Eggs Fertilities Detection System on the Image of Kampung Chicken Egg Using Naive Bayes Classifier Algorithm

Aris Diantoro, Irwan Budi Santoso

Abstract—Losses in chicken eggs hatchery make breeders income declined. The main cause of these things because it is less effective and efficient in distinguishing the state of fertilities in the eggs. The detection of fertile and infertile eggs will automatically provide ease of selection and removal of the eggs are fertile and infertile eggs. This will bring more profits for breeder as well as time efficiency more and selling power. Infertile eggs will give breeders the sale price if it is known as early as possible in order not to fail hatching. A method fuzzy c means and naive bayes classifier is designed to identify the state of the fertility of eggs. By putting eggs near the source light and black background in a dark room, then taked of image with a high qualities camera. From the resulting camera image, then extracted features or take characteristics that distinguish between fertile and infertile eggs. The total amount of data used in this study of 450 eggs image sourced from the field survey. Training data is used 250 data, 125 fertile eggs image data and 125 infertile eggs image data. As for testing the data using the 200 data, the image data 150 fertile eggs and 50 infertile eggs image data. Based on trial results of training data is obtained the best accuracy is equal to 80% at intervals of 5, 86.4% at intervals of 5 and dimensions 70x60, and 99.6% on 1x2 resize. The accuracy of the results obtained by 78%, 82% and 94% in trials testing data.

Index term—Egg, Fertil, infertile, Fuzzy C Means, Naive Bayes classifier.

I. PENDAHULUAN

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.

II. NAIVE BAYES CLASSIFIER

Naive Bayes is a development of the bayesian classification method. Bayesian classification is a statistical classification that can be used to predict the probability of membership of a class. Bayesian classification is based on a bayes theorem that has similar classification capabilities to the decision tree and neural network. In many applications it is often assumed that the opportunity function of each attribute or feature is Normal distribution (Gaussian)[8]

If known object features, $X = \{X_1, X_2, X_3, ... X_d\}$ And each feature is distributed Normal (Gaussian) then the chance of features or attributes with the known condition of the to-j class (C_j) is

$$p(x|\omega_j) = \prod_{k=1}^{d} N(x_k; \hat{\mu}_{jk}, \hat{\sigma}_{jk}) \qquad (1)$$

with $\hat{\mu}_{jk}$ and $\hat{\sigma}_{jk}$ Is the result of distribution parameter estimation (Normal) μ and σ for featrues to-k and k to j class. Under discriminant rules f on class C if $g_i > g_j$, for by j ≠ i Then obtained the formulation[9]

$$g_j(x) = \log\left(p(\omega_j)\right) - \sum_{k=1}^{d} \log(\hat{\sigma}_{j,k}) -$$
$$\frac{1}{2}\sum_{k=1}^{d} \frac{(x_k - \hat{\mu}_{jk})^2}{\hat{\sigma}_{j,k}^2} \qquad (2)$$

where ω_j Known as feature vectors in class j with vector mean parameters $\hat{\mu}_{jk}$ And covariance matrix. $\sum k$ Which is the result of estimation based on training data. Next to the process of classification of the pattern of data testing then used the rule of Equation 3

$$\hat{C} = arg_{cj}\, max\, P(C_j) \prod_{k=1}^{d} N(x_k; \hat{\mu}_{jk}, \hat{\sigma}_{jk}) \qquad (3)$$

with Cˆ Is the result of classification or detection[10].

III. RESEARCH METODHS

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

A. System Design

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 croping by taking the required image, segmentation with fuzzy c means with the aim of clustering the selected part and the part that is not, resize, feature extraction to get the characteristic of the object, as well as estimating parameters with mean and variance values as parameters in recognition with the Naive Bayes Classifier.
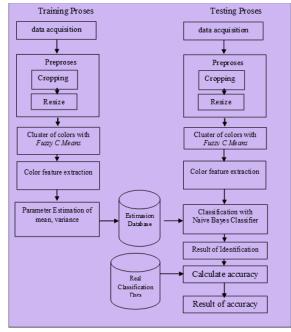

Figure 1. System design

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

B. Data Acquisition Instrumen

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
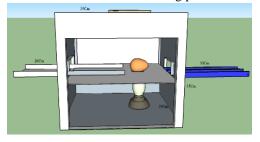

Figure 2. Data Acquisition Design

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.

