Determining Feasibility Level of Beef Quality Based on Histogram and K-Means Clustering

[1)]Sri Winiarti, [2)]Ahmad Azhari, [3)]Khalidia Mega Agusta

Informatics Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
[1)]sri.winiarti@tif.uad.ac.id, [2)]ahmad.azhari@tif.uad.ac.id, [3)]khalidiamega16@gmail.com

Abstract— **Beef is one of the staple foods that are consumed by many people. The rise of beef with poor quality on the market and the lack of knowledge of consumers and the public in distinguishing the quality of meat that is decent and not suitable for consumption when seen in plain view. The Department of Agriculture and Food of the City of Yogyakarta, especially Animal Slaughterhouse (RPH), Giwangan Yogyakarta in determining the quality of beef, still uses laboratory tests which take a long time. This study proposes a system to identify the image of beef that is suitable for consumption based on the color features generated from the image histogram with Hue, Saturation, and Intensity (HSI) color models and the K-Means Clustering method. The results of identification carried out on 40 test data obtained maximum and minimum values of HSI from each feasibility category.**

Keywords—beef quality; feature extraction; HIS; k-means clustering; image processing

I. Introduction

Data from the Director of Veterinary Public Health, Director General of Animal Husbandry and Animal Health Ministry of Agriculture states that the prognosis of beef demand for 2017 reached 604,966 tons calculated based on the assumption of average national consumption of 2.31 kg/capita/year [1]. Meanwhile, the target of domestic meat production in 2017 was 354,770 tons, so it found deficiency of 250,196 tons. This indicates that the people's need for beef is very high. Increasing human awareness of the importance of nutritional needs is one of the reasons for the need for increased beef. One of the nutrients in beef that is very beneficial for the human body is protein.

Increased consumption of beef in Indonesia is used by certain parties to reap large profits by cheating on the quality of beef sold. As reported in [2], it has been found that 300 kilograms of poor quality beef occurred in Magetan, East Java. Of course, the sale of beef with low quality or not suitable for consumption can harm consumers.

References [3] conduct visual inspection in the process of determining the level of meat quality. Factors that influence the determination of the level of meat quality are the consideration of the quality of meat and the level of meat to be traded.

Determination of the quality of beef conducted by the Department of Agriculture and Food of the City of Yogyakarta, especially the Giwangan Yogyakarta Slaughterhouse (RPH) still uses laboratory tests. One of the easiest and simplest determinations of beef quality is by using a pH meter test. Meat pH meter test aims to measure the pH level of the meat by taking a sample of beef then the meat is pierced with a puncture device on a digital pH meter, after that the pH level of the beef will appear. When testing is done, the data is inputted manually to the computer, even there are still some tests that are still not computerized, so the need for innovation in the form of a system that is useful to identify the feasibility of a meat, especially beef computerized.

This research work proposes an identification of beef quality system by using histogram-based Hue, Saturation, and Intensity (HSI) color feature extraction. The results of the extraction will be clustered using the K-Means Clustering method so that the final results will be obtained meat quality information.

II. Literature Review

Reference [4] states that the determination of image quality on animal skin can be done with image segmentation. This study works using 4 types of animal skins. Image segmentation process uses the Darwinian Segmentation of Particle Swarm Optimization (DPSO) method. The next segmentation results are compared with the Harmonic Search Algorithm (HSA) method. The results show a small difference between the DPSO and HSA methods, but it can be concluded that the results of the DPSO method are slightly better than HSA.

Reference [5] analyzes color features on 50 road samples from three color spaces using RGB (Red-Green-Blue), HSV (Hue-Saturation-Value), and CIE L * a * b * (*Commission International de l'Clairage*) and perform feature extraction using GLCM (Gray Level Co-occurrence Matrix). The test results of the road detection method from 150 road test images using the feature limitation of the analysis showed 90.54% accuracy.

Reference [6] applying K-Means clustering for brainwave feature classification. Feature extraction methods using statistical feature including mean, standard deviation, skewness, kurtosis, and entropy. The results show that K-means clustering has been successful divide the feature data into four cluster

groups with small difference between each cluster. This study can be concluded that activity "Math" can provide signal characteristics as the basis of application development of brain computer interface by utilizing single sensor EEG.

The use of K-means clustering is also used in research [7] with the aim of reducing computational burden by reducing the size of high-dimensional data. The assumption of using k-means clustering can put redundant features on the same cluster, while irrelevant features are not included in any cluster. This research work shows that k-means clustering based on feature options can be performed to produce a feature subset and can return more than 80% accuracy of classification results.

