Bigint Computing Journal
Open Access
Publisher Ali Institute of Research and Publication

Implementation of HSV Imagery with K-Nearest Neighbor for Classification of Maturity Levels in Tomatoes

Lailatul Husna Aulia*, Fazar Azhari, Muhammad Dhuha Bimantara

Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Indonesia

A B S T R A C T

The tomatoes used use tomatoes (Lycopersicum esculentum Mill) which is a type of horticultural plant. One type is plum tomatoes. The process of classifying tomato ripeness is carried out manually through direct visual observation. However, this is very difficult to do because it is inconsistent. Therefore, relevant features are needed to classify tomato maturity levels based on HSV features using the KNN method. The method used in classification is K-Nearest Neighbor, this algorithm requires features to build the model. The feature used is HSV feature extraction. Based on the results of research tests carried out, it proves that Euclidean distance k=2 has a percentage value of 85%. Based on the level of accuracy, the color feature k=2 shows the best k value in classifying tomato ripeness levels measuring 400x400 pixels. To achieve a high level of accuracy, image processing time should be less.

Keyword:
classification, KNN, HSV

* Correspondence :
Lailatul Husna Aulia,
Universitas Islam Negeri Sumatera Utara
Email: lailatulhusnaaulia@gmail.com

INTRODUCTION

Tomato fruit (Lycopersicum esculentum Mill) is a type of horticultural plant. One type is the plum tomato (Lycopersicon lycopersicum)[1]. Currently there are no relevant features to accurately determine the ripeness level of tomatoes. Although HSV images provide information about the ripeness level of tomato fruit, variations in color intensity and lighting conditions can affect the quality of color representation. Tomatoes are one of the agricultural products that have a high production level. Tomatoes have a relatively fast maturity time [3]. The distribution

of tomatoes in various regions makes it important to classify tomatoes based on their level of ripeness [4]. Manual observation produces maturity levels that are less uniform and less satisfactory [5]. Because the results of manual selection are less than satisfactory, a method is needed to select and classify tomatoes properly [2]. Digital image processing or image processing has input and output in the form of images [6]. The process of identifying fruit ripeness can use image processing [7]. Fruit maturity parameters in terms of fruit skin color are one of the important factors in determining fruit maturity [8].

The KNN algorithm was chosen as the classification algorithm because KNN is quite simple with good accuracy based on the minimum distance using Euclidean Distance [9]. The K-Nearest Neighbor (KNN) algorithm is a method for classifying objects based on learning data that is closest to the object [10]. Near or far neighbors are usually calculated using Euclidean distance [11]. The main thing to do is determine the k value used for the classification process [12]. K-NN is also a non-parametric classification system that is able to solve probability density problems [13]. One method that can be used to identify colored objects is digital image processing techniques by extracting special features in the image [3]. There are several color analysis models in digital image recognition, including Hue, Saturation, Value (HSV) [14]. Hue is the level of redness and greenness of light, saturation shows the intensity of the white color in the image, and value shows the amount of light received by the eye [15]. The process of getting the value of each color you want to display is through a calculation process by changing the RGB color space to HSV color space [16]. Image segmentation to perform segmentation based on color characteristics, expansion and contraction operations to help eliminate noise, naming fruit objects using the recursive component labeling method [8].

Previous research [17] on the classification of tomato ripeness levels using the HSV algorithm as a color feature and the LVQ algorithm as a classification, obtained an accuracy rate of 83.75%. Research [9] Classification of Tomato Ripeness Levels Based on RGB and HSV Colors Using the KNN Algorithm obtained an accuracy value of 91.25%. The accuracy obtained is quite good, but there are still classification errors because there are still images of tomatoes that have light reflections with different intensities. The distance between training data and test data can be calculated using various methods, including using the Euclidean equation [18]. Classification uses the majority vote among k object classifications [19].

It is important to scale HSV space components to reduce computation and increase efficiency [20]. One of the indicators used to determine the ripeness of tomatoes is the color of unripe, semi-ripe and ripe fruit [21]. The scaled image must still have good quality [22]. Mathematically, the image is a function of light intensity in two dimensions [23]. Color features are used to simplify image processing [24]. This research will conduct research related to how to classify the level of ripeness of tomatoes based on HSV image features using the KNN method. There are labels and attributes, labels are taken from the level of fruit ripeness and attributes are obtained from the results of image feature extraction. These attributes are R(red), G(green), B(blue), H(hue), S(saturation), V(value), and Area [25]. In this study, the level of maturity of tomatoes was classified into 3 classes of maturity, namely unripe, semi-ripe and ripe. The aim of this research is to determine the relevant features in classifying tomato maturity levels based on HSV features using the KNN method so that the resulting accuracy can be better. The difference between this research and several previous studies is that the quality of tomato

images must minimize light reflections when taking pictures so that it can produce a better level of accuracy even though not all images use a white background, by using the HSV feature and the KNN method.

