Classification of total carotene and quality of chili pepper (*Capsicum frutescens*) based on image analysis

Y Hendrawan¹, R Utami¹, D Y Nurseta¹, Daisy², S Nuryani¹, D M Maharani¹ and Sandra¹

¹Department of Agricultural Engineering, Universitas Brawijaya, Malang, Indonesia
²Department of Agricultural Product Technology, Universitas Brawijaya, Malang, Indonesia

E-mail: yusuf.h@ub.ac.id

Abstract. The process of sorting the chili pepper (*Capsicum frutescens*) is done physically by using the human eye, based on the visual color uniformity of its skin. This method is not effective and efficient. This study aims to identify the total content of carotene in chili pepper using the color and textural features approach. Color feature extraction used is the value of RGB, HSV, HSL, XYZ, CMY, CMYK, Lab, LUV, LCH, grey color, and 10 textural features from each color-space. The feature extraction results were used to identify the total carotene content by the image analysis method. The image of chili peppers used was 360, consisting of 300 training data and 60 test data. Classification test results with a level of 20% produce the best parameters as an indicator of total carotene chili pepper i.e. hue mean features with a range for quality A (55.04> hue mean > 32.15), quality B (19.80> hue mean> 12.21), quality C (3.55> hue mean> 1.93), and 80% accuracy using the confusion matrix and mean square error (MSE).

1. Introduction
The problem of large-scale production of horticultural commodities is the problem of post-harvest in terms of grading. So far, the chili pepper grading process is only based on physical quality. The process of grading agricultural products is generally still based on uniformity in colour, uniformity in shape, uniformity in size, and the degree of damage [1]. The method of grading in this way depends on the perception of the grader, so the results are less accurate and vary depending on the perception of each grader. Furthermore, chili pepper's grading process has not been based on its nutritional content. In fact, the maturity level of chili pepper affects the nutrients content including total carotene [2], but so far to identify the total carotene content in chili pepper is still being carried out destructively which is expensive and requires a long time. Therefore, accurate, effective, and non-destructive grading methods are needed. One alternative method that can be used is the grading method using image analysis [3-5].

Several studies that have been carried out using image analysis to identify food products include the classification of olives with colour and morphological features that results in an accuracy of 97% in training data, and 75% in validation data [6]; the determination of carica papaya fruit by using area, shape, and textural features which produce an accuracy of 94.04% [7]; the classification of tomatoes using image features based on the fruit damage and the level of maturity which results in an accuracy of 96.47% [8]; the classification of watermelon by using size and shape features with an accuracy of 97%
the determination of mangoes fruit by using geometrical features include area, perimeter, and roundness with an accuracy of 97% (area), 79% (perimeter), and 36% (roundness) [10]; the classification of chili seeds based on colour, shape, and textural features with an accuracy of 84.94% [11]; brazilian chile classification using morphological features that results in an accuracy of 100% [12].

Carotene is an important component in chili pepper which functions as an antioxidant for the human body [13]. The level of carotene in chili pepper is influenced by the redness of the chili, where the reddish colour, the higher the carotene content, so that the colour factor of the chili when harvested will determine the total level of carotene. The main carotene of red chili is capsanthin which represents 35% of the total carotene [14]. Low levels of capsanthin in the early stages of the growth of chili plants will increase gradually until the chili reaches a mature stage [15].

Chili pepper has nine maturity indexes i.e. index 1 with light green skin colour, index 2 with brighter green skin colour, index 3 with yellowish skin colour, index 4 with yellowish and shiny skin colour, index 5 with yellowish and white skin colour, index 6 with slightly orange and white skin colour, index 7 with orange skin that exceeds white, index 8 with reddish and shiny orange skin, and index 9 with dark red skin. The maturity level can be quantified using the colour chart described by Noichinda [16]. This study aims to identify and classify the chili pepper quality index based on total carotene content using image analysis methods.

2. Materials and Methods
This research method included (1) chile pepper image data acquisition; (2) measurement of total carotene levels of chili pepper using the chromatography method; (3) image analysis which included RGB, HSV, HSL, XYZ, CMYK, CMYK, Lab, Luv, LCH, and grey colour features and 10 textural features on each colour-space, so there are a total of 286 image parameters; (3) identification of the relationship of 286 image parameters consisting of colour and textural features to the chili pepper quality index based on total carotene content; (4) tested the equation model by using a confusion matrix. This study used three classes of chili pepper quality with different maturity indexes i.e. quality A (index 3), quality B (index 6), and quality C (index 9). The research method can be seen in Figure 1.

