The Classification Of Monster And Williams Pear Varieties Using K-Means Clustering And K-Nearest Neighbor (KNN) Algorithm

Indarti1, Novita Indriyani2, Arief Setya Budi3, Dewi Laraswati4, Wina Yusnaeni5, Arief Hidayat6

1, 3, 4, 5 Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia
2 Program Studi Sistem Informasi Akuntansi Kampus Kota Bogor, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia
6 Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Indonesia

E-mail: indarti.ini@bsi.ac.id

Abstract. Pear is a kind of fruits which has a lot of varieties. One of the way to differ pear varieties is by looking at the color, size and shape. This research is aimed at giving assistance to classify two varieties of pear i.e. Monster Pear and Williams Pear. In order to get the purpose, pear image processing is done to ease the classification process of pear varieties. Research method used consists of RGB Color Room to l*a*b, image segmentation, characteristics extraction with K-Means Clustering. Besides, K-Nearest Neighbor (KNN) is used to know the distribution and the classification. The use of practice data will increase the accuracy of pear varieties classification. The data used here are 88 pear images which cover 44 image practice data of Monster pear and 44 image practice data of Williams pear. Meanwhile, the test data taken are 10 images of each variety. The result of this research shows that the pear classification accuracy level is 95% which is very good.

1. Introduction

Image is another term for picture as a multimedia component which has a very important role as a visual information form. Image is often rich of information which can include thousands of words [1]. An image will have a lot of information if the image is served in the form of text. Image is divided into two categories: analogue image and digital image. Analogue image is an image having ‘continue’ characteristic as pictures in television monitor, X-ray, etc. Meanwhile, digital image is the image which can be processed in computer. One of the example is the use of medical image technology such as Magnetic Resonance Imaging (MRI) which has rapidly growing in the last decades. This stimulated special digital image analysis technique development suited with big image data set [2]. Image processing is a study about algorithm which takes image as an input and inverses image as an output. Image processing is a technology rapidly growing nowadays. In computer vision and image processing area, feature is a piece of relevant information to complete computation task [3]. Smart based opiris image processing tool is very important compared to visual feature method [4].
Pear is a tree coming from tropical climate area in West Europe, Asia and North Africa. The tree can reach 10-17 meter tall. However, some of the species are short trees with lush leaves. Pear is a fruit plant included in Pyrus genus in Rosaceae family. It has been cultivated for more than 2,000 years [5]. Pear tree has a lot of varieties in shape and colour. Due to its many varieties, ordinary people will get difficulties in recognizing the varieties of the tree conventionally and with bare eyes. Two of the many varieties which are difficult to differ are monster pear and Williams pear.

By using classification process in image processing technology, it is hoped that people will be able recognize pear varieties in a short time. Image processing application An image processing application which can help to ease classification process of fruit varieties in a short time and accurate complemented with relevant research is K-Means Clustering. K-Means Clustering used to get characteristics extraction value of image which has been divided between object and its background. The result is then classified using K-Nearest Neighbour. K-Nearest Neighbor is not only used for classifying data based on the previous image characteristics extraction result, but also for plotting data to simplify in seeing the data group of classification result.

The above problem brings out an idea of how to classify two varieties of pear, that is monster pear and Williams pear. Pear varieties classification process uses K-Nearest Neighbor (KNN) Algorithm. The purpose of this research is to classify varieties of monster and Williams pear from the processed image. The image is the processed using matlab application to get the classification result.

2. Literature Review

2.1. RGB

The using of colour space or colour model is very precise in image processing. By using colour model, it will be easier to get value of processed image. RGB Model is a colour model consists of three colour channel i.e. Red, Green and Blue. Each channel has colour depth displayed in each image pixel. Mostly used colour depth is 24 bit (true colours) which can directly show RGB colour [6]. RGB colour space is widely used and is usually default colour space to store and represent digital image. We can get other colour spaces from the transformation of RGB or non-linear.

