



GA-SVC BASED SEARCH APPLIED FOR OPTIMIZATION OF IMAGE FEATURES SUBSET IN QUALITY ESTIMATION SYSTEM OF BULK GREEN COFFEE BEAN

Radi^{1,2}, Muhammad Rivai¹ and Mauridhi Hery Purnomo¹

¹Department of Electrical Engineering, Faculty of Industrial Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Department of Agricultural Engineering, Faculty of Agricultural Technology, Gadjah Mada University, Yogyakarta, Indonesia

E-Mail: radi10@mhs.ee.its.ac.id

ABSTRACT

This research aims to develop a quality estimation system of bulk coffee grain based on machine vision technique that was mainly focused on finding the best subset of image feature combination. The subset was defined as the minimum number of features for achieving the reasonable level of identification or interpretation. For this purpose, a heuristic searching method based on genetic algorithm (GA) was applied to find the best feature subset from 26 image features extracted from gray channel (9-color features and 17-co-occurrence-based-textural features). The GA with binary code chromosome was designed with a support vector classifier (SVC)-based fitness function which also played as pattern recognition software for such developed-machine vision system. The experiment was started with data collection of image samples captured by a constant illumination of 200 lux of an imaging system. Besides varied the sample (7-grades for Arabica and 8-grades for Robusta), the study also evaluated some preconditioning treatments for the initial image. With a constant population of 80 chromosomes, the selection step was performed until the 20th generation with standard genetic operations (selection, crossover, mutation, elitism), the algorithm was able to obtain an optimal feature subset consisting in average number of 5-7 features for all tested data sets. Evaluation on the analysis result shows that the best identification level was achieved from directly image processing (without preconditioning). By the pre-processing step, a quality estimation system based on selected feature subset was potentially able to estimate the quality of green coffee beans in bulk with accuracy of 86% for Arabica and 87% for Robusta coffee.

Keywords: green coffee bean, quality estimation, image feature optimization, genetic algorithm, support vector classifier.

1. INTRODUCTION

Secondary coffee processing is the downstream coffee industry which aims to produce ground coffee and other derivate products from the primary coffee grain (green coffee bean). The process includes a series of steps beginning of raw material preparation, continued with the process of roasting, blending, grinding, and packaging as well as product development of coffee derivatives [1]. In the early stages, besides for management control, quality information of the raw materials is necessary for determining an appropriate treatment to be applied. Since the treatment is needed to obtain an optimum flavor of the final product, quality monitoring system of the raw materials becomes main role on the processing of coffee. Although quality classification of coffee grain in detail based on ISO standard can only be done through to the defect test, as well as it requires expert and takes a long inspection time, this technique is less appropriate to be applied in the coffee processing that runs quickly and continuously. With the standard, the raw material quality is determined based on the percentage of grain defects contained in the samples. Because the defective grains (grains burst, black, dirty) appear on the surface when such grains are exposed in bulk, assuming that the whole are homogeneous, quality of the bulk coffee grains is possible to be estimated from visual analysis of the surface appearance. In this issue, the development of machine vision may become an appropriate solution for predicting the quality of the coffee grains quickly so as to follow the processing line in the secondary coffee industry.

As a system designed for providing information or such a specific decision from an image analysis particularly focused on tasks in industry, a machine vision system includes a set of imaging technology, image acquisition, pre-processing, and software for interpretation or classification [2-3]. In other word, it consists of a set of hardware and software. Because the decision-making is inferred solely from information obtained from the image as representation of the analyzed object, the software package plays as a central role for developing a reliable and accurate-machine vision system. As usual, machine vision software includes algorithms of image processing and machine learning for identification or interpretation. Corresponding to the estimation system of green coffee bean, in which decision is interpreted from information obtained from the image surface, thus such image analysis algorithm particularly for feature extraction may become the main task.

Since the feature extraction aims to obtain information from an image sample, by which the final decision is inferred, the features must be relevant to the classification purpose. Considering on the human visual sense, [4] explained that human uses the information of spectral, textural, and contextual for their perception and interpretation for such image sample. Regarding on classification of bulk coffee grains based on the surface information, the spectral and textural features can be used as an analytical study. Although similar classification studies have been carried out by [5], the accuracy for some grades of sample is still low. This is expected because of



the limited number of features used in this study which only focused on combination of three-color features (mean of R, G, B) with four-features of gray level co-occurrence matrix (GLCM), i.e. the feature of energy, entropy, contrast, homogeneity. Several similar researches show that such machine vision applying more complete features provides a better level of identification. Those are the development of a machine vision system for corn variety identification [6], discrimination of wheat grain varieties [7], identification of nine Iranian wheat seed varieties [8], wheat class identification [9]. These studies generally use a fuller number of color and textural features. Therefore, we expect that a reliable quality estimation system for coffee grain can be designed by applying combination of more complete features. Because of the usefulness of color and textural features for such image retrieval, especially for the purpose of classification or identification of agricultural products [10], then the feature combination is deemed sufficient for the quality estimation system of the coffee grains.

