Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018)
IEEE Xplore Part Number:CFP18AWO-ART; ISBN:978-1-5386-4765-3

An Automated Tomato Quality Grading using Clustering based Support Vector Machine

Shweta S. Deulkar
School of Computer Engineering and Technology
MIT Academy of Engineering, Alandi
Pune, India
ssdeulkar@mitaoe.ac.in

Dr.Sunita S. Barve
School of Computer Engineering and Technology
MIT Academy of Engineering, Alandi
Pune, India
ssbarve@comp.maepune.ac.in

Abstract—**This paper focus on representing the technique of fruit grade classification, automated machine vision based technology has become more potential and important to many areas like agricultural sector and food processing industry. The proposed system which calculates the grade of fruit based on its external features. Grading of fruits is one of the most important processes in post harvesting, but this procedure is mostly carried out manually which is not efficient as it requires enormous number of employment, and tends to human error. The grading process is carried out by capturing the fruit image using digital camera and this image is interpreted using image various processing techniques. Color is very prominent feature for recognizing defect and ripeness of the fruit. The major objective is to check the fruit quality with high speed for analyzing maximum number of fruits in least amount of time. The conventional process of fruit quality assessment needs new tools deciding the quality of fruit. This system performs color features and size of fruit and captures the fruit side view image. The Otsu thresholding and K-Means clustering algorithms are used to extract the features of fruit. This system achieve the fruit quality sorting using Support Vector Machine and gives a few advantages over conventional techniques. This system will help in the advancement of an automated non destructive grading system with high speed, high accuracy and low cost. Implementation of this system will have applications in fruit quality grading in field like food science and trades where standardization is essential.**

Keywords—Tomato grading; Support vector machine; Computer vision; Classification, Clustering;

I. INTRODUCTION

India is facing the issue of increasing population, and increased food expectations. The losses in managing and processing of high quality products, there is a requirement of precise, speedy and desired quality identification system of agricultural product. Agriculture sector is one of the leading economic area in India. In economic development agriculture plays a key role in our country. In India because of the increasing population, increasing demand of high quality agricultural products, and losses in post harvesting process, became a reason to develop an accurate and fast quality grading system to obtain the high quality determination of fruits and vegetables.

The study states that post harvesting methods for fruits and vegetables will give more profits to the producer. It is therefore, essential to give equal importance to the post harvesting process as production [1-3].

India is second largest producer of tomatoes after china. Tomato is a climacteric fruit i.e. it continues ripening even after it is harvested. The grading of fruit is done based on its external and internal features, like shape, size, color, weight, texture, softness etc. The demand in fruit market is very selective and of high quality. The increasing demand and supply gives rise to the implementation of an automated fruit grading system for improving the quality [4-6].

In this paper we are proposing an Image processing based tomato quality grading system using external features of Tomato. The system is combination of three process such as binarization, feature extraction and classification considering the damaged percentage. All three units will be introduced further.

II. LITERATURE SURVEY

Momin, M. A., et al. [7] presented, "Geometry-based mass grading of mango fruits using image processing." Developed an image acquirement and preparing framework to remove anticipated zone, border, and roundness highlights. In this framework, pictures were obtained utilizing a XGA arrange color camera of 8-bit dim levels utilizing fluorescent lighting. A picture preparing calculation in view of area based worldwide thresholding color binarization, joined with middle channel and morphological investigation was produced to group mangos into one of three mass evaluations, e.g., huge, medium, and little. To accomplish a better reviewing, two diverse evaluating highlights could be utilized as a part of succession. The picture evaluating framework is straightforward and effective and can be viewed as an appropriate first stage to motorizing the business reviewing of mangos in Bangladesh. In addition, the technique can possibly be connected to different products with reasonable changes.

Nouri-Ahmadabadi, Hosein, et al. [8] presented, "Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine." In this survey, an astute framework in view of joined machine vision (MV) and SVM

(Support Vector Machine) was created for sorting of peeled pistachio. The framework was made out of lighting box, conveyor belt, categorizing unit, image processing unit, and camera. The pictures were digitalized by a catch card and exchanged to a PC for further investigation. At first, pictures were changed over from RGB (Red, Green, and Blue) to HSV color spaces. The segmentation process is carried out based on H-component in the HSV color space and Otsu thresholding method. An element vector of 30 color highlights was removed from the caught pictures. A component choice method in light of affectability examination was completed to choose unrivaled highlights.

