Development of an Embedded Vision Based Fruit Sorting Machine

Submitted By

Clary Norman (2017141960)

Under Supervision of

Mr. Edwin Vans

FIJI NATIONAL UNIVERSITY

SCHOOL OF ELECTRICAL & ELECTRONICS ENGINEERING

COLLEGE OF ENGINEERING, SCIENCE & TECHNOLOGY

DERRICK CAMPUS, FIJI NATIONAL UNIVERSITY

November, 2020

Development of an Embedded Vision Based Fruit Sorting Machine

CERTIFICATE

This is to certify that Project Report entitled "Development of an Embedded Vision Based Fruit Soring Machine" which is submitted by, Clary Norman in partial fulfillment of the requirement for the award of degree of Bachelor of Engineering in School of Electrical & Electronics Engineering of FIJI NATIONAL UNIVERSITY, is a record of the candidate own work carried out by him under my/our supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date: ______________________________

Supervisor's Name: ______________________________

Supervisor's Signature: ___________________________

DECLARATION

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature: CN

Name: Clary Norman

ID No. 2017141960

Date: 10/11/2020

ACKNOWLEDGEMENTS

I would like to express my sincere heartfelt gratitude to the following people for their tremendous support in one way or the other to the successful completion of this PEB802 Capstone Design Project 2.

First and foremost, I would like to thank the Almighty father for his showers of blessing throughout my research work to complete the research successfully.

Secondly, I would like to express my deep and sincere gratitude to my supervisor Mr. Edwin Vans for his support and cognizant efforts throughout the entire course of this project. Without his help, the completion of this paper would have been impossible.

Thirdly, I would like to thank my parents for their support rendered through encouragement and motivation that keeps me focus towards achieving my goal.

Lastly, I would like to acknowledge the contribution of all faculty members of the school of Electrical and Electronics Engineering department for their kind assistance and cooperation which help me a lot in completing the PEB802 Capstone Design Project 2 within the limited time frame.

ABSTRACT

Since the advancement of technology in the area of agricultural and commercial farming, manual operation has been drastically increasing to combat any disadvantage being encountered, for example more handy work has to be doubled in order to produce efficient output. These kind of practice often results in major drawbacks such as; low output, time consuming, inconsistency and so forth. Old fashion operation was now being replaced by vision based system for sorting and grading of fruits due to the disadvantage being mentioned. The transition came about as a result of manual operation fail to meet the high production demand and causing labour cost to increase thus more workers are now being recruited to perform the task. Therefore, in order to meet the high production demand and speed up the process, a proposed design based on Development of an Embedded Vison based fruit sorting machine was designed to increase the production and reduce the cost. So basically, in this project, the main objective was to develop a low cost modular system using vision based technology to distinguish between orange and apple fruit. The system involves training the model with different classifier such as SVM, kNN and binary classifier and comparing the performance of each of the classifier. It has been observed that SVM classifier technique for orange and apple conditions is efficient than other techniques. The results recorded for the accuracy and precision using this technique were 100% respectively. Furthermore, the number of fruits process within 60sec is 39. The source code for the project can be found at: https://github.com/clary045/Apple-and-Orange-Grading-Using-SVM. A video demonstration of the project can be found at: https://youtu.be/IHFCdor68Iw

TABLE OF CONTENTS

Page

List of Tables

List of Figures

List of Acronyms

GMM	Gaussian Mixture Model
FSM	Fruit Sorting Machine
FNU	Fiji National University
SPC	Secretariat of the Pacific Community
HSV	Hue Saturation Value
RGB	Red Green Blue color space
CMY	Cyan Magenta Yellow
IEEE	Institute of Electrical and Electronics Engineers
GUI	Graphical User Interface
PCB	Printed Circuit Board
kNN	k –Nearest Neighbors
SVM	Support Vector Machine
ASM	Angular Second Moment
GLCM	Gray Level Co-occurrence Matrix
ANN	Artificial Neutral Network

CHAPTER 1. INTRODUCTION

1.1 BACKGROUND OF PROBLEM

Before civilization came into place, old fashion style of agricultural and commercial farming has been widely practiced throughout the globe. Although it helps to boost the agricultural sector, there were major drawbacks that exert a gap between the old and this 21st century. Technical engineering evolved rapidly as well, thus there is a need for change in all aspect of agricultural and commercial farming. This paper is aimed to investigate on vision based system for sorting of fruits.

Fruits are essential commodities in today’s world, and they must be checked for the ripeness, firmness, texture and size. Different fruits when shipped across one place to another must be checked for quality controls as well. The manual method of handpicking the best fruit among the stock is a time consuming process.

Commercial agricultural concept requires an adequate and optimum level of modern technology to yield a good quality crop. There were numerous obstacles encountered by orange and apple grower to nurture a standardized product. One of the issues troubling agriculture’s entrepreneur is heavy loss due to damaged and spoilt orange and apple. The common reasons were inadequate knowledge in technical and management wise. For instance, manual plucking and vibrator’s assistance method to harvest goods is a leading factor. Machine vision technique based on shape analysis will be applied to tackle the issues arises. Nevertheless, a compatible color detection algorithm will be complement for the former method to further improve the system’s efficiency. Relevant experiments will be conducted throughout the project to offer a judgment that applied machine vision technique is capable to replace human workforce [1].

Generally, this work utilizes many resources from previous works of researchers. The first part consists of research papers on application of small and embedded computer unit in completing an image processing task. In these papers, more discussion is made over the benefit of using Raspberry Pi in conducting image processing task and the practical design using Solid Work.

1.2 CONTEXT

This study develops an efficient algorithm for grading of orange and apple fruit based on the colour features in real time. The proposed real-time system uses a raspberry pi camera for capturing fruit image when transported on the conveyor belt. The captured image are then analysed with different trained classifier such as binary classifier, SVM and kNN to predict the outcome. Based on the performance of each classifier, the superior classifier was implemented in hardware basis to achieve the scope of the project.

A number of researchers have already proposed an automated fruit sorting machine for different types of fruits. Previously automated fruit sorting machine was done using MATLAB and Arduino where the computation time is often the main issue. However, this proposed paper also focusses on the same field but instead of using those technologies, a raspberry pi was used as an embedded system to accomplish the same task.

In this paper, the context of study will be based on the following guideline;

- Image Processing
- Feature Extraction
- Train Classifier
- Performance Testing

1.3 PURPOSES

The purpose of carrying out this project is to provide a cost – effective and efficient fruit sorting machine for local farmers and supermarket shop owners whom are still practicing manually sorting of fruits.

The goal of this project was to design a fruit sorting machine that is cost efficient, marketable, environmental damage free, reduce risk and solve the problem related to manual sorting of fruit.

Under this scope, the main objective of this project are as follows;

1. Designing a fruit sorting machine to sort two different fruits (e.g. Apple and Orange)
2. Come up with a suitable algorithm that would work best for processing of the image and decision making in less time.
3. Implementing the system and doing testing in real time in unseen cases
4. Provide users with more secure and cost-effective fruit sorting machine system

The project expected overall output are as follows:

- Improvement in the agricultural sector

  The manual sorting of fruits will be eliminated and farmers will be provided with an effective fruit sorting machine at a lower cost. This will help to speed up their production rate and contribute more efficiently in the international market.

- Safe and easy to use

  Since the operation and design specification are not complex as compared to a larger scale type of machine, this will be handier to the user, provided that it is controlled and monitored by an embedded systems such as Raspberry Pi [2]. In addition, most of the job will be done by the system thus user has less exposure to the system.

- Minimal environment impact

  This machine is environmental damage free, meaning it does not contribute to any pollution that is harmful to the environment. Not forgetting, the machine has better control over the defects fruits and much safer to the surrounding environment since manual operation is not being practiced.

- Social impact

  The social impact of this system are as follows.

  Speed and accurate sorting can be achieved, meaning it will help the local farmers to produce good quality fruits to sell while at the same time it will lessen the time for them to choose among the best fruit and the defect ones. Greater product stability, meaning consistency of products in the international markets. Cost effective, meaning it will provide income for the local farmer through their quality products. Job opportunities, meaning it will provide opportunity for local's to go and train how to manage and operate the machine. Social interaction, meaning it will bring local farmers together to discuss new ideas and innovations.

1.4 DESCRIPTIONS OF PROBLEM

Fiji is known to be a country situated in the tropical climate where farming has been their main source of income especially crop farming where land is used for growing fruits and vegetable. According to the SPC land Resources Division report in 2017, an estimated 46,074 tonnes valued at $F56.4 million of fruits and vegetables were imported and around 2,681 tonnes valued at $F4.8 million were exported [3]. The main issue faced by local farmers and supermarket owners was the grading of the fruit. Since most of the imported fruits often comes in different sizes and types, it was very difficult to pack them accordingly. To do this task it requires more workers to deliberately get the job done. However, the main issue was the time it takes to complete the job. Since most of the grading was done by visual interference of the human eye this can pose problem in maintaining consistency in grading and uniformity in sorting. In terms of prizing the fruits, the cashier needs to know the size of the fruit to allocate the right price. This was very difficult to assess with the natural eye and this can often leads to imbalance pricing. Local farmers faced the same issue when production demand were very high. At times when farmers face financial issues such as not enough money to pay the workers, many of their fruits were left unpack for weeks and the fruits started to lose its quality. This has cause so many problem such as high labour cost, slow in work speed and degradation in production demand. Also during this pandemic, COVID-19 cases rise significantly thus, workers are not allowed to work together in large number as it may leads to spreading of the viral disease. This has lead this paper to come up with a cost effective project based on embedded system vision based technology to solve the issue faced by local farmer and local supermarket owners. This system was design such that, the process of sorting the fruit take less time to process and distinguish between fruits coming on the conveyor. It also calculated the size of the fruit and pack it separately accordingly at a faster rate and an accuracy of 90 – 100% (fruit/second) to allow faster production.

CHAPTER 2. REVIEW OF LITERATURE

Over the past years manual operation has been the most commonly practiced system that was being used in the agriculture and food/marketing system. Although this practice was widely used there were major draw backs that associate with the system these includes: low production, unattractive appearance of product and more workload. Furthermore, the manual quality and quantity control of the product (fruits) are often time consuming thus making this method less efficient when it comes to its work ethics [4]. Therefore, development have taken place over the past years up until now and computer aided system or rather automatic systems, were put into place to counter the disadvantage that were faced from the manual operations. Today, there are various techniques being developed by researchers for sorting of a particular fruit depending on the needs and requirements.

Base on these research article report title as “Smart Farm: Automated Classifying and Grading System of Tomatoes using Fuzzy logic [5]”,” Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques [6]” and “An experimental Machine Vision System for sorting sweet Tamarind [7]”, the two research shows that manual sorting of fruit was one of the leading causes of low production, time and money consuming, low precision and low automation thus this current trend needs to be stopped. From previous research article named respectively as “Machine Vision Based Automatic Fruit Grading System using Fuzzy Algorithm [8]” and “Computer Vision: An objective, rapid and non-contact quality evaluation tool for the food industry [9]“, pointed out and justify how the importance of machine learning and vision based technology has impact and solve problem related to manual sorting of fruit.

There are various type of techniques being developed by researcher for sorting of a particular fruit depending on the needs and requirements. Ng Weng Seng et. al [10] introduced a system based on image enhancement. They present new techniques of improving image processing and a prototype of a vision based sorting machine. Their system consist of a Raspberry pi board, Raspberry pi camera module, infrared sensor, DC motors and servo motor. This method uses the intensity of RGB value and a converted binary image for determining the color, shape and size of the work pieces. Furthermore, an additional contrast smoothing algorithm called Contrast limited adaptive histogram specification (CLAHE) were also used to reduce effect of the surrounding illumination. In this system, tested results indicates that vision based automatic segregation system improves the accuracy and efficiency of the work. Unlike various image processing techniques that were used thus, this help to stabilize the capture image when contrast of the surrounding is too low or too bright. However, this system cannot be used in industry where texture features of an object were highly recommended.

Zenan Lin et. al [11] introduced a system based on HSV color space for fruit recognition. In the setup, the image was captured from the acquisition module passed on to the recognition module, where it converts the RGB space of the image to HSV color for extracting the H and S component. Once the H and S component of the HSV color space are utilized, morphological method is then used to determine the quality of the fruit such as Area, the class, size and the calculated position of the fruit. The method are tested on 255 different fruit image and the result shows that the average time taken for recognizing the fruits is 0.0402s with an accuracy of 93.73%. This method was preferred when light intensity changes is becoming a problem because it helps to overcome the influence of light intensity and satisfy different situation. However, the light intensity still has a certain influence when recognizing different classes of fruit.

Prathamesh Avalekar et. al [12] also introduced a system based on a neural network concept. This concept was based on extracting features of images that were captured using several different color model such as RGB, HIS, CMY,YCbCr. These color model techniques were used to get the features for training the neural network. When the training are completed then other fruits were divided into different categories based on above trained network. This system was proven cheap, effective, compact in size and user-friendly. However, neural networks has no specific rule for determining the structure of artificial neural networks. Appropriate network structure is achieved through experience and trial and error.

