020 Classification of Lycopersicon Esculentum Fruit Based on Color Features with Linear Discriminant Analysis (LDA) Method_2019

by Nita Merlina

Classification of *Lycopersicon Esculentum* Fruit Based on Color Features with Linear Discriminant Analysis (LDA) Method

1st Arief Setya Budi*
Computer Science
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jakarta, Indonesia
ariefsetya334@gmail.com

2nd Moh Arie Hasan
Computer Science
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jakarta, Indonesia
hasan.arie@gmail.com

3rd Sri Hadianti
Computer Science
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jakarta, Indonesia
sri.shv@nusamandiri.ac.id

4th Nita Merlina*
Information System
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jakarta, Indonesia
nita@nusamandiri.ac.id

5th Dwiza Riana*
Computer Science
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jakarta, Indonesia
dwiza@nusamandiri.ac.id

***Abstract*— Tomatoes (*Lycopersicon Esculentum*) is a fruit that has many types. One way to distinguish the types of tomatoes can be seen from differences in color, size, and shape. To assist in the classification of tomatoes, a study was conducted which aims to classify two types of tomatoes namely plum tomatoes and beef tomatoes. Therefore, the processing of tomato images is done to facilitate the layman in classifying the two types of tomatoes. The research method used consisted of preprocessing in the form of color conversion l*a*b and HSV for feature extraction, and Linear Discriminant Analysis (LDA) method used to determine data distribution. This method can separate and classify the two types of tomatoes well. The use of adequate training data will further improve classification accuracy. The final results of this study indicate that the level of accuracy in the classification of tomatoes for both types is 90%.**

Keywords*— Lycopersicon esculentum, Linear Discriminant Analysis (LDA), Classification of Tomatoes, l*a*b, HSV.

I. Introduction

In image processing, computer graphics and computer vision can be considered as "translating" the input image into the appropriate output image [1]. An important part of image processing is color. Besides being able to be seen visually, the image also has important information in the presentation of the image quality. One image processing is on tomatoes. External color features and internal feature tightness are the most important factors observed by consumers (wholesalers or retailers) to determine the quality of tomatoes [2].

Tomatoes are one of the most popular fruits in everyday life because they are consumed by millions of people every day. They are usually preferred when cooked in good condition. To maintain the condition of tomatoes, it is necessary to prevent it from damage during distribution. [3]. Tomatoes are known worldwide as one of the most produced plants. Food and Agriculture Organization of United Nations reported that the world produced nearly 170,750 tons of tomatoes in 2014 [4].

Licopersicon esculentum or tomatoes are one of the fruits that are plentiful in Indonesia and often exceeds the market. Tomato fruit has many different colors and shapes. There are many types of tomatoes such as plum tomato, beef tomato, cherry tomatoes, and many more. Therefore, many people have difficulty in classifying the types of tomatoes. This research is the initial research that will be the basis for making an application that allows users such as tomato farmers or ordinary people to easily recognize the classification of tomatoes based on technology.

Many studies on the classification of image processing that have used color features in various images, such as the classification of research [5], [6], [7], [8], [9] and many more. This research will also utilize color features in pre-processing, which has not been done much in previous studies, especially for *Lycopersicon Esculentum* fruit images.

Based on the problems above it obtained an idea of how to classify two types of tomatoes namely tomato plum and tomato beef. Tomato types are classified by using the Linear Discriminant Analysis (LDA) algorithm. The features used are feature b of l * a * b and features h and s of HSV. The purpose of this study was to classify the types of plum tomato and beef tomato from existing images. Then the image was processed by using Matlab application to obtain the classification of plum tomato image and beef tomato image. The contribution of this research is the process of combining the features of l*a*b and HSV in pre-processing for use in classification of *Lycopersicon Esculentum* fruit images.

This paper is divided into several sections. Section 2 discusses about research methods used in the research. Section 3 describes the results and discussion, which will be enclosed with conclusions and further research plans.

