Analysis of Infected Fruit Part using Improved K-Means Clustering Algorithm

Ridhuna Rajan Nair[1] Swapnal Subhash Adsul[2] Namrata Vitthal Khabale[3] Vrushali Sanjay Kawade[4] Anjali Sanjivanrao More[5]

[1,2,3,4]Student [5]Assistant Professor & Head of Dept.

[1,2,3,4,5]Department of Computer Engineering

[1,2,3,4,5]S.B.P.C.O.E. Indapur, SPPU, Pune, India

Abstract— Rapid increase in the import and export of fruits on large scale is increasing the qualitative production and risk in manufacturing the products. As the world is getting atomized, manual work is getting machine oriented. Considering the time dependency accuracy the drastic changes are taking place. Not only the clustering techniques but also the combined study of image processing made it easy for the defect detection in fruits. Using the improved K-means algorithm in this paper we are giving solution to the defect detection and the classification of fruits. Segmentation technique provides a new turning point to the analysis and detection of the fruit parts. Using some of the fruits as a sample for the experimental results.

Key words: Improved K-means Algorithm, Clustering, Image Processing, Defect Segmentation

I. INTRODUCTION

Data mining has a wide range of application in various fields of data extraction and analysis. Agriculture science has reached its extreme end using modern technology. Advanced image processing and photography using the segmentation and clustering technique is helpful in fruit quality assessment and fruit disease detection. Increasing the accuracy than the K-means algorithm considering the time complexity of the proposed system. Maintaining the quality of the product and increasing the efficiency of production is the aim of proposing this system. Feasibility and time consumption than the existing system is developed in this proposed system using the improved K-means clustering algorithm for gaining accuracy.

II. RELATED WORK

Sr. No.	Title of the paper	Year of publication	Technology used	Advantages	Disadvantages
1.	Study on fruit quality inspection based on its surface colour in produce logistics.	2010	Non-invasive and non-destructive fruit quality inspection method RGB-HSI.	Used in quality inspection with feasible result.	Only fraction of accuracy can be maintained.
2.	Detection and classification of Apple fruit diseases using complete local binary patterns.	2012	Use of K-means clustering.	Automation detection and classification of fruit disease.	Only one technique is used.
3.	Infected fruit part detection using K-means clustering segmentation technique	2013	Use of K-means clustering and segmentation technique.	K-means is used to determine the natural groupings of pixels present in an image.	It is more complex and dependent on K.
4.	Quality analysis and classification of bananas.	Jan 2014	Use of image processing.	Digital image processing can classify fruit in speed & accuracy.	It is work with only single banana.
5.	Machine vision applications to locate fruits, detect defects and remove noise: A review.	Mar 2014	Use Of machine vision Method.	Fruit sorting is done using machine vision system.	Sensors are required to detect infected fruit.

Table 1: Related Work

In the base paper the author presents infected fruit part detection using k-means clustering segmentation technique [1]. K-means is used to decide the natural grouping of pixels presents in the image. Clustering and segmentation technique is used to find defected part of fruit. K-means clustering is straight forward and very fast but drawback of using k-means clustering is that the output of k-means algorithm highly depends upon the selection of initial cluster center because the initial cluster are chosen randomly[1] [2]. The other limitation of the algorithm is to input required number of clusters.

In paper [2] author presents a fruit quality inspection based on its surface color in produce logistics. In this paper non-invasive and non-destructive fruit quality method as well as RGB to HIS method is used [2].Drawback of this technology is only fraction of accuracy can be maintained.

In other paper presents detection and classification of apple fruit disease using complete local binary pattern [3].Advantage of using complete local binary pattern is that automation detection and classification of fruit disease. But as we compare to other technologies there is one drawback

that is only one technique is used, can use fusion of technique for more accuracy [3].

Authors in [4] [5] presents a quality analysis and classification of bananas as well as machine vision applications to locate fruits, detect defects and remove noise. Image processing is used for determining infected fruit part [4]. Digital image processing used to find defected part. But the major limitation of this technology is that it is work with only single banana [4]. Machine vision technique is also used for finding infected fruit part [5]. The steps used in machine vision include the capturing the images, analysis and processing of characteristics in food products. The quality attributes such as shape, size, color and other external features are analyzed using machine vision technique. Fruit sorting is done using machine vision method [5]. Limitation of machine vision method is that sensors are required to detect infected fruit part [5].