Pavithra et. al., [13]presented a system based on Machine vision based automation using kNN and SVM for sorting of cherry tomato. In the setup, the system firstly check for maturity of the fruit. The maturity was determined based on the ratio of the red pixel area of the image to the total area. If the ratio is greater than 0.6, the tomato will be classified as mature. Secondly, the system will continue to check for the quality of the fruit such as texture features, color features and shape. For texture features, GLCM used to extract the structural arrangement of the tomatoes such as Entropy and ASM. Similarly, for color features, CLCM used to calculate the Red band. The above features including the shape were fed to the KNN and SVM for classifying the matured fruit into their classes. Although combing KNN and SVM has greater performance in accuracy, however the computation time taken for running the system longer.

Shweta S. Deulkar et. al., [14]proposed an image processing based tomato quality grading system using K-means clustering and SVM. In the setup, the captured image was first represented in RGB format and then transformed to gray scale image format using Otsu thresholding method. This allows the histogram and threshold value of the image to be computed. After this, K-means clustering method was applied to extract the features of the fruit which are relevant for the classify technique to analyze and make decision. Based on the extracted features, SVM method then classify the fruit into three different classed (1^{st} grade, 2^{nd} grade and 3^{rd} grade). Although, this technique has proven to solve the problem related to tradition sorting of fruit, however, the system should utilize other features like weight and size to help improve the outcome and not just the color itself.

Dnyaneshwari Pise et. al., [15]presented a technique for grading mango fruit using Naïve Byes method. This process utilized RGB method for extracting the features of the mango. The acquired image were pre-processed using different method as Gray scale, Binary image representation and max height. Based on the RGB value of the image, the size, shape and surface were calculated. For size and shape detection, an addition method known as edge-detected method was used along with the RGB method for detection of its size and shape. Once all of this features were extracted, Classification using Naïve Byes and Posterior analysis technique were performed to grade the mango. The table 4-0 below shows the result obtained when performing RGB method and Naïve method.

Table 1: Classification of mango

Sample Images	RGB Values	Diameter	Mango Class
	139.189, 182.466,108.846	223	Green mango
	233.162, 200.682, 135.866	274	Yellow Mango
	105.277, 41.619,53.937	356	Red mango

Although this result has obtained greater accuracy as reflected from the result, however, this method only work for a particular surface area and cannot be use if other than that surface exist, thus it is not accountable for outside marketing area.

Arun Kumar R et. al., [16]proposed a robust non-destructive method to sort pomegranates using Wavelet features and ANN training. Their system utilized MATLAB 2010a software for carrying out the process and implementing the algorithm. The system uses Logitech webcam for capturing of the image with the help of the light. Since light has some effect on the image captured, thus they have use Histogram Equalization to neutralize the effect of light. In this way, the intensity values of the image is also neutralize. However, on the other hand using of Histogram Equalization can also reduce the information of the image. Therefore, a Wavelet denoising was employed to improve the information of the image and remove the noise present in the signal. After pre-processing of the image, the image will then be fed into the Frequency domain features for extraction of its features. In this method a total of 252 wavelet features were extracted for each image. In Table 4-1 shows the features that were extracted using frequency domain.

Table 2: Wavelet features extracted

Sl. No.	Feature Name	Remark
Statistical features		
1	Mean	Average of the coefficients
2	Standard Deviation	Standard deviation of the coefficients
Wavelet texture features		
3	Contrast	a measure of the intensity contrast between a pixel and its neighbor over the entire image
4	Correlation	a measure of how correlated a pixel is to its neighbor over the entire image
5	Energy	sum of squared elements in the GLCM
6	Homogeneity	measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal
7	Entropy	a statistical measure of randomness, used to characterize the texture of an image

These features then fed to the ANN classifier for training. Based on the training, a confusion matrix was obtained and the following statistic were given (TP, FN, FP & TN). Based on this statistic, sensitivity (se), specificity (Sp) and accuracy (Ac) were calculate. With this performance parameter, the Pomegranates were sorted into disease and healthy classes. This method has an accuracy of 91.3%, however some of the features may not contributed to the sorting process, as they may contribute to slowing the processes.

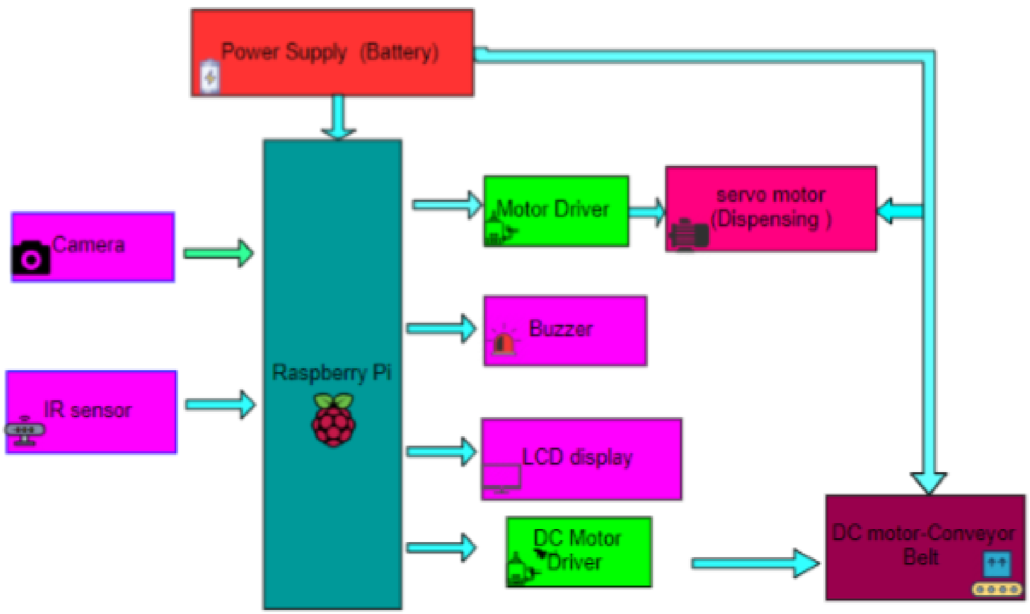
Finally, Chandra Sekhar Nandi et. al., [17] presented the idea of grading fruit using GMM and Fuzzy Algorithm. In the system, the image captured goes through pre-processing and extraction features. This is where the background elimination and contour detection is done. Afterwards, GMM then used to predict the maturity of the fruit as well as its quality. Once the maturity and quality are extracted, fuzzy rule based algorithm then classify the fruits accordingly. In this setup, five varieties of mango were tested with the algorithm. The results shows that the classification performance for the proposed vision based automatic technique are the same as the manual expert technique. This technique was found to be fast, low cost and moreover intelligent. However, the propose system fail to find out the maturity level correctly when the surface of the mango were highly contaminating with scratches and black patches.

In conclusion, overall perception of the studies that were being reviewed in these literatures shows that all researcher have the same goal which was to eliminate the use of manual operation by using vision based technology. However, their method used to achieve their goal were all different. This was because all the methods that were being used had several different circumstances and are not applicable with one another.

CHAPTER 3. SYSTEM DESIGN OVERVIEW

3.1 Block Diagram and Flow chart

Figure 1 shows the block diagram representation of the desired system with their functionality.



The Figure 1: Proposed block diagram of vision based fruit sorting machine

The function of the block diagram main parts were described as follows;

1. The input image is taken from the camera and send to the raspberry pi for processing of the image.
2. In raspberry pi, the input image will be processed and model using GMM algorithm, SVM and KNN. If the image match any of the type of fruit (e.g. apple and orange), the raspberry pi will then send a signal to control the motor movement to sort the fruit into their respective box. Python coded algorithm will be programmed into the raspberry Pi where the user will trained the model and see if the model recognize the fruit.

3. The infrared sensor connected to the raspberry pi function is to sense the fruit coming on the conveyor belt so that the running conveyor belt can be interrupted at any instant when a fruit is detected and give chances for the camera to take a snap shot of the fruit.
4. Fruit detection would also activate the buzzer sound and display a message indicating a fruit is being detected.
5. The purpose of the servo motor was to dispense the fruit and it is control by the raspberry pi.
6. In this project, the conveyor belt was used to transport the fruit and it is control by the raspberry pi.
7. Lastly, in order for the system to run a power supply is setup to run the whole system and charge the battery that is to prevent the system from any power failure occurring.

The figure below shows the conceptual engineering design for the system

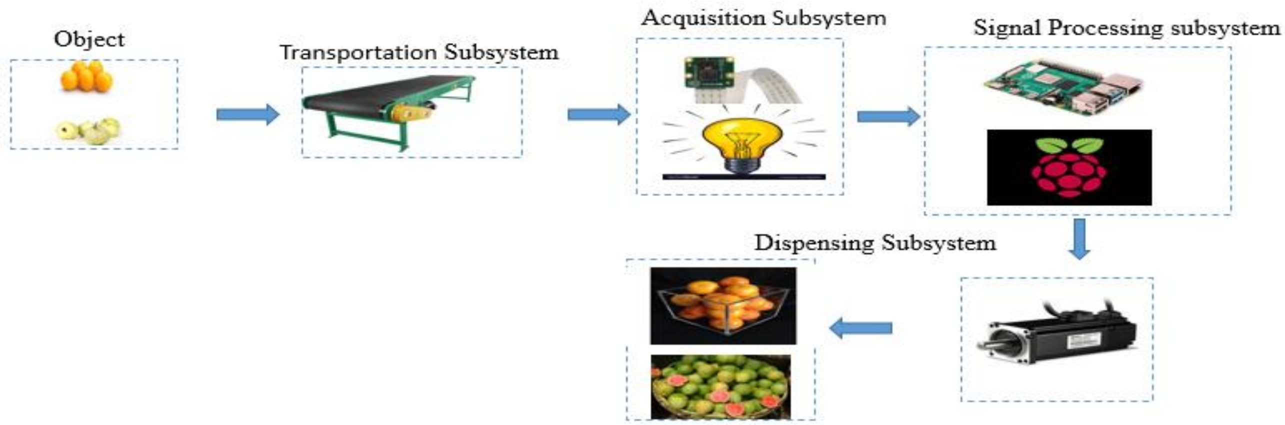


Figure 2: Conceptual Engineering Design for the system

From the figure, the first subsystem is the transportation subsystem and it includes the object (apple or Orange) and the conveyor. The purpose of the first subsystem is to bring the object to the acquisition site. The second subsystem is the acquisition subsystem and it is made up of camera and lightings. The purpose of the second subsystem is data acquisition. The third subsystem is the signal processing subsystem and as the name suggest, this subsystem involves pre-processing, feature extraction and modelling to process the raw Image signal acquired from the acquisition site. It also calculated the size of the object and identify the type of image signal being process. This subsystem contains the Raspberry Pi for image processing and transmission. The fourth subsystem is the dispensing subsystem that consist of the servo motor and the lever. The purpose of the servo motor is to control the movement of the lever when sorting of fruit.

3.2 Flow Chart of the system:

Figure 3 below shows the summary work flowchart of the design

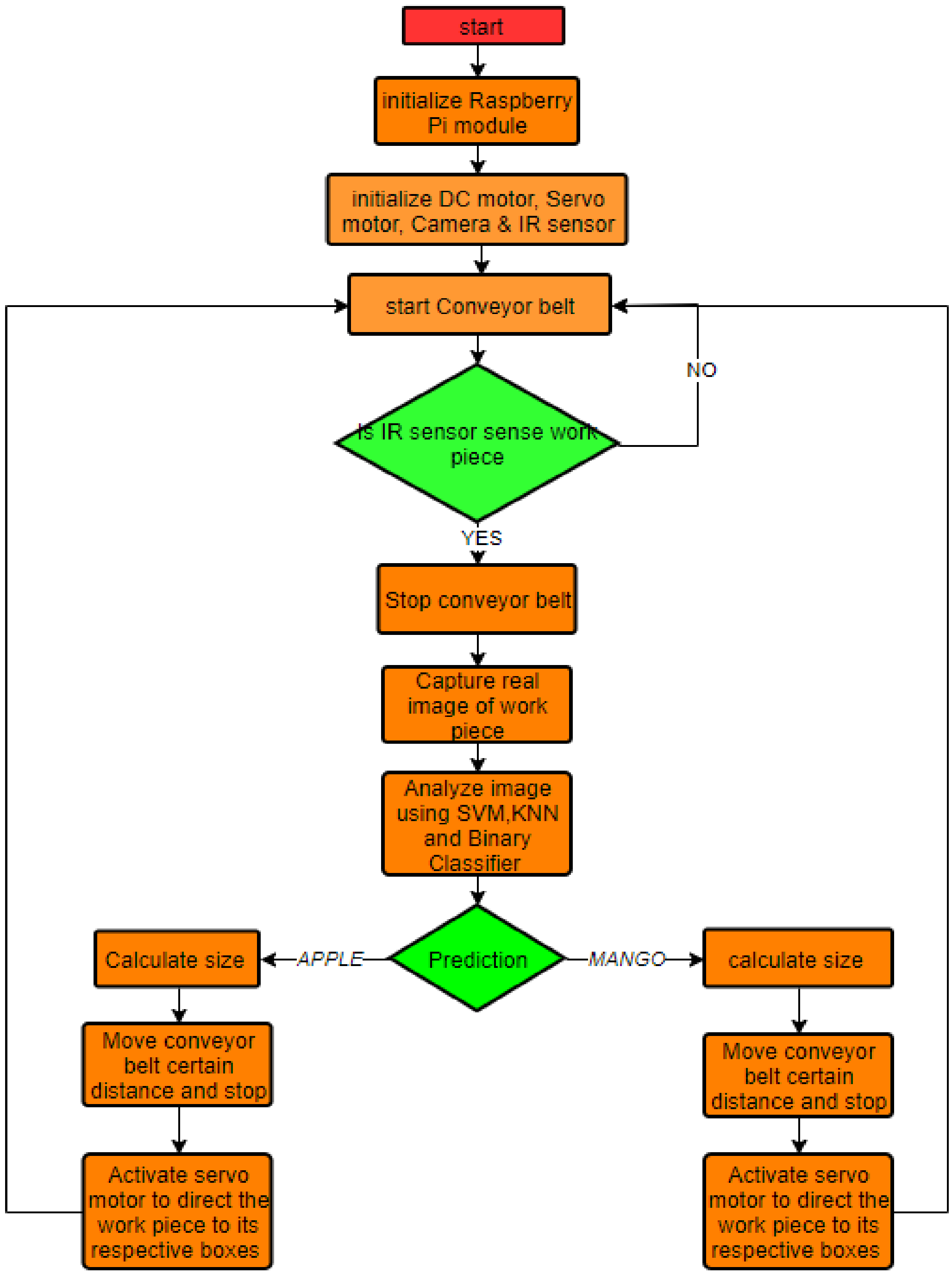


Figure 3: System flowchart

3.3 Schematic Diagram of the system

The figure below shows the overall schematic diagram of the system.

