Freshness Identification of Banana using Image Processing Techniques

A dissertation submitted for the Degree of Master of Computer Science

M. Yanusha

University of Colombo School of Computing

2019

S	
E1	
E2	
For Office Use Only	

Masters Project Final Report

(MCS)

2019

Project Title	Freshness Identification of Banana Using Image Processing Techniques
Student Name	Miss. Yanusha Mehendran
Registration No. & Index No.	2016/MCS/059 16440599
Supervisor's Name	Prof. N. D. Kodikara

For Office Use ONLY

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: M. Yanusha

Registration Number: 2016/MCS/059

Index Number: 16440599

Signature: **Date**: May 01, 2019

This is to certify that this thesis is based on the work of Ms. M. Yanusha under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Prof. N. D. Kodikara

Signature: **Date**:

Abstract

Agriculture is a very important culture which includes several practices for instance cultivating soil, producing crops and raising livestock. In Sri Lankan context, banana is widely consumed as it fits all the occasions and it has export value too. As a result, determining freshness of banana has major influence in defining its quality. The naked eye observation of experts is the main approach adopted for determining the freshness of banana in terms of days. We developed a method to identify the freshness using image processing techniques. For this experiment, images were captured using a professional camera. The fruit's regions were segmented using K-Means clustering and the determination of freshness was done with Support vector machine by training with the selected features from the training set of images. The accuracy level of freshness determination was calculated separately for each category in terms of days from day one to ten. Association among the features as Contrast, Correlation, Energy, Homogeneity, Entropy, Mean, Standard deviation, Skew, and Kurtosis gave the optimum accuracy. This system with high accuracy motivates the other researchers to extend the system with added functionality, which will be a consumer friendly software solution.

Keywords: HSI, K-means, Gray-level co-occurrence matrix, Support vector machine.

Acknowledgement

It is really a great opportunity and benefit to follow the Master degree in the field of "Computer Science" at University of Colombo School of Computing (UCSC), Sri Lanka. Therefore, with gratitude I would like to take this opportunity to thank University of Colombo School of Computing, Sri Lanka. Also, I heartedly present my respectful thanks to all those who supported and motivated me in every domain, directly and indirectly, to conduct this Projet work.

Especially, I wish to present my sincere thanks to my research supervisor Prof. N. D. Kodikara for his valuable guidance and his assistance during all this time. In addition, my sincere thanks go to the other staffs of UCSC.

Last but not the least, I would also like to thank to my parents, friends, colleagues and collaborator for their encouragement towards this project.

M. Yanusha

Table of Contents

Table of Figures

List of Tables

Chapter 1

1. Introduction

1.1. Background

Computer science is the study that concerns the understanding of computer processes, information transfer and the transformation. The most vital characteristic of computer science is problem solving in wide context. Within computer science Image Processing and Pattern Recognition are part of the principals which provide solution for the problem occurs in various fields.

Agriculture is a very important culture which includes several practices for instance cultivating soil, producing crops and raising livestock. Agriculture plays a vital role in developing country. On the other hand, poor quality is one of the major factors for agricultural losses. Freshness has turned into a problem as it can cause bulky reduction in both value and quantity of yields. As a traditional practice, in Sri Lanka the experts are following naked eye observation to identify the freshness of the agricultural yield which is very expensive.

1.2. Problem Domain

As Sri Lanka is a luxuriant tropical land with the potential for the cultivation, agriculture is considered as the best prospect sector for our country. As an evident if we analyze the Central Bank of Sri Lanka Annual Report it is stated that the agriculture sector contributes about 7.5 percent to the national GDP over fisheries sector which contributes around 1.4 percent and livestock sector accounted for 0.8 percent [1]. The tremendous applicability of computer science and its sub fields encourage the potential for interdisciplinary studies. For instance, the success of image processing field expanded its usage in other fields like medical, forensic, engineering, remote sensing, agriculture etc. For the same extent of demand we have for agriculture in country like Sri Lanka, there is a necessity to develop computer vision applications in order to improve the productivity in agriculture as image processing has become an inevitable area in recent decades. Image processing is considered as an art of processing real time images and producing images with more information and other useful parameters. This project focuses on finding the freshness of banana using image processing techniques.

Agricultural products like vegetables and fruits can be efficiently processed and analyzed using the colour, shape, size and texture features [2, 3]. When we think of image processing with vegetables and fruits, the above mentioned features can be vigorously used in several applications such as object detection, maturity analysis, age factor analysis, disease detection, grading and so on. Each application has its own significant necessity as does the freshness measurement.

The economic value of plant-derived food depends on its quality and how it is preserved over the whole production chain, until it reaches the final consumer. The concept of quality is wide and covers several aspects such as, external appearance, nutritional value, presence of health-related compounds, safety and security. On the other hand, freshness is strictly connected to the "age" of the product and though it is an attribute related to quality. On the other hand, freshness presents one of the main attributes of consumer choice of fruits and vegetables as the consumer is first attracted by the appearance of the product but later on he/she has to be satisfied by the taste and aroma. Consequently, the visual appearance is very important since it attracts the consumer, while the edible quality and the correct information about nutraceutical and nutritional aspects, will make the consumer to re-purchase the product [4, 5]. Moreover, fresh fruits and vegetables have a central role in human diet as well [4]. Therefore we can have an understanding that freshness of a fruit directly affects the quality and the taste of the fruit hence the freshness determination is vital with respect to the economic value [6].

In Sri Lankan context, banana is widely consumed as it fits all the occasions and it has export value too [7]. As a result, determining freshness of banana has major influence in defining its quality. In addition to that, banana is selected to this study as banana is widely consumed and it is convenient to find bananas all the time.

Like other fruits, for banana also the freshness can be easily determined by considering colour, shape, size and texture features as they mature every day from the day it is detached from the tree. In general, freshness determination of a fruit requires expertise approach which has several drawbacks like,

- Difficulty in finding expert
- Labour intensive task
- Low productivity
- Instability of grading precision [8]

Hence by considering the mentioned facts we can realize that there is a vital requirement to automate the freshness determination process by developing a computer vision application using image processing techniques which will enable the local customers and the exporters to consume high quality bananas.

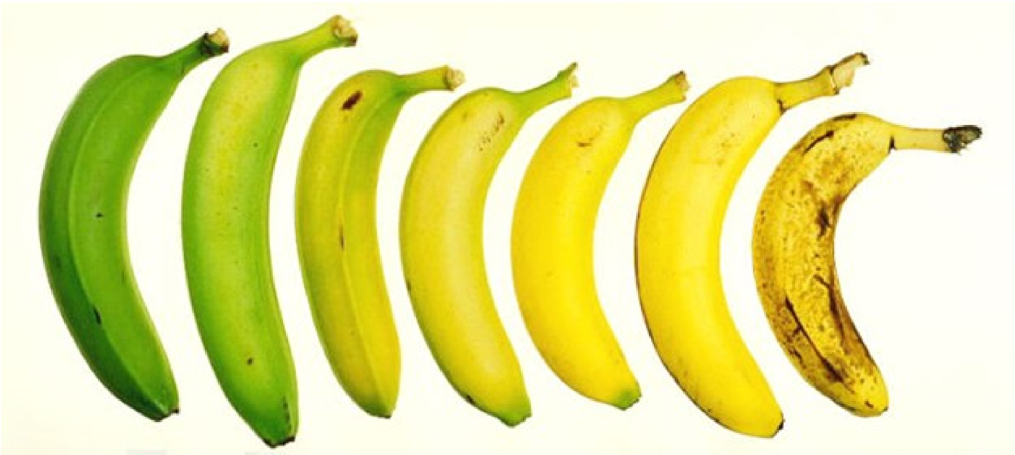

Figure 1: Changes in banana over consecutive days

Moreover, fresh fruits and vegetables have a central role in human diet as well.

1.3. Project Objective(s)

The aim of this project is to automate the process of determining the freshness of banana using image processing techniques by considering colour, shape and texture features. The aim is designed to be achieved by the successful completion of the objectives.

The objectives identified are,

- To correlate the manual results with the automated procedure results in terms of freshness
- To identify the best combination of features to determine freshness of banana
- To determine the freshness of banana in terms of days with high accuracy

1.4.Scope of the Project

This project aims to determine the freshness of banana using various image processing techniques and banana is selected on carry on this project for several reasons as banana is the prominent farming fruit in Sri Lanka that fits all season and has export value. As banana is a climacteric fruit that starts ripening at post harvest, the changes in the colour, texture, and shape can be easily observed over days [9]. The above mentioned three features will be taken into account to develop the application.