METHODOLOGY

The flow or steps of the KNN algorithm for classifying ripeness levels in tomatoes can be seen in the image below:

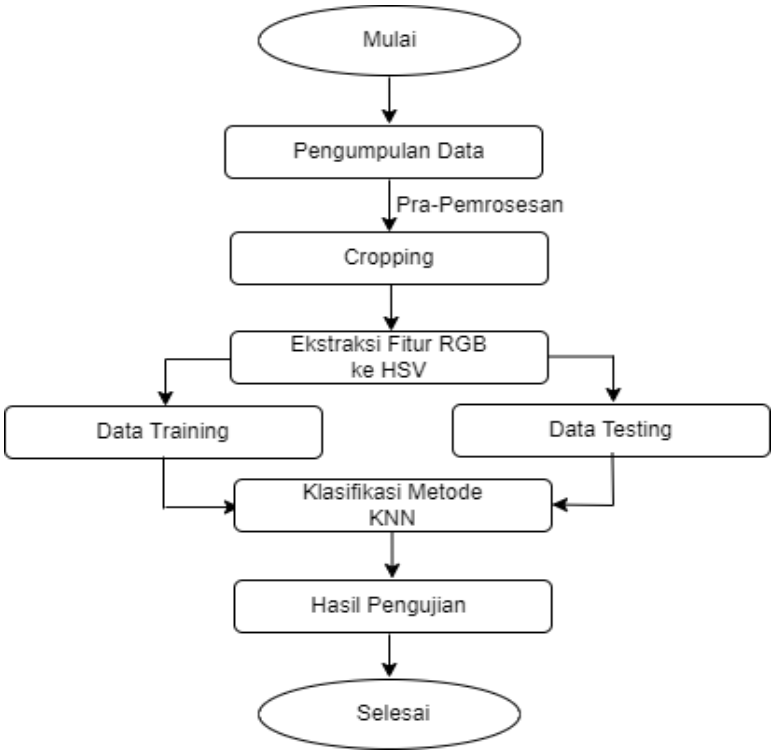


Figure 1. Flow of the KNN Method

1. Data Collection

   Data collection carried out in this research was by observation, which is an important data collection technique in the research process. This observation technique was carried out in markets and tomato gardens to collect image data on tomatoes. Then the tomatoes were acquired using an iPhone 11 camera, and a dataset of images of tomatoes with various levels of ripeness was collected, namely: raw, quite ripe, and ripe.

Figure 2. Raw

Figure 2. Sufficiently Ripe

Figure 3. Ripe

2. Cropping

   Next, we proceed to the pre-processing stage, namely cutting the image of the tomato (cropping). This process is carried out to focus only on the object part of the tomato fruit. Crooping or cutting image areas is a technique used to determine exactly which part of the image contains the object area to be processed, so that it can be cut and separated from areas that are not needed for further processing. This research uses a resize size of 400x400 pixels, the process of cropping the tomato image can be seen as in Figure 4 below.

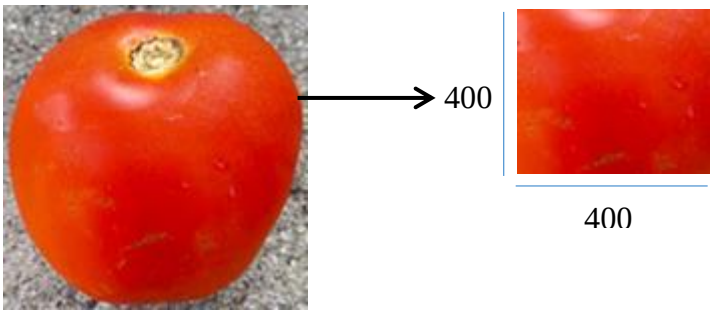


Figure 4. Cropping process

3. RGB to HSV Feature Extraction

   The RGB to HSV feature extraction stage aims to convert the RGB values to HSV in order to obtain the correct values. Where it uses two data, namely training data and test data. The training data will later be used to find appropriate values, while the test data will be used to test what has been obtained in the testing stage. The HSV image model maps color into three main components, namely hue, saturation and value. Calculate the average of each H, S, and V value. The average value is used as an HSV feature. Feature extraction at this stage is carried out by inserting images of tomatoes into the program design that has been designed using the GUI tool in the Matlab R2016a application.

