Exploiting Distance Measurement for Ripeness Identification

Fatma Susilawati Mohamad[1], Azizah Abdul Manaf[2] and Suriayati Chuprat[3]

**[1] Faculty of Informatics, University of Sultan Zainal Abidin, Gong Badak Campus
Kuala Terengganu, Terengganu 21300, Malaysia**

**[2] Advanced Informatics School (AIS), University of Technology Malaysia, UTM International Campus
Kuala Lumpur 54100, Malaysia**

**[3] Advanced Informatics School (AIS), University of Technology Malaysia, UTM International Campus
Kuala Lumpu 54100, Malaysia**

Abstract

Color is one of the main indicators for oil palm ripeness identification. Usually, human can interpret colors differently by just looking into them visually. Therefore, a suitable color model must be chosen in order to determine the right color for the right fruit. This paper exploits the use of Distance Measurements for histogram-based oil palm ripeness identification. In this study, HSV color model is chosen to explore its potential of colors. Four Distance Measurements are selected and compared in this study. Results are shown in later part of this paper. Promising results are obtained for both categories of fruits which is ripe or unripe.

Keywords: *Fruit Ripeness Identification; HSV Color Model; Distance Measurement***.**

1. Introduction

Recently, development of oil palm is increasing due to the high demand of palm oil for cooking needs or for biodiesel applications. According to [1], the use of oil palm is stated as “a road to zero waste”.

Recent development of oil palm research indicates the need of providing a computer-based grading system [2][3]. Currently, manual system conducted by human grader at oil palm mills lead to misconduct and disputes. Since color is main indicator of ripeness, it is important to research a right technique to determine the fruit ripeness identification. As mentioned by [4], color provides valuable information in estimating the maturity and examining the freshness of fruits and vegetables. Thus, the color of the oil palm fruits remains one of the important factors which determine the grade and quality of the palm oil [5],[6].

Many literature reviews on fruits applications measures the fruits ripeness by the quantity of chemical levels in each fruits including their DNAs, protein and PH as explained above. Therefore, for oil palm fruits, the quality bunches are determined by ripeness of fruits which indicate high quality of oil content. The right identifications of ripe fruits are extremely important. Currently, oil palm grading processes are conducted manually with the help of appointed graders at oil palm mills. Therefore, the processes will lead to mistakenly grade the low quality fruits as ripe ones. This can be happened since human eyes perceive things differently and this will lead to mistake. Moreover, graders mood might vary all the time, this will be another factor to propose a computer-aided model to assess the quality of fruits based on their ripeness indicator. Good quality bunches will be evaluated based on their ripeness and a penalty will be imposed for poor quality of fruits.

Although it is not possible to obtain one hundred percent (100%) good quality bunches out of all categories, the selection process for the correct category of fruit ripeness is extremely important. However, this study focuses only on classifying two categories of fruits which are Ripe and Unripe. As explained by [7] Oil Palm Fruit Grading Manual, Fresh Fruit Bunch (FFB) is defined as fresh fruit which is sent to mill for processing within 24 hours after harvested. However in real practice, grading process of oil palm is conducted manually by human grader. Thus this manual grading process leads to misconduct and human error while inspecting the right category of fruits for the purpose of oil palm production at the mill. It is extremely important to identify the degree of ripeness of FFB must be at 95% level of confidence as mentioned by Malaysian Palm Oil Board (MPOB). Therefore, wrong evaluation of graded fruits will result wrong report regarding the oil content. However, the most critical part of oil palm grading is the fruit classification. Error in classifying the right category of FFB will cause error in estimating the oil content.

Research conducted by Federal Land Development Authority (FELDA) at mills show the estimated oil content for ripe fruit is 60%, while under ripe is 40% and unripe is only 20% minus water and dirt. This indicates the importance of the right classification of FFB during grading process is essential to prevent from mistakenly claim low quality fruits as the good ones. Problem will occur while receiving the grading report claiming the high percentage of Basic Extraction Rate by the appointed graders while they have been proven to be poor quality fruits during oil production process.

