

Mobile e-detection of Banyuwangi's citrus fruit maturity using k-nearest neighbor

Chairul Anam^{a,1}, Solehatin^{a,2,*}

^{a,b} Manajemen Informatika, Sekolah Tinggi Ilmu Komputer PGRI Banyuwangi, 648416, Banyuwangi, Jawa Timur, Indonesia

¹ anam.stikom@gmail.com *; ² atin33@yahoo.co.id

* corresponding author

ABSTRACT

Banyuwangi is the largest oranges-producing city in East Java, and the orange produced is Siamese citrus fruit. Siamese is Banyuwangi local citrus fruit often found at the harvest time and has a sweet taste. To determine the citrus fruit level, people can detect it from the color and texture. In this modern era, people can use an application to determine the citrus fruits' maturity level. From the elements of color and texture, this research will add the citrus fruit's contours, namely the pore size of the citrus fruit and the distance between the curve of the tip of the orange. Taking pictures of citrus fruits will be following the application stages that will be used as the image of inputting the data. The detection is then conducted using the K-NN method based on several criteria based on the input image after the feature extraction process. The feature extraction stages are segmentation, normalization, thresholding, and thinning, which will be produced in several criteria: the maximum RGB value, the minimum RGB value, pore size, and the distance between the tip's curve of the orange. The research results that have been carried out are based on the research stages to get a similarity percentage following the inputted data. The E-Detection application can provide information to citrus farmers, especially beginner citrus farmers, to know the level of fruit maturity oranges to be harvested.

Keywords:

E-Detection
Citrus Fruit Maturity
K-NN
RGB Value

I. Introduction

In the previous research, by testing the detection of 10 (ten) tangerines, it was known that the level of maturity at ripe conditions was 2 (two) tangerines with a suitability value between 72.94% to 82.68%. For the condition of Mengkal, as many as 4 (four) tangerines with a suitability value between 59.09% to 81.20% [1]. The raw conditions have four tangerines with a suitability value between 69.02% to 77.60%. For improving system performance in the form of an Android-based application of tangerine fruit ripeness detection, further research is suggested to add other intelligent functions, such as a function that can recognize the shape of an object its texture [2].

Citrus fruit is a non-climacteric fruit obtained from the trees only or cannot be ripened after harvesting. So it is essential to pick or harvest citrus fruits when they are ripe. After harvesting, the fruit with a sour taste will not be sweet because there is no ripening process after harvesting [3]. Color is one of the attributes that could be used to identify specific objects. It is not included in the extraction of information about the object's surface's spectral properties and seeking information from a set of known descriptions to carry out recognition [4]. Recognition of the image objects can use the basic color differences of RGB (Red, Green, Blue) [5], [6].

There are various ways to extract color from an image. The simplest method which is often used is the color histogram [7] and color moment [8]. Both methods can be evaluated using several different color spaces, namely RGB, HSV, YUV, LAB, and YCrCb [9], [10]. The color histogram represents the distribution of colors in an image. The advantage of using color histograms in image analysis is its reliability on rotation, scaling, and relatively fast computation times. On the other hand, the disadvantage of color histograms is that they tend to lose spatial information and high sensitivity to noise disturbance [11]s, such as changes in luminance intensity and quantization errors [12]–[14].

The classification is a critical process to distinguish one thing from another, and it could be in the form of animals, plants, or humans [15], [16]. This identification could be taken by focusing on the



characteristics of something. One way to classify the citrus fruit is using the naive Bayes method in an image-based way [17], [18]. Their texture and color could recognize citrus fruit. Citrus fruit has identical colors; they are green and yellow. In some cases, people sometimes have difficulty in classifying orange and mandarin oranges [19].

In this study, the k-nearest neighbor (K-NN) method was used to detect citrus fruits' ripeness with several criteria for color, contour, and texture. This study was contributed to the model development of the soft computing approach applied to the mobile application.

II. Method

Before processing the input image, the image data (direct image captured by gadgets such as a mobile phone or tablet) are sent to the server. The types of file compression supported in this study are limited to several types, namely jpg, bmp, gif. There are several provisions or rules in shooting fruit, including:

- a. The fruit picture is taken in an upright position.
- b. Besides, ensure that there is no cut-off part of the fruit, and the limitation for the left, right, top, and bottom shootings are right at the border of the leaf fruit.
- c. There will only be one fruit in the image.
- d. The fruit pictures could be detected when it is lighter than the background color.
- e. The system has a minimum and maximum limit for the image resolution.
- f. The largest size of the image is limited to 1300 x 1300 pixels, and the lowest is 80 x 80 pixels.

