IOP Conf. Series: Materials Science and Engineering **434** (2018) 012065 doi:10.1088/1757-899X/434/1/012065

Identification of sugarcane maturity scale based on RGB, Gabor feature extraction and Support Vector Machine

C Rahmad[1], F Rahutomo[1], M A Gustalika[2]* and I F Rahmah[2]

[1] Information Technology Department, State Polytechnic of Malang, Malang, Indonesia
[2] Electrical Engineering Department, State Polytechnic of Malang, Malang, Indonesia

*m_azrino@polinema.ac.id

Abstract. Sugarcane is a unique grass variety that allows many cuts for more than a few years. Plant growth is very slow in the early stages that takes 25-30 days to complete germination and 90-95 more days to complete the tillers. Small scale sugarcane farmers generally do not have direct access to sugar mills but through loggers. Sugarcane farmer must increase cost for the loggers by themself. Sugarcane farmers often forgot the time at the moment when they planted sugarcane. This study proposes a new approach to establish detection of maturity of sugarcane before harvesting by detecting through stem of sugarcane. This study illustrates the maturity of sugarcane. This study used its own datasets of 300 stem images of sugarcane and to differentiate into three categories, immature, semi-mature, and mature. The study is divided into two processes, the first one is extraction using the RGB color feature and Gabor texture feature, and the second is the classification process using the Support Vector Machine, Naive Bayes and Artificial Neural Network methods. Based on the trial it is known that Support Vector Machine has the best results with precision and recall of 87,4% and 85,7% respectively.

1. Introduction
Sugarcane is a unique grass variety that allows many cuts for more than a few years and is considered semi-lasting. It grows rapidly with abundant reproduction and allows for the financial management of many parts of the plant. The expansion cycle consists of stages of sprouting, affiliation, enlargement and maturation. From planting to primary cutting, the plant is called sugarcane, which lasts between 12 to 18 months, depending on the period and area of planting. After the first cut, the cane regrets were replaced by cane rattan, with a standard 12-month cycle [1].

Sugarcane is a long-lasting plant that reaches its maturity in 11-12 months. Plant growth is very slow in the early stages that takes 25-30 days to complete germination and 90-95 more days to complete the tillers. Larger amounts of nutrients and water are applied during the early stages of growth compared to the later stages. Sugarcane set is planted continuously in line with distance of 90 cm to 150 cm. Large amounts of nutrients, moisture, solar radiation and wider distances between plants also support the growth of weeds in sugarcane fields that prevent plants from getting nutrients [2].

Brazil is the world's largest sugarcane producer, followed by India and China, and is the largest producer of sugar and ethanol. Brazil is responsible for about 20% of sugar production and 40% of the world's exported sugar. Cane is also very important in terms of the environment as ethanol is one of the best alternatives to reduce emissions of gas pollutants, which is the fundamental cause of the greenhouse effect [3].

In India, most farmers do not receive predictable crops for a variety of reasons. The cultivation of agricultural products in principle believes in climatic conditions. Rain conditions also affect rice farming. In this background, farmers inevitably require a wise guide to envisioning future efficiency gathering plus investigations to be resolved in arrangements to help farmers harvest crops well in their crops [1].

Knowledge of sugarcane varieties is a difficult thing to do. Data about this knowledge is complex data. Currently, cane varieties can be done manually by sugarcane experts. This is because the information about sugarcane is a qualitative descriptive semantic. This makes the annotation process difficult for the general public as well as computer programs [4].

In practice small scale sugarcane farmers generally do not have direct access to sugar mills but through loggers (contractors). Involvement of loggers in the sugarcane trade had an impact on the increase in production costs borne by sugarcane farmers. Sugarcane farmers often forgot the time at the moment when they planted sugarcane so they do not know the definite time in the harvest of sugarcane, the sugarcane farmers only see from the condition of sugarcane only during the harvest. With this problem, this study aims to make the application using machine learning algorithm is Artificial Neural Network, Naive Bayes and SVM that was useful to know the maturity of sugarcane through stems with certainty and also help increase income from sugar cane farmer.

In previous study, has been experiments of Precision sugarcane monitoring using SVM classifier [5], which monitors parameters (temperature and humidity) and is responsible for the healthy growth of sugarcane plants. The difference with the previous study is we make an application based on the maturity of sugarcane through stem of sugarcane and using one type of sugarcane with code PS 86-2. then there is a previous study on Feature Extraction for Identification of Sugarcane Rust Disease [6], which aims to find the right features in identifying rust diseases in sugarcane by combining color and texture extract features. Other studies have also mentioned that there is no system of monitoring for sugarcane products widely. In a near real time yield estimation for sugarcane in Brazil combining remote sensing and official statistical data [7], it was explained that they used remote sensing technology to estimate the yield of sugarcane maturity. We also created our own datasets that can be used for further studies.

