E-ISSN : 2407 – 3911
P-ISSN : 2686 - 0333

IMAGE PROCESSING BASED TILAPIA SORTATION SYSTEM USING NAÏVE BAYES METHODS

Sukenda[1], Ari Purno Wahyu W[2], Benny Yustim[3], Sunjana[4], Yan Puspitarani[5]
Informatics Departement, Faculty of Engineering[1,2,3,4,5]
Widyatama University
Jl CIkutra no 204A, Bandung
kenda@widyatama.ac.id[1], ari.purno@widyatama.ac.id[2], byustim@widyatama.ac.id[3],
sunjana@widyatama.ac.id[4], yan.puspitarani@widyatama.ac.id[5]

Abstract

Tilapia has a value of export quality and is imported from America and Europe, tilapia is cultivated in freshwater, the largest tilapia producing areas are Java and Bali for the export market in the Middle East, value fish with a size of 250 grams / head (4 fish / kg) in their intact form is in great demand. According to news circulating, fish of this size in the Middle East are ordered to meet the consumption of workers from Asia. the fish classification process is a very difficult process to find the quality value of the fish to be sold to meet export quality. Fish classification techniques can use the GLCM technique (Gray Level Oc-Currance Matrix) classification using images of fish critters with the GLCM method.The fish image data is analyzed based on the value of Attribute, Energy, Homogenity, Correlation, Contrash, from the attribute the density data matrix is generated for each. Fish image data and displayed in the form of a histogram, the data from the GLCM results are then classified with the Naive Bayes algorithm, from the results of the classification of data taken from 3 types of tilapia from the types of gift, Red, and Blue.

Keywords: Fish classification, GLCM, Naive Bayes

I. INTRODUCTION

Fisheries sector currently plays an important role as a support for the country's economy after sector oil and gas, types of fisheries in Indonesia are divided into two parts namely freshwater and saltwater fisheries, conditions Indonesia which has tropical temperatures suitable for aquaculture bargain, one of the cultivated fish that has value high economical is tilapia, this type of fish is already including long cultivated in Indonesia, fish farming areas tilapia many areas that have wide lowlands between Indramayu Area, Cirebon, Cilacap and North Coast Area Java, there are types of value fish cultivated in Indonesia. several types namely best tilapia, srikandi tilapia, tilapia wanayasa and the last one the red tilapia or janti type. take and feed for tilapia is very easy, the fish lives with 21^0 c water conditions, fish sometimes eating insects and other microorganisms so that the feeding process is not too difficult Tilapia fish has a very high protein content and contains lots of omega 3 and very good for consumption by children, tilapia itself is widely exported to various countries in America and Europe and East Asian countries such as Japan.

II. THEORITICAL BASIS

Statistical techniques or classification can use the Naive Bayes Algorithm with a reading and probability value of a class value [1]. The classification capabilities of the Naive Bayes Algorithm are the same as the Decision Tree and Neural Network Algorithms and are able to read large amounts of data in databases [2]. Naive Bayes algorithm is widely used statistical algorithm and induction reference and can be used to predict a class [3]. Calculation of the Naive Bayes Algorithm can use training data that can be used to estimate data by reading the attribute values of a class, Naive Bayes can perform data calculations and calculate estimates from missing data will be ignored [4]. The definition of data mining itself is a process of finding useful information in large databases. Content techniques are also known as Knowlegde Discovery [5]. Training data by predicting existing data by reading a training dataset that will produce data classification techniques through establishing data correspondence or supervised learning [6].The

Naiva Bayes Algorithm theorem can be formulated as follows [7].

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

Figure 1. Naive Bayes theorem formula

the probability value X that is tangent to the Y axis where this value has the probability value of X and YP (X | Y) is the presentation value of the number of members Y and X.