There are various kinds of bananas and precisely variety called "Ambul Banana" will be used as the sample object to build up this application by the way it is considered as one of the very popular banana that has high demand in Si Lanka[10]. For the extended work, other variety will be also added to the developed model.

1.4.1. Assumptions

- As freshness is measured in terms of days from the date the fruit is detached from the tree, specifically first ten days are considered where we can differentiate the changes.
- Combined consideration of colour, shape and texture features will enhance the accuracy of the classification based on freshness measure.
- Images of hundred different bananas will be captured every day to have well-formed test data.
- Support Vector Machine (SVM) will be used as the classifier for better classification performance.
- Nearly 1000 photographs will be taken and from that 700 photographs will be considered as training data and rest of them will be considered as testing data.

1.5.Structure of the Dissertation

Chapter 1 - Introduction

This Chapter describes about the state the general topic and gives some background knowledge about the study which has been carried out. Moreover, it defines the terms and scope of the study as well as to identify the importance of the proposed project by stating the aims and objectives of the study.

Chapter 2 - Literature Survey

This chapter clearly specifies the available literature sources and the surveys that are currently available and it highlights the drawbacks in the existing systems and the further developments in the existing systems and practices followed.

Chapter 3 - Methodology

This chapter signifies the selection of the best techniques and practices that suits the research flow in the right manner and it also analyses the right tools and the timely tasks to be carried out throughout the process.

Chapter 4 - Design and Implementation

This chapter includes how the technique and the practices were developed and what are the sources behind the development of such techniques and practices. In addition, this chapter signifies the implementation of the developed systems and technology to showcase the desired output.

Chapter 5 – Testing and Evaluation

This includes the testing of the practices and the technology based systems. Furthermore, it includes the mechanisms to test the system developed against all the constraints and presents the evaluation plan to grade the system's accuracy.

Chapter 6 – Results and Discussion

This chapter includes the results obtained throughout the process carried and every individual result is discussed to make sure that the objectives are met.

Chapter 7 – Conclusions and Further Works

This chapter includes summary of the project and future works of the developed convertor.

Chapter 2

2. Literature Survey

2.1. Chapter Introduction

For past few decades, researchers are interested in doing some researches that provide fruitful solutions to the agriculture field and the farmers. Even though agriculture provides invaluable contribution to the economic growth, farmers in developing countries still lack the equipment, money and ability to respond to the agricultural challenges. This system provides an efficient solution for the farmers and other buyers by determining the freshness of banana in terms of days with high accuracy.

RGB images are used here as an input image as they are considered as the true color image. Region of interest from the RGB image is extracted that reduces the memory consumption and the processing time and moreover it helps to improve the accuracy of the results.

The initial images were taken through several steps until the freshness parameter is extracted. This chapter further explains the similar researches followed in the same domain of the project as well as the techniques used in brief. Further, this chapter discusses about the novelty of the project carried out over the other similar researches. In recent decades several studies were carried out that incorporated image processing techniques to serve several purposes in agricultural field. Among them modest amount of work is done to grade the quality of fruits and vegetables.

2.2. Assessment of Fruit Maturity using Direct Colour Mapping

This study carried out by K. D. Raut and Prof. V. Bora, aimed on the maturity stage of fresh fruit is an important factor that affects the fruit quality during ripening and marketability after ripening. This study used a reasonable method to predict sweet cherries and strawberry's colour constraints by image processing. The colour measuring technique consisted of a CCD camera for image acquisition and MATLAB software for image analysis. In order to demonstrate the worthy of this technique, changes of fruits color during ripening will be analyzed. Evaluation of L*, a* and b* values showed the possibility of reliable use of this system for determination of absolute color values of food items in automatic packing systems for export with a much lower cost [2].

2.3. Identification of Age Factor of Fruit (tomato) using Matlab- Image Processing

Prof. Pramod and Shilpa propose through their study that it is now essential constraint to obtain the advance conception techniques for the safety of fruit and vegetables. It requires the implementation of a set rule for product based on its various features called upon as an algorithm for safe keeping. This research paper represents the analysis of an age factor of tomato based on one of the graphical feature of tomato that is based on colour. The performance of the product is been examined based on the colour of tomato i.e. Red, Green or Orange. The performance of an algorithm was examined and discussed by perceiving the various sample products [3].

2.4. Analysis of Fruits by Image Processing Algorithms

The research study followed by C. Velappan and Subbulakshmi, found that manual sorting of fruits and vegetables is only carried out at many places. This type of manual sorting will not be precise in real time processing. The detection accuracies are also prominently affected by many factors such as time, bruise type, bruise harshness, fruit type, and fruit pre-harvest and post-harvest conditions. Manual sorting has many shortcomings as it requires plenty of labors to investigate it results in low productivity and grading standard is difficult to carry out. Hence the manual sorting is replaced by Machine Vision (MV) system with the advantages of high accuracy and high automatization which is designed by using image processing algorithms [8].

2.5. Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry

M. P. Arakeri and Lakshmana suggested that fruit grading by human is inefficient, labor intensive and prone to error. The automated grading system not only accelerates the time of processing, but also reduces error. There is a great demand for tomatoes in both local and foreign markets. This paper proposes an automatic and effective tomato fruit grading system based on computer vision techniques. The proposed quality evaluation method consists of two phases: development of hardware and software. The hardware is developed to capture the image of the tomato and direct the fruit to the appropriate bins without manual intervention. The software is developed using image processing techniques to analyze the fruit for defects and ripeness [11].

2.6.Fruit Grading using Image Processing Techniques

T. Guray and A. K. Yakut proposed new safe and fast methods for ranking of fruits have vital role in agricultural economy. At the present time, traditional grading methods have still been used widely. However, high cost and some contradictions guide post-harvesting industry to automate the applications in classification operations. Recently, enterprises aligned towards to automation systems for increasing working capacity and decreasing costs. As known, size, shape, color and tissue are base criteria in the classification process. In this research, automatic apple grading by size and colour using digital cameras and computerized image processing techniques were analyzed. The assembled system has achieved basic tasks but it needs to be developed further [12].

2.7.Quality Detection of Fruits by Using ANN Technique

M. Kaur and R. Sharma developed an automated system for grading and classification of fruits is based on observations and through experiences. The system utilizes image processing techniques to classify and grade quality of fruits. Images of fruits are classified based on shape and colour based analysis methods. However, different fruit images may have identical colour and shape values. Hence, using colour and shape features analysis methods are still not effective enough to identify and differentiate fruits images. Therefore, they used a mechanism to increase the accuracy of the fruit quality detection by using colour, shape, and size based method with the support of artificial neural network. Proposed method grades and classifies fruit images based on extracted feature values by using cascaded forward network [13].

2.8.Fruit Quality Inspection using Image Processing

Prof. S. M. Shirshath and S. S. Telang established information that harms in handling and processing and the increased expectation of food products of high quality and safety standards, there is a need for the growth of and objective and accurate, fast quality determination of food and agricultural products. Computer vision is a rapid, economic, consistent and objective inspection technique, which has expanded into many various industries. The process consists of speed and accuracy satisfies ever increasing production and quality necessities, hence supporting in the development of totally automated processes. The non-destructive method of inspection has found applications in the agricultural and food industry, including the inspection of quality and grading of fruit and vegetable [14].

2.9. Fruit Quality Management using Image Processing

Prof. S. M. Shirshath and S. S. Telang again worked on the concept of fruit quality where they found that colour is most outstanding feature for identifying disease and maturity of the fruit. The main prominence is to do the quality check with a short span of time so that maximum number of fruits can be inspected for quality in minimum amount of time. The absolute reference point is the way to perceives and interpret the quality of fruit. The present evaluation method of fruit quality requires new tools for size and colour measurement and capturing the fruit side view image, some fruit characters is extracted by using detecting algorithms. This system performs the sorting using MATLAB software and gives some advantages over traditional practices [15].