After taking the image, the resizing process is carried out in 150 x 150 pixels. For speeding up the recognition process, an example of resizing can be seen in Fig. 1.

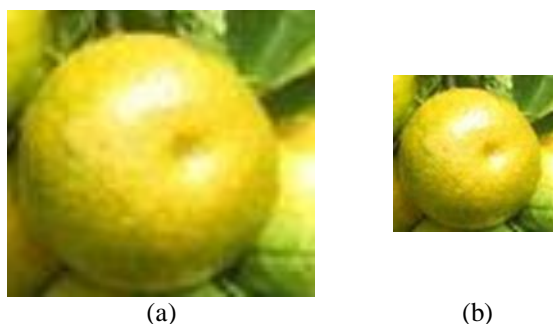


Fig. 1. The Example of Resizing Process, (a) the original image size 650 x 650, and (b) the result image size 150 x 150.

A. Algorithm Design

The first step of Algorithm design in this study is processing the input image to be processed in a different process, known as preprocessing. The preprocessing process starts from browsing the file name; then, the filtering process is carried out. Then, in the image segmentation, the thresholding process is conducted after separating the image from the background.

The second process calculates the minimum and maximum points for each red, green and blue color component. This calculation will be used as the criteria that use the k-nearest neighbor (K-NN) method [20]. The preprocessing process is used to find the curve points to the centroid. The data will be used in the recognition process using the K-NN method. The next stage is to calculate the K-NN method using 8 (eight) criteria, including the minimum and maximum Red, Green, and Bluepoints, the minimum distance of the curve point to the centroid, and the curve points maximum distance to the centroid.

The final result is a percentage value of the closeness between the input image and the database's image (data training). In the K-NN method, the citrus fruit data matches the 8 (eight) criteria for each citrus fruit. Last, the information about the citrus fruit maturity will be shown in the final result. The block diagram is shown in Fig. 2.

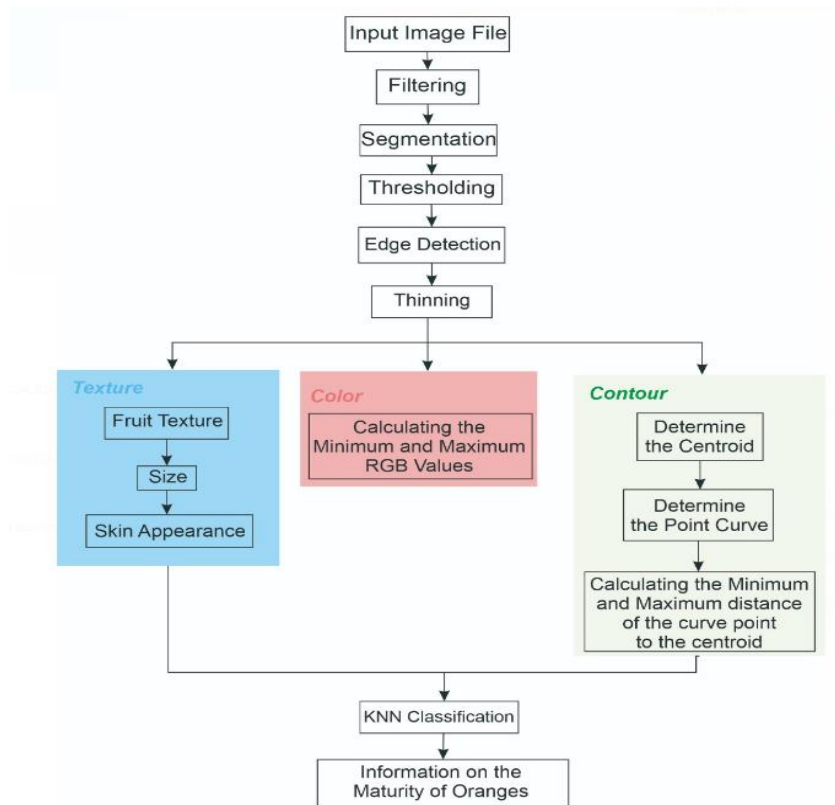


Fig. 2. Proposed algorithm

B. Software Development Design

The system architecture created explains the relationship between an admin, user, and database in an application. After the detection process, the system architecture flow's resizing process could be seen in Fig. 3.

