Model for identification of rice type using combination of shape and color features

J Jumi*, A Zaenuddin and T Mulyono
Politeknik Negeri Semarang, Semarang, Indonesia

*jumi_mail06@gmail.com

Abstract. Rice is an agricultural commodity that is a staple food in Indonesia with hundreds of types of rice that have different characteristics. The type of rice can be distinguished from color and shape. The main feature that is dominant and can distinguish each type of rice is the color and shape. This feature is the main key in identifying types of rice. Identification is done by comparing the similarity of rice images using the value of color and shape features. The similarity can be determined through the difference in feature values between the query image and the database image. The closer the difference is to zero, the higher the level of similarity. The degree of similarity will affect the accuracy of image recognition at the time of identification. In this study, an analysis of the accuracy of image identification and measurement of computation time was carried out. Improved identification accuracy using the weighting of color and shape feature values. Extraction of the two value features using the invariant moment and color moment. Preprocessing before extraction using Grayscale, resize, edge Enhancement, Histogram Equalization. Clustering of rice image data using K-Means clustering. The results showed that the accuracy of identification with 400 rice image test data, reached more than 95% in the weighting scheme Ws (weighted Shape) = 40% and Wc (weighted color) = 60% with an average computing time of 5 milliseconds at 10 the cluster.

1. Introduction
Rice is one of the agricultural commodities that are needed as the main food ingredient. Rice has a variety of types with different characteristics-different. The type of rice has different quality and economic value so that it can influence the perception of customers who consume it [1]. The variety of rice types of rice necessary database for easy identification and tracking information about the types of rice. Identification is done through the introduction of rice images by using the image feature as a key field.

The value of the feature extracted from the object image affects the accuracy of the identification of the object which in this study is rice as the object. Identification through the introduction of image features is influenced by various variables including accuracy and accuracy in the process of calculating the value of features. One other variable that can affect accuracy is the use of feature value weighting schemes.

Generally, the image on an object dominates different features so that a feature weighting scheme is needed by the object. The image of an object in rice has a more dominant color feature to distinguish the similarity between one image of rice with another image of rice. The shape and texture features of rice images are as supporting features for image similarity measurement [2]. Image similarity can be grouped in three main groups in the measurement namely the similarity of shape, color similarity and...
the last is the similarity of texture [3]. This similarity is used as a key introduction to image-based data identification.

If more nearly zero, then have a higher level of similarity [3]. Data that are grouped based on the value of features that have been weighted is a process to improve accuracy and shorten the time for information retrieval.

The stages of feature weighting scheme with different percentages and clustering were carried out in this study. The stages are carried out to increase the accuracy value and speed up image identification time. Identification is done by comparing the feature values between the query image with the image in the database [4].

Previous research that discusses image similarity search based on feature or color feature extraction. The research has been carried out without regard to feature weights and clustering to produce a similarity accuracy value of more than 75% [5]. A computer vision-based approach is used to classify six types of rice using twenty-one features of rice grains. The classification uses seven features based on color and fourteen morphological features. The method used is a neural network that reaches an accuracy of 88.30% [6]. Support multi-class vector machines (SVM) are also used to classify rice into premium classes, grade A, grade B, grade C and achieve 86% accuracy [7]. The neural network method in 4 groups of a total of 52 types of Philippine rice grains can achieve an average accuracy of 70%, by classifying rice groups from rice images [8]. Here, images of individual rice grains are obtained by placing rice grains one at a time on the scanner. After scanning it continues using the color and morphological features to classify the six types of rice once again using neural networks. Their model achieves an average accuracy of 84: 8%. In Silva et al. [9] Singh et al. [10] classifying 9 different types of rice using thirteen morphologies, fifteen textures, and six color features and achieving 92% accuracy.

This study aims to analyze the identification of image-based rice types using feature value weighting schemes and the calculation of search computing time after clustering using the K-Means technique. The K-Means algorithm is used to determine the cluster position in each image by first calculating the image distance with all centroids using the Euclidean distance method.

2. Methods
This research uses a method by combining color and shape features. The method used consists of three stages, namely data acquisition, recognition, and identification shown in Figure 1. This study uses a merging method that was not found in previous studies, namely the method of weighting features and clustering based on color and shape features for the process of identification of rice types.

