Identification Strawberry Maturity using Naïve-Bayes and Image Processing

Wahyu Winarno[1], H Y Riskiawan[2], T Rizaldi[3], D P S Setyohadi[4]

[1]Department of Agricultural Production, Politeknik Negeri Jember, Indonesia
[2,3,4]Department of Information Technology, Politeknik Negeri Jember, Indonesia

[1]yufit@polije.ac.id, [2]taufiq_r@polije.ac.id

Abstract. Strawberry is one of the fruits that can grow and develop well in a country that has a tropical climate such as Indonesia, the interest of countries with moderate climate to strawberries makes this fruit become one of the export commodities that are quite popular. Indonesia has an average value of strawberry exports reaching 19 tons per year in 2009-2013. The high export value raises several problems, one of which is the selection of strawberry fruit to produce quality products. The selection process carried out with visual perception raises the problem of accuracy to measure the maturity of strawberry fruit. The solution that can be applied to overcome these problems is to use image processing. This study extracts percentage values of RGB (Red, Green, and Blue) color parameters from the image of strawberry fruit. The color extraction results are then classified with the help of the Naive Bayes method to determine whether the strawberries are ripe or not.

Keyword-Classification, Image processing, Naive Bayes, Strawberry.

1. Introduction

The average value of Indonesian exports for strawberries reached 19 tons per year in 2009-2013 making strawberry one of the export commodities that has a pretty good market potential. Strawberry fruit can be a source of foreign exchange if strawberry fruit cultivation is developed, in addition to cultivating fruit, another thing that must be considered is maintaining the quality of the fruit. One of the things that affect the quality of strawberry juice is an accurate selection process so that the strawberries are really ripe to be thrown into the market. The rapid development of information technology allows for the identification and selection of strawberries by utilizing color parameters.

The use of RGB color parameters that are converted into HSI color parameters is applied to classify tomatoes according to the Tomate maturity index (TMI) [1]. The application of the HSI color feature combined with the nearest neighbors algorithm produces an accuracy rate of 85% - 90% in the classification of apples [2]. The level of accuracy of the classification of bananas based on their level of maturity by using color intensity algorithm and area algorithm reaches 99.1% - 85% [3]. The application of image processing and K-Means algorithm in utilizing machine vision technology that is used to determine the level of maturity of strawberries based on the fruit color results in an accuracy of 88.8% [4].

In this study, the color features R, G, B (Red, Green, Blue) are used to do a comparison with the training data where the comparison parameter is full green and full red. Picture of strawberry fruit with sweet charlie type for extraction process is obtained on the conditions and time of image take that has been determined to produce accurate data. Comparative results based on color parameters are then classified using the naïve Bayes algorithm by utilizing training data that has been stored in the database and compared with new input data to get the classification results whether the fruit is mature or not.

2. Material and Methods

Classification to determine whether strawberries enter the mature class or not determined based on the extraction of RGB (Red, Green, Blue) color format. The image of the strawberry fruit used is the fruit image with a resolution of 375x500 pixels. Preliminary data obtained from the extraction of strawberry images into RGB format and are continually changed to discrete data, with the data type change rule if the color value is between 147-220 then the value will change to a high state, if the value is green 74 -146, the value will be changed to the middle state and if the blue value 1-73 will be changed to a low state [5].

3. *Sample collection*

To get the image results that can be detected for color features, data retrieval using a camera with 8Mp resolution and image retrieval is done using a white background [6] with a distance of up to 20cm and bright lighting conditions. The results of the image retrieval process of strawberries are stored using bitmap or bmp format to facilitate the introduction of each pixel contained in the image of strawberry fruit [7]. The training data was used as many as 20 pieces of data, in which the training data was divided into 2 classes namely mature and raw with a range of colors is fully green and fully red as shown in figure 1. The strawberry fruit variety used was sweet Charlie variety.

