Novel Feature Extraction for Oil Palm Bunches Classification

Hui Hui Wang¹,²*, Yin Chai Wang¹,², Bui Lin Wee¹, Shiang Wei Sim¹

¹ Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kuching, Sarawak, Malaysia
² Institute of Tourism Research and Innovation, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

ARTICLE INFO

This paper presents research on an image processing approach for oil palm fruit branches classification on automating the process of classifying palm fruit branches based on their visual characteristics, which can have applications in the agriculture and palm oil industry. The research aims to classify the bunches into four categories which are unripe, under ripe, ripe and over ripe. Additionally, this approach can reduce the manual labour involved in palm oil grading and provide a more objective and consistent method for grading the oil. The proposed approach consists of the process of image acquisition, image pre-processing, colour processing, image segmentation and classification. The ripeness of the oil palm fruit bunches is determined based on the percentage of ripeness areas masked on the fruit surface. The proposed algorithms successfully classified the palm oil bunches and improved the accuracy of grading palm oil, achieving 85% accuracy from the experiment results.

Keywords: Palm Oil Classification; Image Processing; Ripeness; Feature Extraction

1. Introduction

The agriculture industry plays an important role in Malaysia and had become a major source of income for the country, with oil palm being the main commodity. In 2020, Malaysia accounted for 25.8% and 34.3% of world’s palm oil production and exports, respectively [1].

The global demand for palm oil will continue to rise, owing to the growing population and economy [19]. Due to the high demand in palm oil industry, Malaysian Palm Oil Board (MPOB) needs to regulate the fresh fruit bunch (FFB) quality in accordance with the ripeness of fruit, to ensure that the produced palm oil products are in high quality [2]. To increase the production of good quality crude palm oil for food industry and cosmetic industry, one of the challenges is to harvest the oil palm fresh fruit bunches at the optimum ripen stage [3,4].

Currently, human graders conduct the palm fruits classification process of the oil palm fruit manually. However, this manual grading process can lead to misconduct and human error when inspecting the category of oil palm fruits [2]. Human subjective judgments are unable to accurately analyse the content and quality of oil palm fruits. Moreover, the visual similarity of the fruits makes manual inspection challenging and also poses problems for conventional computer vision techniques.
This has led to low-quality palm oil and kernel oil production. Additionally, Malaysia is facing a shortage of expertise in manual grading of the oil palm fruit due to the tremendous production of the oil palm. Furthermore, manual grading is a tedious and time-consuming process.

Unfortunately, oil palm fruit ripeness classification based on computer vision has not gained many satisfactory results [5]. To ensure the quality of the oil palm fruit is classified accurately, a novel feature extraction for oil palm fruit branch’s classification is needed and proposed to reduce the inaccuracy of quality classification of oil palm fruit.

The main objective of this research is to classify the oil palm fruit branches into four different categories which are unripe, under ripe, ripe and over ripe. The aim of study is to focus on the extraction of oil palm fruit branches features for classification.

2. Related Works

Choong et al., [6] used the content of palm fruit to highly correlate it with the redness of the palm oil fruits under controlled environments. The images were manually edited for consistency using image processing techniques. Ghazali et al., [7] used the RGB colour components as features for palm oil classification. The captured images captured were processed by eliminating the non-red colours, and the resulting images into 3 categories: ripe, under ripe and unripe. This study has provided valuable insights into the use of image processing techniques and feature extraction for oil palm fruit classification.

Shabdin et al., [9] then improved by using the hue, saturation, and colour intensity as the main features. Classification is only based on two classes (underripe and ripe) while images were captured in a controlled environment. Septiarini et al., [10] proposed a method for ripeness classification of oil palm fruit into three classes, including raw, under-ripe and ripe. This method is applied against images that contain a single palm fruit by combining RGB and grey colour features and SVM that are preceded by threshold-based segmentation using the Otsu method. Artificial Neural Networks [ANN] is integrated for analysis and achieved accuracy of 70% [9]. Advanced artificial intelligence tools such as machine learning or deep learning are also used [5,11-16]. Images are collected, pre-processed and trained under the artificial neural network (ANN), and the ANN learns the features from those images [9,18]. Moreover, the support vector machine (SVM) [6] and naive Bayes classifier are applied to maturity grading after colour images are captured [19].

Ghazali et al., [20] processed images with image segmentation and colour feature extraction. Hue measurement was used as a feature to determine the ripeness of FFB. They also trained and tested the model using a support vector machine. In addition, they used the bag of visual word model. The average accuracy for the 4 classes of the palm oil fruit branch by colour features was 57%, while the accuracy for ripeness classification using the bag of visual word was 70%.

