

Faculty of Electrical and Electronic Engineering Technology



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ISZUZULDIN AMIRULL BIN MOHD JANUDIN

Bachelor of Electrical Engineering Technology (Industrial Automation & Robotics) with Honours

ESTIMATING THE WEIGHT OF MANGOES USING IMAGE PROCESSING AND ANALYSIS TECHNIQUES

ISZUZULDIN AMIRULL BIN MOHD JANUDIN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DECLARATION

I declare that this project report entitled "Estimating The Weight of Mangoes Using Image Processing and Analysis Techniques" is the result of my research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.



APPROVAL

I hereby declare that I have checked this project report and in my opinion, this project report is adequate in terms of scope and quality for the award of the degree of Bachelor of Electrical Engineering Technology (Industrial Automation & Robotics) with Honours.

Signature :	Ahmad Idil
Supervisor Name : TS. AH	MAD IDIL BIN ABDUL RAHMAN
Date : 11 JAN	UARY 2022
Signature مليك	lever michaline
Co-Supervisor	KNIKAL MALAYSIA MELAKA
Name TS. JOH	IAR AKBAR BIN MOHAMAT GANI
Date : 11 JAN	UARY 2022

DEDICATION

I want to thank everyone who has helped me with this thesis. Furthermore, I am grateful to both my parents and siblings for their constant support and encouragement in helping me finish this bachelor's degree project (BDP). I wouldn't have made it this far without them. Apart from that, I'd want to thank all of my fellow lecturers for their guidance, instruction, and mentoring during my studies. Not to mention all of my dear friends I've met along the way. Finally, I'd like to express my gratitude to everyone who has helped me complete this project, whether directly or indirectly.



ABSTRACT

One of the most important factors that consumers associate with mango quality is its size. The weight of the Mango is commonly used to assess its size. This project aims to offer research that uses image processing and analysis techniques to estimate the weight of mangoes. The mango picture produced from the image acquisition system was processed and analyzed using the MATLAB program. There have two methods were used to find the number of pixels of mangoes, which are area and perimeter. This project offers an application that can choose which method is most suitable for the estimated weight of mango based on observations of r-squared values. The relationship between mango pixels and mango weights was examined using linear regression as a statistical approach. As a consequence, methodologies for mango weight assessment using image processing and analysis are practical, viable, and successful.



ABSTRAK

Salah satu faktor terpenting yang dikaitkan pengguna dengan kualiti mangga ialah saiznya. Berat Mangga biasanya digunakan untuk menilai saiznya. Projek ini bertujuan untuk menawarkan penyelidikan yang menggunakan pemprosesan imej dan teknik analisis untuk menganggarkan berat mangga. Gambar mangga yang dihasilkan daripada sistem pemerolehan imej telah diproses dan dianalisis menggunakan program MATLAB. Terdapat dua kaedah yang digunakan untuk mencari bilangan piksel mangga iaitu luas dan perimeter. Projek ini menawarkan aplikasi yang boleh memilih kaedah yang paling sesuai untuk anggaran berat mangga telah dikaji menggunakan regresi linear sebagai pendekatan statistik. Akibatnya, metodologi untuk penilaian berat mangga menggunakan pemprosesan dan analisis imej adalah praktikal, berdaya maju dan berjaya.



ACKNOWLEDGEMENTS

Alhamdulillah, all praises to Allah for the strength and His blessing in completing this project. This project necessitates a great deal of patience and knowledge to end. This project is significant because it is one of the requirements for completing the degree level of electrical engineering technology. Second, I'm grateful to Ts. Ahmad Idil Bin Abdul Rahman, my supervisor, and Ts. Johar Akbar Bin Mohamat Gani, my co-supervisor, contributed to my project by sharing how to manage the project and research themes connected to my project. This project may not be completed on time if they do not provide guidance and encouragement. In addition, I'd like to express my gratitude to my entire family, particularly my father, Mohd Janudin Bin Salleh, and my mother, Noraisah Binti Ahmad. They've been there for me the whole time I've been working on my senior project. Thank you for all of their encouragement and support as we worked through the ups and downs of completing this project. Finally, I'd like to express my gratitude to my friends for their unwavering support. Finally, I appreciate all of the assistance and support received throughout this final year project adventure.

اونيۈم سيتي تيڪنيڪل مليسيا ملاك UNIVERSITI TEKNIKAL MALAYSIA MELAKA

TABLE OF CONTENTS

		PAGE
DEC	CLARATION	
APP	ROVAL	
DED	DICATIONS	
ABS	TRACT	i
ABS	TRAK	ii
ACK	KNOWLEDGEMENTS	iii
ТАВ	BLE OF CONTENTS	i
LIST	T OF TABLES	iv
LIST	r of figures	v
LIST	r of symbols	viii
LIST	T OF ABBREVIATIONS	ix
LIST	r of appendices	x
CHA 1.1 1.2 1.3 1.4	PTER 1 INTRODUCTION Background Problem Statement TI TEKNIKAL MALAYSIA MELAKA Project Objective Scope of Project	11 11 12 13 13
CHA 2 1	APTER 2 LITERATURE REVIEW	15
2.1 2.2	Introduction Image Acquisition and Preprocessing 2.2.1 Digital Camera 2.2.2 Color camera 2.2.3 Camera with a video adapter	15 15 15 18 19
2.3	 Image Segmentation 2.3.1 Thresholding method 2.3.2 Interpretation of the form of an image histogram 2.3.3 The traditional method for image enhancement, de-noising, and edge 	20 20 24
	detection 2.3.4 Image blurring 2.3.5 Curve fitting 2.3.6 Grayscale method	26 28 29 30
2.4	Weight Estimation 2.4.1 Estimate weight using volume estimation	32 32

	2.4.2 Estimating weight using linear regression analysis2.4.3 Estimate weight using neural network	35 36
	2.4.4 Estimating weight using visible imaging	38
2.5	Comparison of Previous Studies	39
2.6	Summary	43
CHA	PTER 3	45
3.1	Introduction	45
3.2	Project Implementation	45
3.3	Project Development	50
	3.3.1 Sampling	50
	3.3.2 Image acquisition and pre-processing	51
	3.3.3 Image processing and analysis	52
	3.3.3.1 Grayscale image	53
	3.3.3.2 Graythresh image	53
	3.3.3.3 Remove noise	54
	3.3.3.4 Finding area	5/
	2.2.4 Statistical analysis	50
	3.3.4 Statistical analysis	59
	3.3.4.1 Collection coefficient	59
	3 3 5 Performance evaluation	66
34	Component Involves	69
5.4	3.4.1 Web camera with stand	70
	3.4.2 Mini studio box	70
	3.4.3 Scales	72
	3.4.4 Mango	73
3.5	Summary	74
CHA	PTER 4 NIVEFRESULTS AND DISCUSSIONS A MELAKA	75
4.1	Introduction	75
4.2	Image Analysis	75
4.3	System Implementation	78
	4.3.1 Training session	79
	4.3.2 Testing session	80
4.4	Data Analysis	81
	4.4.1 Area	82
	4.4.2 Perimeter	85
4.5	Discussion	88
4.6	Performance Evaluation	89
4./	Summary	91
CHA	PTER 5 CONCLUSION AND RECOMMENDATIONS	92
5.1	Introduction	92
5.2	Conclusion	92
5.3	Recommendation for Future Project Improvement	93
5.4	Commercialization Relevancy	93
REFE	ERENCES	95

APPENDICES



LIST OF TABLES

TABLETITLE	PAGE
Table 2.1 Reference values for weight-based fruit classification	34
Table 2.2 Chosen parameter.	37
Table 2.3 Comparison of previous studies.	39
Table 3.1: Example 1 calculating the correlation coefficient	62
Table 3.2: Example 1 Calculating the Equation of Regression Line	63
Table 3.3: Example 2 calculating the correlation coefficient	63
 Table 3.4: Example 2 Calculating the Equation of Regression Line Table 4.1: Mango area, actual, and estimated weight of mangoes in the reset 84 Table 4.2: Mango perimeter actual and estimated weight of mangoes 	64 eference
reference set	87
Table 4.3: Weight estimation equation and they are Respective r2 values	89

LIST OF FIGURES

FIGURE TITLE	PAGE
Figure 2.1 The image acquisition setup	16
Figure 2.2 Digital photos of Mango	17
Figure 2.3 The setup used to capture the image	17
Figure 2.4 Schematic diagram of the machine vision system for image acquisition	18
Figure 2.5 Schematic diagram of the machine vision system for image acquisition for image acquisition	19
Figure 2.6 Front diffuse illumination scheme	20
Figure 2.7 Binarization on H, S, and V channel images and result comparison	24
Figure 2.8 Bimodal Histogram	25
Figure 2.9 Relation between CB and Cr channel in the YCbCr color space	26
Figure 2.10 Original image and CB channel information	27
Figure 2.11 Histogram of a threshold segmentation and the resulting image	28
Figure 2.12 Top profile of mango	29
Figure 2.13 Threshold with no blurring	29
Figure 2.14 Threshold with blurring window	29
Figure 2.15 Curve fitting the mango boundary with two polynomial equation	30
Figure 2.16 (a) Original image (b) Grey-scale image (c) Binary image (d) Edge detection (e) Contour filling (f) Filtere image	31
Figure 2.17 Pixel value for the segmentation image	31
Figure 2.18 A cross-section of the fruit	32
Figure 2.19 Ratio analysis at different points	33
Figure 2.20 Major and minor axis on fruit image	34
Figure 2.21 Mango dimension measurement	38
Figure 3.1 Flowchart of project implementation	49

Figure 3.2: Image acquisition system	52
Figure 3.3: RGB image or input image	52
Figure 3.4: Binarize and thresholded image	54
Figure 3.5: Fill holes image	55
Figure 3.6: Remove blob on the border-image	56
Figure 3.7: Remove small blob	56
Figure 3.8: Output image	56
Figure 3.9: Number of white region area	57
Figure 3.10: The properties of region prop	58
Figure 3.11: Number of perimeters	59
Figure 3.12: Sketched regression line example 1	65
Figure 3.13 Sketched regression line example 2	65
Figure 3.14: Comparison linear regression line between example 1 and 2	66
Figure 3.15: Flowchart of the overall system	67
Figure 3.16: Flowchart of image acquisition	68
Figure 3.17: Flowchart of image processing _ MALAYSIA MELAKA	68
Figure 3.18: Statistical Analysis	69
Figure 3.19: ANBIUX USB webcam	70
Figure 3.20: Webcam stand	71
Figure 3.21: Mini studio box	71
Figure 3.22: Mini scale	72
Figure 3.23: Mango	73
Figure 4.1 Original image for web camera	77
Figure 4.2 Adjust YCbCr	77
Figure 4.3 Adjust HSV value	77
Figure 4.4 Algorithm to find number of black pixels Mango	78

Figure 4.5: Image analysis result	78
Figure 4.6: Training application interface in MATLAB	80
Figure 4.7: Testing application interface in MATLAB	81
Figure 4.8: The dependency between area and actual weight	84
Figure 4.9: 95% confident interval for intercept and slope of the linear model of the area	85
Figure 4.10: Calculating all the area data using excel	85
Figure 4.11: The dependency between perimeter and actual weight	87
Figure 4.12: 95% confident interval for intercept and slope of the linear model of the perimeter	88
Figure 4.13 Calculating all the perimeter data using excel	88
Figure 4.14: Actual weight and estimated weight method 1: area	90
Figure 4.15: Actual weight and estimated weight method 2: perimeter	90
اويوم سيبي ميت يت ميسيا مارد	