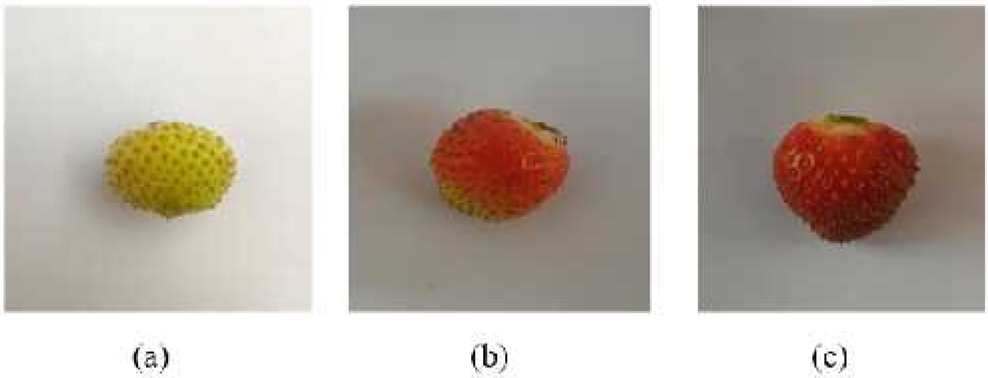

Figure 1. Color changes on strawberry a) full green, b) green red, c) full red

4. *Naïve Bayes*

The classification conducted by Naïve Bayes is a classification of opportunity-based statistics. The basis of the classification is training data and conditional probability. The training data obtained from the image of strawberry fruit is still continuous, to be able to do the classification process, the data needs to be changed into discrete data. Continuous data obtained from the results of extraction and reading

RGB values on the image of strawberries are shown in Table I, there are 2 class M for Mature and R for Ripe.

Table 1. Training Dataset (Continuous)

Data	R	G	B	Class
1	144	19	17	M
2	167	32	28	M
3	150	27	29	M
4	137	27	26	M
5	126	23	26	M
6	147	36	19	M
7	143	21	16	M
8	148	43	39	M
9	129	20	25	M
10	194	47	27	M
11	163	47	47	M
12	141	120	77	R
13	215	214	210	R
14	215	212	205	R
15	135	100	8	R
16	214	187	129	R
17	156	131	45	R
18	213	211	206	R
19	206	180	120	R
20	151	125	40	R

Changes from training data sets that become continuous into discrete training data sets are shown in Table 2.

Table 2. Training Dataset (Discrete)

NO	R	G	B	Class
1	Medium	Low	Low	Y
2	High	Low	Low	Y
3	High	Low	Low	Y
4	Medium	Low	Low	Y
5	Medium	Low	Low	Y
6	High	Low	Low	Y
7	Medium	Low	Low	Y
8	High	Low	Low	Y
9	Medium	Low	Low	Y
10	High	Low	Low	Y
11	High	Low	Low	Y
12	Medium	Medium	Medium	N
13	High	High	High	N

14	High	High	High	N
15	Medium	Medium	Low	N
16	High	High	Medium	N
17	High	Medium	Low	N
18	High	High	High	N
19	High	High	Medium	N
20	High	Low	Low	N

From discrete training datasets, the naïve Bayes method is used. The formula from naïve Bayes is

$$P(a_i|V_j)=\frac{Nc+m*p}{n+m}\quad\ldots\ldots(1)$$

where:

N = number of state where V = Vj

Nc = number of correct state on the attribute

p = $\frac{\text{amount of data available}}{\text{amount of attributes}}$

m = number of attributes

5. Result

To get a classification of what a strawberry fruit image into the class Mature or Ripe, conducted a comparison process between test data with training data that has been stored in the database. Test image data entered into the application to get each pixel value of RGB colour that shown in Figure 2.

Figure 2. Display data value determination continue one of the image of strawberry fruit.

RGB value that has been obtained is still as data continue so it needs to be converted into data with the discrete type. In Table 3, the test data has been converted to discrete data.

Table 3. Change of Continue data TYPE into Discrete Data type

Data	RGB	Value	Class
1	Red	High	
2	Green	Low	?
3	Blue	Low	

The discrete test dataset is matched with the training dataset if the value of the dataset is the same as the value of the training dataset, the resulting value is correct. If not, the resulting value is false as shown in Table 4. The next step is to calculate the same attributes as table 5.

Table 4. Result between 1st Test Data and Train Dataset

Data	R	G	B
1	False	False	True
2	True	True	True
3	True	True	True
4	False	True	True
5	False	True	True
6	True	True	True
7	False	True	True
8	True	True	True
9	False	True	True
10	True	True	True
11	True	True	True
12	False	False	False
13	False	False	False
14	False	False	False
15	False	False	False
16	False	False	False
17	False	False	False
18	False	False	False
19	False	False	False
20	False	False	False

Table 5. Matching Value

<table>
<tr><td colspan="2"></td><td>R</td><td>G</td><td>B</td></tr>
<tr><td rowspan="4">Yes</td><td>N</td><td>13</td><td>11</td><td>14</td></tr>
<tr><td>n_c</td><td>6</td><td>11</td><td>11</td></tr>
<tr><td>P</td><td>0,5</td><td>0,33</td><td>0,33</td></tr>
<tr><td>M</td><td>3</td><td>3</td><td>3</td></tr>
</table>

No	N	13	11	14
	n_c	7	0	3
	P	0,5	0,33	0,33
	M	3	3	3

Then from the result in Table 5, the values for V(yes) and V(no) to be calculated.