![Figure 1. Overview of image search stages.](image-url)
The details of the stages of this research are as follows:

2.1. Preprocessing
Preprocessing is the stage of improving image quality before feature extraction to improve the accuracy of image feature extraction results. There is a difference between preprocessing in color feature extraction and shape feature extraction. The difference in preprocessing is to get a quality image before feature extraction.

2.1.1. Preprocessing color features
a. Resize
The larger the image size, the longer the extraction time, so that the resize step is needed to speed up the computation process. At this stage, the image is resized to 200 x 200 pixels.

b. Histogram Equalization
At this stage, histogram leveling is done so that the image quality becomes more contrasted to get a quality color feature value. Histogram leveling results can be seen in Figure 2 and Figure 3.

![Figure 2](image1.png) ![Figure 3](image2.png)

**Figure 2.** Comparison of the image before and after the histogram process, (a) Original Image, (b) Image of Histogram Equalization (HE).

**Figure 3.** Comparison of histograms, (a) Original image histogram, (b) Image histogram of HE results.

2.1.2. Preprocessing shape features
a. Resizing
The size of the image will affect the length of feature extraction time, the smaller the size, the faster the computation of feature extraction. At this stage, the image is resized to 200 x 200 pixels.

b. Grayscale
Especially in calculating the features of color elements are not counted so that the image is changed to grayscale so that the computation process is faster.

c. Edge Enhancement dan Histogram Equalization (HE)
The edge enhancement stage will produce a new image with clearer or sharper edges of the object, so the shape of the image will be clearer. This stage uses the convolution method with the Sobel operator [3]. The edge sharpening results can be seen in Figure 4.
2.1.3. Feature extraction. Color and shape feature extraction is performed on the image after going through the preprocessing stage. Color feature extraction using color moment and form feature extraction using invariant moment.

The color moment uses three main moments of the color distribution of the image, namely the mean, standard deviation, and skewness, so this method produces three values for each color component [11]. Color Moments method is a method used to distinguish images based on color features [12]. The color moment assumes that the color distribution of an image can be expressed as a probability distribution. The invariant moment produces feature values that are not susceptible to image changes caused by Rotation, Scale, and Translation (RST) [11]. Analysis of the set of moments of a function $f(x, y)$ of two variables can also be done with [2].

2.1.4. Weighting of features. This study uses the weighting of different feature values for identification accuracy analysis. Weighting varying feature values will result in different levels of identification accuracy. Weighting is carried out with four variations namely the W1 scheme (0.25, 0.75) is a 25% weighting scheme for the color feature weights and 75% the shape feature weights. The W2 scheme (0.6, 0.4) is a 40% weighting scheme for color feature weights and 60% for shape feature weights. The W3 scheme (0.75, 0.25) is 75% the color feature weight and 25% the shape feature weight. And in the W4 scheme (0.4, 0.6) is a 60% weighting scheme for the color feature weights and 40% the shape feature weights.

2.1.5. Clustering. Data clustering to speed up the identification process. In this study, the K-Means Clustering and Euclidean distance methods are used as a measurement of image similarity. The similarity measurement equation uses equation 1.

$$D(Q, M) = \sqrt{\sum_{n=1}^{k} (Q_n - M_n)^2}$$ (1)

Which $Q_n$ and $M_n$ are features of the query image and database image in the nth dimension. Image similarity is determined by the difference in the value of the identified image features with the image in the database. The difference in feature values close to zero has the highest level of similarity.

3. Results and analysis
Tests in this study used 5 types of rice there are IR-35, IR-50, IR-64, Sticky Rice, and Mentik with a total of 400 rice images measuring 200 x 200 pixels.
| No | Image of Rice | Name     | Grain Type | Level |
|----|--------------|----------|------------|-------|
| 1  |              | IR 35    | Whole seed | 2     |
| 2  |              | IR 35    | Whole seed | 1     |
| 3  |              | IR 50    | Whole seed | 2     |
| 4  |              | IR 50    | Whole seed | 1     |
| 5  |              | Ketan    | Whole seed | 3     |
| 6  |              | Ketan Merah | Whole seed | 3     |
| 7  |              | IR 64    | Whole seed | 3     |
| 8  |              | IR 64    | Whole seed | 5     |
| 9  |              | Menthik  | Whole seed | 2     |
| 10 |              | Menthik  | Whole seed | 1     |

Table 1. Data on rice image acquisition.
In Table 1 an example of rice image data with names, grain types, and quality levels as the identification of rice quality. Quality at level 1 is poor quality rice, level 2 is medium quality rice, level 3 is pretty good quality, level 4 is good quality rice, and level 5 is premium quality rice.

