

PAPER • OPEN ACCESS

## Corn quality identification using image processing with k-nearest neighbor classifier based on color and texture features

To cite this article: M Effendi *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **230** 012066

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

# Corn quality identification using image processing with k-nearest neighbor classifier based on color and texture features

**M Effendi, M Jannah and U Effendi**

Department of Agro-industrial Technology, Faculty of Agricultural Technology,  
Universitas Brawijaya, Malang, Indonesia

E-mail:mas.ud@ub.ac.id

**Abstract.** Corn is food crop commodity that is widely used, including as raw material for animal feed. Determination of corn quality at the farm level is often associated with drying time. This method has weaknesses, namely low efficiency, objectivity and level of consistency and also can lead to conflicts between traders and farmers. This study aims to identify the quality of corn using digital image processing based on color and texture features. This research uses Pertiwi-3 and Pertiwi-6 corn varieties. The corn quality identification system uses 7 features input (hue, saturation, value, contrast, correlation, energy, homogeneity) and K-NN algorithm as classifiers. The number of image data used are 500 images with a test ratio of 70: 30. This research is able to classify the quality of corn into 10 quality categories. The highest accuracy is obtained at 90.00% when the k value (the nearest neighbor) is 5 and the distance calculation method is Cityblock.

## 1. Introduction

Corn acts as a food source for some Indonesian people. However, corn is used more as raw material for animal feed and other industries. Data from the Indonesian Agency for Agricultural Research and Development states that the highest utilization of corn is used as animal feed, namely 22% as direct animal feed and 44% as feed industry raw materials, 25% for food industry raw materials and 9% for direct household consumption [1]. Corn quality problems at the farm level are still high in water content and the level of damage and contamination that occurs during the postharvest handling process and during the piping process with many corn products at the farm level that are not absorbed by the industry, especially the feed industry. The feed industry cannot be supplied directly by farmers because the resulting moisture content of the corn is too far from the SNI standard. Therefore, the fulfillment of corn for the needs of the feed industry is carried out through collectors or wholesalers.

Determination of the quality of corn by farmers is based on the drying time which is categorized into wet, half dry and dry corn which will affect the purchase price decision. The method is less efficient and subjective depending on each trader, and can cause conflict due to price differences between traders. To accelerate price determination based on water quality, it is necessary to develop a good and accurate method for identifying corn quality based on digital image processing. Image processing aims to improve image quality so that it is easily interpreted by humans or machines (in this case computers). Digital image processing involves visual perception and has the characteristics of input data and output information in the form of digital image files [2]. Sophisticated digital image processing capabilities enable the process of identifying the quality of agricultural commodities [3] to



be more effective and efficient. This study aims to design a system to identify the quality and varieties of corn based on color and texture features and also determine the accuracy of the K-Nearest Neighbor (K-NN) method in estimating the quality of corn.

In this study, an application system for identifying quality and corn varieties was developed using HSV (hue, saturation, value) color features and GLCM (Gray Level Co-Occurance Matrix) texture using the K-NN classification [4]. The color of HSV is the result of transformation from RGB using geometric methods [5], where Hue is a measure of the wavelength found in the dominant color received by sight, Saturation states the purity of the color of light [6], and Value states the brightness of the color [7]. The choice of HSV colors because these colors represent colors according to human vision features [8]. Texture is one of the most important characteristics for image analysis, where texture provides information about the structure arrangement on the surface, changes in intensity, or brightness of the color [9]. The use of texture features as a parameter in estimating water content because the level of water content can affect the texture of corn. Texture analysis was carried out using the GLCM method. GLCM is a matrix that represents the relationship of proximity between pixels in the image in various directions of orientation and spatial distance [10]. The calculation of the GLCM feature consists of the value of contrast, correlation, energy, and homogeneity. The use of K-NN algorithm is because it is able to achieve higher accuracy results [11] and is easier and more reliable to represent [12] compared to other classification algorithms such as Decision Tree and Naive Bayessian. K-Nearest Neighbor (K-NN) is a simple algorithm that belongs to a strong classifier, but requires a k value and the use of the right distance to produce high accuracy [12]. In addition, K-NN requires a large memory allocation because it does not build a classification model in the process.