2. Material and method

2.1. System Description

Following system description in this study

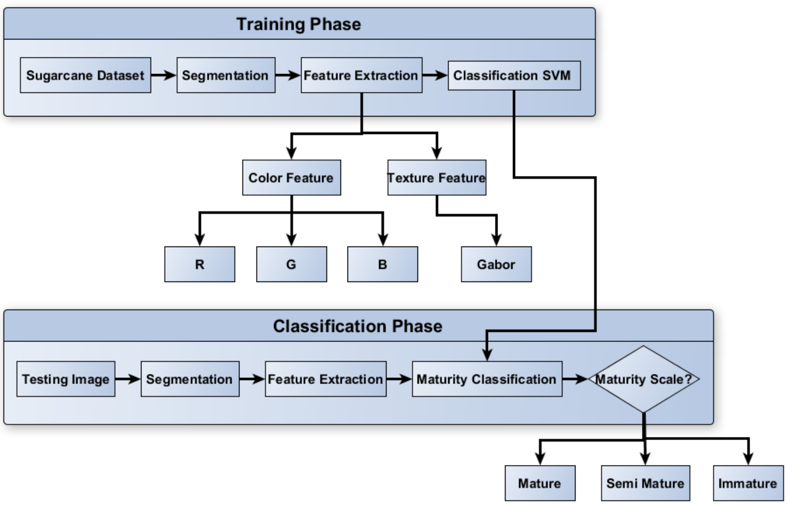

Figure 1. System description.

IOP Conf. Series: Materials Science and Engineering **434** (2018) 012065 doi:10.1088/1757-899X/434/1/012065

Figure 1 shows the description system we created. Our system description consists of two main phases: the training phase and the classification phase. In the training phase, the cane dataset in segmentation uses otsu then looks for its feature extraction, there are 2 feature extractions in use, color extraction and texture. On color extraction using RGB while on textural extraction using Gabor. Next classification of training data that has been in the extraction by using SVM. In the second phase is the classification by using the dataset of the data testing, it will compare the results of the SVM classification of training data so it will produce a percentage of maturity between the mature, half-mature and immature.

2.2. Machine based approach

Machine learning algorithms are the processes used to adapt the model to a dataset, through training or learning. The model studied is then used against independent datasets. Basic machine learning algorithms require a pair of documents - training and testing of data. The training phase includes self-training using various images and then tested using test data. There are several Machine learning algorithms, such as Artificial Neural Network (ANN), Naïve Bayes (NB) and Support Vector Machines (SVM), data testing requires the dataset to produce classification.

2.2.1 Support vector machine. SVM is a machine-based learning approach based on statistical theory, which can find the optimal solution of classification results with limited information about small sample datasets. Unlike other machine learning approaches, SVM uses the idea of kernel functions to convert nonlinear problems into linear problems, and reduce the complexity of mapping [8]. By using functional transformation and kernel function optimization, the optimal problem can be converted into a convex quadratic functional extremism problem. In theory, SVM is bound to get global optimal solutions SVM Formula is:

$$\emptyset(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \frac{\gamma}{n}\sum_{i=1} \varepsilon_i$$

$$\text{s.t. } y_i\, g(x_i) = y_i(w^T x_i + b) \geq \boldsymbol{\varepsilon_i} - 1,\ \ i = 1,2,3,\ldots..n \quad (1)$$

where ε_i are slack variables and $\frac{\gamma}{n} > 0$ acts as a term of regularization.

2.2.2 Naive Bayes. The Naïve Bayes classification is a probabilistic classification based on the Bayes Theorem. For each sample, the possibility of the class is determined. It is assumed that all attributes are independent [9]. NB is a popular machine learning tool for classification. Naive Bayes works very well on various issues. Naive Bayes formula is :

$$Z_{nb} = argmax\ P(c)\ \Pi\ P\ (a|c) \quad (2)$$

Explanation of symbols
a = word
c = category
P(a|c) = word probability in category *c*
P(c) = probability of category *c*
P(a) = probability of *a* word.

