PALM OIL FRESH FRUIT BUNCH RIPENESS GRADING IDENTIFICATION USING COLOR FEATURES

N. Sabri\textsuperscript{1,}\textsuperscript{*}, Z. Ibrahim\textsuperscript{2}, S. Syahlan\textsuperscript{3}, N. Jamil\textsuperscript{2} and N. N. A. Mangshor\textsuperscript{1}

\textsuperscript{1}Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Jasin 77300 Merlimau, Melaka, Malaysia
\textsuperscript{2}Faculty of Computer and Mathematical Science, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
\textsuperscript{3}Faculty of Plantation and Agrotechnology, Universiti Teknologi MARA, Jasin 77300 Merlimau, Melaka, Malaysia

Published online: 05 October 2017

ABSTRACT

This research investigates the ripeness grading identification of the palm oil FFB using color features that are color histogram, color moment and color correlogram. Palm oil is harvested during the optimum stage of its ripeness since it improves the FFB oil quality and quantity. Harvesting wrong bunches decreases the oil extraction rate of the palm. A preliminary research on the palm oil FFB grading identification is conducted. Each ripeness stage has its own unique color. A study on color features is investigated. A new dataset of images of FFB is constructed. A comparative study between Support Vector Machine (SVM) and Naïve Bayes classifiers has been performed using the values of color histogram, color moment and color correlogram. The results of the experiments indicate that color moment with SVM produce a higher palm oil FFB ripeness grading identification accuracy compared to color histogram and color correlogram.

Keywords: color features, fresh fruit bunch (FFB), Naïve Bayes, ripeness stage, support vector machine.

Author Correspondence, e-mail: zaidah@tmsk.uitm.edu.my
doi: http://dx.doi.org/10.4314/jfas.v9i4s.32
1. INTRODUCTION

Malaysia is one of the largest countries that produce and export palm oil and palm oil production, accounting for 41% of world production and 47% of world export [1]. Elais guineensis of nigrescens (African palm oil) is the typical species found in Malaysia [2-3]. It can only be harvested after three to four years of planting and it can reach up to 30 feet high. The palm oil has a spherical shape and it consists of hard seed (kernel) encased in shell (endocarp), which is encompassed by a fleshy husk (mesocarp). This mesocarp is extracted to produce palm oil and the palm oil kernel is extracted from the kernel seed. To obtain high quantity of oil extracted from mesocarp, it is important to harvest the palm oil at its optimum stage [3-4]. In [5] state about the importance of collecting the palm oil Fresh Fruit Bunch (FFB) at the right phase of maturity to ensure ideal quality and the amount of oil production, thus profitability to the oil palm industry. Palm oil Fresh Fruit Bunch (FFBs) contains various phases of maturity level. The higher the stage of FFB, the higher the quantity of oil that can be produced. Fig. 1 illustrates the parts of the spherical palm oil and FFB.

Expert harvester can determine the maturity level of palm oil based on loses fruits. The higher quantities of loose fruits found dropped beneath the oil palm tree, the higher the possibility that they are mature. However, this practice is unreliable due to the possibilities that the loosed fruit might come from different trees. The probability of the dropped fruit took or brought by animals or flood can increase the chance of errors in determining the level of maturity [4]. Therefore, it is important to find a solution to identify the maturity phases before the palm oils falls to the ground and evade losses [6].
In [8] stated that the quality of palm oil FFB can be tenacious by their conditions such as maturity or bruises. Therefore, the value of FFBs should be based on cluster or bunch appearance and assumption of the quality and quantity of oil that can be extracted from the cluster. Mature bunches are captivating since these bunches have more oil compared to unripe ones which have lower free fatty acid (FFA) than overripe bunches. Thus, FFB ripeness is important to ensure optimum quality and quantity of oil production, hence increases profit to the industry.

In [9] studied that the composite of the oil in mesocarp happens from 15 to 16 weeks after anthesis and ended around 20 to 22 weeks. The FFB ought to be harvested at its pinnacle of ripeness. The FFBs has a different color on each maturity level. According to [10], underripe fruit has a lower oil content compared to the ripe fruit. However, when it is overripe, the oil content is lower compared to the ripe fruit. Therefore, it is important to harvest the FFB during the ripe stage but before it begins to overripe. The color of the unripe fresh fruit bunch is purplish black, covering more than 90% of the surface of the bunch. Meanwhile, the ripe FFB is red in color [11, 4]. The samples of two stages of FFB are shown in Fig. 2.