2. Methodology

Beginning of process started from image acquisition captured by certified grader as shown in Figure 1. In this process, images of Fresh Fruit Bunch (FFB) are captured using a digital camera and stored as JPEG images. Then a color model is selected for image processing since color is one of the main indicators for oil palm ripeness. Next, feature extraction process is conducted by using histogram as feature vector. In this process, mean feature is extracted from each FFB. Then, computation for feature matching is performed using Distance Measurements as will be explained further in this chapter. Lastly in the final phase, image is identified by matching process for the similarity count. In this process, recognition rate is obtained and used as a ripeness indicator.

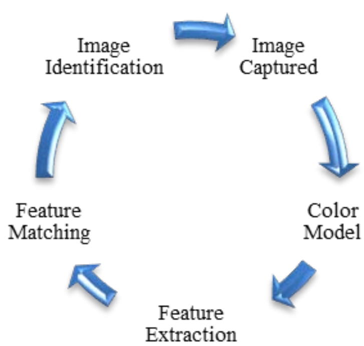

Fig. 1 Proposed Framework.

Referring to Figure 1, images of oil palm in the form of JPEG format are captured at oil palm mill. Images are then converted to HSV color model from RGB color model. Then the histogram features are extracted for every element of Hue, Saturation and Value of HSV color model. Next, calculations are performed for the defined features extracted (average mean). Then, H, S, and V values are calculated for every bunch using selected Distance Measurements. After that, range values are computed for each H, S and V for every bunch of oil palm for ripe and unripe category. The same processes are then repeated for an unknown sample of ripe FFB (30 bunches) and unripe FFB (30 bunches). Lastly, comparisons are made to match the unknown fruits with known category of fruits and the correct matches are then calculated.

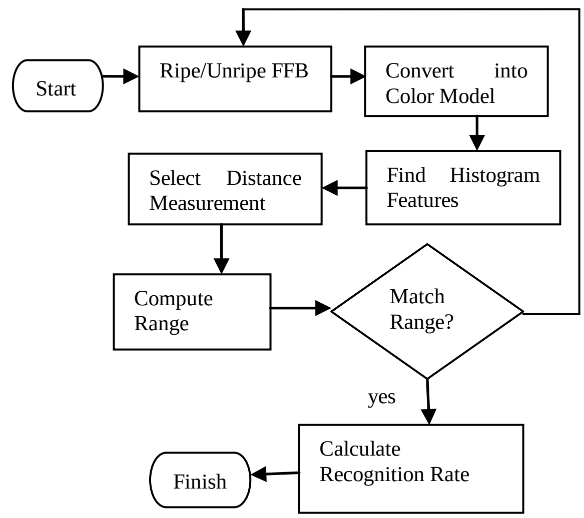

Fig. 2 Process Flow.

In this study, four Distance Measurements are used as classifiers for feature matching process. Euclidean Distance, Nearest Neighbor, Furthest Neighbor and Mean Distance are chosen because of their simplicity and robust for oil palm applications. The objective is to find the best distance measurement for oil palm ripeness identification. Comparisons are made between these Distance Measurement and recognition rates are ranked based on the highest score. Details of the distances are explained further in the next process.

In feature matching process, Distance Measurement is selected as a technique of comparisons. The chosen Distance Measurement techniques used are Euclidean, Nearest, Furthest and Mean Distance. The process of comparison is then performed within two main categories, known and unknown. Known is a category of FFB which is verified by certified grader which comprises of ripe and unripe FFB. Meanwhile unknown is an undetermined category of FFB which FFB images are taken randomly also by certified grader.

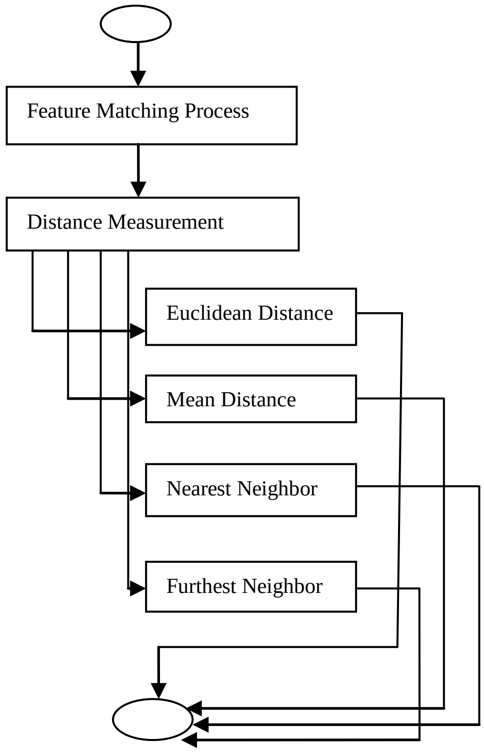

Fig. 3 Features Matching via Distance Measurement.

