ILKOM Jurnal Ilmiah Vol. 14, No. 1, April 2022, pp. 25-31
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Detection system of strawberry ripeness using k-means

Dolly Indra [a,1,*]; Ramdan Satra [a,2]; Huzain Azis [a,3]; Abdul Rachman Manga' [a,4]; Harlinda L [a,5]
[a] *Faculty of Computer Science, Universitas Muslim Indonesia, Makassar and 90231, Indonesia*
[1] *dolly.indra@umi.ac.id;* [2] *ramdan@umi.ac.id;* [3] *huzain.azis@umi.ac.id;* [4] *abdulrachman.manga@umi.ac.id;* [5] *harlinda@umi.ac.id*
* *Corresponding author*

Article history: Received November 27, 2021; Revised April 25, 2022; Accepted April 25, 2022; Available online April 30, 2022

Abstract

Strawberry is one type of fruit that is favored by the people of Indonesia. The detection process of strawberries can be done by utilizing advances in computer technology such as digital image processing. This study develops a strawberry ripeness detection system using the values of Red, Green and Blue as the reference values. On the other hand, K-Means algorithm using the Euclidean distance difference as the reference will be used to determine the type of classification. The test on 51 strawberry images consisting of ripe, half-ripe and unripe fruit using the K-Means algorithm resulted in an accuracy rate of 82.14%. In addition, tests other than the strawberry image on 8 images resulted in an accuracy rate of 100%.

Keywords: strawberry; detection system; RGB; k-means; Euclidean distance

Introduction

Strawberry is one of the famous fruits which has been known as the "Queen of Fruits" [1]. This fruit is liked by most people because of its nutrition, sweet taste, sufficient water content, and affordable prices. Several studies on strawberries have been carried out in various ways such as Accurate Strawberry Plant Detection System based on Low-altitude Remote Sensing and Deep Learning Technologies which aimed to monitor in real-time [2]. Real-time Visual Localization of the Picking Points for a Ridge-planting Strawberry Harvesting Robot aims to harvest ripe strawberries via a robot [3]. A Strawberry Detection System using Convolutional Neural Networks which aimed to measure accuracy was implemented on the Raspberry Pi 3B [4].

The Strawberry Maturity was identified by applying Naïve-Bayes while the Image Processing Identification of Strawberry Maturity was conducted by using Naïve-Bayes and Image Processing which aims to extract the percentage values of RGB (Red, Green, and Blue) color parameters from strawberry fruit images. The color extraction results were then be classified by using Naive Bayes method [5]. Detection of Strawberry Flowers in Outdoor Field by Deep Neural Network aimed to estimate strawberry flower yields that are accurate, fast and reliable [6]. Strawberry Ripeness Identification using the RGB Extraction Feature and K-Nearest Neighbor aimed to classify strawberry images with an accuracy rate of 85% [7]. This study aims to classify the strawberry image using another method, in this case, the K-Means to categorize the strawberry image as the ripe, semi ripe and raw.

Method

The Strawberry Ripe Detection System proposed in this study is shown in **Figure 1**. This study preprocessed the strawberry image first by changing the image dimensions from 768 x 1024 to 128 x 128.