RGB formulation:

\[
\begin{align*}
    r &= \frac{R}{R+G+B} \\
    g &= \frac{G}{R+G+B} \\
    b &= \frac{B}{R+G+B}
\end{align*}
\]

Figure 1. RGB Colour Model
2.2. CIE L*a*b
After conducting RGB colour space process, the colour space is then transformed into CIE l*a*b. Lab colour model illustrate systematically all colours which can be seen in three dimension. In Lab colour model, ‘L’ represents light (positive number for light), ‘a’ represents green-red colour contrast (positive number for red and negative number for green), and ‘b’ represent blue-yellow colour contrast (positive number for yellow and negative number for blue).

2.3. HSV
HSV colour model defines colour in Hue, Saturation and Value. Hue is an actual colour, like green, white and black. Hue is used to differ colour and determine reddish, greenish light. Hue is related to wavelength of light. Saturation states the purity level of a colour, which shows how much is given to white colour. Value is an attribute stating light amount received by eyes [8]. HSV is also used to recover blur picture dealing with brightness and saturation [9].

2.4. K-Means Clustering
K-means Clustering Method is the most popular method and used widely in data clustering process. To increase accuracy of fault diagnosis algorithm, samples must be filtered [10]. K-means Clustering is used to get characteristics extraction value of processed image to be classified in the next process.

2.5. K-Nearest Neighbour (KNN)
K-Nearest Neighbour (KNN) algorithm is a method to classify new object based on the nearest neighbour (K). This method is used to classify object on exercise sample in feature space. It counts distance between test data and exercise data [11]. Beside to classify, K-Nearest Neighbour (KNN) is also used to plot data to ease the grouping. Not only using two varieties of pear, this research can also use more varieties with suitable classification method and therefore ease the research development. Generally, this research can be a step to use for other fruits or objects.

3. Methods
The data of this research are the images of Monster and Williams pear. The data used here are 88 pear images which cover 44 image practice data of Monster pear and 44 image practice data of Williams pear. Meanwhile, the test data taken are 10 images of each variety. Final step of proposed method is classifying image based on extracted feature using KNN algorithm. Pear image varieties classification process can be seen on Figure 2:

![Pear Image Data](image-url)
Figure 3 shows the steps generally of the method:

![Research Diagram](image)

**Figure 3.** Research Diagram

4. Result and Discussion
The beginning process of pear variety classification is inputting pear image from the test data of the pear in this research. Test data can be seen on Figure 4 below:

![Pear test data image](image)

**Figure 4.** Pear test data image

Pear practice data image is then processed and transformed from RGB colour image to l*a*b to soften. Then image segmentation is conducted to separate object from the background. After that, characteristics extraction is done using K-Means Clustering to get the value R, G, B, H, S, V and Area which can be seen on Figure 5. After getting R, G, B, H, S, V and Area values from the pear image, K-Nearest Neighbour (KNN) is conducted to get image classification result of monster and Williams pear. Test result can be seen from processed image data plotting graphic. Practice data distribution graphic in each class is shown on Figure 6.

![Characteristics Extraction](image)

**Figure 5.** Characteristics Extraction.

![Practice data distribution graphic](image)

**Figure 6.** Practice data distribution graphic.

After that, practice data distribution graphic and test data with K-Nearest Neighbour (KNN)
method can be seen on Picture 7.
To get accuracy from pear variety classification, a test using GUI Matlab application that has been designed is conducted.

Figure 7. Graphic of practice data and test data distribution

Figure 8. Application of pear variety classification

Accuracy test by inserting a test image of the image data of each variation of the pear into the application Matlab which further do the calculation with the method of K-means Clustering and classification KKN to know fit into the category of what kind of test images inserted. from the test results by entering the image data then can be a classification of types of pear, as in show in figure 8. puts pictures of pear of monsters and tested in the method with applications Matlab generate a classification of the pear into the type of pear monsters. and this is done against the image of the pear other to determine the classification of types of pear, Monster Pear or Williams Pear. the test results with the application of Matlab shown in the table below:

| No | Image Name | Type of Pear | Result       |
|----|------------|--------------|--------------|
| 1  |            | Monster Pear | Monster Pear |
| 2  |            | Monster Pear | Monster Pear |
| 3  |            | Monster Pear | Monster Pear |
| 4  |            | Williams Pear| Williams Pear|
| ...|            | ................| ...............|
| 20 |            | Williams Pear| Williams Pear|

Table 1. Pear Varieties Classification.
Based on the above table, the result of test accuracy shows that 90% from 19 (correct number)/20 (test data number) * 100%. The high number of accuracy value resulted shows that K-Means Clustering and K-Nearest Neighbor (KNN) methods are very good to apply in classifying the fruit variety.