4. K-Nearest Neighbor Method Classification

   The next stage is to classify the tomato image using the K-Nearest Neighbor Method, which is a method that uses a supervised algorithm where the results of the new query instance are classified based on the majority of the categories in the K-Nearest Neighbor. The classification process with KNN is to determine the level of maturity of tomatoes. This classification uses the most votes among the classifications of k objects. The k values used in this research are 1, 2, 3, 4, 5, 6, 7, and 8 with the search for the distance between training data and testing data being Euclidean Distance. There are several stages in using KNN, namely:

   a. Sets the K value used.
   b. Calculating the Euclidean distance between test data and training data.
   c. Sort the results from smallest to largest.

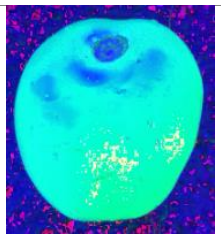
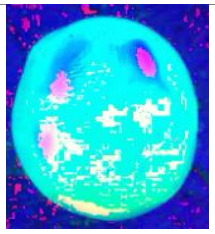
d. Grouping data based on predetermined K values
e. Use the class that has the most results from new data classification.

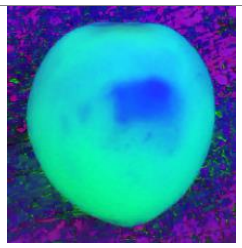
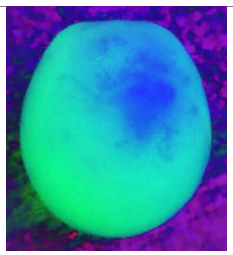
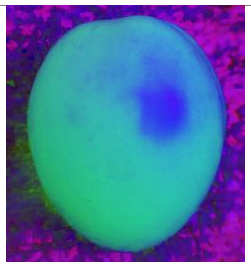
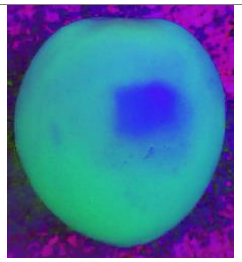
5. Test result
   The test results stage is to test the influence of image size and the influence of environmental value (k) on KNN. This stage is to test the training data and test data to determine the level of accuracy resulting from testing tomato images.

RESULTS AND DISCUSSION

Pictures were taken using a non-white background, because the background color does not affect the value of each image characteristic. The image classification results are test data that has been processed in a classification system based on the characteristics of each image with 9 different distance values. In this research, 30 test data were used with each class having 10 original images. After that, manual calculations will be carried out to determine the level of accuracy of each distance/k value. The feature extraction results have an average HSV value of 100 data. Based on the results of feature extraction processing, the higher the image pixel value, the longer the time required. Based on the existing dataset that has gone through the preprocessing process, the next experimental stage carried out is the Hue Saturation Value (HSV) Color Extraction Method and Classification using the K-Nearest Neighbor (KNN) method to test the dataset on the dataset. prepared model. The following is evidence of experiments that have been carried out on the dataset to be tested, totaling 30 Test Data. The table below shows the results and attached evidence, totaling 6 experimental data, namely raw, moderately mature, and mature:

Table 1. Appendix Results of Research Experiment Evidence

Data	Original Image	HSV image	Class	Result	Information
1			1	1	Right, because class 1 or above ripe
2			1	1	Right, because class 1 or above ripe

Data	Original Image	HSV image	Class	Result	Information
3			2	2	Right, because class 2 or enough ripe
4			2	2	Right, because class 2 or enough ripe
5			3	3	Right, because class 3 or raw
6			3	3	Right, because class 3 or raw

In Table 1, each original image was tested using the GUI tool in the Matlab R2016a application. The original image of the tomato is extracted from the RGB color features to HSV to use the color characteristics as a limit or benchmark for whether a tomato image can be said to be ripe or not based on the previous process which used the red color as a reference for this determination.

Table 2. System Test Accuracy Results

k value	Appropriate data	Overall Data	Accuracy
1	23	30	56%
2	37	30	85%
3	31	30	78%
4	21	30	53%
5	19	30	48%
6	18	30	45%
7	19	30	48%
8	21	30	53%

Based on Table 2 above, the level of accuracy obtained from the test data produces the highest accuracy value of 85% with a distance k=2. So that it can be used as a reference in carrying out classification using the KNN method, the best K value is 2. Meanwhile, the lowest accuracy value is obtained at a distance of k=6, namely 45%. To predict and test the KNN classification model that has been created, the results obtained are that all the images tested produce the prediction results are the same as the class so they match the experimental dataset.