C. Features Model

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:

$$Grayscale = 0.299R + 0.587G + 0.144B \qquad (4)$$

3) Distribution frequency: 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 [11]. 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.

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]

$$\mu_{ic=\frac{1}{n}} \sum_j^n F_{ij} \qquad (5)$$

while the equations used to calculate variants [8]

$$\sigma_{ic}^2 = \frac{1}{n-1} \sum_{i=j}^n (F_{ij} - \mu_{ic})^2 \qquad (6)$$

IV. RESULTS AND ANALYSIS

A. Data Description

The data to be used in this study is primary data, obtained from the data retrieval independently by the researcher. The data were data of chicken eggs obtained from several respondents from various farms that have two types of fertility, it is fertile 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 was done by taking samples of chicken eggs definitely in the hamlet Tejo, Nyawangan village, Kras subdistrict, Kediri district.

The total data that will be used in the research is 450 data, the training data are 250 image of chicken eggs include 125 fertile eggs and 125 infertile eggs. A total of 200 data testing.

B. Test of the system

In the research of Eggs Fertilities Detection System on the Image of Kampung Chicken Egg Using Naive Bayes Classifier Algorithm, there are 2 stages in the experiment. First, the take of data fertile egg and infertile egg with the data acquisition tool in Figure 2. Second after obtaining good egg data Fertil and egg infertil, then from the data is sorted into 2 different directories with data categories, training data is data that will be studied system for get estimation of parameters and data testing, data will be tested by system not yet previously data system it.

The system test outline covers 3 ways, each way has a value comparison between RGB and Grayscale. The first trial is a test with frequency tabulation without resizing the image dimension, the second test is a test with frequency tabulation and change the image dimension, and the third test is a test by changing the image dimension without frequency tabulation.

C. Result of Training test

The results of training trials with 3 methods test on 250 data in the system studied for each class. Different training results are obtained. For each method of testing the chicken egg image by treatment of RGB data feature and Grayscale data feature

Table 1 Results Frequency distribution training without changing the image dimension

Interval	RGB		Grayscale	
	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
3	62.80	37.20	16.40	83.60
5	80.00	20.00	15.20	84.80
7	71.20	28.80	13.60	86.40
9	70.80	29.20	12.40	87.60

Based on Table 1 it is found that the most optimal color data training feature is on RGB(Red, Green, Blue) color feature at interval 5 with 80.00% accuracy. As for the features of grayscale color has the most accurate accuracy that is with the accuracy of only 16.40% at interval 3.

Table 2 Results Training frequency distribution with changes in image dimensions

Interval	rows	column	RGB		Grayscale	
			Accuracy (%)	Error(%)	Accuracy (%)	Error(%)
3	20	10	65.60	34.40	30.40	69.60
3	30	20	61.60	38.40	26.40	73.60
3	30	30	61.60	38.40	26.40	73.60
3	40	30	61.20	38.80	23.60	76.40
3	50	40	61.20	38.80	20.80	79.20
3	60	50	61.20	38.80	20.00	80.00
3	70	60	61.20	38.80	19.20	80.80
3	70	70	60.80	39.20	19.20	80.80
3	60	60	61.20	38.80	20.00	80.00
3	80	70	61.20	38.80	18.40	81.60
3	80	60	61.60	38.40	18.40	81.60
5	20	10	78.00	22.00	26.80	73.20
5	30	20	64.40	35.60	25.60	74.40

5	30	30	72.00	28.00	22.00	78.00
5	40	30	78.80	21.20	21.60	78.40
5	50	40	76.00	24.00	19.20	80.80
5	60	50	75.00	25.00	18.00	82.00
5	70	60	86.40	13.60	17.20	82.80
5	70	70	76.80	23.20	17.60	82.40
5	60	60	70.80	29.20	17.20	82.80
5	80	70	82.80	17.20	16.40	83.60
5	80	60	77.20	22.80	16.80	83.20
7	20	10	76.40	23.60	23.20	76.80
7	30	20	67.60	32.40	20.80	79.20
7	30	30	76.40	23.60	18.40	81.60
7	40	30	75.20	24.80	18.00	82.00
7	50	40	78.40	21.60	16.80	83.20
7	60	50	73.20	26.80	15.20	84.80
7	70	60	81.60	18.40	14.80	85.20
7	70	70	85.00	15.00	14.80	85.20
7	60	60	71.60	28.40	15.20	84.80
7	80	70	82.40	17.60	14.00	86.00
7	80	60	75.20	24.80	14.40	85.60
9	20	10	70.40	29.60	22.40	77.60
9	30	20	71.20	28.80	19.20	80.80
9	30	30	70.80	29.20	17.20	82.80
9	40	30	72.40	27.60	17.60	82.40
9	50	40	66.40	33.60	15.60	84.40
9	60	50	77.20	22.80	14.80	85.20
9	70	60	54.40	45.60	13.60	86.40
9	70	70	78.40	21.60	14.00	86.00
9	60	60	72.40	27.60	14.40	85.60
9	80	70	77.60	22.40	13.60	86.40
9	80	60	74.80	25.20	13.60	86.40