Generally, the image textural features include four categories, namely statistical feature, model-based texture, structural texture, and transform-based texture; among them the statistical feature plays as the most widely used for classification or identification of such agricultural products [10]. Some methods have been developed for obtaining the statistical features, i.e. based on gray level co-occurrence matrix (GLCM), gray level run-length matrix (GLRM), local binary pattern (LBP), a local similarity number (LSN), and local similarity pattern (LSP). Among them, the GLCM has advantages for representing the surface profile of an image. Since the statistical feature can not be separated from the information of spectral (color), the study will be directed to evaluate the combination of GLCM and color features as descriptor features simultaneously. Because a number of GLCM features can be extracted [4,11], combined with spectral features, as well as the possibility of the use of true color coordinates (RGB) as the basis for analysis, the size of the input feature becomes large enough. Beside effects to the level of identification, a large number of feature is also contributes significantly to the computational load which may decrease performance of the system. Therefore, optimization of the combination of image features is needed in order to maximize the performance of the quality estimation system of the bulk coffee grains. This study aimed to adopt a GA-based heuristic search method for obtaining the best subset of the image features.

2. MATERIAL AND METHOD

Optimization of the feature subset for developing the quality estimation system of bulk coffee grains was carried out as follow. The study was started with designing an imaging system, at the same time, an experimental sample was prepared. The research was then continued with image data collection by capturing each sample grade (7-grades of Arabica and 8-grades for Robusta). The image data obtained from the same coffee variety was

collected as a main data set. They were two main data sets being observed, i.e. data set of Arabica and Robusta. Five pairs of training and testing data set were randomly generates from each main data set. Thus, the total of ten-pairs of training-testing data sets, each consists of a half number of the main data set, were investigated for this experiment. The training data set was needed to optimize combination of the features, while the testing data set was used to evaluate performance of the selected subset as input features of the quality estimation system. Since the testing data set is different to the testing data set, the obtained accuracy can demonstrate the estimation system performance in practice application. Schematically, procedure of this study is presented in Figure-1.

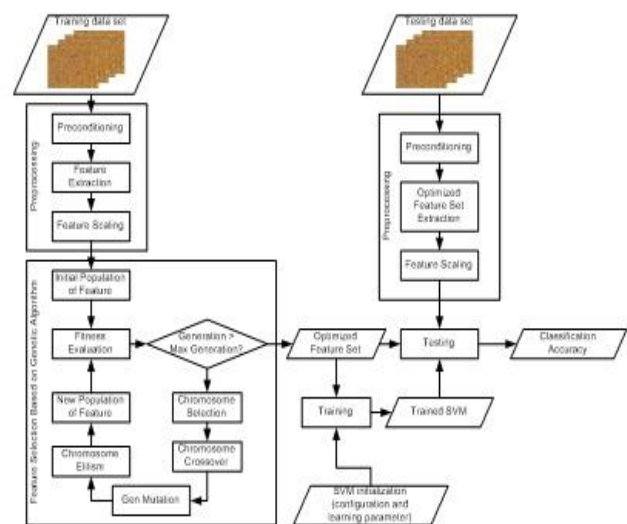


Figure-1. Experimental procedure

a) Imaging system and data collection

This experiment focused on two varieties of coffee prepared with SNI standard of green coffee bean. They are a set of Arabica coffee consisted of 7-categories (Grade I, Grade II, Grade III, Grade IV, Grade V, Grade VI, and Off Grade), and a set of Robusta coffee combined from 8-categories (Grade I, Grade II, Grade III, Grade IVA, Grade IVB, Grade V, Grade VI, and Off Grade). For image acquisition, we prepared a closed imaging chamber with size of 60x60x60 cm designed under a constant illumination of 200 lux. A camera webcam (Logitech HD C920) mounted on the centre-top of the chamber with a round of 20 cm distance to the sample. For each capturing, each sample was prepared on the area of 40x30 cm on a sample board. Each sample was captured thirty times with randomization prior to measurement. Thus, the data set of Robusta consists of 240 images, while the Arabica consists of 210 images (bitmap with size of 640x480 pixels). For this experiment, five data sets were prepared randomly from each sample variety, thus we had 10-pairs of training and testing data set (each grade consists of 15 images).

b) Preprocessing

Prior to be analysed, each image should be prepared with pre-processing stage. This process was



initialized by determining region of interest (ROI) of the image being analysed. Since it was obtained by cropping out 10 pixels in width of the image sample outside, the ROI of each had size of 620x460 pixels. For optimizing this stage, we investigated four different methods for preconditioning to the initial image. Suppose that $I[i,j]$ is the initial image being processed; $A[i,j]$ is the preconditioned image, $i=0,1,...,M-1$; $j=0,1,...,N-1$, G is the gray level, the preconditioning operations are defined as.

a. Preconditioning 1 (P-1):

$$A[i,j] = I[i,j] \quad (1)$$

b. Preconditioning 2 (P-2):

$$A[i,j] = \frac{I[i,j] \times (G-1)}{a-b} \quad (2)$$

Where

c. Preconditioning 3 (P-3):

$$A[i,j] = \frac{H_K[I[i,j]] \times (G-1)}{M \times N} \quad (3)$$

$$H_K[x] = H[x-1] + H[x] \quad (4)$$

d. Preconditioning 4 (P-4):

$$A[i,j] = \frac{H_K[I'[i,j]] \times (G-1)}{M \times N} \quad (5)$$

$$I'[i,j] = I[i,j] \times F \times \text{random}[F] \quad (6)$$

$$H_K[x] = H[x-1] + H[x] \quad (7)$$

Color features are also known as first order statistical feature obtained with the following operation. Firstly, the preconditioned image $A[i,j]$ should be represented as a gray scale image, then the histogram $p[i]$ ($i=0,1,...,G-1$) is extracted and normalized with the following equation.

$$P[i] = \frac{p[i]}{\sum_{i=0}^{G-1} p[i]} = \frac{p[i]}{M \times N} \quad (8)$$

Although many features can be extracted from the normalized histogram $P[i]$, this study only focuses on 9 color features, namely feature of energy (F1), entropy (F2), mean (F3), variance (F4), skewness (F5), kurtosis (F6), smoothness (F7), standard deviation (F8), and spot (F9), defined in Table-1.

