IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN : 2278-0661, p-ISSN : 2278-8727
PP 37-41
www.iosrjournals.org

Analysis And Detection of Infected Fruit Part Using Improved k-means Clustering and Segmentation Techniques

Ridhuna Rajan Nair[1] , Swapnal Subhash Adsul[2] , Namrata Vitthal Khabale[3] , Vrushali Sanjay Kawade[4]
[1 2,3,4]*Computer Engineering, SBPCOE,India,*

Abstract: *Drastic increase in the overseas commerce has increased nowadays .Modern food industries work on the quality and safety of the products. Fruits such as oranges and apple are imported and exported on large scale. Identifying the defect manually become time consuming process. The combined study of image processing and clustering technique gave a turning point to the defect defection in fruits. This paper gives a solution for defect detection and classification of fruits using improved K-means clustering algorithm. Based on their color pixels are clustered. Then the merging takes place to a specific no of regions. Although defect segmentation is not depend on the color, it causes to produce different power to different regions of image. We have taken some of the fruits for the experimental results to clarify the proposed approach to improve the analysis and detection of fruit quality to minimize the precious and computational time. The proposed system is effective due to result obtained.*

Keywords *– Improved K-means Algorithm, Clustering, Image Processing, Defect Segmentation.*

I. Introduction

Data mining has wide application in several areas for the extraction of data such as medical diagnosis, traffic system monitoring, finger print recognition etc. Now-a-days digital image processing is one of the medium for displaying the required information. In now a days to days world modern technology in the agriculture science has reached its extreme height. The quality of the agricultural products is to be checked by the experts and analysis done is time consuming. The quality of fruit decides its value. Advanced photography and image processing using the image segmentation and clustering in the fruit disease detection can be helpful in the fruit quality assessment. To maintain the quality of fruit using the clustering technique we are interested only in the defected portion of the image. Using the improved K-means clustering algorithm in the proposed system for the image clustering and analysis of the defected part the quality of the fruit can be preserved. The accuracy is increased than the existing K-means clustering algorithm considering the time complexity of the proposed system. The main aim of proposing this system is to minimize the time and maintain the quality of product by image processing and segmentation using improved K-means clustering technique. Analysis and detection of infected fruit part using improved K-means clustering and segmentation technique is more feasible and less time consuming than the existing K-means clustering technique for gaining the accuracy.

II. Working of K-means clustering algorithm--

The K-means is the simplest and most commonly used algorithm employing Euclidean distance. It classifies a given data set into certain number of clusters (K in K-means represents number of clusters required). The procedure starts with initialization of K clusters with K centroids, one for each cluster. The clusters and their centroids are recomputed until all the data points in each cluster are at the minimum distance from their centroids. The distance here is taken as Euclidean distance as given in Equation 1

$deuc$ (x, y) $\sum_{i=i}^{n} \sqrt{(xi - yi)^2}$

The basic algorithm works as follows—

Algorithm 1: Basic K-means

Input: Dataset (D), Number of clusters (K)

Output: Elements of Dataset classified into K clusters

l. Select K initial Centroids (cluster centers) randomly form the given data set.

2. Repeat
   a) Assign all points in the dataset to the closest cluster center (centroids) to form K clusters.
   b) Recalculate centroids for each cluster to improve accuracy. Until no further improvement in accuracy.

We observe, K-means provides iterative convergence process to c1assity the entire numerical data into distinct clusters on some similarity parameters.

III. Literature Survey

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 includes 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].

IV. Proposed System

This paper gives a solution for defect detection and classification of fruits using improved K-means clustering algorithm. Based on their color pixels are clustered. Then the merging takes place to a specific no of regions. Although defect segmentation is not depend on the color, it causes to produce different power to different regions of image. We have taken some of the fruits for the experimental results to clarify the proposed approach to improve the analysis and detection of fruit quality to minimize the precious and computational time. The result obtained revel that proposed system is effective.

V. Improved K-means 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.

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

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 ... Y_p } be the set of outliers obtained in step 8 (value of k depends on number of outliers).
10) Repeat until (B==<D)

a) Create a new cluster for the set B, by taking mean value of its members as centroid.
b) Find the outliers of this cluster, depending on the objective function in step 6.
c) If no. of outliers = p then

i) 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.

ii) Find the outliers if any

d) 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.

e) B= {Z_1, Z_2 .... Z_q} be the new set of outliers. (Value of q depends on number of outliers).

VI. Flow of project

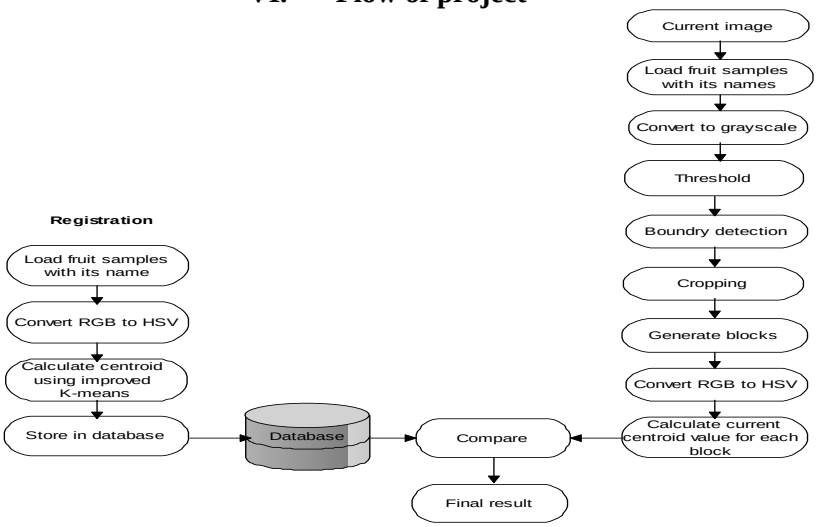

Figure1: Flow of project

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:
4.1)The input image will get converted from RGB to HSV value.
4.2)The centroid value can be calculated by using the Improved K-Means clustering algorithm.
4.3)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 mean-time 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.

Advantages
[1] No dependency on K –
As our software is used by technical as well as non-technical person no dependency of K value is very advantageous. In K-means clustering algorithm the input of required number of clusters i.e. K is must. And in our proposed improved K-means clustering algorithm we need not have to give value of K.

[2] Requires less time complexity -
As the value of K is predefined massive data gets classified rapidly as per the original K-means clustering algorithm. And time complexity decreases providing qualitative products.

[3] Increases qualitative production -
As the time required to clustering is on less amount and defects are detected efficiently, the qualitative production of fruits are increases on huge amount.

Applications
The huge range of applications of our proposed system are as follows,

[1] In import and export 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.

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

[3] In food Industry -

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

[4] In medical field –

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.

[5] In cosmetic production 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.

As per the above discussion our software helps to detect defected fruits for best choice. And plays vital role in the quality maintenance purpose.

VII. Conclusion

The inspection of agricultural products, fruits in particular, is an important process. Manual inspection is time consuming process. Automated inspection of reduces human interaction with goods, classify and detect defect of fruits faster than humans. The proposed approach used improved k-means clustering and segmentation technique. 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.

Acknowledgement

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 guidance and moral 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, Iss 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.