International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 8, No. 2, (2024), pp. 130-138

Application of The K-Means Clustering Algorithm to Identify Strawberry Fruit Ripe

Muhammad Rizki[1], Mhd Furqan[2], Sriani[3]
[1,2,3]Department of Computer Science, Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Indonesia
Email: Mhdrizki545@gmail.com[1], mfurqan@uinsu.ac.id[2], sriani@uinsu.ac.id[3]

Abstract

Fruit ripeness will usually be determined by several parameters, including size, weight, color characteristics, fragrance, etc. The parameter of fruit ripeness in terms of fruit skin color is one of the important factors in identifying fruit maturity. Segmentation is a method in digital image processing to differentiate objects in an input image. The general classification process is carried out by looking directly at the fruit object. The purpose of this research is to create an analysis in identifying the ripeness of strawberry fruit and designing an application system that can identify the ripeness of strawberry fruit. This application was built with the MATLAB application. The methods used include K-Means Clustering segmentation, labeling and feature extraction. The detection of the type of fruit is carried out using feature matching at the level of shape and color. Before classifying the name of the type of fruit and the level of maturity, the fruit training must be carried out first and then continued with fruit detection and identification of maturity. Based on the results of the strawberry image maturity identification test with six test strawberry images consisting of three types of maturity levels, the results were obtained, namely mature test one and mature test two levels of ripeness and correct identification results, half ripe test one and half ripe test two levels of ripeness and results Correct understanding, raw test one and raw test two mature levels and correct recognition. Meanwhile, the accuracy test results obtained an accuracy value of 100% for identifying the maturity of strawberry images. From the results of the tests carried out, it can be concluded that identification of ripeness in strawberry fruit images was successfully applied using the K-Means Clustering method on images of ripe, half-ripe and unripe strawberries. And from testing the identification of ripeness of strawberry fruit with test data of six images and training data of twelve images, it gave an accuracy result of 100%.

Keywords: *detection, identification, MATLAB, fruit, strawberry, k-means*

1. Introduction

The strawberry plant is a fruit plant that has high economic value. Its allure lies in its striking red fruit with a small, attractive shape and sweet and fresh taste [1]. Strawberry farmers hope that the fruit they grow will produce good quality strawberries so that they will have a high selling price in the future. In ensuring the quality of the fruit, you can also look at the level of ripeness, which can be seen from the color of the strawberries. Fruit maturity is usually determined by several parameters, including the parameters of size, weight, color characteristics, fragrance of the fruit, and others. Fruit maturity parameters in terms of fruit skin color are one of the important factors in identifying fruit maturity [2].

Observations are carried out visually by the farmer's eyes, of course this is not very effective, observations using the eyes have weaknesses, namely human error such as being distracted, looking wrong, visual disturbances, drowsiness and others, this has an impact on the level of ripeness of the strawberry fruit, causing the fruit to be picked not yet just in time so that the taste of the fruit becomes sour and the color obtained is less attractive [3]. This error can be detrimental to farmers because it has an impact on the quality of the fruit produced, thereby reducing the quality of the selling price, and

ISSN : 2580-7250
Copyright © 2024 IJISTECH

reducing consumer confidence. Therefore, we need a system to detect the ripeness of strawberries so that the fruit is picked on time and the fruit tastes sweet and the fruit color is attractive so that it can help farmers in picking and harvesting the strawberries they plant.

As technology evolves, humans have developed many tools including the field of image processing. Segmentation A method used for image segmentation is K-means to separate objects in an input image [4]. K-Means is a non-hierarchical data clustering method that divides the current data into one or more groups or clusters, where data with similar characteristics are classified into one group and points b different 'characteristics' are classified into other groups. One or more groups that use consumer political knowledge/useful information in the decision-making process [5].

Cluster analysis can also be used for image segmentation. Image segmentation using a cluster analysis approach is grouping multidimensional image pixel data into several clusters based on the closeness of the distance between pixels [6]. Image segmentation is very necessary to understand image characteristics completely and is the first step in image analysis. The k-means clustering algorithm is a cluster analysis technique that is often used in image segmentation because of its ease and ability to group large data very quickly [7].