Figure 1. Image analysis research method.

Chili pepper was captured using a scanner (EPSON L360, 600 × 1200 dpi). Image acquisition was done on one side of the chili for each sample of chili pepper. Each quality group consisted of 120 samples, so the total sample of images in this study were 360 samples which were then divided into 300 data for training and 60 data for validation. The image was saved in the bitmap (.bmp) format. Measurement of total carotene content was carried out using the high-pressure liquid chromatography (HPLC) method. Equation 1 was used to determine the total carotene content [17]:

\[
Total\ carotene\ (\mu g\ g^{-1}) = \frac{A \times V \times 10^4}{A_{1\ cm}^\% \times P}
\]

where: A = absorbent; V = total volume of sample extract (ml); P = sample weighted (g); \(A_{1\ cm}^\% = 2592\) (B-carotene coefficient in Petroleum Ether).
The extraction stage of the chili pepper image parameters was performed with a computer program that had been developed with Visual Basic 6.0. The self-built program of feature extraction had the ability to extract colours from RGB and convert RGB colours to hue-saturation-value (HSV), hue-saturation-lightness (HSL), Lab, XYZ, CMY, CMYK, LCH, Luv, and grey. From all colour-space, textural features were calculated. So that the total image features obtained were 286 features [18]. From 286 image features, several image features can be selected to classify the chili pepper quality index accurately.

3. Results and Discussion

Measurement of total carotene was performed five times for each chili pepper quality index. Based on the measurement of total carotene, each class of chili pepper quality had a different total carotene content. The total carotene content for each chili pepper category can be seen in Figure 2.

![Figure 2](image)

Figure 2. (a) the total carotene content in each chili pepper quality index; (b) the relationship of the chili pepper index to the total carotene.

Based on figure 2(a), it is shown that the category of quality A had a total carotene content ranging from 0.5 to 2 μg/g, the quality category B had a total carotene content ranging from 9-30 μg/g, whereas for the quality category C, it had a total carotene content ranging between 100-160 μg/g. Next, the total carotene in each category was calculated on average. The average value of total carotene in each chili pepper quality and the relationship of the chili pepper index to total carotene can be seen in figure 2(b).
Based on figure 2(b), it was known that the total carotene content increased in each quality, that was from quality A increased to the quality B and increased again to quality C. Based on the linear regression test, a linear equation was obtained to show the correlation between total carotene content and the quality index of chili pepper, the equation was \( y = 57.37x - 69.15 \) with a correlation coefficient of 0.863. The correlation coefficient value showed that the relationship between independent and dependent variables was very strong because the correlation between variables was perfect. The correlation value \((R)\) which was in the range 0.8-1 showed a very strong correlation.

Figure 2 shows that quality A had an average total carotene content of 1.39 ± 0.30 μg/g, the quality B had an average total carotene content of 19.23 ± 7.76 μg/g, and the quality C had an average total carotene content of 116.13 ± 2 μg/g. This data showed that the amount of total carotene content in chili pepper was influenced by its maturity index. The higher the maturity index of chili pepper, the value of the total carotene content was also greater [19]. Based on data analysis that had been done regarding the relationship of digital image features with the quality index of chili pepper, using 286 image features, there were 18 features can distinguish all quality index (A&B, A&C, B&C), 29 features were able to distinguish two pairs of the quality index (A&B and A&C), 21 features were able to distinguish two pairs of the quality index (A&B and B&C), 26 features were able to distinguish two pairs of the quality index (A&C and B&C), 30 features were able to distinguish one pair of the quality index (A&B), 31 features were able to distinguish one pair of the quality index (A&C), 14 features were able to distinguish one pair of the quality index (B&C), and 117 features were not able to distinguish the chili pepper quality index. This inability to distinguish was due to the standard deviation of the image features values between the two-quality index of chili pepper coinciding or overlapping. Of the 286 image features consisting of colour features and textural features that can be used as indicators of chili pepper determination, there were 18 image features i.e. hue mean, \( H_{(LCH)} \) mean, green sum mean, grey sum mean, hue sum mean, \( Y_{(XYZ)} \) sum mean, \( M_{(CMY)} \) sum mean, \( L_{(Lab)} \) sum mean, \( a_{(Lab)} \) sum mean, \( H_{(LCH)} \) sum mean, \( u_{(Lab)} \) sum mean, hue variance, \( a_{(Lab)} \) variance, \( u_{(Lab)} \) variance, \( V_{(Lab)} \) variance, and hue cluster tendency. From 18 image features which can differentiate all of chili quality index, five best image features were then selected and tested, the results showed high accuracy values as shown in figure 3, including hue mean, \( H_{(LCH)} \) mean, \( V_{(Lab)} \) variance, \( a_{(Lab)} \) cluster tendency, and \( a_{(Lab)} \) variance.