The difference between this research and previous research [21], the method used is the same but the resulting level of accuracy is higher. The KNN environmental value in the classification of tomato maturity levels influences the accuracy results, but is not very significant. On the other hand, image size has a significant effect on accuracy, because the average test results at the highest image size also produce high accuracy. The highest accuracy reached 92.5% on a tomato image measuring 1000x1000 pixels and an environmental value of 3. In this study the image used was 400x400 pixels in size. Research [22] regarding the classification of ripeness levels in tomatoes produced an accuracy level of 92% with a value of k=3. However, some tomato images have light reflections. In this research, the image of tomatoes obtained must minimize light reflection. Meanwhile, research using another method [23], namely the Support Vector Machine method for classification of maturity in tomatoes, provides accurate results. The HSV color model produces an accuracy rate of 95%, YCbCr 74% and CIElab 100%. However, it is necessary to use other methods and color models to maximize the accuracy of the system results.

CONCLUSION

Based on the results of research that has been carried out using 30 test data using a dataset that was tested as many as 3 images of tomatoes, namely images of unripe, quite ripe and overripe tomatoes. From the experimental results above, the K-Nearest Neighbor (KNN) method and Hue Saturation Value (HSV) feature extraction to classify the level of ripeness in tomatoes are suitable as algorithms and can be used well because of the resulting level of accuracy. that's enough to suit your wishes. namely 85% with a distance k=2 on a tomato image measuring 400x400 pixels. Even though it doesn't use a white background.

There are several things that influence the process of classifying tomato ripeness levels, so it is recommended to use other pre-processing methods and images with higher sizes to produce high accuracy, but require less image processing time. Additionally, it is recommended to develop appropriate applications to use the best test results scenarios

REFERENCES

[1] A. Marliah, M. Hayati, and I. Muliansyah, "Utilization of Liquid Organic Fertilizer on the Growth and Yield of Several Varieties of Tomato (Lycopersicum esculentum L.)," J. Agrista, vol. 16, no. 3, pp. 122–128, 2012.

[2] S. Y. Riska, "Classification of Tomato Ripeness Levels Based on Differences in Image

Improvement Using Average RGB and Pixel Index," J. Ilm. Technol. Inf. Asia, vol. 9, no. 2, pp. 18–26, 2015.
[3] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto, and F. Syafria, "K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors," Indonesia. J. Artif. Intel. Data Min., vol. 2, no. 2, p. 101, 2019, doi: 10.24014/ijaidm.v2i2.7975.
[4] S. Y. Riska and P. Subekti, "Classification of Tomato Fruit Ripeness Levels Based on Color Features Using Multi-Svm," J. Ilm. Inform., vol. 1, no. 1, pp. 39–45, 2016, doi: 10.35316/jimi.v1i1.442.
[5] N. Atrianda, "Classification of Tomato Fruit Ripeness with Variations in Color Models Using Support Vector Machine," VOCATECH Vocat. Educ. Technol. J., vol. 1, no. 2, pp. 45–52, 2020, doi: 10.38038/vocatech.v1i2.27.
[6] R. Pratama, A. Fuad, and F. Tempola, "Detection of Tomato Fruit Ripeness Based on Color Features Using the His Color Space Transformation Method," JIKO (Journal of Information and Computers), vol. 2, no. 2, pp. 81–86, 2019, doi: 10.33387/jiko.v2i2.1318.
[7] N. Nafi'iyah, "Classification of Mango Fruit Ripeness Based on HSV Image with KNN," J. Elektron. List. and Technol. Inf. Terap., vol. 1, no. 2, pp. 4–7, 2019.
[8] A. Chan, P. Liem, N. P. Wong, and T. Gunawan, "Fruit Segmentation Using the K-Means Clustering Method and Identification of Ripeness Using the Color Content Comparison Method," J. SIFO Mikroskil, vol. 15, no. 2, pp. 91–100, 2014, doi: 10.55601/jsm.v15i2.156.
[9] L. Ningsih and P. Cholidhazia, "Classification Of Tomato Maturity Levels Based on RGB And HSV Colors Using KNN Algorithm," RIGGS J. Artif. Intel. Digits. Bus., vol. 1, no. 1, pp. 25–30, 2022, doi: 10.31004/riggs.v1i1.10.
[10] I. A. A. Angreni, S. A. Adisasmita, and M. I. Ramli, "Regarding the Accuracy Level of Road Damage Identification," vol. 7, no. 2, pp. 63–70, 2018.
[11] D. Syahid, J. Jumadi, and D. Nursantika, "Classification System for Philodendron Leaf Ornamental Plant Types Using the K-Nearest Neighbor (KNN) Method Based on Hue, Saturation, Value (HSV)," J. Online Inform. , vol. 1, no. 1, p. 20, 2016, doi: 10.15575/join.v1i1.6.
[12] M. G. R. Miftahus Sholihin, "Egg Quality Classification Based on Color Features Using the K-Nearest Neighbor Method," Fak. Technol. Inf. UNMER Malang, pp. 1188–1193, 2018.
[13] H. Parvin, H. Alizadeh, and B. Minati, "A Modification of K-Nearest Neighbor Classifier," Glob. J. Comput. Sci. Technol., vol. 10, no. 14, pp. 37–41, 2010.
[14] M. Manchanda and R. Sharma, "Fusion of visible and infrared images in HSV color space," 3rd IEEE Int. Conf. , pp. 1–6, 2017, doi: 10.1109/CIACT.2017.7977384.
[15] S. Sanjaya, "Application for Recognizing Ripeness Levels of Tomatoes Using Android-Based Hsv Color Features," J. Teknoinfo, vol. 16, no. 1, p. 26, 2022, doi: 10.33365/jti.v16i1.1489.
[16] A. Salsabila, R. Yunita, and C. Rozikin, "Flower Type Image Identification using the KNN Algorithm with HSV Color Extraction and GLCM Texture," Technomedia J., vol. 6, no. 1, pp. 124–137, 2021, doi: 10.33050/tmj.v6i1.1667.
[17] M. A. Anggriawan, M. Ichwan, and D. B. Utami, "Recognition of Tomato Ripeness Levels Based on Color Images in a Case Study of Automatic Selection System Development," J. Tek. Inform. and Sis. Inf., vol. 3, no. 3, pp. 550–564, 2017, doi: 10.28932/jutisi.v3i3.688.
[18] J. Harefa and M. Pratiwi, "Comparison Classifier: Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) In Digital Mammogram Images," Juisi, vol. 02, no. 02, pp. 35–40, 2016.
[19] A. P. N, M. M. R. Wulanningrum, M. Kom, U. Nusantara, and P. Kediri, "IDENTIFICATION OF TOMATO FRUIT MATURITY USING THE K-NEARST NEIGHBOR (K-NN) ALGORITHM IKA WAHYUN NINGSIH The progress of human information is developing at a rapid pace "In technology, life and agriculture are getting better in terms of quality," vol. 01, no. 08, pp. 3–8, 2017.
[20] M. Lumb and P. Sethi, "Texture Feature Extraction of RGB, HSV, YIQ and Dithered Images