According to [8], understanding the use of appropriate approaches in data analysis becomes specific to get the expected results. K-Means analysis provides good performance records in comparative analysis between K-Means and K-Medoids for Statistical Clustering.

III. METHODS

A. Digital Image Processing

Reference [9] states that digital image processing is the study of several things related to image quality improvement, image transformation, image feature selection, object recognition in images, compression or data reduction, data transmission, and data processing time. The purpose of digital image processing is for analysis, information retrieval, object description, and pattern recognition. In general, the steps in digital processing:

1) Image Acquisition

Image acquisition is the initial stage to get digital images. This stage starts from the object that will be taken, the preparation of tools, and imaging. Imaging is an activity of transformation from visible images into digital images.

2) Preprocessing

In the preprocessing stage, important things that are done at this level include image quality improvement, eliminating noise, image restoration, image transformation, determining the part of the image to be observed.

3) Segmentation

This stage aims to partition the image into key parts that contain important information.

4) Representation and Description

This stage aims to present an area as a list of coordinate points in a closed curve, with a description of the area or perimeter. The image description using feature extraction and selection

5) Introduction and Interpretation

The recognition stage aims to label objects whose information is provided by the descriptor, while the interpretation stage aims to give meaning to the group of objects that are recognized.

6) Knowledge Base

The knowledge base is useful for guiding the operations of each process module and controlling the interaction between the modules. In addition, the knowledge base is also used as a reference for the template matching and pattern recognition processes

B. RGB Color Model

The RGB color model is the standard color component of light, where the primary colors are red (R), green (G) and blue (B). Other colors are defined as a mixture or combination of the three basic colors. RGB color models are obtained by normalizing the three components' colors. Normalization in the red (r), green (g), and blue (b) models is shown in equation 1, equation 2, and equation 3.

$$r = \frac{R}{R + G + B} \tag{1}$$

$$g = \frac{G}{R + G + B} \tag{2}$$

$$b = \frac{B}{R + G + B} \tag{3}$$

R, G, and B represent the values of each red, green, and blue that have not normalized

C. HSI color model

The HSI color model consists of hue, saturation and brightness. Hue is a color attribute that describes pure colors (yellow, orange or pure red), while saturation provides a measure of the degree to which pure colors are weakened by white light. Intensity (gray-level) is the most useful descriptor in monochromatic imagery. This quantity can be measured by definition and can easily be interpreted. The HSI color model separates the intensity components from the color information carried (hue and saturation) in the image color. The color subspace for the RGB model is shown in Fig. 1.
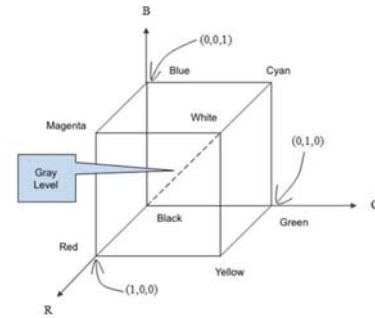

Fig. 1. Color subspace.

D. Convert RGB model to HIS

If an image is given in RGB color format, then the H component of each RGB pixel is obtained using (4), (5) and (6).

$$H = \begin{cases} \theta & if\ B \leq G \\ 360 & if\ B > G \end{cases}$$

$$\theta = cos^{-1}\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\} \tag{4}$$

The saturation component is determined by:

$$S = 1 - \frac{3}{(R+G+B)}[min(R,G,B)] \tag{5}$$

The intensity component is determined by:

$$I = \frac{1}{3}(R + G + B) \qquad (6)$$

Assuming that the RGB value is the normalized value in the range [0.1] and the angle θ is measured by the degree to the red axis of the HSI space. Hue can be normalized in the range [0,1] by dividing by 360 ° all values resulting from the hue equation. Two other HSI components have been obtained in this range if the RGB value is given in intervals [0.1]

E. Histogram

Image histogram is a graph that describes the distribution of pixel intensity values from an image or certain parts in the image. From a histogram can be known the frequency of occurrence of relative (relative) of the intensity of the image. The histogram can also show many things about the brightness (brightness) and contrast (contrast) of an image. Therefore, histograms are valuable tools in image processing both qualitatively and quantitatively.

F. K-Means Clustering

Many studies have been conducted by researchers related to the use of k-means algorithms in grouping. K-Means clustering is one of the most widely used grouping techniques in research [6]–[8], [10]–[15]. K-means shows more accurate results for solid datasets compared to rare datasets. The performance is influenced by significant feature subset. This method partitioning data into groups so that different characteristic data are grouped into other groups. The purpose of grouping this data is to minimize variation within a group and maximize variation between groups. Flowchart identification of beef quality can be seen in Fig. 2.