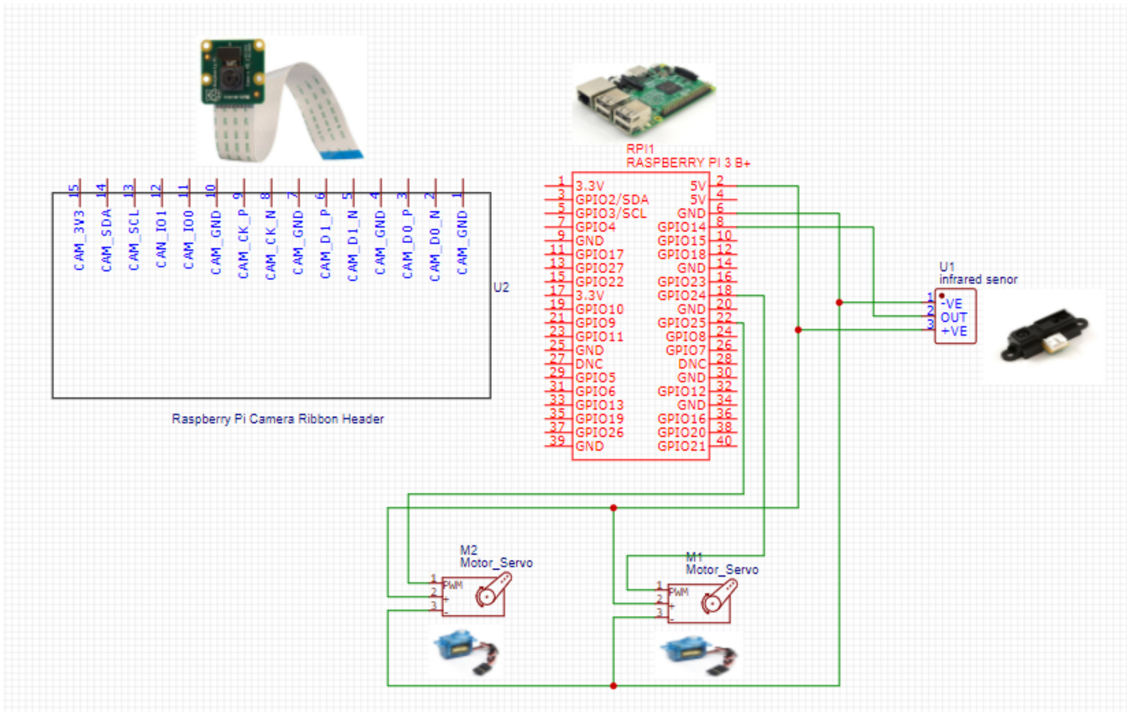


Figure 4: Overall Schematic Diagram

CHAPTER 4. MATHMATICAL MODELLING

4.1 Accuracy based on training and test data

In supervised learning algorithm, train and test data are very important, as they act as a baseline for training the model and testing the model performance in terms of accuracy [18]. The test and train data are a collection of images.

- **Training data**

A training data is a set of dependent variables or parameters that make the machine learn about the environment completely. It is also the data set on which the model is built on.

- **Test data**

The test data is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set. If the test set does contain samples from the training set, it will be difficult to assess whether the algorithm has learned to generalize from the training set or has simply memorized it.

4.2 Evaluation Metrics

The metrics considered for the evaluation of the classification were accuracy, precision, recall and f1-score. Accuracy is the most common metric used for classification and it refers to the fraction of samples that are predicted correctly. Recall (Sensitivity) is the fraction of events that are predicted correctly. Precision (Positive Predictive Value) is the fraction of predicted positive events that are actually positive. The f1_score is the harmonic mean of recall and precision with higher score as a better model.

The mathematical modelling used for determining the evaluation metrics were presented below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where, TP is the number of orange correctly classified as orange

TN is the number of apple correctly classified as apple

FP is the number of orange correctly classified as orange when actually apple

FN is the number of apple correctly classified as apple when actually orange

$$recall = \frac{TP}{TP+FN} \quad (2)$$

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$f1_{score} = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} \quad (4)$$

This is important as it will be used for evaluating the accuracy of the model. This will be used with SVM and KNN algorithm to compare the performance.

4.3 Mean of Training data

Mean is the most common statistic used to measure the centre of a numerical data set. The equation below shows the general formula used to calculated the mean of the training data.

$$mean = \frac{sum\ of\ all\ data\ points}{Number\ of\ all\ data\ points} \quad (5)$$

This is important as it speak of each class centre in a distribution. The extracted pixels of each class collected are represented in RGB color space, where each component of R, G and B were calculated to get its mean. The result of the mean is a 3 by 1 matrix.

4.4 Covariance of Training Data

Another important parameter of a supervised learning algorithm is the variance. Variance is the measure of how two variables are related to each other. The equation below shows the formula used to calculate the variance of the training data.

$$q_{jk} = \frac{1}{N}\sum_{i=1}^{N}(x_{ij} - \mu_j)(x_{ik} - \mu_k) \qquad (6)$$

This is important as it tells us how data points in a specific population are spread out. Each of the class will have different variance and mean.

4.5 Multivariate Gaussian Mixture Model

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal -tract related spectral features in speaker recognition systems [19]. GMM parameters are estimated from training data using the manually calculated method from the two formulae above.

The probability density function of the *d*-dimensional multivariate normal distribution is given by:

$$P(x, \mu, \Sigma) = \frac{1}{\sqrt{2\pi^{d}|\Sigma|}}\exp(\frac{1}{2}(x - \mu)\Sigma^{-1}(x - u)') \qquad (7)$$

Where x and μ are 1-by-d vectors and Σ is a d-by-d symmetric, positive definite matrix.

Note that:

- x is the input test image.
- μ is the mean of the extracted pixels of each class.

- Σ is the covariance of the extracted pixels of each class.

The probability density function is used to describe, at least approximately, any set of (possibly) correlated real-valued random variables each of which clusters around a mean value. This is important when setting the threshold value to segment the target color out from the background.

CHAPTER 5. BINARY CLASSIFICATION USING MULTIVARITE GAUSSIAN MIXTURE MODEL

5.1 Multivariate Gaussian mixture Model

A multivariate Gaussian mixture model is a parametric probability density function that used to cluster a features of a new image through the mean and the covariance of a training data. It is used to detect what type of classes a new image belongs to by calculating the probability and setting the threshold.

5.2 Training and testing procedure

The figure shown below shows the detail steps involved in training of the model as well as testing of the model.

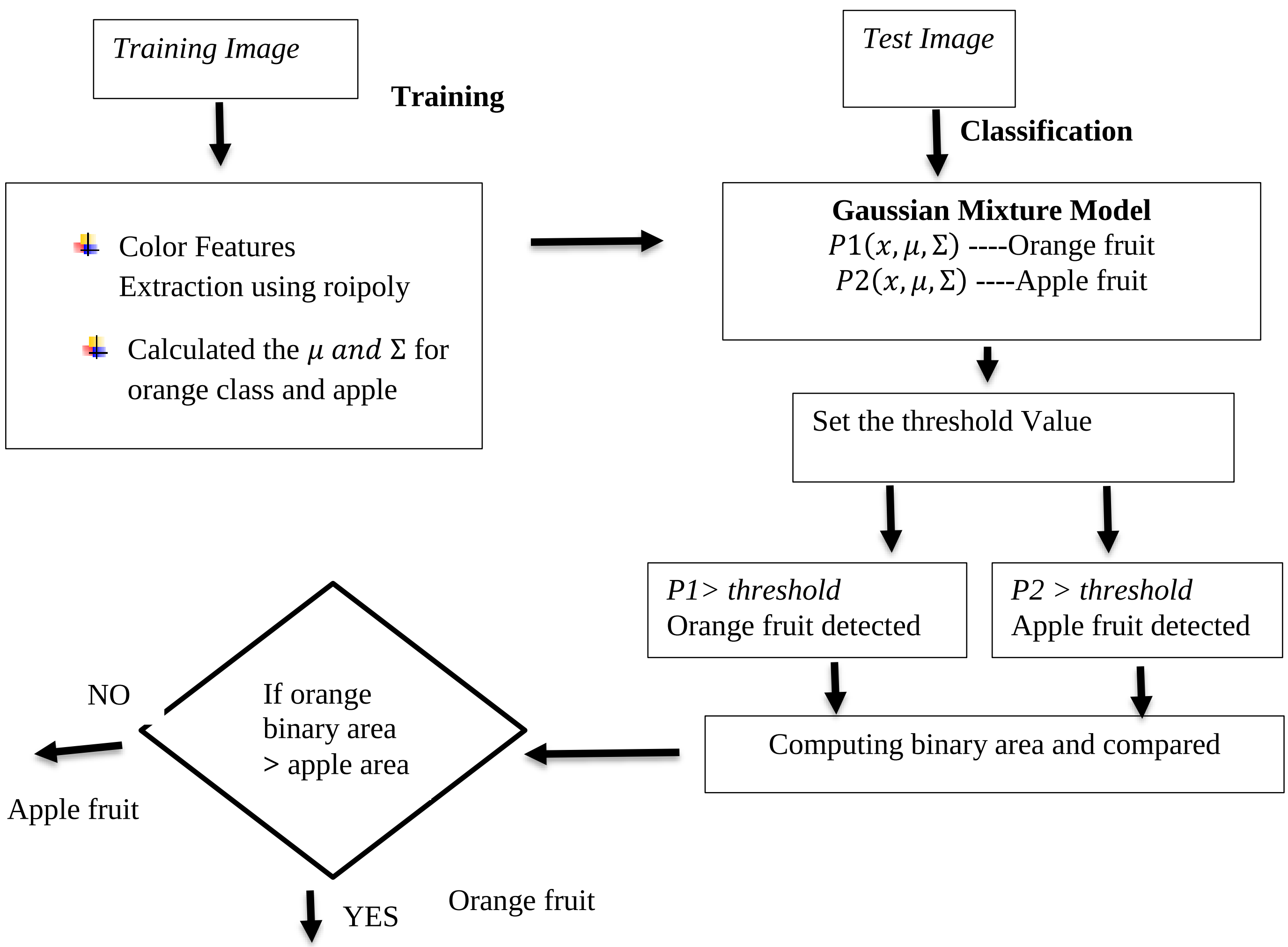


Figure 5: Binary classifier detail training and testing of the model

5.1.1 Dataset

For training and testing of the model, the following dataset were used to train and test the model.

