Automated Rubber Seed Clones Identification Using Imaging Technique And Statistical Analysis

Muhammad Fahmi Bin Harun
Faculty of Electrical Engineering, Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia
e-mail:fahmiharun88@gmail.com

***Abstract*— This paper describe research work to recognize selected rubber seed series clones using image processing techniques based on perimeter, area and radius. There are five types of rubber seed clone have been used as samples in this project which are RRIM2002, RRIM2015, RRIM2020, RRIM2023 and RRIM2024. Sample of rubber tree seeds are captured using digital camera where the RGB color image are stored and processed. Processing involves segmentation algorithm which includes thresholding and application of morphological technique to extract the shape features. Another 225 samples are used for testing and the analysis is done using SPSS software to identify the clones correctly. From the observed one-way ANOVA and error plot measurement, it shown that all of clones series significantly different from each other for perimeter classification but only two series shows significantly different for area and radius classification. As a conclusion, perimeter of rubber seed clone can be used in order to recognize all selected best rubber tree clones compare with area and radius can be used for RRIM2002 and RRIM2015only.**

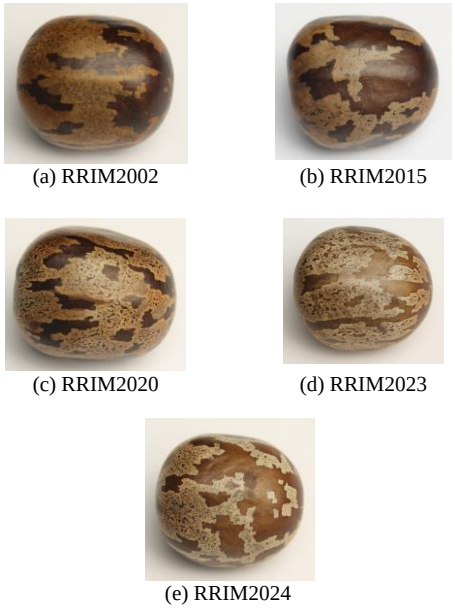
Keywords: Digital Image Processing, Rubber seed clones, SPSS, MATLAB.

I. INTRODUCTION

Rubber is one of the important agricultural products of Malaysia. In order to ensure continuous supply of quality latex and heveawood to the manufacturers, replanting with higher yielding rubber tree clones is hence very important [1]. Therefore, seeds planted must be from the quality rubber tree series clones *e.g*. RRIM2000 Series. There are about 33 types of clone within the RRIM2000 Series. The fundamental of seed identification is by looking at the seeds and try to match its appearance to the closest appearance photo from a library text. Experienced workers will just look at the shape and texture pattern of the seeds in order to decide. However, all of these methods will consume time, percentage accuracy and as well as high labor cost. Since seeds presentation can also be presented in terms of digital images, therefore they can be processed and measured to produce important quantitative features information [2].

This project is extension from the previous project by Hajar Mohd. Salleh, the analysis is still based on shape recognition but in different rubber seed clone and different analysis to prove whether this technique can be used or not to recognize rubber seed series. Application from this factor was used in order to achieve the objective of this project.

A. Clonal characteristic



(a) RRIM2002 (b) RRIM2015

(c) RRIM2020 (d) RRIM2023

(e) RRIM2024

Figure 1: Types of rubber seed clones

Figure 1 show the five type of rubber seed clone. RRIM2002 is medium in size, brownish with smooth and shining seed coat. It has square shape. Overall growth of this clone is considered good. This clone was recommended for both latex and timber production. While RRIM2015 seed size is medium. It was smooth, shining with light brownish seed coat. The shape is square to slightly ovoid. Overall growth and seed production of the clone is good. This clone was highly recommended for latex and timber production. Then RRIM2020 seed features are most likely same with RRIM2015 but it has square to rectangular shaped of seed. Overall growth and seed production of the clone is vigorous. This clone was also highly recommended for latex and timber production. Meanwhile, RRIM2023 seed size is medium. It was smooth, shining with faint brownish seed coat. The shape is round to almost ovoid. Overall growth and seed production of the clone is good. This clone was also recommended for latex and timber production. Finally, RRIM2024 seed features of this clone are most likely same with RRIM2002 but the shape is almost round.Overall growth and seed production of the clone is very good. This clone was also recommended for latex and timber production [1].