Table 1. Color features extracted from normalized histogram.

Feature	Equation
Energy (F1)	$F1 = \sum_{i=0}^{G-1} (P[i])^2$
Entropy (F2)	$F2 = \sum_{i=0}^{G-1} P[i] \log P[i]$
Mean (F3)	$F3 = \frac{\sum_{i=0}^{G-1} i p[i]}{\sum_{i=0}^{G-1} p[i]} = \sum_{i=0}^{G-1} i P[i]$
Variance (F4)	$F4 = \sum_{i=0}^{G-1} (1 - F3)^2 P[i]$
Skewness (F5)	$F5 = \sum_{i=0}^{G-1} (1 - F3)^3 P[i]$
Kurtosis (F6)	$F6 = \sum_{i=0}^{G-1} (1 - F3)^4 P[i]$
Smoothness (F7)	$F7 = 1 - \frac{1}{1+F4}$
Standard deviation (F8)	$F8 = \sqrt{\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (A[i,j] - F3)^2}{M \times N - 1}}$
Spot (F9)	$F9 = \frac{\sum_{i=0}^{G-1} P[i] + \sum_{i=0}^{G-1} P[i] \times 7h}{\sum_{i=0}^{G-1} P[i]}$

These color features will be combined to the second order features extracted from GLCM analysis. The GLCM analysis is started by representing the image A into a co-occurrence matrix $CM[i,j]$ which defined as a matrix whose elements contains of a number of relative couples of color intensity in the analyzed image with certain distance of d and direction of θ . Since an image with G -gray level obtains a $G-1 \times G-1$ size of CM , it is considered too large for analysis, thus, for simplification, the original G -gray level of the initial image needs to be converted into a new image $A'[i,j]$ with a lower H -gray level defined as.

$$A'[i,j] = \frac{A[i,j]}{G/H} \quad (9)$$

Generally, the use of 16-gray level [0..15] is sufficient for this analysis. Then the new co-occurrence matrix obtained from $A'[i,j]$ is formulated as follows.

$$CM[k,l|d,\theta] = Z_{[k,l|d,\theta]} \quad (10)$$

Where Z is the pair number of $k-l$ color-level at the distance of d and direction of θ . k and l is the new gray level of the scaled image A' , so that the value of k,l are $0,1,2,...,H-1$. The CM matrix is then normalized as following equation.

$$Q[k,l] = \frac{CM[k,l]}{\sum_{k=0}^{H-1} \sum_{l=0}^{H-1} CM[k,l]} \quad (11)$$

The textural features are then extracted from the normalized co-occurrence matrix of $Q[k,l]$. This study focused on 17 textural features as summarized from [4] and [11]. Those features include energy (angular second moment, F10), entropy (F11), contrast (F12), inverse difference moment (homogeneity, F13), autocorrelation (F14), dissimilarity (F15), sum of average (F16), sum of entropy (F17), sum of variance (F18), difference of entropy (F19), information measures of correlation (F20, F21), difference variance (F22), correlation (F23), sum of square (F24), cluster shade (F25) and cluster prominence (F26). Those can be obtained as equation listed in Table-2.

