Identification of Ripeness of Tomatoes

Vaibhav G B
Dept. of Computer Science & Engineering Sahyadri
College of Engineering Mangalore, India

Mohammad Safwan K
Dept. of Computer Science & Engineering Sahyadri
College of Engineering Mangalore, India

Shreehari K
Dept. of Computer Science & Engineering Sahyadri
College of Engineering Mangalore, India

Bhuvananda
Dept. of Computer Science & Engineering Sahyadri
College of Engineering Mangalore, India

Vanishree B S
Asst. Prof Dept. of Computer Science
Sahyadri College of Engineering Mangalore, India

Abstract -- **In present era both fruit and vegetable market are subjected to choose. Ripeness checking and grading is a well-known process that takes place in all the industries. The inspection done by the humans might go wrong because of laziness, inaccuracy. Due to this, farmer might suffer by not getting the right amount of price to their goods. Presently the grading and ripeness checking devices are too costly and farmers are unable to afford it. So here a cost-effective system is developed. The images are taken using raspberry pi and the features are extracted which will help to find the ripeness and grading. Machine learning algorithm SVM is used for classification and to get the desired output. The above ripeness and grading information can be passed on to different devices for large scale estimation.**

Keywords—Machine learning, SVM, cost effective, ripeness, grading

I.   INTRODUCTION

Agriculture is the backbone of the Indian economy. Two-third population of the India is dependent on agriculture. As the world is moving towards digitalization, but the agriculture development is sluggish. So, agriculture has to be made smart and efficient. Tomato is universally cultivated because of its better nutrient. Tomatoes can be used as raw vegetables or which can also be used to prepare products like soup, ketch-up, paste, juices, etc. In present era farmers are awarded fair price only for the good quality fruits. As the quality of the fruit increases, farmers are able to get fair price for their product. Therefore, ripeness checking is necessary for the food industries and the farmers to ensure that the tomatoes are not defect and are ripe.

The fruit market and business is a matter of choice. The manual inspection done by the humans result in various problems with respect to accuracy in ripeness and less speed. In addition to that due to large quantity, poor decision by human eyes will result in loss. Usually people will arrange the fruits and vegetables based on the color, size and defectness. But, doing this work people will get tired and they will lose their concentration after some period of time which will causes the inaccurate, inconsistent outputs. Therefore, it is the necessary of industry to provide high speed and low-cost automated quality detection and grading system. Then, the work will be faster, free from error, saves time and reduces human resources. Neural Network, Genetic Algorithm, Fuzzy Logic Multiclass SVM Based Classification are some of the techniques used for checking the fruit quality.

In this paper, as the color being sign of tomatoes maturity, we specified the red tomatoes belongs to high ripeness, green tomatoes belongs to low ripeness and defect fruit will be separated for the consumptions.

Tomato maturity is divided into three cases based on the percentage of ripeness:

CASE 1: Green color of the tomatoes will indicate the unripe fruit.

CASE 2: Red color of the tomatoes will indicate the well ripen fruit.

CASE 3: Black dots or shades of the tomatoes will be indicated as defected fruits.