3. Computation

First step of matrix formulation is by computing a mean value for every image for both category of FFB. The computation for the mean value is performed by the following equation.

$$\mu = FFB\frac{\|i1+i2+....i30\|}{30} \quad (1)$$

Where μ is average mean of all mean of FFB indicated by i1, i2...i30. Table 4.1 shows the mean value extracted for every FFB in known ripe category. In this process, mean value was extracted from histogram of r1, r2, r3 until r30 where r1 refer to ripe FFB1 until ripe FFB30.

Second step is by developing matrix for every Distance selected in feature matching process. In this process, every FFB is compared within same category of FFB. 30 ripe FFB of known category is compared between each other. The computation is performed through this process where i iequal to FFB, i=1 means i starts with FFB1. Then i ≤ 30 means i equal to 1 to 30 FFB. Then i++ means after every comparison of FFBi, i is equal tu i+1. Meanwhile for j is also same as i, j=1 means FFBj starts with 1 then j ≤ 30 means FFBj is uqual to 1 to 30. Next, j++ is the process of comparing of FFB equals to j=j+1. KFFBRi and is ripe FFB within known category while KFFBRj is another FFB in the same group for the purpose of comparison as the comparison matrix is i x j.

In third step, an average mean is computed for every group of 30 FFB. Then, range value is derived from the grouped of average mean. After that image identification process is performed by comparing the range value of known category of FFB with unknown.

4. Result

The following results show comparisons for four selected Distance, Euclidean, Nearest Neighbor, Furthest Neighbor and Mean Distance are compared based on their performance between ripe and unripe as shown in the following.

a. Ripe FFB

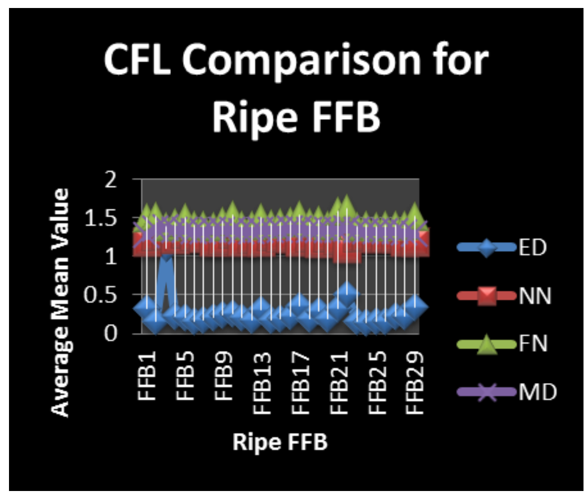

Figure 4 Comparison for Ripe FFB

Figure 4 shows comparison for ripe FFB among four classifiers. Euclidean Distance (ED) results is shown in blue line while Nearest Neighbor (NN) in red line, then Furthest Neighbor (FN) in green line and lastly Mean Distance (MD) in purple line. By looking at the result, MD is more consistent and stable for ripeness identification while NN is less reliable, followed by FN and lastly ED is the least reliable Distance for ripeness identification.

a. Unripe FFB

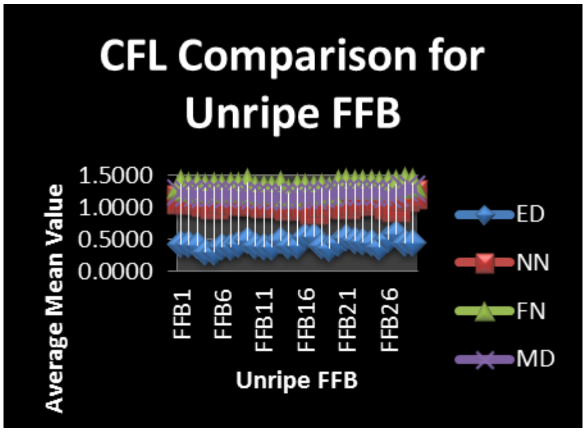

Figure 5 Comparison for Unripe FFB

Figure 5 shows a comparison for unripe FFB among four classifiers. Euclidean Distance (ED) results is shown in blue line while Nearest Neighbor (NN) in red line, then Furthest Neighbor (FN) in green line and lastly Mean Distance (MD) in purple line. By looking at the result, the most consistent Distance is MD while FN is still reliable, followed by NN and lastly, the least reliable Distance Measurement with inconsistency value of the average mean is ED.