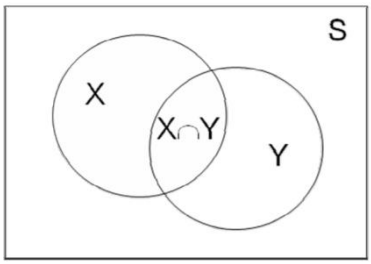


Figure 2. Probability Of Tangents

II.1 B.Teknik Glcm(Gray Level Confusion Matric)

To analyze a classification process, you can use three-stage texture analysis, the method is the Spactral method, the texture analysis method has three stages, namely the spactral, structural and statistical methods, these three methods are commonly used for classification by reading the value of the degree of gray which will be displayed in the form of a histogram. [8]. The direction of the pixel value is represented by a value of 1350,900,450 and 00, the pexel value illustrates the matrix value.The illustration of statistical feature extraction is shown in Figure 1 [9].

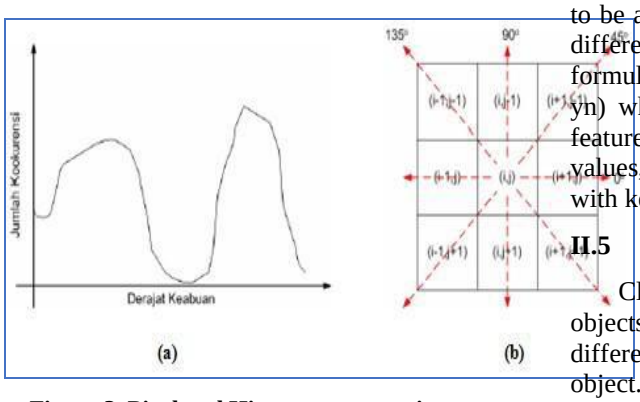


Figure 3. Pixel and Histogram extraction results

Color Segmentation

Color segmentation can be called a part of a color partition tool that is separated by components or pixels or represents a different image and has the meaning that the data must be able to be analyzed based on boundary or line or line values, color segmentation itself is the process of labeling to describe the characteristics of an object [10].

II.2 GLCM

The GLCM method is a method that shows a statistical form of data from the elements of a pair of an image pixel, the pixel distance is separated in an angle called the inclination, the probability value of an image is assessed from the grayscale attribute. [11].

II.3 Tecture Analysis

Tecture can be interpreted as a surface property of an object, to read the tecture of an object is very difficult because each object has a different data variable value, the value of a tecture is strongly influenced by several factors, namely structural, probabilistic, or functional. Texture images can be fine, coarse, smooth, granular, random, linear, motled, irregular, hummocky is a common way of analyzing statistical textures [12].

II.4 Support Vector Machine (SVM)

Support Vector Machine is a method that can be used for classification by reading statistical data, and has a high level of accuracy and is very effective even though it uses very little training data. [13].

$$h_j = \begin{cases} 1 \;\; if \;\; p_j f(x) < p_j \theta_j \\ 0 \;\; otherwise \end{cases} \qquad (1)$$

SVM is the best method for classification with pattren techniques or pattern matching and image classification, the algorithm is designed to be able to calculate separate training data for different data classes which can be explained in formula (1) above (x 1, y1), (x2, y2), ..., (xn, yn) where the xiin Rd value, the d-dimension feature space, and yi in {-1, + 1}, the class label values, where i = 1..n. SVM value optimization with kernel addition (K) [14].

II.5 Classification

Classification is a technique for analyzing objects with function values and models from different objects distinguishing classes in an object. how to find the average accuracy value on an object can use the Confusion Matrix

attribute. basic measurement to measure the quality of the text invention [15] [16]:

- Precision: whether a method is appropriate for classifying an object.
- Recall: pengalan an object as a whole from an object that has been recognized.
- F-Measure is

  is a formula value that is described from the measurement results of the pressian and class recall:

$$\text{Precision } (i,j) = \frac{n_{ij}}{n_j}$$

$$\text{Recall } (i,j) = \frac{n_{ij}}{n_i}$$

Figure 4: Precision and Recall Formulas

II.6 2.7 Clustering

A clustering process can be likened to previous objects, the separation of an object can be done by dividing some data partitions into several clustering can be displayed visually which have been grouped into specific target data. [17].

