Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2024. Vol. 3. No. 2; e-ISSN : 2808-4519

Segmentation Algorithm K – Means Based on The Maturity Level Of Blueberries

Aditya Putra Prananda[1*], Akim M.H Pardede[2], Rahmadani[3]

[1,2]STMIK Kaputama, [3]Universitas Pembangunan Panca Budi
adityaprananda72@gmail.com [1]; akimmhp@live.com[2]; rahm4dani@gmail.com [3]

Abstract

Traditionally, farmers and consumers have determined the ripeness of blueberries by manual means, such as observing the color, pores, and skin of blueberry fruits. Such recognition takes a relatively long time and gives rise to different levels of maturity because people have visual limitations in recognition, fatigue levels and differences of opinion about good maturity. Consumers tend to pay attention to aspects such as the striking color and size of blueberries, but do not know how ripe and nutritious the fruit is for consumption. Several image processing techniques were used in this study, including image segmentation for segmentation based on color features, expansion and contraction operations to remove noise, and naming fruit objects using recursive component labeling methods. This is followed by separation and training of geometric features and colors. At the time of testing, a classification of the degree of maturity and type of fruit of the object is carried out. To more accurately identify the degree of ripeness of the fruit, check the geometric features and characteristic values of the color content. Range values are determined by statistical methods such as mean and standard deviation. Using the K-Means algorithm to segment blueberry imagery, the study aimed to develop an efficient method to distinguish blueberry ripeness levels automatically. It can help classify and group blueberries according to their degree of maturity. In addition, the results of this study can also be used for quality control of blueberries in the agricultural industry or for fruit marketing applications.

Keywords*: *Blueberry, Clustering, K-Means

1. Introduction

Traditionally, farmers and consumers have determined the ripeness of blueberries by manual means, such as observing the color, pores, and skin of blueberry fruits. Such recognition takes a relatively long time and gives rise to different levels of maturity because people have visual limitations in recognition, fatigue levels and differences of opinion about good maturity. Consumers tend to pay attention to aspects such as the striking color and size of blueberries, but do not know how ripe and nutritious the fruit is for consumption. Several image processing techniques were used in this study, including image segmentation for segmentation based on color features, expansion and contraction operations to remove noise, and naming fruit objects using recursive component labeling methods. This is followed by separation and training of geometric features and colors. At the time of testing, a classification of the degree of maturity and type of fruit of the object is carried out. To more accurately identify the degree of ripeness of the fruit, check the geometric features and characteristic values of the color content. Range values are determined by statistical methods such as mean and standard deviation. The K-Means algorithm is a method of grouping data into groups based on the similarity of certain attributes. The algorithm works by finding the center of the group (midpoint), which is the average representation of the group members. In blueberry segmentation, the K-Means algorithm can be used to group image pixels into groups that describe the maturity of the fruit. Using the K-Means algorithm to segment blueberry imagery, the study aimed to develop an efficient method to distinguish blueberry ripeness levels automatically. It can help classify and group blueberries according to their degree of maturity. In addition, the results of this study can also be used for quality control of blueberries in the agricultural industry or for fruit marketing applications.

2. Research Methods

2.1. Previous Research

The first study entitled "Fruit Segmentation Using K-Means Clustering Method and Identification of Its Maturity Using Color Content Comparison Method" [1]. To create an application to detect fruits and identify their ripeness. The application is built using the Delphi 7 programming language. The methods used include K-Means clustering segmentation, expansion and shrinkage, component labeling to feature extraction. Fruit type detection is done using matching features of fruit roundness level while identification of maturity based on color features of the fruit. Before classifying the name of the type of fruit and the level of maturity, fruit training must be carried out first then proceed with fruit detection and identification of maturity. The application made is believed to be able to classify the type of fruit and

the degree of maturity. This is evidenced by training and testing of 3 (three) types of fruit as many as 30 (thirty) samples each for raw, calc and ripe maturity levels (a total of 90 samples) which show a success rate of identification above 90%.
The second study is entitled "Segmentation of Orange Fruit Maturity Based on Color Similarity Using Algorithm-Means" [2]. Segmentation of orange ripeness based on color similarity using the k-means algorithm results obtained from the process of the stages carried out, using a dataset of 8 oranges taken from 6 angles. Top, bottom, right, left, front and back corners and produce images of 48 datasets to be tested. With *.jpg format and size 124 x 165pixel. Each T value is obtained automatically from the system, then it will be processed with k-means clustring at each image input 1 to image input 6, from image input 1 to image input 6 will be united, if the mature value is greater than 3 it is declared mature. However, if the mature value is less than or equal to 3, it is declared raw. The results obtained from the dataset of 8 oranges measured from 6 angles resulted in a dataset of 48, namely raw data totaling 2 and mature data totaling 6.

