Proceeding of International Conference on Green Technology
Vol.8, No.1, October 2017, pp. 446-453
p-ISSN: 2580-7080 – e-ISSN: 2580-7099

Classification of Egg Fertility on the Image of Kampong Chicken Egg Using the Frequency Distribution Feature and Naive Bayes Classifier Algorithm's

Aris Diantoro[1], Irwan Budi Santoso[2]
Departement of Informatics Engineering, UIN Maulana Malik Ibrahim Malang
[1]arisdiantoro1@gmail.com,[2]irwan@ti.uin-malang.ac.id

Article Info

Article history:

Received Jul 12th, 2017
Revised Aug 20th, 2017
Accepted Oct 26th, 2017

Keyword:

Eggs
Fertile
Infertile
Frequency Distribution
Fuzzy C Means
Naive Bayes Classifier

ABSTRACT

Losses in hatching chicken eggs make breeders income decreased. The main cause of this because it is less effective and efficient in distinguishing the fertilities state in eggs. Detection of fertile and infertile eggs will automatically facilitate the selection and removal of fertile eggs and infertile eggs. This will bring more benefits to farms such as time efficiency and more selling power. Infertile eggs will provide the selling price if known as early as possible so as not to fail hatched. A Fuzzy C Means method and the Naive Bayes Classifier are designed to identify the fertility state of the egg. By putting eggs near a light source as well as a black background in dark spaces, then take the image with a high quality camera. From the resulting image the camera extracted features or distinctive features that distinguish between fertile and infertile eggs. The total amount of data used in this study is 350 eggs from the field survey. Training data used 250 data, 125 fertile egg image data and 125 infertile egg image data. While for data testing using 100 data, 50 data fertilized egg image and 50 data image infertile egg. Based on the results of the training data trials obtained the best accuracy is 93.2% at interval 4 with RGB feature, 50% at interval 3 on Grayscale feature. Accuracy results in the data testing test obtained by 87% on RGB features, 35% on Grayscale features.

Copyright © 2017 Green Technology.
All rights reserved.

Corresponding Author:

Aris Diantoro,
Departement of Informatics Engineering,
UIN Maulana Malik Ibrahim Malang,
Jl. Gajayana No. 50 Malang, Jawa Timur, Indonesia 65144
Email: [1]arisdiantoro1@gmail.com

1. INTRODUCTION

The Food in the form of chicken meat for the community is inseparable. Chicken meat is one commodity that has the potential to improve the economy of the community [1]. According to data from BKP (Food Security Agency) Ministry of Agriculture in 2013 to 2015 chicken eggs and poultry meat has ranked number 2 of 3 commodities of animal food needed by the community. Projected in the higher than the existing realization, this proves that the needs of both foodstuffs from year to year higher. The high demand that is not accompanied by optimal production hence the welfare of society so decreasing.

According to BPS (Central Statistics Agency) data from 2009 to 2015 stated that the production of eggs and poultry meat (not race) is less than the chicken race. This makes the price of eggs and chicken meat (not race) to soar. High demand is higher than the amount of production then the worst thing that happens is the commodity of domestic poultry (not race) into an animal commodity that is difficult to be fulfilled. in general, poultry chicken is traditionally managed to result in low productivity of chicken poultry. On the other hand, the increasing population and nutritional needs for the citizen, the need for domestic poultry, both meat and eggs are increasing. Therefore, traditional chicken breeding is done traditionally, need to be intensified again in its maintenance system.

In order to maintain the population of chickens, both laying and broilers, steps taken by hatching eggs [2]. Hatching eggs need quality seeds to get quality chicken results and resistant to disease. Failure in hatching eggs into its own problems in the world of animal husbandry, on a large scale that has been intensively breeding domestic poultry (not race). Not hatching eggs on the 21st day makes losses breeders, since eggs that can not hatch should be thrown away in vain and do not have a sale price. Losses that are not comparable with the issued capital makes the farmers switch to easy-ranch chicken that is resistant to disease.

knowing the state of fertilities as early as possible becomes the best solution for avoiding hatch failure, fertilized egg failure due to fungal disorders from unidentified infertile eggs and throwing egg is vain. Egg fertility checking is done by candling the egg itself, the egg is brought close to a light source with a certain intensity sufficient to penetrate the eggshell, but the accuracy of detection depends on the experience of the observing workforce, if hatching is done on an industrial scale it will take much labor experienced. In addition, the use of time becomes inefficient. Subjective assessment makes observed eggs not maximal in differentiating egg fertilities.