5. Conclusion and Future Work

The objective of this study which is to find the best Distance Measurement for oil palm ripeness is met. HSV color model was proven to be the best color model for oil palm ripeness identification due to its ability in defining color intensity. This is also supported by result shown in previous study [8][9] regarding the most dominant feature for ripeness which is Value. In addition, Value is also well separated to discriminate ripeness by providing the least overlap of mean value compare to Saturation and Hue. On the other hand for Distance Measurement, Mean Distance proved as the best Distance Measurement for both category of FFB which is ripe and unripe.

Acknowledgments
Special thanks to University of Sultan Zainal Abidin (UniSZA) and Ministry of Higher Education Malaysia for funding the study leave.

References

[1] Road to Zero Waste – MPOB's Research & Development and Commercialization of Products from Oil Palm Biomass, MPOB, 2009

[2] A. Jaffar, R. Jaafar, N. Jamil, C.Y. Low, B. Abdullah, Photogrammetric Grading of Oil Palm Fresh Fruit Bunches, "International Journal of Mechanical & Mechatronic Engineering", IJMME. Vol:9 No:10, 2009

[3] N. Jamil, A. Mohamed, S. Abdullah, "Automated Grading of Palm Oil Fresh Fruit Bunches (FFB) using Neuro-Fuzzy Technique", 2009 International Conference of Soft Computing and Pattern Recognition, 2009.

[4] M.S.M. Alfatni, A.R. Mohamed Sharif, H.Z. Mohamad Shafri, Oil palm fruit bunch grading system using red, green and blue digital number, "Journal of Applied Science", 8(8): 1444-1452, 2008.

[5] S. Rashid, A.A. Nor, M. Radzali, M. Shattri, H. Rohaya, and G. Roop, "Correlation between oil content and DN values", GISdevelopment.net, 2002.

[6] W.I. Wan Ismail, M.Z. Bardaie, A.M. Abdul Hamid, Optical Properties for Mechanical Harvesting of Oil Palm FFB, "Journal of Oil Palm Research", 12, 38-45, 2000.

[7] Malaysian Palm Oil Board (MPOB).(1995). Oil Palm Fruit Grading Manual.

[8] F.S. Mohamad, A.A. Manaf, S. Chuprat, "Nearest Neighbor for Histogram-based Feature Extraction", Procedia Computer Science, 4, p.1296-1305, Jan 2011. Elsevier

[9] F.S. Mohamad, A.A Manaf, and S.Chuprat, "The Use of Elimination Method and Nearest Neighbor for Oil Palm Fruit Ripeness Indicator". Part 1,CCIS 179, pg: 674-686, 2011. Springer-Verlag.

First Author Fatma Susilawati Mohamad is currently a PhD candidate at UTM International Campus, Kuala Lumpur. She has obtained her B.Sc from Oklahoma, USA in 1997 and a M.Sc in Computer Science from National University of Malaysia in 2004. To date, she has published in plenty number of journals and conference proceedings particularly in Computer Science area. Currently, her interest is in Statistical Pattern Recognition especially in similarity or dissimilarity representations also in pattern matching and identification. She is an academic staff of University of Sultan Zainal Abidin in Terengganu, Malaysia and having a fourteen years of teaching experiences in Computer Science area.

Second Author Azizah Abdul Manaf is a professor in Advanced Informatics School (AIS) at UTM international Campus, Kuala Lumpur. Currently, she pursues her interest in Digital watermarking as well as Pattern Recognition area.

Third Author Suriayati Chuprat is a senior lecturer at Advanced Informatics School (AIS), UTM International Campus, Kuala Lumpur.