2.2. Image

Imagery is a two-dimensional representation of objects from the visual world, involving a wide variety of disciplines that include art, human vision, astronomy, engineering, and so on. It is a collection of colored pixels or dots that are two-dimensional in shape. Literally, an image is an image on a bimatra (two-dimensional) plane. Viewed from a mathematical point of view, the image is a continuous function of the intensity of light in the dual plane. The light source illuminates the object, the object reflects back part of the light beam. This reflection of light is captured by optical devices, such as human eyes, cameras, scanners, and others so that the image of objects in the form of images can be recorded. Images can be grouped into two parts, namely still images and moving images. A still image is a single, motionless image. While moving images are a series of still images that are displayed in a row (sequential), thus giving the impression to the eye as a moving image. Each image in the series is called a frame. The images that appear on feature films or television basically consist of hundreds to thousands of frames [3].

2.3. Blueberry

Blueberries are native to North America and Europe. There are about 50 species of blueberries, both grown and wild, with many different varieties grown in different states in the U.S. and Canada. Blueberries are now also widely grown in Asia and Australia.Blueberry varieties include cranberries (Vaccinium macrocarpon) and blueberries (Vaccinium vitis-idaea). The pods are thin (semi-transparent), dark in color; from dark red-green to dark purple, with a surface coated with a "powder" that serves as a protective layer. Kumquats are collected in groups, ranging in shape from round to oblong and ranging in size from peas to marbles in size. Small seeds are removed from blueberries. Blueberries are rich in vitamin C and have been used since ancient times as a food and medicinal source [4]

2.4. Types of images

2.4.1. Analog Imagery

Analog imagery is generated from analog image acquisition tools, examples are the human eye and analog cameras. Images captured by the human eye and photographs or film captured by analog cameras are examples of analog imagery. A digital image is a representation of the function of light intensity in discrete form on a two-dimensional plane. An image is composed of a set of pixels (picture elements) that have coordinates (x,y) and amplitudes f(x,y). The coordinate (x,y) indicates the location/position of pixels in an image, while the amplitude f(x,y) indicates the color intensity value of the image. Image processing is the process of processing pixels in a digital image that is used for a specific purpose. Initially, image processing was carried out to improve image quality, with the development of the computing world which was characterized by increasing computer capacity and capabilities allowing humans to retrieve information from an [7].

2.4.2. Digital Imagery

Digital image is a 2D function, f(x,y), which is a function of light intensity, where the x and y values are spatial coordinates and the value of the function at each point (x,y) is the grayness of the image at that point. Digital images are expressed by a matrix in which rows and columns represent a point in the image and matrix elements (referred to as image elements or pixels) represent the grayness of that point [3].

2.5. JPEG (.jpg)

JPEG (.jpg) is a very commonly used format today especially for image transmission. this format is used to store imagery using the JFEG strategy [2].

2.6. Pengolahan Citra Digital

Digital image processing is a form of information processing with input in the form of images (images) and outputs that are also in the form of images or can also be part of the image. The purpose of this processing is to improve the quality of the image so that it is easily interpreted by humans or computer machines [3].

2.7. Image segmentation

Segmentation is a technique that means having the option to isolate an image into multiple locales and each region has the same characteristics [2]. The technique of dividing or separating a picture into several regions (regions) by looking at the similarity of attributes is called segmentation. Segmentation is also considered a cycle that separates an image into various parts or objects. Segmentation is certainly not a single interaction that is completed in digital image processing. Nonetheless, segmentation is an important cycle in digital image processing. In segmentation interactions, objects that turn into goals will be retrieved for the next system. There are two qualities of image degree values in this division technique, specifically discontinuity and similarity. In discontinuity, an image is isolated depending

on a noticeable change in its brightness, usually applied to the discovery of edges, lines, areas and sides of the image. While image similarity will be isolated depending on thresholding, area development and local separation and merging, it is generally applied to static and dynamic images.