## 2. Materials and Methods

The research was conducted at the Laboratory of Food and Agricultural Production Engineering, Department of Agricultural Engineering, while processing data at the Computing and System Analysis Laboratory, Department of Agro-Industrial Technology, Faculty of Agricultural Technology, Universitas Brawijaya Malang. Corn material was obtained from corn farmers in Tuban regency, East Java. The shelled corn used is Pertiwi 3 and Pertiwi 6, each consisting of wet, half dry and dry shelled corn. In addition to whole shelled corn, shelled corn with broken seeds and moldy seeds are also used. The equipment used is an image-acquisition device and moisture tester. The stages of the research include sample preparation, image acquisition device design, development of image processing algorithms, GUI design, training and system testing and system performance analysis.

### 2.1. Sample preparation and image acquisition

The initial stage is preparing samples of wet, half dry and dry shelled corn of Pertiwi 3 and Pertiwi 6 varieties. There are 10 classes of categories consisting of Pertiwi 3 and Pertiwi 6 each of 5 categories (categories 1-5 are Pertiwi 3 varieties, while categories 6-10 are Pertiwi 6) based on PERMENDAGRI No. 27/M-DAG/PER/5/2017. Each category consists of 5 groups of water content variations namely Category 1 water content variations: 11% -15%, Category 2 water content variations: 16% -20%, Category 3 water content variations: 21% -25%, Category 4 water content variations: 26% -30%, Category 5 water content variation: 31% -35%. Likewise, Category 6-10 has the same variation in water content as Categories 1-5. Each group of water content variation weighed  $20 \pm 2$  grams consisting of whole seeds, and a maximum of 2% of each broken seeds and moldy seeds. Furthermore, the image of each sample group is taken 10 times by randomizing the position of the grain, resulting in a different image. The total corn image datasets obtained are 500 image data. Image acquisition illustration can be seen in Figure 1. The results of image acquisition are stored in JPEG format.



**Figure 1.** Image acquisition.

### 2.2. Development of the image processing algorithm

There are several processes that must be carried out to develop the algorithm, starting from the image acquisition process, cropping and resizing images to 300x300 pixels, extracting HSV color and GLCM textures features, which include contrast, correlation, energy and homogeneity, then classify using K-NN. This study uses variations in the values of k (closest neighbors) 1, 3, 5, 7, and 9 with the Euclidean and Cityblock distance methods.

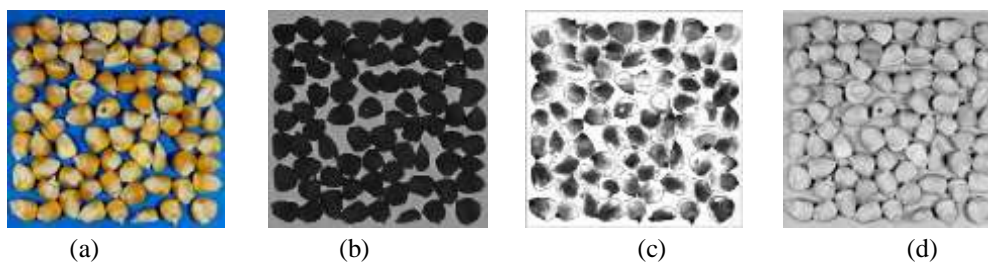
### 2.3. System training and testing

The system training and testing phase aims to determine the level of accuracy and the ability of the system to classify the quality of corn. The system input consists of 7 parameters, namely 3 color features H, S, V and 4 GLCM texture features, namely contrast, correlation, energy and homogeneity. The ratio of training data and testing data used is 70:30. From a total of 500 corn image data taken, 350 data consisting of each category class are used as training data, and 150 data are used as testing data. Furthermore, the image is analyzed by comparing the closest distance between the testing data and the k closest neighbor value in the training data using the K-NN method. The output of the test results is in the form of corn class categories which are represented in varieties and water content.