III. OVERVIEW OF PROPOSED SYSTEM

A. Functional Steps:

1) The project will start with the Registration process.
2) Number of users can be sign in and use the software.
3) The main and very first task is to load the images of the fruit samples with its name.
4) The input image will undergo the processing as follows:
   – The input image will get converted from RGB to HSV value.
   – The centroid value can be calculated by using the Improved K-Means clustering algorithm.
   – The processed data will be compared with the database to obtain the final result.
5) The current input image is to be taken into consideration while processing.
6) While the fruit sample is being loaded to the software processing in the meantime provide the fruits name simultaneously.

For example: Take image of an Apple.
Specify name as "Apple".

7) Conversion of the image into the gray scale is necessary to get all the shades of the loaded image from 0 to 255. Due to the gray scale we can get the total shades transmitted or the shades of reflected light with visible wavelength.
8) Next Threshold process is to be carried out with the loaded image so as to isolate the relevant image from the whole digital image. Commonly threshold is carried out on the processed Gray scale image. The commonly used threshold techniques are histogram and multi-level threshold.
9) Boundary detection is to be carried out after the Threshold process to get the required area to undergo the processing.
10) Cropping of the required portion of the image is to be done.
11) Generation of the blocks after the segmentation is to be carried out in the image processing for the proper evaluation of the portion with defect.
12) The conversion of the RGB to HSV is to be carried out for the further processing of the centroid value.
13) The current centroid value for each block is to be calculated for the comparison with the data in the data base.
14) After the comparison and proper evaluation of the data items with the data set the result can be processed.
15) The final result is to be obtained with the defect detection and the disease caused to the infected fruit with Improved K-Means clustering algorithm.

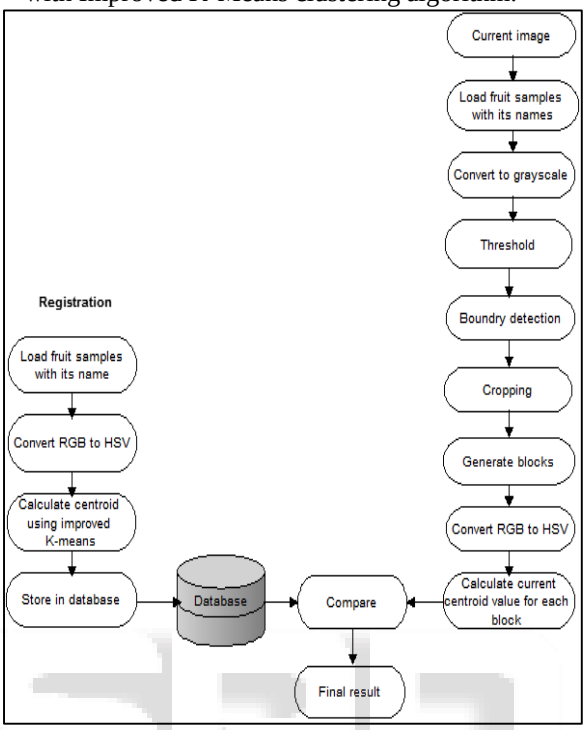

Fig. 1: Flow of Project

IV. PROBLEM STATEMENT

Analysis and detection of infected fruit part using image processing and segmentation technique using improved k-means and clustering.

V. SCOPE

The wide scopes of our proposed system are as follows,

A. In overseas trading of fruits

Manual checking of fruit quality takes large amount of time and at the time of importing and exporting of fruits, quality of fruits have to be better.

B. In Marketing and Sales of Fruits

In fruit markets and malls buying and selling of fruits is done on huge amount. For the quality maintenance purpose.

C. In Industrialization of Fruits

In food industries making foods like jams, chocolates, pickles etc. qualitative fruits have to be used. Food products directly relates to human health.

D. In Medicinal Productivity

Many of Ayurveda medicines are made up of fruits. Also health drinks are made by fruits therefore qualitative fruits have to be used in production of those.

E. In Cosmetic Industry

Cosmetics likes face wash, creams, lip balm's etc. are made up of fruits and these all are used on human skin that's why qualitative fruits have to be used at the time of production of cosmetics.

VI. INPUT OUTPUT BLOCK DIAGRAM

A. Input:

We are giving input as a image which we want to detect infected fruit part.

B. Processing:

We processed on the image by using improved k-means clustering and segmentation algorithm.