2.8. K-Means Algorithm

K-Means is a non-hierarchical data grouping technique that seeks to group existing data into one group or set so that data of comparable quality are aggregated into the same group and data with various attributes are aggregated into another [2].

2.9. Euclidean Distance

To measure the distance between the data and the focal point of the cluster, Euclidean distance is used, then at that point a distance matrix will be obtained as follows:

X = Center of cluster $$\mathbf{d}(x, y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (1)$$

Y = Data (2)

2.10.Segmentation Using K-Means

In the segmentation process using K-Means, the initial step is to determine the number of clusters on the previously handled image and count the number of centroids randomly. Then, at that point, confirm the distance of the pixels to the centeroid and collect the pixels that depend on the closest distance. After the pixel values of the image are collected based on their closest distance, the center is recalculated as another center of mass by counting the normal pixels per cluster as another centeroid and uniting the pixels according to the centeroid. If there are still pixels that move clusters, then a new centeroid calculation is carried out, but if there are no pixel values that move clusters, the clustering system is complete [2].

2.11.UML

UML is one of the tools / models for designing object-oriented software development. UML itself also provides standards for writing a blueprint system, which includes business process concepts, writing classes in specific program languages, database schemas, and components needed in software systems [7]. Unifield Modeling Language is one of the visual modeling methods used in designing and making object-oriented software. UML is a writing standard or a kind of blue print diamna in it including a business process, writing classes in a specific language [8]

2.12.Matlab

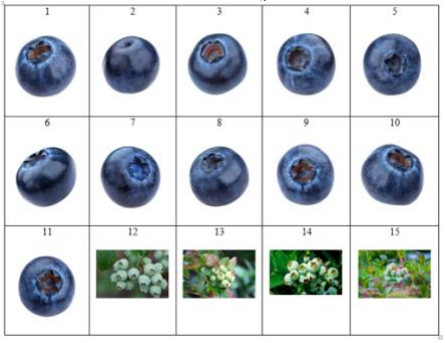
Matlab uses the concept of array/matrix as a standard component variable without the requirement for array declaration as in other languages. It can also be coordinated with outside applications and programming languages such as C, Java, .NET and Microsoft Excel. Matlab programming has a wide range of applications, especially in applications that require numerical calculations. Realize that Matlab does all numerical calculations in matrix form. All numerical activity in Matlab is a matrix operation. Matlab can display computational results in the form of graphs and can be planned according to our wishes by utilizing the GUI that we create ourselves [2].

3. Results And Discussion

3.1. Algorithm Analysis

In data analysis for k-means segmentation based on the maturity level of blueberries, data collection and needs analysis were carried out. Data collection is carried out to obtain some information related to the manufacture of a testing system to determine the k-means segment based on the level of maturity of blueberries, namely in the form of data sets from blueberries. Requirements analysis consists of process requirements, input needs and output needs. Process needs analysis, which explains how the system will work, what processes are used, starting from the entry of input data which is then processed by the system to become output data (the final display of the system). The following dataset of blueberries used in the k-means segmentation process is based on the maturity level of blueberries:

Table 1: Blueberry Dataset



3.2. Calculation of the K-Means Method

Dividing the Image into 4 x 4 Pixel Size to adjust to K-Means, then the size of the image to be processed is divided by blocks with a size of 4 x 4 pixels per block. The author used a grayscale blueberry image dataset with an image size of 4 x 4 pixels with a depth of 8 bits with *.jpg format. For blueberry picture 1 of grayscale grades as follows:

207	172	133	111
169	132	94	78
126	107	84	83
91	86	76	84

Fig. 1: 4x4 Grayscale Image

The steps of segmenting grayscale images with the k-means algorithm are as follows:
f (x,y) = 1,jika f $(x,y) \geq$ T
f (x,y) = 0,jika f $(x,y) \leq$ T

With this method, the Threshold T value is:

$$T = \frac{fmaks + fmin}{2}$$

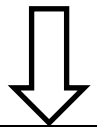
$$T = \frac{207 + 76}{2}$$

T = 141.5

From the results of determining the thresholding value in determining the binary value, then as a reference it is determined by the value of the threshold. If the matrix value is less than the threshold value then the result is 0 and if the matrix value is greater than the threshold value then the result is 1. It can be seen like image 2 of Grayscale images and image 3 of segmented images with the k-means algorithm.

