Special issue
Sci.Int.(Lahore),29(2),121-125,2017 ISSN 1013-5316;CODEN: SINTE 8

IDENTIFICATION OF DORSAL AND VENTRAL SURFACE OF RUBBER SEED USING IMAGE PROCESSING AND MACHINE LEARNING APPROACH

Siti Nurul 'Afiah M. Johari*, Siti Khairunniza-Bejo*, W. Ishak Wan Ismail
Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor
*For correspondence; Tel. + (60) 132717976, E-mail: sitinurulafiah@gmail.com
*For correspondence; Tel. + (60) 389464332, E-mail: skbejo@upm.edu.com

ABSTRACT: *Natural rubber tree is one of major plantation crops in Malaysia. To increase the production and germination of the rubber, proper placement of seeds is needed. Ventral surface of rubber seed needs to be placed downward attached to the soil. Nowadays, it is necessary to use an automatic detection technology in order to reduce labor intensity and improve the production efficiency. Therefore, this study was conducted to identify the dorsal and ventral surface of rubber seeds using image processing and machine learning approach. Canny edge detection was used to identify features at the center of the seed namely maximum length of detected edge, ratio major and minor axis, number of pixel, maximum convolution, and number of intersection. These features were used as the input parameters in classifying the dorsal and ventral surface at horizontal position. A new prediction model using decision rule was developed for identification of the dorsal and ventral surface. Support Vector Machine (SVM) and Artificial Neural Network (ANN) were also used for the classification. Based on the results, a new prediction model gave the best percentage of classification with 83.5% successful compared to ANN (67.9%) and SVM (61.5%).*

Keywords: Image processing, edge detection, rubber seed, decision rule, Artificial Neural Network, Support Vector Machine

1. INTRODUCTION

Rubber is one of the major plantation crops in Malaysia. It is an important economic resources for the world. Malaysia is one of the leading producer and exporter of natural rubber. However, according to the Malaysia Rubber Board, the production of the rubber in Malaysia was declined drastically over a million tons in 2016 (187,690) from 2015 (722,122). High germination of rubber is needed to increase its germination. Therefore, the seed need to be placed in a proper position during planting.

This research was done due to the main problem faced by the labor in the rubber nursery industries. They need to determine the dorsal and ventral surfaces manually during planting. The germination of the rubber seeds will be high if the seeds were planted correctly as the dorsal surface in on the top meanwhile, ventral surface was at the bottom, facing to the soil with the suitable germination requirement such as sunlight, water and fertilizer. It is necessary to use an automatic detection technology to reduce such a huge labor intensity and improve the production efficiency. It is significant to develop a model which is a high speed and worth cost system in detecting the surface of rubber seeds.

Machine technology and spectral analysis are widely applied in agriculture, especially to inspect the quality and grading the agriculture product [1]. Seed recognition system has been developed by applying the image processing methods. Granitto *et al.* [2] had applied morphological operations ,namely color and texture characteristics to identify weed seeds with precision rate 98%. Meanwhile, Chanjiang and Guangrong [3] had extracted three main weed seed features i.e. seed area, seed perimeter, and the axis of the enclosing seed rectangle. The system used a back-propagation neural network (BPNN) to recognize the weed seed image with 94% accuracy. Zhong-zhi and You-gang [4] had developed a peanut seed recognition system. Features involved were texture and color of the peanut images. Artificial neural network (ANN) has been applied to recognize the peanut seeds with the precision of 93%. Wencang and Junxin [5] used morphological characteristics and a back propagation network (BPNN) to identify weed seeds with 98% precision rate.

Jinwei *et al*. [6] had developed the rapeseed recognition system. The system only used color features such as RGB (red, green, blue). HSV (hue, saturation, value) and NCM (nine color model). The system employed a rule-based technique to recognize the rapeseed and the result achieved 92.72% successful identification. Hadzli *et al*. [7] proposed a method of classification of rubber clones using image processing technique, including thresholding and morphological operations. Two models were proposed using Principal Component Analysis (PCA) and trained by an artificial neural network (ANN) using Levenberg-Marquardt (LM) algorithm to recognize rubber seed. The selectivity for both models achieved 86% and 96% accuracy, respectively. Currently, there is no image processing technique available to differentiate between dorsal and ventral surface of rubber seed. The current method of identification is human based,which is having the disadvantages of low efficiency, low speed and high cost of labor.