III. REASEARCH METHODS

Modeling is a phase that directly involves data mining techniques, namely by selecting data mining techniques and determining the algorithm that will be used, the purpose of this final project is to build a system model that processes data starting from tilapia image processing to producing analysis output of image pattern recognition results. processed into digital form or numbers (normalization matrix). The image is converted into digital form so that it can be processed using Principle Component Analysis (PCA), namely the eigen image algorithm which produces dominant characteristics so that it represents the structure of the image pattern. Then, testing and validating whether the output results are in accordance with the image we are analyzing in the form of a histogram.

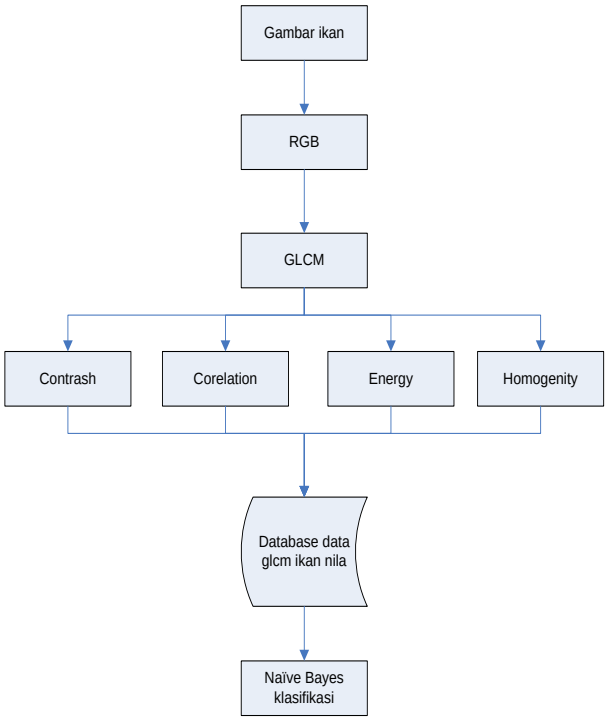


Figure 4 Stages of Research Methods

III.1 Research Method Stages

1. Image capture of fish

   Taking pictures of tilapia critra can be taken through a photo shoot or using an existing tilapia image as a sample, samples are taken from three types of tilapia namely GIFT, RED, and BLUE, for each type 10 samples are taken.
2. RGB data processing

   In this section the tilapia crita taken has a white color or image background, through this stage the background color will be removed using the Grayscale and Treshold methods so that the noise image will be reduced and does not affect the main image that will be processed using the GLCM method.
3. Fish image database

   The collected image data of the tilapia is then processed by the GLCM method where the data from the fish image will be displayed in a histogram form of the pixel density of each image using several attribute parameters, namely;
4. Entropy Entropy "is used in this process. Entropy is a measure of data disorder. Entropy is measured in bits. This method is called the measurement

of uncertainty in any random variable. (Girja, Bhargava & Mathuria, 2013)

$$Entropy(p) = -\sum_{j=1}^{n} \frac{|pj|}{|p|} \log \frac{|pj|}{|p|}$$

$$Entropy(j|p) = \frac{|pj|}{|p|} \log \frac{|pj|}{|p|}$$

Figure 5 Entropy formula

5. information Gain is information for measuring input and output between associations. (Girja, Bhargava & Mathuria, 2013)

$$Gain(p,j) = Entropy(p - Entropy(j|p))$$

Figure 6 Information Gain formula

- Confusion Matrix

  Confusion Matrix contains actual information and predictive classification is carried out by classification system. The system performance uses the data in the matrix. (Girja, Bhargava & Mathuria, 2013)

- Confidence

  Confidence is a measure that assesses the degree of certainty the association is detected, it is taken as the conditional probability P (Y | X), i.e. the probability that a transaction containing X also contains Y.

- Support

  Support is a measure that represents the percentage of transactions from the transaction database given satisfactory rules, this is taken to be probability P (XUY), where XUY indicates that the transaction contains both X and Y, i.e. a set of X and Y itemsets.