http://dx.doi.org/10.33096/ilkom.v14i1.1054.25-31

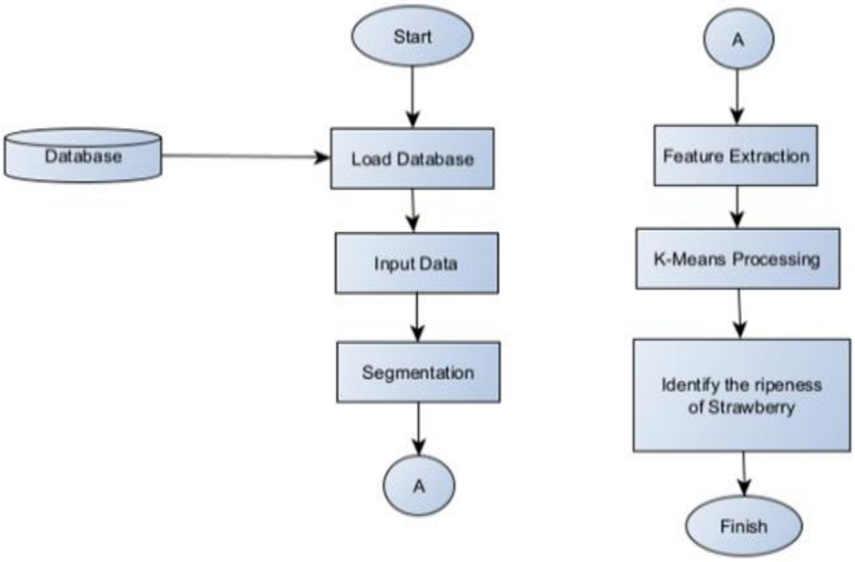

Figure 1. Proposed System

A. Input Data

The input image used in this study is a strawberry image which is classified into 3 classes, namely ripe, semi ripe, and raw as can be seen in Figure 2. The input image is in RGB form where the RGB color space consists of three color components, namely Red, Green and Blue [8]. Each component consists of 8 bits [9].

(a) (b) (c)

Figure 2. Strawberry image classification: (a) ripe, (b) semi ripe, (c) raw

The input image data are ripe strawberries of 16 images, semi ripe strawberries of 17 images, and raw strawberries of 18 images. These 51 images of strawberry images were used as training data.

B. Segmentation

Color segmentation is the process of separating areas in an image based on the colors contained in the image [10] which aims to separate the object area (foreground) from the background area for easy analysis [11]. This separation process is very important for content analysis and image understanding [12]. The Segmentation algorithm is shown as follows:

Step 1: Image data input
Step 2: Extract each component R, G, and B
R = Img(:,:,1); G = Img(:,:,2); B = Img(:,:,3);
Step 3: Perform image segmentation using the thresholding method
bw = im2bw(B);
Step 4: Set the background value to zero
R(bw) = 0; G(bw) = 0; B(bw) = 0;
Step 5: Perform image complement
bw = imcomplement(bw);
Step 6: Display the segmented binary image
imshow(bw);

Step 7: Display segmented RGB images
RGB = cat(3,R,G,B); imshow(RGB);

C. Feature Extraction

Feature extraction converts pixel data into representations of shapes, movements, colors, textures [13]. Feature extraction is an important stage in the construction of each pattern classification to obtain relevant information from the characteristics of each class so that it can be used in the next stage, namely pattern recognition in an image [14]. The feature extraction algorithm is shown as follows:

Step 1: Call variables bw, R, G, and B
bw = handles.bw; R = handles.R; G = handles.G; B = handles.B;
Step 2: Calculate the mean value of the R, G, and B components of the segmented image
Red = (sum(sum(R)))/(sum(sum(bw))); Green = (sum(sum(G)))/(sum(sum(bw)));
Blue = (sum(sum(B)))/(sum(sum(bw)));
Step 3: Performing morphological operations, namely filling holes and opening areas to remove objects that have an area of less than 500
bw = imfill(bw,'holes'); bw = bwareaopen(bw,500);
Step 4: Calculate the value of area, roundness, & slenderness of the segmented image
stats = regionprops(bw,'All');
luas = cat(1,stats.Area);
keliling = cat(1,stats.Perimeter);
lebar = cat(1,stats.MinorAxisLength);
panjang = cat(1,stats.MajorAxisLength);
[~,m] = max(luas);
luas = luas(m);
keliling = keliling(m);
lebar = lebar(m);
panjang = panjang(m);
kebulatan = (4*pi*luas)/(keliling^2);
kerampingan = lebar/panjang;
Step 5: Display color and morphological data
Step 6: Combining the mean values of R,G,B
Step 7: Combining morphological values

D. Identify Strawberry Ripeness

K-Means is a data clustering analysis method. The K-Means method tries to classify existing data into several groups which are included in the unsupervised learning method [15], where data in one group has different characteristics from data in other groups. Euclidean distance is used to measure the distance from 2 (two) points [16] or a special case of p-Minkowski distance which uses metrics to evaluate the difference between 2 (two) points [17], the closest calculation result will be the identification result. The identification algorithm is shown as follows:

Step 1: Calling the variable characteristics of color, morphological, area, roundness, slenderness
ciri_warna = handles.ciri_warna; ciri_morfologi = handles.ciri_morfologi;
luas_obj = ciri_morfologi(1); kebulatan_obj = ciri_morfologi(2);
kerampingan_obj = ciri_morfologi(3);
Step 2: Input test image
Step 3: load centroid
Step 4: Calculate the Euclidean distance between the data and each centroid
dist1 = mean(sqrt(sum(sum((ciri_warna-C(1,:)).^2))));
dist2 = mean(sqrt(sum(sum((ciri_warna-C(2,:)).^2))));
dist3 = mean(sqrt(sum(sum((ciri_warna-C(3,:)).^2))));
Step 5: Finding the closest distance
jarak = [dist1;dist2;dist3]; [m,n] = min(jarak); jarak_min = m;
Step 6: Changing the output value into an output class, namely “semi ripe”, “raw”, “ripe” and “unrecognized”

Step 7: Display the output class
Step 8: Initialize data_kmeans variabel variable
Step 9: Fill in the data_kmeans variable with data kmeans
Step 10: Display data_kmeans

Results and Discussion

The first stage conducted was the training process on the strawberry image which aimed to create a dataset that functions as training data or data reference. The data used for the training process were 51 images divided into 3 categories, namely Ripe, Semi ripe and Raw.

The training process for each strawberry images consisting of 16 images for ripe strawberries, 17 images of semi ripe strawberries, and 18 images of raw strawberries was shown by **Table 1, Table 2,** and **Table 3** respectively. The images use RGB format with .jpg extension.

Table 1. Training Process for Ripe Strawberries Images

No	Image	R	G	B	Ct R	Ct G	Ct B	Euclidean Distance
1	matang1.jpg	0.39782	0.21062	0.14452	0.4312	0.2089	0.1512	0.0341
2	matang2.jpg	0.44735	0.22395	0.14655	0.4312	0.2089	0.1512	0.0225
3	matang3.jpg	0.51606	0.2264	0.16279	0.4312	0.2089	0.1512	0.0874
4	matang4.jpg	0.45815	0.21452	0.15328	0.4312	0.2089	0.1512	0.0276
5	matang5.jpg	0.4935	0.23135	0.15949	0.4312	0.2089	0.1512	0.0667
6	matang6.jpg	0.50493	0.25281	0.16597	0.4312	0.2089	0.1512	0.0870
7	matang7.jpg	0.44436	0.2225	0.16152	0.4312	0.2089	0.1512	0.0215
8	matang8.jpg	0.49003	0.21628	0.14442	0.4312	0.2089	0.1512	0.0596
9	matang9.jpg	0.40561	0.21997	0.18665	0.4312	0.2089	0.1512	0.0451
10	matang10.jpg	0.36036	0.15116	0.11631	0.4312	0.2089	0.1512	0.0979
11	matang11.jpg	0.3766	0.18972	0.13256	0.4312	0.2089	0.1512	0.0608
12	matang12.jpg	0.4176	0.1788	0.17012	0.4312	0.2089	0.1512	0.0381
13	matang13.jpg	0.38171	0.19025	0.17604	0.4312	0.2089	0.1512	0.0585
14	matang14.jpg	0.38062	0.17594	0.14328	0.4312	0.2089	0.1512	0.0609
15	matang15.jpg	0.45002	0.21249	0.14032	0.4312	0.2089	0.1512	0.0220
16	matang16.jpg	0.44864	0.16968	0.13806	0.4312	0.2089	0.1512	0.0449