**Table-2.** Textural features extracted from GLCM analysis.

Feature	Equation
Energy (F10)	$F10 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} Q[k, l]^2$
Entropy (F11)	$F11 = - \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} Q[k, l] \log(Q[k, l])$
Contrast (F12)	$F12 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} (k-l)^2 \times Q[k, l]$
Inverse diff. moment (F13)	$F13 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} \frac{Q[k, l]}{1+(k-l)^2}$
Autocorrelation (F14)	$F14 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} k \times l \times Q[k, l]$
Dissimilarity (F15)	$F15 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} k-l \times Q[k, l]$
Sum of average (F16)	$F16 = \sum_{i=0}^{2H-2} i \times Q_{x+y}[i]$
Sum of entropy (F17)	$F17 = - \sum_{i=0}^{2H-2} Q_{x+y}[i] \times \log(Q_{x+y}[i])$
Sum of variance (F18)	$F18 = \sum_{i=0}^{2H-2} (1-F17)^2 \times Q_{x+y}[i]$
Diff. of entropy (F19)	$F19 = - \sum_{i=0}^{2H-2} Q_{x-y}[i] \times \log(Q_{x-y}[i])$
Information measures of correlation (F20, F21)	$F20 = \frac{F11 - C1}{\max(B_x, B_y)}$ $F21 = \sqrt{1 - e^{-2(C2 - F11)}}$
Difference variance (F22)	$F22 = \sqrt{\frac{\sum_{i=0}^{H-1} (Q_{x-y}[i] - Q_{x+y}[i])^2}{H-1}}$
Correlation (F23)	$F23 = \frac{\sum_{k=0}^{H-1} \sum_{l=0}^{H-1} k \times l \times Q[k, l] - \mu_x \times \mu_y}{\sigma_x \times \sigma_y}$
Sum of square (F24)	$F24 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} (k - \mu_x) \times (l - \mu_y) \times Q[k, l]$
Cluster shade (F25)	$F25 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} (k + l - \mu_x - \mu_y)^2 \times Q[k, l]$
Cluster prominence (F26)	$F26 = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} (k + l - \mu_x - \mu_y)^4 \times Q[k, l]$
Where:	
$p_{x[k]} = \sum_{l=0}^{H-1} Q[k, l]$	$Q_{x+y}[i] = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} Q[k, l], \quad i = k + l$
$p_{y[l]} = \sum_{k=0}^{H-1} Q[k, l]$	$Q_{x-y}[i] = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} Q[k, l], \quad i = k - l $
$\mu_x = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} k \times Q[k, l]$	$C1 = - \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} Q[k, l] \times \log(p_{x[k]} p_{y[l]})$
$\mu_y = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} l \times Q[k, l]$	$C2 = - \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} p_{x[k]} p_{y[l]} \times \log(p_{x[k]} p_{y[l]})$
$\sigma_x = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} (k - \mu_x)^2 \times Q[k, l]$	
$\sigma_y = \sum_{k=0}^{H-1} \sum_{l=0}^{H-1} (l - \mu_y)^2 \times Q[k, l]$	
$B_x = \text{entropy of } p_x$	
$B_y = \text{entropy of } p_y$	

c) Support vector classifier (SVC)

SVC is known as powerful data classification algorithm working to optimize the best partition vector (hyperplane) called as a support vector (SV). The description of the SVC algorithm for such classification task can be found in [12-14]. In simple explanation, the algorithm can be rewritten as follows. Suppose P is a set of two-class data consisting of k-pairs, written as follows.

$$P = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^k = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_k, \mathbf{y}_k)\} \quad (12)$$

Where \mathbf{x}_i is n -dimensional input vector and \mathbf{y}_i is the target of two-classes data [-1,1]. For linearly separable data, classification of the data can be performed by obtaining a classifier vector $f(\mathbf{w}, b) = \mathbf{w} \cdot \mathbf{x}_i + b$, by which the data can be separated as

$$\mathbf{y}_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \text{ for } i = 1, \dots, k \quad (13)$$

For simplify analysis, assuming that the data lies in the function, then two boundary vector are obtained, namely as function $\mathbf{w} \cdot \mathbf{x}_i = 1 - b$ for $\mathbf{y}_i = 1$ and function of $\mathbf{w} \cdot \mathbf{x}_i = -1 - b$ for $\mathbf{y}_i = -1$. In this case, a diverse function of hyperplane $f(\mathbf{w}, b)$ can be obtained from a range of the both boundary vectors. Supposing the best hyperplane is defined as a function located on the centre of both functions, determining the origin can be obtained by

computing the margin (d_M) from both functions, which is defined as

$$d_M = \left| \frac{(1-b)}{\|\mathbf{w}\|} - \frac{(-1-b)}{\|\mathbf{w}\|} \right| = \frac{2}{\|\mathbf{w}\|} \quad (14)$$

Therefore, the distance of the boundary to the origin can be calculated as $d_o = \frac{1}{\|\mathbf{w}\|}$. Because the hyperplane is on the origin, the vector can be computed by maximizing the distance d_o . Algorithmically, the best classifier can be determined from any function which is then optimized by maximizing the distance d_o . Mathematically, it can be wrote as the following optimization functions

Minimize

$$J_S(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (15)$$

$$\text{S.t. } \mathbf{y}_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \text{ for } i = 1, \dots, k \quad (16)$$

In fact, many classification tasks are conducted for non-linearly separable data. In this case, the margin will never be found. Thus, the grouping task practically requires a tolerance for accommodating some infeasible data. This strategy can be done by modifying the function of boundary into softer equation which can be conducted by adding a variable of tolerance, known as non-negative slack variable $\{\xi_i\}_{i=1}^k$. Therefore, the new boundary function can be written as

$$\mathbf{y}_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \text{ for } i = 1, \dots, k \quad (17)$$

Since the amount of data being accommodated on the slack variable needs to be determined, it requires a non-negative constant of C as a multiplier. Thus, the optimization problem becomes the following function.

Minimize

$$\text{Minimize } J_{NS}(\mathbf{w}, \xi_i) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (18)$$

$$\text{S.t. } \mathbf{y}_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i; \xi_i \geq 0 \text{ for } i = 1, \dots, k \quad (19)$$

Optimization problem with objectives can be formulated into a non-constrained quadratic programming function by the non-negative Lagrange multipliers α_i, μ_i which can be written as follows.

Minimize

$$L(\mathbf{w}, b, \xi_i, \alpha, \mu) := \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^k \xi_i - \sum_{i=1}^k \alpha_i \{\mathbf{y}_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i\} - \mu_i \sum_{i=1}^k \xi_i \quad (20)$$

Solving of the optimization problem can be done by minimizing the function to the variable of \mathbf{w}, b, ξ_i and simultaneously it should be maximized to the variable of α_i, μ_i . At the saddle points where the partial derivatives of the function to the variable \mathbf{w}, b, ξ_i are equal to zero, this optimization problem can be easily solved into dual problem. Thus it can be written as follows.