Baigvand, Mehrdad, et al. [9] presented, The framework hardware was made out of a CCD camera, a belt conveyor, a feeder, a partition unit, and a lighting framework. Three quality lists, to be specific part size, and color, were first categorized by fig-processing specialists into the five different groups. Then, the pictures of the fig tests were caught utilizing a machine vision framework. In the first place, the pixels length in each picture and longitudinal directions of the focal point of gravity of fig pixels were separated. This algorithm decided color intensity and width of each fig as the identifier of its size and color, individually. For computing the split region, first the pictures were binarized by utilizing the color intensity distinction between the split and other parts product with a specific end goal to decide the region of the split area. A classification algorithm was additionally coded in Lab-VIEW for categorizing figs in view of their quality indices extricated by the picture processing algorithm into five subjective evaluations. In this method, the estimations of these highlights were contrasted and the threshold esteem that was foreordained by a specialist. Results represent that the created framework enhanced the sorting precision for every one of the classes up to 95.2%. The framework's mean rate was 90 kg/h for classifying figs.

Moallem [10] states, a vision-based algorithm for apple grading is presented which is implemented six stages. Background pixels as foundation are initially expelled from input pictures. Then, stem end is recognized by mix of Mahalanobis classifier and morphological technique. Calyx area is additionally distinguished by implementing K-implies bunching on the Cb component in YCbCr color space. From that point forward, deserts segmentation is carried out utilizing MLP (Multi-Layer Perceptron) neural system. Then, measurable, textural and geometric highlights from refined surrendered locales are removed. At long last, for apple grading, a correlation between execution of MLP, SVM, and KNN (K-Nearest Neighbor) classifiers is tested. Classification is implemented in two conducts. Firstly, an apple information is classified into two classes of healthy grade and defected grade. In the second way, the information is classified into three classifications of first rank grade, second rank grade and harmed grade ones. In both classification steps, SVM classifier performs the best one with acknowledgment rate of 92.5% and 89.2% for both the steps. In addition, the exactness of the segmentation algorithms including stem end discovery and calyx identification are assessed for two diverse apple picture databases.

Zaborowicz, Maciej, et al. [11] presented, "Application of neural image analysis in evaluating the quality of green house tomatoes." Presents the exploration on the utilization of strategies for artificial neural modeling and computer image analysis during the time spent surveying the nature of green house tomatoes assortment Capriccio. The subject of the investigation was tomatoes of the sizes from 40 mm to 67 mm and the colors: 1– 6, incorporate middle colors. The RMSE rate of system preparing for the preparation set was 0.075986, for the approval set it was 0.072194, and for the test set 0.061714.

Bhange, Manisha, and H. A. Hingoliwala. [12] presented, In this paper, we propose a web based apparatus that helps agriculturists for recognizing organic product infection by transferring natural product picture to the framework. The framework has an effectively trained dataset of pictures for the pomegranate organic product. Info picture given by the client experiences a few preparing ventures to distinguish the seriousness of sickness by contrasting and the trained dataset pictures. To start with the picture is resized and after that its highlights are extricated on parameters, for example, morphology, CCV, and color and clustering is finished by utilizing k-means algorithm. Next, Support vector Machine (SVM) is utilized for arrangement to group the picture as tainted or non-contaminated. A goal seek system is likewise given which is exceptionally helpful to discover the client intension. Out of three highlights separated we got best outcomes utilizing morphology. Test assessment of the proposed approach is compelling and 82% exact to distinguish pomegranate ailment.

Sofu, M. M., et al. [13] presented, "Design of an automatic apple sorting system using machine vision." This examination proposes a programmed apple arranging and quality assessment framework, which depends on real-time handling. Brilliant and Starking Delicious, and Granny Smith apple cultivars are arranged into various classes by their color, size and weight. It additionally distinguishes apples influenced by rot, scab and stain. The proposed framework comprises of a class conveyors, a roller and transporter joined with an encased lodge with machine vision, stack cell and control board units. The roller and transporter conveyors have two channels. With a precise end goal to dissect the visual properties of apples, two indistinguishable mechanical color cameras are determined to the roller transport. Four pictures of any apple moving on the transport can be caught and prepared utilizing picture handling programming in 0.52 s.