Table 3: shows detail of apple and Orange database

Fruit type	**Number of training image**	**Number of testing image**	**Image size**	**Image format**	**Total image**
Apple	20	8	150 x 150	jpg	28
Orange	20	8	150 x 150	jpg	28

The figure 6 and 7 shown below shows the training image as well as the test image for apple class and orange class

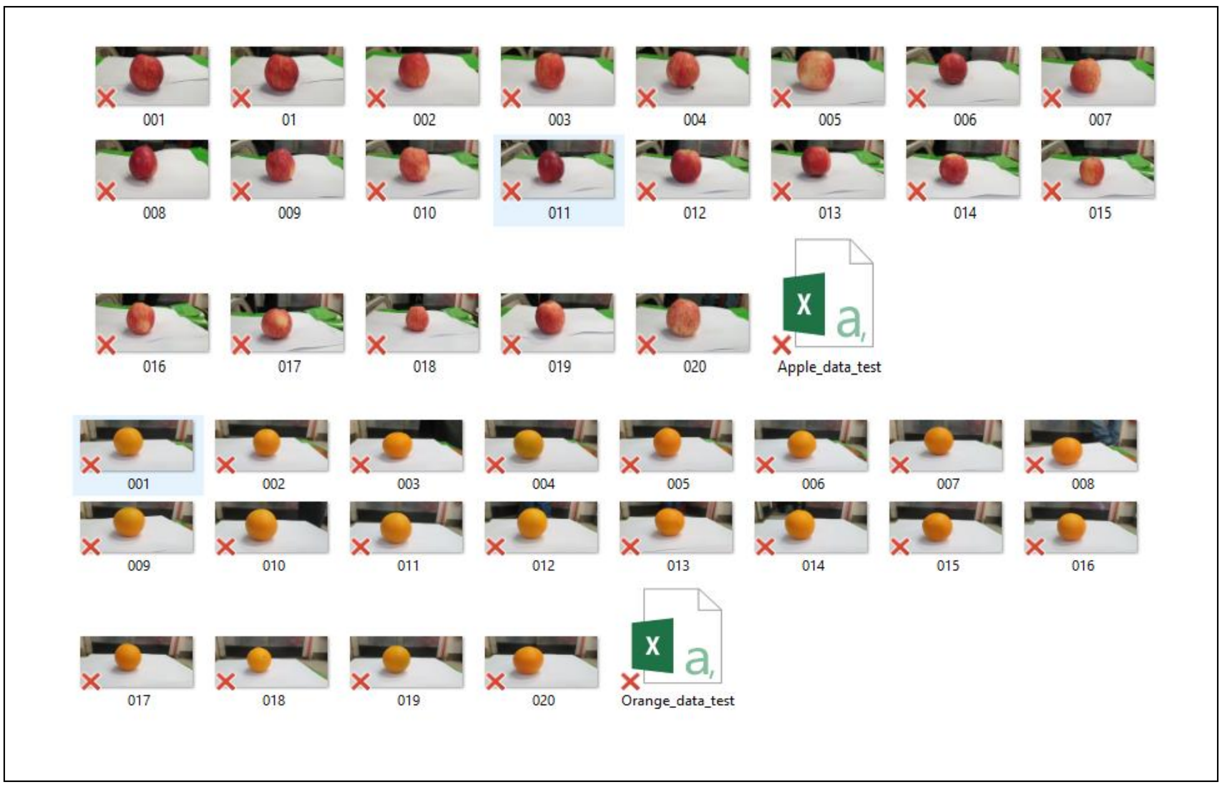


Figure 6: Training Image

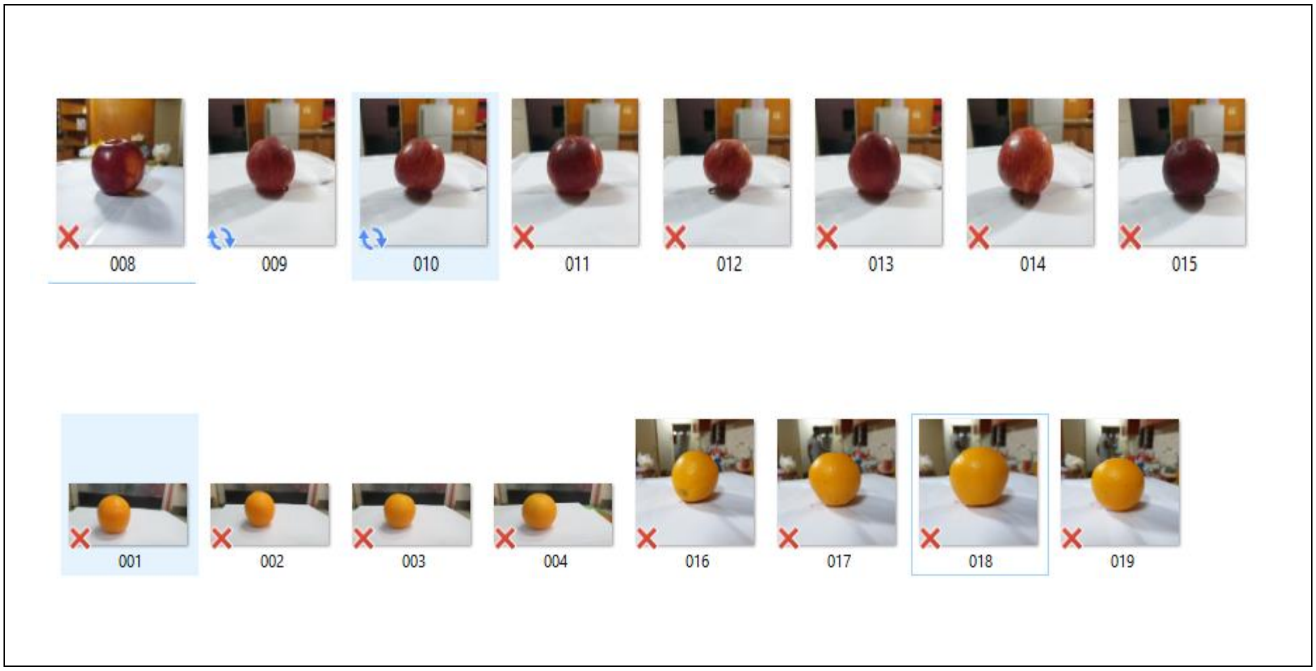


Figure 7: Test image

5.1.2 Image Pre-processing/ Feature extraction

Image Pre-processing

Image Pre-processing is a common name for operations with image at the lowest level of abstraction, both input and output are intensity image [20]. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortion or enhance some image features important for further processing.

In this phase, images were resized into 100×100 pixel resolution for reasonable computation speed. In addition, in this phase for image which were not clear or blurry the contrast adjustment were applied to enhance the image. The goal here was to enhance the entire image before extracting the ROI pixels for each class.

Feature Extraction

In feature extraction, the goal is to extract the color features that contribute most to the classification and to eliminate the rest. For extracting the label color pixel samples of orange class and apple class, first image processing was done to enhance the image and then extracting pixels of region of interest using roipoly command in MATLAB and then stored as csv file. This is important as it will be used to build the model and train the model to classify orange and apple correctly. This was done for all the 28 images for both apple class and orange class. The main purpose of storing the data in CSV file is because python can read file in CSV which is easier to import the data in python.

The figure shown below shows the collected color samples of orange class and apple class samples from the training images.

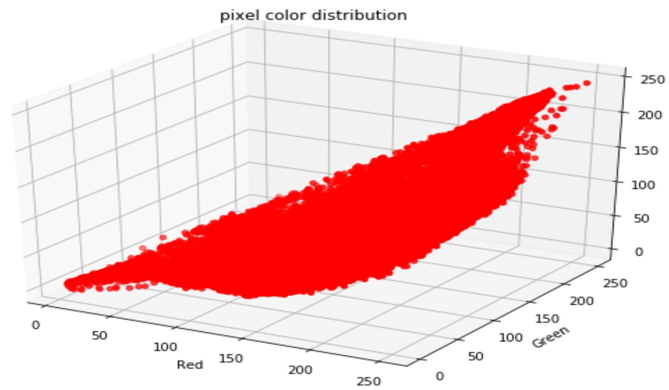


Figure 8: Color distributions of apple class

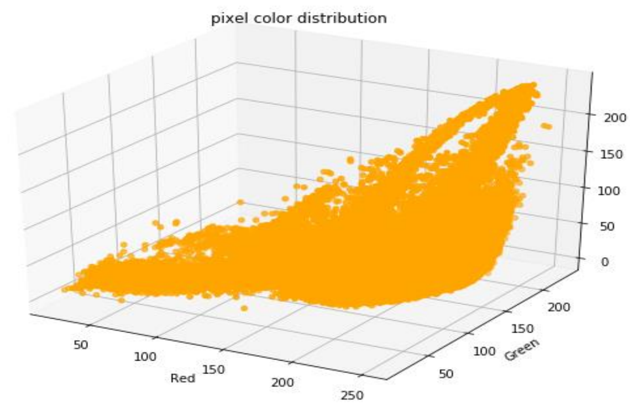


Figure 9: Color distributions of orange samples

5.1.3 Calculating μ *and* Σ

After extracting the ROI label pixels of orange and apple samples, the mean and the covariance were calculated. The mean and the covariance were important as it will be used in Multivariate Gaussian mixture model to calculate the probability of a new test image. Based on the calculated probability, the threshold will be set to segment only the fruit image from its background. The result will be a binary image. The goal of the statistical model ($\boldsymbol{\mu}$ ***and*** $\boldsymbol{\Sigma}$) was to represent the statistics of each class, which allows the classes to be separated from each other.

The table 4 given below present the calculated mean and covariance of orange class and Apple class.

Table 4: Calculated mean and covariance of orange and apple class

Apple		**Orange**	
Mean	$mu = [\,167.978 \quad 76.6507 \quad 65.574\,]$	Mean	$mu = [\,200.224 \quad 123.543 \quad 32.4314\,]$
Covariance	$Covarianc = \begin{matrix} 1068.4 & 1003.08 & 640.939 \\ 1003.08 & 1468.42 & 999.001 \\ 640.939 & 999.001 & 784.846 \end{matrix}$	Covariance	$Covariance = \begin{matrix} 746.829 & 613.434 & 114.998 \\ 613.434 & 724.818 & 380.201 \\ 114.998 & 380.201 & 698.326 \end{matrix}$

5.1.4 Calculating the probability using Gaussian Mixture Model

To calculate the probability of the new test image, the previously calculated μ and Σ from the extracted features of each class was used. To obtain the probability of a new test image the given formula below was used to calculate.

$$P(x, \mu, \Sigma) = \frac{1}{\sqrt{2\pi^d |\Sigma|}} \exp(\frac{1}{2}(x - \mu)\Sigma^{-1}(x - u)')$$

This was done so that when setting the threshold value the model is able to detect which classes of the two fruit the new image belongs to. The result of this is a binary image where the background is represented as 0 and the foreground is represented as 1.

5.1.5 Setting the threshold Value

To segment an apple and orange image from its background, the threshold was set based on the calculated probability obtained from the Multivariate Gaussian formula. The segmented image then converted into binary and applied morphological operation to fill the possible holes inside the apple and orange fruit. Then, the binary image is multiplied with original image to segment the apple image as shown below in the *fig 9.*

Figure 10 below show the outcome of the model after setting the threshold.

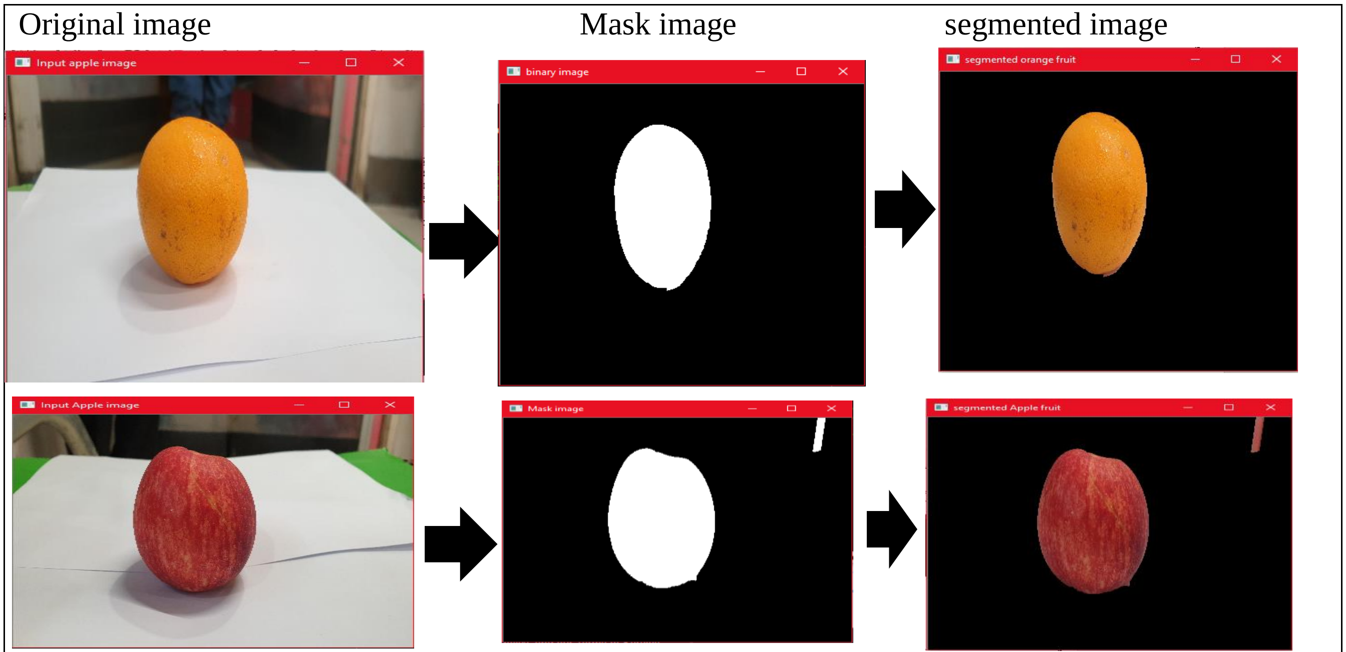


Figure 10: Segmented binary image

4.3.7 Calculating the size

To calculate the size of the fruit, the number of pixels covering the object was counted in binary after setting the threshold value in the previous step.