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF SYMBOLS

t - ton % - Percentage cm - Centimeter mm - Millimeter s - Seconds dB - Decibel T - Threshold ρ - Density v - Volume m - Mass Df - Depth factor g - Gram cm^3 - Cubic centimeter r^2 - Coefficient of Determination r - Correlation Coefficient L - Lenght W - Width H - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ - Mean s - Sample Standard deviation $\sum_{i}^{i} i$ UNIVES Standard deviation ϵ_{rms} - Root Mean Square φ - Accuracy	δ	- Voltage angle
%-Percentagecm-Centimetermm-Millimeters-SecondsdB-DecibelT-Threshold ρ -Density v -Volumem-MassDf-Depth factorg-Gram cm^3 -Cubic centimeter r^2 -Coefficient of Determinationr-Correlation CoefficientL-LenghtW-WidthH-Height \hat{y} -Predicted Valuem-Slopeb-y-intercept μ -Mean s -Sample Standard deviation $\sum i$ iIntervent ϵ -Error ϵ_{rms} -Root Mean Square φ -Accuracy	t	- ton
cm - Centimeter mm - Millimeter s - Seconds dB - Decibel T - Threshold ρ - Density v - Volume m - Mass Df - Depth factor g - Gram cm ³ - Cubic centimeter r ² - Coefficient of Determination r - Correlation Coefficient L - Lenght W - Width H - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ - Mean s - Sample Standard deviation $\sum_{i} i$ INVESTIGATION CONTRACTION ϵ_{rms} - Root Mean Square φ - Accuracy	%	- Percentage
mm - Millimeter s - Seconds dB - Decibel T - Threshold ρ - Density v - Volume m - Mass Df - Depth factor g - Gram cm^3 - Cubic centimeter r^2 - Coefficient of Determination r - Correlation Coefficient L - Lenght W - Width H - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ Mean s - Sample Standard deviation $\sum_{i} i$ i i i i i i i i i i i i i i i i i	cm	- Centimeter
s - Seconds dB - Decibel T - Threshold ρ - Density v - Volume m - Mass Df - Depth factor g - Gram cm^3 - Cubic centimeter r^2 - Coefficient of Determination r - Correlation Coefficient L - Lenght W - Width H - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ Mean s - Sample Standard deviation $\sum_{\epsilon} i$ - Error ϵ_{rms} - Root Mean Square φ - Accuracy	mm	- Millimeter
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	S	- Seconds
T-Threshold ρ -Density v -Volume m -MassDf-Depth factor g -Gram cm^3 -Cubic centimeter r^2 -Coefficient of Determination r -Correlation CoefficientL-LenghtW-WidthH-Height \hat{y} -Predicted Valuem-Slopeb-y-intercept μ Mean s -Sample Standard deviation $\sum i$ Intercept ϵ - ϵ Error ϵ_{rms} - φ	dB	- Decibel
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	Т	- Threshold
$v = Volume$ $m = Mass$ $Df = Depth factor$ $g = Gram$ $cm^{3} = Cubic centimeter$ $r^{2} = Coefficient of Determination$ $r = Correlation Coefficient$ $L = Lenght$ $W = Width$ $H = Height$ $\hat{y} = Predicted Value$ $m = Slope$ $b = y-intercept$ $\mu = Mean$ $s = Sample Standard deviation$ $\sum_{s} i = V$ $Fror$ $\epsilon_{rms} = Root Mean Square$ $\varphi = Accuracy$	ρ	- Density
$ \begin{array}{rcl} m & - & \text{Mass} \\ Df & - & \text{Depth factor} \\ g & - & \text{Gram} \\ cm^3 & - & \text{Cubic centimeter} \\ r^2 & - & \text{Coefficient of Determination} \\ r & - & \text{Correlation Coefficient} \\ L & - & \text{Lenght} \\ W & - & \text{Width} \\ H & - & \text{Height} \\ \hat{y} & - & \text{Predicted Value} \\ m & - & \text{Slope} \\ b & - & y\text{-intercept} \\ \mu & - & \text{Mean} \\ s & - & \text{Sample Standard deviation} \\ \hline \sum_{i} i & i & \text{Intercept} \\ \epsilon & - & \text{Error} \\ \epsilon_{rms} & - & \text{Root Mean Square} \\ \varphi & - & \text{Accuracy} \\ \end{array} $	v	- Volume
Df - Depth factor g - Gram cm^3 - Cubic centimeter r^2 - Coefficient of Determination r - Correlation Coefficient L - Lenght W - Width H - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ - Mean s - Sample Standard deviation $\sum_{i}^{i} i$ INIVESUMENT TEKNIKAL MALAYSIA MELA ϵ - Error ϵ_{rms} - Root Mean Square φ - Accuracy	m	- Mass
$\begin{array}{rcl} g & - & Gram \\ cm^3 & - & Cubic centimeter \\ r^2 & - & Coefficient of Determination \\ r & - & Correlation Coefficient \\ L & - & Lenght \\ \hline W & - & Width \\ H & - & Height \\ \hat{y} & - & Predicted Value \\ m & - & Slope \\ b & - & y-intercept \\ \mu & - & Mean \\ s & - & Sample Standard deviation \\ \hline \sum_{k=1}^{N} i & UNIVESUME THEKNIKAL MALAYSIA MELA \\ \epsilon & - & Error \\ \epsilon_{rms} & - & Root Mean Square \\ \varphi & - & Accuracy \\ \end{array}$	Df	- Depth factor
cm^3 -Cubic centimeter r^2 -Coefficient of Determinationr-Correlation CoefficientL-LenghtW-WidthH-Height \hat{y} -Predicted Valuem-Slopeb-y-intercept μ -Sample Standard deviation $\sum_{i} i$ Image: Sum term of terms ϵ -Error ϵ -Error ϵ - φ -Accuracy-	g	- Gram
r^2 Coefficient of Determinationr-L-L-WidthH-Predicted Valuem-Slopeb-y-intercept μ Sample Standard deviation $\sum_{\epsilon} i$ Error ϵ -Error ϵ -Root Mean Square φ -Accuracy	cm^3	- Cubic centimeter
r-Correlation CoefficientL-LenghtW-WidthH-Height \hat{y} -Predicted Valuem-Slopeb-y-intercept μ -Sample Standard deviation $\sum_{i} i$ Intercept ϵ - ϵ -Error- ϵ -Error- φ -Accuracy	r^2	- Coefficient of Determination
L - Lenght W - Width H - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ Mean s - Sample Standard deviation $\sum_{i} i$ INTERNIKAL MALAYSIA MELA ϵ - Error ϵ_{rms} - Root Mean Square φ - Accuracy	r	- Correlation Coefficient
W - Width H - Height \hat{y} Predicted Value m - Slope b - y-intercept μ - Mean s - Sample Standard deviation $\sum_{i} i$ UNIVERSITITEKNIKAL MALAYSIA MELA ϵ - Error ϵ_{rms} - Root Mean Square φ - Accuracy	L	🚪 - Lenght 🖻
H \hat{y} - Height \hat{y} - Predicted Value m - Slope b - y-intercept μ Mean s - Sample Standard deviation $\sum_{i}^{s} i$ - Error ϵ_{rms} - Root Mean Square φ - Accuracy	W	- Width
$ \hat{y} = Predicted Value m = Slope b = y-intercept \mu = Mean s = Sample Standard deviation \sum_{i} i = \frac{Sum}{VEVENTITEKNIKAL MALAYSIA MELA} \epsilon = Error \epsilon_{rms} = Root Mean Square \varphi = Accuracy $	Η	🗧 - Height
$m - Slope$ $b - y-intercept$ $\mu - Sample Standard deviation$ $\sum_{i}^{s} i - Error$ $\epsilon - Error$ $\epsilon_{rms} - Root Mean Square$ $\varphi - Accuracy$	ŷ	- Predicted Value
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	m	- //Slope
μ s $Sample Standard deviation$ $\sum_{\epsilon}^{s} i$ ϵ	b	- y-intercept
$Simple Standard deviation$ $Sum = \frac{s}{\epsilon} i$ $First - Error$ $\epsilon_{rms} - Root Mean Square$ $\varphi - Accuracy$	μ	او يومر سيخ بيڪنيڪ ملي Mean مالاك
$\sum_{\substack{\epsilon \\ \epsilon \\ rms}} i UNIVESUMITI TEKNIKAL MALAYSIA MELA \\ \epsilon \\ \epsilon \\ rms \\ \varphi \\ - \text{ Root Mean Square} \\ \epsilon \\ \epsilon \\ \varphi \\ - \text{ Accuracy} $	S	- Sample Standard deviation
$\frac{1}{\epsilon} - \text{Error}$ $\frac{\epsilon_{rms}}{\varphi} - \text{Root Mean Square}$ $\frac{\varphi}{\varphi} - \text{Accuracy}$	\sum_{i}	UNIVE ^{SUM} ITI TEKNIKAL MALAYSIA MELAKA
ϵ - Error ϵ_{rms} - Root Mean Square φ - Accuracy	` ل	CHIVERSTITTERNINAL MALATSIA MELANA
ϵ_{rms} - Root Mean Square φ - Accuracy	ε	- Error
φ - Accuracy	ϵ_{rms}	- Root Mean Square
, ,	arphi	- Accuracy

LIST OF ABBREVIATIONS

V	-	Voltage
BDP	-	Bachelor Degree Project
MATLAB	-	Matrix Laboratory
Ts.	-	Technologist
FAMA	-	Federal Agricultural Marketing Authority of Malaysia
USB	-	Universal Serial Bus
JPG	-	Joint Photographic Experts Group
XGA	-	Extended Graphics Array
CCD	-	Charge-coupled device
PL	-	Polarized light
IC	-	Integrated Circuit
BMP	-	Basic Metabolic Panel
IEEE	-	Institute of Electrical and Electronics Engineers
RGB	- 15	Red, Green, Blue
HSI	1	Hue, Saturation, Intensity
CMYK	Š -	Cyan, Megenta, Yellow and Black
HSV	2 -	Hue, Saturation, and Value
YCbCr	F -	Luminance and Chroma components
2D	El-	2 Dimensional
3D	20	3 Dimensional
MLP	241	Multilayer perceptron
SVM	1.5	Support Vector Machine
CNN	ملاك	Convolute Neural Network
ANN	-	Artificial Neural Network
RMSE		Root Mean Square Error
FPS	UNIVE	Frames per second AL MALAYSIA MELARA
LED	-	Light Emitting Diode
App	-	Application

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	(Gantt Chart) BDP 2	97
Appendix B	(MATLAB Application Interface)	98
Appendix C	(Image Acquisition Setup and Demonstration)	99
Appendix D	(Output Image and Result)	101



CHAPTER 1

INTRODUCTION

1.1 Background

One of the essential fruits produced in Malaysia, mangoes were valued at 126,000 t in 2002, representing a 12% share of total exports. Post-harvest handling is a crucial step in the fruit production process since it helps maintain its quality. Cleaning, grading, and packing are all part of the post-harvest handling process. Fruit grading is necessary for identifying the quality of the fruit as well as its cost. Many local farmers are still heavily reliant on human labor to grade their fruits, following the grade recommendations set out by the Federal Agricultural Marketing Authority of Malaysia (FAMA). The fruits are assessed according to their size, maturity index, and the presence of exterior flaws (Teoh and Syaifudin, 2007).

Computer vision systems may be utilized for automated fruit inspection and grading and other applications. These systems have been extensively employed in the food and agricultural industries for review and evaluation because they offer an appropriate speedy, economical, consistent, and objective assessment method (Sun, 2000). According to (Kanali et al, 1998) the use of machine vision to automate the inspection of products not only results in labor savings, but it also enhances the objectivity of the inspection process. Advancements have spurred research on constructing a system to assess the quality of various and processed meals in hardware and software for digital image processing and analysis during the last decade. (Gerrard et al. 1996; Locht et al. 1997, 1996). Image processing and analysis are at the heart of computer vision, allowing it to perform the necessary categorization and measurement tasks (Krutz et al, 2000). (Nagata et al, 1997) The use of image processing and analysis to sort fresh strawberries based on size and shape was examined. The findings revealed that the system could accurately sort 600 strawberries into three classes of condition and five classes of length with 94–98 percent accuracy.

Size is one of the significant parameters that consumer identifies together with the quality of fruits (Aleixos et al, 2002). According to the Florida Agricultural Marketing Association, weight may affect the size of tropical fruits (i.e. Mango, pineapple, watermelon, papaya and starfruit). Manual weighing of fruits to determine their size is time-consuming, labor-intensive, and expensive. Instead of relying on weight to determine fruit size, image processing and analysis methods in computer vision systems may be utilized to make an automated determination of fruit size. The goal of this project was to estimate the weight of the mangoes and estimate the error to assess the ability of the linear model generated in the linear regression analysis to evaluate the importance of the mangoes, which was the primary purpose.

1.2 Problem Statement

The size of the Mango is one of the most critical factors that the customer has identified as being associated with mango grade. According to the Malaysian Federal Agricultural Marketing Authority (FAMA), the size of Mango is defined by its weight. It is, however, difficult to determine the grade of mango fruits because of the difficulty in collecting the fruit at the proper size. Manual weighing of fruits to determine their size is time-consuming, labor-intensive, and expensive. Furthermore, a computer vision system can automate fruit inspection and grading to achieve the Mango fruit's desired weight. The mango pictures collected from the image capture equipment will be processed and analyzed using the MATLAB program in the computer environment. It is the software's job once image processing and analyzing operations have been completed to determine how many pixels are included inside the mango area of the acquired picture. With the help of the linear regression approach, we will discover a link between mango pixels and mango weights.

1.3 Project Objective

In this study, the primary objectives were to investigate several image processing and analysis methodologies for determining the weight of mangoes, specifically:

- a) To develop a vision-based system for mangoes' weight estimation.
- b) To select an appropriate feature from the image analysis that can be used as a parameter in the regression technique.
- c) To perform testing and analysis in term of the performance of the completed Usystem.RSITI TEKNIKAL MALAYSIA MELAKA

1.4 Scope of Project

In this study, a primitive machine vision system was used to improve mango fruit grading in the country of Malaysia. The methodology is straightforward compared to other state-of-the-art technologies now accessible, but it has limits in analysis and practical use.

 a) Image processing and analysis were performed by filtering and thresholding processes using MATLAB software.