2.10. Identification of Quality Index of Vegetable using Image Processing

Prince, R, Sathish, H, Kumar et al carried out a research which deliberates the basic concepts and technologies associated with computer vision system and automatic vision-based technology, a tool used in image analysis and automated sorting and grading are highlighted. The proposed system starts the process by capturing the fruit or vegetable image. Then, the image is transferred to the processing level where the fruit features like colour, shape, and size of fruit samples are extracted. After that by using ANN fruit images are going through the training and testing. In this proposed method neural network is used to detect shape, size, and colour of fruit and with the combination of these three features, the results obtained are very promising [16].

2.11. Quality Identification of Tomato using Image Processing Techniques

Tamakuwala, S., Lavji, J and Patel, R developed an automated quality identification using some image processing techniques is there that can be done using some image features which help in quality detection of tomato like shape, colour and size. This research work presents identification of the condition of vegetable is focused on the image processing techniques like segmentation and classification. First extract certain features from the input vegetable image, later using different method like thresholding, segmentation using k-means clustering and classification using SVM (Support Vector Machine) and ANN [17].

2.12. The Fruit Quality Identification System in Image Processing Using Matlab

Phakade, S V, Vaishali, E, Pooja, C et al presented a hierarchical grading method applied to the Fruits. In this work the identification of quality and defective fruits is focused on the methods using MATLAB. First they extract certain features from the input images, later using different method like thresholding, segmentation, k-means clustering and thus they get related databases. Relating several trained databases, they get a specific range for the quality and defective fruits. From the proposed range they can identify the normal and defective fruits. Thus this paper analysis the normal and defective fruits with a very high accuracy successfully using image processing [18].

2.13. The Quality Identification of Fruits in Image Processing using Matlab

Alok M, Pallavi A and Pooja K focused on the identification of good and bad fruits are focused on the methods using MATLAB. First they extract certain features from the input fruit image, later using different method like thresholding, segmentation, k-means clustering and thus they get related databases. Comparing several trained databases, they get a specific range for the good and bad fruits. From the proposed range they can detect the good and bad fruits. Thus this paper analysis the good and bad fruits with a very high accuracy successfully using image processing [19].

2.14. Summary

As a summary, we can clearly understand that there is a strong positive correlation between fruit ripening and the physical parameters such as colour, shape and texture. In most of the previous studies the models were not trained in order to examine the models via test samples. This project will consider all three types of features to determine the freshness of banana in terms of days and intended to train the system using SVM to enhance the system accuracy and performance. All the features regarding shape, colour and texture will be calculated for every image and they will be critically evaluated to find out the freshness of the banana and later the freshness measure resulted from the automated method will be compared with the manual results to determine the accuracy of the model developed. In addition to this, the project is also intended to find out the best feature or feature combination which can be applied in the scenario where to find out the freshness of a fruit.

Chapter 3

3. Methodology

3.1. Chapter Introduction

Now-a-days, in modern agricultural system, various computational methods have been developed to help farmers to monitor the proper growth of their crops. In our traditional agricultural system, during harvesting process of the crops, the naked eye observation of farmers or experts is the main approach adopted in practice for detection and identification of crop freshness. However, this requires continuous monitoring of experts which might be prohibitively expensive in large farms.

In this research study, the freshness determination of bananas considered to be classified using image processing techniques. Therefore, a methodology is developed in order to achieve the goal of the study. In this chapter, the developed methodology is described along with the techniques which are used.

3.2. Technology Adopted

System for freshness determination of bananas is developed using MATLAB. MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

MATLAB is used here for the image processing techniques and the SVM classification.

Techniques and Methodologies used here are listed below:

- Image enhancement
- Image color space transformation
- K-means clustering
- Thresholding
- GLCM matrices for feature extraction
- SVM classifier
- NN tool

3.3.Flowchart

As an initial step, the RGB images of the bananas are acquired using a digital camera. Then the RGB image was segmented into number of clusters and image processing techniques were used to extract the useful information from the images. The extracted features were used as an input and the freshness is determined in terms of days based on a classifier. Figure 2 illustrates the basic procedure of the proposed methodology for freshness prediction solution.

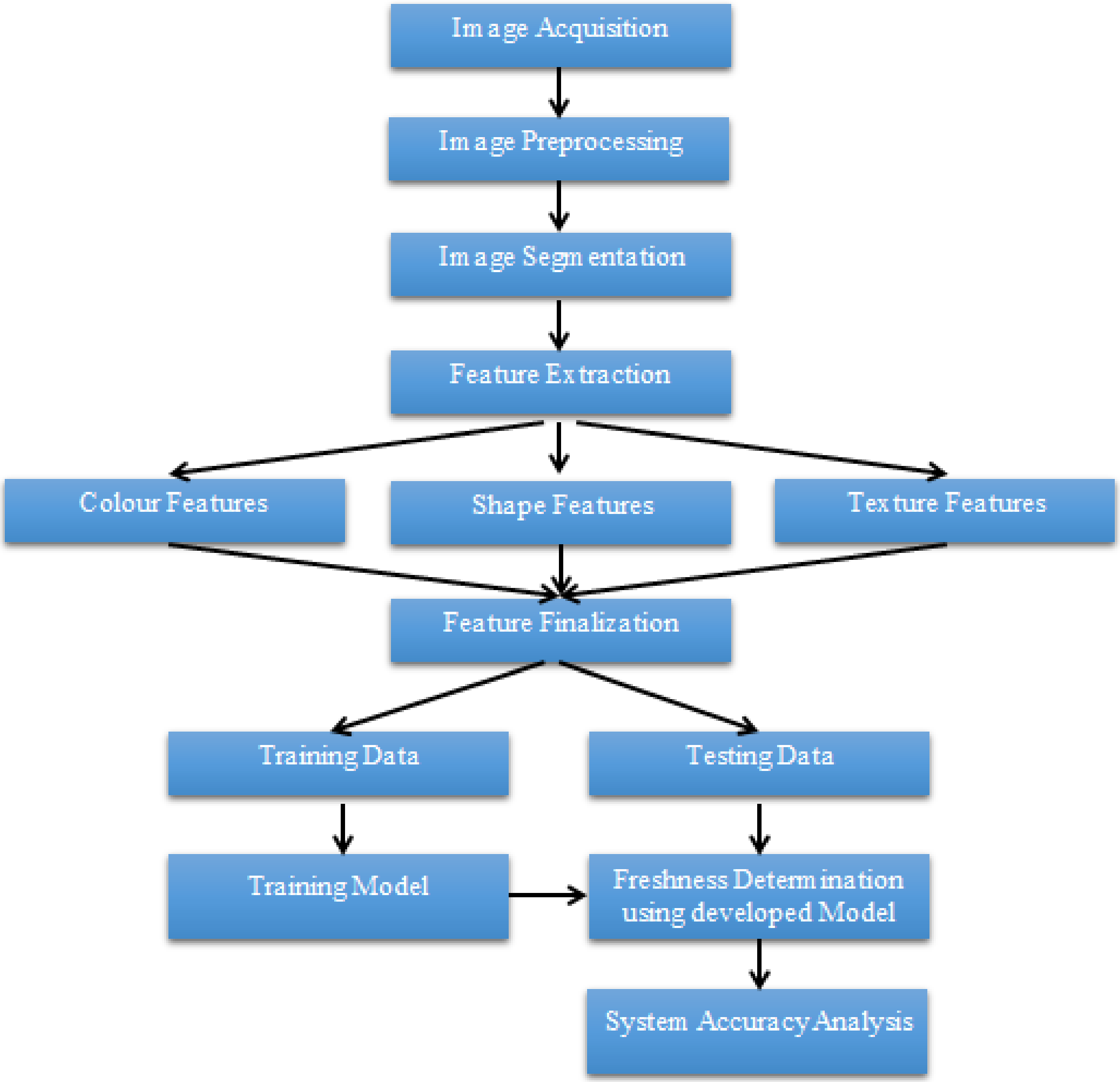

Figure 2: Flowchart of the methodology of the project

The basic steps of the proposed algorithm are listed below.