II. Research Methods

The data used in this study consists of 50 *Lycopersicon Esculentum* images which are 25 training data on plum tomato images and 25 training data on beef tomato images. The test data consists of 5 plum tomato images and 5 beef tomato images. An example of the image of both types can be seen in Fig. 1.

(a) (b)

Fig. 1. Plum Tomato (a) and Beef Tomato (b)

The research methods for classification process of tomato image types can be seen in Figure 2.

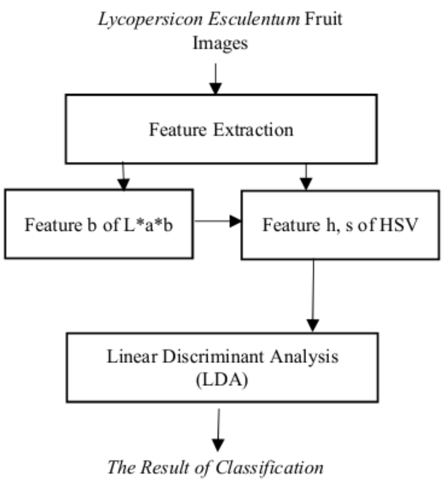


Fig. 2. Classification Design of *Lycopersicon Esculentum* Fruit Types

Figure 2 shows the flow process of classification. From inserting RGB of *Lycopersicon Esculentum* Fruit images. The initial stage is feature extraction process. Consists of two processes namely converting RGB images to L*a*b and converting RGB images to HSV. The process of RGB images to HSV to reduce the intensity of light from the outside to obtain Hue, Saturation and Value. The l*a*b and HSV feature are extracted from the original image. In this study, feature b is used to represent a dark image of a tomato image. In this process also done performing morphological operations to improve the structure and shape of images into binary images. The next step is data analysis and classification using the Linear Discriminant Analysis method to separating two types of objects. In this process data plotting distribution is also carried out.

A. RGB Image Input

RGB color space is widely used and is usually the default color space for storing and representing digital images (Fig.3). We can get other color spaces from RGB or non-linear transformations. RGB color space is the color space used by computers, graphics cards, and monitors or LCDs [10]. This process aims to display the RGB (Red, Green, Blue) color space from the image of tomatoes that are managed.

RGB formula:

$$r = \frac{R}{R+G+B} \tag{1}$$

$$g = \frac{G}{R+G+B} \tag{2}$$

$$b = \frac{B}{R+G+B} \tag{3}$$

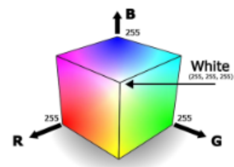


Fig. 3. RGB Color Model

B. CIElab (CIE l*a*b

After processing the RGB color space, then the color space is changed to CIE l*a*b. CIELAB or known as L*a*b is the uniform CIELab color space where the Euclidean distance among two different colors is close to the color difference felt by human eye [11]. L*a*b is very often used as a reference for color space because this color space is able to describe the entire color space that can be seen by human eye. The process of color space changing aims to obtain lighting values and colors in managed images so that the input image looks smoother and the details are better. To convert RGB images to CIElab, the color space must be changed into XYZ like the following formula:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.431 & 0.342 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.020 & 0.130 & 0.939 \end{bmatrix} \begin{bmatrix} R_{norm} \\ G_{norm} \\ B_{norm} \end{bmatrix}, \tag{4}$$

Where Rnorm, Gnorm, and Bnorm show three channels after normalization by (R þ G þ B). Normalization helps reduce nonuniformity of illumination. CIELab Channels are calculated by:

$$L^* = \begin{cases} 116 \times \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, & \frac{Y}{Y_n} > 0.008856 \\ 903 \times \left(\frac{Y}{Y_n}\right), & \frac{Y}{Y_n} \leq 0.008856 \end{cases}. \tag{5}$$

$$a^* = 500 \times \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right]. \tag{6}$$

$$b^* = 200 \times \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right]. \tag{7}$$

Where Xn, Yn, and Zn are tristimulus values of CIE XYZ from white color given by f:

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & t > 0.008856 \\ 7.787 \times t + \frac{16}{116}, & t \le 0.008856 \end{cases} \tag{8}$$

C. Hue Saturation Value (HSV)

The input image in the RGB color space is converted to HSV color space by using transformation. HSV images are a collection of three different images as hue, saturation, and value [12]. HSV has proximity to RGB system in describing colors that humans can see. HVS serves to reduce the intensity of light from outside and is able to detect certain objects. The following is RGB to HSV formula:

$$\begin{aligned} C_{max} &= \max(R'G'B') \\ C_{min} &= \min(R'G'B') \\ \Delta &= C_{max} - C_{min} \end{aligned} \tag{9}$$

Cmax functions to determine the largest constant value in the RGB value while Cmin determines the smallest value in the RGB value.

Hue calculation formula:

Hue Calculation:

$$\begin{cases} 60^0 \mathrm{x} \left(\frac{G'-B'}{\Delta} \bmod 6\right), C_{max} = R' \\ 60^0 \mathrm{x} \left(\frac{B'-R'}{\Delta} + 2\right), C_{max} = G' \\ 60^0 \mathrm{x} \left(\frac{R'-G'}{\Delta} + 4\right), C_{max} = B' \end{cases} \tag{10}$$

Hue is to look for the dominant color that can be accepted by human eye based on the size of wavelength.

The saturation calculation formula:

Saturation calculation:

$$S = \begin{cases} 0, & \Delta = 0 \\ \frac{\Delta}{C_{max}}, & \Delta <> 0 \end{cases} \tag{11}$$

It functions to measure the amount of white light found in hue.

Value calculation formula:

Value calculation:

$$V = C_{max} \tag{12}$$

Value (V) is the brightness value of color. The value ranges from 0-100%. If the value of 0, then the color will be black. The greater the value is, the brighter it will be, and new variations emerge from that color.

D. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a type of supervised learning method where the input area is divided into decision areas whose boundaries are called decision surfaces or decision boundaries [13]. The use of Linear Discriminant Analysis (LDA) method aims to group data into several classes, and then the grouping is based on a straight line of linear equations so that it can classify the image of plum tomato and beef tomato.

E. Plotting Data Distribution

Plotting data distribution is carried out with the aim of testing and seeing the graph of image data distribution processed based on hue and saturation values to see the results on the test accuracy of tomato image processing. Plotting data distribution to be displayed is the distribution of training data in each class, the distribution of training data for each class along with the boundary lines, and the distribution of test data in each class.

III. The Results and Discussion

The initial process of classification on tomato types is the input of tomato images from the test data in this study. The examples of training data can be seen in Figure 4.

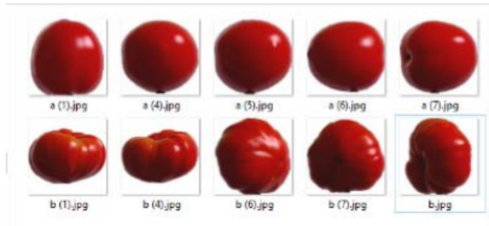


Fig. 4. The examples of Training Data Images of Tomatoes

The training data image of the tomato was then processed and transformed from an RGB color image to l*a*b to refine and show the image details of tomatoes as shown in Figure 5.

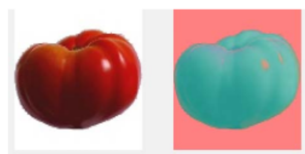


Fig. 5. RGB image to L* a*b

In this study, segmenting images was conducted by using thresholding method and morphological operations to perfect segmentation results and converting grayscale images to determine the foreground area and background area with binary image values. The results of segmentation are shown in Figure 6.

Fig. 6. Binary Image

Transforming the color space from RGB image into HSV image (Hue, Saturation, Value) was used as a reference to do color recognition of an object on a digital image and to reduce the intensity of light from outside that can be seen in Figure 7.