Several techniques have been used in research to detect fertile and infertile eggs among others Least aquare Support Vector Machine by Zhiu Zhu et. al[3], Hyperspectral imaging and Predictive modeling System oleh Smith et. al[4], K-Means Clustering by L. Liu and M. O. Ngadi [5], Neural Network oleh Qiaohua Wang, et al[6], jaringan syaraf tiruan sebagai pendeteksi telur infertil by Nawawi[7].

Naive Bayes Classifier algorithm as fertility identification in eggs. Research contributions can identify egg types based on their fertility, reducing losses to farming. Further development can be directed to the selection of eggs automatically and based on mobile for the learning process.

1.1. Theorem of Naive Bayes

Bayes Rule is a rule that is correct or revise a probability by utilizing additional information. That is, from initial probability (prior probability) that have not been repaired formulated based on information currently available, and then formed the next probability (posterior probability) [9]. When the feature is known ω_j class j, and x is the object features (pattern x), then the probability ω_j bila unknown can be formulated in equation [8]:

$$p(\omega_j|x) = \frac{p(\omega_j)\, p(x|\omega_j)}{p(x)} \quad (1)$$

value $p(\omega_j|x)$ equivalen with of value $log\left(p(\omega_j|x)\right)$, so :

$$log\left(p(\omega_j|x)\right) = log\left(\frac{p(\omega_j)\, p(x|\omega_j)}{p(x)}\right) = log\left(p(x|\omega_j)\right) + log\left(p(\omega_j)\right) - log(p(x)) \quad (2)$$

1.2.Singularitas Naïve Bayes Against Gaussian Classifier

Naïve Bayes can be assumed into the special model p (x | θ), hereinafter naïve Bayes method to estimate population distribution parameter θ = (μ, Σ) to model Multivariate Normal distribution (Gaussian). One from the simplest approach to avoid a single value parameter covariance matrix is to use diagonal covariance matrix. Based on this it can be assumed that each feature is independent so that equation [9]:

$$p(x|\omega_j) = \prod_{l=1}^{d} N(x_1; \hat{\mu}_{jl}, \hat{\sigma}_{j,l}^2) \quad (3)$$

Where $\hat{\mu}_{jl}$ a component to-l from $\hat{\mu}_j$ (component to-l based on the mean of sample data from class ω_j), $\hat{\sigma}_{j,l}^2$ is the diagonal elements all l from Σ*j* (component to-l based variants of data samples from class ω_j) and N (*x*, *μ*, $\hat{\sigma}^2$) is the probability density function from the distribution model Univariate Normal (Gaussian) with mean μ and variance $\hat{\sigma}^2$. The theorem can direpsentasikan into the discriminant function, so by using conditional probability x against ω_j if $g_i > g_j$, for all $i \neq j$ obtained discriminant function [9] as follows:

$$g_j(x) = log\left(p(\omega_j)\right) - \sum_{l=1}^{d} log(\hat{\sigma}_{j,l}) - \frac{1}{2}\sum_{l=1}^{d} \frac{(x_l - \hat{\mu}_{jl})^2}{\hat{\sigma}_{j,l}^2} \quad (4)$$

with

$$p(\omega_j) = n_j / \sum_{i=1}^{c} n_i$$

Based on the equation 4, the rules for classifying the *pattern x* is:

$$\hat{C} = \arg max\left(g_j(x)\right), j = 1, \ldots, C \quad (5)$$

Where $\hat{C}$ is the chosen class id [10].

1.3.Frequency distribution

The data that have been obtained from a research that is still a random data that can be made into data in groups, ie data that have been compiled into specific classes. A list containing grouped data is called frequency distribution or frequency table. The frequency distribution is the arrangement of data by class of specific intervals or by certain categories in a list [11].