- b) Image is captured using a digital camera and transferred to MATLAB software.
- c) The position of the Mango must remain in a static setting during the capturing process.
- d) Just one Mango should be placed at a time.



CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

An overview of estimating the weight of mangoes using image processing and analysis techniques was covered in Chapter 2. This project aims to create image processing and analysis tools to estimate the weight of mangoes, which will assist local producers in classifying the fruits following the FAMA's suggested classification. It was also done in this project to compare and contrast the advantages and disadvantages of past research findings. It discusses the source and supports the claim with facts from similar research or study known as the literature review.

2.2 Image Acquisition and Preprocessing

Image Acquisition and Preprocessing for Machine Vision Systems is a detailed, in-UNIVERSITITEKNIKAL MALAYSIA MELAKA depth guide text covering any element of image acquisition and preprocessing, from scene lighting to image shaping optics, from CCD and CMOS image capture to image transformation

2.2.1 Digital Camera

(Khairul Adilah and Sharifah Lailee, 2016) Indicate that the method of capturing photographs using a digital camera is known as image acquisition. For this experiment, the distance between the camera and the fruit was adjusted at 30 cm. Until being processed into grayscale files, the captured images were saved in RGB color format. According to (Ganiron,

2014) Venvi Enterprise took all of the digital images, which were then stored later. The Canon Digital IXUS 400 camera was used to take 140 photos of green mangoes at a resolution of 640×480 pixels each. Natural light was used to capture the photographs, which were taken between 9 a.m. and 2 p.m. Before the study, through the Universal Serial Bus (USB) adapter, the photos were transferred to the computer's hard drive. The segmentation process was made more accessible by using a white backdrop. Figure 2.1 shows the image acquisition of the setup.



The mangoes were photographed using digital cameras set up at a 45-centimeter vertical distance from the plane surface on which the Mango was resting and a 40-centimeter horizontal distance from the fruit's core. One Mango was picked as a reference point, and all subsequent photos were measured on the same scale. The data were plotted with the help of the Origin 7G computer program (Spreer and Müller, 2011). The images were manually corrected and evaluated using "Measure 2.1" The computer code was used to generate a polygon around the Mango that was proportionate to the fruit's original size.

Using the top view's from the graph, the values for L and Wmax were computed by taking the difference between the highest and lowest values in each of the x and y directions, respectively. The minimum and maximum Feret diameters were comparable to the oblong

form of the mangoes. Tmax was calculated from the maximum in the y-direction of the displayed graph of the side view Figure 2.2 (Spreer and Müller, 2011).



Figure 2.2 Digital photos of Mango

A digital camera served as the primary picture acquisition device (Canon Inc.'s EOS 550D) to capture high-quality photographs with a resolution of 0.03 mm/pixel and 1200 × 800 pixels resolution. Due to internet limits, fruit photographs were saved in JPG format. Each sample was put in an examination room equipped with a camera and lighting equipment. Figure 2.3 depicts the vision system that was utilized to take the images. The samples were 20 cm away from the camera, which was positioned at a distance of 20 cm. A total of four lamps with two fluorescent tubes were used to provide illumination. With the lens axis and illumination sources at around 45 degrees, the diffuse reflection that produces the hue happened at approximately 45 degrees from the entering light.



Figure 2.3 The setup used to capture the image

Cross-polarization was employed in combination with polarising filters positioned in front of the lights and camera lenses to mitigate the effect of these specular reflections. The fluorescent lights were driven by a high-frequency electronic ballast, which reduced the flickering impact of the other current and provided more constant illumination. Each Mango was photographed from two different angles A and B (Sahu and Potdar, 2021).

2.2.2 Color camera

(Momin *et al.*, 2017) indicated that there had been reports on photographs were captured using an XGA format ¹/₂" Sony CCD ICX205AK color camera with 8-bit gray levels and a C-mount lens. Because of a cuticle coating on the top, fluorescent lamps induce halation (purple color pixels) on the fruit skin. A PL filter was fitted to the camera lens to remove halation. The image acquisition device diagram depicts the camera setup, illumination, sample configuration, and image transfer to the personal computer show in Figure 2.4. The photographs were taken using a camera with a CCD camera with 6 mm focal length and polarized light (PL) filter.



Figure 2.4 Schematic diagram of the machine vision system for image acquisition

During the picture processing, which took place in a secure environment, dust and stray light was prevented. The downward-looking camera was positioned 220 mm above the samples, with a 150 mm \times 150 mm field of view. Between the center of a fruit item and the

lighting plate, a distance of about 200 mm was maintained. For the fluorescent lights illumination panel, a shutter speed of 1/30 s, iris of 1.4, a gain of 1000 dB, and gamma correction of 120 were regarded as appropriate camera specifications. I was using The Imaging Source's IC capture picture acquisition programme version 2.3.382.1796 with a resolution of roughly 0.17 mm pixel1 and a size of 980 880 pixels, photos were obtained at a resolution of about 0.17 mm pixel1 and stored in bitmap (BMP) format. (Momin *et al.*, 2017).

According to (Teoh and Syaifudin, 2007), following that, Figure 2.5 shows an image acquisition system that featured a color camera, a computer equipped with a screen for picture collection and a lighting system for shooting mangoes. The images were shot in a controlled environment with illumination and 22 cm between the camera and the Mango. Every picture was captured at a pixel resolution of 0.3 millimeters.



Figure 2.5 Schematic diagram of the machine vision system for image acquisition for image acquisition

2.2.3 Camera with a video adapter

The suggested approaches begin with a set of fruit photographs that have been collected. The camera connects to the device through video adapters to capture the photo. The IEEE 1394 transmission protocol is utilized for this connection, and the serial port is used. We picked 480x352 pixels as the image resolution since higher resolution allows for a

more comprehensive examination, but boosting the image typically causes processing stress. For the picture acquisition depicted in Figure 2.6, a four white-light 6-watt lights were used as a diffuse front lighting source. (Bermúdez, 2013).



Figure 2.6 Front diffuse illumination scheme

2.3 Image Segmentation

Picture segmentation is the process of different an image into various sections. This component is typically associated with something that humans can readily distinguish and interpret as separate objects. Computers cannot recognize objects intelligently, so many different methods for segmenting images have been created. The segmentation process is dependent on the image's other characteristics. It may be color information used to make histograms, or information about pixels that show borders or borders, or texture data.

2.3.1 Thresholding method

As commented by (Momin *et al.*, 2017), there are three types of picture segmentation algorithms: thresholding, identification of patterns, and deformable models. The goal of threshold segmentation is to discover neighboring pixels with similar attributes that fall inside specific threshold value ranges. There are thresholding algorithms that are hybrid, edge-based, and region-based. Area-based algorithms function by comparing the information of adjacent or related pixels within the same area, such as colour or grey level intensities. A comprehensive analysis of multiple image segmentation algorithms for diverse fruits based on colour, texture, and form has been published. To get more information at the pixel level than greyscale images, this research used a region-based global thresholding colour segmentation technique and looked for pixels with same feature values to the corresponding pixels.

Other colour spaces, such as RGB (Red, Green, and Blue) and HSI (Hue, Saturation, and Intensity), were investigated for thresholding purposes in order to improve the image processing approach. The HSI colour model outperforms RGB and CMYK for several reasons: first, the intensity variable is decoupled from the colour information expressed by the hue and saturation components; second, the hue and saturation components are inextricably linked to how humans perceive colour; and third, the hue value is invariant to changes in light intensity. Due to these qualities, the HSI model is an excellent starting point for developing algorithms for digital image processing. This colour representation was chosen for this investigation since it is the most extensively used in computer vision. The RGB values of the fruit's surface and the photograph's background were identified and then converted to HSI using the provided colour space conversion relationships.

The following measures were used for the RGB to HSI conversion:

- a) In the [0, 255] range, read the RGB image.
- b) Divide each RGB component by 255 to display the images in the [0, 1] range.
- c) HSI components should be measured.
- d) Represent saturation and strength in the [0, 100] and [0, 255] ranges for simplicity.

By viewing the HSI colour values of the mango surface and background, a histogram based on colour details was constructed. The appropriate colour channel was chosen as the threshold limit (T) for segmenting the target's source image from the context based on the Histogram. For transforming a source picture to a binary image, the global threshold segmentation technique is provided as:

$$g(x,y) = \begin{cases} 1, for f(x,y) > T\\ 0, for f(x,y) \le T \end{cases}$$
(2.1)

The threshold image is represented by g(x, y), the source image is represented by f(x, y), and the threshold value is represented by T. All pixels with color information equal to or less than T in the original picture are set to zero and considered background (black). On the other hand, pixels containing color information higher than T are assigned to one and are termed object pixels (white) (Momin *et al.*, 2017).

Using a thresholding procedure to process the background from foreground items is an excellent way to accomplish that goal.

$$p(x,y) = \begin{cases} c(x,y) \text{ if } c(x,y) \ge \alpha \\ 0 \text{ otherwise} \end{cases}$$
(2.2)

The ultimate intensity value of the pixel at coordinate (x, y) will be p(x, y), while the initial intensity value will be c(x, y). The limit γ is defined in various ways, but there are two forms of thresholding in general: global limit and local limit. The adaptive thresholding utilized in this work fits into the latter type, with the limit determined by averaging the pixels in the surrounding region. The bulk of the background will be removed, and the transition between background and foreground items will be accentuated (Vo, Duc Nguyen and Dang, 2019).

After the input photographs have been converted to grayscale, the blurring procedure will be performed to remove as many intensity spikes as feasible without affecting the pictures' more sensitive elements. Once the thresholding method has been completed, the effect of blurring on noise reduction will be apparent. At this moment, all objects and backgrounds will have an intensity value of one, except the transition between two things. Because the shape does not have to be a closed curve, various open loops and lines correspond to the residual noise after thresholding; a contour may be constructed by randomly selecting a pixel of intensity 0 and exploring the area around it. They may, however, be ignored by simply selecting the form with the most pixels on each image. The contours themselves do not need to be closed loops since we will be concentrating on the mangoes' bounding boxes, making blurring and thresholding methods more straightforward (Vo, Duc Nguyen and Dang, 2019).

The mango photographs were taken using a low-cost web camera. RBG (Red, Green, and Blue) color space was used to capture all the images. The image processing algorithm was created in MATLAB. The object must first be identified from the background to extract the required features from a mango photograph. The objective image segmentation algorithms are utilized. The acquired image's RGB color space is converted to the Hue, Saturation, and Value (HSV) color system. Three pictures of the H, S and V channels are shown after the mango image has been transformed from RGB to HSV color space. In all three-channel images, Otsu's Global Thresholding technique is employed to identify acceptable thresholds. The photos are then binarized using the points that have been given to them. Fill a gap in the mango border, the flood fill approach is used, and groups of pixels smaller than 1000 pixels are eliminated. The best results for this segmentation assignment are considered to be obtained by binarizing the saturation channel image with its threshold. The binarization of a saturation channel image may be explained using the equation below (Aung *et al.*, 2020).

$$I_{bin}(i,j) = \begin{cases} 0, I_s(i,j) < Th \\ 1, I_s(i,j) \ge Th \end{cases}$$
(2.3)

Where Th stands for threshold obtained by using Otsu's method, this process is shown in Figure 2.7.



Figure 2.7 Binarization on H, S, and V channel images and result comparison

2.3.2 Interpretation of the form of an image histogram

As commented by (Teoh and Syaifudin, 2007), to distinguish the mango entity from the context of the filtered mango file, image thresholding was used. The filtered mango image can be thresholded using an interpretation of the form of an image histogram. The image histogram is created by graphing the number of pixels with a certain grey level against that value.



Figure 2.8 Bimodal Histogram

An image histogram with two peaks and a valley separated by a threshold is shown in Figure 2.8. (T). A bimodal histogram is a name for this kind of picture histogram. Because the image histograms for all 100 filtered mango photos in this investigation are bimodal, a simple thresholding approach was used to split the filtered mango picture into mango object and context areas. The following is the thresholding algorithm:

$$g(x,y) = \begin{cases} 0, if f(x,y) \le T \\ 1, otherwise \end{cases}$$
(2.4)

Where f(x, y) represents the original filtered mango picture, and g(x, y) represents the picture after thresholding. The letter T stands for the threshold. Object pixels (1) were identified as pixels with grey level values larger than T in the filtered mango picture. On the other hand, background pixels were defined as pixels with grey levels equal to or less than T. (0). The number of pixels in the mango region of the thresholded image was counted for regression analysis using the mango weight (Teoh and Syaifudin, 2007).