2. Reseach Methodology

Image is a continuous function of light intensity in a two-dimensional plane. The light source illuminates the object, the object reflects back all or part of the light beam and is then captured by an optical or electro-optical device [8]. In other journals, the definition of image is a representation, similarity, or imitation of an object or thing. An image contains information about the object being represented. Images can be grouped into visible images and invisible images. To be seen by the human eye, an invisible image must be converted into a visible image, for example by displaying it on a monitor, printing it on paper and so on. One of the invisible images is a digital image. Image can also be defined as a two-dimensional image produced from a continuous two-dimensional analog image into a discrete image through a sampling process. Analog images are divided into N rows and M columns so that they become discrete images. The intersection between certain rows and columns is called a pixel [8]. Image is one component that plays an important role as a form of visual information [9].

The meaning of processing according to the Kamus Besar Bahasa Indonesia (KBBI) is a method or process of trying to make something different or more perfect. Meanwhile, according to the KBBI, image means a form or image, in this case an image obtained using a visual system. Overall, image processing means a way of turning an image into another, more perfect or desired image. In other words, image processing Digital image processing is processing an image using a computer so that the image is easily interpreted by humans or machines [5].

Digital image processing was initially only used to convert analog images to digital and improve image quality. As equipment supporting image processing develops, the use of digital image processing becomes more diverse. Through image processing algorithms, it is hoped that the function of human vision sensors can be replaced with artificial vision sensors (cameras). The rapidly increasing computer processing speed allows digital image processing to be carried out in real time. Likewise, the development of memory allows analog images to be encoded into digital color images that are close to the original color [10]. Digital image processing is a scientific discipline that studies image processing techniques. The image referred to here is a still image (photo) or a moving image (which comes from a webcam). Meanwhile, digital here means that image processing is done digitally using a computer [11].

Image segmentation is the first stage carried out before the image analysis stage in the image recognition process from a particular input. The function of image segmentation is

to divide the image into regions based on similarities in the form of texture, color, shape, and so on. Common applications of image segmentation can be found in facial recognition applications, fruit quality detection, canned food industry, batik industry, and so on [12].

Strawberry is a fruit plant in the form of herbs that was first discovered in Chile, America. One species of strawberry plant is Fragaria Chiloensis L, which is spread to various countries in America, Europe and Asia. Furthermore, another species, namely Fragaria Vesca L, is more widespread than other species. This type of strawberry was also the first to enter Indonesia. Strawberry (Fragaria L) is a fruit commodity that is very popular with people around the world, including in Indonesia [13].

2.1. K-Means Algorithm

The following is how the K-Means algorithm partitions a data set into clusters [14], [15]:

1. The algorithm receives the number of clusters to group the data into, and the dataset to be clustered as input values.
2. The algorithm then creates K initial clusters (K= number of clusters required) from the dataset, while selecting K data records randomly from the dataset. For example, if there are 10,000 rows of data in the dataset and 3 clusters need to be formed, then the first initial K=3 clusters will be created by randomly taking 3 records from the dataset, as initial clusters. Each initial cluster formed has only one data record.
3. The K-Means algorithm calculates the arithmetic average of each cluster formed in the dataset. The average of a cluster is the average of all records contained in the cluster. Because in all the first K clusters there is only one record, the average is the average of that record. The average of a record is a collection of values that make up the record. For example, if in dataset S there is a record P that accepts values for the Height, Weight and Age fields, then we can write P = {Age, Height, Weight}. If John has Age = 20 years, Height = 1.70 meters and Weight = 80 Pounds then we write John = {20, 170, 80}. Because there is only one record in each initial cluster, the average of the cluster where John is located is = {20, 170, 80}.
4. Next, K-Means sends the K records in the dataset to only one of the initial clusters. The 4th to 6th records is passed to the nearest cluster (nearest cluster, namely the cluster that is very similar to that record) using a distance or similarity measure such as the Euclidean or Manhattan/City-Block distance measure.
5. K-Means recalculates the arithmetic mean of all clusters. The average of a cluster is the average of all records in that cluster. For example, if a cluster contains two records John = {20, 170, 80} and Henry = {30, 160, 120}, then the average P(average) is expressed as P(average) = {Age (average), Height(average), Weight(average)). Age(average) = (20 + 30)/2, Height(average) = (170 + 160)/2 and Weight(average) = (80 + 120)/2. The arithmetic mean of this cluster is {25, 165, 100}. This average becomes the center of this new cluster. Following the same procedure, new cluster centers are created for all existing clusters.
6. K-Means sends each record in the dataset again to only one of the newly formed clusters. A record or data points are passed to the nearest cluster, as before.
7. The previous steps are repeated until stable clusters are formed and the K-Means procedure is complete. Stable clusters are formed when iteration or repetition of K-Means does not create a new cluster as the center of the cluster or the arithmetic mean value of all new clusters is the same as the old cluster. There are several techniques to determine when a stable cluster is formed or when the K-Means algorithm ends.