The steps - steps in making the frequency distribution table is as follows:

a. Specifying Range (R)

If there are n data that is x_1, x_2 ,, ..., x_n ,, then R = (maximum x value) - (minimum x value). Max value x means that the data has the largest value while the minimum x is the data that has the smallest value.

b. Determine the number of class intervals (k)

The number of class intervals, usually best between 5-15 class intervals or by using the Sturgess rule k = 1 + 3.3 log (n), where n is the number of data.

c. Specifies the interval width (i)

The interval width (i) is obtained by dividing the range (R) by the number of class intervals (k) and mathematically written i = R / k. From the width of this interval, we can then set the lower limit of the interval and the upper limit of the interval for each class interval.

d. Determine the amount of data that goes into each class interval (tally).

e. Determining the frequency (f)

Frequency is the number of incoming data at each class interval.

2. RESEARCH METHOD

Stages in this study include several stages in the outline of system design, data collection, system implementation, and testing.

2.1.Framework System

The design of the system to be constructed in this study includes 2 main parts is egg fertility detection (training and testing) as shown in Figure 1.

In the training process, the first step is to take the egg image data, then the initial process is cropping by taking the required image, segmentation with fuzzy c means with the aim of clustering the selected part and the part that is not, feature extraction to get the characteristic of the object with frequency distribution, as well as estimating parameters with mean and variance values as parameters in recognition with the Naive Bayes Classifier.

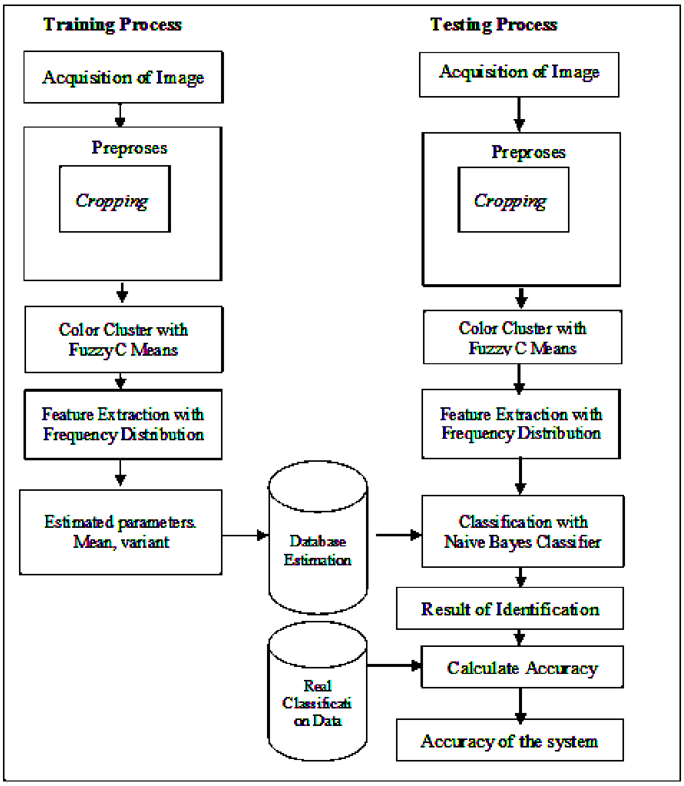

Figure 1. System design

The process of testing with the first step of data retrieval, cropping process to take the specified area, segmentation to chose the part used, then the feature extraction process to take the characteristics of the object with frequency distribution, and then classification Naive Bayes Classifier in accordance with the parameter values obtained from the results of parameter estimation In the training process.

2.2.Data Collection Instrument

The research instrument used to obtain egg image data. A specially designed tool that adjusts the camera's ability and accuracy of object position. This research instrument can also be called candling process

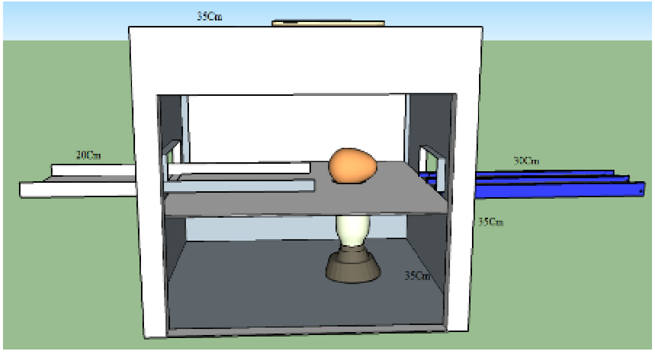

Figure 2. Data Acquisition Design

The tool is designed have with three parts, namely the bottom as the center of irradiation, the middle as a place to lay the object of research, and the top where the camera as data retrieval tool. Such a tool has a fixed provision that the object is placed above the light rays of the lamp, the room with a light impermeable from the outside, then the inside of the egg will a visible the egg components to be taken with the camera.