II. METHODOLOGY

The process recognition of rubber seed clone can be simplified as below.

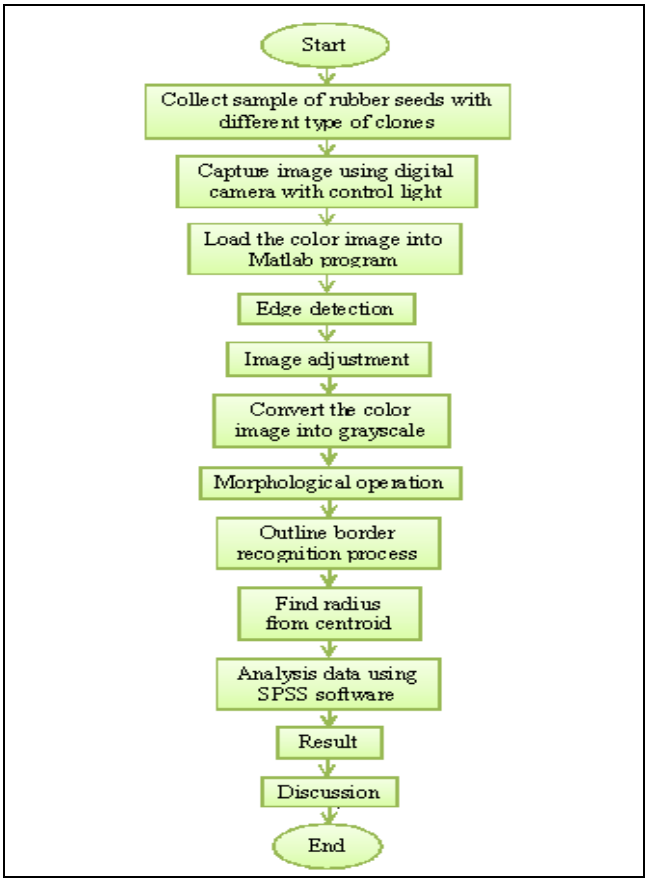


Figure 2: Methodology process

A. Data Collection

In this project, there are five types of rubber seed clones have been selected for recognition of its clone. These five clones are RRIM2002, RRIM2015, RRIM2020, RRIM2023 and RRIM2024. From these five types of rubber seed clones, there are 45 for each sample of data has been collected. All samples were collected at Rubber Research Institute of Malaysia (RRIM), Sungai Buloh and were captured at ICS Room.

The basic RGB color images were obtained using Canon 1000D digital camera, with pixel resolution 1280x960 in JPEG format. This size is suitable for analysis [3]. The image capturing process was done in Image Capturing Studio Room (ICS Room) under control environment at the Advanced Signal Processing (ASP) Lab, Faculty of Electrical Engineering UiTM Shah Alam. For the camera settings, the International Standards Organization (ISO) was set at 800 to gain the sensitivity of image sensor, the shutter is 40, the aperture is F8 and zoom is 55mm.

The camera was placed at a distance of 9 inch directly above the rubber seed along the photo session, at the same time the lighting source was provided by spotlight which is (Digicolor K-2500) at 45° angle from the sample. The light intensity was controlled by a standard low flash having mean lux of 4.31 $\pm$ 0.241 Klux by using Heavy Duty Light Meter. A detailed illustration of the image acquisition and classification process is shown in Figure 3 and 4 below.