5. Conclusion
Based on the research, application use to classify two varieties of pear (monster and Williams pear) through image processing gets good accuracy. Test data distribution and practice data shows maximum result by using K-Nearest Neighbor algorithm. Due to this result, the application has shown good accuracy which ease people who are not familiar with pear to classify pear varieties. The quantities of practice data and image quality impact the result of this research test very much. The more practice data, the better test result will be gotten. The data used here are 88 pear images which cover 44 image practice data of Monster pear and 44 image practice data of Williams pear. Meanwhile, the test data taken are 10 images of each variety. This research is limited only for two varieties of pear, i.e. Monster and Williams pear. However, this method can be conducted for further research applied for more varieties of pear or for other objects—with suitable classification method, of course. The result of test accuracy shows that 90% from 19 (correct number)/20 (test data number) * 100%. The high number of accuracy value resulted shows that K-Means Clustering and K-Nearest Neighbor (KNN) methods are very good to apply in classifying the fruit variety.

Acknowledgment
The authors would like to thank for all of the prayers and the help of some friends as well as a thank to the family for the all of supports.

References
[1] M. Lv, X. Wang, Y. Xue, L. Yu, and M. Jin, “Color Compression of RGB Image Based on K-means Clustering,” 2019.
[2] J.-D. Tournier et al., “MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation,” Elsevier.
[3] X. Liu, Z. Deng, and Y. Yang, “Recent progress in semantic image segmentation,” Artif. Intell. Rev., vol. 52, no. 2, pp. 1089–1106, Aug. 2019.
[4] A. Nashat, N. M. Hussain Hassan, and A. A. Nashat, “Article in Multidimensional Systems and Signal Processing,” Springer, vol. 30, no. 2, pp. 571–589, Apr. 2018.
[5] C. Xue et al., “PbrmiR397a regulates lignification during stone cell development in pear fruit,” Plant Biotechnol. J., vol. 17, no. 1, pp. 103–117, Jan. 2019.
[6] L. Fu, E. Tola, A. Al-Mallahi, and Y. Cui, “A novel image processing algorithm to separate linearly clustered kiwifruits,” Elsevier, 2019.
[7] B. Lei, N. Wang, P. Xu, and G. Song, “New Crack Detection Method for Bridge Inspection Using UAV Incorporating Image Processing,” J. Aerosp. Eng., vol. 31, no. 5, Sep. 2018.
[8] Y. A. Gerhana, W. B. Zulikar, A. H. Ramdani, and M. A. Ramdhani, “Implementation of Nearest Neighbor using HSV to Identify Skin Disease Recent citations Designing academic advising information system using prototyping method A D Supriatna et al-Architecture of human resource management system at universities R Setiawan-Conceptual model of executive information system data (A Case Implementation of Nearest Neighbor using HSV to Identify Skin Disease),” iopscience.iop.org.
[9] T. Zhang, H. Hu, B. L.-I. Access, and undefined 2018, “A naturalness preserved fast dehazing algorithm using HSV color space,” ieeexplore.ieee.org.
[10] J. Liu and Q. Li, “A discrete hidden Markov model fault diagnosis strategy based on K-means clustering dedicated to PEM fuel cell systems of tramways,” Artic. Int. J. Hydrog. Energy, 2018.
[11] M. Zuhaer et al., “Face Recognition System Based on Kernel Discriminant Analysis, K-Nearest Neighbor and Support Vector Machine,” Int. J. Res. Eng., vol. 5, no. 3, pp. 335–338, 2018.