FRUIT QUALITY EVALUATION USING K-MEANS CLUSTERING APPROACH

[1]SURAJ KHADE, [2]PADMINI PANDHARE, [3]SNEHA NAVALE, [4]KULDEEP PATIL, [5]VIJAY GAIKWAD

[1,2,3,4,5]Department of Electronics, Vishwakarma Institute of Technology 666,
Upper Indira Nagar, Bibwewadi, Pune-400 037
E-mail: [1]khade.555@gmail.com, [2]mimipandhare@gmail.com, [3]sneha.navale13@gmail.com [4]kuldeeppatil20@gmail.com, [5]vijay.gaikwad@vit.edu

Abstract— Manual identification of defected fruit is very time consuming. In previous years, several types of image analysis techniques are applied to analyse the fruit quality. This work presents a novel defect segmentation of fruits based on texture feature with K-means clustering algorithm. This approach thus provides a feasible robust solution for defect segmentation of fruits.

The proposed method can process, analyze, classify and identify the fruits images, which are selected and sent in to the system based on color and texture feature of the fruit. The recognition system that has been developed is able to recognize all the test fruit images which are being selected by a user from the fruit selection menu which is based on GUI block in MATLAB on the system.

Apple is taken as a case study and evaluated the proposed approach using defected apples. The experimental results clarify the effectiveness of proposed approach to improve the defect segmentation quality in aspects of precision and computational time.

Keywords— Segmentation, K-means Clustering, GLCM.

I. INTRODUCTION

The naked eye observation of experts is the main approach adopted in practice for detection and identification of fruit diseases [2]. In traditional reliable quality method, human operators are used. It is boring and time consuming. For example most of the food products are clustered by seasonal workers [2]. Work shift is sometimes long and working condition is hard and mostly there is time limitation for reliability of products freshness .In order to improve the fruits' quality and production efficiency, to reduce labor intensity, it is necessary to research nondestructive automatic detection technology. Fruit nondestructive detection is the process of detecting fruits' inside and outside quality without any destructive, using some detecting technology to make evaluation according to some standard rules. Nowadays, the quality of fruit shape, default, color and size and so on cannot evaluate on line by using traditional methods [7]. With the development of image processing technology and computer software and hardware, it becomes more attractive to detect fruits' quality by using machine vision detecting technology. Appearance is a major factor in the judgment of quality and human eye has historically done this. The color indicates parameters like ripeness, defects, damage etc. and texture indicates the shape [8].

Defect segmentation of fruits can be seen as an instance of the image segmentation in which we are interested only to the defected portion of the image. In another way, segmentation of the image is nothing but pixel classification [4] [5].

The difficulty to which the image segmentation process is to be carried out mostly depends on the particular problem that is being solved. It is treated as an important operation for meaningful interpretation and analysis of the acquired images. It is one of the most crucial components of image analysis and still is considered as most challenging tasks for the image processing and image analysis [3]. It has application in several areas like analysis of Remotely Sensed Image, Medical Science and Fingerprint Recognition and so on.

This paper presents an efficient image segmentation approach using K-means clustering technique based on texture features from the images [1]. Defect segmentation is carried out into two stages. At first, the pixels are clustered based on their texture and spatial features, where the clustering processing is accomplished. Then the clustered blocks are merged to produces a high discriminative power for different regions of the image [6].

II. DESIGN METHODOLOGY

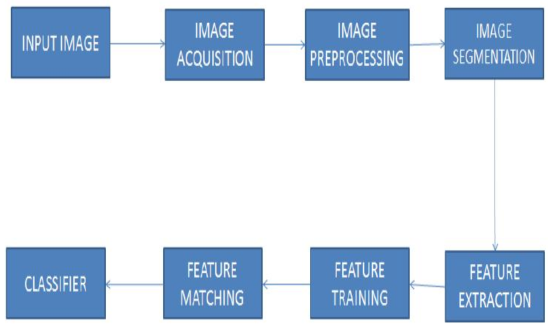

Fig.1 Block diagram of design methodology.

A digital image is produced by one or several image sensors, which besides various types of cameras. The image is acquired from cameras and files and grabbers directly into MATLAB. Incomplete, noisy and inconsistent data is removed, noisy data is smoothed out, tuples are ignored, and missing values are filled up and object reorganization is done in preprocessing.