- RGB image acquisition
- Image preprocessing
- Region of interest extraction
- Convert the RGB input image into LAB color space
- Apply K-means clustering for image segmentation
- Masking and removing green pixel in the image boundaries
- Convert the infected cluster to HSI format
- SGDM Matrix generation
- Extract features using GLCM function
- Texture feature analysis
- Determination of freshness based on SVM classifier

3.3.1. Image Acquisition

This is the initial step of the entire process. As a part of it, light box was developed in order to capture the photograph of the banana fruit in order to discard the shadow and the effects from light source. In the top of the light box, camera was placed where the height remained constant throughout the image acquisition process. Images were taken with the help of a 10 megapixel camera to have quality pictures. Even though the DSLR cameras are in capable of producing quality picture, by considering the processing power and the computation power a mobile camera with 10 megapixels was selected. Pictures of banana were taken for the first ten days from the day it is detached from the tree. For each day pictures 100 different bananas were captured which sum up to 1000 pictures for further processing.

3.3.2. Image Preprocessing

Image pre-processing is an important process relative to feature extraction and image recognizing. The images which have input are always not satisfactory regardless of what image acquisition devices are adopted. For instance, there are noises in the image, the region of interest in the image is not clear or other objects' interference exists in the image. Image enhancement is one of the most interesting and important issues in digital image processing field. The ultimate objective of image enhancement technique is to extract out details that are hidden in an image, or to elevate the contrast in the image where the contrast is considerably low. Image enhancement produces an output image that subjectively looks better than the original image by changing the pixel's intensity of the input image. Here the initial inputs are RGB images and the suitable preprocessing mechanisms were selected.

3.3.3. Image Segmentation

Since the size of the image is high, it is hard to process the entire image. Therefore we can perform image segmentation to take only the image segments which has relevant information. In order to remove the background, segment the image based on colour and retrieve a particular segment of image, several techniques can be used such as k-means clustering, Otsu's method and watershed segmentation. Image segmentation was done for both training and testing data.

3.3.4. Color Transformation

The RGB color space of the input image is transformed and the template image to Lab color space. Lab color space is a 3-axis color system with dimension L for lightness, a and b for the color dimensions and the Lab color space is the most exact representation of color. This step is inevitable for carrying out the next step. In next step the images in hand are going to be segmented using K-means clustering.

3.3.5. Apply K-means Clustering

K-mean is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters fixed a priori. K-means clustering technique targets to partition n observations into k clusters in which each observation belongs to the cluster is considered with the nearest mean, serving as a model of the cluster.

3.3.6. Feature Extraction and Analysis

For determining the freshness, three different features were taken into account initially; colour, shape and texture features. All the possible features were studied and only the values of essential features were extracted from images. After extracting the values for selected features, they were analyzed to check whether they can be used to differentiate each class. After the analysis, mandatory features will be finalized. Furthermore, the data was divided into two major classes such as training and testing dataset in 7:3 proportion accordingly in order to train as well as to test the model developed.

3.3.7. Determination of freshness based on SVM classifier

In machine learning, support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. The results of the selected features were used as the attribute to the Support Vector Machine for training and testing process. Using the Support Vector Machine, first the system was trained with training sample set and then the new data set was classified or predicted using the trained Support Vector Machine. Based on the classification performance, manual labeling and the SVM labelling were correlated to assure that the automated process perfectly mapped the conventional classification method.

3.3.8. System Accuracy Analysis

Results of the model were used to examine the accuracy of the system and the feature set, classifier and the proportion of training and testing data were reconsidered in order to enhance the accuracy as required.

Chapter 4

4. Design and Implementation

4.1. Chapter Introduction

In this research study, the banana freshness is considered to be identified using image processing techniques. The RGB images were collected and then they were enhanced in preprocessing technique. Then the fruit area was extracted based on the region of interest and k-means clustering was used to cluster the images so that the interested area was separated based on the color. Then the useful segment was considered and the while pixels in the background were masked and removed. For colour features, GLCM matrices were calculated for the resultant image and the statistics were calculated. The statistics were exported to a text file and is used as a testing and training data set for the Support Vector Machine classifier. Finally, the accuracy of the system was calculated.

4.2. Initial Input

The RGB images were used here as the initial input as RGB is the true color image which represents the Red, Green and Blue components in the image. Then the image processing techniques were used to enhance the image quality. This is an essential step as we have to remove the noise from the initial input as the RGB images are not reached the satisfactory level. Then the region of interest was extracted since it reduces the memory load and the processing time. At the same time it increases the accuracy level for the further processing as well. Then k-means clustering was applied to segment the images based on the colour so that the fruit part is isolated in a cluster. It is used widely in cluster analysis for that the K-means algorithm has higher efficiency and scalability and computational time is very less.

The white pixels in the background were masked and removed so that it reduces processing time and increases the processing speed. Then for the colour features, the GLCM matrices for energy, correlation, homogeneity and contrast were calculated. These three types of features are considered as they are the most preferable properties of Gray level Co-occurrence method which is followed by other researchers as well. Then the statistics were calculated from the texture features. The statistics calculated were stored in a text file and the data in the text file is separated into training data set and testing data set. The SVM classifier was trained with the training data set and classified the tested data set based on the training data set and the category. SVM was used here because they are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

4.3.Image Enhancement

Image enhancement techniques were used to elevate the quality of an image, where the noise was removed and made certain features easier to see by modifying the colours or intensities. Intensity adjustment is an image enhancement technique that maps an image's intensity values to a new range.

The MATLAB functions were used to adjust the intensity values of an image into a new range based on a threshold value selected. The patterns of the intensity values of the images were analyzed and the threshold value is selected as 120. In order to make the background white, the pixels lay in the background with range of intensity values were brought into a single value as 255 without affecting the relevant details of the image.

Figure 3: Image Enhancement

4.4.Region of Interest Extraction

Region of Interest (ROI) concept is used as the tool to extract a selected circle shaped area of input images. A roi concept is a designated subsection of samples within an image recognized for a particular purpose.

Image = imread('E:/MCS/IndividualProject/Sample/Day1/Photo 2-20-19, 8 47 53 PM.jpg');

ROI = roicirclecrop(image);

Roicirclecrop method is used extract the region of interest from the banana fruit samples which increases the processing speed. This method selects two points using mouse click where the first point would be mid-point of the circle and the second point would be the radius of the circle.

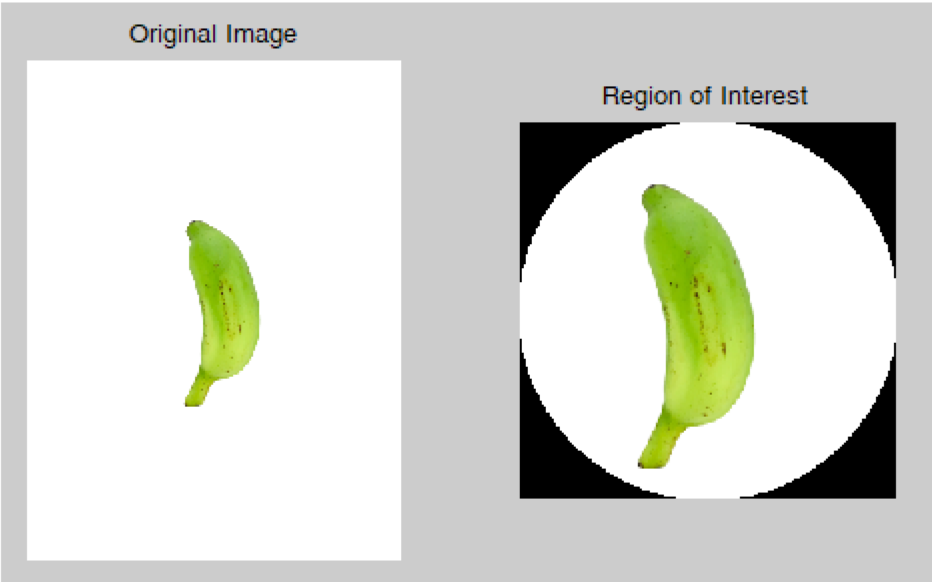

Figure 4: Region of Interest Processing

4.5.K-means Clustering for Segmentation

K-means clustering aims to partition number of observations into k clusters in which each observation belongs to the segment with the nearest mean, serving as a pattern of the cluster.

K-means clustering is used to segment the input image into 2 clusters so that the banana is displayed in one cluster. In this experiment multiple values of number of clusters have been tested. Best results were observed when the number of clusters was 2.

When the K-means is called in the default way, it is not guaranteed to generate the same initial points each time you call the algorithm. So the color segmentation results are varied every time. As a solution two additional parameters are added to the k-means call. The flag start means that we are specifying the initial points, and seeds are a k x p array where k is how many groups we want.