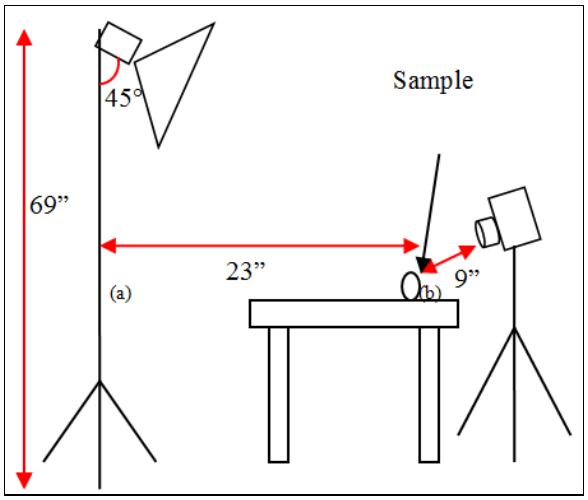


Figure 3: Equipment setting for image capturing.

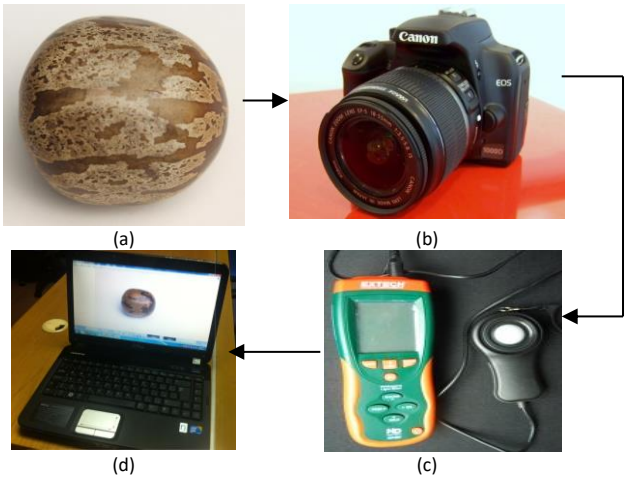


Figure 4: Image acquisition setup
(a) Sample seed (b) camera (c) lux meter and (d) PC

B. Image processing

Image processing is the most critical part since the RGB component will be used to get the perimeter, area and radius. If it is done in improper process, the recognition process will becomes inaccurate. It consists of resizing an image, converting RGB image to grayscale image, edge detection and morphological operation as showed in Figure 5.

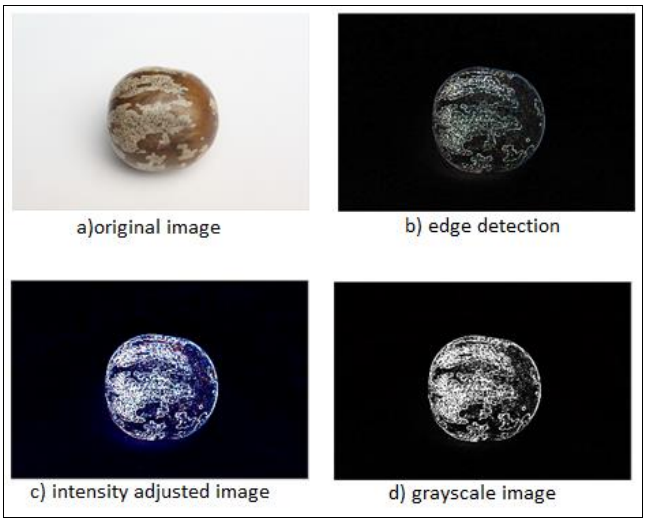


Figure 5: (a) Original image, (b) Edge detection using Sobel, (c) Image adjustment, (d) grayscale image

Morphology is an image processing that relates to the shapes and form of objects. The algorithm involved opening and closing to remove any small objects or noise from an image. It allows object to expand, do closing and opening, thus potentially filling in small holes and connecting disjoint objects [4]. Figure 6 shows when morphological thresholding is done in order to produce the outline of the seed. Area calculation of the outline can be done using the following equation:

$$Area = \sum_{i=0}^{n} pixel_i^{white} \quad (1)$$

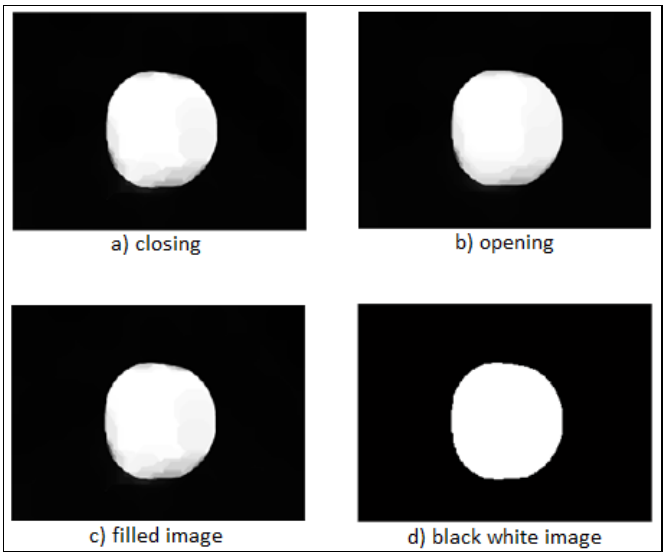


Figure 6: Resulted images of morphological process

Figure 7 shows the image that went through outlined border recognition to performs the masking operation between original image (RGB and grayscale) and outlined border image. The previous procedures must be followed sequentially to find a good outlined border for a rubber seed images. The finding from the outlined border can be used for perimeter calculation of the image as shown:

$$Line_{border} = \sum_{i=0}^{n} pixel_i^{white} \quad (2)$$

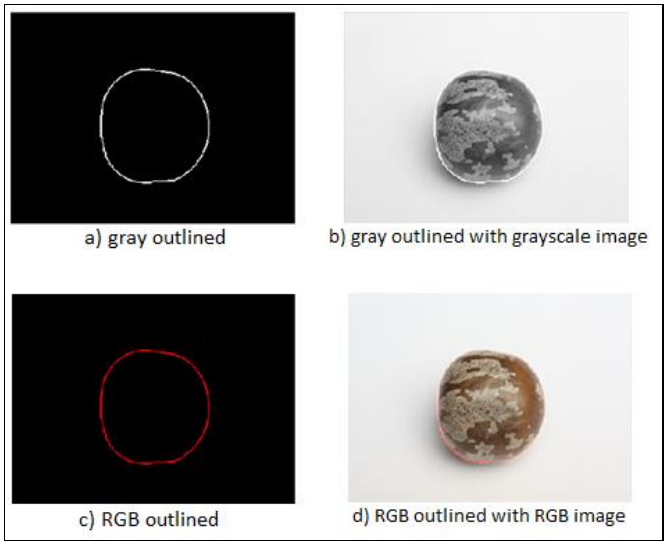


Figure 7: (a)-(d) shows the process of finding the outlined border for rubber seed images.

Figure 8 describe the final process of image processing. It shows the distance of each point of outlined border from centroid. Radius compute the average great circle distance of the data points from the centroid (Xc,Yc) [5]. The resultant of the radius can produce significant result of recognition. Altogether, there were 36 points being produced. The calculation involves the following steps:

$$x_{center}\,(x_c) = \frac{\sum_{i=1}^{row} x_i}{no.of\ pixel} \quad (3)$$

$$y_{center}\,(y_c) = \frac{\sum_{i=1}^{column} y_i}{no.of\ pixel} \quad (4)$$

$$Radius,\ r = \sqrt{\left(y_i - y_c\right)^2 + \left(x_i - x_c\right)^2} \quad (5)$$