The main objective of this research is to identify the dorsal and ventral of rubber seeds using an image processing approach. An appropriate features will be extracted and a new prediction model will be developed by using all extracted features as the input parameters in identifying the surface of dorsal and ventral in horizontal position. This method will be developed based on suitability in automated rubber seeds.

2. EXPERIMENTAL DETAILS

There were seven steps of research methods and five types of features being analyzed as shown in figure (1). It started with image acquisition and followed by an image pre-processing method includes conversion from RGB (red, green, blue) to HSV (hue, saturation, value) color image, noise removal i.e. morphological operation and canny edge detection. Features namely maximum length of the detected edge, the ratio major and minor axis, number of pixel, maximum convolution, and the number of intersection were analyzed and used to create a prediction model to identify the surface of dorsal and ventral.

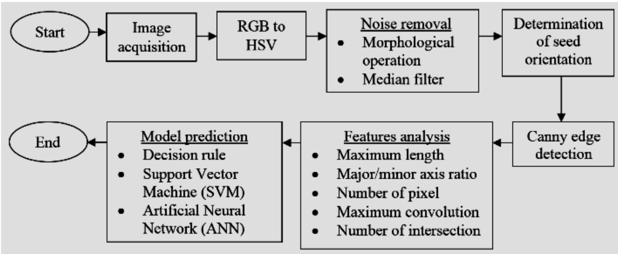

Fig (1) research flowchart

Rubber seeds were collected from Ladang Getah Chuping, Perlis. There were 2720 images of dorsal and ventral surfaces acquired using camera model SM-P605 Samsung Galaxy Tab in RGB (red, green, blue) color format. It was taken perpendicularly at 80 cm height with fixed setting and constant light intensity. Figure (2) shows images of rubber seeds in horizontal position. From the images, it can be seen clearly the significant difference between the dorsal and ventral surface appears at the center of the seed. There was a straight valley at the center of ventral surface whereby numerous darker mottles on the dorsal surface.

Fig (2) images of rubber seed. (a) Dorsal. (b) Ventral surface

All sample images were then converted into HSV (hue, saturation, value) color space as it provides more convenient representation, especially for background as shown in figure (3), (4), (5). It was clearly show that hue band image in figure (3) was clearer to separate seed image (object) from the background), whereby, saturation band image in figure (4) was suitable for edge extraction during edge detection. Therefore, hue band image was selected for background separation meanwhile saturation band image was used during edge detection. Histogram of hue band image has been analyzed to get suitable threshold value for segmentation. Threshold value of 0.1 was chosen to separate seed and background for both dorsal and ventral surface.

Fig (3) hue band image

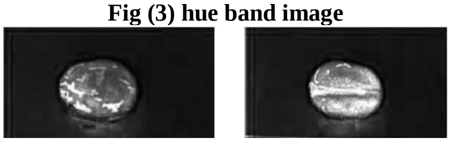

Fig (4) saturation band image

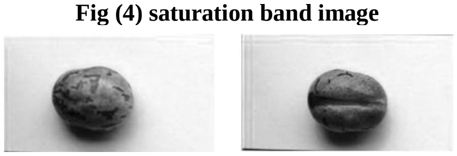

Fig (5) value band image

After performing segmentation process, there were still some noises existed as it were, not properly segmented as shown in figure (6). Area based filter was done by removing all the area of noises that have less than 1000 pixels in the background. Morphological operation with structuring element of disk-shaped with radius 5 pixels was then applied to remove small dot inside the seed. Figure (7) shows the final thresholded image after removing small dot in the seed image.