207	172	133	111
169	132	94	78
126	107	84	83
91	86	76	84

Fig. 2: 4x4 Grayscale Image



T = 141.5

1	1	0	0
1	0	0	0
0	0	0	0
0	0	0	0

Fig. 3: Image of Segmentation Results with K-Means Algorithm

The steps of the process of identifying blueberry image by calculating Euclidean distance using the k-means algorithm are as follows:

1. Dataset

Table III.2 is a dataset table of 15 blueberries. Of the 15 pictures of blueberries will be grouped into 2 parts, namely ripe and unripe. The results of testing segmentation datasets with thresholding using the k-means algorithm can be seen in table III.2 of the dataset.

Table 2: Dataset

No	Dataset	Image Segmentation Results With k-means
1	Blueberry 1	141.5
2	Blueberry 2	156.5
3	Blueberry 3	83
4	Blueberry 4	85
5	Blueberry 5	92

2. Determine the number of clusters to be formed.

Cluster 1(C1) = Mature
Cluster 2 (C2) = Raw

3. Set C the center of the initial cluster randomly. In determining the initial cluster center C randomly taken from the Blueberry fruit dataset, namely Blueberry 1 and Blueberry 2.

C1 = 141.5
C2 = 156.5

4. Allocate all data/objects into the nearest cluster. Here's the result of data allocation to cluster distance. As for the result of the distance to the cluster, it is obtained from calculations with the formula:

Iteration 1 :

d1.1 = $\sqrt{(141.5 - 141.5)^2}$ = 0
d1.2 = $\sqrt{(141.5 - 156.5)^2}$ = 15
d2.1 = $\sqrt{(156.5 - 141.5)^2}$ = 15
d2.2 = $\sqrt{(156.5 - 156.5)^2}$ = 0
d3.1 = $\sqrt{(83 - 141.5)^2}$ = 58.5
d3.2 = $\sqrt{(83 - 156.5)^2}$ = 73.5
d4.1 = $\sqrt{(85 - 141.5)^2}$ = 56.5
d4.2 = $\sqrt{(85 - 156.5)^2}$ = 71.5
d5.1 = $\sqrt{(92 - 141.5)^2}$ = 49.5
d5.2 = $\sqrt{(92 - 156.5)^2}$ = 64.5

After calculating iteration 1, the following results are obtained:

Table 3: Distance to Cluster

No	Dataset	Jarak ke Cluster		Hasil
		C1	C2	
1	Blueberry 1	0	15	1
2	Blueberry 2	15	0	1
3	Blueberry 3	58,5	73,5	1
4	Blueberry 4	56,5	71,5	1
5	Blueberry 5	49,5	64,5	1

5. Redetermine the center point of the new cluster based on the average of the new cluster obtained from the formula: result value / many results.

C1 = (141.5 + 156.5 + 83 + 85 + 92 + 75.5 + 155 + 101 + 103 + 99 + 95 + 81 + 94 + 105.5 + 95) / 15 = 1562 / 15 = 104.13
C2 = 0

6. Redo step 4 until the center point of each cluster does not change.

Iteration 2:

d1.1 = $\sqrt{(141.5 - 104.13)^2}$ = 37.37
d1.2 = $\sqrt{(141.5 - 0)^2}$ = 141.5
d2.1 = $\sqrt{(156.5 - 104.13)^2}$ = 52.37
d2.2 = $\sqrt{(156.5 - 0)^2}$ = 156.5
d3.1 = $\sqrt{(83 - 104.13)^2}$ = 21.13
d3.2 = $\sqrt{(83 - 0)^2}$ = 83
d4.1 = $\sqrt{(85 - 104.13)^2}$ = 19.13
d4.2 = $\sqrt{(85 - 0)^2}$ = 85
d5.1 = $\sqrt{(92 - 104.13)^2}$ = 12.13
d5.2 = $\sqrt{(92 - 0)^2}$ = 92

After calculating iteration 2, the following results are obtained:

Table 4: Distance to Cluster

No	Dataset	Jarak ke Cluster		Hasil
		C1	C2	
1	Blueberry 1	37,37	141,5	2
2	Blueberry 2	52,37	156,5	2
3	Blueberry 3	21,13	83	2
4	Blueberry 4	19,13	85	2
5	Blueberry 5	12,13	92	2

7. Redetermine the center point of the new cluster based on the average of the new cluster obtained from the formula: result value / many results.