V(Yes):

P(R|Yes) $= \frac{n_c + m * p}{n + m} = \frac{6 + 3 * 0{,}5}{13 + 3} = 0{,}2812$

P(G|Yes) $= \frac{n_c + m * p}{n + m} = \frac{11 + 3 * 0{,}33}{11 + 3} = 0{,}33$

P(B|Yes) $= \frac{n_c + m * p}{n + m} = \frac{11 + 3 * 0{,}33}{14 + 3} = 0{,}2717$

Then, V(Yes) = P(Ya) * P(R|Yes) * P(G|Yes) * P(B|Yes)

V(Yes) = 0,0126

V(No):

P(R|No) $= \frac{n_c + m * p}{n + m} = \frac{7 + 3 * 0{,}5}{13 + 3} = 0{,}3125$

P(G|No) $= \frac{n_c + m * p}{n + m} = \frac{0 + 3 * 0{,}33}{11 + 3} = 0{,}0707$

P(Be|No) $= \frac{n_c + m * p}{n + m} = \frac{3 + 3 * 0{,}33}{14 + 3} = 0{,}1164$

Then, V(No) = P(No)*P(R|No)*P(G|No)*P(B|No)

V(No) = 0,0012

V(Yes) = 0,0126 > V(No) = 0,0012

The result shows that value V(Yes) > Value from V(No), so we can get the conclusion that data number 1 class is Yes or M (Mature) class. Table 6 shows the result from comparison the result that gets using the application comparing the result from an actual dataset, from that result it shows that accuracy from the application that uses RGB and Naïve Bayes is 80%.

Table 6. Comparison Result Between Real Result and Application Result

Data No	Real Result	Application Result
1	Mature	Mature
2	Mature	Ripe*
3	Ripe	Ripe
4	Mature	Mature

5	Mature	Ripe*
6	Ripe	Ripe
7	Ripe	Ripe
8	Ripe	Ripe
9	Ripe	Ripe
10	Mature	Mature

6. Conclusion

From the test results using 20 training datasets and 10 test datasets, the application of the naive Bayes method and RGB color parameters is obtained by an accuracy rate of 80%. The division of RGB color pixel groups into high, medium and low will cause errors that occur at the level of accuracy if the difference in red and green color values is not significant. The use of other color features such as CMYK or HSI can be used to increase the accuracy.

Acknowledgment

The researcher would like to express the gratitude and appreciation to Department of Information Technology, Politeknik Negeri Jember, who has helped support for this research.

References

[1] Choi, K., G.Lee, Y.J. Han, J. M. Bunn. 1995. TOMATO MATURITY EVALUATION USING COLOR IMAGE ANALYSIS. Transactions of the ASAE. 38(1): 171-176. American Society of Agricultural and Biological Engineers, St. Joseph: Michigan

[2] Arifianto, AS., Taufiq Rizaldi, Muhammad Yuris. 2016. Klasifikasi Citra Buah Apel Batu Menggunakan Teknik Digital Image Processing Berdasarkan Fitur Warna HIS. Seminar Hasil Penelitian dan Pengabdian Masyarakat 2016 Politeknik Negeri Jember.

[3] Prabha, D. Surya & J. Satheesh Kumar. 2013. Assessment of banana fruit maturity by image processing technique. Journal of Food Science and Technology March 2015, Volume 52, Issue 3, pp 1316–1327.

[4] Liming, Xu & Zhao Yanchao. 2010. Automated strawberry grading system based on image processing. Computers and Electronics in Agriculture Volume 71, Supplement 1, April 2010, Pages S32-S39.

[5] Riskiawan, HY., Taufiq Rizaldi, Dwi Putro S Setyohadi, Tri Leksono. 2017. Nitrogen (N) fertilizer measuring instrument on maize-based plant microcontroller. 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).

[6] Ayuninghemi, R. and Setyohadi, D.P.S., 2015, October. Optimum Feature for Palmprint Image Authentication. In *Prosiding International conference on Information Technology and Business (ICITB)* (pp. 134-138).

[7] Destarianto, P., Riskiawan, H.Y., Agustianto, K. and Kautsar, S., 2017, November. Developing food sensory test system with preference test (Hedonic and Hedonic quality) wheat bread case study. In *Sustainable Information Engineering and Technology (SIET), 2017 International Conference on* (pp. 430-434). IEEE.