To compute the Pixels to total image area percentage ration, the formula given below has been utilized to estimate the area.

$$Area_{ratio} = \frac{Pixels}{Image_{area}} \times 100\%$$

5.1.6 Classification

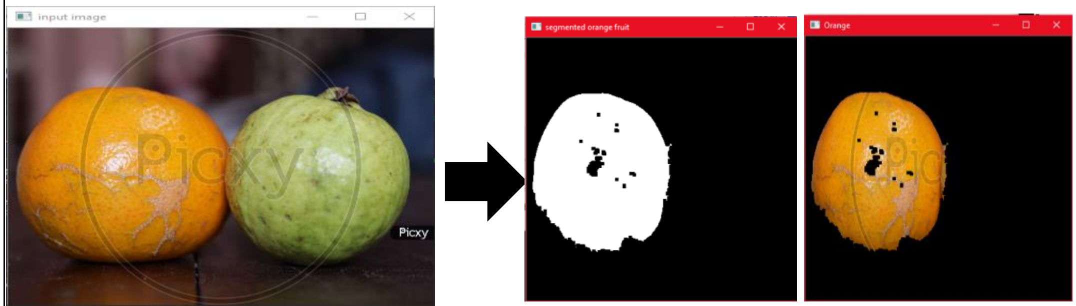
For classification, a simpler method was utilized to classify orange and apple fruit. This method counts the non-zero number of pixels in binary to classify the fruit after a threshold is set for both apple and orange. Since, both apple and orange classes have different mean and covariance, therefore their probability will also be different. Setting the threshold will determine which fruit will be detected base on the mean and covariance. The fruit with the highest non-zero number of pixels greater than the other fruit will classify as the selected fruit.

The figure 11 below illustrates the testing of the model to see if the model can differentiate and classify orange and apple fruits.

Orange fruit

Input image

Processed image after classification



number of pixels for orange:
23.15625
number of pixels for apple:
11.324375

Calculated non-zero-pixel value for both orange and apple for the input image above is shown below

Apple fruit

Input image

Processed image after classification

number of pixels for orange:
2.461875
number of pixels for apple:
12.767500000000002

Calculated non-zero-pixel value for both orange and apple for the input image above is shown below

Figure 11: Classification of orange and apple fruit

CHAPTER 6. SVM AND KNN CLASSIFIER

6.1 Support Vector Machine (SVM)

A support vector machine is a supervised machine learning model that uses classification algorithm for two group problems [21]. It is a type of algorithm that trained on a set of labelled training data and predicting a new image to one category or the other, making it a non-probabilistic binary linear classifier. Because it is a fast and dependable classification algorithm that performs very well with a limited amount of data, it is often used in many applications such as in text and hypertext categorization, classification of images etc.

Given a set of Labelled training data, an SVM takes data point and output the hyperplane that best separates the classes. This is often called the decision boundary. The goal here is to come up with a hyperplane that maximizes the margin from the nearest training data point of any class, since in general the larger the margin, the lower the generalization error of the classifier. Therefore, whenever a new image is fitted it will mapped that image in to that decision boundary margin and prediction will be carryout base on which side of the gap the image fall. In this way classification of new image is done. This type of hyperplane decision boundary is called linear classification.

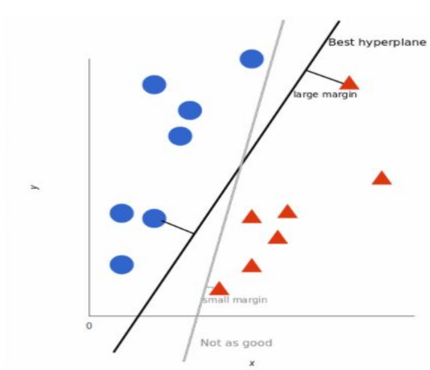


Figure 12: Linear classification

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. This is very important when dealing with non-linear training data.

6.2 K-Nearest Neighbors

KNN [22] is also one of the simplest of classification algorithms available for supervised learning. The idea is to search for closest match of the test data in feature space. The differences between SVM and KNN is that, SVM assumes there exist a hyper-plane separating the data points, while kNN attempts to approximate the underlying distribution of the data in a non-parametric fashion. In kNN, the approach for classification of a new test image depend on voting of the nearest neighbor. For instance, considering the figure below, let say k = 3, the green dot can be classified as belonging to the triangle class since it has two Red and one Blue.

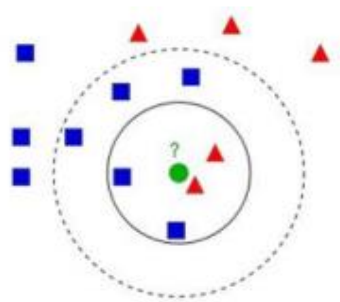


Figure 13: KNN Classifier

6.3 Training and testing procedure

The figure shown below shows the detail steps involved in training of the SVM and kNN classifier as well as testing of the model.

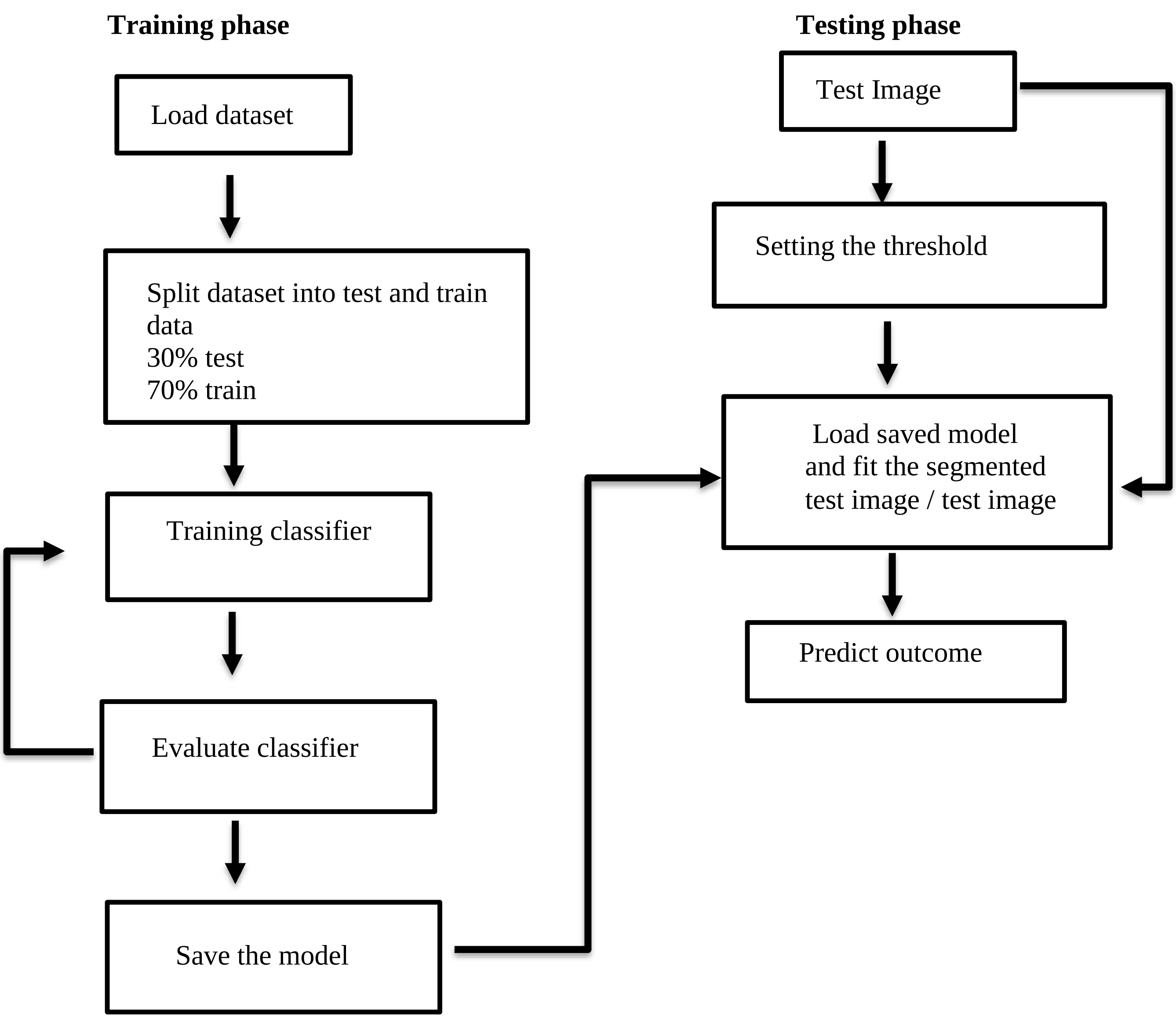


Figure 14: SVM and KNN detail training and testing of the model

6.3.1 Dataset

For training and testing images, a total of 640 images of orange and apple were gathered for training the model. Upon the 640 images, 30 percent of the total image were used for testing. The training and testing image should not mixed together, as this will pose difficulties in assessing whether the algorithm has learned to generalize from the training set or has simply memorized it.

Table 5: Apple class and Orange class database

Fruit type	**Number of training image**	**Image size**	**Image format**
Apple	640	100 x 100	jpg
Orange	640	100 x 100	jpg

6.3.2 Training classifier

For training each of the classifier, the sklearn package in python was utilized to perform the operation. This package or library was implemented for easier computation and analysis. So after splitting dataset into xtrain, ytrain, xtest and ytest, the model was build based on the xtrain and ytrain data (features and label). In this phase, the model tries to fit the given data such that it creates a decision boundary between the two classes. Once the fitting was done, the model then predicts the xtest data (features) and crosscheck with the ytest (labels) data for accuracy. Based on this, the accuracy was determining and the confusion matrix was obtained.

6.3.3 Evaluating algorithm

Confusion matrix, precision, recall and F1 score were the most commonly used metrics to evaluate the performance of the classifier. Based on the performance, each classifier were compared and the superior classifier was implemented to test in an unseen cases. Also in this phase, adjustment were also done to improve the performance of the classifier once the accuracy obtained has a satisfactory result. This can only be done once the data used is a non-linearly separable data. This involves setting the kernel tricks and increasing the number of K. Once the accuracy obtain has greater result, the model was saved using pickle library package.

6.3.4 Test image / segmented image and predictions

To test the model with real time image, there are two ways that can be done.

1. Fitting the captured image directly to the saved model
2. The test image was first threshold before fitting the image to the saved model. This was done to obtained best result when predicting, so that it does not fall between the two classes where the decision line was drawn. This can make the classifier find it very hard to predict and it can lead to misclassification.

The figure 15 below shows test image used to fit to the classifier model

Test Image | Segmented test image to be fitted to the classifier

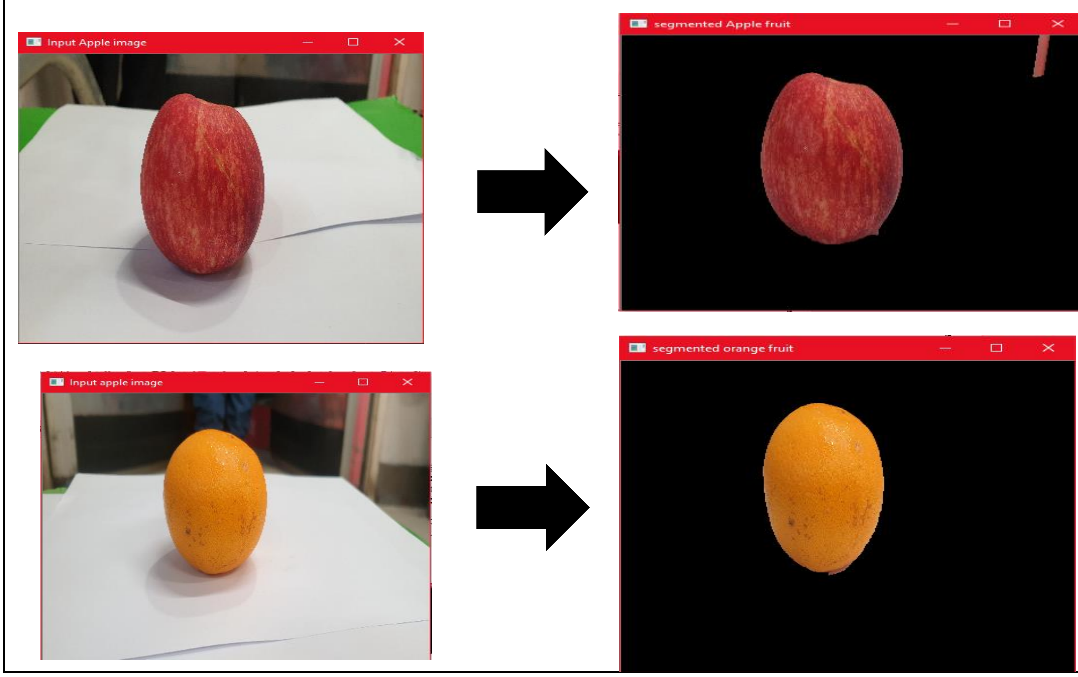


Figure 15: Segmented test Image

CHAPTER 7. HARDWARE RESOURCES AND SOFTWARE RESOURCES

This section outlines the hardware resources used to achieved the scope of the project

7.1 Hardware Resources

7.1.1 Raspberry pi

The Raspberry Pi 3 B+ is like a little Computer that can do things like spreadsheets, word processing, and games that your desktop PC normally does. The design is based on a Broadcom BCM2835 SoC with an ARM1176JZF-S 700MHz processor, VideoCore IV GPU, and 512Mb of RAM. The model does not include a built-in hard disk drive or solid state drive, but relies on a microSD card for booting and storage in the long term. This board is intended to run Linux kernel based operating systems.