2.3.3 The traditional method for image enhancement, de-noising, and edge detection

Adopted a sequence of measures to evaluate the key characteristics of the fruit from the images: preprocessing, weight determination, degree of maturation, and spot calculation. In each level, the processes made use of MATLAB 2011a software capabilities and Statistics Toolbox features. Picture improvement, de-noising, and edge detection are both done using conventional techniques. This stage aims to distinguish the fruit from the background picture. Color segmentation was used to make this distinction possible. Researchers studied several color spaces to build a more accurate way to accomplish this aim, particularly those in which color information can be isolated from the intensity variable, such as HSI (Hue, Saturation, and Intensity), YCbCr (Luminance and Chroma components), color spaces. The YCbCr color system was selected because it allows for more precise segmentation than other color spaces, resulting in a clear distinction of fruit and background colors. Mango fruit is often green to red in hue. The negative meaning portion [-1, 0] of the Cb channel is found in the second and third quadrants of the YCbCr color space, as illustrated in Figure 2.9. When looking at this region, this means that distinguishing the fruit from the surroundings is straightforward (Bermúdez, 2013).



Figure 2.9 Relation between CB and Cr channel in the YCbCr color space

26

Because just the Cb channel information is required, it is unnecessary to execute a complete translation from an RBG image. Instead, the Cb channel information from an RGB image is extracted using the following function for efficiency:

$$b(x,y) = \frac{1}{255} \left(-37.797 * R(x,y) - 74.203 * G + 112 * B(x,y) \right) + 128$$
(2.5)

R(x,y), G(x,y), and B(x,y) are all integers in the range [0 255]. This technique was used to normalize the output of the primary function in the range of 0 to 255. In this case, the negative range in Figure 2.9 is between 0 and 127.



$$Imseg = g(x, y) = \begin{cases} 1, siCb(x, y) < U \\ 0, siCb(x, y) \ge U \end{cases}$$
(2.6)

Each image pixel (x, y) in which Cb(x, y) U is a pixel belonging to the fruit; otherwise, the pixel is part of the context, as shown in Figure 2.10. Using histogram analysis, I calculated the U value from a series of photos and set it to U = 120. It's consistent with prior findings on the Cb channel's fruit variety.



Figure 2.11 Histogram of a threshold segmentation and the resulting image

Features specific to fruit geometry, such as weight and length, may be derived from the resulting picture. However, it is crucial to identify an enhanced vision for estimating the degree of maturation and the study of spots. This image (IM1) is the product of multiplying the

2.3.4 Image blurring
$$R(x, y) = R(x, y) * Imseg(x, y)$$

$$G(x, y) = G(x, y) * Imseg(x, y)$$

$$B(x, y) = B(x, y) * Imseg(x, y)$$
(2.7)

There are evident, generally equally distributed intensity spikes in practically all

RGB images obtained by customers from digital cameras. These mistakes result from mechanical or electronic faults in the camera's lens, mechanical or electronic flaws during analog to digital conversion, and problems that arise during digital to analog conversion. The blurring operation is particularly useful in boosting the accuracy and precision of image processing processes, and by eliminating or reducing picture defects, this feature expands image processing's precision and precision. A square window in the shape of a rectangle with a value of an odd number on one side will traverse, with the color values of the central pixel equaling the result of calculating those values using the values of other pixels according to a specific formula. For the sake of this article, Shall employ Gaussian blurring since it is
simple and easily implementable. Example outcomes of image processing stages in the proposed model are shown in Figures 2.12 to 2.14. (Vo, Duc Nguyen and Dang, 2019).



Figure 2.12 Top profile of mango



Figure 2.14 Threshold with blurring window

2.3.5 Curve fitting

Curve fitting is the process of describing a collection of data points using a mathematical function. In this study, curve fitting describes Mango's upper and lower curves using two polynomial equations. One per each angle, two sets of data points are needed. The mango contour is precisely contained within the picture border once the segmented image is cropped. Each data point in the mango contour is an array of pixel distances calculated from

the bottom edge of the cropped image along the length of the Mango. The vertical distances between the bottom image border and pixels in the top half of the contour, as illustrated in Figure 2.15, are y_{u1} and y_{u2} . The vertical distances between the bottom image border and pixels in the bottom half of the contour are represented by y_{u1} and y_{u2} . These two sets of data points are used to build two polynomial equations for the top and bottom curves of the mango shape. The MATLAB built-in function "polyfit" may be used to find two sixth-order polynomial equation coefficients. The generated curves are in significant agreement with the mango shape, as shown in Figure 2.15 (Aung *et al.*, 2020).



Figure 2.15 Curve fitting the mango boundary with two polynomial equation

اويونرسيتي تيڪنيڪل مليسيا ملاک

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Greyscale image processing is used to transform the raw RGB mango picture to greyscale. Figure 2.16 (a) displays the original RGB picture, whereas Figure 2.16 (b) shows the converted greyscale picture. The preliminary background subtraction serves two purposes. The first step is to remove as many background pixels as possible to identify the coarse mango regions. The second objective is to determine if the mango pixels are darker than the overall background. After preprocessing, the greyscale image is converted to a binary image using the 0.15 threshold value, as shown in Figure 2.16. (c). The remainder of the land is white, except for the mango area, which is black. Finally, when the mango region

is white, figure reversal is implemented, which increases the classifier's performance by reversing the black and white color areas.



Figure 2.16 (a) Original image (b) Grey-scale image (c) Binary image (d) Edge detection

(e) Contour filling (f) Filtere image

The photographs of fruit were separated from the context using this algorithm. As seen in Figure 2.17, the product of this step is a single picture of a single contour of fruit with pixel values of "0" and "1." These numbers are used to extract the fruit's characteristics (Khairul Adilah and Sharifah Lailee, 2016).



Figure 2.17 Pixel value for the segmentation image

2.4 Weight Estimation

The weight in grams of Mango may be calculated using the equation $m = \rho \times v$, where v is the Mango volume, and ρ is the density of Mango. The mangoes' weight and volume were measured and documented. These values, it is feasible to determine the desired value. The formula becomes $\rho = \frac{m}{v}$. The average density and previously collected estimated volume V is used to calculate the Mango's estimated weight or mass.

2.4.1 Estimate weight using volume estimation

Volume estimate may be used to compute the weight of fruit. Volume is the sum of all sections made by a transverse cut along the length of the fruit, and the size of the fruit is illustrated in Figure 2.18. For example, suppose we select a small enough h value (height of each cross-section). Then, the significance of each section will be approximated to the volume of an elliptic cylinder, as follows:



 $Vc = \pi * a * b * h$

(2.8)

Figure 2.18 A cross-section of the fruit

Figure 2.19 shows a ratio study of the semi-axes a and b at different positions along the length of the Mangifera indica L fruit, which leads to the conclusion that the semi-axes values have a stable connection, which is known as Depth Factor (Df):



Figure 2.19 Ratio analysis at different points



Count the number of pixels on each row and divide by two to get the b axis value, representing half of the fruit's length in pixels in this picture segment. The region properties function of the Matlab image toolbox was used to measure the size of the minor axis of each preprocessed picture acquired before finishing the volume estimate procedure. This function may be used to create the MinorAxisLength property, which returns the length in pixels of the minor axis of the ellipse with the same normalized central moments as the item in the image:

$$Minoraxis = regionprops(B, 'MinorAxisLength')$$
(2.11)



Figure 2.20 Major and minor axis on fruit image

It's simple to determine the other axis once have the little axis length for each graphic is shown in Figure 2.20. This axis combination determines the front vision of the fruit. The depth factor was determined by the relationship between the primary and minor axes of a group of pictures Df (Bermúdez, 2013).



D is the average density determined from a set of known-weight and-volume natural fruits. Based on the numbers in Table 2.1, the weight-based categorization approach is used (Bermúdez, 2013).

Caliber	40	28	24	18	15
Weight (g)	<100	101-130	131-160	161-200	>200

Table 2.1 Reference values for weight-based fruit classification

Weight in grams of Mango can be calculated from the equation m equals ρ times V, where ρ is the density of the Mango and V is the volume of the Mango. The actual values of weight and volume of the mangoes have been recorded. With these values, it is possible to evaluate the required value of ρ . The equation becomes ρ equals m divided by V. The average value of ρ for 24 mangoes samples is found to be 0.98g/cm3. The estimated weight or mass of the Mango is the product of average density ρ and previously acquired estimated volume V (Aung *et al.*, 2020).

2.4.2 Estimating weight using linear regression analysis

The number of pixels in the mango region of the thresholded image was counted for regression analysis using the mango weight. The Mango 2-dimensional (2D) area of the thresholded picture and mango weight was studied in linear regression analysis. The 2D location is determined by the number of pixels in the mango region of the thresholded image. In the regression study, a linear model was built to predict the weight of the Mango (dependent variable) from the counted mango pixels (independent variable). A 95 percent confidence and prediction intervals graph is used to verify the linear model and demonstrate that it properly fits the findings. (Teoh and Syaifudin, 2007).

The confidence region was obtained using the formula: (Estimate) $\pm t * S(D)^{1/2}$, while the prediction region was calculated by the formula: (Estimate) $\pm t * S(1 + D)^{1/2}S$ is the standard deviation (estimated by mean square error), t is the t-value from the Student's t-distribution, which depends on confidence and degree of independence. D is the regression line's value. The distinction between the two formulas is that one is applied to the quantity under the square root sign in the forecast field estimation with a coefficient of correlation (r2) of 0.9769. This coefficient of correlation indicated that a good dependency exists between counted pixels and weight (Teoh and Syaifudin, 2007). The regression model describing this dependency is:

$$Weight = 0.0029 * Counted Pixels - 17.084$$

$$(2.14)$$

The 95% confidence and forecast intervals indicate that the weight prediction has a high level of confidence. The projection bands are more comprehensive than the associated because the linear model predicts the value of a random variable rather than computing a parameter. The findings of statistical analysis must meet such data assumptions to be considered accurate. Wherever the linear model is supposed to be meaningful, the residuals must be independent and normally distributed with the same variance. The breach of assumptions was detected using a plot of residuals against different values. The disparity between the actual and projected weights was used to calculate the residual. It increased the likelihood of the data being distributed normally.

As a consequence, it implies that expectations have been met. Outliers or an incorrect problem may also be observed using the map. The plot indicated no very high residuals (and hence no apparent outliers), and no pattern in the residuals implies the linear model is unsuitable. As a result, image processing and measurement methods can be used to measure the weight of a mango with a high degree of precision (Teoh and Syaifudin, 2007).

2.4.3 Estimate weight using neural network

The length, width, and height of the mangoes assuming they are confined inside minimum volume bounding boxes should be L, W, and T in millimeters, respectively. The weight and measurements of 92 Hoa Loc mango samples were collected at random from different retailers. L, W, and T are then utilized to extract other parameters for further analysis.

Table 2.2	Chosen	parameter.
-----------	--------	------------

Base Parameters	Derived Parameters			
L	L^2	L ³		
W	W ²	<i>W</i> ³		
Т	T^2	T^3		
	$L \times W$	$W \times T$		
	$T \times L$	$L \times W \times T$		

These parameters are part of a set of geometric dimensions that relate to the Mango's form and, by extension, its weight. A basic multilayer perceptron (MLP) with one hidden layer of 10 neurons will estimate the weight using all 13 geometric attributes. Olden and Jackson's connection weight methodology may be used to calculate and rank the relative significance and contribution of the input (Olden & Jackson, 2002). The assistance and relative contribution for an MLP with one hidden layer is calculated using the following formulae after normalizing the input and output:

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

$$C_{i,k} = W_{i,j} \times W_{j,k}$$
(2.15)

$$R_i = \sum_{k \in b} \frac{|c_{i,k}|}{\sum_{i \in a} |C_{i,k}|}$$
(2.16)

The input and hidden layers are designated by A and b, respectively, while a neuron of each layer is defined by I or j, while the output neuron is represented by k. Thus, C (i,k) denotes input neuron I's contribution to hidden neuron k, whereas R I denotes its relative contribution to the network. The validity of this 3method compared to other methods to this issue was shown in (Olden et al., 2004), where (Olden and Jackson, 2002) is shown to have the best accuracy and precision when tested using simulated data.

As you can see, tests with simulation were used to compare the performance of all the approaches mentioned above. However, this strategy will not function as well with the mangoes data. Instead, re-estimate each input.

2.4.4 Estimating weight using visible imaging

In this project, the volume of Mango is estimated using the Cylinder Approximation analytical methodology. In this case, the method employed was a simple circular volume technique. In this study, more than a thousand Harumanis mangoes were used. The value of h and the radius of the fruit, r, were calculated using photographs of mangoes. Figure 2.21 shows how the cylinder approach was used to redesign the cylindrical component of the



Figure 2.21 Mango dimension measurement

The volume is determined using the cylinder volume formula, in which the width, W, and length, L are taken from the mango sample that was measured:

$$V = \pi r^2 h \tag{2.17}$$

$$r = \frac{W}{2} \tag{2.18}$$

$$h = \frac{L}{2} \tag{2.19}$$

The scatter plot showed the association between the actual weight and the estimated volume of mangoes utilized the accurate weight throughout. So, here's how it can get the mangoes' weighted mass: After that, using the best scatter plot between the actual weight and volume, which is computed in a different method for each measurement, you may acquire the weight mangoes. Then, the equation below was used

$$y = mx + c$$

$$x = \frac{y - c}{m}$$
(2.20)
(2.21)
(2.21)
(2.21)
(2.21)

Table 2.3 Comparison of previous studies.