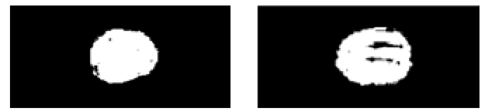

Fig (6) after performing image segmentation

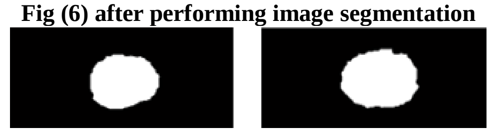

Fig (7) final thresholded image

Ventral surface can be identified based on the straight valley at the center of the seeds. Hence, canny edge detection was done to extract the center edges for both surfaces using saturation band image (as mentioned previously). Region of interest (ROI) with window size of 65 height and 65 pixels width was selected based on the centroid of the saturation image. Therefore, only significant extracted edges were remained for both surfaces. Median filter with window size of 5 by 5 pixels was performed to improve the detected edges and remove small noises. Image after performed edge detection and ROI was shown in figure (8).

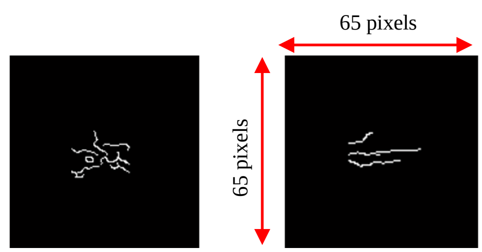

Fig (8) after performing canny edge detection and ROI

Another ROI was selected around the centroid of detected edges with window size 20 pixels height and 25 pixels width as shown in figure (9). This is due to some cases which the center features were not located at the center. Thus, figure (10) shows the final output image was formed which was more focus at the center. Straight line appeared at the center for ventral surface, meanwhile, curvy line was appeared for dorsal surface.

Fig (9) after performing canny edge detection and ROI

Fig (10) final output image

It can be concluded that canny edge detector successfully extracted important center edges of the seeds. Both edge images of dorsal and ventral were then used in developing a new prediction model for automatic detection dorsal and ventral surface of rubber seed.

2.1 FEATURES EXTRACTION

Maximum length of detected edge

Maximum length of detected edges was defined as the longest edges at the x-axis in the image. Only one edge with maximum length value was selected to represent the sample. The length of edges was calculated from one end to another end indicated by dashed line as shown in figure (11).

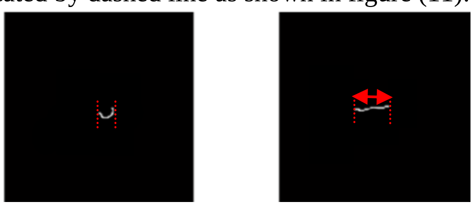

Fig (11) maximum length of detected edge

From the figure above, it clearly shows that the length of center edge for ventral surface was longer compared to dorsal surface. It was 40 and 18 for ventral and dorsal surface, respectively. Therefore, value of 40 was selected as threshold value to identify the surface image in te prediction model.

Ratio major and minor axis of detected edge

A bounding box was created for the detected edge to calculate the ratio of the major (y-axis) and minor (x-axis) as shown in figure (12). Value of y-axis represent the height of the detected edge and value of x-axis represent the width of the detected edge. The ratio of the ventral surface of detected edge will be smaller as the y-axis value is smaller compared to dorsal surface which has higher value of y-axis. Thus, the ratio will be higher for dorsal surface.

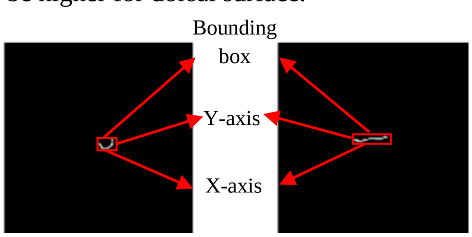

Fig (12) ratio of major and minor axis of detected edge

From figure (12), the ratio value of dorsal surface was 0.3, whereby for ventral surface was 0.06. Therefore, value of 0.3 was selected as threshold value to identify the surface image in prediction model.

Number of pixels

Number of pixels was defined as te total area of edge region inside a single detected edge of the image. The number of pixel in the ventral surface with clear long edge line will be higher compared to dorsal surface. Thus, a value of 50 was selected as threshold value to identify the surface seed.