Maximize



$$L(\alpha) = \sum_{i=1}^k \alpha_i - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j y_i y_j x_i x_j \quad (21)$$

$$\text{s.t. } \sum_{i=1}^N \alpha_i y_i = 0 \text{ dan } 0 \leq \alpha_i \leq C \text{ for } i = 1, \dots, k \quad (22)$$

By providing optimum limits of the Lagrange multipliers, we can calculate the optimum weighting vector w with $w = \sum_{i=1}^k \alpha_i y_i x_i$, then

$$f(x) = \text{sign}(w \cdot x + b) \quad (23)$$

Since the number of non-zero points lie on the SV is smaller than the amount of the training data, the value of b can be determined from any support vector data as follows

$$b = y_{SV} - \sum_{i=1}^k \alpha_i y_i x_i x_{SV} \quad (24)$$

In addition, the SVC algorithm offers a possibility for transforming the input data space into a feature space with higher dimension. Suppose that a linear classifier vector is not found in the input space, the SVC offers a way to extend the analysis of non-linear input vectors into a higher dimensional feature space. Since a linear classifier vector is obtained in the feature space, it can be used to separate the non-linear separable data in the initial input space. This solution can be done by modifying the multiplication of dot product of the input vectors with a specific function known as kernel trick illustrated as.

$$\Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j) \quad (25)$$

This kernel is then used to modify the equation-21.

Some functions of K have been applied by researchers, namely function of linear, polynomial, radial basis, and exponent. Among them, the radial basis function (RBF) has advantages when applied for such non-linearly separable data. Thus, this study uses the RBF defined as follows

$$K(x_i, x_j) = e^{r(x_i - x_j)^2} \quad (26)$$

For such real problem, the classification issue is not only limited on the two-class data as well as on this experimental task. Since this study is conducted to classify sample in multi-classes data (7-grades of Arabica and 8-grades for Robusta), modification of two-classes classifier is necessary. There are three-methods widely applied, namely with methods of one-against-one (OAO), one-against-all (OAA), and q -class SVM. Since its possibility to be simply applied, this study uses the OAA method which can be described as follows. The OAA applies q -classifiers to separate q -different classes of data. Each classifier functions to separate one class with the other retained classes. Therefore, the i^{th} -classifier separates the i^{th} -class data with the other $(q-1)$ -classes data. Mathematically, the decision function of i^{th} -classifier is

$$f_i(x) = w_i \cdot x + b_i \quad (27)$$

Minimize:

$$J(w, \xi^i) = \frac{1}{2} \|w_i\|^2 + C \sum_{q=1}^k \xi_q^i \quad (28)$$

$$\text{s.t. } y_n^i (w_i \cdot x + b_i) \geq 1 - \xi_n^i; \xi_n^i \geq 0 \text{ for } n = 1, 2, \dots, k \text{ and } i = 1, 2, \dots, q \quad (29)$$

Where

$$y_n^i = \begin{cases} +1 & \text{if } y_n = i^{\text{th}} - \text{class} \\ -1 & \text{for otherwise} \end{cases} \quad (30)$$

Finally the class of unknown sample x is recognized from the highest value of $f_j(x)$.

GA based feature optimization

A genetic algorithm is one of heuristic searching methods based on the mechanism of biological evolution. In this process, the probability of an organism to survive and reproduce strongly depends on the level of its fitness. Adopting to the terminology of genetic evaluation, finding a solution from a number of possible answers known as population is conducted through to series of genetic operations, namely selection, crossover, mutation, and elitism. In early stages, population containing a number of initial solutions is determined randomly, while in the next generation, members of the population are obtained from the evolution procedure. By the evolution routing, the population at a certain generation will be dominated by fit-individuals potentially applied as solution of the problem being solved.

Implementation of GA for optimization the feature subset potentially applied on the quality estimation system of the coffee grains is conducted as follows. Initially, combination of features is represented as chromosome. The GA operation is used to optimize the chromosomes, by which, optimal identification accuracy can be obtained. Design of the GA for this optimization can be described as follows. For example, there are n_f -features potentially to make up a solution expressed as $F = [F1, F2, F3, \dots, F_{n_f}]$.

a. Chromosome design and initial population

Chromosome is group of genes representing subset of features, so that, each chromosome consists of n_f -genes. Supposing that each feature has an equal chance to be selected as member of the subset, the chromosome is represented as binary code ('1' is for selected feature and '0' for others). For this study, the population is fixed at 80 chromosomes which is initialized by randomly function.

b. Evaluation function

Evaluation function is an instrument for measuring the fitness of individual (chromosome). It is usually derived from the expected-goal function. In this case, because the purpose is to determine the optimal feature subset for classifying the grade of coffee samples, the optimum subset should contain of minimal number of features by which the estimation system can accurately classify the samples. Since the quality estimation system is carried out by a SVC method, the evaluation function f_{Fit} is also derived from the classifier combined with the number of selected feature described as follows.