II. RELATED WORK

The study conducted here helps us understand the ripeness based on color. Further use of fuzzy logic, genetic algorithm, neural network and SVM technique s are used and are studied, to judge the ripeness of the fruit. The following three criteria are taken into consideration ripeness (red), Non ripeness (green), and Defect black dots on tomato. Here the three RGB Colors cannot be used in machine learning because it is less sensitive, so we make use of HSV values for further processing in machine learning. Further preprocessing the images and extracting these features, we can use classification algorithm for calculating the ripeness of the tomato [1]. The study conducted here helps us understand the use of color extraction, the normal and diseased images are collected for pre-processing. Furthermore, the color extracted here uses RGB color values for classification and the use of MATLAB is done to process out these values. These values classified using the MATLAB software is further processed into the SVM for binary classification, for this we see that the classification maintains an accuracy of 87%. Here the steps taken are involves using images of normal and diseased sugarcanes for finding the features of the sugarcane [2]. Here the classification is based on five stages; these five stages involve the color identification of the tomato. The initial green color is taken into consideration further breaking and tuning of the image is done for further stages the color may change to pink to light red to red. These stages involve the percentage calculation for each stage which increases accordingly. Designing and development of these identification stages is done using PyCharm IDE to support the working of the designed software. The particular set of steps taken involve data collection, image augmentation, feature extraction, creating a CSV file, training the images from the collected CSV values, classifying these images for identification. Furthermore, verification and validation are done on the results obtained to verify the robustness of the results obtained [3]. This system the ripped tomato is segmented by K-means clustering and extract a single ripe tomato using mathematical morphology method to identify the situation of tomato. Here shape and color features are also combined to recognize the ripe tomatoes [4]. Color maturity of the tomato is divided into six different stages namely Green stage, breakers stage, turning stage, pink stage, light red stage and red stage. Here fuzzy logic is used to identify the color size and shape of the tomatoes [5]. Using image processing approach glucose content and level of maturity tomato is predicted. This system uses image of tomato skin which determine the red histogram, RGB, and HSV. Here, backpropagation with red histogram parameter is used to determination of tomato maturity level [6].

III. METHEDOLOGY

The methodology is divided into two phases. The first part contains the training phase and the second part contains the testing phase. Before starting up the first phase collecting datasets of tomatoes is important. The below diagram represents the system architecture.
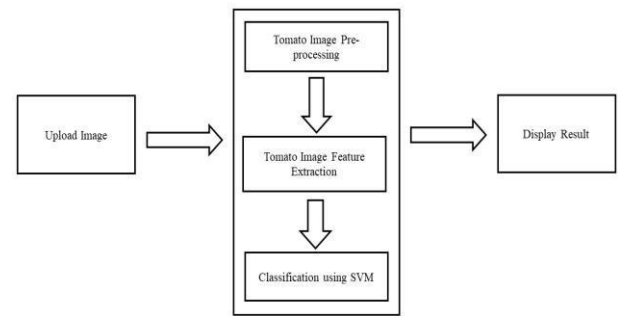

Fig. 1. System architecture

A. Data Collection

The collection of data for the system is in the form of images. The images are further classified into red and green images of tomatoes. For better accuracy we need to have a greater number of images. Images with higher resolution leads to more accurate value. The dataset background should be white because there should not be any problems regarding intensity. The collected images are loaded to the system for further use.

B. Pre Processing

Pre-processing stage aims at converting the real colored images into gray scale images. The conversion into gray scale helps in texture feature extraction. MATLAB provides an easier way built in function for conversion to grey scale.

C. Shape and Size estimation

Size will help to find small or big the fruit is, and shape will help to get the boundary information.
Algorithm to find the size and shape of the tomatoes

1. Find the outer boundary using gray scale
2. The outer boundary is represented by white color line
3. Divide the image into many parts
4. Compare the area of each image
5. Find the biggest area
6. Include the smaller areas to the larger area and label it

D. Defect Detection

Ripeness is one part of the maturity checking. As well as checking the maturity, we need to check the defected tomatoes which in turn helps in grading and saving from loss. In order to do the defect checking we need a clustering algorithm. The highest used clustering algorithm is K-means. It is a straight forward approach and easy way or clustering the data. The value of the clusters can be set according to the user choice.

Algorithm for defect detection

1. Initialize the center points for the different clusters

2. To a closest K we assign our data points

3. Once all the objects are assigned, the positions of the k centroids are calculated again.

4. We continue with steps 2 and 3 until the centroids stop moving.

E. Color Feature Extraction

Color is a distinctive feature that represents the image. Mean, skewness and kurtosis are the features of color. The values of the above features are extracted and stored in a .mat file. Average value of each color is represented by mean. The distribution of each color is done by skewness and kurtosis. Combination of all three explains the color feature.

F. Texture Feature Extraction

The features of texture are extracted from gray scale. Gray level Co-occurrence Matrix (GLCM) extracts the texture features. Features includes contrast, correlation, energy, homogeneity. Contrast measures contrast between neighbors. Correlation of pixels within neighbors is measured using correlation. Energy is a sum of GLCM. Homogeneity computes similarity of GLCM.