## 3. Results and Discussion

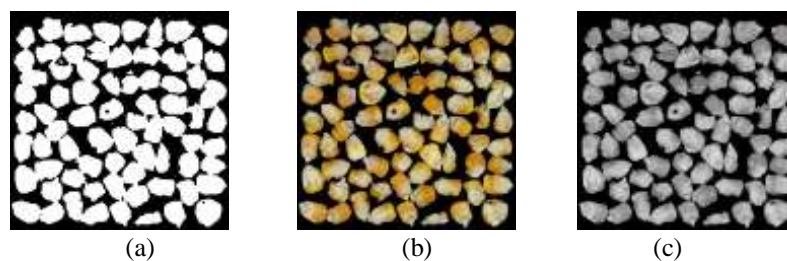
### 3.1. Image acquisition and processing algorithm

Image acquisition and processing algorithm in the corn quality identification system starts from reading the image, cropping, resizing, feature extraction and classification using the K-NN algorithm. Cropping to facilitate data processing at a later stage. Resizing aims to change the size of the cropping image into a processed image. The cropping result input image can be seen in Figure 2(a). Feature extraction is done to get features from corn kernels. Color feature extraction is done using HSV color by calculating the mean value of each component H, S, V. The mean or average pixel value is a value that indicates the dispersion size of an image. The results of component extraction are Hue, Saturation, Value can be seen in Figure 2(b), Figure 2(c), and Figure 2(d).



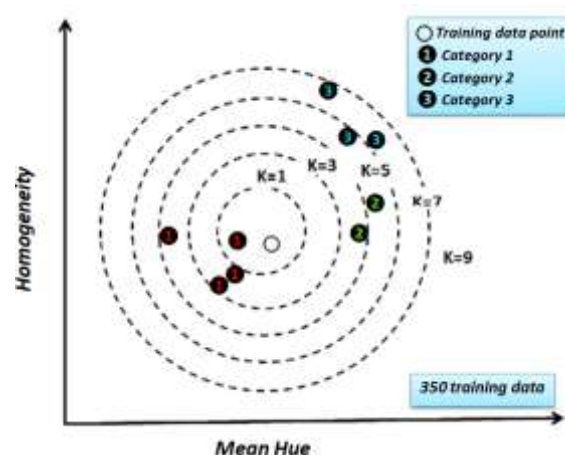
**Figure 2.** (a) Input image, (b) extraction of Hue value, (c) extraction of Saturation values, (d) extraction of Value values.

Texture feature extraction uses the Gray Level Co-occurrence Matrix (GLCM) through preprocessing, segmentation and extraction stages. The GLCM features used are contrast, correlation, energy, and homogeneity. The preprocessing stage or image quality improvement is done using image enhancement and median filters. Image enhancement serves to enhance certain features in the image so that the success rate of the image processing process will be higher. The technique used is intensity adjustment (imadjust) to get a new contrast that is better than the original image. Median filters are used to eliminate noise in the imadjust image [13]. Preprocessing stage facilitate image when feature extraction is done [14]. The segmentation phase uses the thresholding method by dividing the image region that tends to be dark made darker (perfect black), while the image region that tends to be bright is made brighter (perfect white). The threshold value used is 130, where this value is obtained through trial error until the parts of the object and background can be completely separated. The segmentation results are binary images with pixel intensity values of 0 (background) or 1 (objects) shown in Figure 3 (a). The segmented binary image is then reassembled with the original object so as to produce a background color image (Figure 3 (b)). To get the value of the GLCM feature, the segmented RGB image is converted to grayscale. The grayscale image from the RGB image conversion is shown in Figure 3 (c).