Septiarinia et al., [10] used the contour-based approach for segmenting oil palm fruit, as the fruit can have various shapes and colours. This method involved implementing the Canny algorithm along with several morphology and reconstruction operations to eliminate noise. The segmentation accuracy achieved an average of 90.13%, with false positive and false negative error rates of 2.92% and 5.20%, respectively.

Wong et al., [21] used linear segmentation to classify bunches based on mean hue value and achieved an accuracy of 85%. However, they only made predictions across two classes, ripe and unripe. Teh et al., [24] then performed classification by extracting mean RGB values and colour ratios, and they were the first to explicitly use edges as a supplementary feature to colour data through the use of Canny edge detection. This resulted in an accuracy of 68.75% over three classes.
Although many studies have been conducted to determine the ripeness of fruits and vegetables, there has not been much research on classifying the ripeness stage of fresh fruit bunches in an outdoor environment based on fruit detachment [20].

3. Proposed Solution

The proposed solution consists of six stages: image acquisition, image pre-processing, colour processing, image segmentation, classification, and lastly results and evaluation.

3.1 Image Acquisition

The images of the oil palm fruit bunches were taken from different angles of the fruit for each ripeness category, including unripe, underripe, ripe and overripe. A total of 100 images were captured for the 50 samples of oil palm fruit bunches used in the prototype for this research. Two images for each category were taken simultaneously from the front and back sides of each fruit bunch. The images were then converted to JPEG format.

3.2 Image Pre-Processing

To reduce unnecessary noise, the background colour of the images was eliminated, as shown in Figure 2. This was done by applying a user-created mask over the oil palm fruit bunches image to extract the background. The mask was created by converting the oil palm fruit bunches image into binary and enhancing it using several morphological operations, including dilation, erosion, and reconstruction, to extract the whole oil palm fruit bunches from the images. The oil palm fruit bunch object was then extracted, as shown in Figure 3, but there were some small holes in the masked object that needed to be filled. The background of the binary mask was then converted to black, and the masked object was converted to white, as shown in Figure 4. Since the image in Figure 4 was in binary condition, it needed to be converted to a colour image for the next stages of colour processing. The white extracted object was transformed back to the original colour image, as shown in Figure 5.

3.3 Feature extraction

HSV colour spaces was chosen as standard specification for defining colour. The three planes of HSV explains on the colour vibrancy and the brightness respectively. The Figure 6, Figure 7 and Figure
8 show on the hue, saturation and value mask from one of the converted HSV image of the sample fruit bunch.

![Fig. 6. Hue Mask](image1)
![Fig. 7. Saturation Mask](image2)
![Fig. 8. Value Mask](image3)

The code below shows on masking the hue saturation and value mask.

```matlab
hsvImage1 = rgb2hsv(extractfruit1);
% Extract out the H, S, V image individually
hImage1 = hsvImage1(:,:,1);
sImage1 = hsvImage1(:,:,2);
vImage1 = hsvImage1(:,:,3);

hueMask1 = (hImage1 >= hueThresholdLow1) & (hImage1 <= hueThresholdHigh1);
saturationMask1 = (sImage1 >= saturationThresholdLow1) & (sImage1 <= saturationThresholdHigh1);
valueMask1 = (vImage1 >= valueThresholdLow1) & (vImage1 <= valueThresholdHigh1);
ripeRegionsMask1 = uint8(hueMask1 & saturationMask1 & valueMask1);
axes(handles.firstbinarySegment);
imshow(ripeRegionsMask1, []);
```

The mask shown above indicates that the saturation component affects most of the area of the oil palm fruit bunches. Therefore, saturation was selected as the threshold value to indicate the ripeness of an oil palm fruit bunch before proceeding to the classification stage. A total of 30 fruit samples were taken to determine the value of the saturation threshold. The threshold value obtained will be the predefined value for the proposed solution to distinguish the ripe fruit values. The threshold value will be used to segment out the region of the ripeness part from the whole oil palm fruit bunch. The code to mask the ripeness part is shown below.

```matlab
hueMask1 = (hImage1 >= hueThresholdLow1) & (hImage1 <= hueThresholdHigh1);
saturationMask1 = (sImage1 >= saturationThresholdLow1) & (sImage1 <= saturationThresholdHigh1);
valueMask1 = (vImage1 >= valueThresholdLow1) & (vImage1 <= valueThresholdHigh1);
ripeRegionsMask1 = uint8(hueMask1 & saturationMask1 & valueMask1);
axes(handles.firstbinarySegment);
imshow(ripeRegionsMask1, []);
```

The region will be converted back to the original colour image before proceeding to the segmentation part. The ripe region masks use the cast function to change the image type back to the original data type. Then, the ripe region masks are used to mask out the ripe part of the RGB image, and finally, the masked region colour bands are concatenated to form the RGB image.
3.4 Masked Region Extraction

After the parameter of the saturation had been determined, masking the region of the ripeness part of the oil palm fruit bunches is carried out. The masked areas of the front and back side of the images will be calculated. The masked region area will compare with the total area of the whole fruit bunch in percentage to be used for classification process.

3.5 Classification of the Ripeness Category

The classification of the oil palm fruit bunches is based on the percentage of the masked area to the whole fruit bunch. The table below shows the classification of the ripeness category:

<table>
<thead>
<tr>
<th>Category</th>
<th>Range of the area (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unripe</td>
<td>&lt;=20</td>
</tr>
<tr>
<td>Under ripe</td>
<td>21-45</td>
</tr>
<tr>
<td>Ripe</td>
<td>46-65</td>
</tr>
<tr>
<td>Over ripe</td>
<td>&gt;65</td>
</tr>
</tbody>
</table>

The threshold value of the saturation is chosen to differentiate between the different categories of ripeness of the fruit bunch. Saturation represents the vibrancy of the colour, and different saturation values will give different results for the ripeness category. The threshold value used is within the range of 0.65-1.0. The Figure 9 shows the result of the initial test on the threshold value.

The output of the masked region of the image after the image processing stages will be compared with a range of area values in order to identify which category the fruit bunch falls under.
4. Experiments and Threshold Value Setting

Before experiments are carried out, the threshold settings were needed to mask the region for the ripe category of the oil palm fruit bunches. A total of 30 samples were taken to test the saturation threshold value. The result of the testing value was determined by referring to the colour of the masked region. If the red colour of the fruit bunch was masked, then the threshold value was identified. The following table shows some of the samples of the threshold value on the colour of the masked region. The saturation represents the vibrancy of the colour, and different saturation values will give different results of the ripeness category. The threshold value used in this research is within the range of 0.68-1.0. Figure 10 shows the scatter plot of the result of the initial test on the threshold value. The threshold value of the saturation was chosen to differentiate between the different categories of ripeness of the fruit bunch. After the threshold value of the ripeness category was determined, the ripe category range was selected for the proposed solution for the classification of the fruit bunches.
From the above implementation algorithm of the images, each ripeness category of the oil palm fruit bunch can be identified based on the area of the region that is masked on the whole fruit bunch. The Oil palm fruit bunch extraction and classification process is shown in Figure 11 below. The segmented images, the percentage of the front part, back part and overall palm oil fruit bunch ripeness are calculated. The classification of the palm fruit bunch category is classified and shown in the Figure 11.

**Fig. 10.** Ripeness threshold distributions

**Fig. 11.** Oil Palm fruit bunch extraction and classification process
5. Results and Analysis

Before testing, the fruit bunch images were visually graded by a local grader from the Sarawak Land Consolidation & Rehabilitation Authority (SALCRA) to ensure that the ripeness of the fruit bunches was known prior to the image analysis process. A total of 50 sample images were graded by the graders. From the manual grading process, 30 fruit bunches were classified as ripe, 10 as underripe, and 5 as overripe and unripe fruit bunches respectively. The results from the manual grading process were then compared to the output from the proposed solution for each individual fruit.

The masked areas of the images were calculated and recorded as tabulated in Table 2. The table shows the results for the 50 samples of fruit bunches, including the number of fruit samples, the total area masked for the front and back images of each fruit bunch, the average area masked, ripeness category, and respective comments from the proposed solution.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Area Masked Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of fruit samples</td>
<td>Percentage of area masked on front part of the fruit (%)</td>
</tr>
<tr>
<td>1</td>
<td>16.58</td>
</tr>
<tr>
<td>2</td>
<td>56.15</td>
</tr>
<tr>
<td>3</td>
<td>66.49</td>
</tr>
<tr>
<td>4</td>
<td>26.45</td>
</tr>
<tr>
<td>5</td>
<td>25.15</td>
</tr>
<tr>
<td>6</td>
<td>30.50</td>
</tr>
<tr>
<td>7</td>
<td>40.51</td>
</tr>
<tr>
<td>8</td>
<td>64.16</td>
</tr>
<tr>
<td>9</td>
<td>47.15</td>
</tr>
<tr>
<td>10</td>
<td>57.85</td>
</tr>
<tr>
<td>11</td>
<td>10.85</td>
</tr>
<tr>
<td>12</td>
<td>64.16</td>
</tr>
<tr>
<td>13</td>
<td>52.36</td>
</tr>
<tr>
<td>14</td>
<td>64.19</td>
</tr>
<tr>
<td>15</td>
<td>47.50</td>
</tr>
<tr>
<td>16</td>
<td>61.18</td>
</tr>
<tr>
<td>17</td>
<td>36.14</td>
</tr>
<tr>
<td>18</td>
<td>80.30</td>
</tr>
<tr>
<td>19</td>
<td>39.48</td>
</tr>
<tr>
<td>20</td>
<td>54.61</td>
</tr>
<tr>
<td>21</td>
<td>11.67</td>
</tr>
<tr>
<td>22</td>
<td>30.26</td>
</tr>
<tr>
<td>23</td>
<td>65.15</td>
</tr>
<tr>
<td>24</td>
<td>25.48</td>
</tr>
<tr>
<td>25</td>
<td>41.36</td>
</tr>
<tr>
<td>26</td>
<td>57.15</td>
</tr>
<tr>
<td>27</td>
<td>66.49</td>
</tr>
<tr>
<td>28</td>
<td>50.18</td>
</tr>
<tr>
<td>29</td>
<td>58.15</td>
</tr>
<tr>
<td>30</td>
<td>10.32</td>
</tr>
<tr>
<td>31</td>
<td>51.37</td>
</tr>
<tr>
<td>32</td>
<td>81.37</td>
</tr>
<tr>
<td>33</td>
<td>53.36</td>
</tr>
<tr>
<td>34</td>
<td>56.70</td>
</tr>
<tr>
<td>35</td>
<td>70.33</td>
</tr>
</tbody>
</table>
There was a total of seven fruit bunch samples that were misclassified. The table above indicates misclassification for sample numbers 6, 13, 16, 25, 41, 43, and 44. Most of the misclassifications occurred in the ripe and overripe categories, possibly because some of the root parts of the fruit bunch may have masked the reddish-orange (ripe) and reddish-yellow (overripe) colours of the fruit. Figure 12 and Figure 13 show examples of fruit bunches where the masked region includes the surface as well as the root part of the fruit in the ripe and unripe categories, respectively.

Hence, the area of the masked regions might include which will affect the accuracy of the grading. The misclassification on this two sampling of the average percentage on the masked area is nearly close to the range of the ripe and over ripe category. Most of the fruit surface was actually ripe if manually inspected and the result from the proposed solution is actually correct. The fruit images should be manually graded again by the grader to avoid for the confusion and to clarify that the ripeness category is processed by the proposed solution is correct. The fruit sampling of number 6 and number 25 might due to the reason where there is a huge different of the percentage of the front and back side of the images. This may due to the reason where the front and back sides of the images were taken in the uncontrolled environment. The light intensity of the image captured is also one of the reasons where there is an inaccurate in classification of the fruit.

The best solution would be to take the fruit bunch images in the rotating condition so that all sides of the images will be captured and to increase the accuracy of the result. The accuracy of pineapple ripeness classification will be calculated using the equation below.
Classification accuracy = \frac{\text{Total number of correct classification}}{\text{Total number of experiments on fruit samples}}

The experiment results achieved accuracy of 86%.

6. Conclusions

In this paper, the proposed algorithms have succeeded in classifying palm oil bunches and improving the accuracy of grading palm oil. The proposed algorithm is able to classify palm oil fruits into four main groups: unripe, under ripe, ripe, and overripe, achieving an overall accuracy of 86%. One of the main advantages of this approach is that it doesn't require a training phase, which can be resource-intensive and time consuming. The solution relies on expert knowledge in palm oil fruit characteristics, which can lead to accurate feature extraction. The accuracy of the proposed solution is a significant improvement over the traditional manual grading process, which is prone to errors and subjectivity. With this new solution, farmers and agriculturalists can determine the grade of fruit categories more effectively and efficiently. In the future, the proposed algorithm can be deployed in mobile and cloud-based platforms, allowing farmers and agriculturalists to access the grading system from anywhere and at any time. This can help to democratize access to advanced agricultural technologies and promote more equitable and sustainable agricultural practices.

Acknowledgement

This research was funded by an Industry grant (Decision Support Dashboard Architecture and Design for Smart Farming), Universiti Malaysia Sarawak (IG/F08/U-GLOBAL/02/2021) and supported by the Institute for Tourism Research and Innovation, Universiti Malaysia Sarawak.

References


[22] Chew, Chien Lye, Chun Yong Ng, Wai Onn Hong, Ta Yeong Wu, Yee-Ying Lee, Liang Ee Low, Pei San Kong, and Eng Seng Chan. "Improving sustainability of palm oil production by increasing oil extraction rate: A review." *Food and Bioprocess Technology* 14 (2021): 573-586. [https://doi.org/10.1007/s11947-020-02555-1](https://doi.org/10.1007/s11947-020-02555-1)