\[
\text{normalized hue mean} = \frac{(\text{hue mean}-1.117752)}{(93.09783-1.117752)} \tag{2}
\]

\[
\text{normalized } H_{(LCH)}\text{mean} = \frac{(214.379-37.64878)}{(600.7512-23.74606)} \tag{3}
\]

\[
\text{normalized } V_{(Lab)}\text{variance} = \frac{(a_{(Lab)}\text{cluster tendency}-23.74606)}{(150.371-6.97932)} \tag{4}
\]

\[
\text{normalized } a_{(Lab)}\text{cluster tendency} = \frac{(a_{(Lab)}\text{variance}-6.97932)}{(600.7512-23.74606)} \tag{5}
\]

\[
\text{normalized } a_{(Lab)}\text{variance} = \frac{(a_{(Lab)}\text{variance}-6.97932)}{(150.371-6.97932)} \tag{6}
\]

The hue mean feature was then normalized with a range value from 0 to 1. Based on the 300 data, the hue mean feature had a maximum value of 93.098 and a minimum value of 1.118. Based on the maximum and minimum values, the normalization results were determined based on the normalization equation. Equation 2 was the normalization equation for the hue mean feature. The pattern of the hue mean graph showed that the higher the quality of chili pepper, the hue mean value decreased, this was because the hue colour index had a colour angle range of 0–360°. A value of 0° represents a red colour, so the redder the colour of an image, the smaller the value. Quality A to quality B decreased by 0.30 (from 0.46 to 0.16) and from quality B to quality C decreased by 0.14 (from 0.16 to 0.02). The highest data distribution was quality A of 0.25 followed by quality B of 0.08 and the lowest was quality C of 0.02. The hue mean feature can be used as a parameter in determining the quality category of chili pepper because the resulting data distribution was not overlapping, and the error rate was relatively small 27.3%.

Besides the hue mean feature, the other image features that can be used to determine the chili pepper quality index was the \( H_{(LCH)} \) mean. Based on the 300 data, the \( H_{(LCH)} \) mean feature had a maximum value
of 214.18 and a minimum value of 37.65. Equation 3 described the normalization calculation of $H_{(LCH)}$ mean. The $H_{(LCH)}$ mean graph pattern showed the same pattern as the hue mean graph in which the decreasing quality of chili pepper followed by the decreasing of the $H_{(LCH)}$ mean value. Quality A to quality B decreased by 0.29 (from 0.52 to 0.23). Quality B to quality C decreased by 0.22 (from 0.23 to 0.01). The highest data distribution was quality B of 0.29 followed by quality A of 0.22 and the lowest was quality C of 0.01. The $H_{(LCH)}$ mean feature can also be used for chili pepper quality determination because there was no overlapping in the distribution of the data and had a relatively small error rate of 26.3%.

The $V_{(Luv)}$ variance feature had a maximum value of 16546.61 and a minimum value of 502.204. Equation 4 described the normalization calculation for the $V_{(Luv)}$ variance feature. The higher the quality of chili pepper, the $V_{(Luv)}$ variance value also increased, this was because the colour value of $V_{(Luv)}$ in an image shows the level of brightness. The range of $V_{(Luv)}$ was 0-1, where $V_{(Luv)} = 0$ indicated that the colour had no light, while $V_{(Luv)} = 1$ indicated the colour had the maximum light. Quality A to quality B increased by 0.627 (from 0.021 to 0.648) and from quality B to quality C increased by 0.337 (from 0.648...
to 0.985). The highest data distribution on quality B was 0.288 followed by quality A of 0.035 and the lowest was quality C of 0.016. The parameter $v_{\text{LUV}}$ variance can be used to determine the chili pepper quality index because there was no overlapping in the distribution of the data and had a relatively small error rate of 24.67%.