III. System Architecture

A. Architecture Overview

The system works on features, extracted from a binary image with the help of k-means clustering [5] [12]. The proposed image processing system is developed to differentiate the Tomato into three different categories like 1st grade, 2nd grade, and 3rd grade. For accurate and effective feature extraction, Tomato image is clustered and for better performance only discriminatory features are extracted after clustering. The proposed technique obtains high accuracy for Tomato grading of different grades when the SVM is implemented with linear kernal function.

Figure 1 shows a system architecture of the proposed grading system which classify the Tomato into three different categories. All these preprocessing steps are explained in the following subsections.

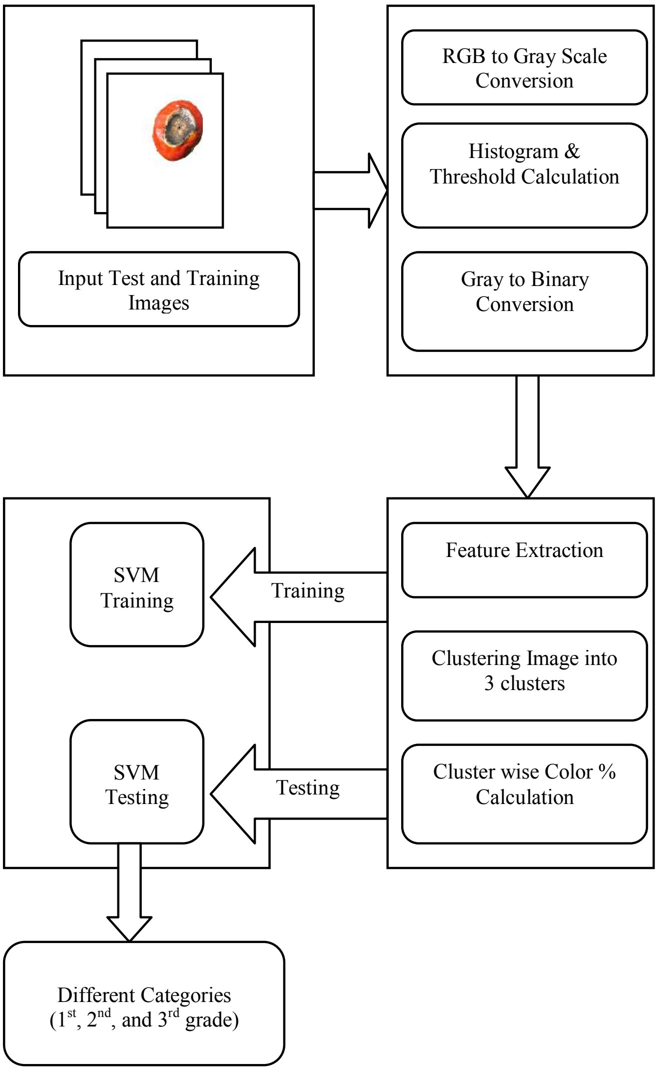

Fig. 1. System architecture.

B. *Sample collection and Image Acquisition*

Tomato Samples are collected from local market; the group of every sample fruit is determined from survey of 12 people. Every person response is recorded as one of the grade from fresh, medium damaged, damaged, and utmost votes obtained for a sample considered for particular group as label. Sample images are captured by a Canon D1100 camera and its specification is detailed below.

- Lens- 3x zoom
- 12.2 mega pixels
- Resolution of 4272 × 2848
- Interface- hi speed USB
- Sensor - CMOS

Tomato fruit were positioned at uniform background to reduce the efforts of background removal and camera was situated at distance of 30 cm from tomato and vertically at angle of 45°– 55° and manually slanting the side of fruit. Images are captured using two CFL of 25W and image size was fixed to 225*225 pixel, and in image format is of PNG. Total of 90 images are gathered successfully comprising of fresh, medium damaged and damaged. Below Table I show the sample dataset of all type grades [14] [15].