Maximum convolution

Convolution operation was performed to convolute a single edge line by using kernel window size of 3 by 3 pixels. The purpose of convolution was to identify number of the edge intersection. Higher maximum value represent more pixels inside a 3 by 3 kernel window which indicates existence of the edge intersection and belongs to the dorsal surface. Thus, maximum convolution of four was selected as a threshold value.

Number of intersection

Number of intersection was defined as any convolution value which is above three. Some intersections might occur in ventral surface. It was defined as ventral surface if the convolution value was less than five. If the convolution value more than five and above, it was defined as dorsal surface. Therefore, value of 5 was selected as threshold value of the seed identification.

3. RESULTS AND DISCUSSION

Prediction model

A prediction model was developed by using conditional method (classification rule) in MATLAB. All the features namely maximum length of detected edges, the ratio major and minor axis, number of pixels, maximum convolution, and the number of intersection was used as an input parameters to develop a new prediction model as shown in figure (13).

It was started with maximum length of detected edge and followed by others features. If the maximum length was more than 40, it was detected as ventral surface, otherwise, it moved to next features which was maximum convolution. If the maximum convolution was less than 4 and number of intersection was less than 5, it detected as ventral surface. If the maximum convolution was less than 4, number of intersection was more than 5 and the ratio major and minor axis was more than 0.3, it detected as dorsal.

Overall classification for dorsal and ventral surface was shown in table (1). In average, there were 83.5% successful detection and 16.5% unsuccessful detection.

Table (1) overall performance using prediction model

	Dorsal	Ventral	Average
Correct	87%	78%	83.5%
Incorrect	13%	22%	16.5%

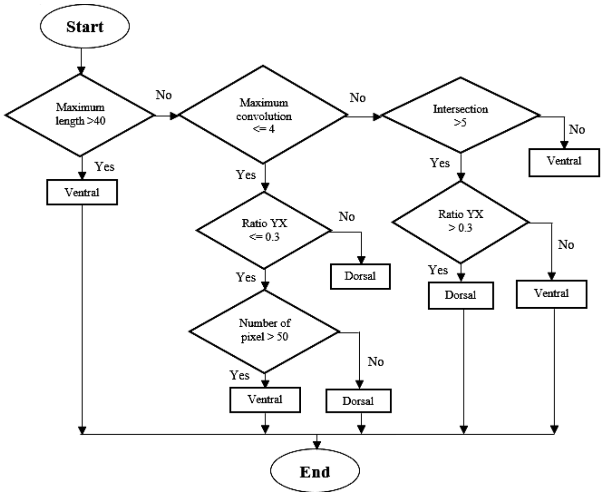

Fig (13) flowchart of decision rule

Support Vector Machine (SVM)
Linear classifier was chosen to separate the data into their respective classes. All the features was used as the input data. It will classify all the data into two classes (dorsal and ventral surface). 1020 images for both surfaces were trained and another 1700 were used for classification.
Overall classification using SVM was summarized in table (2). The classification for dorsal surface was good compared to ventral surface. Average for correct and incorrect classification was calculated in table (2) below:

Table (2) overall performance using SVM

	Dorsal	Ventral	Average
Correct	92%	31%	61.5%
Incorrect	8%	69%	38.5%

Artificial Neural Network (ANN)
Neural network acts as proficient classifier and well suited for addressing non-linear problem such as classification. All five features act as the input data to the neural network and two classes (dorsal and ventral surface) act as target. The number of hidden layer was fixed into one and various number of neuron were used i.e. 5 neurons, 10 neurons, 15 neurons, 20 neurons, 25 neurons, and 30 neurons. The results were summarized in table (3) and only one number of neuron with the highest correct classification was selected.