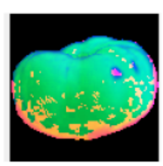


Fig. 7. HSV Image

The test results can be seen from the image data plotting graph processed. The graph of training data distribution in each class is shown in Figure 8.

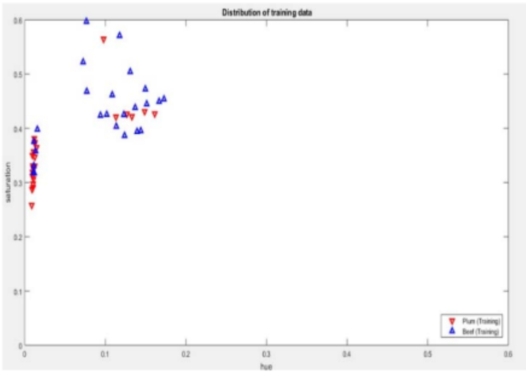


Fig. 8. Training Data Distribution Graph

While the graph of training data distribution in each class along with the boundary lines is shown in Figure 9.

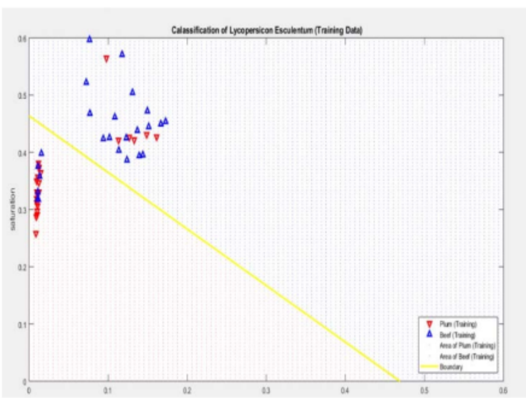


Fig. 9. Training Data Distribution Graph

Next, the graph of training data and test data based on the boundary line with the Linear Discriminant Analysis method can be seen in Figure 10.

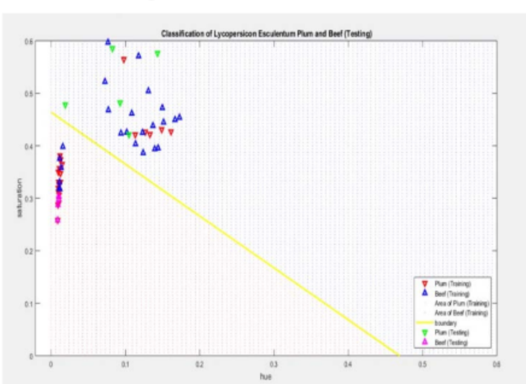


Fig. 10. The Distribution Graph of Test Data and Training Data

To get the accuracy on the classification of tomato types, a test from Graphical User Interface (GUI) application designed was conducted. The following is the appearance of the application that has been created (Fig.11 and Fig.12).

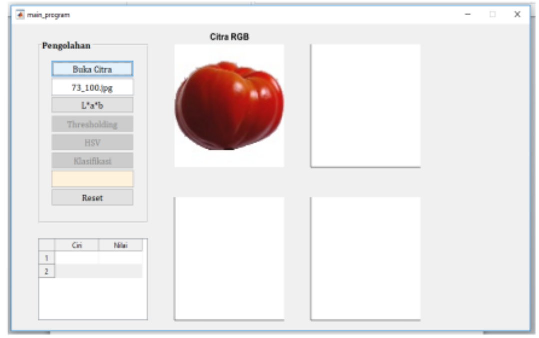


Fig. 11. Display of Input Image on Matlab GUI Application

The following is the appearance of the application on RGB image to L*a*b (Fig. 12):