The a$_{\text{Lab}}$ cluster tendency feature had a maximum value of 600.751 and a minimum value of 23.746. Equation 5 was the normalization calculation for a$_{\text{Lab}}$ cluster tendency feature. The pattern of a$_{\text{Lab}}$ cluster tendency graph was fluctuating, it was increasing and decreasing. The increased pattern occurred from quality A to quality B by 0.371 (from 0.193 to 0.564), while from quality B to quality C decreased by 0.541 (from 0.564 to 0.023). The highest data distribution on quality B was 0.445 followed by quality A of 0.155 and the lowest was quality C of 0.024. The a$_{\text{Lab}}$ cluster tendency feature can be used to determine the quality index of chili pepper because the resulting data distribution had no overlapping data and had a relatively small error rate of 19%.

The a$_{\text{Lab}}$ variance feature had a maximum value of 150.371 and a minimum value of 6.979. Equation 6 was the normalization calculation for a$_{\text{Lab}}$ Variance feature. The graph pattern of a$_{\text{Lab}}$ Variance was fluctuating. The increased pattern occurred from quality A to quality B that was 0.373 (from 0.193 to 0.565), while from quality B to quality C decreased by 0.543 (from 0.565 to 0.023). The highest data distribution on quality B was 0.444 followed by quality A of 0.155 and the lowest was quality C of 0.024. The a$_{\text{Lab}}$ variance feature can be used to determine the quality index of chili pepper because there was no overlapping in the distribution of data and had a relatively small error rate of 20%.

$$X_d = X_n \times (93.09783 - 1.117752) + 1.117751 \quad (7)$$
$$X_d = X_n \times (214.179 - 37.64878) + 37.64878 \quad (8)$$
$$X_d = X_n \times (16546.61 - 502.2044) + 502.2044 \quad (9)$$
$$X_d = X_n \times (600.512 - 23.74606) + 23.74606 \quad (10)$$
$$X_d = X_n \times (150.317 - 6.979732) + 6.979732 \quad (11)$$

The explanation of the denormalization equation on the five best image features can be seen in equations 7-11. While the error rate and mean square error (MSE) results of the identification test of the chili pepper quality index can be seen in Table 1.

| No. | Image features        | The error of validation data | MSE  |
|-----|-----------------------|-----------------------------|------|
| 1   | hue mean              | 20.00 %                     | 0.20 |
| 2   | H$_{\text{LCH}}$ mean | 40.00 %                     | 0.40 |
| 3   | $v_{\text{LUV}}$ variance | 41.67 %                 | 0.42 |
| 4   | a$_{\text{Lab}}$ cluster tendency | 46.67 %                  | 0.47 |
| 5   | a$_{\text{Lab}}$ variance | 46.67 %                  | 0.47 |

### 4. Conclusions

The quality index of chili pepper was developed based on different maturity indexes i.e. quality A (index 3), quality B (index 6), and quality C (index 9). The total carotene content in the three chili pepper quality index ranged from 0.5-120 μg/g, where the highest total carotene content was quality C of 152.22 μg/g and the lowest was quality A of 0.9 μg/g. Of the 286 digital image features, consisting of colour and textural features that can be used as indicators of chili pepper quality determination, there were 18 image features i.e. hue mean, H$_{\text{LCH}}$ mean, green sum mean, gray sum mean, hue sum mean, $Y_{\text{XYZ}}$ sum mean, M$_{\text{CMY}}$ sum mean, L$_{\text{Lab}}$ sum mean, a$_{\text{Lab}}$ sum mean, H$_{\text{LCH}}$ sum mean, $v_{\text{LUV}}$ sum mean, hue variance, a$_{\text{Lab}}$ variance, $u_{\text{LUV}}$ variance, V$_{\text{LUV}}$ variance, and hue cluster tendency. Of the best 18 image features, after testing the validation data, the best image feature as chili pepper classification was hue mean with a range for quality A (55.04 > hue mean > 32.15), quality B (19.80 > hue mean > 12.21),
quality C (3.55 > hue mean > 1.93), and the resulting accuracy was 80% using the confusion matrix and mean square error (MSE) methods.