2.2.3. Artificial Neural Network. Artificial Neural Networks are trained in the image. The ANN architecture is based on advanced feed networks. This architecture is trained in accordance with Back propagation algorithm. The results obtained are further improved by using the Radial function network Base The performance of the radial base network functions outperformed the ANN [10].

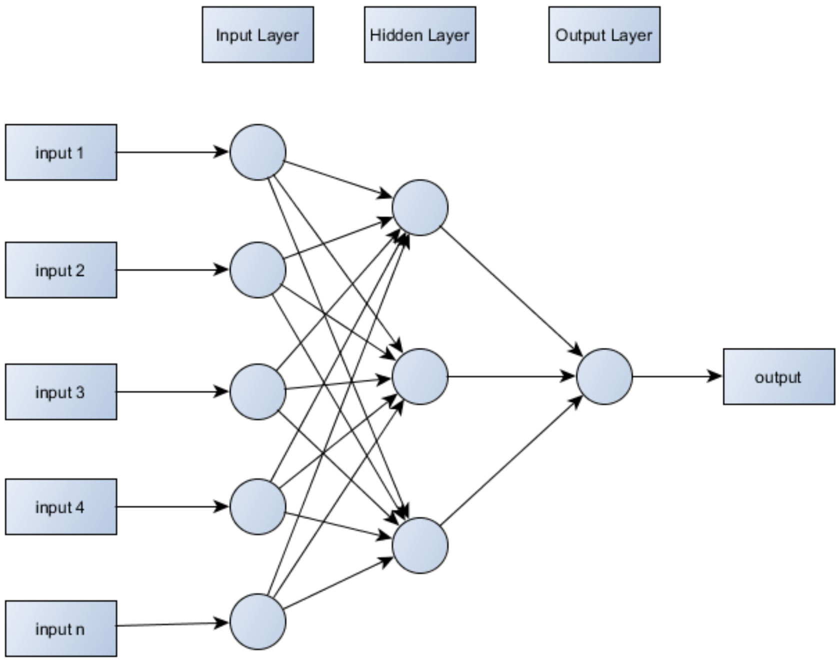

Figure 2. Artificial neural network for classification of cane.

As shown in Figure 2. It consists of a set of interconnected neurons to map the input to output features. An advanced three-nerve neural network is implemented with an input, hidden h and 1 neuron output indicating the cane as mature or not. The learning of neural networks is monitored and the weights are adjusted according to the back propagation procedure

2.3. Sugarcane sample

The sugarcane samples used in this study were (code name = 86-2) or person call it (62).A total of 300 samples, with maturity levels of immature, semi-mature, and mature, were picked in December 2017 in Wates, Kediri. For supporting sugarcane color and texture rating, then training dataset divede 70 dataset training into 3 category. The remnant is shared for data testing. Table 1 shows the details of the training data and the data testing of the sugarcane dataset.

Table 1. The details of data training and data test of Sugarcane.

Category	Data Training	Data Testing
Imaature	70	30
Semi Mature	70	30
Mature	70	30
Total	210	90

Table 1 shows the image in each category of size 156 x 208 pixels where the image has been through the process of segmentation

2.4. Feature extraction

Extraction features are generally based on two pixel criteria, a). Equations and b) Proximity of pixel values. The criteria for similarity are based on the gray range of features corresponding to the threshold to separate background image data. So the feature of extraction can be interpreted as a process to distinguish features that distinguish objects from other objects [11].

2.4.1. Color feature. In order to detect whether the cane is immature, semi-mature, and mature, the color statistical and color texture features for each Red (R), Green (G) and Blue (B) channels are extracted as given below

$$\text{Color Mean}(\mu) = \frac{1}{M}\sum_{i=1}^{M} Xi \tag{3}$$

Sugarcane maturity is strongly associated with color. Therefore, the color feature is extracted from the cane image to classify it as immature, semi-mature, or mature. The values of R, G, B of the extracted and average images are calculated based on the following equation. Thresholds are selected by experiment and R is average compared to threshold. If it is larger than the threshold, then the mature cane is considered immature [12].

$$\text{Mean R} = \frac{R}{N} \tag{4}$$

$$\text{Mean G} = \frac{G}{N} \tag{5}$$

$$\text{Mean B} = \frac{B}{N} \tag{6}$$

Where, Mean R = the mean value of the red layer, Mean G = the mean value of the green layer, Mean B = the mean value of the blue layer, R = Red pixel, G = Green pixel, and B = Blue pixels.