**Figure 3.** (a) Binary image, (b) RGB image segmentation result, (c) Grayscale image.

The illustration of the K-NN classification is shown in Figure 4. The testing data that class is unknown is described as a black query point, while the training data points are assumed to be meeting points of the seven feature dimensions. Because the dimensions and the amount of data are too large, the example of the Mean Hue and Homogeneity features is taken. Figure 4 shows when using  $k = 1, 3, 5, 7$  or  $9$ , the testing data is classified into category 1, because classes majority in the closest neighbors of  $1, 3, 5, 7$  and  $9$  are category 1.



**Figure 4.** K-NN classification results using  $k = 9$  in the Mean Hue and Homogeneity features.

### 3.2. System training and testing using K-Nearest Neighbor (K-NN)

The system training process is carried out to train the K-NN algorithm to be able to identify the quality of corn. K-NN built a classification system through learning from previously classified training data. In the case of quality and corn varieties identification, the K-NN algorithm training process was carried out by storing HSV color feature and GLCM texture features vectors and class labels from each category class. Data extraction in the form of 3 color features and 4 corn texture features are used as test parameters. The number of image data used is 500 images with a test ratio of 70: 30. Training image database planted on a system is 350 data. The database is used as a reference in the system testing process.

System testing aims to determine the system's ability to predict the accuracy of the identification of corn quality based on HSV color features and GLCM texture features. The number of testing data used is 150 data consisting of 15 data for each category. 15 data were taken randomly as much as 3 data per group of water content variations. The test focuses on the accuracy of the classification results. For each variation of the value of  $k = 1, 3, 5, 7, 9$  and the Euclidean and Cityblock distance method used, the accuracy is calculated. Overall, correctly classified results using K-NN are shown in Table 1.

**Table 1.** Correct classification results using K-Nearest Neighbor.

K	Number of Correct Data										Total	Accuracy (%)
	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Cat. 8	Cat. 9	Cat. 10		
Euclidean method												
1	13	10	11	13	12	15	14	8	14	15	125	83.33
3	14	10	9	12	12	15	15	10	14	15	126	84.00
5	14	10	13	11	12	15	14	7	13	15	124	82.67
7	14	9	13	10	14	14	13	9	15	15	126	84.00
9	14	10	14	11	10	14	15	9	14	14	125	83.33
Cityblock method												
1	12	10	13	12	13	15	14	10	15	15	129	86.00
3	14	10	12	11	14	15	15	11	14	15	131	87.33
5	14	11	14	12	14	15	15	10	15	15	135	90.00
7	13	11	14	9	14	15	15	13	15	15	134	89.33
9	14	10	13	12	11	14	15	10	15	14	128	85.33

In Table 1 can be seen the results of the system classification using K-NN for each variation of the  $k$  value and the method of calculating the distance used. In this system, the highest accuracy results are obtained when using Cityblock distance method with a value of  $k = 5$  that is equal to 90.00%. The value of  $k = 5$  means that there are five adjacent vectors that are used as a comparison of the testing data which can represent the feature vectors of various classes. From a total of 150 testing data, the system is able to classify the quality of corn correctly as much as 135 data. The lowest accuracy is obtained when using Euclidean with a value of  $k = 5$ , which is 82.67%, where the data is classified as true as 124 data. Based on the results of the best K-NN accuracy, the system is able to classify all data correctly in Category 6, Category 7, Category 9 and Category 10. The data that is least correctly classified is Category 8. Images that are not properly classified can be influenced by light factors, the dirt on the object and the condition of the camera that is less focused during the image acquisition process so that the image has a lot of noise and is difficult to recognize. The image acquisition process affects the results of accuracy generated from the system [3]. The results of this study can be used as a basis for further development of corn quality identification methods at the farm level based on image processing in form of computer or mobile applications. With 90% accuracy, this system is able to identify corn quality and variety well. In the identification of agricultural commodities with digital image processing, this study resulted in higher accuracy than previous study [14].