2.3.Model Feature

1) *RGB (red, green, blue):* The main component in an image consisting of 3 colors ie red, green and blue. Extract feature by taking the RGB value directly, without changing the value by making the RGB value on be average.

2) *Grayscale:* Image with gray color representation becomes an alternative in describing feature to light intensity at object. to Make the main component of RGB color into 1 channel color that is grayscale color. So it can be written an equation:

$$\text{Grayscale} = 0.299R + 0.587G + 0.144B \tag{6}$$

3) *Frequency Distribution:* The frequency distribution tables is very suitable for large data sizes. Large data is data with very features many , usually up to thousands and even millions, so it is effective to be presented with a frequency distribution table [12]. Steps in making the frequency distribution table among others determine the range, determine the number of class intervals, determine the interval width, determine the amount of data that goes into each class interval, determine the frequency.

The determination of the upper and lower limits in this study, among others, is based on testing of a relevant egg image and then searched for the maximum value and minimum value in the pixel value containment in the image. The Red value has a lower limit of 0 and the upper limit of 255, the Green value has a lower limit of 0 and the upper limit of 134, the Blue value has the lower limit of 0 and the upper limit of 61 and the Gray value has a lower limit of 0 and the upper limit of 135. The following is shown Figure 3 Boundary determination Upper and lower bounds.

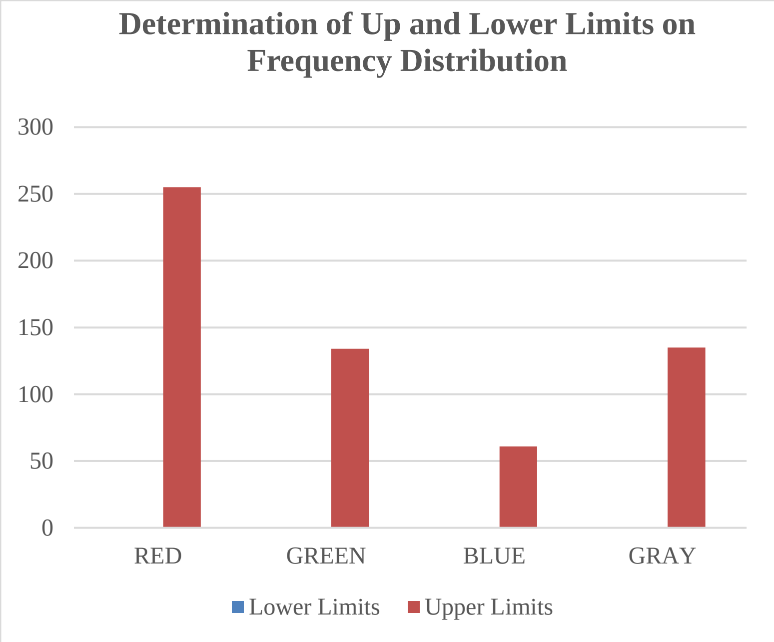

Figure 3. Determination of Up and Lower Limits on Frequency Distribution

4) *Parameter Estimation:* If dealing with continuous data, Object features are assumed to be normally distributed (gaussian), with parameter models of mean and variance. So can get the equation of the mean parameter[8]

$$\mu_{ic=\frac{1}{n}} \sum_{j}^{n} F_{ij} \tag{7}$$

while the equations used to calculate variants [8]

$$\sigma_{ic}^{2} = \frac{1}{n-1} \sum_{i=j}^{n} (F_{ij} - \mu_{ic})^{\ 2} \tag{8}$$

.

