Oil palm fruit ripeness detection using K-Nearest neighbour

I F Astuti, F D Nuryanto, P P Widagdo, D Cahyadi

Computer Science Department, Faculty of Computer Science and Information Technology, Mulawarman University, Gunung Kelua Campus, Samarinda, Indonesia

*Corresponding Author: indahfitriastuti@gmail.com

Abstract. Palm oil plant (Elaeis guineensis jacq) is one of the most popular plants in Indonesian plantation and also a species of the palm family. Harvesting process should obtain fresh fruit bunches at optimal ripeness. Some farmers are lack of ripeness knowledge and do not understand which color that represents optimal ripeness of palm fruit to harvest. Another issue on the scarcity of such system provides ripeness identification also encourage a study to develop an image-processing-based expert system. The aim of this study is to provide a tool for examining oil palm fruit ripeness level. Identification system consists of several image processing phases, namely image edge detection and feature value of the image to be calculated such as mean value, standard deviation, skewness, entropy, energy, homogeneity, contrast and intensity. In the proposed work, K-Nearest Neighbor (KNN) is applied as a computation method which allows the system to classify data of ripeness level based on user-entered image. The output is the identification of the ripeness level of the oil palm fruit.

1. Introduction

The amount and quality of oil produced by oil palm depend on various factors. One of the most important factors is fruit ripeness when harvested and its handling to the Palm Oil Mill. The harvest should produce fresh fruit bunches at optimal ripeness because harvesting on raw fruit slices would result in a reduced amount of oil production. Moreover, overripe, slow handling or rotten fruit will produce oils with a high free fatty acid. Identification of oil palm fruit ripeness level is conducted by looking directly at the color of the fruit, but this can only be done by experts, so it is necessary to build a tool to transfer expert knowledge into a machine, called expert system. The expert system is a computer-based system that uses expert knowledge, facts, and reasoning to solve problems. The development of identification systems must be followed by the learning process in the recognition of color patterns by computer systems.

Color pattern recognition is defined as raw data retrieval so that the system could act based on previous data. Classification of the mutually class image plays a significant role in engineering and other computer vision application. Some important fields where these types of classification technique are widely used include image processing in medical, robotics based on classification, pattern recognition. In KNN classification, the output is the class membership. The classification of the object is conducted by the most common of its K nearest neighbors.

K-nearest neighbor (KNN) classification is conventional non-parametric classifier, which is widely used as the basic classifier in pattern classification cases. This classification is based on metric of the
distances between the test image and each of the training images to decide the final classification output.

The phases in KNN classification of patterns are: (a) First, put the train data features and labels obtained in a two dimensional space, (b) Clustering or grouping of the train data features are done, (c) The test data is then located on to the same space, (d) For obtaining the nearest neighbor related to test data, calculate the distance of train data using Euclidean distance. The distance calculation is given by

\[dE(x,y) = \sqrt{(x_i - y_i)^2},\]

Choose the closest distance vector with respect to both test and train data, (f) Find the closest distance by choosing or averaging the majority of data points. Choosing is the method of arranging the data in ascending order (sorting the data points). The arranged data is then compared with labeled trained data to return the related class. Consequently, the classified result is obtained. The mean is determined by calculating the arithmetic mean. This KNN searching method is for the most common images which represent the right level ripeness.

Based on this background, an expert system is developed to identify the ripeness level of oil palm fruit using expert knowledge, image processing, and the K-Nearest Neighbor classification method. With the K-Nearest Neighbor classification method, each motive image has a feature value that will compare its value according to the smallest difference of each train and test data.

2. Development details

A set of training images are involved and then combined in a single matrix. The correlation is calculated for these matrices and is sent to a K-Nearest Neighbor Classifier to collect the classification of the test image.

This system is built to calculate only the value of the image feature from the color of the palm oil varieties in the wide cultivated varieties of Tenera in East Kalimantan. Trained images and test images used are the images of the palm fruit using a predetermined size (i.e., 300x300 pixels). The value of defined features is mean, standard deviation, skewness, brightness, hue, saturation, entropy, and energy.

Those images are all in digital form. Digital imaging contains a number of basic elements. These basic elements are manipulated in image processing and further exploited in pattern recognition. The basic elements are brightness, contrast, contour, color, and shape.

Development of this expert system is following these steps: (1) Identify requirements and problems, (2) Determining suitable problems, (3) Consider alternatives, (4) Calculates the return on investment, (5) Choose a tool to develop, (6) Engineering knowledge, (7) Designing the system, (8) Finalize development.

Developing an overall system needs some steps according to the system development life cycle. There are six stages namely: planning stage (system engineering and analysis), analysis stage, design stage, application stage (implementation), evaluation stage, and maintenance stage.

The planning stage is understanding the problem arises in the developing system so that it is able to determine the aim of building the system and not difficult to recognize constraints such as data collection of oil palm fruits, application of methods and coding. Data analysis of oil palm fruit and KNN method by discussing more intensively, objectives and functions carried out by the system so that the system is able to detect colors accurately. Design stage determines what configurations are needed by the system and the methods used in the calculation. Designing that includes the process, rules used, user interface, menu arrangement and required input such as flowchart diagram design and form design on user-friendly systems. Application stage is an activity to implement the system by translating the system into a computer or a program. Edge detection will be carried out on the oil palm fruit color image and the KNN method is applied to identify the image of the oil palm fruit. Evaluation Stage is testing the program so that if an error occurs, it can be repaired. This trial process is needed to determine whether the system that has been made is correct and in accordance with the characteristics that have been set. In addition, the testing process was carried out to see the accuracy of color
detection of oil palm fruit. Maintenance stage is a trial of the system that has been made, the system maintenance stage is needed to overcome the possibilities that will occur such as correcting errors, maintaining the latest and improving the system.