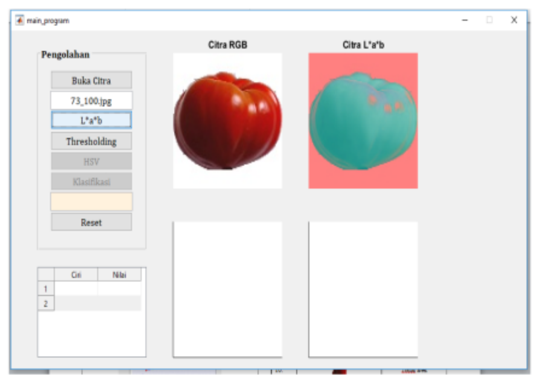


Fig. 12. Display of application on RGB Image to L*a*b

Display of conversion image process and morphological operations with binary image values in the application (Fig 13):

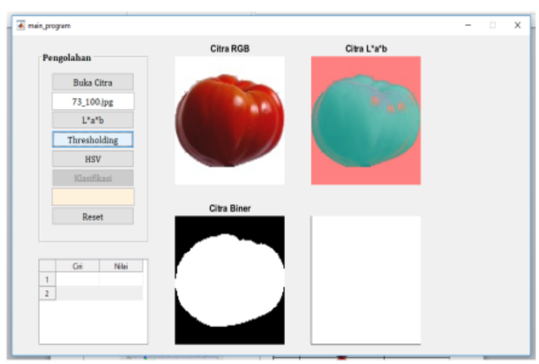


Fig. 13. Display of conversion images

Figure 14 is appearance of application for the color space transformation from RGB image into HSV image (Hue, Saturation, Value):

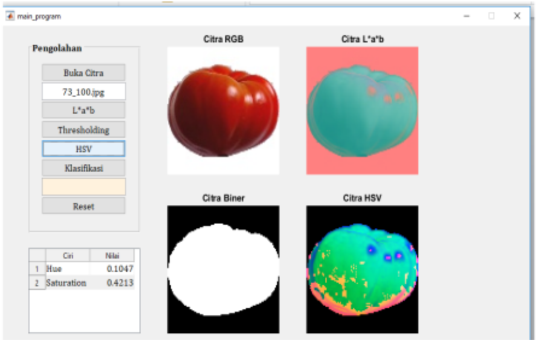


Fig. 14. Display of RGB to HSV transformation

The display results from the classification of tomatoes with the Linear Discriminate Analysis algorithm (Fig. 15 and Fig. 16):

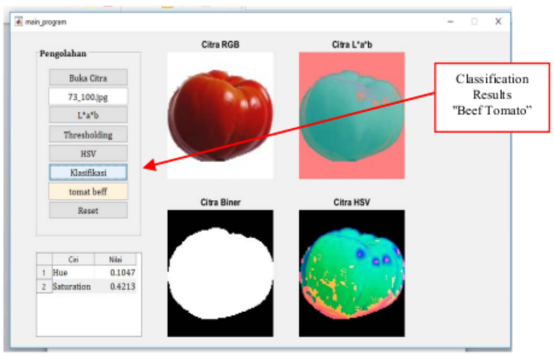


Fig. 15. Display of Beef Tomato Classification Results

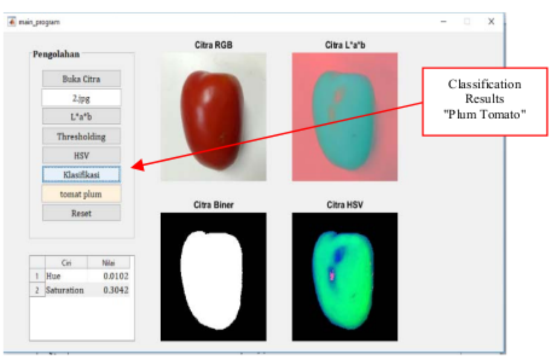
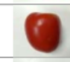


Fig. 16. Display of Plum Tomato Classification Results

The test results from the application can be seen in Table I.