C. Output:

At output block we got result. Infected fruit part get detected.

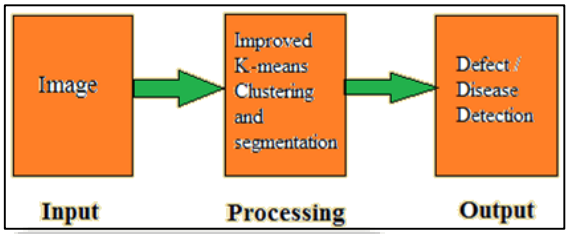

Fig. 2: Block Diagram

VII. ALGORITHM

Below we have proposed the improved k-means algorithm which does not require number of cluster (k). In this algorithm two clusters are created initially by choosing two initial centroids which are farthest apart in the data set. This is done so that in the initial step itself we can create two clusters with the data members, which are the most dissimilar ones.

A. Input:

D: The set of n tuples with attributes $A_1, A_2, \ldots, A_m$ where m = no. of attributes. All attributes are numeric.

B. Output:

Suitable number of clusters with n tuples distributed properly Method:

1) Compute sum of the attribute values of each tuple (to find the points in the data set which are farthest apart).
2) Take tuples with minimum and maximum values of the sum as initial centroids.
3) Create initial partitions (clusters) using Euclidean Distance between every tuple and the initial centroids.
4) Find distance of every tuple from the centroid in both the initial partitions. Take d=minimum of all distances. (Other than zero)
5) Compute new means (centroids) for the partitions created in step 3.
6) Compute Euclidean distance of every tuple from the new means (cluster centers) and find the outliers depending on the following objective function:
   - If Distance of the tuple from the cluster mean<d then not an Outlier.
7) Compute new centroids of the clusters.
8) Calculate Euclidean distance of every outlier from the new cluster centroids and find the outliers not satisfying the objective function in step 6.
9) Let B= $\{Y_1, Y_2 \ldots Y_p\}$ be the set of outliers obtained in step 8 (value of k depends on number of outliers).
10) Repeat until (B==<D)
   - Create a new cluster for the set B, by taking mean value of its members as centroid.
   - Find the outliers of this cluster, depending on the objective function in step 6.
   - If no. of outliers = p then
   - Create a new cluster with one of the outliers as its member and test every other outlier for the objective function as in step 6.
   - Find the outliers if any.
     - Calculate the distance of every outlier from the Centroid of the existing clusters and adjust the Outliers in the existing which satisfy the Objective function in step 6.
   - B= {Z1, Z2 .... Zq} be the new set of outliers.

(Value of q depends on number of outliers).

VIII. MODULE EXPLANATION

1) Login: Number of users can login to our software using their username and password. Only authorized users can login to our software. So, Security get maintain.
2) Registration: Number of users can register to software so as to use project. If any user is new to our system then user needs to register first.
3) Save dataset: In this module we are saving the dataset of different fruits. Dataset contains different features of different fruits such as color of fruits, disease of fruits. We are saving the different types of images of fruits.
4) Manage Dataset: This module is used to manage the manage the dataset. Given input image and saved dataset of image get compared with each other.
5) Prediction: Infected fruit part get predicted. Output get obtained.

IX. EXPERIMENTAL SETUP

For implementing proposed system we are using java language.

A. Hardware Requirement:

1) Dual core i3/i5 processor.
2) RAM- 512MB and above.
3) Memory- 20GB and above.

B. Software Requirement:

1) Netbeans 7.1
2) Jdk.

X. COMPARATIVE STUDY

Existing system	Proposed system
Existing system is more complex and dependent on K.	Proposed system is less complex and independent on K.
In existing system only	In proposed system more

fraction of accuracy can be maintained.	accuracy can be maintained.
Some existing systems are work with only single fruit.	Proposed system can work with different category of fruit.
Some existing system requires sensors to detect infected fruit.	Proposed system does not require sensors.
Time complexity is more.	Time complexity is less.

Table 1: Comparative Study

XI. EXPERIMENTAL RESULT

To validate the performance of proposed system, we have taken apple and mango as a case study. The proposed system estimates the infected fruits.