2.5

No	Title/Author	Problem	Solution/Method
1.	Image processing	One of the most important	The mango picture obtained
	and analysis	factors that consumers associate	from the image capture
	techniques for	with mango quality is its size.	equipment was processed and
	estimating the	Fruit grading is necessary for	analyzed using the computer
	weight of	identifying fruit quality as well	PCI program. The application
	Chokanan	as price. Many local producers	counted the number of pixels
	mangoes (Teoh	still depend on human labor to	in the mango area of the
		grade their crops according to	acquired image after

	and Syaifudin,	Malaysia's FAMA (Federal	processing and analysis. The
	2007)	Agricultural Marketing	relationship between mango
		Authority) grading guidelines.	pixels and mango weights was
		The size, maturity index, and	investigated using a statistical
		outward faults of the fruit are	regression methodology.
		used to evaluate the fruits.	
2.	Automatic	Fruit farmers have a difficult	Image capture, picture
	Preharvest	time harvesting the fruit at the	preprocessing, image
	Grading of	proper size. One of the most	segmentation, feature
	Harumanis	crucial characteristics of	extraction, and classification
	Fruits, (Khairul	Harumanis fruit grading is the	were all part of the computer
	Adilah and	size of the fruit. On the other	vision approach used in this
	Sharifah Lailee,	hand, fruit farmers have a	study. According to the
	2016)	difficult time picking the fruit at	statistical research, the size
		the proper size. This study aims	has a strong association with
	TEL	to see whether image	the Harumanis' weights and
	LIS	processing methods can be used	grades, which employed a
	NILLE &	to assess Harumanis fruits	linear regression model.
	sMal.	before they are harvested.	laint with
3.	Geometry-based	Mango (Mangifera indica) is an	Researchers developed an
	mass grading of	essential and popular fruit in	image collection and
	mango fruits	Bangladesh. However, most	processing system that
	using image	post-harvest processing is still	extracts projected area,
	processing	done by hand, which is	perimeter, and roundness
	(Momin et al.,	inefficient in terms of accuracy	parameters. Images were
	2017)	and throughput.	obtained using an XGA
			format color camera with 8-bit
			gray levels and fluorescent
			lighting. In addition,
			researchers developed a
1			method to divide mangos into
			method to divide mangos into

4.	Automatic visual	In the agricultural business, the	The classification process
	model for	physical attributes of fruits are	follows the Norma Técnica
	classification and	the most critical information in	Colombiana (Colombian
	measurement of	determining quality for	Technical Norm) NTC 5139
	quality of fruit:	operations such as exporting.	standard. It involves using
	case Mangifera		Principal Components
	indica L,		Analysis and an ellipsoidal 3-
	(Atencio,		D model of the fruit.
	Sánchez T and		
	Branch, 2009)		
5.	Using machine	Mangoes and other fruits are	This research proposes a
	learning to grade	currently classed based on	unique method of assessing
	the Mango's	human perceptions of low yield,	interior quality based on
	quality based on	which is a poor classification	Mango's outward
	external features	quality.	characteristics and weight.
	captured by the		The fruit grading is
	vision system		implemented using four
	(Long and Thinh,		machine learning models.
	2020)	la 15:5:	Data standardization and
	-/~ 5		outlier removal result in a
	UNIVER	SITI TEKNIKAL MALAYS	better-equipped dataset for
			machine learning techniques.
6.	Image analysis	To fulfill export regulations,	In this research, the
	for the automatic	fruit attributes such as weight,	methodology for predicting
	feature of the	maturity level, and certain spots	these properties from a set of
	Maginfera indica	must be determined.	pictures taken of mango fruits
	fruit (Bermúdez,		is provided. The weight
	2013)		measuring methodology had
			an error margin of fewer than
			6 grams. Techniques for
			determining the degree of
			development and the group of

			spots had an overall accuracy
			of more than 90%.
7.	Shape and	Fruit growers have a hard time	This paper presents research
	weight grading	picking fruit that is the right	on the use of visible imaging
	of mangoes using	size. Therefore, one of the most	as a method for grading
	visible imaging	crucial characteristics of	mangoes. A Fourier-
	(Sa et al., 2015)	Harumanis fruit grading is the	descriptor methodology was
		size of the fruit. On the other	utilized to analyze the
		hand, fruit producers have a	mangoes based on their
		hard time getting the right size	shapes. The forecast's high
		of fruit. This study aims to see	accuracy illustrates that this
		whether image processing	simple technique can
	MALA	methods can be utilized to	accurately anticipate fruit
		classify Harumanis fruit grades	weight and volume.
	TEK	before they are harvested.	
8.	Weight	The Mango (Mangifera Indica)	In this study, a total of 24
	estimation of	grows abundantly in various	mangoes were used as
	Mango from a	parts of Myanmar. Many	samples. Just a single web
	single visible	farmers grow various mango	camera is used to record the
	fruit surface	varieties for profit, and some	Mango's top perspective. The
	using computer	are in high demand for export.	visible surface area is
	vision (Aung et	The grading technique is	calculated using curve fitting
	al., 2020)	currently carried out manually.	and the area between the two
		However, image processing	curves method. The thickness,
		techniques and computer vision	which is not visible in the top
		technologies have progressed to	view shot, is significant for
		the point where it is now	estimating weight and
		feasible to automate.	volume.
9.	Analysis on	Most fruit species are assessed	Various weight estimate
	mangoes weight	in the food sector based on	methods and approaches were
	estimation	various criteria, one of which is	performed utilizing 2-D and 3-
	problem using	weight. The majority of weight	D image analysis tools. In

	neural network	estimating techniques may be	Vietnam, using these methods
	(Dang <i>et al.</i> ,	broken down into two	can offer significant
	2019)	categories. The first is a direct	advantages to the agricultural
		estimate using image	sector by automating the
		processing measures, which is	process of sorting produce to
		shown in this article.	other nations.
10.	Estimating the	For export, mangoes must have	The length, maximum
	mass of mango	a homogeneous fruit mass and	breadth, and maximum
	fruit from its	shape to make handling and	thickness of over 1000 mango
	geometric	shipping easier, as well as to	'Chok Anan' fruits were all
	dimensions by	match client preferences.	measured. Photographs of 30
	optical	Because of its irregular shape,	fruits were shot using digital
	measurement,	which cannot be approximated	cameras, manually corrected,
	(Spreer and	using typical geometrical	then evaluated using a
	Müller, 2011)	figures, the Mango is more	computer algorithm. To
	TEX	challenging to model than	compute the mass of mangoes
	E	round or oval-shaped fruits.	based on three geometric
	S'ANINO		dimensions, an equation was
	abl (created, which subsequently
	يا ملاك) بیکسیکل ملیسہ	demonstrated a high degree of
	UNIVERS	SITI TEKNIKAL MALAYS	explanatory power.

2.6 Summary

As a conclusion to this chapter, numerous strategies may be adopted in building this image processing and analysis methods for estimating the weight of mangoes, based on previous studies of this project and papers and journals linked to this project. This is due to the development of the method or approach itself, which has increased the characteristics that aid in completing the project. This research was covered in this chapter under image processing and analysis methodologies for calculating mango weight. Following that, based on the research conducted in this chapter, it is evident that some research has benefits over others, while others need improvement to be a better application.

Finally, conclude in this chapter that image processing and pixel modification are essential for creating a successful detection system for determining the number of pixels in each Mango. After performing linear regression between the actual weight and the number of pixels, the correlation equation can determine the estimated weight. The percentage error between the actual weight and the projected weight must then be calculated. After that, select the method's accuracy. Finally, the mangoes must be sorted into grades.



CHAPTER 3

METHODOLOGY

3.1 Introduction

Following the theoretical explanations explored in the previous chapter, this chapter will discuss the practical implementation and analysis of image processing and analysis techniques for estimating mango weight, including how to recognize image processing capabilities, and how to avoid mistakes during the estimation process. Finally, this chapter will focus on the process and techniques for building the actual project, which will involve software and hardware development and any issues that may emerge during the testing phase.

This methodology chapter is the most important in ensuring that the project's objectives are accomplished. Project implementation, which will discuss the various models available, project development, which will include a flowchart depicting the overall operation of the application, and finally, the requirement analysis phase, which is considered the project development flow, are the components discussed in this chapter. Finally, this chapter covers all of the procedures necessary to complete the project, from project implementation through development and documentation.

3.2 Project Implementation

The sequence of the technique that will be employed will be shown in the diagram throughout this phase of project implementation. The stages of development for this project

are shown in the chart below. Figure 3.1 depicts the processes involved in the creation of this project.

a) Literature Review

A literature review is a selective examination of existing research relevant to the image processing and analysis techniques for estimating mango weight, demonstrating how it connects to the investigations. It is necessary to research journals to determine the best method for the project by comparing all of the options.

b) Project Design

ALAYS

The general structure and design of the image processing and analysis methodologies for estimated mango weight are mainly determined during the project design phase. The identification of design objectives is the initial step in this phase. The following things are required:

- a. The types of camera.
- b. The position of the camera is placed according to the order of the picture taken.
 - c. Places that are less distracted from sunlight cause the picture to be blurry.
- c) Develop Image Acquisition

The essential item to prioritize during this phase is image processing. All images are captured in the image acquisition process. MATLAB was used to create the image processing algorithm. The item must first be separated from the background picture before extracting essential characteristics from a mango picture. In this section, two methods will be used, namely area and perimeter. The purpose of these two methods is used to obtain the value of the r-square. Which method is closer to one when making linear regression as an analysis technique.

d) Programming

The programming includes the programs utilized to extract the area and perimeter values of the mangoes, while MATLAB is used to construct image processing techniques. However, there are a few commands to use before getting a good image (no interference). Using Matlab commands like rgb2gray, 'graythresh', 'imbinarize', 'imfill', 'imclearborder', 'strel', and 'imopen', the process is completed successfully. More information will be provided in the section on project development.

e) Functionality Test

This stage is critical to this project's success. Must first do a functionality test to detect the Mango using an image analysis method, after which, techniques analysis was performed to obtain an estimate of mango weight based on the area values and parameters that had been obtained may display the estimated weight. If this phase is not successful, must return to develop an image analysis process before comparing the actual weight and pixels number into the correlation equation.

f) Collect Data

The data gathered should be used to determine the picture capturing system's accuracy. In this approach, data can be ranked by stability and how fast

and reliable MATLAB image division is. The area and perimeter data from picture segmentation will be saved to obtain the linear regression line and rsquare value. The correlation equation can be used to calculate all of the data from the capture and estimated weight. It will be selected which r-square is the finest.

g) Analysis and Discussion

The data will then be examined to check if there are any errors or mistakes. Then, to overcome the problem, a new approach will be devised. Improvements for future projects should also be suggested.

h) Documentation

The result will be compiled and saved into documentation complete reports.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA



Figure 3.1 Flowchart of project implementation

3.3 **Project Development**

To guarantee that the project is completed effectively and following the specified objectives, a thorough grasp of the project is essential in this research. Furthermore, to ensure the proper use of the hardware and software components, a thorough knowledge of the hardware and software employed in this project was required. The appropriate selection of parts will influence the project's development since it will aid in the project's smooth development.

3.3.1 Sampling ALAYSIA

Employees in the agricultural and agro-based sectors may benefit from this technique, especially when assessing the size of a mango fruit based on its weight. The flowchart in Figure 3.15 depicts the first way for sampling the mangoes. At the Mydin Wholesale Hypermarket, Melaka International Trade Center, a total of ten random samples of physiologically and green mature mangoes were purchased. Digital scales are used to weigh fruits. The samples were divided into two groups: a training set of ten samples and a test set of eleven or more mang. A linear regression model for weight estimation was developed using 10 mango samples from the training set. Another sample of 11 and above was used to evaluate the linear regression model of the training set, referred to as the test set application. Finally, the mango training set can also be used as a test set application. This is because the method to estimate the weight of mango uses statistical analysis.

Although the mango had been in the training set to obtain a linear regression line, there was also an error even though it had been made into 10 samples selected for the training set. Unless the project uses a Support Vector Machine (SVM), Convoluted Neural Network (CNN), and Artificial Neural Network (ANN), the sample can no longer be used in test set applications due to errors for all samples used in this method will be blank percent. To enhance our method, utilize machine learning to grade this Mango automatically in the future. Some of the approaches that may be employed are Support Vector Machine (SVM), Convolution Neural Network (CNN), and Artificial Neural Network (ANN).

3.3.2 Image acquisition and pre-processing

All the mangoes samples were immersed in water and cleaned using a soft sponge. Following that the fruits were exposed to an image acquisition system as shown in Figure 3.2 below consisting of a web camera, a laptop equipped with an image acquisition box, and a lighting system for capturing mangoes images. The image was captured in a hall with a proper control environment such as lighting and the height of 22 cm from camera to mangoes was fixed. The image captured was saved in RGB color format before they are converted into grayscale.