$$f_{Fit} = Acc + 5 \frac{n_f - s}{n_f} \quad (31)$$

Where the Acc is the accuracy of the SVC for classifying the training data set, n_f is the number of features, and s is the number of selected features.

c. Searching algorithm

Searching optimum feature subset with GA is performed by a series of process including selection, crossover, mutation, and elitism. In this study, selecting feasible individuals is conducted by a roulette wheel selection procedure. After that, the genetic process is continued by selective crossover, which begins with selection of the parent-candidates. Selected couple-parent is crossed each other to generate a new chromosome (offspring) containing a gene-combination from both parents. The new chromosome is then used to substitute the corresponding parental chromosome in the new population. After crossover, the process is continued with mutation applied by changing the randomly selected genes. Then, the next step is known as elitism which is conducted by duplicating some high-fitness chromosomes. Finally, the new population obtained after the four-genetic operators is used to replace the previous population. The new population is called as next generation of the evolution. The genetic operation is repeated to the new population until such termination operator stops the process. After the algorithm is terminated, the fitness of the last population is then evaluated. The highest fitness of chromosome is selected as the problem solution by decoding it into a feature subset.

The subset containing a number of selected features is then used to simulate such quality estimation system of the coffee green beans. A feature subset extracted from the training data set is then trained to the corresponding SVC. A similar extraction method is also applied to the testing data set. The testing feature set is then projected into the trained SVC. The classification accuracy is obtained from the true-classified testing data.

3. RESULT AND DISCUSSION

a) Selection process

This research is focused on optimizing the combination of 26 features consisting of color and GLCM features extracted from gray channel as feature descriptors of bulk grain image of coffee bean. This optimization was conducted by searching the best feature subset, by which, such designed quality estimation system can achieve high classification performance of the green coffee bean. Based on the procedure, searching of the subset working at a feature space by applying a machine learning system can be categorized as wrapped technique [15]. As a machine learning algorithm, the GA requires some input parameters that influence its convergence. For this reason, in order to evaluate all treatments regarding to the appropriate feature extraction method, the GA was designed with constant input parameters, i.e. population size of 80 chromosomes, length of chromosome of the number of features (26),

selection, crossover, mutation, and elitism probability are 80%, 25%, 1%, and 2% sequentially. While the fitness function is formulated from combination of SVC accuracy and the number of non-selected features. Since obtaining a hyperplane as classifier vector using SVC may be completed with some methods, we consider to apply a sequential optimization proposed by [15]. For all data sets, the SVC was run under constant parameters including the RBF for kernel operation, 1 for lambda, 0.05 for gamma, 100 for C, and 1000 epochs for training iteration. While searching, the number of feature regarding to the formed subset is listed. The change in feature composition considering to the feature number at specified generation of the GA is listed on Table-3, while the average amount of feature subset at the entire generation is presented in Figure-2.

Table-3. Feature number of selected subset at different generation.

Sample	Pre-conditioning	Gen-1 Avr (min,max)	Gen-5 Avr (min,max)	Gen-10 Avr (min,max)	Gen-15 Avr (min,max)	Gen-20 Avr (min,max)
Arabica (Set 1-5)	P-1	15.0 (14,16)	10.8 (7,13)	8.2 (6,9)	6.0 (4,8)	5.2 (4,7)
	P-2	15.2 (13,16)	10.4 (7,13)	8.2 (6,10)	6.8 (6,8)	6.6 (5,8)
	P-3	15.4 (14,16)	10.0 (7,12)	8.2 (6,9)	6.0 (5,7)	5.6 (4,7)
	P-4	15.4 (14,16)	9.4 (7,12)	7.6 (6,9)	6.0 (5,7)	5.4 (4,7)
Robusta (Set 1-5)	P-1	15.4 (15,16)	10.0 (8,12)	8.2 (7,10)	6.0 (5,7)	5.6 (4,7)
	P-2	14.4 (13,16)	11.4 (10,13)	8.8 (8,10)	7.2 (6,98)	6.6 (5,8)
	P-3	15.6 (15,16)	10.2 (8,12)	8.6 (6,11)	7.8 (5,10)	7.0 (5,9)
	P-4	15.4 (15,16)	10.0 (8,12)	8.2 (7,10)	6.0 (5,7)	5.8 (5,7)

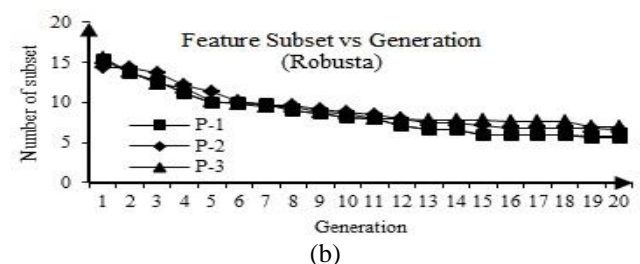
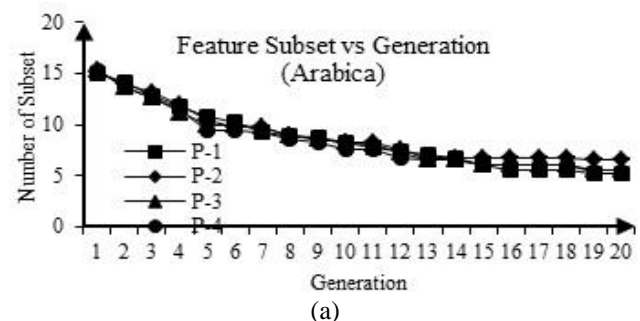


Figure-2. Feature number of the selected subset in all generation during iteration: (a) Arabica (b) Robusta.