The Raspberry Pi compared to other microprocessors has been chosen due to the highlights:

- Improved power management that improves USB support and allows for high-drain devices such as external USB hard drives.
- Larger bank of accessible GPIO ports compared to Model B board allows for direct hardware device control.
- Smaller, compact design

Features:

- More Energy Efficiency (Less Power Required)
- Improved Power Management: Manage More Devices from Your Pi!
- Bigger and Better projects via an Expanded GPIO Header (40 pins vs. 26)
- Increased connectivity - 2 Extra USB ports (making a total of 4) and a new 4-pole connector replace the existing analogue and composite video port on the Model B.

Specifications:

- Dimensions: 85mm x 56mm
- Chip: Broadcom BCM2835 SoC full HD multimedia applications processor
- CPU: 700 MHz Low Power ARM1176JZ-F Applications Processor
- GPU: Dual Core VideoCore IV® Multimedia Co-Processor
- RAM: 512 MB SDRAM @ 400 MHz
- Storage: MicroSD
- USB 2.0: 4x USB Ports
- Ethernet: 1x 10/100mb Ethernet RJ45 Jack
- Video Connections: HDMI, Composite RCA (shared with audio jack)
- Supported Resolutions 640×350 to 1920×1200, including 1080p, PAL & NTSC standards
- Audio: Multi-Channel HD Audio over HDMI, Stereo from 3.5 mm jack Operating Systems: Raspbian, RaspBMC, Arch Linux, Risc OS, OpenELEC, Pidora
- Power Draw / voltage: 600mA up to 1.8A @ 5V
- GPIO 40
- Other Connectivity: 1x CSI-2 for Raspberry Pi camera modules
- 1x DSI for Raspberry Pi displays
- Power Source: 1x Micro-USB

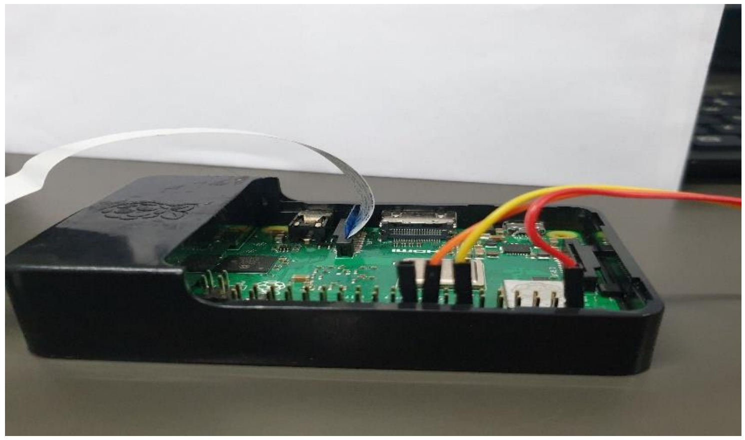


Figure *16*: Raspberry pi

7.1.2 Conveyor belt (Full package including Motor etc.)

Conveyor belts are used mainly in industrial settings that include tripping mechanisms such as trip cords along the length of the conveyor. They are commonly used in material handling such as moving boxes. Belt Conveyers are used to transport large volumes of resources and agricultural materials. The main intention of having the conveyor in this project is to transport the fruits to its respective boxes. This will reduce the amount of time when parking the fruits.

Specification:

- Material: MS
- Conveyor speed: As per customer's requirement
- Motor: Default 12W
- Belt: PVC
- Dimension: 550x130x160 mm

Features:

- Can set conveyor as per the requirement
- Motor speed is by default 12W
- It has PVC belt

Figure *17*: Conveyor belt

7.1.3 Servo Motor

A servomotor is a rotary actuator or linear actuator that allows for precise control of angular or linear position, velocity and acceleration. It consists of a suitable motor coupled to a sensor for position feedback. It also requires a relatively sophisticated controller, often a dedicated module designed specifically for use with servomotors. In this project the servo motor was used to sort fruit after classification is done. The main intention is to separate the fruits coming on the conveyor belt.

Features:

- High speed and high torque digital servos
- Metals gears for maximum strength and durability

Specification:

- Length 40.77mm
- Width 19.7mm
- Height 42.9mm
- Weight 55g
- Servo size: standard
- Servo Torque 13(6V)/10.5(4.8V)kg.cm
- Servo Horns included

- Dual Bearing
- Servo Arduino compatible
- DC voltage : 4.8V to 7.2V
- DC max current 1.2A
- No load current 170mA
- Full load current 1.2A
- No load RPM 58RPM

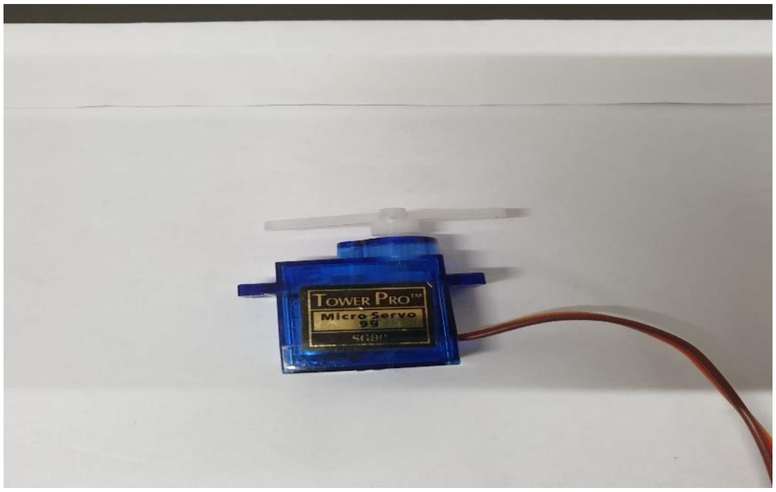


Figure *18*: Servo Motor

7.1.4 Servo motor driver

In this project, SunFounder PCA9685 16 Channel 12 Bit PWM Servo Driver was used to control the switching of the servo motor. It has the following specification.

- Contain an I2C-controlled PWM driver with a built-in clock. It means, unlike the TLC5940 family, you do not need to continuously send it signals tying up your microcontroller; it is completely free running!
- 5V compliant, which means you can control it from a 3.3V microcontroller and still safely drive up to 6V outputs (this is good when you want to control white or blue LEDs with 3.4+ forward voltages)

- Support using only two pins to control 16 free-running PWM outputs – you can even chain up 62 breakouts to control up to 992 PWM outputs
- 3 pin connectors in groups of 4, so you can plug in 16 servos at one time (Servo plugs are slightly wider than 0.1" so you can only stack 4 next to each other on 0.1" header
- 12-bit resolution for each output - for servos, that means about 4us resolution at 60Hz update rate

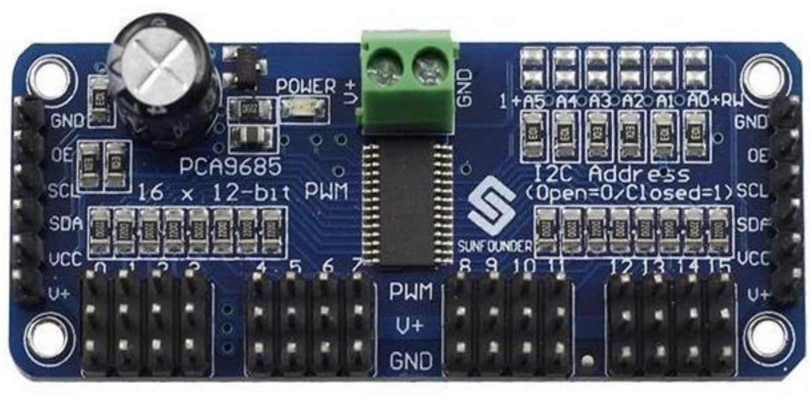


Figure *19*: Servo motor driver

7.1.5 Infrared sensor

The GP2Y0A41SK0F is a short range Sharp analog distance sensor. The detection range is (4 cm to 30 cm). The shorter range gives you higher resolution measurements, and the lower minimum detection distance makes this sensor great for detecting very close objects. The distance is indicated by an analog voltage, making this sensor very easy to use.

The used of infrared sensor is to sense the fruit coming on the conveyor belt so that the running conveyor belt can be interrupted at any instant when a fruit is detected and give chances for the camera to take a snap shot of the fruit.

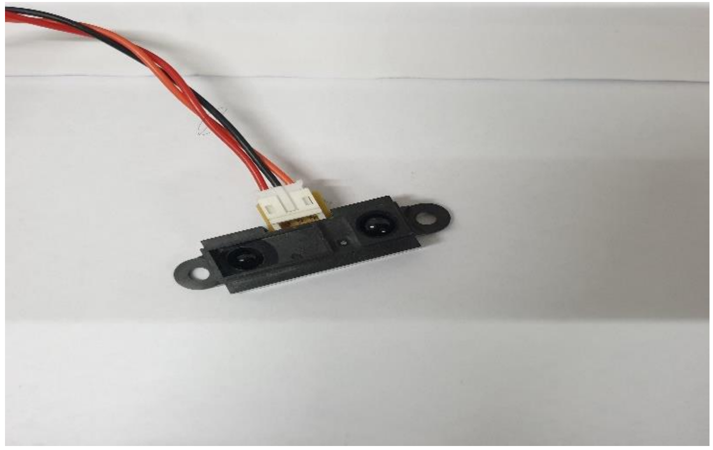


Figure *20*: Infrared sensor

7.1.6 Raspberry pi camera

The Raspberry Pi Camera Module v2 is a high quality 8 megapixel Sony IMX219 image sensor custom designed add-on board for Raspberry Pi, featuring a fixed focus lens. It's capable of 3280 x 2464 pixel static images, and also supports 1080p30, 720p60 and 640x480p60/90 video. It attaches to Pi by way of one of the small sockets on the board upper surface and uses the dedicated CSi interface, designed especially for interfacing to cameras. The board itself is tiny, at around 25mm x 23mm x 9mm. It also weighs just over 3g, making it perfect for mobile or other applications where size and weight are important. It connects to Raspberry Pi by way of a short ribbon cable.

In this project, the used of the raspberry pi camera is to capture the fruits when coming on the conveyor belt. This is important as this fruit image were used for classification purposes.