Table (3) results of ANN based on number of neurons

Number of neurons	Correct classifications	Incorrect classification
5	49.3%	50.7%
10	67.9%	32.1%
15	48.7%	51.3%
20	61.9%	39.1%
25	50.5%	49.5%
30	47.5%	52.5%

From table (3), it shows that number of neurons of 10 has the highest value with 67.9% correct classification compared to other neurons. The confusion matrix of 10 neurons was shown in figure (14). Green box represent percentages of correct classification, meanwhile, red box represent percentages of incorrect classification. Blue box represent overall accuracy. Therefore, the overall accuracy of correct and incorrect classification were 67.9% and 32.1%, respectively.

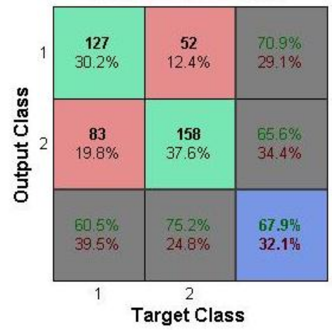

Fig (14) confusion matrix using 10 neurons

Comparison between prediction model, SVM and ANN
There were three models have been used to identify the dorsal and ventral surface of rubber seeds. All the results includes all three models were summarized in table (4). From table (4), it shows that new prediction model gave the highest performance of 83.5% followed by Artificial Neural Network (ANN) 67.9% and Support Vector Machine (SVM) 61.5%.

Table (4) overall classification performance of prediction model, SVM and ANN

Classes/Models	Prediction model (%)	SVM (%)	ANN (%)
Correct	83.5	61.5	67.9
Incorrect	16.5	39.5	32.1

From the result in table (4), it shows that a new developed prediction model was more likely suitable in identifying the surface of dorsal and ventral as the model was simple and easy to understand and interpret. Furthermore, it used a white box model which the result was provided by the model itself compared to ANN and AVM. Both of them used too much of a black box which make them very difficult to train.

4. CONCLUSIONS

This research study was conducted to identify the surfaces of dorsal and ventral surface in order to place them correctly during planting. A new prediction model was developed for the identification of dorsal and ventral based on the significant different at the center of the seed using image processing technique. By using all the features as an input parameter, it was tested with Artificial Neural Network (ANN) and Support Vector Machine (SVM).
Based on the performance from all three models, it shows that a new prediction model gave the best successful identification of 83.5% compared to ANN and SVM. For future research, the study can be improvised by using 3D image whereby it will be more accurate in identifying the surfaces of the rubber seeds.

REFERENCES

[1] C. Yud-Ren, C. Kuanglin, Moon S. Kim. “Machine vision technology for agricultural applications”, *Journal Computers and Electronics in Agriculture*, **36:**173-191, (2002).

[2] M. Granitto, F. Verdes, and H. Alejandro. “Large-Scale Investigation of Weed Seed Identification by Machine Vision”, *Journal of Computer and Electronics in Agriculture*, **47**(1), (2005).

[3] S. Changjiang and J. Guangrong. “Recognition Method of Weed Seeds Based on Computer Vision”, in *Proceeding of International Congress on Image and Signal Processing*, 1-6, (2009).

[4] H. Zhong-Zhi and Z. You-gang. “A Method of Detecting Peanut Cultivars and Quality Based on the Appearance Characteristics Recognition”, in *Proceeding of the International Conference on Information and Computer Science*, 21-24, (2009).

[5] Z. Wencang, and W. Junxin. “Study of Feature Extraction-Based Visual Invariance and Species Identification on Weed Seeds”, in *Proceeding of the International Conference on Natural Computation*, 631-635, (2010).

[6] L. Jinwei, L. Guiping, and F. Xiao. “Rapeseed Seeds Colour Recognition by Machine Vision”, in *Proceeding of Chinese Control Conference*, 146-149, (2008).

[7]H. Hadzli, O. Fairul, J. Syed Abdul Mutalib, H. Muhammad Adib and S. Hajar. “An Intelligent Classification Model for Rubber Seed Clones Based on Shape Features through Imaging Techniques”, in *Proceeding of the International Conference on Intelligent Systems, Modeling and Simulation*, 25-31, (2010).

*For correspondence; Tel. + (60) 132717976, E-mail:sitinurulafiah@gmail.com

*For correspondence; Tel. + (60) 389464332, E-mail: skbejo@upm.edu.com