using GLCM, Wavelet Decomposition Techniques," Int. J. Comput. Appl., vol. 68, no. 11, pp. 25–31, 2013, doi: 10.5120/11623-7087.

[21] S. Sanjaya *et al.*, "K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue , Saturation , and Value Colors," vol. 2, no. 2, pp. 101–106, 2019.

[22] U. K. N. N. Algorithm, "Classification Of Tomato Maturity Levels Based on RGB And HSV Colors Using KNN Algorithm 1," vol. 1, no. 1, pp. 25–30, 2022.

[23] V. Education and T. Journal, "Vocational education and technolgy journal," vol. 1, no. 2, 2020.

[24] R. Pratama, A. Fuad, andF. Tempola, "Detection of Tomato Fruit Ripeness Based on Color Features Using the His Color Space Transformation Method," IKO (Journal of Informatics and Computers)., vol.2, no. 2, pp. 81-86, 2019, doi: 10.33387/jiko.v2i2.1318.

[25] S. Aprilisa and Sukemi, "Classification of Tomato Fruit Ripeness Levels Based on Color Features," Pros. Annu. Res. Semin. 2019, vol. 5, no. 1, pp. 978–979, 2019.

[26] K. A. Pratama, W. P. Atmaja, and V. Lusiana, "Classification of Kersen Fruit Maturity Levels Using HSI Image with the K-Nearest Neighbor (KNN) Method. Kersen is a plant that has small red and sweet fruit like cherries. Kersen plants are a very common type of tree," vol. 11, no. 1, pp. 105–108, 2022.

[27] D. I. Muhammad, E. Ermatita, and N. Falih, "Using K-Nearest Neighbor (KNN) to Classify Starfruit Images Based on Color Features," Inform. J. Computer Science., vol. 17, no. 1, p. 9, 2021, doi: 10.52958/iftk.v17i1.2132.

[28] S. P. Adenugraha, V. Arinal, and D. I. Mulyana, "Classification of Ambon Banana Ripeness Using KNN and PCA Methods Based on RGB and HSV Images," J. Media Inform. Budidharma, vol. 6, no. 1, p. 9, 2022, doi: 10.30865/mib.v6i1.3287.