2.4.2. Texture feature. In this study using Gabor feature as texture extraction on sugarcane. This filter has received considerable attention in processing the proposed image. Also, since these filters can have optimal localization properties in both the spatial and frequency domains, they are perfectly suited for texture segmentation issues. Gabor filters can be viewed as sinusoidal fields of specific frequency and orientation, modulated by Gaussian envelope [13]. Gabor Filter formula is:

$$\psi_{(\boldsymbol{\omega},\theta)}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{\sigma^2}\right)\exp(2j\pi\boldsymbol{\omega}x) \tag{7}$$

where $j = \sqrt{-1}$

Gabor filters focus on a certain frequency range. If the input image contains two different texture areas, different local frequencies between regions will detect the texture in one or more image output sub-filters. The basic Gabor function can perform space decomposition. Each Gabor Filter is determined by basic Gabor [11].

2.5. Classification

Features extracted in the previous step are classified into 3 algorithms, naive bayes, artificial neural networks, and support vector machines. Then from the three algorithms are compared, which has the highest accuracy in determining the maturity of sugarcane. After that will get the results whether the sugarcane is in the category immature, semi-mature, and mature.

3. Experiment and result

Detection of sugarcane maturity can be seen with the color difference in each class. Among the mature, semi-mature, and immature classes, it is clear that striking color differences to determine their maturity.

The evaluation stage of sugarcane detection is done by calculating the precision-recall curve value, where the precision-recall curve formula

Pricision :

$$\frac{\textit{the number of correct maturity of sugarcane detection}}{\textit{the}\ \text{number of sugarcane maturity detection}} \tag{8}$$

IOP Conf. Series: Materials Science and Engineering **434** (2018) 012065 doi:10.1088/1757-899X/434/1/012065

Recall :

$$\frac{the\ number\ of\ correct\ maturity\ of\ sugarcane\ detection}{the\ number\ of\ true\ maturity\ of\ sugarcane} \quad (9)$$

Here are the results precision and recall from sugarcane maturity detection

Table 2. Precision and recall values of Sugarcane

Method	Precision	Recall
SVM	87,4%	85,7%
Naïve Bayes	62,3%	89,0%
ANN	86,6%	83,3%

Table 2. showed using a collection of sugarcane datasets can yield that SVM method has the highest precision value of 87.4%, while Naive Bayes has the highest recall value of 89% in the dataset set.

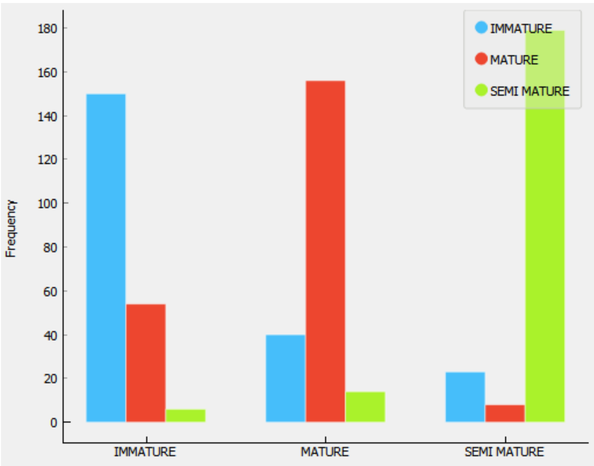

Figure 3. The classification graph using SVM.

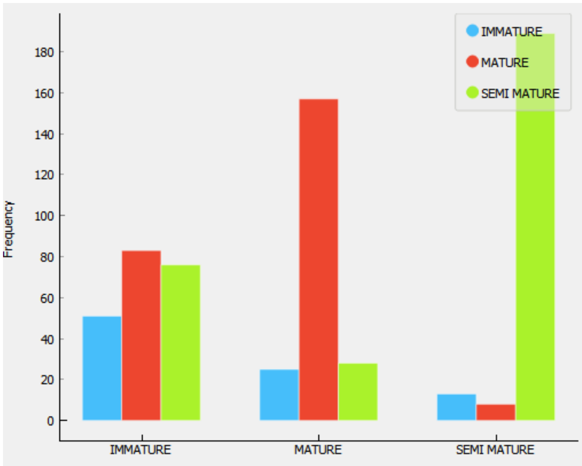

Figure 4. The classification graph using Naive Bayes.