3. RESULTS AND DISCUSSION

3.1.In Data Descripsions

The data to be used in this study is the primary data, obtained from the data retrieval independently by the researcher. The research data are data of chicken eggs obtained from several respondents from various farms that have two types of fertility, namely fertilized chicken egg and infertile chicken egg. Determination of fertility of this egg manually, after the candling process appears directly determined type, this data is then used for the learning process. So the survey is done by taking a sample of chicken eggs in the of Nyawangan Village, Kras district, Kediri district.

The total data that will be used in the research is 350 data, the training data are 250 image of chicken eggs including 125 fertile eggs and 125 infertile eggs. A total of 100 data testing.

3.2.Test of System

In the research of Fertilities Detection System of Egg on Image of Chicken Egg Using Algorithm of Naive Bayes Classifier, there are 2 stages in the experiment, first, data of result capture fertile egg and infertile egg with data acquisition tool in Figure 2. Second after getting good data, of egg Fertil and egg infertil, then the data is sorted into 2 different directories with data categories, training data is data that will be studied system to get estimation of parameters and data testing, data will be tested system previously that data has never been studied system.

The system test outline covers 2 ways, each way has a value comparison between RGB and Grayscale. The first test is a test with frequency tabulation by changing the interval value on the RGB feature, the second test is a test with frequency tabulation by changing the interval value in the Grayscale feature.

3.3.Results of Training Process

The results of the training test with 2 ways of testing on 250 data in the system studied for each class. Different training test results are obtained for each of the experimental ways to image the chicken eggs by treatment of RGB data features and Grayscale data features.

Table 1 Frequency Distribution Training Results RGB feature

Interval	Accuracy (%)	Error (%)
3	80.00	20.00
4	93.20	6.80
5	72.40	27.60
6	70.00	30.00
7	50.00	50.00
8	68.40	31.60
9	50.00	50.00
10	68.40	31.60

Table 2 Frequency Distribution Training Results Grayscale feature

Interval	Accuracy (%)	Error (%)
3	50.00	50.00
4	17.60	82.40
5	8.00	92.00
6	11.20	88.80
7	50.00	50.00
8	17.60	82.40
9	50.00	50.00
10	19.20	80.80

Based on table 1 it is finded that the RGB feature in the training process of the system using the difference in value at the interval has the most optimal value at interval 4 that has an accuracy value of 93.20%. Has the least accuracy value at intervals 7 and 9 that is 50.00%.

Based on table 2 it is finded that the Grayscale feature in system training process that use difference value at interval has the most optimal value at intervals 3, 7 and 9that is 50.00%.

From the analysis results of table 1 and table 2, it is finded that the various accuracy values are required to choose the most optimal for the classification process or the testing process. So the data parameter estimation that will be used classification process with naive bayes classifier method that is interval 4 on RGB feature and interval 3 on Grayscale feature as comparator value of test process accuracy.

3.4.Result of Testing

The test results are done with data that have been prepared to be tested with the amount of test data of 100 data. Data testing in this study has never experienced the previous learning process. The testing process uses parameter estimation results from the most optimal training process. Next will be compared to then calculated the great degree of accuracy and error. The results of identification of testing based on trial variation are shown in Table 3 below

Table 3 Test results

Type of trial	Accuracy (%)	Error (%)
RGB Features	87.00	13.00
Grayscale Features	35.00	65.00

Results test against 2 ways of comparing the accuracy of RGB features and Grayscale features, the best accuracy value is RGB feature with 87.00% accuracy.

Accuracy in this testing process as the final value of the test phase of this research. So from the trials from intervals 3 to 10 and features RGB and Grayscale have varying values and only one will be selected as the best final value.

3.5.Analysis of Systems Work

The analysis obtained from the training process and testing test shows that RGB feature is superior to Grayscale feature. Some factors that can make RGB feature superior to Grayscale feature is, RGB has 3 color chanel that is red, green and blue. 3 color chanel will be read one by one against the frequency distribution tabulation so that the value of details is higher. The way the frequency distribution works is to calculate the feature value of the image data feature to the interval. For example the RGB feature on a fertile egg image that has matrix number 300,000 at interval 4 each RGB interval as follows 0-255 red intervals with 4 range, green interval 0-134 with 4 range, and blue interval 0-61 with the 4 range. Each container will be filled with values one by one according to the same value and contained in the arrays. So RGB has 12 grid values, 12 values are obtained from 3 x 4, 3 features and 4 intervals. This is different from Grayscale which only has 1 color chanel so its accuracy is less accurate.