Expert system of palm fruit ripeness level identification is the process of image resize, in which image will be adjusted in accordance with the specified size, which is 300x300 pixels. Image resize algorithm as follows:

1. Insert the image of the oil palm fruit
2. Set the width and height of the image to be used to 300x300 pixels
3. If the image has not been inserted, re-request the image to be inserted
4. If the image has been inserted, apply the width and height to the image
5. Apply the bicubic interpolation function
6. Save the resized result image

Descriptive statistics are utilized in this image processing, pixels are processed in the image. Some of the values used:

- Mean, a group explanation technique based on the average value of the group, expressed by a formula
- Skewness is the degree of asymmetry of a distribution. If the frequency curve of a distribution has a more elongated tail to the right (seen from the mean) then it is said to be right (positive) and otherwise left-handed (negative)
- Entropy is used to measure the randomness of the intensity distribution, by the formula
- while Energy states the size of the couple concentration with a certain gray intensity on the pixel matrix in the image, calculated by the formula

3. Result and discussion

The testing form in this system has nine features: Edge Detection, Clear All Database, Feature Extraction, Save Database, Save Image, Feature Value Process, Load Image, Image Resize, and Exit. The Edge Detection button is used for conversion of palm fruit image to edge line mode. The Feature Extraction button is used to calculate the feature value of the palm fruit image entered by the user. The Feature Process Feature button is used to calculate the Euclidian value between the image being tested with the image in the database. The Testing form is the main form of this application, used for testing the ripeness level of oil palm fruit. The Sobel edge detection is represented by Figure 1.
In this system, 80 data images of palm oil are trained with 7 levels of ripeness. Testing is also done by involving 7 trained data with each motive has 1 test data that has never been trained previously in the system. An example of one detection resulted in these values; mean 30.8318333, skewness 6148.530057, intensity 2774865, energy 0.2288966, deviation standard 44.585062474, entropy 2819.322, homogeneity 308.9574563, and contrast 156523835.

The flow of the detection system is illustrated in Figure 2, where the image of the palm fruit is inserted first and then checked for size. Sobel edge detection is then applied and feature values are calculated, such as middle value, standard deviation, skewness, entropy, energy, intensity, contrast, and homogeneity.
Figure 2. Flowchart of Oil Palm Fruit Ripeness Detection

The image will be categorized into one type, namely test image or training image. For the test image, the Euclidian value and ripeness level will be calculated based on the selection of the nearest neighbor from K, while the training image will be stored in the database for the training image.

The feature values for the training data of the oil palm fruit are shown in table 1.

Table 1. Features Value for Trained Data
| Features   | Value            |
|------------|------------------|
| Mean       | 22.1946222       |
| Deviation Standard | 32.8302788468147 |
| Skewness   | 4328.93077       |
| Entropy    | -1506.651        |
| Intensity  | 1997516          |
| Homogeneity| 233.158769167521 |
| Energy     | 0.1793772        |
| Contrast   | 109365262        |

Combo box displays all ripeness levels in the database, namely Very Raw F00, Raw F0, F1 Underdeveloped, F2 ripe, F3 ripe, F4 Overripe and F5 Overripe. In this example, the F2 ripeness level is used, as shown in Figure 3. Users should choose the ripeness level to be trained for the oil palm fruit.

Figure 3. List of Ripeness Level

Figure 4 displays an example of ripeness level detection from an image of oil palm fruit inserted. In this calculation, system has a result of F2 level, which is ripe enough.
4. Conclusion
As the research has demonstrated, it shows that using K-Nearest Neighbour method to a palm fruit ripeness detection could be one solution to help farmers harvesting at optimum ripeness. The test conducted in this study obtains effectiveness and accuracy at 65% by applying Sobel edge detection.

Acknowledgment
We would like to express our appreciation to Mr. Sugiyanto his patient guidance and valuable support for this research work.

References
[1] Patidar B C, Shah M R, Mishra 2014 Performance Analysis of K Nearest Neighbor Image Classifier with Different Wavelet Features, *Int Conf on Green Computing Communication and Electrical Engineering* (ICGCCEE)
[2] Li-Yu H, Min-Wei H, Shih-Wen K, and Chih-Fong S 2016 The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets *SpringerPlus* 5 p 1304.
[3] Tomaszewski J E, Hipp J, and Tangrea M A 2014 Digital Imaging Fundamentals, *Pathobiology of Human Disease* 6
[4] Gray G L, Chiu V, Liu Q, and Li P 2014 The Expert System life cycle in AIS Research: What does it mean for future AIS research?, *International Journal of Accounting Information Systems*.
[5] Bhumika G, Pushkar S, and Ankus M 2016 K-Nearest Correlated Neighbour Classification for Indian Sign Language Gesture Recognition using Feature Fusion, *International Conference on Computer Communication and Informatics*.
[6] Ajin M and Mredhula L 2017 Diagnosis of Interstitial Lung Disease By Pattern Classification, *7th International Conference on Advances in Computing & Communications*. 

Figure 4. An Example of Detection Result (F2 = ripe enough)