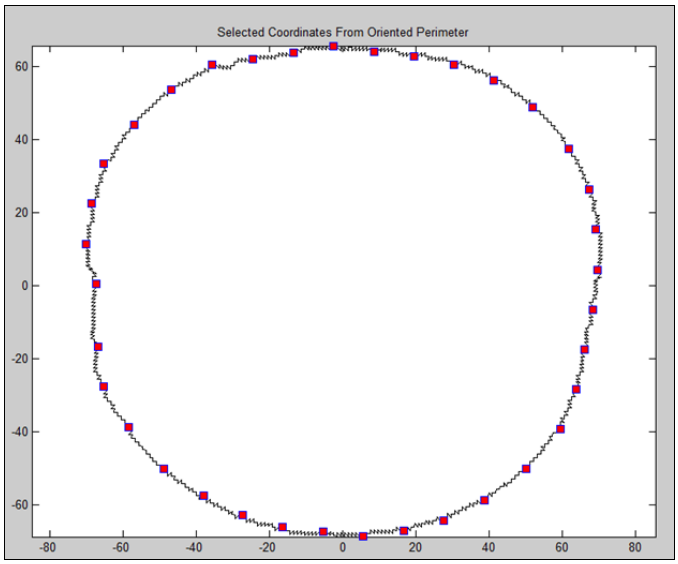


Figure 8: Radius distance of rubber seed image from centroid

III. RESULT AND DISCUSSION

An analysis on the data has been done by using the SPSS software. As a result, two major tests which are one-way ANOVA and error plot are able to prove that:

A. One-way ANOVA

Critical value,α usually set as 0.05.

H_0 = There are no significant differences in area between the rubber seeds clones' mean

H_1 = There is a significant difference in area between the rubber seeds clones' mean

TABLE 1(a): ANOVA TABLE FOR AREA

		Sum of Squares	df	Mean Square	F	Sig.
AREA	Between Groups	474.160	4	118.540	70.168	.000
	Within Groups	371.661	220	1.689		
	Total	845.821	224			

Table 1(a) shows there is a significant difference in area between the types of clones for all rubber seed series, it have strong evidence to conclude that the null(H_0) hypothesis is not true and the alternative (H_1) hypothesis is accepted since the F_{AREA} (4, 220) = 70.168 is greater than $F_{0.05}$ (4, 220) = 2.45 as shown in statistical table. It also show that p-value is lower than 0.05.

H_0 = There are no significant differences in perimeter between the rubber seeds clones' mean

H_1 = There is a significant difference in perimeter between the rubber seeds clones' mean

TABLE 1(b): ANOVA TABLE FOR PERIMETER

		Sum of Squares	df	Mean Square	F	Sig.
PERIMETER	Between Groups	.674	4	.169	80.590	.000
	Within Groups	.460	220	.002		
	Total	1.134	224			

Table 1(b) shows there is a significant difference in perimeter between the types of clones for all rubber seed series, null(H_0) hypothesis is not true and the alternative (H_1) hypothesis is accepted since the F_{AREA} (4, 220) = 80.590 is greater than $F_{0.05}$ (4, 220) = 2.45 from statistical table. P-value also lower than 0.05.

H_0 = There are no significant differences in radius between the rubber seeds clones' mean

H_1 = There is a significant difference in radius between the rubber seeds clones' mean

TABLE 1(c): ANOVA TABLE FOR RADIUS

		Sum of Squares	df	Mean Square	F	Sig.
RADIUS	Between Groups	1.941E7	4	4852153.809	74.884	.000
	Within Groups	1.426E7	220	64795.456		
	Total	3.366E7	224			

Table 1(c) shows there is a significant difference in radius between the types of clones for all rubber seed series, the null(H_0) hypothesis is not true and the alternative (H_1) hypothesis is accepted since the F_{AREA} (4, 220) = 74.884 is greater than $F_{0.05}$ (4, 220) = 2.45 as shown in statistical table. P-value also lower than 0.05.

B. Multiple Comparisons

Table 2(a) below showed the comparison between mean of area and significant values for all series clones with each other. The result shows that only RRIM2002 and RRIM2015 have the entire mean significant below than 0.05. So, all the evidence is proved to assume only RRIM2002 and RRIM2015 are significantly different between each other. Whereas others series clones are no significantly different between each other.