2.2. Data Preparation

Before implementing the K-Mean Clustering method, first prepare strawberry image data for the training process and test data which will be converted into grayscale form. Then segmentation is carried out using k-mean thresholding to separate the object and image background. This aims to get the value of the object which will be clustered into the categories cooked, semi-cooked and raw. In the process of applying the manual count to identify the ripeness of strawberries, 3 sample images of strawberries were determined with the identification of maturity levels as follows:

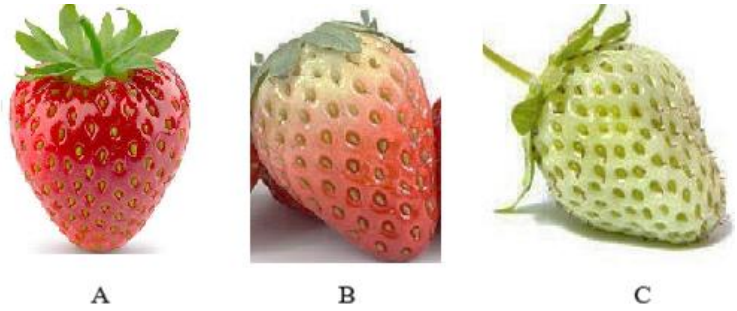
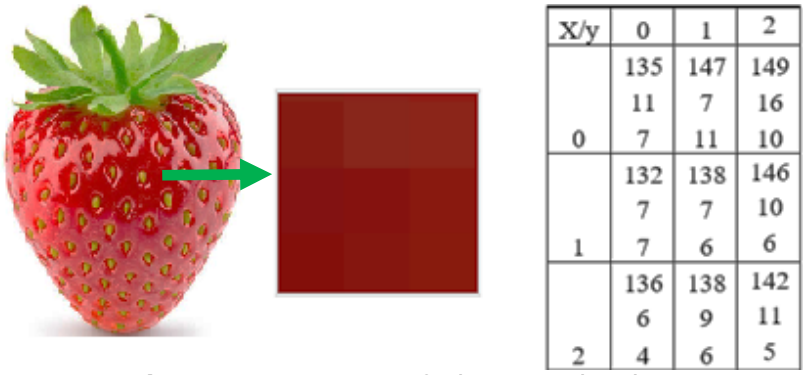


Figure 1. A=Ripe, B=Semi-Ripe, C=Raw

Based on the image above, the image of a ripe strawberry, an image of a half-ripe strawberry and an image of an unripe strawberry are known. To facilitate manual calculations in identifying the ripeness of strawberry fruit images using the K-Mean Clustering method, 3x3 pixel samples were taken from each image of ripe, semi-ripe and unripe. Here is the data:

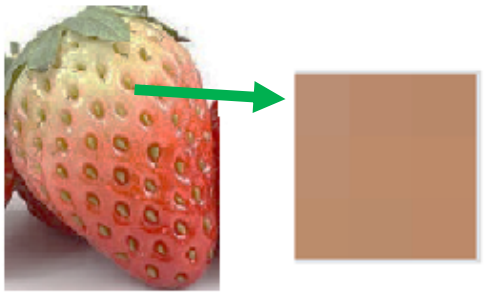
1. 3x3 Image of Ripe Strawberries
   The following is a 3x3 image for the ripe strawberry category used in the manual calculation of the training process:



X/y	0	1	2
0	135 11 7	147 7 11	149 16 10
1	132 7 7	138 7 6	146 10 6
2	136 6 4	138 9 6	142 11 5

Figure 2. 5x5 Image of Ripe Strawberries

   Based on Figure 2, it is known that the 3x3 pixel image of a ripe strawberry sample along with the RGB value for each pixel coordinate is obtained from the MATLAB application.
2. Image of 3x3 Half Ripe Strawberries
   The following is a 3x3 image for the half-ripe strawberry category used in the manual calculation of the training process:



X/y	0	1	2
	215	215	213
	153	145	142
0	107	93	88
	216	218	218
	152	148	146
1	100	90	88
	216	214	217
	150	149	145
2	92	91	88

Figure 3. Citra 5x5 Half Ripe Strawberries

Based on Figure 3, it is known that the image of a half-ripe strawberry sample is 3x3 pixels along with the RGB value for each pixel coordinate obtained from the MATLAB application.