K-means call in the default way:

[cluster_idx, cluster_center] = kmeans(ab,2,'distance','sqEuclidean', 'Replicates',3);

Fixed K-means call:

[cluster_idx, cluster_center] = kmeans(ab,2,'distance','sqEuclidean', 'Replicates', 3, 'start', seeds);

Figure 5 depicts the image which is segmented into two clusters based on the colour where the second cluster was considered for feature extraction since the banana is segmented in the second cluster.

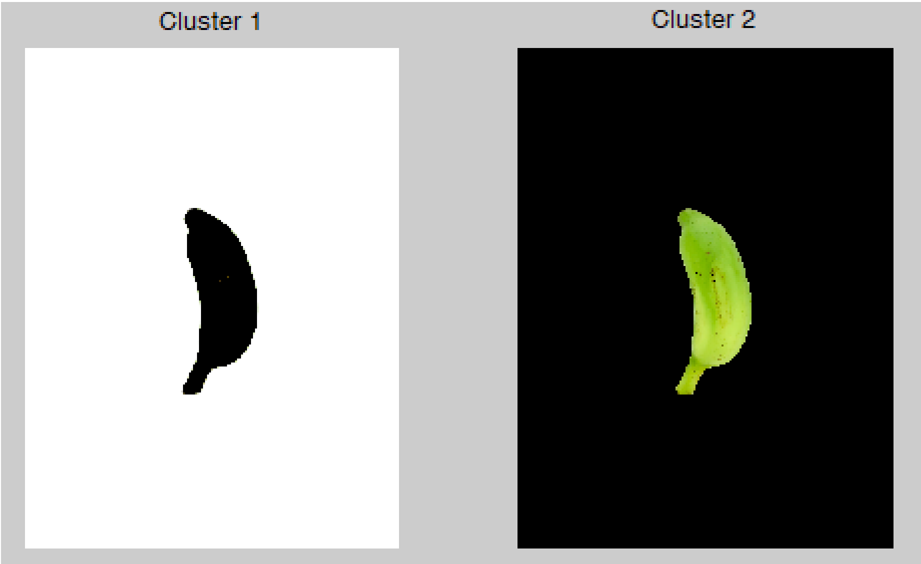

Figure 5: K- means clustering for segmentation

4.6. Feature Extraction using GLCM

By calculating how often couple of pixel with certain specified values and in a specified spatial relationship occurs in an image, the GLCM functions describe the texture of an image, creating a GLCM, and then extracting statistical parameters from this matrix.

For this research study four texture features are taken into account such as correlation, contrast, homogeneity and energy.

```
glcm_matrix = graycomatrix(image);
% extracting statistics from GLCM
statistic = graycoprops(glcm_matrix,'Contrast Correlation Energy Homogeneity');
Contrast = statistic.Contrast;
Correlation = statistic.Correlation;
Energy = statistic.Energy;
```

Homogeneity = statistic.Homogeneity;

From the above code the statistics for the GLCM matrices are calculated and moreover here contrast, correlation, homogeneity and energy are calculated.

Table 1: GLCM matrix for a sample image

glcms

8x8 double

1	2	3	4	5	6	7	8
256883	0	5	15	71	131	7	0
0	0	0	0	0	0	0	0
1	0	3	12	5	2	0	0
7	0	7	36	77	14	0	0
28	0	7	67	2923	436	4	0
163	0	1	11	386	5994	212	0
30	0	0	0	3	190	1669	0
0	0	0	0	0	0	0	0

```
>> [Contrast Correlation Energy Homogeneity]
ans =
  Columns 1 through 3
        0.0448329621380846        0.978504283380338        0.909891547220059
  Column 4
         0.995911293173543
```

Figure 6: Texture feature extracted for a sample image

4.7.Support Vector Machine Classification

The statistics calculated were stored in a text file and the data in the text file was separated into training data set and testing data set. The SVM classifier was trained with the training data set and classified the tested data set based on the training data set and the category. SVM was used here because they are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

function [trdata,tedata]=splitdata(data,traindata,testdata)

Using the above code the total data set was split into two categories; one is training data set which contains the 70% of the data and testing data set contains rest of the data. Here the training data set was randomly selected rather than selecting the first 70% of the data.

Then the SVM was trained with the training data set and then the label for the testing data set was removed and when the SVM was tested the labels for the testing data set were again generated. The newly generated labels are checked along with the stored data set and the accuracy of the classification was calculated for each and every category in terms of days. The table 2 below, partially displays the values of the selected features extracted from the input images after the preprocessing and segmentation.

Table 2: Portion of extracted feature set

0.012460	0.976276	0.918337	0.996299	3.245886	17.335646	0.544028	2.030315	276.215949
0.010029	0.980500	0.916408	0.997581	3.301041	17.151895	0.549159	2.091308	269.843982
0.010702	0.976136	0.923468	0.997331	2.877772	15.947575	0.508303	1.829812	231.301032
0.011191	0.971841	0.921787	0.997372	2.903262	15.575564	0.510137	1.82923	7218.996788
0.014501	0.972191	0.910881	0.996392	3.512441	17.579052	0.572688	1.943972	275.077621
0.011593	0.968709	0.924840	0.997021	2.613995	14.694053	0.501569	1.823909	195.995547

Chapter 5

5. Evaluation

5.1. Chapter Introduction

This chapter describes to what extent the aim of the project is achieved. Testing process was carried out in this phase to check whether the functional and non-functional requirements are achieved.

Totally thousand (1000) images were taken in for ten days of hundred bananas. For these images, statistics were calculated through GLCM matrices and they were stored in a text file. Then the text file data was fed as an input to SVM so that it can be trained. The testing data set was fed into SVM classifier and they were tested against the training set and the accuracy level of the system was calculated.

5.2. User Evaluation

In this research study the accuracy for the prediction of the freshness of banana is tested. Its functionality is obvious which is the system should function to suit the user's conditions and needs.

Main functionalities in this system:

- Segmentation of image
- GLCM matrices generation
- Colour feature analysis
- Shape feature analysis
- Freshness Classification

5.3. Evaluation Approach

5.3.1. Data Set

In this project data set is considered as two different sets; one is training data set and the other one is testing data set. Images of hundred (100) bananas were taken for ten (10) consecutive days and the selected features falling under color, shape and texture categories were extracted from the bananas and labelling was done manually. The features extracted along with the labels were constructed as the training data set and used for computer vision-based model training. The other set is testing data set; for that thirty (30) bananas were roughly selected for feature extraction and as training data set and the collection of this feature set was constructed as testing data set. The testing data set was used for freshness prediction and determination of model accuracy.

5.3.2. Label Prediction

As the training data set, testing data set also comprises with labels entered manually by myself. As I know the freshness of the banana in terms of day, the labels were manually entered by me in the testing data set. Once the model is trained using training data set, the testing data set was fed to the system where the system or the model developed removed the labels and calculate the labels again. Thereafter the system compared and correlated the manually entered labels and the predicted labels. Once a new banana image is fed into the system, it predicted the label and since the tester knows the freshness the accuracy of the system can be evaluated.

5.3.3. Computational time

A separate module was developed to find out the computational time of the model to predict the label correctly using Matlab. Once the model is trained using training data set, the testing data set was fed to the system and the system was designed to remove the labels from testing data set and reproduce the data for every field of features. The time taken for this process was considered as the computational time and further used for system response evaluation.

5.3.4. Accuracy evaluation of the system

The accuracy of the system is decided to be calculated using Confusion Matrix which is a table often used to describe the performance of a classification model.

Table 3: Confusion Matrix

	Predicted Class	
Actual Class	True Positive	False Negative
	False Positive	True Negative

For each and every testing image, the accuracy was calculated using the concept of confusion matrix and the same process was continued up to 500 iterations to get the optimum accuracy evaluation of the model developed.

5.4. Performance Evaluation

This phase describes the performance characteristics of this system such as modifiability, reusability, reliability, performance and accuracy.

- Modifiability

This system can be modified and extended with additional features. For this research study only freshness of a particular type of banana is calculated and this system can be further developed as a freshness predictor for other types of bananas and other fruits.

- Reusability

This system can be reused as a component and other feature like Mobile application also can be developed.

- Reliability

The accuracy of the system depends on the number of input data that is used as input to Support Vector Machine. When the number of inputs increases the accuracy of the system increases which increases the reliability.