TABLE 2(a): MULTIPLE COMPARISONS OF AREA

(I) CLONE	(J) CLONE	Mean Difference (I-J)	Std. Error	Sig.
2002	2015	-4.48371*	.27401	.000
	2020	-1.41600*	.27401	.000
	2023	-1.76945*	.27401	.000
	2024	-1.95636*	.27401	.000
2015	2002	4.48371*	.27401	.000
	2020	3.06772*	.27401	.000
	2023	2.71426*	.27401	.000
	2024	2.52735*	.27401	.000
2020	2002	1.41600*	.27401	.000
	2015	-3.06772*	.27401	.000
	2023	-.35346	.27401	.698
	2024	-.54037	.27401	.283
2023	2002	1.76945*	.27401	.000
	2015	-2.71426*	.27401	.000
	2020	.35346	.27401	.698
	2024	-.18691	.27401	.960
2024	2002	1.95636*	.27401	.000
	2015	-2.52735*	.27401	.000
	2020	.54037	.27401	.283
	2023	.18691	.27401	.960

*. The mean difference is significant at the 0.05 level.

Table 2(b) showed the comparison between mean of perimeter and significant values for all series clones with each other. The result shows that only RRIM2002 and RRIM2024 doesn't have the entire mean significant below than 0.05. So, all the evidence is proved to assume only RRIM2002 and RRIM2015 are no significantly different between each other. Whereas others series clones are significantly different between each other.

TABLE 2(b): MULTIPLE COMPARISONS OF PERIMETER

(I) CLONE	(J) CLONE	Mean Difference (I-J)	Std. Error	Sig.
2002	2015	-.11988*	.00964	.000
	2020	-.05096*	.00964	.000
	2023	.04418*	.00964	.000
	2024	-.02195	.00964	.156
2015	2002	.11988*	.00964	.000
	2020	.06891*	.00964	.000
	2023	.16405*	.00964	.000
	2024	.09792*	.00964	.000
2020	2002	.05096*	.00964	.000
	2015	-.06891*	.00964	.000
	2023	.09514*	.00964	.000
	2024	.02901*	.00964	.024
2023	2002	-.04418*	.00964	.000
	2015	-.16405*	.00964	.000
	2020	-.09514*	.00964	.000
	2024	-.06613*	.00964	.000
2024	2002	.02195	.00964	.156
	2015	-.09792*	.00964	.000
	2020	-.02901*	.00964	.024
	2023	.06613*	.00964	.000

*. The mean difference is significant at the 0.05 level.

Table 2(c) shows the comparison between mean of radius and significant values for all series clones with each other. The result shows that only RRIM2002 and RRIM2015 have the entire mean significant below than 0.05 same as the analysis for mean of area. So, all the evidence is proved to assume only RRIM2002 and RRIM2015 are significantly different between each other. Whereas others series clones are no significantly different between each other.

TABLE 2(C): MULTIPLE COMPARISONS OF RADIUS

(I) CLONE	(J) CLONE	Mean Difference (I-J)	Std. Error	Sig.
2002	2015	-900.62210*	53.66375	.000
	2020	-295.81423*	53.66375	.000
	2023	-285.85252*	53.66375	.000
	2024	-388.59082*	53.66375	.000
2015	2002	900.62210*	53.66375	.000
	2020	604.80786*	53.66375	.000
	2023	614.76957*	53.66375	.000
	2024	512.03127*	53.66375	.000
2020	2002	295.81423*	53.66375	.000
	2015	-604.80786*	53.66375	.000
	2023	9.96171	53.66375	1.000
	2024	-92.77659	53.66375	.418
2023	2002	285.85252*	53.66375	.000
	2015	-614.76957*	53.66375	.000
	2020	-9.96171	53.66375	1.000
	2024	-102.73830	53.66375	.313
2024	2002	388.59082*	53.66375	.000
	2015	-512.03127*	53.66375	.000
	2020	92.77659	53.66375	.418
	2023	102.73830	53.66375	.313

*. The mean difference is significant at the 0.05 level.