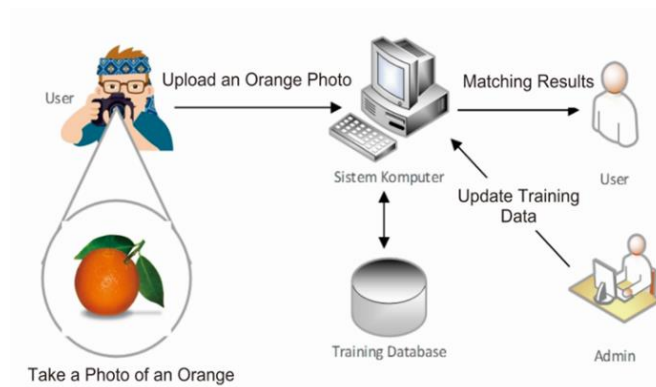


Fig. 3. System Architecture

1. Filtering Process

The filtering process has a function to smooth the image in the image by using the linear filtering method with the low pass filter method by taking low frequencies, removing high frequencies and carrying out blur effects, and noise the reduction.

2. Edge Detection

Edge detection is an image that produces the edges of image objects. The purpose of doing edge detection is to mark the part that becomes the detail of the image and improve the blurred image's detail. This study's edge detection process is from the segmentation process results using the Sobel method applied to the respective RGB values with gray-level values between 0-255.

3. Thinning

The thinning process was first carried out at the coordinate point (x, y), namely the coordinates (0.75) of the pixels at the center point of the coordinate axis on the 150x150 pixel image. The search is carried out if the first point is found and will be continued at the next point by tracing clockwise based on calculating the 8-connected component approach.

C. System Analysis

In this study, the system's preparation was made in 2 stages: system analysis and system design. For the system analysis, the author uses the Unified Modeling Language (UML), and the author describes the form design of the application.

1. Business Process

The business process describes the flow of the application for admins and operators. During an admin using the application, he logs in first to use the application; then, he could input the testing's orange data. After the orange data has been entered into the database, the operator can operate the application. Without the login step, the operator can use the application by taking photos of oranges and citrus data. Then, based on the orange data that has been inputted by the operator using the citrus data collection process and the existing methods, the application can provide the information about the citrus fruit maturity for the operator.

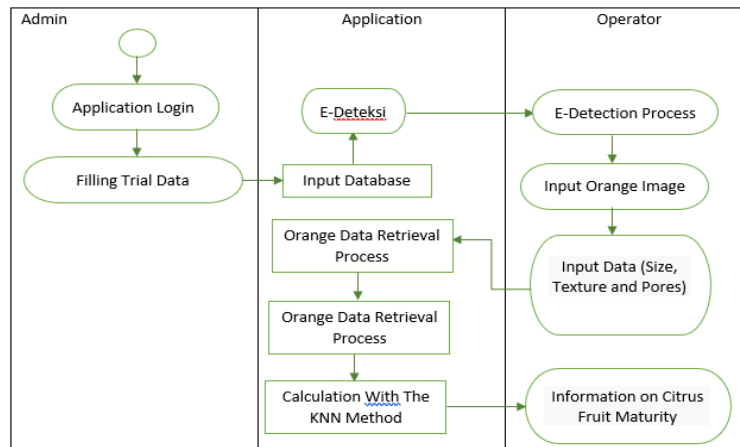


Fig. 4. Business Process.

2. Use Case Diagram

Admins and operators use the E-detection application following the access rights that exist in the application. An admin can access all applications, including inputting, processing, and editing both user data and orange data, by logging in to access them. Meanwhile, the operator can only use the application to detect citrus fruit's maturity by inputting the required data about citrus fruits. The use case image can be seen in Fig. 5.