This paper is organized as follows. Section 2 explains about the literature review related to this research. Section 3 discusses the methodology results obtained and section IV explains the result analysis. Section V concludes the finding of this research and future work.
2. LITERATURE REVIEW

As mentioned by [10], color is a popular characteristic to recognize FFB ripeness [12-14]. The most popular color features are color histogram [15-16], color moment [17] and color correlogram [15]. These colors are extracted from the Red Green and Blue (RGB) image of FFB. Color histogram has less computational complexity compared to color moment and color correlogram [18]. It is less sensitive to a small color variation in the image [15]. Color correlogram is a spatial extension of the histogram. It is usually used for image retrieval [18-19]. On the other hand, color moment is simple and robust since it is insensitive to color variations [17-18].

Various FFB ripeness recognition techniques have been investigated by researchers [14, 8, 12]. In [12, 9] propose an automated palm oil grading using Artificial Neural Network (ANN). Both researches use Hue, Saturation and Value (HSV), Hue, Saturation and Intensity (HSI) color models. A good result produced by HSV but not HIS color model due to illumination and noise. One drawback of ANN is that it takes a relatively long processing time to obtain accurate results. This is due to the training process needed. It also faces an over-fitting issue. Besides that, a large size of training data is also necessary to obtain good results [20-21].

A rule-based technique to identify FFB has been proposed by [22] where it selects three types of Region Of Interest (ROI) for analysis. However, this technique is based on certain cases of FFB ripeness and not for all situations of FFB ripeness. Meanwhile, an artificial intelligence algorithm has been implemented using Neuro-Fuzzy to identify FFB ripeness stage. This algorithm used mean value of RGB image and implement Hebbs algorithm to identify the best-fit value represents the FFB image and 73.3% ripeness identification accuracy is
achieved [8]. The automatic grading machine has been suggested by [23]. This machine is developed for the FFB grading system according to Indonesian Oil Palm Research Institute (IOPRI) standard and the first grading machine that works on-site. The drawback of this RGB FFB grading machine is that it was built with high computational complexity and produced 93% ripeness identification accuracy.

Color features have been utilized for classification in most of the FFB grading research [14, 8, 12, 23, 21]. One of the advantages of color features is color uniformity. It has been used in database retrieval. However, it may produce a large set of color due to oversampling. It may lead to the increasing number of bins in the histogram [18]. The HSI color model has been used by [12] but this color model is too simple and some important information is eliminated [24].

Support vector machine (SVM) is a well know classifier and has been used in many image processing applications [25-27]. The utilization of Naïve Bayes has grown rapidly for image retrieval and classification [28-29]. Naïve Bayes surpass SVM in research on brain tumor recognition to identify two classes of brain tumor [30]. On the other hand, SVM successfully achieves higher classification accuracy in bruise and cultivar of apple detection in agriculture product [31].

The objective of this research is to identify the suitable color features for automated palm oil FFB ripeness grading identification. A comparative study between SVM and Naïve Bayes classifiers is performed to evaluate the ripeness grading accuracy for each input feature. These experimental results can be used as a preliminary research for automated FFB grading identification.

3. METHODOLOGY

Fig. 3 illustrates the methodology of FFB ripeness grading identification. The first phase is image acquisition, followed by image segmentation, color features extraction and finally ripeness grading classification. The image will be classified according to their ripeness stage that is either ripe or unripe.
3.1. Image Acquisition
The data sample used in this research was collected from the palm oil estate in Johor, Malaysia. Asus Zenfone 3 Max was used to capture the images of FFB. A total of 264 images of ripe and unripe images has been captured. These images were collected by a FFB expert grader. It was captured between 8.00 am to 12.00 pm. Fig. 4 shows a sample image of an oil palm tree as a result of image acquisition.

3.2. Image Segmentation
Image segmentation is important to remove noise or spike to obtain the accurate color features from the FFB. In this stage, K-Means clustering is selected to remove the non-fruit region.
The non-fruit region contains green and brown spike. Therefore, this operation is used to eliminate the non-fruit regions that are in green and brown colors. Before the implementation of K-Means clustering, the image is cropped into the size of 439 pixels height and 367 pixels width. Fig. 5 shows a sample of an image of a ripe FFB before and after the crop operation. The process of K-Means clustering used in this research is shown in Fig 6. It starts with the selection of the value k (starting point) that contains the palm oil fruit color [32]. This algorithm then segments the region of interest (ROI) that has the similar cluster with k (starting point). This process continues until no more pixel with the similar cluster, k is found. Fig. 7 illustrates the image of FFB after performing K-Means clustering algorithm.

![FFB Image before and after cropping operation](image1)

![Process for K-Means clustering](image2)
K-Means clustering targets to partition \( n \) into \( k \) sets \((k \leq n)\), \( S = \{ S_1, S_2, \ldots, S_k \} \) to minimize the within-clustering sum of squares (WCSS). Equation (1) shows K-Means clustering algorithm.

\[
\arg \min_S \sum_{i=1}^{k} \sum_{j \in S_i} ||x_j - \mu_i||^2
\]  

(1)

The \( \mu_i \) is the mean of points in \( S_i \) [32-33]. K-Means clustering is executed by calculating the distance for each \( k \) object to the centroids. Each object with the minimum distance is clustered together. These steps are repeated until there is no more clusters created and centroids are idle. The distance is calculated using Euclidean Distance as shown in Equation (2).

\[
d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}
\]  

(2)

where \( x \) and \( y \) are \( n \)-dimensional vectors. Euclidean Distance determines the square root difference between vectors \( x \) and \( y \) [32].