Table 2. Training Process for Semi Ripe Strawberries Images

No	Image	R	G	B	Ct R	Ct G	Ct B	Euclidean Distance
1	mengkal1.jpg	0.50853	0.29166	0.17474	0.4548	0.3219	0.1994	0.0664
2	mengkal2.jpg	0.48417	0.2693	0.16237	0.4548	0.3219	0.1994	0.0707
3	mengkal3.jpg	0.50209	0.28834	0.16285	0.4548	0.3219	0.1994	0.0685
4	mengkal4.jpg	0.46868	0.36292	0.19964	0.4548	0.3219	0.1994	0.0433
5	mengkal5.jpg	0.47557	0.28115	0.16404	0.4548	0.3219	0.1994	0.0578
6	mengkal6.jpg	0.48611	0.24913	0.15322	0.4312	0.2089	0.1512	0.0681
7	mengkal7.jpg	0.56865	0.36301	0.21071	0.4548	0.3219	0.1994	0.1216
8	mengkal8.jpg	0.49049	0.29215	0.17002	0.4548	0.3219	0.1994	0.0549
9	mengkal9.jpg	0.36739	0.30097	0.19156	0.3237	0.3008	0.2361	0.0624
10	mengkal10.jpg	0.46416	0.36036	0.20052	0.4548	0.3219	0.1994	0.0396
11	mengkal11.jpg	0.40634	0.3626	0.23772	0.4548	0.3219	0.1994	0.0740
12	mengkal12.jpg	0.41411	0.36764	0.24403	0.4548	0.3219	0.1994	0.0758
13	mengkal13.jpg	0.39078	0.31358	0.19672	0.4548	0.3219	0.1994	0.0646
14	mengkal14.jpg	0.421	0.29487	0.2001	0.4548	0.3219	0.1994	0.0433
15	mengkal15.jpg	0.43695	0.28779	0.19688	0.4548	0.3219	0.1994	0.0386
16	mengkal16.jpg	0.40634	0.3626	0.23772	0.4548	0.3219	0.1994	0.0740
17	mengkal17.jpg	0.38877	0.30764	0.1797	0.4548	0.3219	0.1994	0.0704

Table 3. Training Process for Raw Ripe Strawberries Images

No	Image	R	G	B	Ct R	Ct G	Ct B	Euclidean Distance
1	mentah1.jpg	0.34096	0.36417	0.37133	0.3237	0.3008	0.2361	0.1503
2	mentah2.jpg	0.33176	0.29506	0.29396	0.3237	0.3008	0.2361	0.0587
3	mentah3.jpg	0.37972	0.37093	0.37235	0.3237	0.3008	0.2361	0.1631
4	mentah4.jpg	0.33514	0.31199	0.29416	0.3237	0.3008	0.2361	0.0602

No	Image	R	G	B	Ct R	Ct G	Ct B	Euclidean Distance
5	mentah5.jpg	0.32465	0.30537	0.30802	0.3237	0.3008	0.2361	0.0721
6	mentah6.jpg	0.21404	0.32795	0.25869	0.3237	0.3008	0.2361	0.1152
7	mentah7.jpg	0.3353	0.2943	0.29262	0.3237	0.3008	0.2361	0.0580
8	mentah8.jpg	0.4508	0.34405	0.25281	0.4548	0.3219	0.1994	0.0580
9	mentah9.jpg	0.32197	0.281	0.17787	0.3237	0.3008	0.2361	0.0616
10	mentah10.jpg	0.3041	0.24958	0.14041	0.3237	0.3008	0.2361	0.1103
11	mentah11.jpg	0.33774	0.29145	0.18605	0.3237	0.3008	0.2361	0.0528
12	mentah12.jpg	0.28487	0.26355	0.17318	0.3237	0.3008	0.2361	0.0828
13	mentah13.jpg	0.32077	0.30825	0.22568	0.3237	0.3008	0.2361	0.0131
14	mentah14.jpg	0.33398	0.30382	0.19133	0.3237	0.3008	0.2361	0.0460
15	mentah15.jpg	0.32666	0.28795	0.18336	0.3237	0.3008	0.2361	0.0544
16	mentah16.jpg	0.34428	0.28632	0.16482	0.3237	0.3008	0.2361	0.0756
17	mentah17.jpg	0.30041	0.27152	0.18859	0.3237	0.3008	0.2361	0.0605
18	mentah18.jpg	0.30298	0.22445	0.12655	0.4312	0.2089	0.1512	0.1315

Each image of strawberries for all categories shown in the Table 1-3 strawberries was measured by the value of R (Red), G (Green) and B (Blue). The calculation process was then carried out using K-Means to calculate the value of each Centroid Red (Ct). R), Centroid Green (Ct G) and Centroid Blue (Ct B). Next, the closest distance was measured by using Euclidean distance. The results of the training setup was good, then the stage could be continued to the testing stage.