TABLE I. RESULTS OF TOMATO IMAGE PROCESSING

No	The Result of Classification	
	Image testing	***Name of fruits***
1.		Beef Tomato
2.		Plum Tomato**
3.		Plum Tomato**
4.		Plum Tomato**
5.		Plum Tomato**
6.		Beef Tomato**
7.		Beef Tomato**
8.		Beef Tomato**
9.		Beef Tomato**
10.		Beef Tomato**

** Classification is true

The table above shows that the final results of testing accuracy is 90% obtained from 9 (correct number) / 10 (total test data). The accuracy value shows that the linear discriminant analysis method is very well applied to the classification of fruit types.

IV. CONCLUSION

The classification on the image processing of plum tomato and beef tomato obtained an accuracy of 90% meaning that the application of Linear Discriminate Analysis method is very well used in the process of tomato classification. The image quality used and the amount of image training data is very influential on the results of image processing accuracy. The more training data used, the better the accuracy of the classification of tomatoes. The classification of tomato still uses two types of tomatoes, namely plum tomato and beef tomato. It is suggested to develop further research using more than two types of tomatoes.

REFERENCES

[1] P. Isola, J. Zhu, … T. Z.-P. of the I., and undefined 2017, “Image-to-image translation with conditional adversarial networks,” openaccess.thecvf.com.

[2] P. Sehgal and N. Goel, “Auto-annotation of tomato images based on ripeness and firmness classification for multimodal retrieval,” in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1084–1091, 2016.

[3] L. Zhang, J. Jia, G. Gui, X. Hao, W. Gao, and M. Wang, “Deep Learning Based Improved Classification System for Designing Tomato Harvesting Robot,” IEEE Access, vol. 6, pp. 67940–67950, 2018.

[4] R. G. de Luna, E. P. Dadios, and A. A. Bandala, “Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition,” in TENCON 2018 - 2018 IEEE Region 10 Conference, pp. 1414–1419, 2018.

[5] Ji, Yuhui, Qinjun Zhao, Shuhui Bi, and Tao Shen. "Apple Grading Method Based on Features of Color and Defect." In 2018 37th Chinese Control Conference (CCC), pp. 5364-5368. IEEE, 2018.

[6] Riana, Dwiza, Herdian Tohir, and Achmad Nizar Hidayanto. "Segmentation of Overlapping Areas on Pap Smear Images with Color Features Using K-Means and Otsu Methods." In 2018 Third International Conference on Informatics and Computing (ICIC), pp. 1-5. IEEE, 2018.

[7] Sabrol, H., and K. Satish. "Tomato plant disease classification in digital images using classification tree." In 2016 International Conference on Communication and Signal Processing (ICCSP), pp. 1242-1246. IEEE, 2016.

[8] Kartika, Dhian Satria Yudha, and Darlis Herumurti. "Koi fish classification based on HSV color space." In 2016 International Conference on Information & Communication Technology and Systems (ICTS), pp. 96-100. IEEE, 2016.

[9] Hlaing, Chit Su, and Sai Maung Maung Zaw. "Tomato Plant Diseases Classification Using Statistical Texture Feature and Color Feature." In 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), pp. 439-444. IEEE, 2018.

[10] S. Kolkur, D. Kalbande, P. Shimpi, C. B. preprint arXiv, and undefined 2017, “Human skin detection using RGB, HSV and YCbCr color models,” arxiv.org.

[11] Y. Yang, C. Liu, H. Yu, D. Shao, F. Tsow, and N. Tao, “Motion robust remote photoplethysmography in CIELab color space,” J. Biomed. Opt., vol. 21, no. 11, p. 117001, 2016.

[12] K. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, J. M. M. Jenitha, “Comparative study of skin color detection and segmentation in HSV and YCbCr color space” Procedia Computer Science, pp. 41-48, 2015.

[13] N. Mahmoudi and E. Duman, “Detecting credit card fraud by Modified Fisher Discriminant Analysis,” Expert Syst. Appl., vol. 42, no. 5, pp. 2510–2516, 2015.

020 Classification of Lycopersicon Esculentum Fruit Based on Color Features with Linear Discriminant Analysis (LDA) Method_2019

ORIGINALITY REPORT

13%	13%	11%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

6%

★ www.researchgate.net

Internet Source

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		