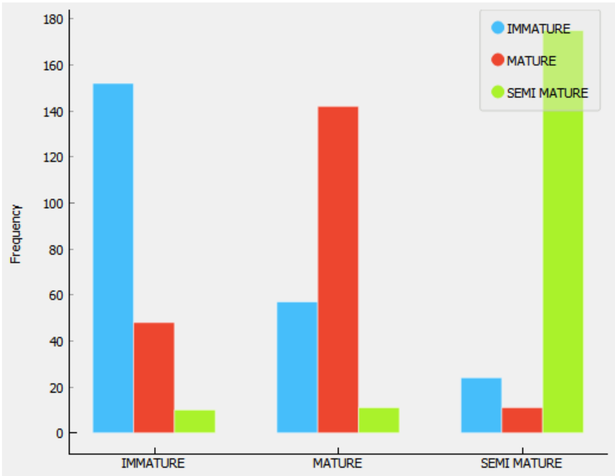

Figure 5. The classification graph using ANN.

Figure 2-4 shows the graphic results of the three classification methods used. Classification using SVM yields high accuracy values for each class. The naive bayes method produces high frequency values in the mature and semi-mature classes but has no high accuracy in the immature or raw classes. In the artificial neural network method, it produces high accuracy values for each class, but if the accuracy value is compared with the SVM method, then the SVM method has the highest accuracy between the three methods. Based result on the trial and analysis above it is known that Support Vector Machine has the best results with precision and recall of 87,4% and 85,7% respectively.

4. Conclusion

Sugarcane is a long duration crop which reaches its maturity in 11-12 months. Crop growth is very slow at the initial stage i.e. it takes 25-30 days to complete germination and another 90-95 days to complete tillering. Knowledge of sugarcane variety is a difficult thing to do. The data on this knowledge is a complex data. Currently, varieties of sugarcane can be done manually by expert sugarcane. This study proposes to build maturity stem detection of cane.

Image processing needs a dataset includes several stages produce 300 image cane. Based result on the trial and analysis above it is known that Support Vector Machine has the best results with precision and recall of 87,4% and 85,7% respectively. In this study proposes all third algorithm for classification method to compare which algorithm has the highest accuracy.

This dataset was good enough to be used for classification because not only the dataset affects but amount of percentage of training data and testing data also affects the classification results.

References

[1] M P R Beulah and O A M Jafar 2017 Prediction of Sugarcane Diseases using Data Mining Techniques *2016 IEEE Int. Conf. Adv. Comput. Appl.* **6**(7) pp. 313–318
[2] M Sujaritha, S Annadurai, J Satheeshkumar, S Kowshik Sharan, and L Mahesh 2017 Weed detecting robot in sugarcane fields using fuzzy real time classifier *Comput. Electron. Agric.* **134** pp. 160–171
[3] E Santoro, E M Soler and A C Cherri 2017 Route optimization in mechanized sugarcane harvesting *Comput. Electron. Agric.***141** pp. 140–146
[4] R V H G Utomo, Adi Heru, and Riyanarto Sarno 2016 Weighted Directed Acyclic Graph Similarity *2016 Int. Conf. Information, Commun. Technol. Syst.* pp. 226–230
[5] S Kumar, S Mishra, P Khanna and Pragyan 2017 Precision Sugarcane Monitoring Using SVM Classifier *Procedia Comput. Sci.* **122** pp. 881–887
[6] R K Dewi and R V H Ginardi 2014 Feature extraction for identification of sugarcane rust disease *Proc. 2014 Int. Conf. Information, Commun. Technol. Syst. ICTS 2014* **14** pp. 99–103
[7] M P Mello, C Atzberger and A R Formaggio 2014 Near real time yield estimation for sugarcane in Brazil combining remote sensing and official statistical data *Int. Geosci. Remote Sens. Symp.* pp. 5064–5067
[8] M Wang, Y Wan, Z Ye and X Lai 2017 Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm *Inf. Sci. (Ny).* **402** pp. 50–68
[9] Z E Rasjid and R Setiawan 2017 Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques *Procedia Comput. Sci.* **116** pp. 107–112
[10] S Kaymak, A Helwan and D Uzun 2017 Breast cancer image classification using artificial neural networks *Procedia Comput. Sci.* **120** pp. 126–131
[11] A Premana, A P Wijaya and M A Soeleman 2017 Image segmentation using Gabor filter and K-means clustering method *2017 Int. Semin. Appl. Technol. Inf. Commun.* pp. 95–99
[12] M P Arakeri and Lakshmana 2016 Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry *Procedia Comput. Sci.* **79** pp. 426–433
[13] A Younesi and M C Amirani 2017 Gabor Filter and Texture based Features for Palmprint Recognition *Procedia Comput. Sci.***108** June pp. 2488–2495