#### 4. Conclusion

Corn quality identification system using digital image processing based on HSV color input and GLCM texture features consisting of contrast, correlation, energy, and homogeneity successfully identified the corn quality and varieties with good accuracy. Based on the results of the test, the system made was able to classify the quality and varieties of corn into 10 categories consisting of Pertiwi 3 and Pertiwi 6 each of 5 categories of quality. The highest accuracy obtained when  $k = 5$  uses Cityblock distance with an accuracy value of 90.00%. For further research, it is recommended to use other parameters that have a good influence in recognizing the characteristics of corn features to improve the accuracy of the quality identification system and corn varieties, such as adding shape features as test parameters.

#### References

- [1] IAARD 2015 Harmonisasi data produksi dan konsumsi beras dan jagung dalam kaitan dengan luas lahan dan produktivitas (Harmonization of data on rice and corn production and consumption in relation to land area and productivity) Indonesian Agency for Agricultural Research and Development (IAARD) - Ministry of Agriculture Jakarta Indonesia. [In Indonesian]
- [2] Daryanto 2016 Aplikasi pembesaran citra menggunakan metode nearest neighbor interpolation (Image enlargement application uses the nearest neighbor interpolation method) *JUSTINDO* **1** 1 31-35. [In Indonesian]
- [3] Sela E I, Ihsan M 2017 Deteksi kualitas telur menggunakan analisis tekstur (Detection of egg quality using texture analysis) *IJCCS* **11** 2 199-208. [In Indonesian]
- [4] Zhang S, Cheng D, Zong M, Gao L 2016 Self representation nearest neighbor search for classification *Neurocomputing* **195** 137-142.
- [5] Deswal M, Sharma N 2014 A simplified review on fast HSV image color and texture detection and image conversion algorithm *IJCSMC* **3** 5 1216-1222.
- [6] Semary N, Hadhoud M, Abdul-Kader H, Ababs A 2013 Novel compression system for hue-saturation and intensity color space *Int. Arab J. Inform. Technol.* **10** 6 546-552.
- [7] Kour H 2015 Analysis on image color model *Int. J. Adv. Res. Comp. Comm. Eng.* **4** 12 233-235.
- [8] Chitra S, Balakrishnan G 2012 Comparative study for two color spaces HSCbCr and YcbCr in skin color detection *Applied Mathematical Sciences* **6** 85 4229-4238.
- [9] de Siqueira F R, Schwartz W R, Pedrini H 2013 Multiscale gray level co-occurrence matrices for texture description *Neurocomputing* **120** 336-345.
- [10] Albrechtsen F 2008 Statistical texture measures computed from gray level cooccurrence matrices Image Processing Laboratory Department of Informatics University of Oslo Norway.
- [11] Gu Q, Song Z 2009 Image classification using SVM, KKN and performance comparison with logistic regression Final Project Report Department of Computer Science, Dartmouth College Hanover USA.
- [12] Arifin A D 2012 Implementasi algoritma K-Nearest Neighbor Berdasarkan One Pass Clustering untuk Kategorisasi Teks (Implementation of the K-Nearest Neighbor algorithm Based on One Pass Clustering for Text Categorization) Thesis Institut Teknologi Sepuluh November Surabaya Surabaya Indonesia. [In Indonesian]
- [13] Sulistyow W, Yos R B, Filipus F 2009 Analisis penerapan metode median filter untuk mengurangi noise pada citra digital (Analysis of the application of the median filter method to reduce noise in digital images) Konferensi Nasional Sistem dan Informatika Bali Indonesia. [In Indonesian]
- [14] Effendi M, Ullivia F, Usman E 2017 Identifikasi jenis dan mutu kopi menggunakan pengolahan citra digital dengan metode jaringan syaraf tiruan *Jurnal Ilmiah Teknologi Pertanian AGROTECHNO* **2** 1 140-146. [In Indonesian]