G. Creating a CSV File

CSV stands for Comma separated values. It is a file format used to store large set of data regarding color and texture feature values. Each image values are stored in a particular row assigned to it.

H. Training and Classification

After the extraction of all the features from the images the training of dataset has to be done. In order to train we need classifiers. Classifier he*lps* to classify the data into different stages. The process of training needs a machine learning algorithm to learn from the dataset.

I. SVM

SVM stands for support vector machine algorithm. It is a machine learning technique used for classification purpose. The classification results in percentage classification which is mentioned above in the introduction. Classification is labeling of extracted feature. The label includes different percentages. The support vector machine is separable by a hyperplane an N-D coordinates. (N is the number of features). The hyperplane helps to take the decision. The data falling on both the sides of the hyperplane results in two different classification. The coordinates depend on the number of features. We plot a graph of labeled training data points on the plane. SVM considers these data points and provides us outputs. The line is decision boundary which puts a clear line between sets that belong to different category based on the inputs.

J. Testing and Validation

In order to find the wellness of the proposed model we split the dataset into two parts. 20% for testing and 80% for classification. The test datasets are tested against the trained dataset and the best result is achieved.

K. Working

The images are uploaded and are read by the system and converted to grey scale. Size and shape estimation of the same can be done by using two methods i.e.: using inbuilt functions or by programming method. Defect detection is done on real image by using k means algorithm. Color feature are taken from the real image and are stored in CSV file. Likewise, the text features are taken from the grey scale image are stored. Later-on the values are passed to SVM classifier for classification purpose. SVM is a supervised learning methodology used for labeling of data. Test images are tested with respected to trained data set and the percentage are given as the output.

IV. RESULT ANALYSIS

The output will be in the form of percentage and information about the size of the tomatoes. This information can be used to grade the tomatoes, to find the defect tomatoes and information regarding size. The model will help to detect the defects and ripeness in tomatoes, estimate the cost of the system and to differentiate tomatoes based on size.

V. FUTURE WORK

All the above information and steps can be used by the industries for large scale ripeness, size detection. Can develop a hardware model consists of conveyor belt system, pi camera and sorting system and motor. The conveyor belt is supported by motor while helps to move the tomatoes in a large scale and pi cam is used to take the images of the tomatoes used for further process in real time. Sorting system has storage bins to store red green bins.

CONCLUSION

The traditional method for checking the ripeness of tomatoes is very expensive and is filled with drawbacks. This system provides a method which is cost effective, and also helps to grade the tomatoes which will provide a better information to the farmers about the quality of tomatoes they have grown and to what price they can sell.

REFERENCES

[1] Ruchitha R.Mhaski,P B Chopade,M P Dale, "Determination of Ripeness and Grading of Tomato using Image Analysis on Raspberry Pi",CCIS 2015.

[2] Ratih Karthika Devi, R.V. Hari Ginardi, "Feature extraction for identfication of sugarcane rust disease", 2014

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTESIT - 2019 Conference Proceedings

[3] Kamalpreet Kaur ,O.P. Gupta "A Machine Learning Approach to Determine Maturity Stages of Tomatoes" , 2017

[4] H. Yin, Y. Chai, S. X. Yang and G. S. Mittal, "Ripe Tomato Recognition and Localization for a Tomato Harvesting Robotic System," 2009 International Conference of Soft Computing and Pattern Recognition, Malacca, 2009, pp. 557-562.

[5] J. B. U. Dimatira *et al*., "Application of fuzzy logic in recognition of tomato fruit maturity in smart farming," 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016, pp. 2031-2035

[6] K. Herwanda Tandrio, S. Palgunadi Yohanes and E. Suryani, "Prediction of Glucose Content and Level of Tomato Maturity with Backpropagation Based on Image Vector Feature," 2018 International Seminar on Application for Technology of Information and Communication, Semarang, 2018, pp. 429-434.