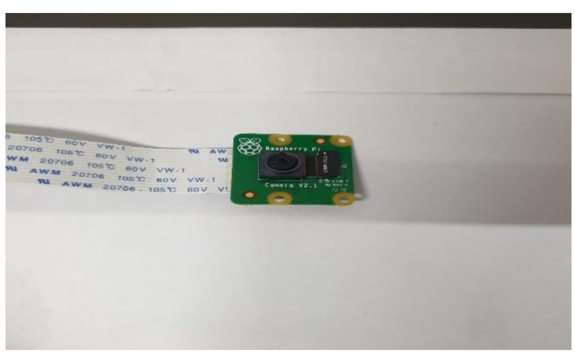


Figure *21*: Raspberry Pi camera

Figure 22 below shows the developed machine vision system

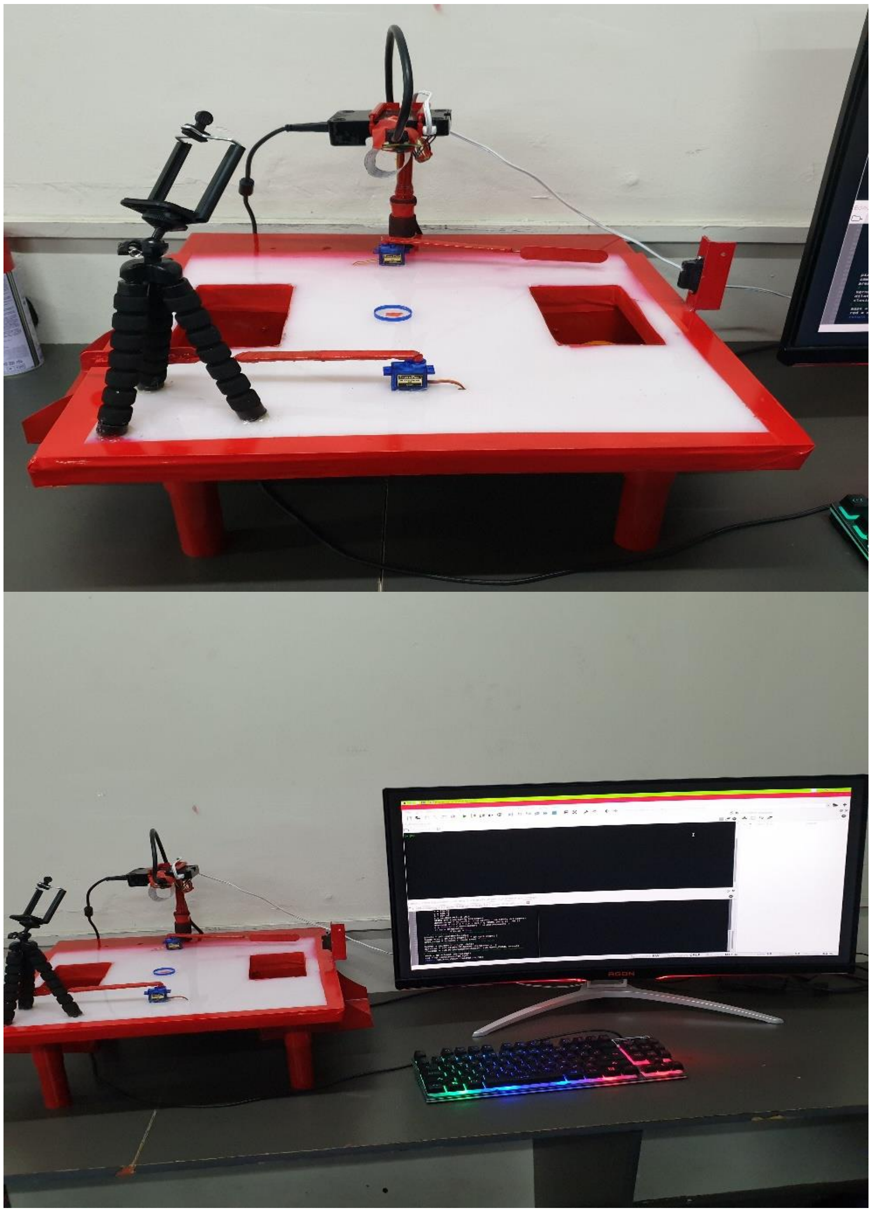


Figure 22: Developed machine vision system

7.2 Software Resources

In order to test and identify the errors present in the system, various software resources are utilize to correctly test the program script and configure before transferring it to the real world for implementation. The table 6 below show the list of software used for implementation of the project.

Table 6: Software Resources

Software	**Description**
MATLAB	MATLAB software was utilized for extracting the ROI of all the pixels of the apple class and orange class and stored it as CSV file.
Python	Python is a high level programming language, which were used for developing desktop GUI applications, websites and web applications. Python software was utilized for compiling the whole working process of the system. This includes the image processing method, filter and the implementation of various classifier as binary classification, SVM, KNN and random forest.

7.3 Modern IT tools

The data analysis of this vision based system can be done with MATLAB or Python software. Since Raspberry Pi uses Python language, the analysis for this project is done using Anaconda, a free and open-source Python distribution for computing data science and machine learning. The major reasons for using Anaconda or Python for the analysis of this project is described below.

- **Open Source**

  Python is open source that means it is free and python open source are not different programming languages and it is available to everybody to use it freely for any kind of purpose.

- **Essential for Data analysis**

  Python is a programming language that can be used in the development of desktop and web applications. It is very useful for complex numeric and scientific applications such as data analysis for vision based system. Since it is open-source, it uses community-based model for development.

- **Machine learning libraries**

  Anaconda has highly optimized libraries for machine learning such as scikit, numpy, scipy, matplotlib, Pandas and tensorflow. The scikit package for example, is a machine-compatible learning tool and scipy is a scientific computing library that contains modules for optimization, Fast Fourier Transform and Signal Processing. Also in comparison to other platforms, Anaconda is easier to add libraries to it.

- **High Readability**

  The Python codes in Anaconda are usually readable, hence making it easier to process and maintain. The troubleshooting is also easier due to python documentation and lots of troubleshooting content on websites such as Stack Overflow, kaggle and github.

CHAPTER 8. IMPACT OF THE SYSTEM

This section outlines the environmental impact, social impact and ethical issues regarding the design project.

8.1 Environmental impact:

Fruit sorting machine can have both advantages and disadvantages on the environment depending on the material used and the hazard the machine causes.

Advantage of Fruit sorting machine on environment.

1. This machine is environmental damage free.
2. Better control of defect fruit.
3. The material used were environmental damage free. This means it does not contribute to any pollution that is harmful to the environment
4. Much safer to the surrounding environment since manual operation is not being practiced.

Disadvantage of fruit sorting machine on environment.

- Once damage parts of the machine are not disposed off properly, it can create rubbish pollution to the environment.

8.2 Social Impact

The social impact of this system includes;

- Speed and accurate sorting can be achieved, meaning it will help the local farmers to produce good quality fruits to sell while at the same time it will lessen the time for them to choose among the best fruit and the defect ones.
- Greater product stability, meaning consistency of products in the international markets.
- Cost effective, meaning it will provide income for the local farmer through their quality products.

- Job opportunities, meaning it will provide opportunity for locals to go and train how to manage and operate the machine.
- Social interaction, meaning it will bring local farmers together to discuss new ideas and innovations.

8.3 Ethical Issues:

Since this machine will be used by local farmers who does not have proper knowledge of electrical principal, this can create issues regarding repair, troubleshooting and fixing of the machine when fault arises.

Some common problems or conflicts are;

- It can cause death since the machine is operating on power. This can arise if workers are not following the proper instruction given or safety procedure.
- It can cause conflict among local farmers if the machine is not working. This is a common problem in the society today. People turn to blame each other for somebody’s fault and this is a real problem.
- At times if the machine needs repair, people turn to neglect to work together for servicing the machine.

Solution to ethical Issues:

- The machine will be given safety measures as on how to operate, troubleshoot and do the repair if there is any technical problem arise. In addition, a well train professional will be assign for daily check-up and servicing of the machine.

CHAPTER 9. RESULT AND DISCUSSION

The table given below show the accuracy of all the classifier used.

For SVM, and KNN, the number of samples were taken from the x_test data after splitting. The accuracy is obtaining after predicting the x_test data and validating it with y_test data. For Binary classifier, since the method used is different, the testing for accuracy was done individually for each image in real time. There were total of 30 fruits were gather for testing. Each of the 30 fruits gathered were not included in the training phase as this may leads to a bias result. The accuracy is obtaining by manually calculating how many fruit classified correctly using the evaluation metrics given above.

The table given below show the performance of each classifier.

Table 7: Classification performance of SVM, kNN and binary classifier

<table>
<tr><th>Classifier</th><th>Evaluation metrics</th></tr>
<tr><td>SVM</td><td>Accuracy: 1.0
Confusion matrix [[150 0]
[0 142]]

precision recall f1-score support
class 0 1.00 1.00 1.00 141
class 1 1.00 1.00 1.00 151
micro avg 1.00 1.00 1.00 292
macro avg 1.00 1.00 1.00 292
weighted avg 1.00 1.00 1.00 292</td></tr>
<tr><td>KNN</td><td>Accuracy: 0.9622641509433962
Confusion matrix [[26 2]
[0 25]]

precision recall f1-score support
class 0 1.00 0.93 0.97 30
class 1 0.92 1.00 0.96 23
micro avg 0.96 0.96 0.96 53
macro avg 0.96 0.97 0.96 53
weighted avg 0.97 0.96 0.96 53</td></tr>
</table>

Binary Classifier	Accuracy: 1.0 Confusion matrix [[150 0] [0 142]] precision recall f1-score support class 0 1.00 1.00 1.00 141 class 1 1.00 1.00 1.00 151 micro avg 1.00 1.00 1.00 292 macro avg 1.00 1.00 1.00 292 weighted avg 1.00 1.00 1.00 292

The above results were obtained using the evaluation metrics formula given in mathematical section. From the table it can be seen that SVM and binary classifier has the highest accuracy compared to kNN classifier. In terms of performance evaluation of the SVM and binary model, there is no misclassification as can be observed from the classification result compare to kNN that has 2 misclassifications.

Table 8: Show the computation time of all the classifier when testing in real time.

Classifier	**Number of fruits processed**	**Computation time**
SVM	39	60 sec
KNN	40	60 sec
Binary Classifier	25	60 sec

The above results show the number of fruits process within 60secs. The results were obtained by creating a for loop function that runs for 60 sec. Therefore, in that period, the camera continuously captured image one at a time and processing each image captured. This was done 10 number of times to reduce random error. The average number of fruit process was recorded in the table above. The result shows that kNN has output both the classifier in terms of how many fruits captured within the timeframe followed by SVM with 39 fruits and binary classifier with 25 fruits.

From the result above, it can be observed that the classification performance for SVM classifier as good as the manual expert based technique in terms of accuracy and computation time. This is because SVM can handle non-linear training dataset and in real-time not all images are clear, some affected by noise which SVM along can handle by setting the kernel tricks.

For binary classifier misclassification may occur when similar object from background have similar color pattern. However, this is overcome by setting the background to different color pattern.

In real time implementation, other factors such as changes in ambient light, camera resolution and distance of the camera also contributes to affect the system performance.

The classification accuracy of proposed machine vision based system and the classification accuracy achieved by Zenan Lin et. al [11] were presented in Table 9

Classifier	Accuracy
Proposed method (SVM and GMM)	100 %
Zenan Lin et. al (HSV)	93.73 %

CHAPTER 10. CONCLUSION, LESSON LEART AND RECOMMENDATION

The results and observation shows that SVM and binary classifier perform really well when testing in real time. Although kNN has the highest number of fruits processed, when tested in real time a lot of misclassification has been observed. This is because the data is a homogenous type meaning it is a non-linear datasheet. In addition, for most practical problems, kNN is a bad choice because it scales badly - if there are a million labelled examples, it would take a long time to find K nearest neighbours.

The automated sorting system for Apple and orange was designed, constructed and tested in this study. The system results verified that the system can be used in real-time which lay foundation for fully automatic production of fruits. The overall system was portable, inexpensive and the results show that the sorting system met the 100% reliability requirement.

LESSON LEARNT

While working on this project, there were many positive things that I have learnt. Not only did I get comfortable in the lab and got used to working with electrical equipment, but I also learned how to use programming software like Anaconda or Python. The design of this project has taught me the following key points:

- Learn to analysed image data and represent them.
- Implementing SVM and kNN classifier and tunning their parameters to get best performance.
- Think logically and responsible while solving problem.
- To be consistent when carryout project.
- how to used online resources like (GitHub and YouTube) to store information and display project.
- Learned python programming using software called Anaconda
- Learned how to apply management tools in the project

- Learned how to do literature study and write reports based on criteria given
- Learned about the safety principles that need to be considered to get optimal results or output

RECOMMENDTION FOR FUTURE RESEARCHER

The future recommendation that can be done for this project are as follows:

- The system can be further improved by finding the defects area of the fruit and considered the weight and maturity to make the system robust. Also future researcher should work on grading of the fruit base on the above mention features.
- The system can be further improvise by incorporating a robotic arm for picking and placing of the fruit.
- Further research should be carried out on the security of the system to allow only the authorized people to use.
- The system could be enhanced using embedded devices with faster processor than Raspberry Pi.

References

[1] M. J. Villaseñor-Aguilar, J. E. Botello-Álvarez, F. J. Pérez-Pinal, M. Cano-Lara, M. F. León-Galván, M.-G. Bravo-Sánchez and A. I. Barranco-Gutierrez, “Fuzzy Classification of the Maturity of the Tomato Using a Vision System,” *Journal of Sensors,* vol. 2019, p. 1–12, 7 2019.

[2] V. Vujovic and M. Maksimovic, “Raspberry Pi as a Wireless Sensor node: Performances and constraints,” in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO),* 2014.

[3] M. F. Brown, “Value chain study of Tropical Fruits in Fiji,” *Facilitating Agriculture Commodity in the Pacific [FACT],* 2017.

[4] S. Yogitha and P. Sakthivel, “A distributed computer machine vision system for automated inspection and grading of fruits,” in *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE),* 2014.

[5] L. C. Dorado, J. I. Aguila and R. B. Caldo, “Smart Farm: Automated Classifying and Grading System of Tomatoes using Fuzzy Logic,” *LPU-Laguna Journal of Engineering and Computer Studies* , vol. 3, no. Octomber 2016, 2016.

[6] D. Pise and G. D. Upadhye, “Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques,” 2015.

[7] B. Jarimopas and N. Jaisin, “An Experimental Machine Vision System for Sorting Sweet Tamarind,” *Journal of Food Engineering,* vol. 89(3), pp. 291-297, 2016.

[8] C. S. Nandi, B. Tudu and C. Koley, “Machine Vision Based Automatic Fruit Grading System using Fuzzy Algorithm,” *2014 International Conference on Control, Instrumentation, Energy & Communication(CIEC),* 2014.

[9] D. W. Sun, “Computer vision - An objective, rapid and non-contact quality evaluation tool for the food industry,” *Journal of Food Engineering* , p. 61, 2004.