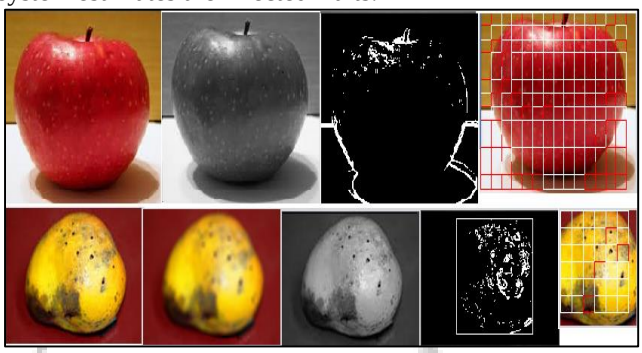

Fig. 3: Result Analysis

XII. CONCLUSION

As per the experimental setup proposed system detect infected fruit part using improved k-means clustering. Considering the wide range of application in the small scale and the large scale industries of the proposed system, we can conclude that the proposed system is more feasible with less time complexity and dependency.

XIII. FUTURE WORK

Accuracy of the proposed system may be increased in future using any other algorithm. In future android application can be implemented of proposed system.

ACKNOWLEDGMENT

We are thankful to our project guide and HOD Ass. Prof. More A.S., and project coordinator Ass. Prof. Gavali A. B., of computer department, S.B. Patil, College of Engineering, Indapur, for their valuable guidance and extreme support in the successful completion of this study. We are also thankful to all our friends for their positive support.

REFERENCES

[1] Shiv Ram Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta , ―Infected fruit part detection using K-means clustering segmentation technique‖, International journal of artificial intelligence and interactive multimedia,Vol.2,No.2.2013.
[2] Yizhong Wang, Yanhua Cui, George Q. Huang, Ping Zhang, Shaohui Chen, ―Study on fruit quality inspection based on its surface color in produce logistics‖, International conference on manufacturing automation,2010.
[3] Shiv Ram Dubey, Anand Singh Jalal, ―Detection and classification of apple fruit diseases using complete local binary patterns‖, Third international conference on computer and communication technology, 2012.
[4] Sanjay Chaudhary, Bhavesh Prajapati, ―Quality analysis and classification of bananas‖, IJARCSSE, Vol.4, Issue 1, Jan.2014.
[5] P. Vimala Devi and K. Vijayarekha, ―Machine vision applications to locate fruits, detect defects and remove noise: a review‖, RASAYAN journal, Vol.7, No.1, 104-113, Jan.-Mar. 2014.
[6] Madhuri A. Dalal, Nareshkumar D. Harale, Umesh L. Kulkarni, ―An iterative improved k-means clustering‖, ACEEE DOI: 02.ACE.2011.02.183.
[7] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta, ―A Comparative Study of Various Clustering Algorithms in Data Mining‖, IJERA, Vol. 2, Issue 3, pp.1379-1384, May-Jun 2012.
[8] Chunfei Zhang, Zhiyi Fang, ―An improved K-means clustering algorithm‖, Journal of information & computational science 10: 1 (2013)193–199.
[9] Anurag Bhardwaj, Ashutosh Bhardwaj, ―Modified K-Means Clustering Algorithm for Data Mining in Education Domain‖, IJARCSSE, Vol. 3, Issue 11, Nov. 2013.
[10] Anupama Chadha, Suresh Kumar, ―An improved K-means clustering algorithm: a step forward for removal of dependency on K‖, ICROIT 2014, India, Feb 6-8 2014.
[11] Shiv Ram Dubey, Anand Singh Jalal, ―Adapted approach for fruit disease identification using image‖, International journal of computer vision and image processing (IJCVIP), Vol.2, no.3, 2014.
[12] Vani Ashok, Dr. D.S. Vinod, ―Using K-means cluster and fuzzy c means for defect segmentation fruits‖, International journal of computer engineering and technology (ijcet), Vol.5, Iss.9, pp.11-19, Sep.2014.
[13] Ridhuna Rajan Nair, Swapnal Subhash Adsul, Namrata Vitthal Khabale, Vrushali Sanjay Kawade ―Analysis and detection of infected fruit part using improved k-means clustering and segmentation techniques, International Organization of Scientific Research (IOSR_JCE), PP 37-41, Dec 2015.
[14] Ridhuna Rajan Nair, Swapnal Subhash Adsul, Namrata Vitthal Khabale, Vrushali Sanjay Kawade, Anjali Sanjivrao More ―Analysis and detection of infected fruit part using improved k-means clustering and segmentation techniques, International Journal For Scientific Research & Development(IJSRD), Vol.3, Issue 10, 2015.