Segmentation of one or multiple image regions which contain a specific object Image features at various levels of complexity are extracted from the image data. Typical examples of such features are: color, shape, texture. Classifying a detected object into different categories of interest.

III. PROPOSED METHODOLOGY

A specific number of regions. Using this two-step procedure, it is possible to increase the computational efficiency.

3.1 Image Segmentation

Image segmentation refers to the process of partitioning a digital image into N number of parts. The images are segmented on the basis of set of pixels or pixels in a region that are similar on the basis of some homogeneity criteria such as color, intensity or texture, which helps to locate and identify objects or boundaries in an image. Practical applications of image segmentation include object identification and recognition, facial recognition, medical image processing, criminal investigation, airport security system, satellite images, quality assurance in factories, etc. Due to the importance of the image segmentation, large number of algorithms has been proposed but the selection of the algorithm purely depends upon the image type and the nature of the problem.

3.3.1. Thresholding based segmentation

Thresholding is probably the most frequently used technique to segment an image. The thresholding operation is a grey value remapping operation g defined by:

$$g(v) = \begin{cases} 0 \ \ if \ \ v < t \\ 1 \ \ if \ \ v \geq t \end{cases}$$

Where v represents a grey value and t is the threshold value. Thresholding maps a grey-valued image to a binary image. After the thresholding operation, the image has been segmented into two segments, indented by the pixel values 0 and 1 respectively. If we have an image which contains bright objects on a dark background, thresholding can be used to segment the image. Since in many types of images the grey values of objects are very different from the background value, thresholding is often a well-suited method to segment an image into objects and background.

3.2 K-means Clustering K-means is one of the simplest learning and efficient algorithms that solve the well-known clustering problem which is color based segmentation. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters/subset). Clustering technique classifies the objects into different groups, partitioning of a dataset into cluster, so that the data in each cluster shares some common information according to some defined distance measurement. Data partitioning is a usual technique for the analysis of statistical data, which is used in many areas such as pattern recognition, image analysis and data mining. There are various methodology of clustering designed for a wide variety of purposes. K-means is a typical clustering algorithm. K-means is generally used to determine the natural groupings of pixels present in an image. It is an very effective because it is fast ,robust and easier to understand .it also Gives best result when data set are distinct or well separated from each other. K-means clustering computes the distances between the inputs and centers. These centers should be placed in a cunning way because of different location causes various result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to recalculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center .This algorithm assumes that a vector space is formed from the data features and tries to identify natural clustering in them. The objects are clustered around the centroids. A loop has been generated. As a result of this loop we may notice that the k center there location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm at minimizing an objective function known as squared error function given by :

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\| x_i - v_j \| \right)^2$$

Where,

$\| x_i - v_j \|$ is the Euclidean distance between x_i and v_j

'c_i' is the number of data points i^{th} cluster.

'c' is the number of cluster centers.

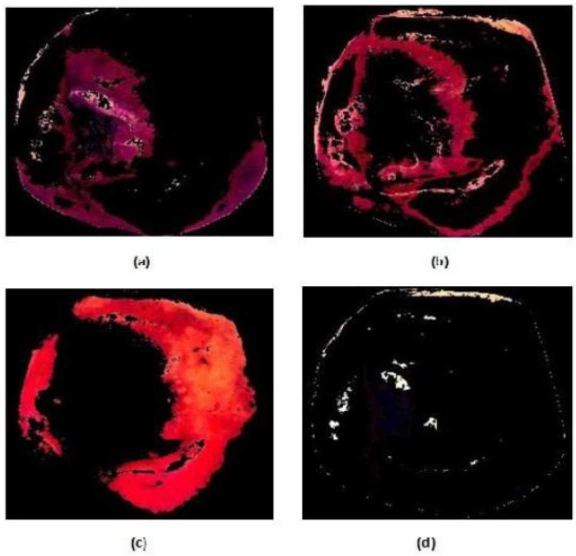

Fig.2.(a) Cluster number one (b) Cluster number two (c) Cluster number three and (d) Cluster number four.

3.21. Algorithmic Steps for K-means Clustering

Let $X=\{x_1,x_2,x_3........x_n\}$ be the sets of data points and $V=\{v_1,v_2,v_3..........v_c\}$ be the set of centers.