IV. RESULT

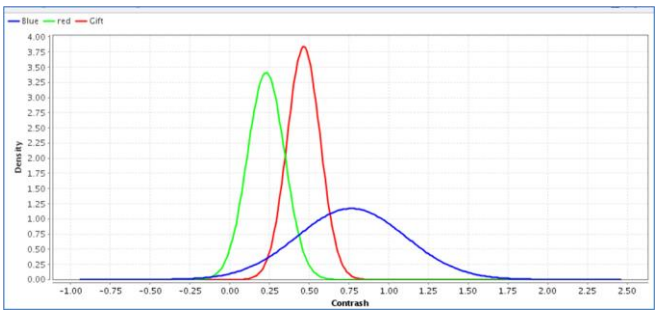
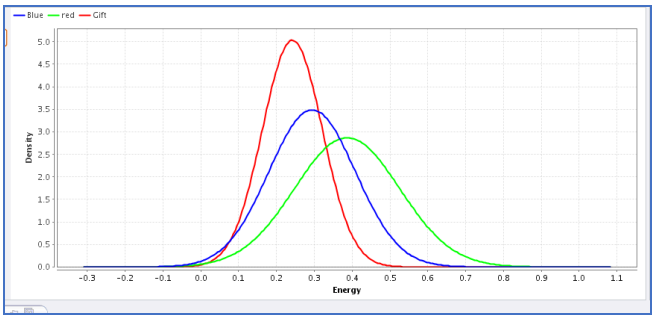


Figure 7 Contrash Calculation on Tilapia Image Nila

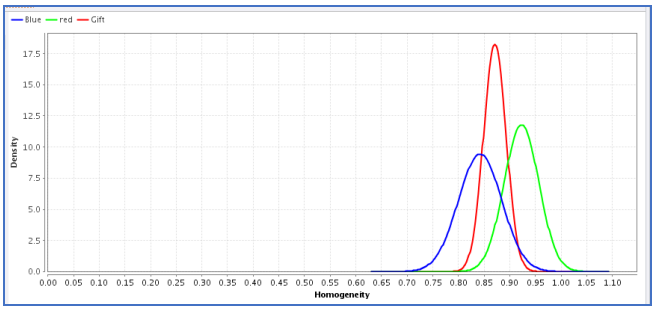
Information : Measuring the spatial frequency of the image and the GLCM moment difference. The difference in question is the difference between the height and the low pixel. The contrast will be 0 if the neighboring pixels have the same value

Table 1.1 Calculation of Energy in Indigo



Note: In table 1.1 measuring uniformity or often called the angular second moment. Energy will be high when the pixel values are similar to each other, on the other hand, it will be small, indicating that the value of normalized GLCM is heterogeneous. The maximum value of energy is 1 which means the distribution of pixels is in constant condition or its shape is periodic (not random)

Table 1.2 Calculation of Homegenity Value on the image of tilapia



Note: In table 1.2 measuring homogeneity. This value is very sensitive to values around the main diagonal. High value when all pixels have

the same value / uniform. The opposite of contrast is that it will be of great value if it has the same pixel value when energy is fixed

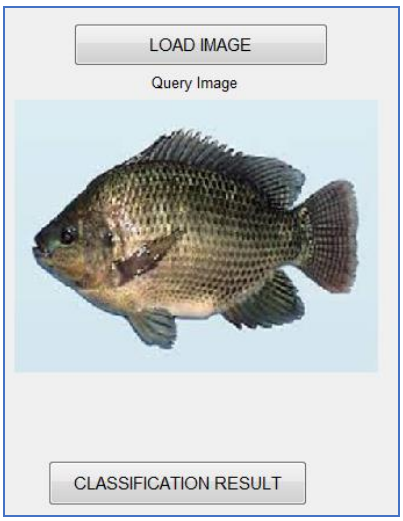


Figure 8. Sorting Testing

Description: In the picture 8 above is the process of testing a system that is connected to a database system to test accuracy, in this accuracy process the system uses the naïve Bayes and neural-network algorithms, in the sorting process the system shows an accuracy of 98%.