[10] N. W. Seng, A. Shahrizam and A. Ghani, “Vision Based Smart Sorting Machine,” pp. 1-4, 2019.

[11] Z. Lin, J. Shao, X. Zhang, X. Deng and T. Zheng, “Multiclass Fruit Packing System Based on Computer Vision,” *Journal of physic: Conference Series,* vol. 1449, pp. 1-5, 2020.

[12] P. Avalekar and L. S. Admuthe, “Tomato grading system based on colour model by using neural network,” *international Research Journal of Engineering and Technology,* vol. 05, 2018.

[13] V. Pavithra, R. Pounroja and B. S. Bama, “Machine vision based automatic sorting of cherry tomatoes,” in *2015 2nd International Conference on Electronics and Communication Systems (ICECS),* 2015.

[14] S. S. Deulkar and S. S. Barve, “An Automated Tomato Quality Grading using Clustering based Support Vector Machine,” in *2018 3rd International Conference on Communication and Electronics Systems (ICCES),* 2018.

[15] D. Pise and G. D. Upadhye, “Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques,” in *2018 International Conference on Smart City and Emerging Technology (ICSCET),* 2018.

[16] R. A. Kumar, V. S. Rajpurohit and V. B. Nargund, “A neural network assisted machine vision system for sorting pomegranate fruits,” in *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2017.

[17] C. S. Nandi, B. Tudu and C. Koley, “Machine Vision Based Automatic Fruit Grading System using Fuzzy Algorithm,” *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC),* 2014.

[18] A. W. . N. K. S. Pan Junjun and M. A. Mughal , “Recognition of Ripe, Unripe and Scaled Condition of Orange Citrus Based on Decision Tree Classification,” *2018 International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018* , no. 2018, 2018.

[19] C. S. Nandi, B. Tudu and C. Koley , “An Automated Machine Vision Based System for Fruit Sorting and Grading,” *2015 Sixth International Conference on Sensing Technology (ICST),* no. 2015, 2015.

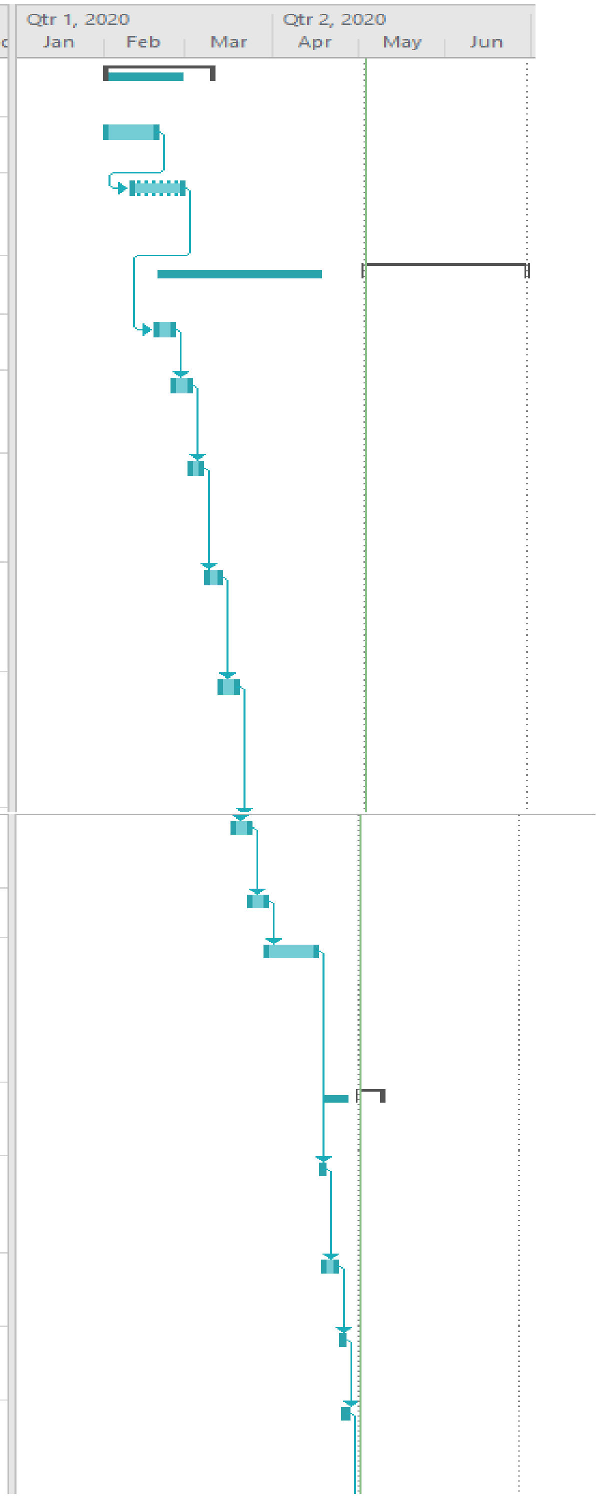
[20] D. . K. KIM , “COLOR DETECTION USING GAUSSIAN MIXTURE MODEL,” *Journal of Theoretical and Applied Information Technology* , 2017.

[21] V. Wasule and P. Sonar, “Classification of brain MRI using SVM and KNN classifier,” in *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS),* 2017.

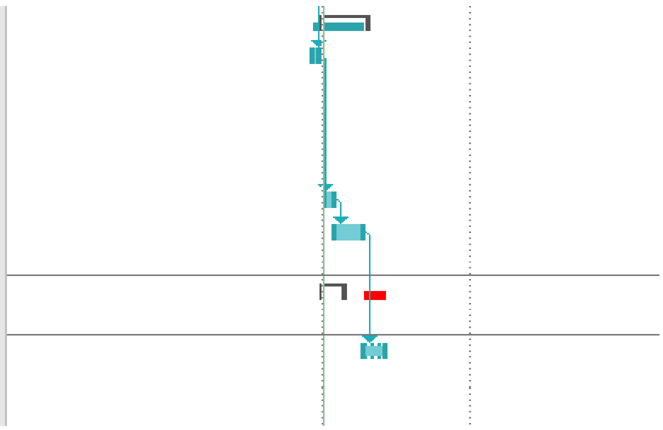
[22] H. I. Dino and M. B. Abdulrazzaq, “Facial Expression Classification Based on SVM, KNN and MLP Classifiers,” in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019.

APPENDIX A: GANTT CHART

Task Name	Duration	Start	Finish	Prec
Project Synopsis and Expression of interest	**28 days**	**Sun 02/02/20**	**Tue 10/03/2**	
Selection of topic with the project supervisor	14 days	Sun 02/02/20	Wed 19/02/2	
Liaise with the project supervisor to produce a synopsis of the project	14 days	Tue 11/02/20	Fri 28/02/2	2
Proposal and literature publication review	**42 days**	**Sun 03/05/20**	**Mon 29/06/2**	
Learn about Machine learning	4 days	Thu 20/02/20	Tue 25/02/2	3
Learn about different image pre-processing technique and filters	4 days	Wed 26/02/20	Mon 02/03/2	5
Learn about different machine learning algorithm to model or cluster an image signal	4 days	Tue 03/03/20	Fri 06/03/2	6
learn about different feature extraction method and familar with Gaussian mixture model	4 days	Mon 09/03/20	Thu 12/03/2	7
Identify the drawback of previous system and proposing new approach and combining with existing approaches	4 days	Fri 13/03/20	Wed 18/03/2	8
Selection of peer-reviewed journal, conference paper and	4 days	Thu 19/03/20	Tue 24/03/2	9
Compilation and analysis of review data	4 days	Wed 25/03/20	Mon 30/03/2	10
Develop a draft proposal that includes background, literature review, research gap and objective, engineering design and project plan	14 days	Tue 31/03/20	Fri 17/04/2	11
Poster Design/Part List/MATLAB /Learn basic on how to code in python	**7 days**	**Sun 03/05/20**	**Mon 11/05/2**	
Design of proposed system and come up with require parts of the system and the budget	1 day	Mon 20/04/20	Mon 20/04/2	12
Design poster, which display the concept, support principal and	4 days	Tue 21/04/20	Fri 24/04/2	14
Learning python with DataCamp's free intro to python	1 day	Mon 27/04/20	Mon 27/04/2	15
learning basic of coding and implementing the concept of image processing using MATLAB	1 day	Tue 28/04/20	Tue 28/04/2	16

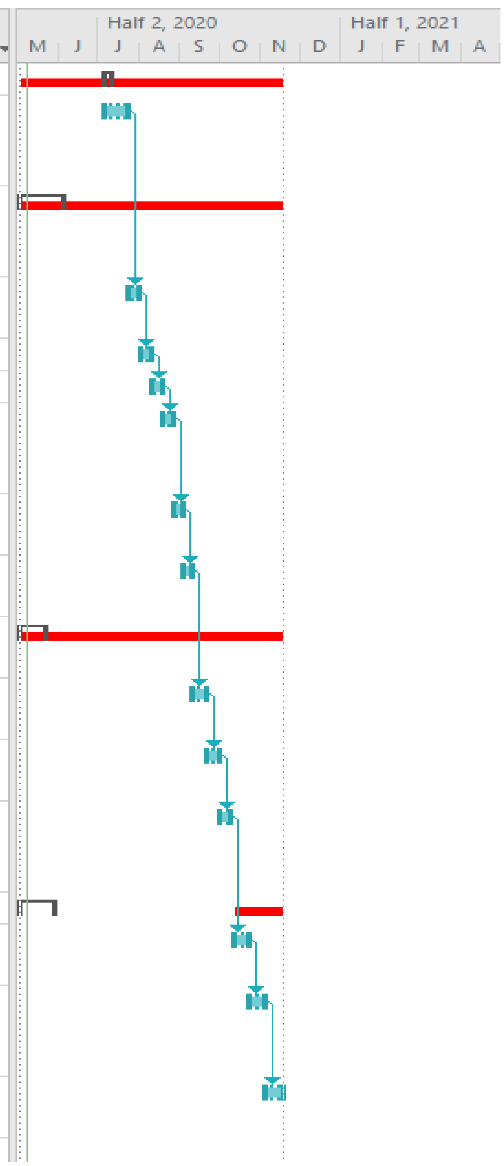


Mid-year Progressive Report	**14 days**	**Sun 03/05/2(**	**Wed 20**	
Continue on fimiliarizing with the concept of image processing and implementation of GMM using MATLAB software	3 days	Wed 29/04/20	Fri 01/05/2	17
Review of project proposal	4 days	Mon 04/05/2	Thu 07/	19
Provide a detailed account of the project progress so	7 days	Fri 08/05/20	Mon 18/05/2	20
Mid semester project Presentation	**7 days**	**Sun 03/05/20**	**Mon 11/0!**	
A ten (10) minutes presentation plus 5 minutes question and	7 days	Tue 19/05/20	Wed 27/05/2	21



Semester 2

Task Name	Duration	Start	Finish	Predeces
Codding and debugging	**1 wk**	**Mon 06/07/20**	**Fri 10/07/20**	
writing python scrip for automated fruit sorting machine	14 days	Mon 06/07/20	Thu 23/07/20	
Image signal acquisition and Data processing and analysis	**5 wks**	**Mon 04/05/20**	**Fri 05/06/20**	
Image signal acquisition and image pre-processing	6 days	Fri 24/07/20	Fri 31/07/20	2
size calculation	6 days	Mon 03/08/20	Mon 10/08/20	4
Extraction of features	6 days	Tue 11/08/20	Tue 18/08/20	5
Model and cluster the image signal using GMM to identify the type of fruit	6 days	Wed 19/08/20	Wed 26/08/20	6
Training and testing of image signal	6 days	Thu 27/08/20	Thu 03/09/20	7
Coding and testing of the dispensing system	5 days	Fri 04/09/20	Thu 10/09/20	8
Hardware interfacing and testing	**3 wks**	**Mon 04/05/20**	**Fri 22/05/20**	
Testing of different type and size of fruit	7 days	Fri 11/09/20	Mon 21/09/20	9
Interface the hardware with software	7 days	Tue 22/09/20	Wed 30/09/20	11
identify, modify and removing error in the script/codes	7 days	Thu 01/10/20	Fri 09/10/20	12
Design analysis	**4 wks**	**Mon 04/05/20**	**Fri 29/05/20**	
Project Modelling and prototyping	9 days	Mon 12/10/20	Thu 22/10/20	13
Remodifying prototype based on supervisor feedback	9 days	Fri 23/10/20	Wed 04/11/20	15
Final Prototype for demonstration	10 days	Thu 05/11/20	Wed 18/11/20	16



APPENDIX B: COSTING

No.	Description	Qty	Unit Cost	Total Cost
1	Raspberry Pi 4	1	$314.80	$314.80
2	Buzzer	1	$1.10	$1.10
3	HDMI LCD display monitor	1	$92.12	$92.12
4	Raspberry pi Camera	1	$57.95	$57.95
5	Conveyor Belt (Full package including the Gear DC Motor)	1	$319.18	$319.18
6	Infrared Sensor	1	$24.22	$24.22
7	Servo Motor	1	$50.10	$50.10
8.	Servo Motor Driver	1	$27.72	$27.72
			Total Estimated Cost :	**$828.14**