In the testing stage, a strawberry image that is different from the training image was used as shown in **Table 4**. The test results on the strawberry image produced an accuracy rate of 82.14% as used in equation (1). Our system is able to identify very well for the image of ripe strawberries and raw strawberries, however this not the case for the semi ripe strawberry images. This is because the image allows it to be in the ripe or raw category only.

$$Accuracy = \frac{Number\ of\ true\ identifikasition}{Total\ of\ Testing\ Data}\ x\ 100\% \quad (1)$$

Table 4. Testing Process for Stroberry Images

No	Testing Image	Observed	System	Identifation
1	ST1.jpg	Ripe	Ripe	True
2	ST2.jpg	Ripe	Ripe	True
3	ST3.jpg	Ripe	Ripe	True
4	ST4.jpg	Ripe	Ripe	True
5	ST5.jpg	Ripe	Ripe	True
6	ST6.jpg	Ripe	Ripe	True
7	ST7.jpg	Ripe	Ripe	True
8	ST8.jpg	Ripe	Ripe	True
9	ST9.jpg	Ripe	Ripe	True
10	ST10.jpg	Ripe	Ripe	True
11	ST11.jpg	Ripe	Ripe	True
12	ST12.jpg	Semi Ripe	Raw	False
13	ST13.jpg	Semi Ripe	Raw	False
14	ST14.jpg	Semi Ripe	Raw	False
15	ST15.jpg	Semi Ripe	Raw	False
16	ST16.jpg	Semi Ripe	Raw	False
17	ST17.jpg	Raw	Raw	True
18	ST18.jpg	Raw	Raw	True
19	ST19.jpg	Raw	Raw	True
20	ST20.jpg	Raw	Raw	True
21	ST21.jpg	Raw	Raw	True
22	ST22. jpg	Raw	Raw	True
23	ST23. jpg	Raw	Raw	True
24	ST24. jpg	Raw	Raw	True
25	ST25. jpg	Semi Ripe	Semi Ripe	True
26	ST26. jpg	Semi Ripe	Semi Ripe	True
27	ST27. jpg	Semi Ripe	Semi Ripe	True
28	ST28. jpg	Semi Ripe	Semi Ripe	True

Moreover, tests on not-strawberry images were also conducted as shown in **Table 5**. In this test, the system is able to identify these images as unrecognizable images with an accuracy rate of 100% as in equation (1).

Table 5. Testing Process for Not Stroberry Images

No	Testing Image	Observed	System	Identifation
1	apel1.jpg	Red Apple	Unrecognized	True
2	aple2.jpg	Green Apple	Unrecognized	True
3	ceri1.jpg	Cherry	Unrecognized	True
4	jambu.jpg	guava	Unrecognized	True
5	tomat.jpg	tomatoes	Unrecognized	True
6	handphone,jpg	handphone	Unrecognized	True
7	topi.jpg	hat	Unrecognized	True
8	laptop.jpg	laptop	Unrecognized	True

Conclusion

Based on the results of the tests on the strawberry images using the K-Means algorithm, an accuracy rate of 82.14% was obtained and the system was able to recognize the not-strawberry images with an accuracy rate of 100%. The system applied was not very good at recognizing the image of semi ripe strawberries leaving future studies to be carried out.

References

[1] L. Ma *et al.*, "Algorithm of Strawberry Disease Recognition Based on Deep Convolutional Neural Network," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/6683255.

[2] H. Zhang, P. Lin, J. He, and Y. Chen, "Accurate Strawberry Plant Detection System Based on Low-altitude Remote Sensing and Deep Learning Technologies," *2020 3rd Int. Conf. Artif. Intell. Big Data, ICAIBD 2020*, pp. 1–5, 2020, doi: 10.1109/ICAIBD49809.2020.9137479.