C. Error plots

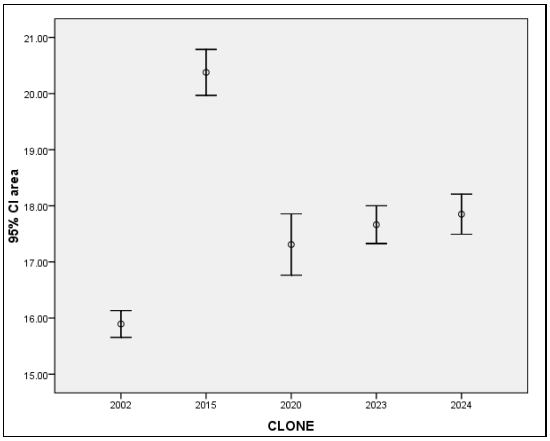


Figure 3(a): Error plot for area comparison

Figure 3(a) shows the error plots based on the sample area of rubber seeds. From observation, RRIM2020, RRIM2023 and RRIM2024 show almost similar each other, so there is no

significantly different between rubber seed series but it is significantly different with RRIM2002 and RRIM2015 clone series. The author noted that only seed series for 2002 and 2015 can be recognize by using area of seed.

TABLE 3(A): DESCRIPTIVE STATISTICS OF AREA

area

	N	Mean	Minimum	Maximum
2002	45	15.8942	13.63	17.24
2015	45	20.3779	17.73	23.49
2020	45	17.3102	13.27	20.29
2023	45	17.6636	15.86	20.45
2024	45	17.8505	15.83	21.00
Total	225	17.8193	13.27	23.49

Table 3(a) shows the Descriptive Statistics area due to minimum, maximum and mean value for all type of clone.

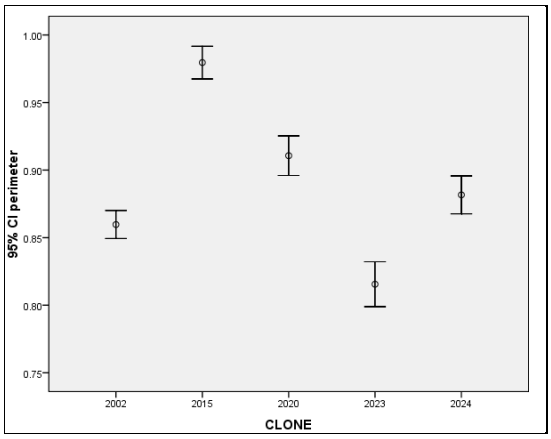


Figure 3(b): Error plot for perimeter comparison

Figure 3(b) shows the error plots based on the sample perimeter of rubber seeds. From observation, each clone series show significantly different each other. The author noted that the different of perimeter each clone series which means the perimeter of seed series is different among their size. All of this rubber seed clone can be recognize by using perimeter of seed.

TABLE 3(b): DESCRIPTIVE STATISTICS FOR PERIMETER

perimeter

	N	Mean	Minimum	Maximum
2002	45	.8597	.73	.90
2015	45	.9796	.90	1.06
2020	45	.9107	.80	.99
2023	45	.8155	.73	.93
2024	45	.8817	.77	.97
Total	225	.8894	.73	1.06

Table 3(b) shows the Descriptive Statistics of perimeter due to minimum, maximum and mean value for all type of clone.

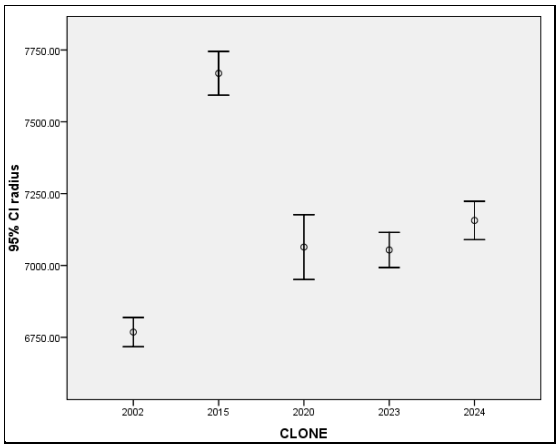


Figure 3(c): Error plot for radius comparison

Figure 3(c) shows the error plots based on the sample radius of rubber seeds. From observation, RRIM2020, RRIM2023 and RRIM2024 show almost similar each other, so there is no significantly different between rubber seed series but it is significantly different with RRIM2002 and RRIM2015 clone series. The author noted that only seed series for 2002 and 2015 can be recognize by using radius of seed same as area comparison.