TABLE I. DATASET TOMATO SAMPLES REPRESENTING FIRST GRADE, SECOND GRADE, AND THIRD GRADE TOMATOES

1st grade Tomatoes			
2nd grade Tomatoes			
3rd grade Tomatoes			

C. *Image Preprocessing*

A digital camera is used to acquire the Tomato uniform background so as to get avoid color fluctuations in a images. The captured tomato images are in PNG format. For defected area detection based on color intensity histogram the color images are disintegrated into R, G and B color component. The captured colored images are transformed into gray scale image format using Otsu thresholding method. In this technique, a color intensity based histogram is calculated and displayed. Further the histogram is analyzed for different threshold values to obtain the optimal threshold value of the input image. The threshold value with minimum class variance is considered as optimal threshold for that image.

A thresholding algorithm is executed to get the binary image of tomato, where white represents tomato area and black represents background. Figure 2 shows are the results of histogram and binarization techniques [15-19].

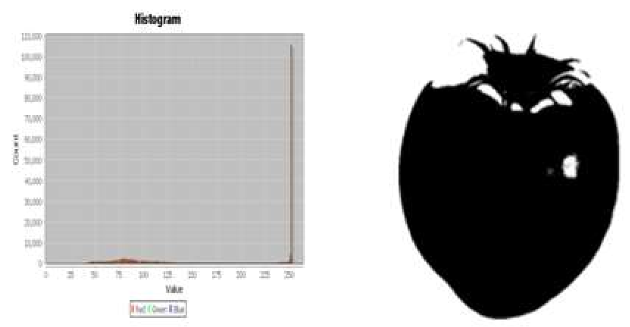

Fig. 2. Binarization results.

D. Kmeans Clustering

This algorithm groups the input data vectors into a predefined number of clusters. Initially the centroids of the predefined clusters are assigned randomly. The dimensions of the centroids are same as the dimensions of data vectors. Euclidian distance measure is used to determine the proximity of pixels and accordingly assign them to clusters. Mean of each cluster is re-calculated after assigning all the pixels to respective clusters. This is repeated either until no changes is noticed for all cluster means or for a fixed number of iterations [20-22].

The algorithm for K –means Clustering:

Input: Input Color image
Output: Three part clustered image
Process:
Step 1: Assign random centroid (k-centroid)
Step 2: Calculate distance from every point for every centroids [15].

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (1)$$

Step 3: Assign points to those centroids with m distance and make a clusters.
Step 4: Calculate mean for each cluster and make them new centroid.
Step 5: Repeat step 2 to 4 till we get same clusters.

E. Support Vector Machine Classifier

SVM is work in high-dimensional spaces seeking an optimal hyper-plane to separate the categories. SVM comprise of a set of related supervised learning methods used for classification and regression [14] [23]. SVM are supervised learning models with related learning calculation that break down information utilized for classification and regression analysis. Support vector machine depends on augmenting the base separation from the isolating hyper plane to the closest instance. Just binary classification is supported in essential SVM, yet for multi-class classification case expansion also can be achievable [13].

In these expansions, extra imperatives and parameters are added to enhancement issues for dealing with the detachment of the various classes. SVM is a binary classifier that implies the class marks can just take two qualities ±1 [24-26].

We are using two step SVM to categorize image feature into three category, In first step SVM image is categorized in to Damaged and Fresh fruit category. Further in the 2nd step of SVM. The fresh fruit category is classified into 1st grade and 2nd grade depending upon the fresh and damage area percentage.

Input: *Three part clustered image damage and fresh cluster area.*
Output: *Showing Actual three stage grading result (showing fruit category in the sense of 1st grade, 2nd grade, damaged grade)*
Process:
Step 1: categorize Image sectors into 2grade according to ratio i.e Fresh Percentage and Damage Percentage.
Step 2: check Damaged Ratio () in dataset
Step 3: check First Grade Ratio () in dataset
Step 4: check Second Grade Ratio () in dataset
Step5:if(OverallDamagedPercent > OverallFreshPercent) then check for damaged ratio as in step 2 if(available) then grade='damaged grade'.
Step 6: else then Grade='Second Grade'
Step7:if(OverallDamagedPercent=< OverallFreshPercent) then check for fresh ratio as in step 3
Step 8: if (available in range) grade='First Grade'
Step 9: else check for second grade ratio if (available) then grade='Second grade'
Step 10: else check for damaged grade ratio if (available) then grade='Damaged grade'
Step 11: Check Grading By Standardize range (Freshper Value)