The result shows that the algorithm can optimize the subset demonstrated from its ability to minimize the number of selected features. The result also appears that no applied treatment effects on the ability of the algorithm to find the best combination subset. The algorithm ability on optimizing the feature subset is also showed from the accuracy of the fitness function. At the last iteration of searching (20th generation), with the composition of the selected subset, the fitness function can classify the training data sets with accuracy of upper than 80% for all treatments and data sets. As final solution of this searching, the highest value of the fitness obtained at the last generation is selected. The accuracy of the system extracted from the selected fitness value for the entire data set is listed in Table 4, whereas Table-5 shows the number of feature subset obtained from the same generation. Based on distribution of the accuracy level, it is known that the preconditioning does not affect to the ability of the selected subset for classifying grades of coffee sample.

Table-4. The classification accuracy extracted from the fitness value at 20th generation.

Sample	Pre-conditioning	Training data set					Avr
		Set 1	Set 2	Set 3	Set 4	Set 5	
Arabica	P-1	83.04	82.23	82.23	83.85	84.65	83.20
	P-2	84.65	83.04	84.65	83.85	85.46	84.33
	P-3	83.04	84.65	83.85	82.23	83.85	83.52
	P-4	83.04	83.04	84.65	83.85	82.23	83.36
Robusta	P-1	83.04	84.65	83.85	82.23	83.85	83.52
	P-2	83.04	85.46	85.46	83.85	83.85	84.33
	P-3	83.04	83.85	86.27	87.08	83.04	84.65
	P-4	83.04	84.65	83.85	83.04	83.85	83.68

Table-5. Feature number of the selected subset at the 20th generation.

Sample	Pre-conditioning	Training data set					Avr.
		Set 1	Set 2	Set 3	Set 4	Set 5	
Arabica	P-1	5	4	4	6	7	5.2
	P-2	7	5	7	6	8	6.6
	P-3	5	7	6	4	6	5.6
	P-4	5	5	7	6	4	5.4
Robusta	P-1	5	7	6	4	6	5.6
	P-2	5	8	8	6	6	6.6
	P-3	5	6	9	10	5	7.0
	P-4	5	7	6	5	6	5.8

Estimation accuracy

After the searching process is completed (20th generation), the obtained subset is then tested as an input feature vector for developing the quality estimation system. The SVC as a pattern recognition algorithm is then trained to classify the training data set based on the selected subset. To demonstrate the subset ability as

feature descriptor for the quality estimation system, the trained SVC is then tested to recognize the corresponding testing data set. The recognizing procedure is started by extracting the feature subset from the testing data set using the appropriate preconditioning. The subset is then projected on the trained SVC. The identification accuracy from the testing illustrates the performance of such selected subset for estimation of bulk coffee grain when it is applied as feature descriptor on such quality estimation system. The estimation result for recognizing the testing data set is listed in Table-6. Based on the data, it appears that preconditioning treatments influence on the accuracy of the system. Generally, it can be concluded that the best accuracy can be obtained when the feature extraction is conducted directly from the initial image. In other word, there is no preconditioning treatment needed for the feature extraction. It may be caused due to several reasons. Supposing that the fixed illumination installed in the imaging system is able to yield homogenous image samples, by which, a reliable feature is generated, thus the initial process is not required. In addition, since that the statistical features is computed from the color differenced with specified distance and angle, the applied preconditioning may obtains several vague surface properties that potentially effect on the extracted features.

Computation time

Although the searching method is easily applied, selecting a subset from highest feature space potentially causes some problems. This is especially caused by the load of computational searching. The time cost for this selection per each data set is listed on Table-7. Table-7 illustrated how much time is required for finding subset from only 26 features. It becomes one drawback of this searching method.

Table-6. Identification accuracy of the quality estimation system demonstrated with obtained feature subset.

Sample	Pre-conditioning	Testing data set					Avr
		Set 1	Set 2	Set 3	Set 4	Set 5	
Arabica	P-1	86.67	90.48	88.57	90.48	77.14	86.67
	P-2	60.00	61.90	52.38	78.10	69.52	64.38
	P-3	75.24	57.14	55.24	40.95	54.29	56.57
	P-4	70.48	67.62	60.00	65.71	75.24	67.81
Robusta	P-1	85.83	85.00	90.00	88.33	86.67	87.17
	P-2	58.33	75.83	55.00	85.00	57.50	66.33
	P-3	72.50	50.83	41.67	39.17	60.00	52.83
	P-4	64.17	63.33	60.83	68.33	62.50	63.83

**Table-7.** The average of searching time (s).

Sample	Preconditioning	Training data set					Avr
		Set 1	Set 2	Set 3	Set 4	Set 5	
Arabica	P-1	639.35	662.00	648.37	648.37	650.10	653.64
	P-2	649.28	651.53	647.76	659.91	644.74	650.64
	P-3	644.31	645.44	648.04	647.29	650.41	647.10
	P-4	644.33	644.98	676.64	656.54	660.13	656.52
Robusta	P-1	954.52	968.98	950.40	961.77	955.49	958.23
	P-2	951.79	970.73	969.98	958.74	961.84	962.62
	P-3	964.88	954.41	969.26	965.81	946.38	960.15
	P-4	965.39	966.37	960.90	969.18	966.80	965.73