TABLE 3(c): DESCRIPTIVE STATISTICS FOR RADIUS

radius

	N	Mean	Minimum	Maximum
2002	45	6.7682E3	6273.72	7056.55
2015	45	7.6688E3	7153.20	8244.48
2020	45	7.0640E3	6196.20	7661.76
2023	45	7.0540E3	6647.86	7565.99
2024	45	7.1568E3	6773.14	7756.72
Total	225	7.1424E3	6196.20	8244.48

Table 3(c) shows the Descriptive Statistics of radius due to minimum, maximum and mean value for all type of clone.

IV. CONCLUSION AND FUTURE RECOMMENDATION

A. Conclusion

This research mainly present a contribution in the field of image processing of recognizing selected best series clones for rubber tree industry. The area, perimeter and radius are useful features use to discriminate between clones. These parameters can be used to produce intelligent model system for rubber tree clones classification. Observation from one-way ANOVA and error plots, it can be concluded that the perimeter of rubber seed clone discriminate each type of clone series. As for as the area and radius, only two type of rubber seed series which is RRIM2002 and RRIM2015 are discriminate each other but for the others type is not discriminate each other. Therefore, it can be concluded that analyzing for perimeter of

rubber seed is the best parameter can be used in order to recognize the selected rubber seed clones but for area and radius it can be conclude that the accuracy is not high enough in order to differentiate the different types of rubber seed clones.

B. Future recommendation

From the result obtained, the accuracy for area and radius is not high enough in order to recognize different type of rubber tree seed clones. So, as a recommendation for future analysis, it needs to be added more parameters in order to increase the accuracy of the result. So, from here we know that more parameters will indicate better result.

ACKNOWLEDGEMENT

The author would like to acknowledge Mr Fairul Nazmie, project supervisor for his valuable advice, ideas and critical guidance throughout the preparation of this project. Also acknowledgement to Miss Nor Ezan for her support in sharing the knowledge in image processing has made this project successful. Not to forget, appreciation to Rubber Research Institute of Malaysia (RRIM) for sharing their knowledge and providing the data prior to the preparation of this project.

REFERENCES

[1] M. Z. A. Aziz, D. R. Othman, D. M. Benong, and D. O. S. Huat, *RRIM2000 SERIES Clones: Characteristic and Description*: Rubber Research Institute Of Malaysia, 1997.

[2] H. M. Salleh, "*An Automated Shape Recognition for Rubber Seed Clones Through Imaging Techniques*," in Faculty of Electrical Engineering, vol. Bsc.Eng (Hons.) Electrical.
Shah Alam: Universiti Teknologi Mara, 2007, pp. 9.

[3] A.Bittorf, M.Fartasch, G.Schuler, and T.L.Diepgen, *Resolution requirements for digital images in Dermatology*, 1997.

[4] R. C.Gonzalez, R. E.Woods, and S. L.Eddins, *Digital Image Processing Using MATLAB*: Prentice Hall, 2004.

[5] MATLAB, "Lavenrberg Marquardt, Radius," in *MATLAB 7.0*, 2007.

[6] M. E. Celebi, Y. A. Aslandogan, and P. A. Bergstresser,"*Unsupervised border detection of skin lesion image*,"
Proc. of the Int. Conf. of Information Technology.

[7] Andy P. Field ,"Primary Text: *Discovering Statistics using SPSS*, 2nd edition.
2005.

[8] M. F. Muhammad, "*An Automated Color Recognition for Rubber Tree Seed Clones Using MCS 600 Spectrometer*," in Faculty of Electrical Engineering, vol. Bsc.Eng (Hons.) Electrical.
Shah Alam: Universiti Teknologi Mara, 2009.

[9] D. C. Montgomery and G. C. Runger "*Applied Statistic and Probability for Engineers*," Fifth Edition
2011

[10] Sobel Edge Detector, Access 20/03/2012
http://homepages.inf.ed.ac.uk/rbf/HIPR2/sobel.htm