet Riyadi, Laila M. Azizah, Tony K. Hariadi, 'Statistical features extraction of discrete curvelet transform for surface quality evaluation of mangosteen," Proceeding 7th IEEE International Conference on Control System, Computing and Engineering 2017. Pg. 238-243.

[14] T. K. Putra, T. K. Hariadi, S. Riyadi and A. N. N. Chamim, "Feature extraction for quality modelling of Malang oranges on an automatic fruit sorting system," in 2nd International Conference on Imaging, Signal Processing and Communication, Kuala Lumpur, Malaysia, 2018.