CONCLUSION

In establishing Classification Application of Egg Fertility on the of Kampung Chicken Egg Image Using Naive Bayes Classifier The first classifier is data collection that will be used for training process and testing process. In the training process there are several processes performed, among others cropping, segmentatasi with fuzzy c means method, resize, feature extraction, tabulation of frequency distribution and parameter estimation. While in the process of testing the data also have the same process, after the tabulation of the frequency distribution is done classification process based on the calculation model with the method using of naive bayes classifier.

Based on the results of the experiment on the image training data chicken eggs, the data used are as many as 250 data where all the data is divided into 2 ie 125 data eggs fertil and 125 data infertile eggs obtained the best accuracy rate is 93.20% with process feature RGB (Red , Green and Blue). The results of testing data of chicken eggs with 50 data of fertil eggs and 50 eggs infertil data obtained the best accuracy is 87.00% with RGB feature process. Some data processing in the image acquisition process results in less accurate accuracy, due to several factors such as lack of intensity of candlesticking and outer shell cleanliness that will affect the segmentation and feature extraction process.

REFERENCES

[1] Supartono W, Yunus M, Yuliando H. “Analisis Kelayakan Finansial Usaha Pemotongan Ayam Tradisional di Daerah Istimewa Yogyakarta” Teknik Industri Pertanian Universitas Gadjah Mada. 2000.

[2] Susanto E, Suliswanto. “Pengaruh Berat Telur terhadap Daya Tetas Telur Ayam Kampung”. Program Studi Peternakan Fakultas Peternakan. ISSN 2086 – 5201

[3] Zhu Z, Ma Meihu. “The identification of white fertile eggs prior to incubation based on machine vision and least square support vector machine”. African Journal of Agricultural Research Vol. 6(12), pp. 2699-2704, 18 June, 2011.

[4] D. P Smith et al. “Detection of fertility and early development of hatching eggs with hyperspectral imaging”. USDA, ARS, Russell Research Center, 950 College Station Road, Athens, GA 30605, USA. 2010

[5] Liu L, Ngadi M. O. “Detection of Chicken Egg Fertility and Early Embryo Development Using Hyperspectral Imaging”. Department of Bioresource Engineering, McGill University, Macdonald Campus.

[6] Wang Qiaohu et al. “egg freshness detection based on digital image technology”. College of Engineering and Technology, Huazhong Agricultural University, Wuhan, 430070, P.R. China. 2009

[7] Nawawi M. Z., dkk. “Klasifikasi Telur Fertil dan Infertil menggunakan Jaringan Syaraf Tiruan Multilayer Perceptron berdasarkan Ekstraksi Fitur Warna dan Bentuk”. Jurnal Teknologi dan Informasi. Univesitas Sumatera Utara. Vol 4 No.2, Desember 2015:100-109

[8] Santoso, Irwan Budi, “Membangun Gaussian Classifier dalam Mengenali Objek dalam Bentuk Image,” Matics, vol. 1, pp. 1–5, 2014.

[9] R. Webb and K. D. Copsey, *Statistical Pattern Recognition*, 3rd ed., Mathematics and Data Analysis Consultancy Malvern, United Kingdom: John Wiley & Sons Ltd., 2011.

[10] Santoso, Irwan Budi. Deteksi Non-RTH(Ruang Terbuka Hijau) Kota Malang Berbasis Citra Google Earth dengan Menggunakan *Naïve Bayes Classifier*. MATICS : Jurnal Ilmu Komputer dan Teknologi. Universitas Islam Maulana Malik Ibrahim Malang. 2015
[11] Hasan, M. Iqbal. 2001. Pokok-pokok Materi Statistik I (Statistik Deskriptif), Bumi Aksara. Jakarta.)
[12] Santoso, Irwan Budi. “Statistika I untuk Teknik Informatika”. UIN-MALIKI PRESS. Malang. 2013.