Chapter 6

6. Results and Discussion

6.1. Chapter Introduction

This chapter discusses the results obtained after the experiments carried out throughout this project. In addition, this chapter emphasis the significance of the project as well as ensures that the objectives are met. Every results obtained is discussed with the support of the screenshot of the frames developed along with the obtained results.

6.2. Feature Set

For this project work to determine the freshness of the bananas, texture, colour and shape features were extracted from the enhanced image and stored as a text file. The Table 4 below shows the values for every feature extracted from the initial input images. Thereafter the data stored in the file is labelled in terms of days and split into two as training sample and the testing sample.

Table 4: Feature set extracted from the initial input

Contrast	Correlation	Energy	Homogeneity	Mean	SD	Entropy	RMS	Variance	Smoothness	Kurtosis	Skewness	Label (Days)
0.022	0.933	0.911	0.996	2.917	14.895	0.573	2.118	202.278	1.000	39.422	5.823	3
0.014	0.965	0.914	0.997	3.084	15.843	0.565	2.134	229.965	1.000	37.159	5.718	3
0.012	0.976	0.919	0.996	3.242	17.095	0.536	2.025	269.056	1.000	40.097	5.908	4
0.011	0.982	0.918	0.997	3.540	18.197	0.534	2.063	304.505	1.000	33.201	5.472	4
0.011	0.975	0.924	0.997	2.952	16.082	0.500	1.818	235.606	1.000	41.516	6.046	4
0.013	0.971	0.921	0.997	3.082	16.201	0.506	1.813	236.522	1.000	34.913	5.619	4

6.3. Feeding Image to the System

A user-friendly graphical user interface is developed using Matlab to feed the initial input image to the system for further processing. For the testing purposes the initial image can be fed to the system through the GUI where the training is already done with the help of a looping structure. Figure 7 depicts the interface where the initial image can be fed to the system.

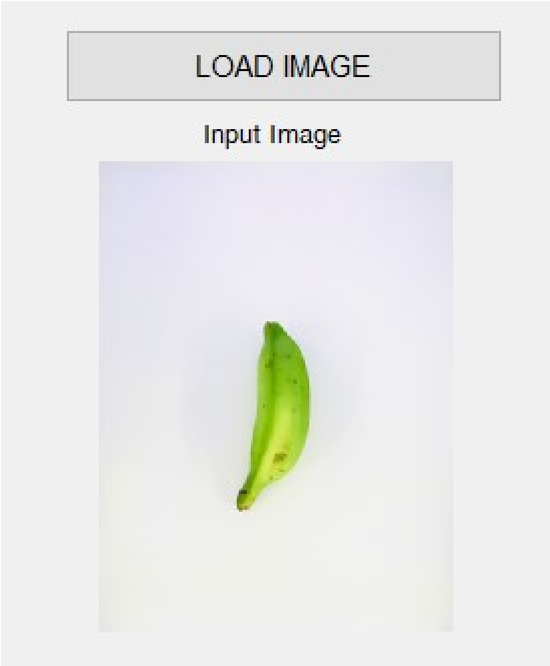

Figure 7: Interface to upload the image

With the help of the "LOAD IMAGE" button the path of the image should be specified and the image can be fed into the system developed to determine the freshness of the banana.

6.4.Image Enhancement

The MATLAB functions were used to adjust the intensity values of an image into a new range based on a threshold value selected. The patterns of the intensity values of the images were analyzed and the threshold value is selected as 120. When the intensity value at a point is greater than 120, then that intensity values will be set as 255 which denotes white colour. The same process was carried out for the training and the testing images. However, for the testing image sample, the same process was carried out through another interface. Figure 8 expresses the interface where the enhanced image will be displayed in a frame.

"ENHANCE CONTRAST" button is used to handle the enhancement of the image and the resultant image will be displayed in the frame after the process.

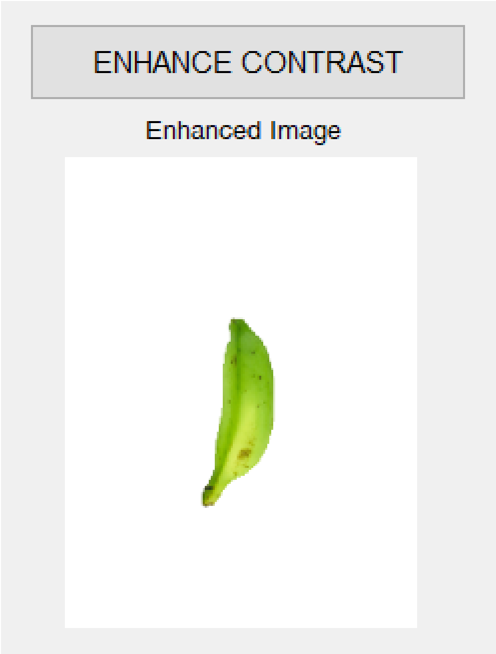

Figure 8: Interface to enhance and display the enhanced image

6.5. Image Segmentation

For the image segmentation for testing images, an interface is developed and through the interface once the initial image is fed into the system, it will go through image enhancement process and then to the image segmentation. For the segmentation processes, k-means method is used and based on the colour the image was segmented into two regions which gave the optimum results for the scenario selected. Figure 9 below expresses the interface where the image can be segmented into two regions and the user has to enter the cluster number where the fruit is isolated using k-means clustering.

Once the cluster number is provided to the system, the selected cluster will be displayed in another panel and from the selected cluster the features will be extracted which is the next process. Figure 10 depicts how the selected cluster is displayed in the frame.