[3] Y. Yu, K. Zhang, H. Liu, L. Yang, and D. Zhang, "Real-Time Visual Localization of the Picking Points for a Ridge-Planting Strawberry Harvesting Robot," *IEEE Access*, vol. 8, pp. 116556–116568, 2020, doi: 10.1109/ACCESS.2020.3003034.

[4] N. Lamb and M. C. Chuah, "A Strawberry Detection System Using Convolutional Neural Networks," *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 2515–2520, 2019, doi: 10.1109/BigData.2018.8622466.

[5] H. Y. Riskiawan, T. Rizaldi, and D. P. S. Setyohadi, "Identification Strawberry Maturity using Naïve-Bayes and Image Processing," pp. 394–400.

[6] P. Lin and Y. Chen, "Detection of Strawberry Flowers in Outdoor Field by Deep Neural Network," *2018 3rd IEEE Int. Conf. Image, Vis. Comput. ICIVC 2018*, pp. 482–486, 2018, doi: 10.1109/ICIVC.2018.8492793.

[7] S. Anraeni, D. Indra, D. Adirahmadi, S. Pomalingo, Sugiarti, and S. H. Mansyur, "Strawberry Ripeness Identification Using Feature Extraction of RGB and K-Nearest Neighbor," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 395–398, 2021, doi: 10.1109/EIConCIT50028.2021.9431854.

[8] D. Indra, "Pendeteksian Tepi Objek Menggunakan Metode Gradien," *Ilk. J. Ilm.*, vol. 8, no. 2, pp. 69–75, 2016, doi: 10.33096/ilkom.v8i2.48.69-75.

[9] D. Indra and L. N. Hayati, "Skin Detection Using Color Distance Measurement And Thresholding," no. 5, pp. 1441–1443, 2019, doi: 10.35940/ijeat.E1209.0585C19.

[10] D. Indra, Purnawansyah, S. Madenda, and E. P. Wibowo, "Indonesian sign language recognition based on shape of hand gesture," *Procedia Comput. Sci.*, vol. 161, pp. 74–81, 2019, doi: 10.1016/j.procs.2019.11.101.

[11] D. Indra, S. Madenda, and E. P. Wibowo, "Feature Extraction of Bisindo Alphabets Using Chain Code Contour," *Int. J. Eng. Technol.*, vol. 9, no. 4, pp. 3233–3241, 2017, doi: 10.21817/ijet/2017/v9i4/170904142.

[12] D. Indra, T. Hasanuddin, R. Satra, and N. R. Wibowo, "Eggs Detection Using Otsu Thresholding Method," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIConCIT 2018*, no. 2, pp. 10–13, 2018, doi: 10.1109/EIConCIT.2018.8878517.

[13] A. Pisal, R. Sor, and K. S. Kinage, "Facial Feature Extraction Using Hierarchical MAX(HMAX) Method," *2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017*, no. figure 2, pp. 1–5, 2018, doi: 10.1109/ICCUBEA.2017.8463755.

[14] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," *Int. Conf. Adv. Comput. Commun. Technol. ACCT*, pp. 5–12, 2014, doi: 10.1109/ACCT.2014.74.

[15] S. Yuqing, Q. Junfei, and H. Honggui, "Structure design for RBF neural network based on improved K-means algorithm," *Proc. 28th Chinese Control Decis. Conf. CCDC 2016*, pp. 7035–7040, 2016, doi:

10.1109/CCDC.2016.7532265.
[16] S. E. Basri, D. Indra, H. Darwis, A. W. Mufila, L. B. Ilmawan, and B. Purwanto, “Recognition of Indonesian Sign Language Alphabets Using Fourier Descriptor Method,” *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 405–409, 2021, doi: 10.1109/EIConCIT50028.2021.9431883.
[17] R. Huang, C. Cui, W. Sun, and D. Towey, “Poster : Is Euclidean Distance the best Distance Measurement for Adaptive Random Testing ?,” pp. 406–409, 2020, doi: 10.1109/ICST46399.2020.00049.