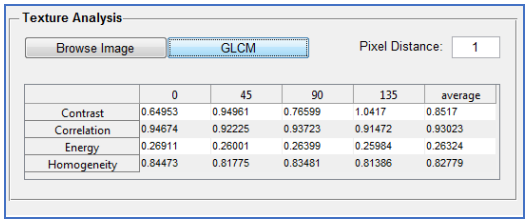


Figure 9 Testing with the GLCM method

Note: In Figure 9 is the process of testing the system on the tilapia image, the fish data is then converted into a pixel by the GLCM method with a different enbtropy value, the entropy data is then stored into a database to be compared with the training data, from which the system will provide response in the form of tilapia fish and accuracy of data checking.

V. CONCLUSION

Based on the results of the analysis and discussion described in the previous chapter, several conclusions can be drawn as follows: Texture analysis method using the statistical gray level co-occurrence matrix method and the Naïve Bayes Algorithm produces a features value consisting of several attributes that can be used as a value for classification of types of tilapia and their quality. From some of the existing data, it can be concluded that each type of tilapia has different features values.

Where the results of the mean and standard deviation obtained have met and are close to the true value in the classification of the quality of tilapia using the naïve Bayes method, it can be implemented. Therefore, 5 attributes, namely Entropy, Energy, Contrast, Homogeneity, and Correlation can be used as variables for classifying high-quality tilapia species. So that the results of texture analysis can be more optimal in determining the attributes for the classification of Tilapia species. And it is also hoped that in the next development this method can present the data from the feature extraction using GLCM into a table to make it easier to learn.

REFERENCE

Fritz Albregtsen, "Statistical Texture Measures Computed from Gray Level Coocurrence Matrices," Oslo, 2008.

Kusrini ; E. T. Luthfi, Algoritma Data Mining, 1 ed., Yogjakarta: ANDI, 2009.

R.W. Conners and C.A. Harlow,, "A theoretical comparison of texture algorithms," IEEE Trans. on Pattern Analysis and Machine Intell, pp. 204-222.

Sofi Defiyanti;Mohamad Jajuli, "Integrasi Metode Klasifikasi Dan Clustering," Konferensi Nasional Informatika (KNIF) 2015, 2015.

Pang-Ning Tan;Michael Steinbach;Vipin Kumar, Introduction to Data Mining Instructor's Solution Manual, Pearson Addison-Wesley, 2006.

F. Gorunescu, Data Mining Concepts Models and Techniques, Chennai: Springer, 2011, 2011.

E. Prasetyo, Konsep dan Aplikasi menggunakan MATLAB, Yogyakarta: ANDI, 2012.

R. M. Haralick K.; Shanmugam ;I. Dinstein, "Textural features for Image Classification," IEEE Transactions on Systems man and Cybernetics, pp. 610-621, November 1973.

T. W. A. Putra, PENGENALAN WAJAH DENGAN MATRIKS KOOKURENSI ARAS KEABUAN DAN JARINGAN SYARAF TIRUAN PROBABILISTIK, Semarang : Universitas Diponegoro Semarang, 2013.

BISWAROOP GOSWAMI, "TEXTURE BASED IMAGE USING GLCM".

P. Mohanaiah;P. Sathyanarayan; L. GuruKumar;, "Image Approach Tecture Extraction Using GLCM," International Journal of Scientific and Research Publications, vol. 3, no. 5, May 2013.

R. C. Gonzalez ; R. E. .Wood, Digital Image Processing, 2 ed., 2002.

Gidudu Antony;Hulley Greg; Marwala Tshilidzi, "Classification of Images Using Support Vector Machine," Department of Electrical and Information Engineering, University of the Witwatersrand, Johanesburg, SoutAfrica, 2013.

D. Lu, Q. WENG,, "A survey of image classification methode and techniques for improving classification perfromance," International Journal of Remote Sensing, vol. 28, no. 823-870, p. 5, 2007.

Olson ; david ;young;shi, Penggalian Data Bisnis, Jakarta: Salemba empat, 2011.

J. Gua ;J. Chena;Q.M. Zhouc; H.W. Zhanga ;L. Mad,, Quantitative Textural Parameter Selection For Residential Extraction From High-Resolution Remotely Sensed Imagery, Beijing, 2008.

Han, J., ; Kamber, M., Data Mining: Concepts and Techniques, 2 ed., San Francisco: Morgan Kaufmann,, 2006.