IV. RESULT ANALYSIS

In this section, a series of tests have been performed to validate the quality of the automated tomato grading system. Through the tests, the accuracy and real-time classification algorithm executed in the proposed study are validated, and the accuracy and the computation time of system are tested and analyzed. The results obtained are presented in fig.3. The chart columns denote accuracy percentage for classification algorithm of every test dataset. The accuracy is calculated as per the equation 2 [4].

$$Accuracy = \frac{No.of\ correctly\ classified\ test\ samples}{Total\ No.of\ test\ samples} \times 100 \quad (2)$$

Fig.3.Disease detection accuracy for different group.

The Table II represents the accuracy of the classifiers for different test groups. Each test group is unique but equal in size, having 18 images each. The table cells denote the accurate predicted number of tomato for each category.

TABLE II. TEST RESULTS OF THREE DIFFERENT TEST DATASET

Test Groups	Existing System	Proposed System
Group 1	14	17
Group 2	11	16
Group 3	12	16

The Table III denotes the compared accuracy results of existing and proposed system. Here we are correlating the proposed system with the existing system for that we are using existing system represented in Bhange et al.,2015.

TABLE III. GRADE WISE ACCURACY OF TEST DATA

Different fruit category	Existing System (In percentage)	Proposed System (In percentage)
First Grade	87.5	95
Second Grade	72	90
Third Grade	85	92

A. Prediction of defective tomatoes

Tomato images are considered defective after the clustering. Different areas are represented as various color patches. Defected area might be spotted by red color and fresh area by green in clustering. These color features are observed and extracted to gain accurate classification of different grades of tomato.

B. Computational Time

The proper implementation of algorithm helps to improve the computational time. The hardware processing capacity also could directly affect the performance of classification system. Application requirement could give proper specifications to the algorithm. Computation time is calculated as processed data sample verses output time. It shows the speed of result calculations. In our case the computation time is 0.56 sec.

C. Accuracy

High accuracy is measured in terms of high precision and trueness, the proposed grading system satisfies both the terms compared to other existing systems. Our system has 91.66% accuracy.

V. CONCLUSION

The proposed project is the demo version for grading of tomato fruit based on the color. This is the first step towards sorting and grading of fruits based on color. Subsequently fruit size can be considered along with design of a hardware interfaced with the system to enable the sorting process. However the model can be used for large scale application by making the necessary changes in the hardware design as per the quantity of fruit to be processed simultaneously. However the total accuracy of the proposed work is sufficient for almost all of the real-time application, but it required more training to upgrade the accuracy. This system works on the color, texture image features for grading, but to improvise the results, we must consider the other features, like weight and size, tuning these values will help to improve the outcomes. Finally, considering the multi angular captured images will also help to improve the accuracy of the system.

REFERENCES

[1] G.Sharma, H. Singh, "Economic Analysis of Post-harvest Losses in Marketing of," vol. 24, no. December, pp. 309–315, 2011.

[2] P. K. Patel and D. A. Patel, "An economic assessment on post harvest losses in fresh vegetables <," vol. 3, no. 2, pp. 160–165, 2012.

[3] T. Report, N. A. D. View, and R. Hegazy, "Post-harvest Situation and Losses in India," no. May, pp. 1–21, 2011.

[4] N. Goel and P. Sehgal, "Fuzzy classification of pre-harvest tomatoes for ripeness estimation – An approach based on automatic rule learning using decision tree," vol. 36, pp. 45–56, 2015.

[5] H. P. W. Jayasuriya, "Development of a low cost machine vision system for sorting of tomatoes," vol. 15, no. 1, pp. 173–180, 2013.

[6] Mohamed Elhadj, Yahya O.Zemirli, Zouhir Ayyadi, Kamel "Advances in Intelligent Systems and Computing" vol.695, 2018.