3. 3x3 Image of Raw Strawberries
   The following is a 3x3 image for the raw strawberry category used in the manual calculation of the training process:

X/y	0	1	2
	248	252	251
	237	235	233
0	215	216	212
	249	250	253
	235	232	233
1	209	209	207
	249	249	254
	234	231	231
2	206	205	203

Figure 4. Image of 5x5 Raw Strawberries

Based on Figure 4, it is known that a 3x3 pixel image of a raw strawberry sample along with the RGB value of each pixel coordinate is obtained from the MATLAB application.

2.3. K-Means Clustering Training

The training process is a process where each strawberry image in the ripe, semi-ripe and unripe categories is converted into a grayscale image. The mature category is determined to be cluster 1, the half-cooked category is cluster 2 and the raw category is cluster 3. After obtaining the grayscale image, threshold segmentation is carried out to separate objects and background, so that the average (means) value of the image that will be clustered into clusters can be determined. mature images, half-cooked image clusters and raw image clusters. Here is the process:

1. *Clusters*1 Mature Image
   To find the average value for cluster 1 in the mature category, the RGB image value of a 3x3 pixel sample based on Figure 4.2 is converted into grayscale and thresholding k-means segmentation. The following are the RGB values of a 3x3 image in mature cluster 1:

Table 1. RGB Value 3x3 Image Cluster 1 Mature

X/y	0	1	2
	135	147	149
0	11	7	16
	7	11	10

X/y	0	1	2
1	132 7 7	138 7 6	146 10 6
2	136 6 4	138 9 6	142 11 5

Based on table 1, first look for the mean value of the RGB image. Here is how to find the RGB mean value:

Pixels (0,0) = $\frac{135+11+7}{3}\frac{135+11+7}{3}$ = 51

Pixels (0.1) = $\frac{147+7+11}{3}\frac{147+7+11}{3}$ = 55

Pixels (0.2) = $\frac{149+16+10}{3}\frac{149+16+10}{3}$ = 58

Pixels (1,0) = $\frac{132+7+7}{3}\frac{132+7+7}{3}$ = 49

Pixels (1,1) = $\frac{138+7+6}{3}\frac{138+7+6}{3}$ = 50

Pixels (1,1) = $\frac{146+10+6}{3}\frac{146+10+6}{3}$ = 54

Pixels (2,0) = $\frac{136+6+4}{3}\frac{136+6+4}{3}$ = 49

Pixels (2,1) = $\frac{138+9+6}{3}\frac{138+9+6}{3}$ = 51

Pixels (2,2) = $\frac{142+11+5}{3}\frac{142+11+5}{3}$ = 53

*MeansCluster*1

RGB = $\frac{51+55+58+49+50+54+49+51+53}{9} = 52$ (1)

Next is to change the image into grayscale form so that it can be segmented using k-means thresholding. Grayscale changes are carried out using the following equation:

$G(x,y) = (0.2989 \times R) + (0.5870 \times G) + (0.1140 \times B)$

From the equation above, the calculations for matrix (0,0) to matrix (2,2) are:

$G(x,y) = (0.2989 \times R) + (0.5870 \times G) + (0.1140 \times B)$ (2)

G(0,0) = (0. 2989× 135) + (0.5870× 11) + (0.1140× 7) =48

G(0,1) = (0. 2989× 147) + (0.5870× 7) + (0.1140× 11) =49

G(0,2) = (0. 2989× 149) + (0.5870× 16) + (0.1140× 10) =55

G(1,0) =(0. 2989× 132) + (0.5870× 7) + (0.1140× 7) = 44

G(1,1) =(0. 2989× 138) + (0.5870× 7) + (0.1140× 6) = 46

G(1,2) =(0. 2989× 146) + (0.5870× 10) + (0.1140× 6) = 50

G(2,0) =(0. 2989× 136) + (0.5870× 6) + (0.1140× 4) = 45

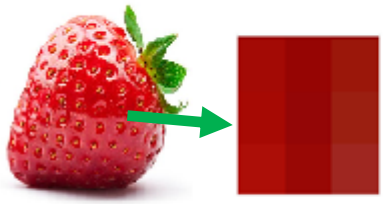
G(2,1) =(0. 2989× 138) + (0.5870× 9) + (0.1140× 6) = 47

G(2,2) =(0. 2989× 142) + (0.5870× 11) + (0.1140× 5) = 49

From the previous calculations, the results obtained are pixel values of 3 x 3 grayscale type which have been rounded.