3.1. Identification results
Identification testing is done by varying the number of clusters and variations in the percentage of feature weights. Test results with 60% weighting variation for color feature weights and 40% form feature weights with 10 cluster variations are shown in Figure 5.

![Figure 5](image)

**Figure 5.** The results of rice tracking the identity of rice (0.6 and 0.4).

In Figure 5, the results of identification of the quality and type of rice are shown using similarity values with the Euclidean distance method. Retrieval ranks similarity 1 to 10 from the rice image database and testing with weighting variations and variations in the number of clusters will produce different retrieval outputs on the same data.

3.2. Discussion
At this stage, retrieval accuracy is calculated in the rice image database. Furthermore, a retrieval accuracy analysis was carried out on the rice image database before and after clustering with the number of clusters that varied from 3 to 15 clusters.

3.2.1. Before clustering. Retrieval accuracy analysis at this stage uses a database of rice images before clustering with weighting variations in shape, color, and texture features. The accuracy value of the retrieval results with the same feature weights and varying feature weights can be seen in Table 2.

| Image of Name | Equal Weight | W1        | W2        | W3        | W4        |
|--------------|--------------|-----------|-----------|-----------|-----------|
| 11           | 75,55%       | 78,59%    | 86,75%    | 84,38%    | 61,33%    |
| 12           | 74,67%       | 79%       | 87%       | 83,87%    | 60,67%    |
| 13           | 75,05%       | 79,45%    | 87,10%    | 83,33%    | 60%       |
| 14           | 74,69%       | 79,15%    | 87,50%    | 86,67%    | 59,33%    |
| 15           | 75,65%       | 78,75%    | 86,67%    | 84,38%    | 60,67%    |
| 16           | 75%          | 78,50%    | 86,36%    | 84%       | 60%       |
| 17           | 75,76%       | 79,45%    | 86,67%    | 83,87%    | 59,33%    |
| 18           | 74,89%       | 78,65%    | 87,33%    | 83,33%    | 60,67%    |
| 19           | 74,87%       | 79,55%    | 86%       | 84%       | 61,33%    |
| 110          | 75,35%       | 79,25%    | 87,33%    | 84,38%    | 60,67%    |

Table 2 shows that the retrieval accuracy with the variation of W2 produces the highest average accuracy and the variation of the weight W4 produces the lowest accuracy.

3.2.2. After clustering. Grouping is done on the image that is based on similarity, the results will narrow the search space for image data information. The process is very beneficial because it will shorten the
retrieval time and increase the accuracy of the retrieval. Retrieval accuracy with variations in weight and number of clusters can be seen in Table 3.

Table 3. Percentage of retrieval accuracy with variations in weight of features and variations in number of clusters.

| Number of clusters | Equal Weight | W1  | W2  | W3  | W4  |
|-------------------|--------------|-----|-----|-----|-----|
| 3                 | 72.55%       | 75.59% | 92.67% | 82.38% | 63.33% |
| 4                 | 72.67%       | 75%   | 92.33% | 82.33% | 63.67% |
| 5                 | 73.05%       | 76.45% | 93.33% | 83.36% | 64.00% |
| 6                 | 73.69%       | 76.15% | 93.67% | 83.67% | 64.33% |
| 7                 | 74.65%       | 77.75% | 94.33% | 84.38% | 65.67% |
| 8                 | 74%          | 77.50% | 94%   | 84%   | 65.00% |
| 9                 | 74.76%       | 77.45% | 94.67% | 83.87% | 64.33% |
| 10                | 75.89%       | 79.65% | 95.67% | 86.67% | 66.93% |
| 11                | 74.47%       | 79.15% | 93.55% | 84%   | 66.23% |
| 12                | 74.15%       | 77.25% | 93%   | 85.33% | 65.67% |
| 13                | 74.20%       | 76.15% | 92.23% | 82.63% | 64.63% |
| 14                | 73.25%       | 75.59% | 92.13% | 82.44% | 64.55% |
| 15                | 73.13%       | 75.29% | 92.09% | 82.23% | 63.43% |

Table 3 shows that the weighting variation of W2 with a weight of 60% of color features, and 40% of color features and variations in the number of clusters 10 produces the highest retrieval accuracy of more than 95%.