References

[1] Hendrawan Y, Amini A, Maharani D M, Sutan S M 2019 Intelligent Non-Invasive Sensing Method in Identifying Coconut (Coco nucifera var. Ebunea) Ripeness Using Computer Vision and Artificial Neural Network PERTANIKA Journal of Science & Technology 27 3 1317-1339

[2] Carvalho L M J, Gomes P B, Godoy R L O, Pacheco S, Monte P H F, Carvalho J L V, Nutti M R, Naves A C L, Vieira A C R A, Ramos S R R 2012 Total Carotenoid Content, α-carotene and β-carotene, of Landrace Pumpkins (Cucurbita moschata Duch): A Preliminary Study Food Research International 47 337-340

[3] Hendrawan Y, Sakti I M, Wibisono Y, Rachmawati M, Sutan S M 2018 Image analysis using color co-occurrence matrix textural features for predicting nitrogen content in spinach TELKOMNIKA. 16 6 2712-2724

[4] Hendrawan Y, Fauzi M R, Khoirunnisa N S, Andreane M, Hartianti P O, Halim T D, Uman C 2019 Development of colour co-occurrence matrix (CCM) texture analysis for biosensing. IOP Conference Series: Earth and Environmental Science 2019 230 012022

[5] Hendrawan Y, Widyaningtyas S, Sucipto 2019 Computer Vision for Purity, Phenol, and pH Detection of Luwak Coffee Green Bean TELKOMNIKA. 17 6 3073-3085

[6] Riquelme M T, Barraci P, Altisent M R, Valero C 2008 Olive Classification According to External Damage using Image Analysis Journal of Food Engineering 87 371-379

[7] Panda S, Sethy P K 2017 Post-Harvest Grading of Carica Papaya Fruit using Image Segmentation and Soft Computing International Journal of Advance Research in Computer Science. 8 7 692-697

[8] Arakeri M P, Lakshmana 2016 Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry Procedia Computer Science. 79 426-433

[9] Sadrnia H, Rajabipour A, Jafari A, Javadi A, Mostofi Y 2007 Classification and Analysis of Fruit Shapes in Long Type Watermelon Using Image Processing International Journal of Agriculture & Biology 9 1 68-70

[10] Momin M A, Rahman M T, Sultana M S, Igathinathane C, Ziauddin A T M, Grift T E 2017 Geometry-based Mass Grading of Mango Fruits using Image Processing Journal of Information Processing in Agriculture 4 150-160

[11] Kurtulmus F, Aillas I, Kavadar I 2016 Classification Pepper Seeds Using Machine Vision based on Neural Network Journal of Agriculture & Biological Engineering 9 1 51 – 62

[12] Ferreira M G, Azevedo A M, Siman L I, Silva G H, Careiro C S, Alves F M, Delazari F T, Silva D J H, Nick C 2017 Automation in Accession Classification of Brazilian Capsicum germplasm through Artificial Neural Networks Journal of Science Agriculture 74 3 203-207

[13] El-Ghoraba A H, Javedb Q, Anjumb F M, Hamedc S F, Shaabana H A 2013 Pakistani Bell Pepper (Capsicum annum L.): Chemical Compositions and its Antioxidant Activity International Journal of Food Properties 16 1 18-32

[14] Curl A L 1962 Red Pepper Carotenoids, The Carotenoids of Red Bell Peppers Journal of Agricultural and Food Chemistry 10 6 504-509

[15] Deli J, Molnár P, Matus Z, Tóth G 2001 Carotenoid Composition in The Fruits of Red Paprika (Capsicum annum var. lycopersiciforme rubrum) During Ripening; Biosynthesis of Carotenoids in Red Paprika Journal of Agricultural and Food Chemistry 49 3 1517-1523

[16] Noichinda S, Bodhipadma K, Moungjomprang D, Thongnurung N, Kasiolarn H 2016 Harvesting indices of chi-fai yai pepper (Capsicum annum L.) fruit The Journal of Applied Science 15 2 20-23

[17] Silva S M, Rocco S A, Sampaio K A, Taham T, da Silva L H M, Ceriani R, Meirelles J A 2011 Validation of a method for simultaneous quantification of total carotenoids and tococols in vegetable oils by HPLC Food Chemistry 129 4 1874-1881
[18] Hendrawan Y, Hawa L C, Damayanti R 2018 Fish swarm intelligent optimization to optimize real time monitoring of chips drying using machine vision and artificial neural network. *IOP Conference Series: Earth and Environmental Science* 2018 131 1 012020

[19] Arimboor R, Natarajan R B, Menon K R, Chandrashekar L P, Moorkoth V 2015 Red pepper (Capsicum annuum) carotenoids as a source of natural food colors: analysis and stability *Journal Food Science Technology* 52 31258-1271