2.4. K-Means Clustering Identification Testing

The following is an example of a test with the image of a test strawberry as follows:



X/y	0	1	2
0	163 10 5	150 5 3	155 23 10
1	166 8 3	148 6 6	155 31 7
2	172 14 7	153 8 5	159 10 6

Figure 5. Test Strawberry Image

Based on the image above, this manual testing was carried out on the same pixels during training, namely 3x3. The following is a summary of the 3x3 pixel values of the test strawberry image:

Table 2. RGB Value 3x3 Test Strawberry Image

X/y	0	1	2
0	163 10 5	150 5 3	155 23 10
1	166 8 3	148 6 6	155 31 7
2	172 14 7	153 8 5	159 10 6

Based on table 2, first look for the mean value of the RGB image. Here is how to find the RGB mean value:

Pixels (0,0) = $\frac{\frac{163+10+5}{3}}{3} = 59$

Pixels (0.1) = $\frac{\frac{150+5+3}{3}}{3} = 53$

Pixels (0.2) = $\frac{\frac{155+23+10}{3}}{3} = 63$

Pixels (1,0) = $\frac{\frac{166+8+3}{3}}{3} = 59$

Pixels (1,1) = $\frac{\frac{148+6+6}{3}}{3} = 53$

Pixels (1,2) = $\frac{\frac{155+31+7}{3}}{3} = 64$

Pixels (2,0) = $\frac{\frac{172+14+7}{3}}{3} = 64$

Pixels (2,1) = $\frac{\frac{153+8+5}{3}}{3} = 55$

Pixels (2,2) = $\frac{\frac{159+10+6}{3}}{3} = 58$

*Means*RGB test image = $\frac{59+53+63+59+53+64+64+55+58}{9} = 59$ (3)

The next step is to change the image into grayscale form so that it can be segmented using k-means thresholding. Grayscale changes are carried out using the following equation:

$G(x,y) = (0.2989 \times R) + (0.5870 \times G) + (0.1140 \times B)$ (4)

From the equation above, the calculations for matrix (0,0) to matrix (2,2) are:

G(x,y) =(0. 2989×R) + (0. 5870×G) + (0.1140×B)
G(0,0) = (0. 2989× 163) + (0.5870× 10) + (0.1140× 5) =55
G(0,1) = (0. 2989× 150) + (0.5870× 5) + (0.1140× 3) =48
G(0,2) = (0. 2989× 155) + (0.5870× 23) + (0.1140× 10) = 61
G(1,0) =(0. 2989× 166) + (0.5870× 8) + (0.1140× 3) = 55
G(1,1) =(0. 2989× 148) + (0.5870× 6) + (0.1140× 6) = 48
G(1,2) =(0. 2989× 155) + (0.5870× 31) + (0.1140× 7) = 65
G(2,0) =(0. 2989× 172) + (0.5870× 14) + (0.1140× 7) = 60
G(2,1) =(0. 2989× 153) + (0.5870× 8) + (0.1140× 5) = 51
G(2,2) =(0. 2989× 159) + (0.5870× 10) + (0.1140× 6) = 54

From the previous calculations, the results obtained are pixel values of 3 x 3 grayscale type which have been rounded.

3. Results and Discussion

Based on the results of the strawberry image maturity identification test with 6 test strawberry images consisting of 3 types of maturity levels, the following results were found:

Table 3. Citra Strawberry Ripeness Level Test Results

No	Image Name	Maturity Level	Identification Results	Results
1	matureuji1.jpg	Ripe	Ripe	True
2	matureuji2.jpg	Ripe	Ripe	True
3	halfcookeduji1.jpg	Half-baked	Half-baked	True
4	halfcookeduji1.jpg	Half-baked	Half-baked	True
5	rawuji1.jpg	Raw	Raw	True
6	rawuji2.jpg	Raw	Raw	True

Based on the test results table above, it was found that 6 test strawberry images were identified correctly without any errors. Next, calculate the level of accuracy based on the test results of 6 strawberry images. The formula is as follows:

$$Accuracy = \frac{True\ Classification\ Data}{The\ total\ of\ data} x\ 100\% \quad (5)$$

$$Accuracy = \frac{6}{6} x\ 100\% = 100\ \%$$

Based on the results of the accuracy test, an accuracy value of 100% was obtained for the process of identifying the maturity of strawberry images for 6 test image data.