While the weighting variation of W4 with 40% weight of color features and 60% weight of color features with variations in the number of clusters 3 produces the lowest accuracy which is less than 64%.

Retrieval accuracy graphs in the rice image database using the same and varied feature weights namely W1, W2, W3, and W4 with variations in the number of clusters ranging from 3 to 15 are shown in Figure 6.

![The percentage of retrieval accuracy with weight variation of features and the number of clusters](image_url)

Figure 6. Retrieval accuracy graph with feature weight variation and cluster number variation.

Figure 6. is the retrieval accuracy that tends to rise until reaching variations in the number of clusters of 10 and tends to decrease after the number of clusters is greater than 10.

3.3. Computation time
The computational time that takes place to retrieve an image in a rice image database is influenced by the size of the image database and its database management techniques. The average computational time for searching in an image database before clustering is with the variation in the number of clusters shown in Figure 7.
Figure 7. Time computation of retrieval.

Figure 7 explains that the greater the number of clusters, the faster the computing time.

4. Conclusion
The level of accuracy in the process of retrieval of rice images is influenced by several variables including image acquisition, preprocessing, feature extraction methods, and identification techniques in the database. This study proves that the use of different feature weighting schemes for each normalized feature value and using a variety of cluster numbers will affect the increase in accuracy and retrieval speed in the process of identifying rice types. The highest level of accuracy obtained in retrieval is the variation of the number of clusters 10 with a weighting variation of W2 that is 60% color features and 30% shape features. From this study, it was found that the color features of rice images had a more dominant factor in determining the level of similarity in the process of identifying types of rice. As for the form features as a complementary feature. The average computing time required for retrieval is 5 milliseconds.

References
[1] Vadivel A, Majumdar AK and Shamik S 2004 Characteristics Of Weighted Feature Vector In Content-Based Image Retrieval Applications International Conference on Intelligent Sensing and Information Processing IEEE.
[2] Gonzales R C and Woods R E 2008 Digital Image Processing ThirdEdition (New Jersey: Pearson Prentice Hall)
[3] Castleman K R 1996 Digital Image Processing (New Jersey: Prentice Hall Inc.)
[4] Jumi and Harjoko A 2012 Image Similarity Analysis Based on Shape, Color and Texture Feature of Asset Image International Conference on Computer Science Electronics and Instrumentation, Yogyakarta, Indonesia.
[5] Xing-yi Huang, Jian Li and Song Jiang 2004 Study on identification of rice varieties using computer vision [J]” In: Journal of Jiangsu University (National Science Edition) 2.003
[6] Harpreet Kaur and Baljit Singh 2013 Classification and grading rice using multi-class SVM” In: International Journal of Scientific and Research Publications 3 1–5
[7] Guzman J D and Peralta E K 2008 Classification of Philippine rice grains using machine vision and artificial neural networks In: World conference on agricultural information and IT, IAALD AFITA WCCA 2008, Tokyo University of Agriculture, Tokyo, Japan, 24-27 August, 2008. Tokyo University of Agriculture. 2008, pp. 41–48.
[8] Zhao-yan Liu, Fang Cheng, Yi-bin Ying, and Xiu-qin Rao 2005 Identification of rice seed varieties using neural network”. In: Journal of Zhejiang University. Science. B 6.11 1095
[9] Chathurika Sewwandi Silva and Upul Sonnadara 2013 Classification of Rice Grains Using Neural Networks” In: Proceedings of Technical Sessions. 29 9–14
[10] Singh S K, Bejagam K K, An Y and Deshmukh S A 2019 Machine-learning based stacked ensemble model for accurate analysis of molecular dynamics simulations The Journal of Physical Chemistry
[11] Acharya T, Ray A K 2005 Image Processing Principle and Applications (USA: John Willey &
[12] Susilo A 2007 Web Image Retrieval untuk Identifikasi bunga dengan Pengelompokan Content Warna, Institut Teknologi Sepuluh November, Surabaya.