In MATLAB software, the webcam list command that had been used in this section. Webcamelist will return a list of all webcams connected to the system (laptop). In this section also command webcam had been used to create a webcam object to acquire frames from the laptop. Webcam command has three methods which are a snapshot, preview, and close preview. Finally, snap shoot will acquire a single frame from the webcam. While for the preview command, will activate a live image preview window as shown in Figure 3.16 below.



Figure 3.2: Image acquisition system

3.3.3 Image processing and analysis

Mango images are captured by using a low-cost web camera as shown in Figure 3.3 below. All images are acquired in Red, Green, and Blue (RGB) color space. The image processing algorithm is developed using MATLAB. The method for this section was shown in Figure 3.17 below.



Figure 3.3: RGB image or input image

3.3.3.1 Grayscale image

To extract required features from a mango image, the mango object must be first differentiated from the image background. For this task, the rgb2gray command in MATLAB was used to convert an RGB image to a grayscale image before proceeding to the next process. The function rgb2gray is to convert RGB image or colormap to grayscale. Next, rgb2gray converts RGB image to grayscale by eliminating the hue and saturation information while retaining the luminance. Last but not least, rgb2gray converts RGB values to grayscale values by forming a weighted sum of the R, G, and B components. Example:

 $0.2989 \times R + 0.5870 \times G + 0.1140 \times B$

3.3.3.2 Graythresh image

By using the command graythresh in MATLAB, can get the value of the thresholded image as shown in Figure 3.4 below. For this task, the input image had been converted from RGB color space to gray image color space. After converting the mangoes image from RGB to gray images, the image is displayed. Otsu's Global Thresholding method is applied in all input images to identify respective thresholds. Binarization is then carried out on each image with their respective thresholds.

The graythresh command in MATLAB is global image thresholded using Otsu's method. It computes a global threshold level that can be used to convert an intensity image to a binary image with binarize. The level is a normalized intensity value that lies in the range [0, 1]. Graythresh uses Otsu's method, which chooses the threshold to minimize the intraclass variance of thresholded black and white pixels. All images of mangoes must be

inverted to white color and the background is black color. Lastly, the value of thresholded was shown at the top of the image by using the streat command in MATLAB.



Figure 3.4: Binarize and thresholded image

3.3.3.3 Remove noise

However, using only the greythresh command can not help to get the desired image in a good way due to interference from the noise around the image border. This is because there is a source of lighting that disturbs the weather in the area of the image border fruit. The way to solve this is to use some commands from MATLAB to remove any noise from the image to get a quality image and make it easier for the next process. Among the commands that have been used are, imfill, imclearborder, strel, and imopen.

The imfill command was used to Fill image regions and holes as shown in Figure 3.5 below. In this output, we cannot see the changes in the fill of the hole of the white region because the image does not have any hole in the region area. That command can fill holes in the input image. The use of this function file (hole closer) is to close the hole in the middle of the white region of the image. A hole is a set of background pixels that cannot be reached by filling in the background from the edge of the image. The input image can be numeric or

logical, and it must be real and nonsparse. It can have any dimension. The output image has the same class as the input image.

Next, imclearborder was used to suppress light structures connected to the image border as shown in Figure 3.6 below. Suppresses structures that are lighter than their surroundings and that are connected to the image border. The input image can be an intensity or binary image. The output image is intensity or binary, respectively. For intensity images, imclearboarder tends to reduce the overall intensity level in addition to suppressing border structures.

To remove the blob that is smaller than 20 pixels across was using strel command as shown in Figure 3.7 below. Strel creates a flat disk-shaped structuring element with the specified radius R. R must be a non-negative integer. This command can be petrified during image processing analysis by placing the value of the blob that needs to be filtered to get the output image without any noise.

Last but not least, by using imopen command to open the Morphologically image as shown in Figure 3.8 below. It performs morphological opening on the grayscale or binary image with the structuring element. A structuring element must be a single structuring element object, as opposed to an array of objects.



Figure 3.5: Fill holes image



Figure 3.6: Remove blob on the border-image



Figure 3.8: Output image

3.3.3.4 Finding area

The area is the number of pixels in the region as shown in Figure 3.9 below.

The command used to extract the area value from the binary image is bwlabel and region props. Command bwlabel. it labeled connected in a 2-D image. It can return a matrix L, of the same size as the Binary image, containing labels for the connectors component in the Binary image. The elements of 'L' are integer values greater than or equal to 0. The pixels labeled 1 make up one object, the pixels labeled 2 make up a second object, and so on.

The region props is to measure properties of image regions. It measures a set of properties for each labeled region in the label matrix L. L can be numeric or categorical. When L is numeric, positive integer elements of L correspond to a different region. In the regionprop command, there are various properties as shown in Figure 3.10 below. Just by entering the desired properties, the MATLAB software will display the required value. To get the area value, by using the propregion along with the 'area' properties the value will be displayed based on the input image that has been defined in the algorithm.



Figure 3.9: Number of white region area

```
>> s = regionprops(open, 'All')
 s =
   struct with fields:
                Area: 32849
            Centroid: [137.1356 145.0821]
         BoundingBox: [50.5000 36.5000 174 213]
         SubarrayIdx: {[1×213 double] [1×174 double]}
     MajorAxisLength: 231.2045
     MinorAxisLength: 183.8761
        Eccentricity: 0.6062
         Orientation: -84.9742
          ConvexHull: [265×2 double]
         ConvexImage: [213×174 logical]
          ConvexArea: 33483
               Image: [213×174 logical]
         FilledImage: [213×174 logical]
          FilledArea: 32849
         EulerNumber: 1
             Extrema: [8×2 double]
       EquivDiameter: 204.5107
            Solidity: 0.9811
              Extent: 0.8863
        PixelIdxList: [32849×1 double]
MALAYS/ PixelList: [32849×1 double]
           Perimeter: 677.8700
```

Figure 3.10: The properties of region prop

3.3.3.5 Finding perimeter

Perimeter is the length of its boundary. For the perimeter, the same method is used when finding the area value. By selecting the properties of regionpros 'Perimeter' the value will be calculated in MATLAB software and display value as shown in Figure 3.11 below. However, to get a good value is from a good image input. Finally in image processing and analysis, the overall process is very helpful to obtain both the required values of area and perimeter.



Figure 3.11: Number of perimeters

3.3.4 Statistical analysis

A linear regression analysis was used to analyze and estimate the weight of mango by using the feature obtained from the image analysis was describe before. The relationship between mango 2-dimension (2D) area with perimeter thresholded image and mango weight. The 2D area is determined by the number of mango region pixels in the thresholded image. While for 2D perimeter was determined by the length of its boundary in the thresholded image. These two types of features such as area and perimeter will be used as the parameter for the analysis as shown in Figure 3.18.

3.3.4.1 Correlation coefficient

Following the image processing of the fruit, two types of features of areas and perimeters will be referred to generate to calculate a correlation coefficient using Equation 3.1 as shown below. A relationship between two variables, the data can be represented by ordered pairs (x, y). x is the independent variable and y is the dependent variable. A scatter plot also can be used to determine whether a linear correlation exists between two variables.

A measure of the strength and the direction of a linear relationship between two variables. The symbol r represents the sample of a mangoes correlation coefficient. The weight of the mangoes is the dependent variable in this research, while two types of features such as area and perimeter of each mango are the independent variable and n is the number of 10 samples of mangoes that were used in the training set.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$
(3.1)

$$r^2 = (r)^2 (3.2)$$

3.3.4.2 Linear regression

In regression analysis, a linear regression line was developed to predict the weight (dependent variables) of mango from the counted mango pixel (independent variable) that was determined by using two types of features such as area and perimeter. After verifying that the linear correlation between two variables is significant, next determine the equation of the line that best models the data (regression line). The data can be used to predict the value of y (estimated weight) for a given value of x (actual weight). The regression line for which the sum of the square of the residual is a minimum. The equation of a regression line for an independent variable x and a dependent variable y as is shown in Equation 3.3 below:

$$\hat{y} = mx + b \tag{3.3}$$

$$m = \frac{n\sum xy - (\sum x)(y)}{n\sum x^2 - (\sum x)^2}$$
(3.4)

$$b = \bar{y} - m\bar{x} = \frac{\sum y}{n} - m\frac{\sum x}{n}$$
(3.5)

Where \hat{y} is predicted y-value for a given x-value, m is the slope, and b is the yintercept. Next from Equation 3.5 above \bar{y} is the mean of the y-values in the data, \bar{x} is the mean of the x-values in the data and the regression line always passes through the point (\bar{x}, \bar{y}) . To plot the regression line, use any two x-value within the range of the data and calculate the corresponding y-value from the regression line. From the regression line, the coefficient of determination can be identified. The coefficient of determination is the ratio of the explained variation to the total variation denoted by r^2 . In this project, r-squared is utilized as the main approach for estimating mango weight based on the value of the rsquared method that is closest to one. The concept of combining two ways that have been developed can assist in determining which method is the most appropriate to be supported by evidence.

To better understand the concept, two examples have been made as a way to make a statistical analysis method. Tables 3.1 and 3.3 are shown below, the example value was used to do some example calculations. Before that obtaining the value of the correlation coefficient (r). It describes how strong the linear relationship is. For an example of 1 means, a perfect positive relationship and a value of zero means no relationship at all. Next, scatter a plot data of dependent and independent variables to see the regression line of each feature involved in this project method which is area and perimeter. The r-square value obtained from the linear regression line, after that already guess which features have a coefficient of determination (r-square) a good dependency after the independent and dependent variables were plotted using scatter plots. Instead of linear regression line as well, from scatter, the plot can also extract linear regression equation. The estimated weight can be calculated by using the equation. The correlation coefficients indicate the dependence exists between the calculated pixels from the two methods used in this project. The regression model describing this dependency is:

$$\hat{y} = mx + b \tag{3.6}$$

Based on the calculation example below as shown in Figure 3.14 below, we can see the value of the first r-square example is closer to the value of one (0.9878) compared to the second calculation example is farther than the value of one (0.873). For more clarity, Figures 3.12 and 3.13 are shown below scatter plots for example one for the three-sample data approaching the linear regression line. Meanwhile, in the scatter plot for the second example, the three data samples are slightly away from the linear regression line. In conclusion, the two methods used in this project are very helpful to users to identify which method can estimate the weight of mango with less percentage error of the fruit.

Example 1:

Table 3.1: Example 1 calculating the correlation coefficient

	Wh Nn				
No.	J all	y le	xy.		y ² .
1.	1.6	428.2	685.12	2.56	183355.24
2.	3.6	828.8	2983.68	12.96	686909.44
3.	4.9	1214.2	5949.58	24.01	1474281.64
	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
	= 10.1	= 2471.2	= 9618.38	= 39.53	= 2344546.32

$$r = \frac{3(9618.38) - (10.1)(2471.2)}{\sqrt{3(39.53) - 10.1^2}\sqrt{3(2344546.32) - 2471.2^2}}$$

r = 0.99387936

No.	x	у	xy	<i>x</i> ²	y^2
1.	1.6	428.2	685.12	2.56	183355.24
2.	3.6	828.8	2983.68	12.96	686909.44
3.	4.9	1214.2	5949.58	24.01	1474281.64
	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
	= 10.1	= 2471.2	= 9618.38	= 39.53	= 2344546.32

Table 3.2: Example 1 Calculating the Equation of Regression Line



Example 2:

Table 3.3: Example 2 calculating the correlation coefficient

No.	x	У	xy	<i>x</i> ²	y^2
1.	1.6	520	685.12	2.56	270400
2.	3.6	750	2983.68	12.96	562500
3.	4.9	1325	5949.58	24.01	1755625

$$\sum_{x} x \sum_{y} y \sum_{xy} xy \sum_{x^2} x^2 \sum_{y^2} y^2$$

= 10.1 = 2595 = 10024.5 = 39.53 = 2588525

$$r = \frac{3(10024.5) - (10.1)(2595)}{\sqrt{3(39.53) - 10.1^2}\sqrt{3(2588525) - 2595^2}}$$
$$r = 0.9343287507$$

Table 3.4: Example 2 Calculating the Equation of Regression Line

No.	x	у	xy	<i>x</i> ²	y^2
1.	1.6	520	685.12	2.56	270400
2.	3.6	750	2983.68	12.96	562500
3.	4.9	1325	5949.58	24.01	1755625
83)	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
للك	= 10.1	= 2595	= 10024.5	= 39.53	= 2588525
UNIV	ERSIT	I TEKNIK	CAL MALA	YSIA M	IELAKA

 $m = \frac{3(10024.5) - (10.1)(2595)}{3(39.53) - (10.1^2)}$

m = 233.0518697

$$b = \frac{2595}{3} - (233.0518697) \times \frac{10.1}{3}$$

b = 80.39203868

 $\hat{y} = 233.0518697 \, x + 80.39203868$

 $r^2 = (0.9343287507)^2$

 $r^2 = 0.8729702144$


Figure 3.12: Sketched regression line example 1



Figure 3.13 Sketched regression line example 2



Figure 3.14: Comparison linear regression line between example 1 and 2

3.3.5 Performance evaluation

Last but not least, following the value of the estimated weight of the Mango obtained from the correlation equation, the error between actual and estimated weight must be calculated. To establish if our technique is successful, we utilize the Root Mean Square Error (RMSE) technique to examine system performance. To obtain a high performance out of the system, one method is to estimate the weight with more than 95% accuracy. The accuracy percentage is calculated using the following equations:

$$\epsilon(\%) = \frac{Actual \ weight - Estimated \ weight}{Estimated \ weight} \times 100\%$$
(3.7)





Figure 3.17: Flowchart of image processing 68



This project uses very little hardware in terms of electronics and mechanical parts because it concentrates on image processing. However, a few things are required to make this project a success. Web camera with stand, mini studio box, scales, and mangoes are among them. All the things were acquired from the online shopping website. Finally, the overall cost of this project does not exceed the amount that the faculty allows for project item purchases to be claimed.