### 3.3. Feature Extraction

Feature extraction involves reducing the amount of resources required to describe a large set of data. Color is a relevant feature to determine the stages of FFB since different stages of ripeness can be identified based on different colors. This research investigates three popular color features namely color histogram [15], color moments and color correlogram. Then, the best among these three features for FFB grading is determined [34].

### 3.4. Color Histogram

Color histogram is a frequency distribution of color bin in an image. Equation (3) shows the probabilities for color histogram.

\[
h_{A,B,C}(a, b, c) = N \cdot \text{Prob}(A = a, B = b, C = c)
\]  

(3)
N represents the number of pixels in the image while A, B and C describe the three color channels (R,G,B). This operation represents the colors within an image and counts the number of pixels of each color and then forms the color histogram [18]. Fig. 8 illustrates a sample of color histogram for red, green and blue extracted from a FFB image shown in Fig. 8(a).

![Color Histogram](image)

**Fig. 8.** Sample result of color histogram (a) a sample of a FFB image (b) Result of color histogram in RGB (red, green and blue)
3.5. Color Moment

Color moment is one of the simplest and the most standout color feature among the most generally utilized components of low level features [17]. Color moment shows better dependability and is more insensitive to the pivot and zoom of images compared with shape and texture features [17]. The values of color moment can be computed into mean, standard deviation and skewness. Equation (4) shows the mean of colour values extracted from the original image represented in $E_i$.

$$E_i = \frac{1}{N} \sum_{j=1}^{N} P_{ij}$$ (4)

In Equation (5), standard deviation operation, $\sigma_i$, is defined as a square root of the variance of the distribution.

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (P_{ij} - E_i)^2}$$ (5)

In Equation (6), skewness of FFB image, $S_i$, is a measure of the degree of asymmetry in the distribution.

$$S_i = \frac{3}{N} \sqrt{\frac{1}{N} \sum_{j=1}^{N} (P_{ij} - E_i)^3}$$ (6)

where variable $N$ represents the number of image pixels and $P_{ij}$ indicates the $i_{th}$ color and $j_{th}$ pixel of the FFB image in RGB color model. Fig. 9 illustrates the results of mean, standard deviation and skewness of color moments for ripe (shown in blue line) and unripe FFB (shown in the red line).

3.6. Color Correlogram

Color correlogram is an enhancement from color histogram where this feature includes spatial correlation information. Color histogram captures only the color distribution of an image and does not include any spatial data [16]. Therefore, it is one type of spatial extension of the histogram and is broadly utilized over color histogram. In color histogram, firstly the histogram of an image was computed and stored into four equivalent bins. Every bin was subdivided into four more bins and a maximum of frequencies was computed for each such subdivision. This data is compiled in the form of a correlogram as shown in Equation (7) [19].

$I$ represent the image where $C_1$....$C_n$ describe the color that is available image $I$. $D$ represents the distance in image $I$. For pixel P1 and P2 of colors $C_i$ and $C_j$ separated by distance $k$. The
The color correlogram preserves the spatial relationship of color that is lost during histogram process. The $C^2D$ represents the size of correlogram. Fig. 10 illustrates a result for ripe and unripe where unripe has a different color distance compared to ripe.

![Fig. 9. Values of color moments (mean, standard deviation and skewness of RGB) for ripe and unripe FFB in Fig. 8(a)](image)

![Fig. 10. Sample of color correlogram of ripe and unripe for image in Fig. 8(a)](image)

3.7. Classification

The purpose of classification is to assign a group of data with homogeneous characteristics and separates the multiple data from each other within the same image. The classification of FFB process in this research is based on two classifiers that are SVM and Naïve Bayes.

3.8. Support Vector Machine (SVM)

SVM classifies the FFB images into two groups namely ripe and unripe using color features extracted earlier. SVM achieves optimal separation between two classes using a hyper plane. SVM attempts to obtain the best hyperplane when the labels available on each side of a hyperplane. The optimum margin is produced when the distance between each maximum side of the hyperplane, also known as support vector is achieved. Therefore, SVM works by separating those features into classes (ripe and unripe) until it achieves the maximum
hyperplane [26]. Fig. 11 illustrates the SVM optimum margin of FFB stage. In this research, polynomial kernel is used for the FFB ripeness grading identification between ripe and unripe.