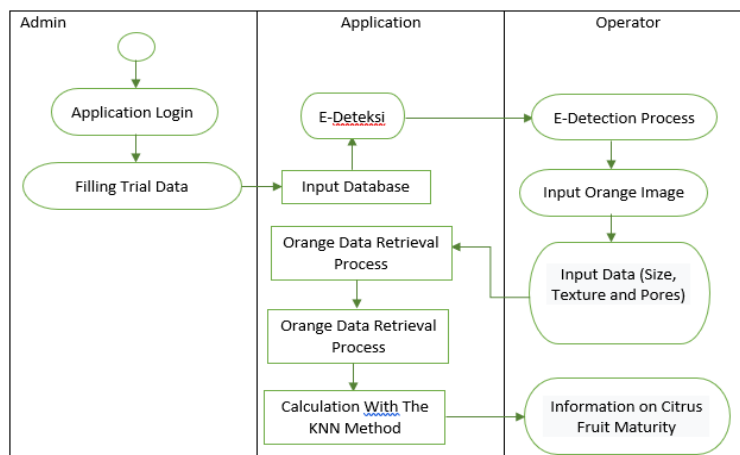


Fig. 5. Use Case Diagram.

III. Results and Discussion

The detection process was previously carried out; several steps needed to be conducted, including inputting photos of oranges to be detected, the cropping process, and inputting texture data, including the size, texture, and pores. After inputting the photos, the next step is the cropping process according to the specified size. It is 150 X 150 as for the cropping process image in Fig. 6. After that, the color image-taking process includes filtering, segmentation and thresholding, and edge detection. The filtering process is carried out after the cropping process takes place on the Mobile Android application, then the segmentation process is carried out, as shown in Fig. 7.



Fig. 6. Cropping Process.

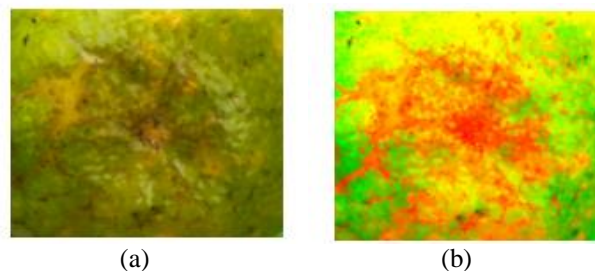


Fig. 7. Segmentation process on (a) original image and (b) the result.

















The thresholding process and edge detection are carried out. However, because the contour only yields one and the sweetness level does not affect, the curve point's minimum and maximum distance to the centroid is not included in the citrus fruit maturity detection criteria. Next, input the texture data, including size, texture, and new pores, press the process button to determine the oranges' level of maturity. Before providing information regarding citrus fruit maturity, some of the resulting criteria are calculated using the K-NN method. The resulting data can be seen in Fig. 8.



Fig. 8. Detailed criteria data on oranges.

The detection results show that the similarity of the data tested with master data has similarity until 98.53%, and the information on the maturity level is 75%. This application is easy to use and can provide online information on the citrus fruits' maturity level that citrus farmers can use before harvesting. The other testing data could be shown in Table 1.

Table 1. Testing of detection of maturity level

Number	Original Image	Segmentasi Image	Size	Texture	Pore	Detection Results
1			0.2000	0.2000	0.8000	Similarity Percentage 82.81%, with a Maturity Level of 75%
2			0.2000	0.2000	0.8000	Similarity Percentage 95.88%, with a Maturity Level of 95%
3			0.3000	0.8000	0.8000	Prosentase Kemiripan 97.70%, with a Maturity Level of 100%
4			0.5000	0.8000	0.8000	Similarity Percentage 98.08%, with a Maturity Level of 100%
5			0.3000	0.8000	0.8000	Similarity Percentage 90.29%, with a Maturity Level of 75%
6			0.2000	0.2000	0.2000	Similarity Percentage 100.00%, with a Maturity Level of 60%
7			0.2000	0.2000	0.2000	Similarity Percentage 100.00%, with a Maturity Level of 40%
8			0.5000	0.8000	0.8000	Similarity Percentage 98.53%, with a Maturity Level of 75%

IV. Conclusion

Based on the research, an Android-based application is easy-to-use and can help citrus farmers detect the citrus fruits' maturity. This study uses the K-Means Clustering method by using the difference between the euclidean distance in the citrus fruit criteria used as a reference to help determine the maturity of oranges. This research can produce an application that can help novice farmers determine the level of maturity of citrus fruits. Citrus farmers will use this application before harvesting as a reference to determine the level of citrus fruit maturity.

Acknowledgment

This research was supported by the PGRI Banyuwangi College of Computer Science and funded by the Higher Education Grant Fund (PDP DIKTI). The author would like to thank for those who have provided financial and moral support for this research.