3.4.1 Web camera with stand

The camera used is an ANBIUX USB Web camera. The high definition 1080 pixels camera brings true-color image and frame rate up to 30FPS will providing a smooth picture for image segmentation as shown in Figure 3.19 below. The selection of a quality camera is very important to help facilitate the image segmentation process. The desktop webcam features a driver-free design. Just need to plug and play without any external driver required to use this camera.

This webcam stand is used to help hold the camera during image segmentation as shown in Figure 3.20 below. The web camera is placed at the top of the mini-studio box to capture pictures of mangoes from the top view side. Finally, the stand used is very flexible easy to position the web camera in the most optimal position for taking pictures of mango.



Figure 3.19: ANBIUX USB webcam



Figure 3.20: Webcam stand

3.4.2 Mini studio box

Mini studio box is an alternative method that was used to succeed in this project as shown in Figure 3.21 below. Good lighting is needed to help the process of taking a good picture. Good picture taking is very important to reduce the process during picture processing is carried out. This is because, the more noise from the image taken, the more process filtering that needs to be done to ensure that the output image is in good condition to continue the next process. Finally, this mini studio box is equipped with a light-emitting diode (LED) to brighten the area of taking pictures of mangoes.



Figure 3.21: Mini studio box

3.4.3 Scales

Mini scales are used to obtain the actual weight of mangoes as shown in Figure 3.22 below. The actual weight of mangoes is one of the main procedures for this project. Among its uses is the actual value of the weight of the mango used during the training of 10 samples of mangoes. This is because the sample training process has two values that are required. The value is a dependent variable and an independent variable. The dependent variable is the weight of the mango fruit sample. While independent variable is one of the pixel values obtained during image analysis carried out from the two methods used. Finally, the actual weight value of mangoes is required during system performance testing, which is to obtain the percentage error value of mangoes. The RMSE formula will be used to obtain the value

of the percentage error.



Figure 3.22: Mini scale

3.4.4 Mango

Various types of mangoes are used in this project, one of them as shown in Figure 3.23 below. This project does not identify the picture through color differences but identifies the picture based on the size of the mango. Therefore, various types of mangoes can be used during the success of this project. This can make it easier to conduct experiments when you can use a variety of mangoes without waiting for the fruit to ripen. However, the mangoes were selected randomly without sorting according to weight grade based on the grade set by FAMA. Mango fruit is classified based on the size of the fruit. Due to a lack of information, fruit classification was not carried out for this project. Finally, the selected mangoes should be wiped with a dry tissue or cloth to ensure that the fruit is dry from any water on the surface of the fruit. This is because the water on the surface of the mango fruit can interfere during the image segmentation process due to the reflection of light on the water on the surface of

the fruit.



Figure 3.23: Mango

3.5 Summary

The approach for implementing 'Image processing and analysis techniques for estimating mango weight' is described and explained in this chapter. The most crucial chapter in project management is the project methodology, ensuring that the project can be finished systematically using the proper project techniques. The development stage includes understanding, choosing, processing, and efficiently analyzing the topic's parameters to minimize errors. Doing research and planning structure, building a project system, defining method and component, and concluding project integration are the four steps of methodology that lead the developer.

Control parameter was established in the project's construction in the design of the project structure plan based on prior research and literature study. Later on, the project's development was enhanced. The most crucial portion of this stage is analyzing and identifying all of the project's components and control elements. The structure of the integration of software and hardware processes is planned at the create project system stage. The determining project is a phase in the mechanical and software design. The material comprises a block diagram of the software (MATLAB) and input devices (web camera) setup and a description of each component's functions. Following that, the whole project integration will be tested and troubleshoot to achieve the project's goals. A Gantt chart is utilized to indicate the order and periodic tasks that must be completed to guarantee effective time management.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter will elaborate further on the result and discussion of the entire project. It includes all the project testing results from simulation until real-time project testing, data analyzing, and project operation condition. The result from these testing and analyses will be the benchmark to determine whether the project objectives are archived or not. This chapter is divided or separated into three subchapters that will be discussed which are related results and final results.

4.2 Image Analysis

The final result of this project, called image processing and analysis techniques for estimating mango weight, will first be started with the preliminary work as shown in Figure 4.1 below. Using color thresholding programs from the image processing toolbox library was the method used during the previous semester. By altering the RGB, HSV, YCbCr, and L*a*b* values in this program determine the threshold for each object as shown in Figures 4.2 and 4.3 below. With the initial product, find the color threshold for this Mango. Export the image and function into a script file once you have reached the required threshold. This function will create the value that was set in the threshold app automatically. Then try to figure out how many black pixels each object has as shown in Figure 4.4 below. This method takes advantage of the web camera's 'show and snapshot picture' techniques to construct an algorithm that calculates the pixel value of an object visible through the web camera.

Next, in the image analysis method, there are some changes in terms of the way to obtain the pixels value of mangoes. This new way is better than the previous way. This is because can apply the knowledge that has learned while taking the Machine Vision course. For this new method, a thresholded image using Otsu's is used during the image analysis. In this section of image analysis, a discussion of the image output that has been obtained from the process that has been carried out. The detailed explanation for this method has been described in chapter 3 in the project development section.

Based on output image analysis as shown in Figure 4.5 below, the thresholded image by using the Otsu method was successful. The picture that has been obtained can show the mango fruit that has been used as an image object. However, there is some noise around the image border. The next process, fill the holes image, or better known as looking at the hole to the white region of the object area. The picture of the result obtained does not show any change because the picture before this process does not have any holes in the white region. To output this process, it will remain the same as the previous process picture.

In the next process, remove any blobs that are in the border-image area. Output results can show, changes in the image for the process. The noise around the border-image was filtered using commands from MATLAB. However, there is a bit of noise that is smaller in size than the pictures obtained from the previous process. The next process is added with a process to filter the small blob image found in the image. As a result of this process, the picture of the visible output has changed. Only placed objects can be displayed using a binary image.

Finally, the output image that is displayed is suitable to be used as the input image to obtain the pixel value in the white area. The value of the pixels can be obtained by using two

methods used to identify which method will obtain the estimated weight value that is closest to the actual weight.





Figure 4.3 Adjust HSV value





4.3 System Implementation

This subtopic is divided into two parts, namely training sessions and testing sessions. these two sessions are separated into two different applications using MATLAB software. The purpose of the application is separated to make it easier for the end-user to use the test application that has been made only to select the load linear regression line that has been trained based on the best dependency.

4.3.1 Training session

10 samples of mangoes were randomly selected for the training data session. A training data session is constructed so that the values for dependent and independent variables can be stored as a reference for the estimated weight of mango when using the testing application. Data from 10 samples that have been trained will be automatically saved as linear regression line data in excel format. After that, this data will be used as memory to reload during the test application. This training application was developed using App Designer from MATLAB software to facilitate experiments as shown in Figure 4.6. This application can show in detail each process that is run. The arrangement of inputs and outputs has been arranged according to the needs of the training application presented.

Four processes are carried out in the training application section. Among them are, image acquisition, image processing and analysis, statistical analysis, and finally linear regression. Firstly, the value of the actual weight of mangoes was inserted as a dependent variable. Next, press either the area or parameter method button to get the value of the independent variable in pixels. Before the independent value is obtained, the system application training will make a detailed image processing analysis and will display a picture of each process that has been carried out as an image output.

Independent values are obtained when the output image can be read well without any noise in the image border. At the same time, the values of the dependent and independent variables will be displayed in the form of tables and linear regression points. Next, repeat the step for another mango sample until the tenth mango is finished in the training session. Next, the reading values of the r-square and linear regression line with the equations can be calculated when the r-square button is pressed. Finally, observe the r-square value that is closest to the regression line or best fits the value of one. The selected r-square will use when a test application is run to test the performance of a buildable system. All buttons and keys in variable data are already programmed by the editor.



Testing application app was developed by using app designer MATLAB interface as shown in Figure 4.7 below. The main purpose of this app is built to evaluate mangoes that have been trained from training applications to identify the performance of the system that has been built. Insert actual weight of mango and click load linear regression line button that method was selected from the result obtained from the training application. This application test can be used on other mangoes other than the ten mangoes that have been selected to be used as samples for training applications. However, ten mangoes that have been sampled for the training set can also be used as an application test.

This is because, although the mango had been in the training set to obtain a linear regression line, there was also an error even though it had been made into 10 samples

selected for the training set. Unless the project uses a Support Vector Machine (SVM), Convoluted Neural Network (CNN), and Artificial Neural Network (ANN), the sample can no longer be used in the test set. Finally, the weight estimate values and error percentages from this test application will be used during data analysis and performance evaluation.



4.4 Data Analysis

The analysis of data is the most critical component of any research. The data analysis procedure summarises the acquired information. It is the process of applying analytical and logical reasoning to data to discover patterns, correlations, and trends.

4.4.1 Area

Table 4.1 below shows the results of the weight measurement for all 10 mangoes in the training set and the weight range from 284 to 391 g. Mean (μ) and standard deviation (s) of the 10 samples, data were 338.40 g and 34.48, respectively by using the equation as shown in 4.2 and 4.3 below. The range of calculating 95% confidence interval for μ by using equations in 4.5 and 4.6 below was between 313.73 and 363.07 g. The number of area mango regions in each of the training set's threshold images is listed in Table 4.1 below. The best fit line for the link between the area counted and the weight of mango is given in Figure 4.11 below, with a correlation of determination (r^2) of 0.9295. This correlation value revealed that there is a strong relationship between the counted area and the actual weight. The regression model that describes this relationship is as follows:

Estimated Weight = 0.001 * Counted Area - 119.81

The 95 percent confidence and prediction intervals result indicate a high level of confidence in the weight prediction. The prediction bands are larger than the associated confidence bands to account for the fact that the linear model predicts rather than estimates **UNVERSITIEEXNIXAL MALAYSIA MELAKA** the value of a random variable. As shown in Figure 4.9 below, the 95 percent confidence intervals for the intercept and slope of the linear model were -222.9546 to -16.6620 and 0.0008 to 0.0012, respectively. The validity of statistical analysis conclusions is contingent upon the fulfillment of specific data assumptions. Particularly when a linear model is supposed to be adequate, the residuals must be independent and normally distributed throughout with the same variance.

Table 4.1 contains the verification results for the built linear model utilizing ten mangoes from the training set. The mean absolute percentage error in calculating the weight of mango using the linear model was 0.77 percent (about 0.78 g for 100 g mango) when

equations 4.7 to 4.9 were used. Thus, image processing and analysis methods may be used to estimate the mango's weight with a high degree of precision. Lastly, all the value was calculated using an excel template that has been created to check the calculation is correct or not as shown in Figure 4.10 below.