4. CONCLUSION AND FUTURE WORK

This study has applied the GA-based heuristic search method to optimize the subset of features on the machine vision-based-quality estimation system of bulk coffee grains. This wrapped searching method is able to find the appropriate feature subset. The result shows that in average of 5-7 feature combination is selected from 26 input spaces for all data sets. With the subset, a fairly high fitness value is obtained. The evaluation shows that the application of those obtained subsets for classifying the testing data sets reaches highest performance for directly preprocessing (without anything preconditioning treatments). With the process, a machine vision system for quality estimation of coffee grains in bulk is able to recognize the grades with accuracy in average of 86-87% for Arabica and Robusta coffee.

Although this method is generally easy to be applied, it needs a high computational cost. Although the number of the population can be reduced in further experiments, it certainly will affect to the ability of the algorithm in achieving the expected solution, especially when it is applied for selecting a subset from a higher feature space. In line to the recent development of such feature extraction method, for obtaining an optimal result, feature subset selection from a large number of feature spaces is required. Therefore, our next study is focusing on searching feature subset particularly by applying a combination of wrapper and filtered methods.

ACKNOWLEDGEMENTS

Some contributors supported this research, therefore we would like to grateful to Indonesian Higher of Education (DIKTI), LPDP-Kemenkeu, Gadjah Mada University, and PT. Aneka Coffee Industry (PT. ACI) Indonesia. We are also thank for any contribution from Prof. T. Nakamoto (Tokyo Institute of Technology, Japan) and his Laboratory's member, Ms. Lily Sujiati (QC and QA Manager of PT. ACI) and hers Laboratory's member, and member's of Industrial Electronic Laboratory, ITS.

REFERENCES

- [1] Sri M., Edy S. and Kaswanto. 2010. Products Development of Coffee. Indonesian Coffee and Cocoa Research Institute. Jl. PB. Sudirman No. 90, Jember 68118 - Indonesia:
- [2] H. Golnabi and A. Asadpour. 2007. Design and application of industrial machine vision systems. Robotics and Computer-Integrated Manufacturing. Vol. 23, No. 6, pp. 630-637.
- [3] D.J. Lee, R. Schoenberger, J. Archibald, and S. McCollum. 2008. Development of a machine vision system for automatic date grading using digital reflective near-infrared imaging. Journal of Food Engineering. Vol. 86, No. 3, pp. 388-398.
- [4] R. M. Haralick, K. Shanmugam and I. Dinstein. 1973. Textural Features for Image Classification. IEEE Trans. Syst. Man Cyber. SMC, Vol. 3, No. 6, pp. 610-621.
- [5] F. Faridah, G. O. F. Parikesit and F. Ferdiansjah. 2011. Coffee Bean Grade Determination Based on Image Parameter. TELKOMNIKA. Vol. 9, No. 3, pp. 547-554.
- [6] X. Chen, Y. Xun, W. Li, and J. Zhang. 2010. Combining discriminant analysis and neural networks for corn variety identification. Comp. and Elec. in Agriculture. Vol. 71(Supplement 1). pp. 48-S53.
- [7] P. Zapotoczny. 2011. Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture. Journal of Cereal Science. Vol. 54, No. 1, pp. 60-68.
- [8] A. Pourreza, H. Pourreza, M.-H. Abbaspour-Fard and H. Sadriani. 2012. Identification of nine Iranian wheat seed varieties by textural analysis with image processing. Comp. and Elec. in Agriculture. Vol. 83, pp. 102-108.
- [9] A. Manickavasagan, G. Sathya, D. S. Jayas, and N. D. G. White. 2008. Wheat class identification using monochrome images. Journal of Cereal Science. Vol. 47, No. 3, pp. 518-527.
- [10] C. Zheng, D.-W. Sun, and L. Zheng. 2006. Recent applications of image texture for evaluation of food qualities—a review. Trends Food Sci. Technol. Vol. 17, No. 3, pp. 113-128.
- [11] L.-K. Soh and C. Tsatsoulis. 1999. Texture analysis of SAR sea ice imagery using gray level co-occurrence



matrices. IEEE Trans. Geosci. Remote Sens. Vol. 37, No. 2, pp. 780-795.

- [12] Sethu V. and Si W. 1999. Sequential support vector classifiers and regression. In Proc. International Conference on Soft Computing (SOCO'99). pp. 610-619.
- [13] E. Phaisangittisagul, H. T. Nagle and V. Areekul. 2010. Intelligent method for sensor subset selection for machine olfaction. Sens. Actuators B Chem., Vol. 145, No. 1, pp. 507-515.
- [14] Y. Liu and Y. F. Zheng. 2006. FS_SFS: A novel feature selection method for support vector machines. Pattern Recognition. Vol. 39, No. 7, pp. 1333-1345.
- [15] Jinjie H., Yunze C. and Xiaoming X. 2006. A Wrapper for Feature Selection Based on Mutual Information. In 18th International Conference on Pattern Recognition (ICPR2006). pp. 618-621.