References

- [1] H. Prabowo, "Deteksi Kondisi Kematangan Buah Jeruk Berdasarkan Kemiripan Warna Pada Ruang Warna Rgb Berbasis Android," *J. Elektron. Sist. Inf. dan Komput.*, vol. 3, no. 2, pp. 9–19, 2017.

- [2] V. Andrearczyk and P. F. Whelan, "Deep learning in texture analysis and its application to tissue image classification," in *Biomedical texture analysis*, Elsevier, 2017, pp. 95–129.
- [3] K. Warman, L. A. Harahap, and A. P. Munir, "Identifikasi Kematangan Buah Jeruk dengan Teknik Jaringan Syaraf Tiruan," *J. Rekayasa Pangan dan Pertan*, vol. 3, no. 2, pp. 248–253, 2015.
- [4] L. Armi and S. Fekri-Ershad, "Texture image analysis and texture classification methods-A review," *arXiv Prepr. arXiv1904.06554*, 2019.
- [5] M. R. Hassan, R. R. Ema, and T. Islam, "Color image segmentation using automated K-means clustering with RGB and HSV color spaces," *Glob. J. Comput. Sci. Technol.*, 2017.
- [6] X. Zhu *et al.*, "Automatic recognition of lactating sow postures by refined two-stream RGB-D faster R-CNN," *Biosyst. Eng.*, vol. 189, pp. 116–132, 2020.
- [7] A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, "Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor," in *2018 international conference on computing, mathematics and engineering technologies (iCoMET)*, 2018, pp. 1–6.
- [8] V. P. Singh and R. Srivastava, "Improved image retrieval using color-invariant moments," in *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT)*, 2017, pp. 1–6.
- [9] B. Farou, H. Rouabhia, H. Seridi, and H. Akdag, "Novel approach for detection and removal of moving cast shadows based on rgb, hsv and yuv color spaces," *Comput. Informatics*, vol. 36, no. 4, pp. 837–856, 2017.
- [10] P. Rosyani, M. Taufik, A. A. Waskita, and D. H. Apriyanti, "Comparison of color model for flower recognition," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 2018, pp. 10–14.
- [11] L. Wendling, I. Debled-Rennesson, and H. Nasser, "Multilevel polygonal descriptor matching defined by combining discrete lines and force histogram concepts," *Multimed. Tools Appl.*, pp. 1–15, 2019.
- [12] J. M. Wandeto and B. Dresp-Langley, "Color Sensitivity of The Quantization Error in a Self-Organizing Map: Dataset."
- [13] J. Xu, J. Miao, Z. Gao, K. Nie, and X. Shi, "Analysis and modeling of quantization error in spike-frequency-based image sensor," *Microelectron. Reliab.*, vol. 111, p. 113705, 2020.
- [14] E. Park, D. Kim, S. Yoo, and P. Vajda, "Precision highway for ultra low-precision quantization," *arXiv Prepr. arXiv1812.09818*, 2018.
- [15] Y. Zhong, E. Dutkiewicz, Y. Yang, X. Zhu, Z. Zhou, and T. Jiang, "Internet of mission-critical things: human and animal classification—a device-free sensing approach," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3369–3377, 2017.
- [16] A. M. Tarazona, M. C. Ceballos, and D. M. Broom, "Human relationships with domestic and other animals: one health, one welfare, one biology," *Animals*, vol. 10, no. 1, p. 43, 2020.
- [17] W. E. Sari, Y. E. Kurniawati, and P. I. Santosa, "Papaya Disease Detection Using Fuzzy Naïve Bayes Classifier," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2020, pp. 42–47.
- [18] A. Wajid, N. K. Singh, P. Junjun, and M. A. Mughal, "Recognition of ripe, unripe and scaled condition of orange citrus based on decision tree classification," in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2018, pp. 1–4.
- [19] R. K. Haba and K. C. Pelangi, "Pengelompokan Buah Jeruk menggunakan Naïve Bayes dan Gray Level Co-occurrence Matrix," *Ilk. J. Ilm.*, vol. 12, no. 1, pp. 17–24, 2020.
- [20] Y. Dang, N. Jiang, H. Hu, Z. Ji, and W. Zhang, "Image classification based on quantum K-Nearest-Neighbor algorithm," *Quantum Inf. Process.*, vol. 17, no. 9, pp. 1–18, 2018.