The Clustering algorithm with K-Means is as follows:

- The first step is to determine the number of clusters to be formed.
- The second step is initialization point of cluster (centroid) randomly from a set of data to be grouped.
- The third step is to calculate the distance of any data to each centroid using Euclidean distance.
- The fourth stage is to choose the shortest distance between each data with the centroid.
- The fifth step is to determine the position of new centroid by calculate the average of data in the same centroid.
- The sixth step is to make sure the algorithm converged with data in the new centroid same with initial centroid. If the two centroids have the same data then the process end. If not, back to step 3.

In this beef quality identification system, there are several steps taken. The first step is image input. The original image that has been acquired in the form of a digital image will be selected based on the image format with the extension file bpm or jpg.

After the original image is displayed on the system, the second step is cropping the original image. Cropping aims to obtain a more detailed image so as to facilitate the categorization of feasibility categories. The third step is to take the red, green, and blue (RGB) values from the image that has been cropped.

The fourth step is image processing. Image processing is done by calculating the average value of the RGB to HSI transformation and also the histogram of each HSI color. After obtaining a histogram from each color, the last step is clustering using K-Means method.
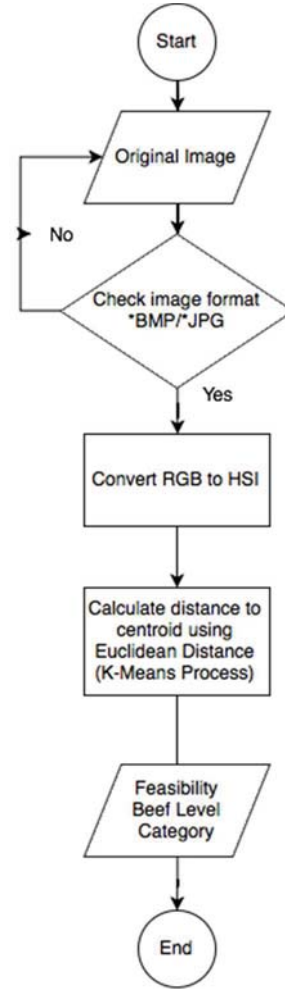

Fig. 2.   General Procedure of Proposed Method

IV.   RESULT AND DISCUSSIONS

The original image is selected based on the extension bmp file or jpg then entered into the system. The next step is the cropping the image and takes the value of red, green, and blue (RGB) from the cropped image. In image processing, RGB images are converted to HSI and calculate the histogram value. Fig. 3 shows the display of beef histogram value. RGB and HIS values are shown in Fig. 4.
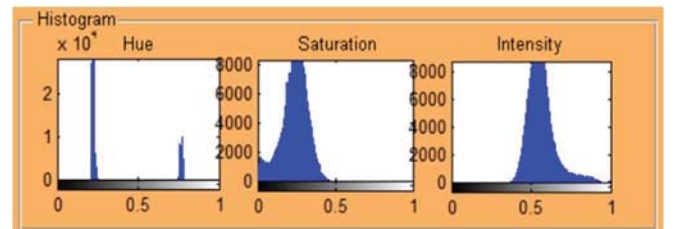

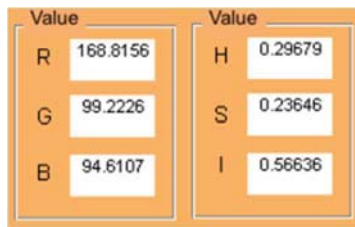Fig. 3.   Histogram of beef identification process.

Fig. 4. RGB and HIS values in beef.

After obtaining a histogram from each color, the last step is clustering using K-Means to get the identification results from the input image, whether it is categorized as very feasible, feasible, less feasible or not feasible. In the image identification process, the image is classified according to the category of the level of feasibility of beef. Determination of categories is obtained based on the Euclidean distance calculation of each image.

After testing 40 data on beef consisting of 10 images of very decent beef, 10 images of decent beef, 10 images of beef that are not feasible and 10 images of unworthy beef obtained maximum and minimum values of HSI from each feasibility category. The "very feasible" category has a minimum value (0.23881, 0.17467, 0.54027) and a maximum value (0.32845, 0.33301, 0.67644). The "feasible" category has a minimum value (0.53421, 0.19735, 0.44565) and a maximum value (0.76843, 0.24434, 0.52883). The "less feasible" category has minimum values (0.23187, 0.18327, 0.41259) and maximum values (0.37045, 0.29206, 0.50205). The "not feasible" category has a minimum value (0.24089, 0.15268, 0.25967) and maximum value (0.32165, 0.23368, 0.33024). Table I represents the appearance of each feasibility category group.