4. Conclusion

Based on the results of tests carried out to identify the ripeness of Citra strawberry fruit, the following conclusions can be drawn:

1. Identification of the maturity of strawberry fruit images was successfully applied using the K-Means Clustering method on images of ripe, half-ripe and unripe strawberries.
2. The segmentation process using K-Means Clustering found 3 clusters with information on cluster 1 mature, cluster 2 half mature and cluster 3 raw. The cluster difference value is obtained from the means of the RGB image and the segmentation image. Based on testing identification of ripeness of strawberry fruit with test data of 6 images and training data of 12 images, it provides accuracy results of 100%.

To develop a better system in further research, the author can provide the following suggestions:

1. For better identification results, it is necessary to add training datasets and test images from this research so that we can see the level of identification of the K-Means Clustering method with a larger dataset.
2. There is a need to compare other identification methods to see the optimality of the K-Mean Clustering method.

References

[1] R. Utari, D. Soedibyo, and D. Purbasari, "Kajian Sifat Fisik Dan Kimia Buah Stroberi Berdasarkan Masa Simpan Dengan Pengolahan Citra," *J. agrotechnology*, vol. 120, no. 02, p. 138, 2018.

[2] A. Chan, P. Liem, N. Wong, and T. Gunawan, "Segmentasi Buah Menggunakan Metode K-Means Clustering dan Identifikasi Kematangannya Menggunakan Metode Perbandingan Kadar Warna," *JSM (SIFO Mikroskil Journal)*, vol. 15, no. 2, pp. 91–100, 2014.

[3] R. . Mayoza, "Menentukan Tingkat Kematangan Buah Pisang Dengan Segmentasi Warna Kulit Pisang Menggunakan K-Means," 2019.

[4] F. Febrinanto, C. Dewi, and A. Wiratno, "Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk," *J. Inf. Technol. Comput. Sci. Dev.*, vol. 2, no. 11, pp. 5375–5383, 2018.

[5] M. Furqan, A. Hasugian, and R. Tanjung, "Digital Image Enhancement Using The Method Of Multiscale Retinex and Median Filter," *INFOKUM J.*, vol. 9, no. 1, pp. 69–76, 2020.

[6] M. Furqan, Sriani, and I. E. Y. Sari, "Penerapan Metode Otsu Dalam Segmentasi Citra PAda Citra Naskah Arab," *J. Manaj. Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 1, pp. 59–72, 2019.

[7] P. Darma, *Pengolahan Citra Digital*. Yogyakarta: Penerbit Andi, 2010.

[8] Sriani, Triase, and Khairuna, "Pendekomposisian Citra Digital Dengan Algoritma DWT," *J. Ilmu Komput. Inform.*, vol. 1, no. 1, pp. 35–39, 2017.

[9] I. E. Y. Sari, M. Furqan, and S. Sriani, "Penerapan Metode Otsu dalam Melakukan Segmentasi Citra pada Citra Naskah Arab," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 1, pp. 59–72, 2020, doi: 10.30812/matrik.v20i1.658.

[10] S. R. Sulistiyanti, *Pengolahan Citra*. 2016.

[11] R. Kusumanto, "Technogenic activity of man and local sources of environmental pollution," *Stud. Environ. Sci.*, vol. 17, no. c, pp. 329–332, 2011.

[12] P. A.MA and M. Murinto, "Segmentasi Citra Batik Berdasarkan Fitur Tekstur Menggunakan Metode Filter Gabor Dan K-Means Clustering," *J. Inform.*, vol. 10, no. 1, 2016.

[13] R. R. Winardi, "Karakter Mutu Strawberry (Fragaria Virginiana) Selama Penyimpanan Dengan Perlakukan Edible Coating Campuran Sorbitol dan Pati Sagu," vol. 1, no. 140–149, 2AD.

[14] F. Anggraeny and S. Munir, "Segmentasi K-Means," 2019.

[15] V. Kumar and P. Gupta, "Importance of Statistical Measures in Digital Image Processing," *J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, 2012.