[7] M. A. Momin, M. T. Rahman, M. S. Sultana, C. Igathinathane, A. T. M. Ziauddin, and T. E. Grift, "Geometry-based mass grading of mango fruits using image processing," *Inf. Process. Agric.*, vol. 4, no. 2, pp. 150–160, 2017.

[8] Nouri-Ahmadabadi, Hosein, et al. "Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine." Information Processing in Agriculture 4.4 (2017): 333-341.

[9] Baigvand, Mehrdad, et al. "Machine vision system for grading of dried figs." Computers and Electronics in Agriculture 119 (2015): 158-165.

[10] P. Moallem, A. Serajoddin, and H. Pourghassem, "Computer vision-based apple grading for golden delicious apples based on surface features," *Inf. Process. Agric.*, vol. 4, no. 1, pp. 33–40, 2017.

[11] M. Zaborowicz, P. Boniecki, K. Koszela, A. Przybylak, and J. Przybył, "Application of neural image analysis in evaluating the quality of greenhouse tomatoes," *Sci. Hortic. (Amsterdam).*, vol. 218, pp. 222–229, 2017.

[12] M. Bhange and H. A. Hingoliwala, "Smart Farming: Pomegranate Disease Detection Using Image Processing," *Procedia Comput. Sci.*, vol. 58, pp. 280–288, 2015.

[13] M. M. Sofu, O. Er, M. C. Kayacan, and B. Cetişli, "Design of an automatic apple sorting system using machine vision," *Comput. Electron. Agric.*, vol. 127, pp. 395–405, 2016.

[14] A. Sanaeifar, A. Bakhshipour, and M. De La Guardia, "Prediction of banana quality indices from color features using support vector regression," *Talanta*, vol. 148, pp. 54–61, 2016.

[15] J. Yousefi, "Image Binarization using Otsu Thresholding Algorithm," . May, 2011.

[16] P. K. Sahoo, S. Soltani, and A. K. C. Wong, "A survey of thresholding techniques," *Comput. Vision, Graph. Image Process.*, vol. 41, no. 2, pp. 233–260, 1988.

[17] J. S. Weszka, "A Survey of Threshold Selection Techniques," *Comput. Graph. Image Process.*, vol. 7, no. 2, pp. 259–265, 1978.

[18] T. Y. Goh, S. N. Basah, H. Yazid, M. J. Aziz Safar, and F. S. Ahmad Saad, "Performance analysis of image thresholding: Otsu technique," *Meas. J. Int. Meas. Confed.*, vol. 114, no. September 2017, pp. 298–307, 2018.

[19] S. Tamakuwala, J. Lavji, and R. Patel, "QUALITY IDENTIFICATION OF TOMATO USING IMAGE," no. 5, pp. 67–70, 2018.

[20] R. Kaur, M. Kumar, and M. Juneja, "Data Generation and Fruit Grading Using Different Classifiers," vol. 5, no. 1, pp. 189–196, 2018.

[21] C. H. Lin, C. C. Chen, H. L. Lee, and J. R. Liao, "Fast K-means algorithm based on a level histogram for image retrieval," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3276–3283, 2014.

[22] N. Di. Kamarudin, C. Y. Ooi, T. Kawanabe, H. Odaguchi, and F. Kobayashi, "A Fast SVM-Based Tongue's Colour Classification Aided by k -Means Clustering Identifiers and Colour Attributes as Computer-Assisted Tool for Tongue Diagnosis," *J. Healthc. Eng.*, vol. 2017, 2017.

[23] B. Zhang *et al.*, "Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier," *J. Food Eng.*, vol. 146, pp. 143–151, 2015.

[24] C. Zhang, C. Guo, F. Liu, W. Kong, Y. He, and B. Lou, "Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine," *J. Food Eng.*, vol. 179, pp. 11–18, 2016.

[25] A. Mizushima and R. Lu, "An image segmentation method for apple sorting and grading using support vector machine and Otsu's method," *Comput. Electron. Agric.*, vol. 94, pp. 29–37, 2013.

[26] J. Clement, N. Novas, J. A. Gazquez, and F. Manzano-Agugliaro, "Clasificadora inteligente de tomates de alta velocidad por color, tamaño y peso," *Spanish J. Agric. Res.*, vol. 10, no. 2, pp. 314–325, 2012.