TABLE I. FEASIBILITY LEVEL CATEGORY OF BEEF

	H	S	I	Category
Min	0.23881	0.17467	0.54027	Very Feasible
Max	0.32845	0.33301	0.67644	Very Feasible
Min	0.53421	0.19735	0.44565	Feasible
Max	0.76843	0.24434	0.52883	Feasible
Min	0.23187	0.18327	0.41259	Less Feasible
Max	0.37045	0.29206	0.50205	Less Feasible
Min	0.24089	0.15268	0.25967	Not Feasible
Max	0.32165	0.23368	0.33024	Not Feasible

V. CONCLUSIONS

Based on the results of the research that has been done, it can be concluded that the results of the identification carried out on 40 test data using a histogram with the HSI color model and using the K-Means Clustering method produces HSI values for the 5 categories of feasibility, among them are very feasible, feasible, less feasible, and not feasible. In the experimental results it can be seen that there are small differences in each feasibility category.

REFERENCES

[1] B. Statistics Indonesia, "Statistik Indonesia 2016," BPS-Statistics Indonesia, p. 680, 2016.
[2] kumparan kumparan, "Penjual Daging Sapi Glonggongan Antar Provinsi Dibekuk Satreskrim Polres Magetan," kumparan.com, Indonesia, 07-Aug-2018.
[3] K. Shiranita, T. Miyajima, and R. Takiyama, "Determination of meat quality by texture analysis," Pattern Recognit. Lett., vol. 19, no. 14, pp. 1319–1324, Dec. 1998.
[4] M. M, S. Winiarti, A. Prahara, and E. Febri, "Evolutionary Optimization Methods for Animal Leather Image Segmentation," Int. J. Comput. Appl., vol. 179, no. 17, pp. 31–35, Feb. 2018.
[5] A. Prahara and A. Azhari, "Analisis Fitur Warna dan Tekstur untuk Metode Deteksi Jalan," vol. 2, no. 2, p. 9, 2016.
[6] A. Azhari and L. Hernandez, "Brainwaves feature classification by applying K-Means clustering using single-sensor EEG," Int. J. Adv. Intell. Inform., vol. 2, no. 3, pp. 167–173, Nov. 2016.
[7] D. P. Ismi and S. Panchoo, "K-means clustering based filter feature selection on high dimensional data," Int. J. Adv. Intell. Inform., vol. 2, no. 1, p. 8, 2016.
[8] N. Arbin, N. S. Suhaimi, N. Z. Mokhtar, and Z. Othman, "Comparative Analysis between K-Means and K-Medoids for Statistical Clustering," 2015, pp. 117–121.
[9] W. Y. Lo and S. M. Puchalski, "DIGITAL IMAGE PROCESSING: Digital Image Processing," Vet. Radiol. Ultrasound, vol. 49, pp. S42–S47, Jan. 2008.
[10] V. S. Chandrawanshi, R. K. Tripathi, and N. U. Khan, "A comprehensive study on k-means algorithms initialization techniques for wireless sensor network," in Signal Processing and Communication (ICSC), 2016 International Conference on, 2016, pp. 154–159.
[11] A. Kapoor and A. Singhal, "A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms," in Computational Intelligence & Communication Technology (CICT), 2017 3rd International Conference on, 2017, pp. 1–6.
[12] P. Novianti, D. Setyorini, and U. Rafflesia, "K-Means cluster analysis in earthquake epicenter clustering," Int. J. Adv. Intell. Inform., vol. 3, no. 2, p. 81, Jul. 2017.
[13] M. A. A. Riyadi, D. S. Pratiwi, A. R. Irawan, and K. Fithriasari, "Clustering stationary and non-stationary time series based on autocorrelation distance of hierarchical and k-means algorithms," Int. J. Adv. Intell. Inform., vol. 3, no. 3, p. 154, Dec. 2017.
[14] B. S. Shedthi, S. Shetty, and M. Siddappa, "Implementation and comparison of K-means and fuzzy C-means algorithms for agricultural data," in Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on, 2017, pp. 105–108.
[15] A. Prahara, I. T. R. Yanto, and T. Herawan, "Histogram Thresholding for Automatic Color Segmentation Based on k-means Clustering," in Recent Advances on Soft Computing and Data Mining, vol. 549, T. Herawan, R. Ghazali, N. M. Nawi, and M. M. Deris, Eds. Cham: Springer International Publishing, 2017, pp. 344–354.