S

MALAYSIA

$$Sum = \sum x_i \tag{4.1}$$

$$Mean = \frac{Sum \ of \ All \ Data \ Points}{Number \ of \ Data \ Points}$$
(4.2)

$$= \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$
(4.3)

$$Confidence Interval = \bar{x} \pm CONFIDENCE$$
(4.4)

$$\epsilon(\%) = \frac{Actual \ weight - Estimated \ weight}{Estimated \ Weight} \times 100\%$$
(4.7)

$$\overline{\text{UNIVERSIT}\epsilon_{rms}(\%)} = \sqrt{\sum_{i=1}^{n} \epsilon_i^2} \text{AYSIA MELAKA}$$
(4.8)

$$\varphi(\%) = 100 - \epsilon_{rms}(\%) \tag{4.9}$$

Samples	Counted	Actual	Estimated	
	mango area	weight (g)	Weight (g)	
1	487199	371	371.4223	
2	414530	298	298.152	
3	441135	334	324.9771	
4	440736	323	324.5748	
5	442172	325	326.0227	
6ALAYSIA	482922	360	367.1099	
7	501417	373	385.758	
8	410989	284	294.5817	
9 9 1/1/10	433486	325	317.2648	
يسياملاك	489891	يتى يېڭىڭ	374.1366	

Table 4.1: Mango area, actual, and estimated weight of mangoes in the reference set



Figure 4.8: The dependency between area and actual weight

SUMMARY OUTPUT								
Regression Si	tatistics							
Multiple R	0.964093332							
R Square	0.929475954							
Adjusted R Square	0.920660448							
Standard Error	9.712334333							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	9945.764494	9945.764	105.436486	6.96372E-06			
Residual	8	754.6355056	94.32944					
Total	9	10700.4						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-119.8082801	44.72946359	-2.67851	0.027990601	-222.9546081	-16.66195209	-222.9546081	-16.66195209
X Variable 1	0.001008275	9.81937E-05	10.26823	6.96372E-06	0.00078184	0.00123471	0.00078184	0.00123471

Figure 4.9: 95% confident interval for intercept and slope of the linear model of the area



4.4.2 Perimeter

Table 4.2 below shows the results of the weight measurement for all 10 mangoes in the training set. For the second approach, the same mango is utilized. As a result, the 10 data samples' weight range, mean, standard deviation, and confidence interval are all the same. Because all of the values are calculated using the weights of all of the samples. The only difference in the second technique is that it now simply uses the existing Matlab function region prop (perimeter) to acquire the perimeter value. The number of mango areas around the perimeter of each threshold picture in the reference set is listed in Table 4.2. Figure 4.11 depicts the best fit line that explains the link between the perimeter counted and the weight of mango, with a coefficient of determination (r^2) of 0.6776. When compared to the first approach, this correlation of determination suggests a poor relationship between the calculated perimeter and the actual weight. The r-squared number published isn't even close to one. As a result, the predicted weight value for mangoes may differ significantly from the actual weight of mangoes. The regression model describing this dependency is:

Estimated Weight = 0.2808 * Counted Perimeter - 378.29

The 95 percent confidence and prediction intervals result indicate a lack of trust in the weighted prediction. The prediction bands are larger than the associated confidence bands to account for the fact that the linear model predicts rather than estimates the value of a random variable. As seen in Figure 4.12 below, the 95 percent confidence intervals for the intercept and slope of the linear model were -781.6510 to 25.0805 and 0.1229 to 0.4387, respectively.

Table 4.2 shows the validation results of the linear model created with the 10-mango training set. When applying a linear model to estimate mango weight, the average absolute percentage error was 2.05 percent (approximately 2.1 g for 100 g of mango) using equations 4.7 to 4.9 above. As a result, using picture processing and analysis tools, the mango weight approach can be approximated with a low level of accuracy. Lastly, all the value was calculated using an excel template that has been created to check the calculation is correct or not as shown in Figure 4.13 below.

Samples	Counted mango	Actual	Estimated		
	perimeter (pixels)	weight (g)	Weight (g)		
1	2669.462	371	371.2523111		
2	2465.69	298	314.0367476		
3	2531.624	334	332.5498454		
4	2470.368	323	315.350247		
5	2475.48	325	324.5245267		
6	2738.568	360	390.6560502		
7 MALAYSI	2572.282	373	343.9658907		
8	2448.664	284	309.2561488		
9	2475.48	325	316.785606		
10	2649.412	391	365.6226267		

Table 4.2: Mango perimeter, actual, and estimated weight of mangoes in the reference set



Figure 4.11: The dependency between perimeter and actual weight

SUMMARY OUTPUT								
Regression S	tatistics							
Multiple R	0.823141179							
R Square	0.6775614							
Adjusted R Square	0.637256575							
Standard Error	20.76722777							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	7250.178007	7250.178007	16.81092526	0.003437225			
Residual	8	3450.221993	431.2777492					
Total	9	10700.4						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-378.2852725	174.9197976	-2.162621256	0.062533462	-781.6510492	25.08050421	-781.6510492	25.08050421
X Variable 1	0.280782264	0.068481594	4.100112835	0.003437225	0.122863424	0.438701104	0.122863424	0.438701104

Figure 4.12: 95% confident interval for intercept and slope of the linear model of the

perimeter



4.5 Discussion

A good estimation equation was selected based on R-square (r^2) obtained for the correlation equation. Table 4.3 shows below, the estimation equations obtained for the contour area and perimeter (in pixels) with their respective r^2 .

Estimation Equations	r^2 values		
Estimated Weight = 0.001 * Area - 119.81	0.9295		
Estimated Weight = 0.2808 * Perimeter - 378.29	0.6776		

Table 4.3: Weight estimation equation and they are Respective r^2 values

The findings demonstrated the characteristics' varied correlation relationships. This is shown by the strong correlation of determination $r^2 = 0.9295$ found for the equations using the weight as the independent variable in the estimate equations involving area measurement through aspect ratio. This suggests that the size measurement calculations utilizing the aspect ratio area functioned well as mathematical descriptors for estimating weight from area measurements used to determine the mangoes' weight. This indicates that the image analysis approach employed in this experiment, which was based on the area of the fruits, was sufficiently reliable to estimate the mangoes' weights.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

4.6 **Performance Evaluation**

The system's performance is evaluated using the Root Mean Square Error (RMSE) approach. This technique is capable of estimating the two ways employed, which are area and perimeter, and compares the real weight of mangoes with a 95 percent accuracy. The percentage of accuracy is obtained using the formulae presented in 4.10 to 4.12 below. Where is the percentage of errors for each sample, ϵ_{rms} is the root mean square error, is the percentage of accurate samples, and n is the total number of mango samples.



Figure 4.14: Actual weight and estimated weight method 1: area



Figure 4.15: Actual weight and estimated weight method 2: perimeter

$$\epsilon(\%) = \frac{Actual \ weight - Estimated \ weight}{Actual \ weight} \times 100\%$$
(4.10)

$$\epsilon_{rms}(\%) = \sqrt{\sum_{i=1}^{n} \epsilon_i^2} \tag{4.11}$$

$$\varphi(\%) = 100 - \epsilon_{rms}(\%) \tag{4.12}$$

4.7 Summary

In this section, the data findings from the data analysis in the previous subchapter would summarize. For this chapter, the data analysis result was divided into two parts, area, and perimeter which these two method was the test in this system. The test was run with 10 samples to test the efficiency of the detection image processing. The test was divided into two parts of the application for testing, which are training application and test application. From the data analysis, it can conclude that both methods are available to use for estimating the weight of mangoes using linear regression.



CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

In this chapter, we emphasize the conclusion of the project and any recommendations that can be done for future development. This part includes the conclusion of all procedures and what can we find in throughout this project that should be avoided in the future to enhance the future development of this project.

5.2 Conclusion

This project, titled "Estimating Mango Weight Using Image Processing and Analysis Techniques," was successfully developed. This image processing was designed using MATLAB to implement all of the methods. The appendices and chapter 3 of this report detail the component utilized in this project. Mangoes may be weighed utilizing picture processing and analytic tools. The findings show that image filtering and thresholding can be used to easily smooth and segment the mango picture. Mango area and perimeter calculated from the thresholded picture show a strong correlation with mango weight.

The region has the best correlation of determination (r^2) , which is 0.9295. The mean absolute percentage error was 0.77 percent when calculating the weight of mango using the linear model. The system's performance is examined using a total of ten mango samples; it is shown that the suggested technique achieves a 95% accuracy for the area and weight estimates with an average processing time of 7 seconds per mango.

5.3 Recommendation for Future Project Improvement

The current effort has completed the first part of determining the mango's weight utilizing picture processing and analysis tools. It is advised that more research be conducted to construct a machine vision system that employs the previously established linear model for automatically rating the weight of mangoes.

Lighting is critical while photographing photographs. In future work, a more controlled environment with correct lighting conditions may be created. Additionally, this algorithm is planned to be paired with a mechanical and electrical system for autonomously sorting and grading mangoes. The image is now processed on a personal computer with the help of the MATLAB program. In future efforts, a microcontroller or microprocessor-based algorithm that is better appropriate for application in a real system should be created.

5.4 Commercialization Relevancy

Mango is one of the most important fruits grown in Malaysia, having a significant export value. Post-harvest handling is a critical step in the fruit production process since it helps maintain the fruit's quality. Cleaning, grading, and packing are all steps in the postharvest handling process. Grading is critical for evaluating the quality of fruit as well as for price. Numerous local farmers continue to rely heavily on human labor to grade the fruits according to the standards set by the Federal Agricultural Marketing Authority (FAMA) Malaysia. The fruits are classified according to their size, maturity index, and exterior flaws (Teoh and Syaifudin, 2007).

This project might be used to automate the inspection and grading of fruits. These systems have been extensively employed in the food and agricultural industries for inspection and evaluation because they provide a speedy, cost-effective, consistent, and objective assessment. Automated product inspection using machine vision not only saves labor but also increases inspection objectivity. Over the last decade, advancements in hardware and software for digital image processing and analysis have prompted various research on the creation of a system for evaluating the quality of a variety of different and processed foods.

Along with the quality of fruits, size is a critical aspect that consumers identify. The size of tropical fruits (mango, pineapple, watermelon, papaya, and starfruit) is determined by their weight, according to FAMA. Manually weighing fruits to determine their size is time-consuming, the labor involved, and expensive. The computer vision system's image processing and analysis capabilities may be utilized to do an alternate automated grading of fruit size based on weight.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

REFERENCES

Atencio, P., Sánchez T, G. and Branch, J. W. (2009) 'Modelo visual automatico para la clasificacion y medida de calidad de fruto: Caso mangifera indica L', *DYNA (Colombia)*, 76(160), pp. 317–326.

Aung, W. et al. (2020) 'Weight Estimation of Mango from Single Visible Fruit Surface using Computer Vision', 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan, SICE 2020, pp. 366–371. doi:

10.23919/sice48898.2020.9240422.

Bermúdez, A. M. (2013) 'Image analysis for automatic feature estimation of the Mangifera indica fruit', *Ingeniería y Desarrollo*, 31(1), pp. 84–104.

Dang, N. T. *et al.* (2019) 'Analysis on Mangoes Weight Estimation Problem using Neural Network', *Proceedings - 2019 19th International Symposium on Communications and Information Technologies, ISCIT 2019*, pp. 559–562. doi: 10.1109/ISCIT.2019.8905118.
Ganiron, T. U. (2014) 'Size properties of mangoes using image analysis', *International Journal of Bio-Science and Bio-Technology*, 6(2), pp. 31–42. doi: 10.14257/ijbsbt.2014.6.2.03.

Khairul Adilah, A. and Sharifah Lailee, S. A. (2016) 'Automatic preharvest grading of Harumanis fruits', *Computing Research & Innovation (CRINN)*, 1(1), pp. 19–26.

Long, N. T. M. and Thinh, N. T. (2020) 'Using machine learning to grade the Mango's quality based on external features captured by vision system', *Applied Sciences*

(Switzerland), 10(17). doi: 10.3390/APP10175775.

Momin, M. A. *et al.* (2017) 'Geometry-based mass grading of mango fruits using image processing', *Information Processing in Agriculture*, 4(2), pp. 150–160. doi:

10.1016/j.inpa.2017.03.003.

Olden, J. D. and Jackson, D. A. (2002) 'Illuminating the "black box": Understanding variable contributions in artificial neural networks', *Ecological Modelling*, 154, pp. 135–150.

Sa, F. S. A. et al. (2015) 'Shape and weight grading of mangoes using visible imaging',

115, pp. 51-56. doi: 10.1016/j.compag.2015.05.006.

Sahu, D. and Potdar, R. M. (2021) 'Defect Identification and Maturity Detection of Mango Fruits Using Image Analysis', (January 2017). doi: 10.11648/j.ajai.20170101.12.

Spreer, W. and Müller, J. (2011) 'Estimating the mass of mango fruit (Mangifera indica,

cv. Chok Anan) from its geometric dimensions by optical measurement', Computers and

Electronics in Agriculture, 75(1), pp. 125–131. doi: 10.1016/j.compag.2010.10.007.

Teoh, C. C. and Syaifudin, A. R. M. (2007) 'Image processing and analysis techniques for estimating weight of Chokanan mangoes', *Journal of Tropical Agriculture and Food Science*, 35(1), pp. 183–190. Available at: http://ejtafs.mardi.gov.my/jtafs/35-1/Chokanan mangoes.pdf.

Vo, M. T., Duc Nguyen, T. and Dang, N. T. (2019) 'Embedded machine learning for mango classification using image processing and support vector machine', *Proceedings -*2019 6th NAFOSTED Conference on Information and Computer Science, NICS 2019, pp. 279–284. doi: 10.1109/NICS48868.2019.9023803.

APPENDICES

Appendix A (Gantt Chart) BDP 2



UNIVERSITI TEKNIKAL MALAYSIA MELAKA



Appendix B (MATLAB Application Interface)

The interface for testing application samples of mango



Appendix C (Image Acquisition Setup and Demonstration)

The image segmentation experiment



Buy 10 samples of mango in Mydin





Demonstrate


Appendix D (Output Image and Result)

Output image for the perimeter





Screen show 10 sample that was recorded 1 by 1 in a linear regression line chart