![Fig. 11. The optimum margin of FFB stage](image)

Fig. 11 illustrates how separation between ripe and unripe works. The ripe represented in blue and unripe green. The object (ripe or unripe) touches the red line is known as support vector and the black line is decision line. The decision line separates between ripe and unripe features. The red line determine based on the distance of the closest object to the decision line.

### 3.9. Naïve Bayes

Naïve Bayes is a statistical classifier where it predicts the probability of classes by using Bayes theorem [35]. This algorithm works by using the probability rules as shown in Equation (8).

\[
P(a | b) = \frac{P(b|a) P(a)}{P(b)}
\]

where \( b \) represents the probability of data occurred in \( a \). \( P(b) \) is the total evidence contributes to the correct class of \( a \). It also can be interpreted as a conditional probability \( P(b|a) \) and prior probability \( P(a) \), also known as priori knowledge.

### 4. RESULTS AND DISCUSSION

This section discusses the tested data for three features namely color histogram, color moment and color correlogram using SVM and Naïve Bayes classifiers. 126 images of ripe FFB and 138 images of unripe FFB have been used in this experiment. The images for both stages are divided randomly into 90% training and 10% testing. The result evaluation is shown using
The confusion matrix. Table 1 shows the result of confusion matrix that contains the values of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP is where an image has a property (in the context of FFB it is correctly identified by the classifier as ripe). FP is one that does not have the property but is recognized anyway and it is wrongly identified as ripe. Meanwhile, TN is correctly identified as unripe and FN is incorrectly identified as unripe.

Confusion Matrix in Table 1 shows higher TN in the identification of unripe FFB using color moment in SVM compared to Naïve Bayes which is 131. It also shows higher TP using the SVM classifier. TN with value 124 also shows a higher result using SVM on identify ripe FFB. It shows that SVM can identify bi-class between ripe and unripe FFB more effectively compared to Naïve Bayes. It also indicates that color moment is a better feature for the identification of the maturity of FFB compared to color correlogram and color histogram.

**Table 1.** Confusion matrix for FFB ripeness grading identification for SVM and Naïve Bayes

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Color Features</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>Correlogram</td>
<td>100</td>
<td>38</td>
<td>40</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Moment</td>
<td>131</td>
<td>7</td>
<td>2</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>120</td>
<td>18</td>
<td>29</td>
<td>97</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>Correlogram</td>
<td>87</td>
<td>51</td>
<td>35</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Moment</td>
<td>110</td>
<td>28</td>
<td>18</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>Histogram</td>
<td>99</td>
<td>39</td>
<td>48</td>
<td>78</td>
</tr>
</tbody>
</table>

The accuracy of the confusion matrix is computed using Equation (9) and the results of accuracy for each feature is illustrated in Table 2.

\[
\text{accuracy} = \frac{(TP + TN)}{\text{TOTAL DATA}} \times 100\%
\]  

**Table 2.** Accuracy for both SVM and Naïve Bayes classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>SVM</th>
<th>Naïve Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Histogram</td>
<td>82.197%</td>
<td>67.0455%</td>
</tr>
<tr>
<td>Color Moment</td>
<td>96.5909%</td>
<td>82.5758%</td>
</tr>
<tr>
<td>Color Correlogram</td>
<td>70.4545%</td>
<td>67.4242%</td>
</tr>
</tbody>
</table>

From Table 2, we can see that the performance of all color features (color histogram, color moment and color correlogram) using SVM produces promising results with the highest
accuracy 96.59% for color moment features. It indicates that color moment is suitable for FFB ripeness grading identification without using high dimensional computation. Furthermore, additional features are not necessary to obtain a high accuracy result. It also demonstrates that SVM is a better classifier compared to Naïve Bayes due to its capability to classify bi-class data.

Fig. 12 shows an example of images that have been incorrectly classified as ripe and unripe. This error is due to the lighting changes, where blur images effect the accuracy of the classification.

![Example of incorrect classification](image)

**Fig.12.** Example of incorrect classification between (a) ripe and (b) unripe class using color moment and SVM

5. CONCLUSION

This paper investigates the performance of color features namely color histogram, color moment and color correlogram for palm oil FFB ripeness grading identification. The identification results show that color moment feature with SVM classifier achieves 96.60% accuracy. It demonstrates that color moment features handle illumination changes better than color histogram and color correlogram. Future work is to perform palm oil FFB grading of ripeness identification using other color models with more categories of ripeness.

6. ACKNOWLEDGEMENTS

This research is financially supported using ARAS grant provided by Institut Pengurusan Penyelidikan dan Inovasi (IRMI), Universiti Teknologi MARA with project code 600-IRMI/DANA 5/3/ARAS (0186/2016).

7. REFERENCES


[34] Ping T D. A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering, 2013, 8(4):385-396


How to cite this article: