Computers and Electronics in Agriculture 118 (2015) 340–349

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A non-destructive oil palm ripeness recognition system using relative entropy

Attaphongse Taparugssanagorn[a,b,*], Siwaruk Siwamogsatham[b], Carlos Pomalaza-Ráez[c]

[a] School of Engineering and Technology/Telecommunications, Asian Institute of Technology, Thailand
[b] National Electronics and Computer Technology Center, Thailand
[c] College of Engineering, Technology, and Computer Science, Department of Engineering, Purdue University, IN, USA

ARTICLE INFO

Article history:
Received 30 July 2014
Received in revised form 4 February 2015
Accepted 18 September 2015
Available online 3 October 2015

Keywords:
Histogram color discrimination
Information theory
Kullback–Leibler divergence
Nigrescens fruits
Virescens fruits

ABSTRACT

This paper introduces a relative entropy based image processing approach for the non-destructive prediction of the maturity of oil palm fresh fruit bunches (FFB) which enables the determination of the correct time for harvesting. The results of an experimental study of applying the Kullback–Leibler distance to the problem of oil palm classification are presented. It is shown that the proposed algorithm has an excellent accuracy and it can be computed very fast. The overall proposed system is simple and useful for oil palm farmers and entrepreneurs.

1. Introduction

Oil palm (Elaeis guineensis) is cultivated on approximately 15 million hectares across the world due to its economic importance (Fitzherbert et al., 2008). It is a highly efficient oil producer, with each fruit containing about 50% of oil. As a result it requires ten times less land than other oil producing crops. Palm oil is a vegetable oil used for a variety of applications, not only edible products, such as cooking oils, margarines, baked goods, but also for non-edible purposes, such as soaps, washing powders, cosmetics, and bio-fuels. A recent trend has been the increasing use of biofuels to reduce the reliance on fossil fuels. This trend has created a demand for palm oil as a feedstock ingredient in the production of a biodiesel, which is also known as palm oil methyl ester, reducing diesel use and consequently reducing CO_2 emissions (Fitzherbert et al., 2008).

In the area of harvesting discipline and quality control for oil palm fruits, color has been an important guide to determine whether the oil content has reached a maximum such that the fruit bunch is ready for cutting (Ng, 1957). Two main indicators, i.e., oil extraction rate (OER) and free fatty acid level (FFA), can indirectly influence the profitability of any plantation enterprise. The national standard stipulates that the theoretical extraction rate is between 21–23% and the FFA content should not be higher than 5% (Ng, 1957). It has been recognized that this is the result of processing poor quality fresh fruit bunches (FFB). In fact for every 1% of unripe bunches present, the OER will decrease by 0.13%, while the FFA content will increase linearly as the percentage of overripe bunches increases (Siregar, 1976).

Ripeness assessment and FFB quality control in the oil palm industry is crucial for the evaluation of the processing results (Wood et al., 1984). In this respect perception of color is very important for indicating fruit maturity and defects. In nature, an oil palm tree continuously produces FFBs and the amount of oil in its fruits reaches a maximum, after which it declines due to the hydrolysis of fat and synthesis of free fatty acid (Southworth, 1976). Mature bunches detach fruit progressively and ultimately become rotten and moldy before breaking-up (Southworth, 1976). Therefore, oil palm FFBs need to be harvested at the optimum maturity. Traditionally, the ripeness level is defined in terms of the number of detached fruits from the bunch. Two different criteria for checking the number of detached fruits were introduced in Hitam and Yusof (2000). The former is the number of detached fruits on the ground before the FFB is cut, and the latter is the number of detached fruit sockets on the bunch. The former method is often used for tall trees, which has been used until today, while the latter is suitable only for short trees.

However, these methods are inaccurate, time consuming and laborious, which lead to higher harvesting and production costs.

* Corresponding author at: School of Engineering and Technology/Telecommunications, Asian Institute of Technology, P.O. Box 4, Klong Luang, Pathumthani 12120, Thailand.

E-mail address: attaphongset@ait.asia (A. Taparugssanagorn).

http://dx.doi.org/10.1016/j.compag.2015.09.018
0168-1699/

In addition, there is not really a common definition to specify which bunch is ripe or unripe (Jalil, 1994; Siregar, 1976; Southworth, 1976). Because of the problems of traditional methods, more consistent image processing techniques have been proposed for replacing these traditional methods. The correlation between the color of oil palm fruits and their oil content has been investigated showing that there is a positive correlation between both attributes, i.e., unripe fruit has the lowest oil content, ripe fruit has the highest oil content, and the oil content deteriorates when the fruit reached the overripe stage (Choong et al., 2006; Alfatni et al., 2006). Computer-based technologies and tools for machine vision, which can mimic human color recognition, have been introduced (Abdullah et al., 2012, 2004; Blasco et al., 2003; Ishak and Hudzari, 2010). However, both hardware components such as personal computers, color frame grabber, and charge-couple-device (CCD) cameras, and software for processing and control are required to provide a complete system, which might not be practical in an actual application. Patter recognition techniques using principal component analysis (PCA) has been proposed to identify different ripeness classes of oil palm FFB (Zhang and Wu, 2011). Three features represented by three RGB values are analyzed to obtain a plot of the principal components. The centroid values is then identified and used for indicating each ripeness class. Finally, the Euclidean distances between the centroid values and the plot of other samples are used to classify the oil palm FFB. This method yielded 75% correct classification for RGB images. A photogrammetric based approach, which correlates the color of the palm oil fruits to their ripeness, is presented in Ahmed et al. (2009). The methodology consists of both a hardware component in the form of an illumination chamber and a software component that calculates the color digital numbers (DN) and classifies the ripeness of the oil palm FFB. However, this method still has difficulty in differentiating the average RGB values for the ripe and unripe FFB since there is no clear-cut distinction of the RGB values for both cases. More sophisticated oil palm FFB grading system using neuro-fuzzy and fuzzy logic are developed in Jamil et al. (2009), May and Amaran (2011). Better accuracy, but more complicated techniques, like artificial neural network (ANN) classifier have been studied for various classification tasks of different agricultural products (Chinchuluun et al., 2009).

This paper proposes a simple non-destructive oil palm ripeness recognition using image processing techniques together with information theory. Images can be taken in a natural light environment. The differences of the distributions of a testing image of palm bunches and of standard scale images, which predefine each level of oil palm ripeness, are computed in terms of relative entropy, also called "Kullback–Leibler distance (KL distance)". The level of the standard scale image, which has the smallest distance from the one of the testing image, is used to decide about the ripeness level of the tested oil palm.

The rest of the paper is organized as follows. Section 2 describes the methodology of this work. All the process including the digital image pre-processing, the information-theoretical color similarity measures, the proposed technique, and the mobile application development are included in this section. Section 3 shows the experimental results. Practical issues are discussed in Section 4. The conclusions are given in Section 5.

2. Materials and methods

2.1. Palm oil grading

As previously stated, two parameters (OER and FFA) indicate the quality of fresh fruit FFB. They can be used to recognize the maturity or ripeness for grading and harvesting process of oil palm. In manual grading of oil palm, the color of FFB is the most important indicator for farmers to determine the ripeness of the oil palm fruit or FFB. Therefore, there should be a relationship between the OER with the color of FFB.

The relationship of Hue values, which is a term to describe the pure spectrum colors, of FFB images with the OER or the mesocarp oil content was developed in Razali et al. (2009, 2012). In these studies the FFB images captured are analyzed using their RGB values. Then, the RGB values are converted to the corresponding Hue digital values. In the next step the OER representing the oil content in the mesocarp parts is calculated. The whole dry weight model is used to calculate oil-to-dry mesocarp ratio (Razali et al., 2009, 2012). The Hue digital values is finally correlated to the mesocarp oil content using regression models as shown in Figs. 6 and 8 in Razali et al. (2009, 2012). The day of harvesting or a number of days before harvest of FFB was also predicted in Razali et al. (2009, 2012). The harvesting days are determined based on the 75% mesocarp oil content which indicated as a ripe stage for FFB (Razali et al., 2012). A linear interpolation technique is used to fix the date for the oil content of mesocarp reaching at 75%. At last, the harvesting days of FFB are determined by the percentage of mesocarp oil content using linear equations as shown in Fig. 14 in Razali et al.. (2012). As a result, the oil is found to start developing in mesocarp fruit at 65 days before fruit at ripe maturity stage of 75% oil-to-dry mesocarp.

In terms of palm oil quality, the FFA is an important parameter. Since fats and oils contain some level of free fatty acid, there is always an increase in acidity with time during transport and storage. The effect of light of different colors on the FFA values of stored crude palm oil was studied in Oyem (2011). In general, the effect of light on stored palm oil is that of increasing not only the rate of oxidation, but also that of hydrolysis since light is a source of energy. Therefore, it is suggested that palm oil samples for storage should be kept to inhibit the effects of light (Oyem, 2011). The change in FFB's color upon ripening due to biochemical reactions was observed through a visible and near-infrared (VIS/NIR) spectroscopy (Muhammad and Peeyush, 2014). In their study, a portable VIS/NIR spectrometer is employed to rapidly measure quality of oil palm FFB on-site, by means of non-contact and non-destructive approach. Two statistical analyses are performed to models FFB quality, i.e., the ripeness, the OER and also the FFA. The former method, i.e., a forward-stepwise method is employed to establish multiple linear regressions. The latter method is a combination between principal component analyses with multilayer perceptron neural network (Muhammad and Peeyush, 2014). In conclusion, the measured color spectral data can be statistically analyzed to predict not only the OER, but also the FFA. Therefore, there should be relationships between the FFA with the color of FFB.

2.2. Digital image pre-processing

All images in this paper are taken by a digital camera with similar specs as the ones found on any smart phone. A digital image is defined as a discrete two-dimensional function, $f(x, y)$, where x and y are the spatial (plane) co-ordinates. The amplitude f at any pair of co-ordinates (x, y) is called the intensity (color) or gray level of the image at that point. A digital image is usually composed of a finite number of elements, each of which has a particular location and values. These elements are called "picture elements" or "pixels". The intensity (color) of each pixel is variable. Typically, a color is represented by three or four component intensities of a color space, for instance, the RGB color space is defined by the red, green, and blue additive primaries and the CMYK color space is defined by the cyan, magenta, yellow, and black colors. In our work, the RGB color space is used since it is a common choice for computer graphics due to its simplicity.

Two image characteristics are enhanced before further processing. First, brightness and contrast, which are the most significant pixel characteristics, is taken into account since an image must have the proper brightness and contrast for easy viewing. Brightness refers to the overall lightness or darkness of the image while contrast is the difference in brightness between objects or regions. The brightness and the contrast of images are automatically adapted to optimum levels. A common technique is based on an existing two-scale decomposition of the image into an automatic image enhancement algorithm based on the bilateral filter and histogram equalization (Wu and Chuang, 2006). This technique has shown to be an effective high-contrast image enhancement due to the capacity of preserving details (Wu and Chuang, 2006). In addition, it is fully automatic and requires no parameter settings (Wu and Chuang, 2006).

Second, image deblurring (or restoration) is considered when the images are blurry, which was possibly caused by camera shaking, dirty or wet lens, wrong focusing, etc. To deblur an image, a blind deconvolution method is used since it works efficiently when no information about the distortion (blurring and noise) is available (Hall and Qiu, 2007). In this case the image is deconvolved using the maximum likelihood algorithm (Hall and Qiu, 2007). Figs. 1 and 2 illustrated the examples of the brightness adjustment and deblurring. Both techniques can automatically enhance the quality of images before further processing.

Fig. 1. Enhancing image by automatic brightness adjustment.

Prior to color analysis of the FFB, unwanted elements or background noise in the image has to be removed if the taken image consists of not only the region of interest, but also unwanted elements or background noise. Since the region of interest is the FFB image, background noise can be automatically extracted by using morphological operations (Géraud et al., 2001) and applying a mask over the FFB image. Fig. 3 shows an FFB image before and after background extraction. However, this part of the process can be an additional option and can be avoided by requesting users to take photos of the part of FFB only, not any background part.

In addition, an FFB image, which usually consists of spikes and fruits, can be improved extracting the color of fruits only for ripeness determination. Therefore, both spikes and fruits pixels can be separated using the k-means clustering algorithm (Kanungo et al., 2002). An example of segmented image is shown in Fig. 4. This part of the process can be an additional option and can be avoided since there is very little impact of the spikes on the proposed color similarity measures, which will be explained in the next section.

2.3. Information-theoretical color similarity measures

Similarity measurement is the most significant issue in this methodology. To measure the similarity between the testing image and the standard scale images stored in a database, the difference between the testing color vector and the database color vectors can be used and is calculated using distance metrics. A small difference between two color vectors indicates a large similarity and a small distance. The image from the standard scale with the minimum distance is considered as the most similar to the testing image.

Several distance formulas for measuring the similarity are defined. In general, the techniques for comparing probability distributions, such as the Kolmogorov–Smirnov test are not appropriate for color distributions because visual perception determines similarity rather than closeness of the probability distributions (Young, 1977). The color distance formulas arrive at a measure of similarity between images based on the perception of color content.

The sum of absolute difference $d_{\text{abs}}(h,g)$ is a very straightforward distance metric and extensively used for checking the color image similarity, it is calculated (Selvarajah and Kodituwakku, 2011) as

$$d_{\text{abs}}(h,g) = \sum_{A}\sum_{B}\sum_{C}(|h(a,b,c)| - |g(a,b,c)|), \tag{1}$$

where $h(a,b,c)$ and $g(a,b,c)$ represent two color histograms, A, B and C represent the three color channels (R, G, B is used here). Computationally, the color histogram is formed by discretizing the colors within an image and counting the number of pixels of each color. Each color histogram is also quantized into L bins. The distance metric is a simple method to search for similar images in the database to the testing image automatically, but it is sensitive to background variations such as illumination and direction of the light.

Histogram Euclidean distance $d_E(h,g)$ is most commonly used for similarity measurement in image retrieval because of its efficiency and effectiveness. It measures the distance between two vectors of images by calculating the square root of the sum of the squared absolute differences and it can be calculated (Szabolcs, 2008) as

$$d_E(h,g) = \sqrt{\sum\nolimits_{A}\sum\nolimits_{B}\sum\nolimits_{C}(h(a,b,c) - g(a,b,c))^2}. \tag{2}$$

In this distance formula, there is only comparison between the identical bins in the respective histograms. Two different bins may represent perceptually similar colors but are not compared crosswise. All bins contribute equally to the distance.

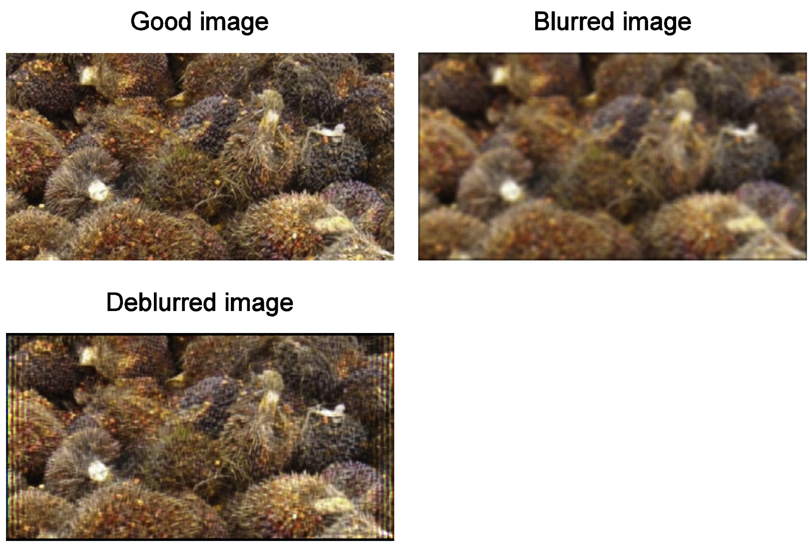

Fig. 2. Deburring image using blind deconvolution.

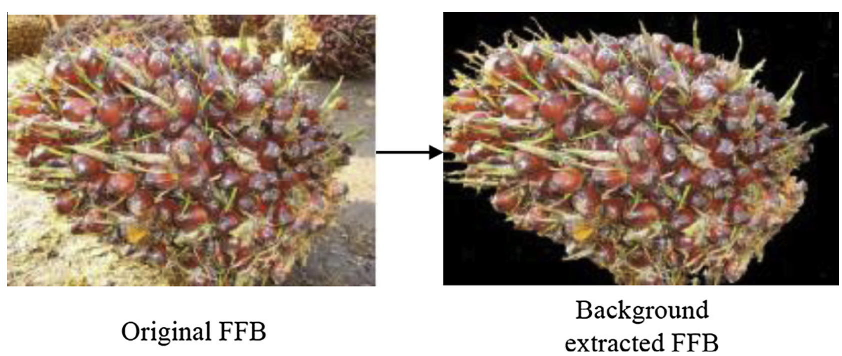

Fig. 3. Background noise extraction using morphological operations.

To overcome the limitations of $d_{\text{abs}}(h,g)$ and $d_E(h,g)$, the relative entropy or also called "Kullback–Leibler distance (KL distance)," which is a non-symmetric measure of the difference between two probability distributions P and Q, is proposed to measure the similarity between a testing image and the standard scale images. The entropy itself measures the randomness of the distribution of not only color levels in bins, but also the textural parameters of the images. The greater the distribution among color levels in an image is, the higher its entropy is. Specifically, the Kullback–Leibler distance of Q from P, denoted $D_{\text{KL}}(P||Q)$, is a measure of the information lost when Q is used to approximate P and is defined as

$$\begin{aligned} D(p||q) &= -E\left[\log_2 \frac{p\{a,b,c\}}{q\{a,b,c\}}\right], \\ &= -\sum_A \sum_B \sum_C p\{a,b,c\} \log_2 \frac{p\{a,b,c\}}{q\{a,b,c\}}, \end{aligned} \tag{3}$$

where $p\{a,b,c\}$ and $q\{a,b,c\}$ denote the joint probability mass functions (PMF) of the intensities of a testing image and one of the standard scale images, respectively. They are referred to their color image histograms as

$$p\{a,b,c\} = h(a,b,c)/N, \tag{4}$$

and

$$q\{a,b,c\} = g(a,b,c)/N, \tag{5}$$

where $h(a,b,c)$ and $g(a,b,c)$ represent two color histograms of a testing image and one of the standard scale images, respectively, and N is the number of pixels in the image.

2.4. Proposed oil palm classification technique

The procedure of the proposed oil palm classification technique is shown in Fig. 5 and is described as follows:

1. Take photos of three levels of the standard color scales of the Nigrescens-type oil palm FFB. The Nigrescens fruits are naturally dark purple young, brownish medium ripe, and red-orange ripe. These correspond to three ripeness levels: completely ripe, medium ripe, and unripe as depicted in Fig. 6. That is, when an oil palm bunch is red-orange in color like the first row, it is supposed to be already developed 100%, completely ripe. And when an oil palm bunch is brownish like the middle row, it is supposed to be already developed 80–90%, medium ripe. Finally, when an oil palm bunch is dark purple in color like the last row, it is supposed to be developed just 60%, unripe.
2. Calculate the color histograms and the joint PMFs of the three standard color scales and store them in the database.
3. Take the photo of a test oil palm bunch.
4. Repeat step 2 for the image in step 3.
5. Calculate the Kullback–Leibler distances, as in Eq. (3), of the joint PMFs in step 2 from the joint PMF in step4.

Fig. 4. Spikes extraction using the k-means clustering algorithm.

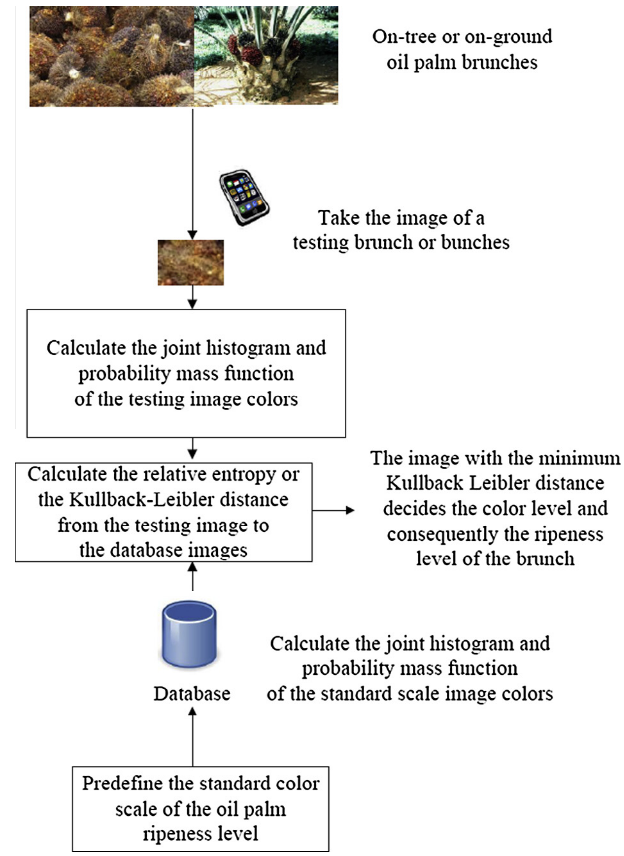

Fig. 5. Block diagram of the proposed procedure.

6. Find the minimum Kullback–Leibler distance from the distances computed in step 5. This means that the corresponding color level of the test oil palm is the one that matches the standard scale with the minimum Kullback–Leibler distance.

2.5. Mobile application development

After the Matlab-implementation, the algorithms were rewritten in Java. To generate the test-results in Matlab at a reasonable speed, the code was optimized as much as possible. This was mainly accomplished by trying to avoid for-loops and replacing them by matrix-based operations. By doing this, the code becomes more efficient. After this code optimization the mobile application on the Android platform was developed. Java is the recommended programming language for Android applications. Besides the inherits basic classes from conventional Java, such as string, container, Math, and network, new classes specific to mobile devices, for instance, camera, telephony, map, location-based services, speech, were also added.

The algorithm was implemented in Android using Eclipse and the Android SDK including an Android Emulator. Our application is called "Ripe or Raw Oil Palm Grader." It is able to take pictures, calculate their feature vectors, i.e., the color histograms, the joint PMFs, and the entropy. It is able to find a match in a database and to display the corresponding results. Each of these tasks were detailed in the previous subsections. In order to find the matches, a database is needed to be built using XML files. This is to improve modifiability and user-friendliness. However, we do not work with the XML files themselves in our application, we only import them into Android's built-in SQLite database system. The database must be able to provide all the data needed to make our application work. Therefore, sufficient tables need to be created with their data fields and the tables are linked in such a way that it can easily provide related data.

The main menu is a basic layout. All parts of the layout are included in the main.xml file. It has five buttons, i.e., creating new activity, continuing the latest activity, loading recent activity, adding standard pictures, and a description about the mobile application, as shown in Fig. 7. These elements are ordered in a Table Layout.

The features of the mobile application are as follows:

1. As shown in Fig. 8, the mobile application can be used for different oil palm breeds. As a default, two different breeds, namely, Nigrescens fruits, which are naturally dark purple young, and red-orange ripe, and Virescens fruits, which are naturally green young, and orange ripe, are provided. An arbitrary oil palm breed can be created on sites selecting "Add Standard Pictures".
2. It can tell different level of ripeness of FFB. As a default, three different levels are provided.
3. It can be used for camera or gallery modes as depicted in Fig. 8.
4. It can be used for any image size.
5. It is assumed that users will take photos of the part of FFB only, not any background part. This action avoids having an additional process for filtering out the background part as explained in Section 2.2. Otherwise, the mode with the additional background filtering is needed causing a larger computation time.

Fig. 6. Three standard scale color levels corresponding to three ripeness levels of oil palm bunches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. It can display in both text and image as shown in Fig. 9, as well as voice formats. Therefore, it is very useful in a real scenario, where one does not want to see the display and just want to hear the result.
7. It provides Thai and English languages.

After finishing an activity, e.g., the pile of all FFBs from a pickup truck, the percentage over the entire pile is calculated for each level of ripeness, i.e., completely ripe, medium ripe, and unripe.

3. Experimental results and evaluations

The proposed approach was tested with one hundred images of the Nigrescens-type oil palm FFB. The built database consists of three different color levels corresponding to three different ripeness levels of oil palm bunches as explained earlier. All the images are in the RGB color space. Although the RGB color space is a very simple to process, it is far from being perceptually uniform. To obtain a good color representation of the image by uniformly sampling the RGB space the quantization step sizes (equivalently the number of histogram bins) must be properly selected to be fine enough such that distinct colors are not assigned to the same bin. At the same time an oversampling can produce a larger set of colors than may be needed. The increase in the number of bins in the histogram impacts the performance of the color similarity measurement.

First, the average KL distances of the joint PMFs of each standard level from the joint PMF of each test sample, using a histogram with 8 bins, from ten out of one hundred samples are shown in Table 1. The level giving the minimum KL distance from

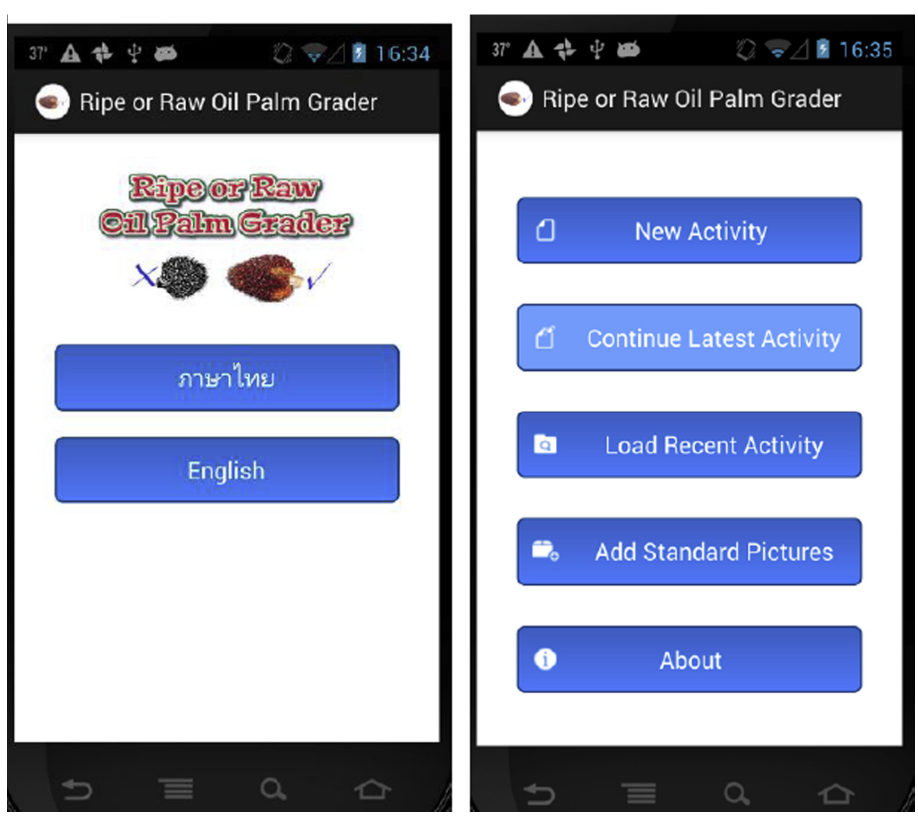

Fig. 7. Front page and main menu of the graphical user interface (GUI) of the mobile application.

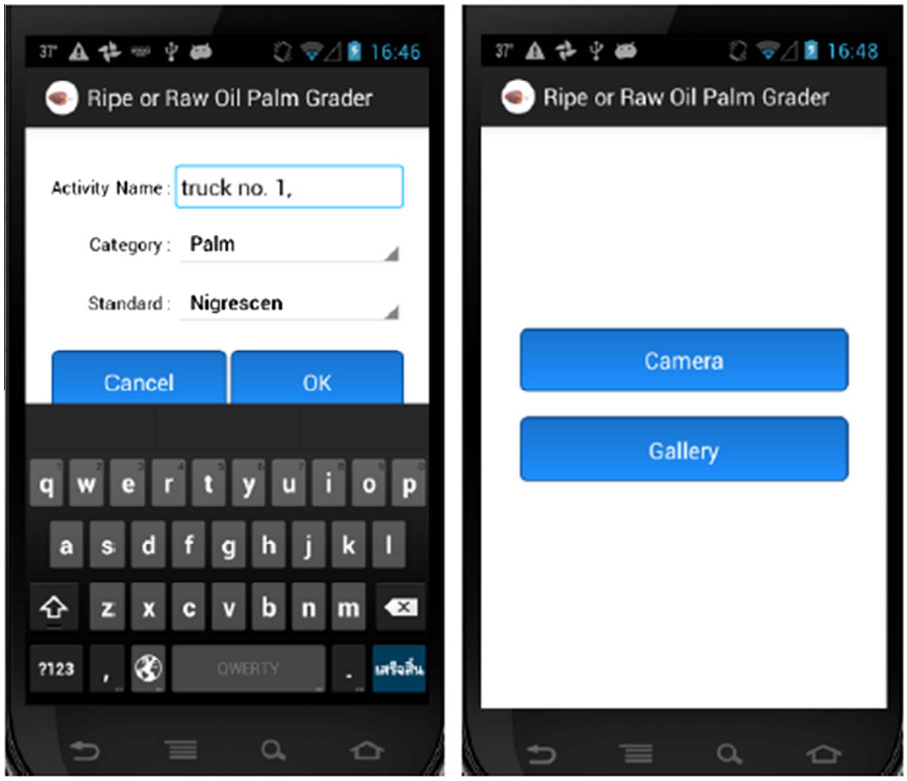

Fig. 8. Creating new activity.

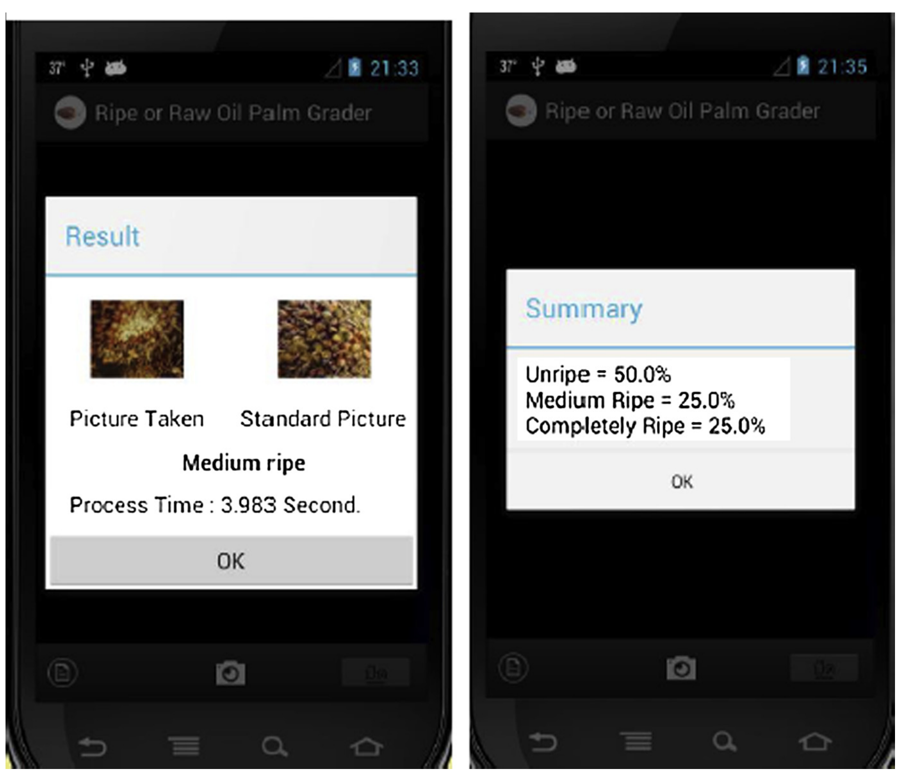

Fig. 9. Result screen of the mobile application.

the standard level set represents the color level of the test image. As an example, three oil palm image samples, i.e., sample 1, 5, and 7 according to Table 1, are shown in Fig. 10. As it can be seen, the system works well even for different image sizes. To statistically evaluate the proposed overall system, the effectiveness of the oil palm classification is based on the performance of the color

Table 1
The average KL distances of the joint PMFs of each standard level from the joint PMF of each test sample using a histogram with 8 bins.

Sample	Average KL distance from level 1 (completely ripe) to sample	Average KL distance from level 2 (medium ripe) to sample	Average KL distance from level 3 (unripe) to sample
1	0.8272	0.2438 (min)	0.3584
2	1.1290	0.6467 (min)	0.6487
3	0.3503	0.1499 (min)	0.2885
4	0.4711 (min)	0.7778	0.5815
5	0.6420	0.7223	0.3372 (min)
6	0.4055 (min)	0.5409	0.5135
7	0.5800 (min)	0.7724	0.6757
8	0.9307	0.6723	0.4475 (min)
9	0.9972	0.5043	0.4271 (min)
10	0.6524	0.5085 (min)	0.6952

Fig. 10. Three testing sample 1 (top), 5 (bottom left), and 7 (bottom right) at histogram bin of 8.

similarity measurement. In this section we use the classification accuracy (or accuracy rate), which is calculated as

$$\text{Accuracy rate} = \frac{\text{Number of correct samples}}{\text{Number of total samples}}. \tag{6}$$

Based on the visual inspection of twenty five oil palm experts, the performance is calculated for quantization of the histograms into 4, 8, 16 and 32 bins separately and all results are averaged. For each quantization of histograms, the KL distance metric is used separately in the experiment to get the results in terms of the classification accuracy or the accuracy rate. Consequently, each KL distance metric is tested for all quantization bins and the results are analyzed for all distance metrics against all quantization bins and shown in Table 2. Moreover, the results are compared with the ones using Euclidean distance which is widely used in the literatures, for instance, in Alfatni et al. (2006, 2012, 2004), Zhang and Wu (2011), Ahmed et al. (2009). We compare our results with the ones from the fuzzy logic system in May and Amaran (2011), which are also evaluated against human graders for accuracy. Their fuzzy logic system achieved 86.67% accuracy, which is outperformed by our proposed system that achieves 96% accuracy with much less complexity. In addition, the average computation time over one hundred testing samples for all quantization bins is also

Table 2
The accuracy rate and the average computation time over one hundred test samples with different quantization bins, 4, 8, 16, and 32, and their average over all quantization bins using both histogram Euclidean distance and KL distance.

	4 Bins	8 Bins	16 Bins	32 Bins	Average
KL distance: accuracy rate (%)	94	96	97	97	96
Histogram Euclidean distance: accuracy rate (%)	82	85	86	86	84.7
KL distance: average computation time (s)	3.21	3.50	3.64	3.71	3.51
Euclidean distance: average computation time (s)	3.00	3.12	3.24	3.29	3.16

shown in the same table. The average performance over all quantization bins is 96% accurate, which is relatively very high. The results from the mobile application are the same as the ones using the Matlab code. The average computation time over all quantization bins is about 3.5 s with the Matlab code. In the same table, the results by using the histogram Euclidean distance are also shown. We can see that the KL distance performs the similarity measure better whereas the average computation time is just slightly longer.

Regarding the mesocarp oil content, the regression models shown in Figs. 6 and 8 in Razali et al. (2009, 2012) are applied to our samples and the results are shown in Table 3. With the same ten samples given in Table 1, their Hue digital values are first calculated from their RGB values. After that, their mesocarp oil contents are estimated.

Finally, the performance with the spikes extraction using the k-means clustering algorithm as described in Section 2.2 is also evaluated. We can see that it does not matter with or without spikes in FFB images, this is because of the fact that the three standard FFB images have always the same conditions, i.e., with or without spikes as the testing FFB images, whereas with spikes extraction process only increases the computation time by about 3 s.

4. Discussion for actual practices

Two potential activities in which the proposed system can be applied are the oil palm grading activity at a ramp site and the harvesting activity in the field. The former begins with the inspection or grading of the FFBs before loading them into the ramp and transferring them using cages into a sterilizer line. Then a crane is used to lift the cages and pour the fruit bunches into the hopper to separate the fruit equally into the conveyor. The grading is the most significant step to determine the throughput of the operation. The proposed system which replaces human's visual inspection can improve the throughput of this operation.

Two possible applications of the proposed system for the oil palm grading at a ramp site are illustrated in Fig. 11a and b. A

Table 3
The Hue digital values and the mesocarp oil content of each test sample based on the regression models in Razali et al. (2009, 2012).

Sample	Ripeness result from KL distance	Hue digital value (0–255)	Average mesocarp oil content (%)
1	Medium ripe	190	70
2	Medium ripe	210	74
3	Medium ripe	185	69
4	Completely ripe	240	75
5	Unripe	162	20
6	Completely ripe	230	79
7	Completely ripe	233	78
8	Unripe	151	11
9	Unripe	155	14
10	Medium ripe	201	71

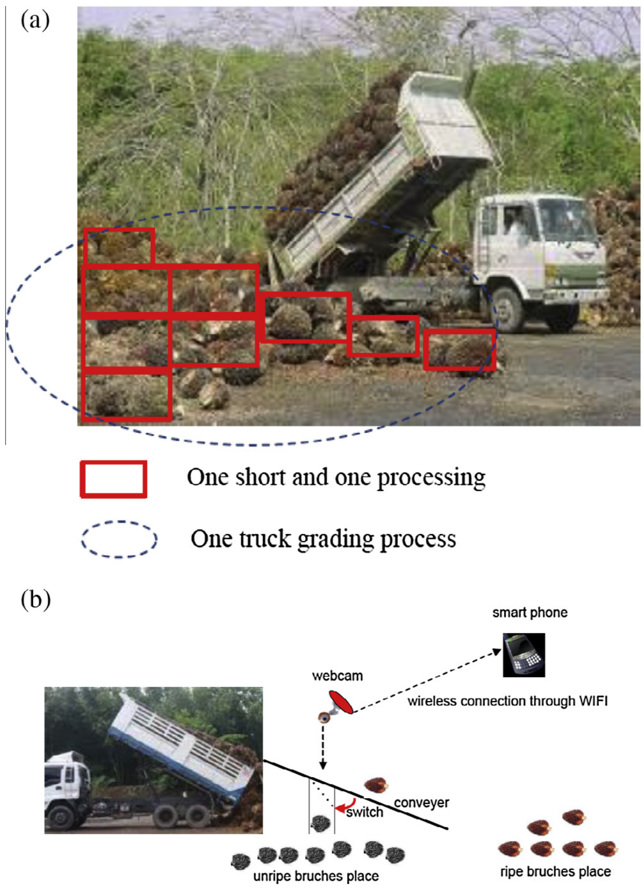

Fig. 11. Possible applications of the proposed system for the oil palm grading at a ramp site.

direct grading application of the proposed system using a smart phone is shown in Fig. 11a). The image of each subgroup of the entire FFB pile is taken and processed to give as a result, completely ripe, medium ripe, or unripe. As a final result, the percentage over the entire pile is calculated for each level of ripeness. Another idea of grading applications at a ramp site is shown in Fig. 11b. The automatic conveyer is attached to the pickup truck. The FFBs are transferred to the conveyer. Their images are taken by the webcam (or more webcams in the case of several lines on the conveyer), and processed by the smartphone. Once the result is unripe, the smartphone sends the signal to control ON/OFF switch to open the hole to separate the unripe bunches to different area from the area of the ripe bunches.

The latter is to apply the proposed system for the harvesting activity. A so called "smart cutter" can be developed attaching the webcam next to the blade and the processor and display (smart phone) at the eye level as depicted in Fig. 12. With this idea, we can efficiently reduce the waste of cutting unripe bunches.

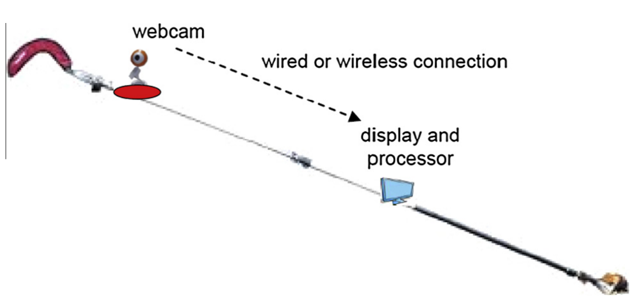

Fig. 12. "Smart cutter", a possible application of the proposed system for the FFB harvesting at a field.

5. Conclusion

We proposed an information theory based image processing approach for a simple non-destructive oil palm ripeness recognition system which enables the replacement of human's visual inspections for oil palm grading activity and for determining the correct time for harvesting activity. The results showed that the proposed system has an excellent accuracy and can be processed very fast. The overall system is simple and useful for improving the benefits of oil palm farmers and entrepreneurs.

The Android operating system is very flexible and provides many tools for developing applications. This approach has allowed us to develop a "ripe or raw oil palm grader" application in a short amount of time. A wide range of Android's possibilities were used. The algorithm in Matlab was optimized to reduce the computation time and translated into Java code. The same results are obtained as in the Matlab code.

Acknowledgement

The authors would like to gratefully and sincerely thank Mr. John Clendon, Dr. Palat Tittinutchanon, Mr. Praiwan Tohdam, and their staffs from the Univanich Palm Oil Public Company Ltd. Thailand for their kindness in giving some suggestions and helping us evaluate the system performance.

References

Abdullah, M.Z., Guan, L.C., Lim, K.C., Karim, A.A., 2004. The applications of computer vision system and tomographic radar imaging for assessing physical properties of food. J. Food Eng. 61, 125–135.
Abdullah, M., Amiruddin, M.D., Alfatni, M.S.M., Marhaban, M.H., Mohamed, S.A.R., Shafie, S., 2012. Fruit Ripeness Grading System, WO 2012039597 A2.
Ahmed, J., Roseleena, J., Nursuriati, J., Cheng, Y.L., Bulan, A., 2009. Photogrammetric grading of oil palm fresh fruit bunches. Int. J. Mech. Mechatron. Eng. (IJMME-IJENS) 9 (10), 7–13.
Alfatni, M.S.M., Shariff, A.R.M., Shafri, H.Z.M., Saaed, O.M.B., Eshanta, O.M., 2006. Oil palm fruit bunch grading system using red, green and blue digital number. J. Appl. Sci., 1444–1452
Blasco, J., Aleixos, N., Molto, E., 2003. Machine vision system for automatic quality grading of fruit. Biosyst. Eng. 85 (4), 415–423.
Chinchuluun, R., Lee, W.S., Bhorania, J., Pardalos, P.M., 2009. Clustering and classification algorithms in food and agricultural applications: a survey. Advances in Modeling Agricultural Systems, vol. 25. Springer, Yardley, PA, USA, pp. 433–454.
Choong, T.S.Y., Abbas, S., Shariff, A.R., Halim, R., Ismail, M.H.S., Yunus, R., Ali, S., Ahmadun, F.R., 2006. Digital image processing of palm oil fruits. Int. J. Food Eng., 838–843
Fitzherbert, E.B., Struebig, M.J., Morel, A., Danielsen, F., Brűhl, C.A., Donald, P.F., Phalan, B., 2008. How will oil palm expansion affect biodiversity? Trends Ecol. Evol. 23 (10), 539–545.
Géraud, T., Strub, P., Darbon, J., 2001. Color image segmentation based on automatic morphological clustering. In: IEEE International Conference on Image Processing, pp. 70–73.
Hall, P., Qiu, P., 2007. Blind deconvolution and deblurring in image analysis. Statistica Sinica 17, 1483–1509.
Hitam, A.H., Yusof, A.M., 2000. Mechanization in oil palm plantations. Advances in Oil Palm Research, vol. 1. Malaysian Palm Oil Board, Ministry of Primary Industries, Malaysia, pp. 653–696.
Ishak, W.I.W., Hudzari, R.M., 2010. Image based modeling for oil palm fruit maturity prediction. J. Food Agric. Environ. 8 (2), 469–476.
Jalil, A.M., 1994. Grading of FFB for palm oil mills in Malaysia. In: The National Palm Oil Milling and Refining Technology Conference, pp. 75–95.
Jamil, N., Mohamed, A., Abdullah, S., 2009. Automated grading of palm oil Fresh Fruit Bunches (FFB) using neuro-fuzzy technique. In: The international conference of Soft Computing and Pattern Recognition (SOCPAR), pp. 245–249.
Kanungo, T., Mount, D.M., Netanyahu, N., Piatko, C., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithm: analysis and implementation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 881–892.
May, Z., Amaran, M.H., 2011. Automated oil palm fruit grading system using artificial intelligence. Int. J. Eng. Sci. 11, 30–35.

Muhammad, M., Peeyush, S., 2014. In situ quality assessment of intact oil palm fresh fruit bunches using rapid portable non-contact and non-destructive approach. J. Food Eng. 120, 248–259.
Ng, S.K., 1957. The Oil Palm: Its Culture, Manuring and Utilisation. International Potash Institute, Berne, Switzerland.
Oyem, H.H., 2011. Monitoring the free fatty acid level of crude palm oil stored under light of different wavelenghts. Am. J. Food Technol. 6, 701–704.
Razali, M.H., Wan, I.W.I., Ramli, A.R., Sulaiman, M.N., Harun, M.H., 2009. Development of image based modeling for determination of oil content and days estimation for harvesting of fresh fruit bunches. Int. J. Food Eng. 5 (2), 1633–1637.
Razali, M.H., Ssomad, M.A.H.A., Syazili, R., Fauzan, M.Z.M., 2012. Simulation and modeling application in agricultural mechanization. Modell. Simul. Eng., 8 pages
Selvarajah, S., Kodituwakku, S.R., 2011. Analysis and comparison of texture features for content based image retrieval. Int. J. Latest Trends Comput. 2 (1), 108–113.
Siregar, I.M., 1976. Assessment of ripeness and crop quality control. In: The Malaysian International Agricultural Oil Palm Conference, pp. 740–754.
Siregar, I.M., 1976. Assessment of ripeness and crop quality control. In: The Malaysian International Agricultural Oil Palm Conference, pp. 740–754.
Southworth, A., 1976. Harvesting. In: Oil Palm Research. Elsevier, Amsterdam.
Southworth, A., 1976. Harvesting – a practical approach to the optimization of oil quantity and quality. In: The Malaysian International Agricultural Oil Palm Conference, pp. 755–768.
Szabolcs, S., 2008. Color histogram features based image classification in content-based image retrieval systems. In: 6th International Symposium on Applied Machine Intelligence and Informatics, pp. 221–224.
Wood, B.J., Ismail, S., Guan, L.S., Chyng, C.H., 1984. A preliminary report on a long-term study of the effect of oil palm harvesting strategy on product recovery, including a comparison before and after weevil pollination. In: The Symposium on Impact of the Pollinating Weevil on the Malaysian Oil Palm I, pp. 261–283.
Wu, J.-L., Chuang, C.-Y., 2006. An efficient method for enhancing high-contrast digital photos automatically. Int. MultiConf. Eng. Comput. Sci., 502–506
Young, I.T., 1977. Proof without prejudice: use of the Kolmogorov–Smirnov test for the analysis of histograms from flow systems and other sources. J. Histochem. Cytochem. 25 (7), 935–941.
Zhang, Y., Wu, L., 2011. Crop classification by forward neural network with adaptive chaotic particle swarm optimization. Sensors 11, 4721–4743.