C1 = 0
C2 = (141.5 + 156.5 + 83 + 85 + 92 + 75.5 + 155 + 101 + 103 + 99 + 95 + 81 + 94 + 105.5 + 95) / 15 = 1562 / 15 = 104.13

8. Redo step 4 until the center point of each cluster does not change.

Iteration 3:

d1.1 = $\sqrt{(141.5 - 0)^2}$ = 141.5
d1.2 = $\sqrt{(141.5 - 104.13)^2}$ = 37.37
d2.1 = $\sqrt{(156.5 - 0)^2}$ = 156.5

d2.2 = $\sqrt{(156.5-104.13)^2}$ = 52.37
d3.1 = $\sqrt{(83-0)^2}$ = 83
d3.2 = $\sqrt{(83-104.13)^2}$ = 21.13
d4.1 = $\sqrt{(85-0)^2}$ = 85
d4.2 = $\sqrt{(85-104.13)^2}$ = 19.13
d5.1 = $\sqrt{(92-0)^2}$ = 92
d5.2 = $\sqrt{(92-104.13)^2}$ = 12.13
After calculating iteration 3, the following results are obtained:

Table 5: Distance to Cluster

No	Dataset	Jarak ke Cluster		Hasil
		C1	C2	
1	Blueberry 1	141,5	37,37	3
2	Blueberry 2	156,5	52,37	3
3	Blueberry 3	83	-21,13	3
4	Blueberry 4	85	-19,13	3
5	Blueberry 5	92	-12,13	3

The results of the 1st iteration and 3rd iteration do not change, so the results are in accordance with the cluster grouping. Here are the results of the grouping.

Table 6: Clustering Results

Dataset	Information
Blueberry 1	Ripe
Blueberry 2	Ripe
Blueberry 3	Ripe
Blueberry 4	Ripe
Blueberry 5	Ripe

4. Conclusions

After doing the explanation in the previous chapter. Then the researcher will provide some conclusions on the results of the program made. The following are the conclusions of research related to k-means segmentation based on the degree of ripeness of blueberries: From the results of the application of image segmentation, the research conducted by the author in analyzing k-means segmentation based on the level of maturity of blueberries resulted in better image quality than before. The results tested that the right segmentation method in carrying out the process of the maturity level of blueberries is Alalgorithm k - means. The system created in segmenting k-means based on the maturity level of blueberries can be used in the selection of the maturity level of blueberries.

References

[1] Andri, Paulus, Wong. N. P., Gunawan. T., (2014), "Segmentasi Buah Menggunakan Metode K-Means Clustering Dan Identifikasi Kematangannya Menggunakan Metode Perbandingan Kadar Warna", vol. 15, no. 2.
[2] Aulia. A., (2021), "Segmentasi Kematangan Buah Jeruk Berdasarkan kemiripan Warna Menggunakan Algoritmak-Means".
[3] Budi S. A, Inna. S., Maulana. H., (2016), "Pengenalan Citra Wajah Sebagai Identifier Menggunakan Metode Principal Component Analysis (PCA)", vol. 9, no. 2.
[4] *Blueberry*. (2022, September 6), *Retrieved* November 1, 2023, *from* https://pertanian.uma.ac.id/sekilas-tentang-buah-blueberry/.
[5] Jumadi. J., Yupiati, Sartika. D., (2021), "Pengolahan Citra Digital Untuk Identifikasi Objek Menggunakan Metode Hierarchical Agglomerative Clustering", vol.10, no. 2.
[6] Prihandoyo. M. T., (2018), "*Unified Modeling Language* (UML) Model Untuk Pengembangan Sistem Informasi Akademik Berbasis Web", vol. 3, no. 1.
[7] Ratna. S., (2020), "Pengolahan Citra Digital Dan Histogram Dengan Phyton Dan Text Editor Phycharm", vol. 11, no. 3.
[8] Sonata. F., Sari. V. W., (2019), "Pemanfaatan UML (*Unified Modeling Language*) Dalam Perancangan Sistem Informasi *E-Commerce* Jenis *Customer-To-Customer*", vol. 8, no. 1, doi: 10.31504/komunika.v8i1.1832.