Based on Table 2 it is found that the most optimal color data training feature is on RGB color feature (Red, Green, Blue) with the dimension of row and column 70x60 at interval 5 with 80.00% accuracy. As for the features of grayscale color has the most accurate accuracy with 20x10 changes in image dimensionswith an accuracy of only 30.40% at interval 3.

Table 3 Training Results with changes in image dimensions

Rows	Column	RGB		Grayscale	
		Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
1	2	81.20	18.80	99.60	0.40
2	2	84.80	15.20	99.20	0.80
2	3	79.60	20.40	98.40	1.60
3	3	82.40	17.60	98.40	1.60
4	3	80.80	19.20	98.00	2.00
5	4	75.20	24.80	92.80	7.20
6	5	70.80	29.20	88.00	12.00
7	6	50.00	50.00	50.00	50.00
8	7	50.00	50.00	50.00	50.00
9	8	50.00	50.00	50.00	50.00
10	9	50.00	50.00	50.00	50.00

Based on Table 3 it is found that the most optimal color data training feature is on the Grayscale color feature with the dimension of row and column 1x2 with 99.60% accuracy. While for RGB color feature has the most optimal accuracy that is with 2x2 dimension changes with an accuracy of 84.80%.

D.Result of Testing Process

The results of the testing process is done with data that has been prepared to be tested with the amount of test data of 200 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 the identification of testing based on trial variation are shown in Table 4 below

Tabel 4 Result of Testing Process

No.	Type of trial	Accuracy (%)	*Error* (%)
1	Color features with Frequency Distribution	78.00	22.00
2	Color features with Frequency Distribution and Image Dimension Changes	82.00	18.00
3	Image dimensions change	94.00	6.00

Testing results with 3 ways that is , the first trial is a test with a frequency tabulation without changing the size of the image dimension, the second test is a test with the frequency tabulation and change the image dimension, and the third test is a test by changing the dimensions of the image Without frequency tabulation.

The most accurate accuracy uses how to change the image dimension on the Grayscale color feature of 94.00%.

V. CONCLUSION

In establishing Eggs Fertilities Detection System on the Image of Kampung Chicken Egg Using Naive Bayes Classifier Algorithm first done 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 the data testing process also has the same process, after the tabulation of the frequency distribution is done classification process based on the calculation model with naive bayes classifier method

Based on the results of the test data on chicken eggs image training data, the data used are as many as 250 data where all the data is divided into 2 ie 125 data eggs fertil and 125 infertile egg data obtained the best accuracy is 80% with RGB color processing process (Red, Green Blue) with interval value 5. In the training trials in RGB (Red, Green, Blue) with dimension size 70x60 at interval 5 has an accuracy of 86.4%. While experiments with dimensional changes in Grayscale with 1x2 image dimension value has an accuracy of 99.6%. The results of testing data of chicken eggs with 150 eggs and 50 infertile eggs data obtained the best accuracy is 94% with grayscale color processing

process with dimension size 1x2. Some of the data processing in image acquisition process resulted in less accurate accuracy, due to several factors such as lack of intensity of candling light and outer shell cleanliness that will affect segmentation process and feature extraction.

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

[9] Purwanto S. D, Santoso I. B. "Sistem Identifikasi Boraks pada Bakso Daging Sapi Berbasis Android Menggunakan Algoritma *Naive Bayes Classifier*". MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi. Volume 9, No. 1(2017), pp 33-37.

[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] Santoso, Irwan Budi. "Statistika I untuk Teknik Informatika". UIN PRESS. Malang. 2013.