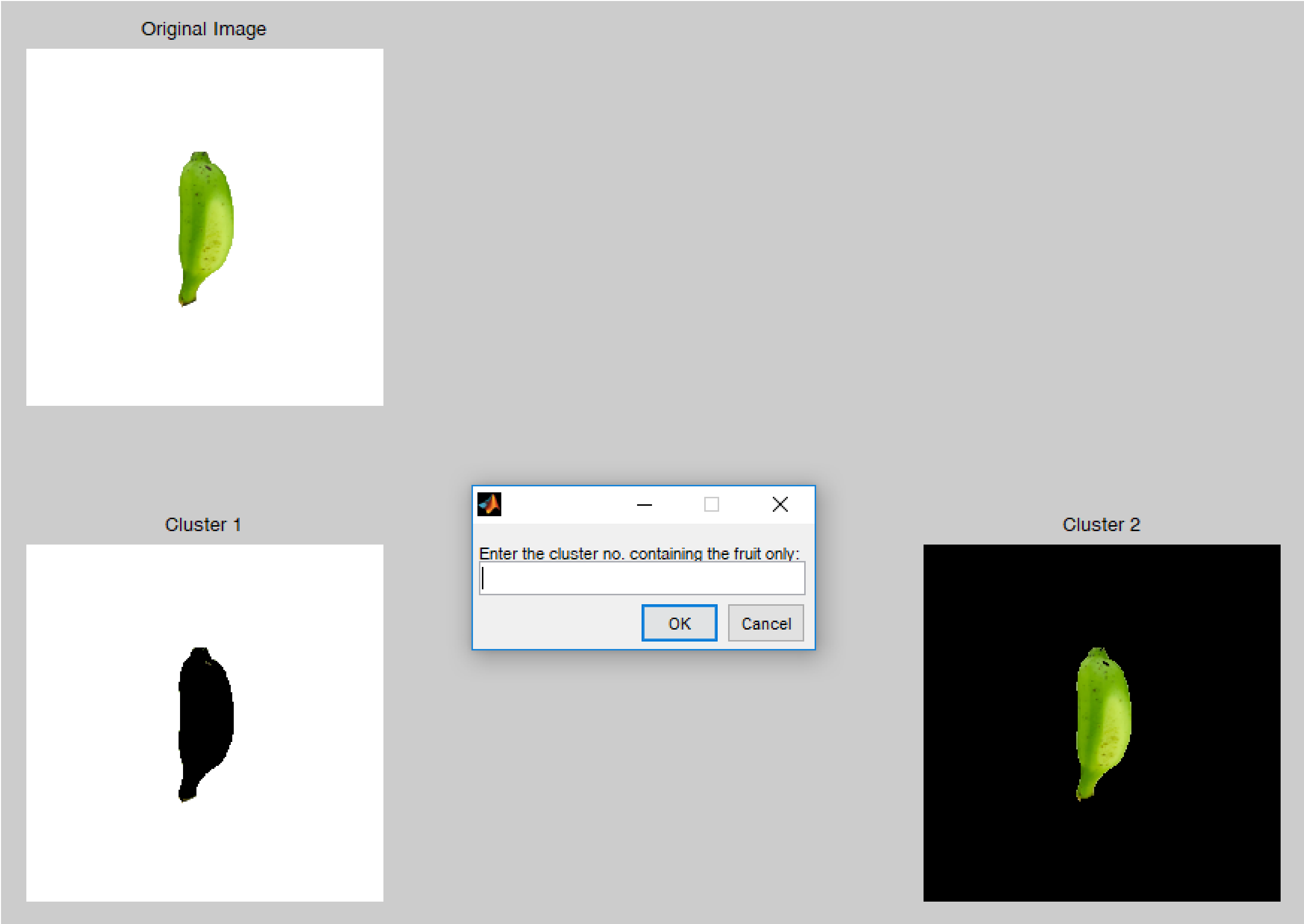

Figure 9: K-means clustering applied to the image

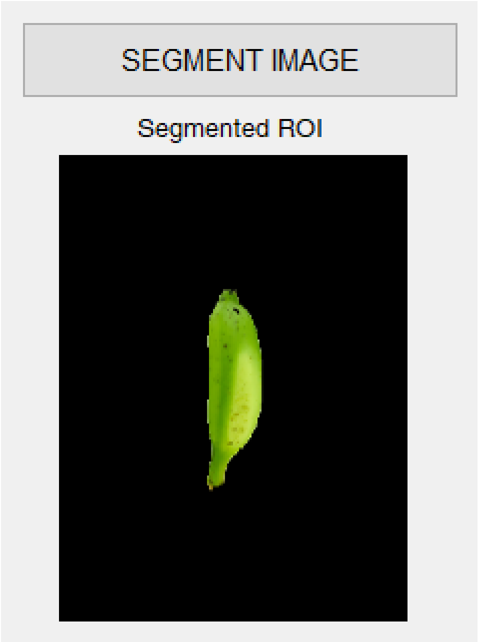

Figure 10: Selected cluster displayed in the frame

6.6. Feature Extraction.

As the above mentioned processes, for both testing and the training image samples the same procedure was followed to extract the values for the selected features such as contrast, entropy, correlation, homogeneity, standard deviation, IDM, mean, root mean square (RMS), smoothness, skewness, energy, variance and kurtosis. However, the feature value extraction is designed to be happened through a developed interface where the values for every feature is displayed in the interface as well as the values are further taken for the class prediction also known as the freshness prediction in terms of days. Figure 11 depicts the interface where the values of the features are extracted from the image fed into the system are displayed.

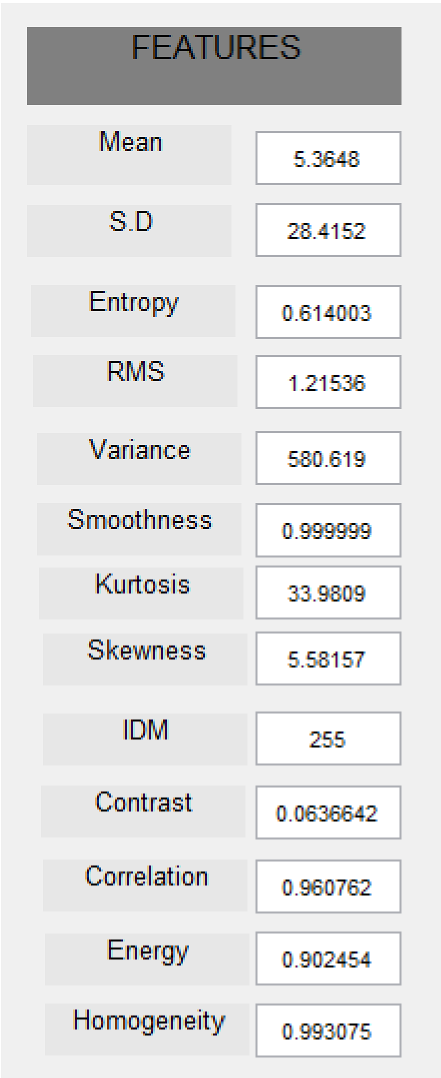

Figure 11: Values of the features extracted from the segmented image

6.7. Freshness prediction

The extracted features from the testing input image will be used for the prediction process. SVM analyzes the patterns of the values with the training data set and the closest similar class will be selected as the results. Here the result is the freshness of the banana in terms of days from the day it is detached from the tree. Figure 12 shows the result predicted in terms of days along with the dialog box which is also displaying the freshness of the fruit.

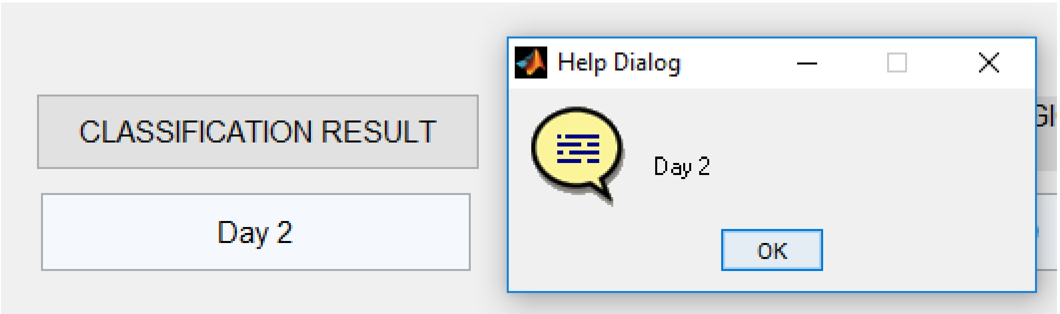

Figure 12: Freshness predicted for image fed to the system

6.8. Accuracy of the System

Accuracy of the prediction also can be calculated where the predictions are designed to take place for five hundred (500) iterations and the number of correct response received was taken into account for calculating the accuracy of every prediction. Figure 13 depicts the accuracy calculation for a particular prediction.

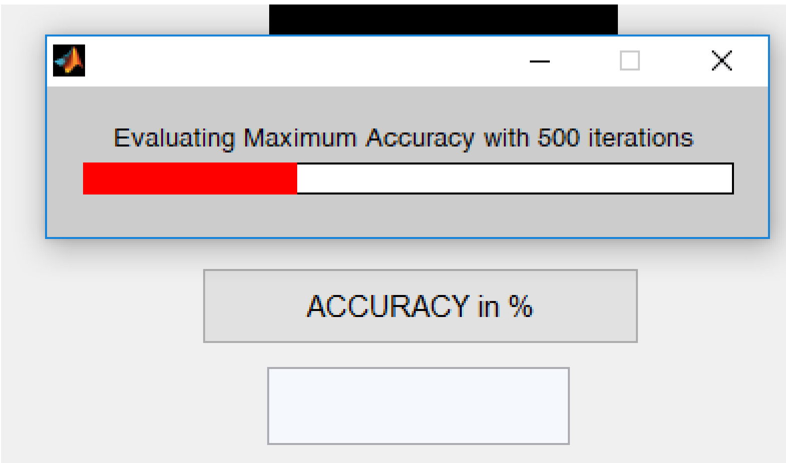

Figure 13: Accuracy evaluation

Chapter 7

7. Conclusion and Future Work

7.1. Conclusion and Remark

This chapter describes summary of the project and a restatement of its main results, i.e. what has been learnt and what it has achieved.

At the end of the project, the accuracy level for the freshness prediction for banana in terms of days was calculated for each and every category (ten different days) separately based on number of iterations. The accuracy of the system is convincing and it can be further elevated through further analysis of features, number of images and various classifiers.

This system with high accuracy motivates the other researchers to extend the system with added functionality which will be a farmer and a consumer friendly software solution.

7.2. Recommendations for Future Work

The future work for this study is planned as a mobile phone application where the farmers, consumers and other interested parties can upload the images to the application and received message stating the freshness of banana in terms of days.

Moreover it is planned to incorporate other types of bananas as well as other fruits and vegetables which have economic value. In addition, feature and classifier analysis can be planned based on new techniques in order to improve the accuracy of the systems.

List of References

[1]"Sri Lanka - Agricultural Sector | export.gov", Export.gov, 2018. [Online]. Available: https://www.export.gov/article?id=Sri-Lanka-Agricultural-Sector. [Accessed: 24- May- 2018].

[2]K. Raut and V. Bora, "ASSESSMENT OF FRUIT MATURITY USING DIRECT COLOUR MAPPING", International Research Journal of Engineering and Technology (IRJET), vol. 03, no. 03, pp. 1540 - 1543, 2016.

[3]P. Devalatkar and S. Koli, "Identification of Age Factor of Fruit (Tomato) using Matlab- Image Processing", International Journal of Recent Trends in Engineering & Research (IJRTER), vol. 02, no. 07, pp. 7 - 13, 2016.

[4]G. Cocetta, "Quality or Freshness? How to Evaluate Fruits and Vegetables during Postharvest", Advances in Crop Science and Technology, vol. 02, no. 03, 2014.
[5]P. Sandrine, "FRESHNESS OF FRUITS AND VEGETABLES: CONCEPT AND PERCEPTION", Ph.D, SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH, 2005.

[6]"A Rapid, Reliable Way to Determine the Freshness of Fruit", Analysis & Separations from Technology Networks, 2017. [Online]. Available: https://www.technologynetworks.com/analysis/news/rapid-reliable-way-to-determine-the-freshness-of-fruit-289704. [Accessed: 31- Jul- 2018].

[7]"export-banana", Cicagri.com. [Online]. Available: http://www.cicagri.com/index.php?page_cat=exportbanana. [Accessed: 24- May- 2018].

[8]C. Velappan and S. K, "Analysis of Fruits by Image Processing Algorithms", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 04, no. 08, pp. 6805-6811, 2015.

[9]"Banana Ripening Chambers", Isopaninsulation.com. [Online]. Available: http://www.isopaninsulation.com/banana-ripening-chambers/. [Accessed: 25- May- 2018].

[10]"Roar Media", Roar.media, 2017. [Online]. Available: https://roar.media/english/life/food/banana-nation-comprehensive-guide-sri-lankan-bananas/. [Accessed: 24- May- 2018].

[11]M. Arakeria and Lakshmana, "Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry", 2016, pp. 426-433.

[12]T. Guray and Y. Ali Kemal, "Fruit Grading Using Digital Image Processing Techniques", Journal of Agricultural Machinery Science, vol. 5, no. 1, pp. 99 - 101, 2009.

[13]K. Mandeep and R. Sandeep, "Quality Detection of Fruits by Using ANN Technique", Journal of Electronics and Communication Engineering, vol. 10, no. 04, pp. 35-41, 2015.

[14] S. M. Shirsath, M. Sumit, and S. Telang, "Fruit Quality Inspection using Image Processing," 2017.

[15] M. Sumit, S. Telang, and S. M. Shirsath, "Fruit Quality Management using Image Processing," *Int. Conf. Ideas, Impact Innov. Mech. Eng.*, vol. 5, no. 6, pp. 727–733, 2017.

[16] R. Prince, H. Sathish, R. S. Kumar, V. Karthik, S. P. Vijay, and V. Reddy, "Identification of quality index of fruit/vegetable using Image Processing," *,* vol. 4, no. 2, pp. 1308 – 1313, 2018.

[17] S. Tamakuwala, J. Lavji, and R. Patel, "Quality Identification of Tomato using Image Processing Techniques," *International Journal of Electrical, Electronics and Data Communication,* vol. 6, no. 5, pp. 68 – 70, 2018.

[18] S. V Phakade, E. Vaishali, C. Pooja, and O. Rupali 4 1 Professor, "The Fruit Quality Identification System in Image Processing Using Matlab," *International Journal of Innovative Studies in Sciences and Engineering Technology,* vol. 3, no. 4, pp. 37 – 41, 2017.

[19] A. Mishra, P. Asthana and P. Khanna , "The Quality Identification Of Fruits In Image Processing Using Matlab," Int. J. Res. Eng. Technol., vol. 03, no. 22, pp. 92–95, Jan. 2015.

Appendix

Colour Transformation

```
cfrm = makecform('srgb2lab');

lab_he = applycform(he, cfrm);
```

K-means Clustering

```
% Classify the colors in a*b* colorspace using K means clustering. Since
% the image has 2 colors create 2 clusters. Measure the distance using
% Euclidean Distance Metric.
a = double(lab_he(:,:,2:3));
nrows = size(a,1);
ncols = size(a,2);
a = reshape(ab,nrows*ncols,2);
nCol = 2;
[clust_idx clust_center] = kmeans(a, nCol, 'distance', 'sqEuclidean', 'Replicates', 3);


pixel_labels = reshape(clust_idx, nrows, ncols);


% Create a blank cell array to store the results of clustering
seg_images = cell(1,3);
% Create RGB label using pixel_labels
rgb_label = repmat(pixel_labels,[1,1,3]);


for k = 1:nCol
    colors = test1;
```

```
    colors(rgb_label ~= k) = 0;

    seg_images {k} = colors;

 end


 figure, subplot(1,2,1);imshow(seg_images {1});title('Cluster 1');
subplot(1,2,2);imshow(seg_images {2});title('Cluster 2');

set(gcf, 'Position', get(0,'Screensize'));

% % Feature Extraction

x = inputdlg('Enter the cluster no. containing the ROI only:');

 i = str2double(x);

 % Extract the features from the segmented image

 seg_img = seg_images {i};

% Convert to grayscale if image is RGB

if ndims(seg_img) == 3

  img = rgb2gray(seg_img);

end
```

Feature Extraction

```
glcms = graycomatrix(img);


% Derive Statistics from GLCM

stats = graycoprops(glcms,'Contrast Correlation Energy Homogeneity');

Contrast = stats.Contrast;

Correlation = stats.Correlation;

Energy = stats.Energy;

Homogeneity = stats.Homogeneity;

Mean = mean2(I);
```

```
Standard_Deviation = std2(I);

Entropy = entropy(I);

RMS = mean2(rms(I));

Variance = mean2(var(double(I)));

a = sum(double(I(:)));

Smoothness = 1-(1/(1+a));

Kurtosis = kurtosis(double(I(:)));

Skewness = skewness(double(I(:)));


% Inverse Difference Movement

m = size(I,1);

n = size(I,2);

in_diff = 0;


for i = 1:m

   for j = 1:n

      temp = I(i,j)./(1+(i-j).^2);

      in_diff = in_diff+temp;

   end

end

IDM = double(in_diff);


%Feature Matrix

feature_set = [Contrast,Correlation,Energy,Homogeneity, Mean, Standard_Deviation,
Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM];
```

SVM Training and Testing Data Separation

```
function [trdata,tedata]=splitdata(data,traindata,testdata)


trdata=[];
tedata=[];
for cls=1:3
   indices=find(data(:,end)==cls);
   temp=data(indices,:);
   r=size(temp,1);
   ind=randperm(r)';
   temp=temp(ind,:);

   ind=round(r*0.7);
   trdata=[trdata;temp(1:ind,:)];
   temp(1:ind,:)=[];
   tedata=[tedata;temp];
end
save(traindata,'trdata','-ASCII');
save(testdata,'tedata','-ASCII');
```

SVM Training and Testing

```
function [trdata,tedata]=mysvmtest(traindata,testdata,interestcls)
trdata=load(traindata);
tedata=load(testdata);
```

```
%training phase

trX=invertData(trdata,interestcls);%change the class label as -1 and 1

trY=trX(:,end); %only last column of training data, (class label)

trX(:,end)=[];% without class label data


options=svmlopt('C',1);%display the parameter C and display the svmlopt value

model=[int2str(interestcls),'vsall']; %this will return as if cls=1==> 1vsall

svmlwrite('svmltrain',trX,trY);%to make as svmlwritable

svm_learn(options,'svmltrain',model);%svm learn (it will learn)


%testing phase

teX=invertData(tedata,interestcls);

teY=teX(:,end);

teX(:,end)=[];

svmlwrite('svmltest',teX,teY);

modeloutput=['predict',int2str(interestcls),'vsall'];


svm_classify(options,'svmltest',model,modeloutput);

predictions=svmlread(modeloutput);

%disp(predictions);

output=Performance(teY,predictions);

disp(output);
```