1) Randomly select 'c' cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
4) Recalculate the new cluster center.

$$v_i = (1/c_i)\sum_{j=1}^{c_i} x_i$$

Where 'ci' represents the number of data point in ith cluster. 5) Recalculate the distance between each data point and new obtained cluster centers. 6) If no data point was reassigned then stop, otherwise repeat from step 3). As a simple illustration of a k-means algorithm, consider the following data set consisting of the scores of two variables on each of seven individuals:

TABLE 1: Illustration of K-means clustering-A

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

This data set is to be grouped into two clusters. As a first step in finding a sensible initial partition, let the A & B values of the two individuals furthest apart (using the Euclidean distance measure), define the initial cluster means, giving:

TABLE 2: Illustration of K-means clustering-B

	Individual	Mean Vector (centroid)
Group 1	1	(1.0, 1.0)
Group 2	4	(5.0, 7.0)

The remaining individuals are now examined in sequence and allocated to the cluster to which they are closest, in terms of Euclidean distance to the cluster mean. The mean vector is recalculated each time a new member is added. This leads to the following series of steps:

TABLE 3: Illustration of K-means clustering-C

	Cluster 1		Cluster 2	
Step	Individual	Mean Vector (centroid)	Individual	Mean Vector (centroid)
1	1	(1.0, 1.0)	4	(5.0, 7.0)
2	1, 2	(1.2, 1.5)	4	(5.0, 7.0)
3	1, 2, 3	(1.8, 2.3)	4	(5.0, 7.0)
4	1, 2, 3	(1.8, 2.3)	4, 5	(4.2, 6.0)
5	1, 2, 3	(1.8, 2.3)	4, 5, 6	(4.3, 5.7)
6	1, 2, 3	(1.8, 2.3)	4, 5, 6, 7	(4.1, 5.4)

Now the initial partition has changed, and the two clusters at this stage having the following characteristics:

TABLE 4: Illustration of K-means clustering-D

	Individual	Mean Vector (centroid)
Cluster 1	1, 2, 3	(1.8, 2.3)
Cluster 2	4, 5, 6, 7	(4.1, 5.4)

Gray Level Co-occurrence Matrix (GLCM) Texture features can be extracted in several methods, using statistical, structural, model-based and transform in-formation, in which the most common way is using the Gray Level Co-occurrence Matrix (GLCM). GLCM contains the second-order statistical information of spatial relationship of pixels of an image. From GLCM, many useful textural properties can be calculated to expose details about the image content.

GLCM is the matrix that describes the frequency of one gray level appearing in a specified linear spatial r relationship with another gray level within the area of investigation. A co-occurrence matrix is defined over an image to be the distribution of co-occurring values at a given offset.

The grayco matrix function creates a gray level co-occurrence matrix by calculating how frequently a pixel with the particular intensity value i occurs in a specified spatial relationship to a pixel with value j.

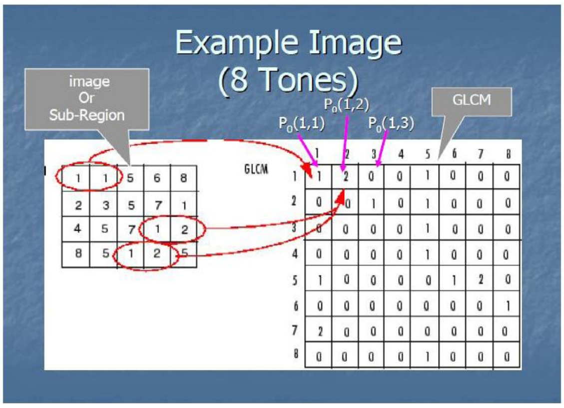

Fig.3 Example of GLCM

An image having gray level from 0 to 8 will create GLCM of size 8 by 8. Sub-region means window size of the image that we have considered. Sub-region means window size of the image that we have considered.

However, we can specify some other spatial relationship between two pixels to create multiple GLCM's, specify some array of offsets to the grayco matrix function. This offsets define pixel relationship of varying direction and distance.

TABLE 5: Calculating statistics from GLCM

Statistics	Description
Contrast	Measures the local variations in the gray-level co-occurrence matrix
Correlation	Measures the joint probability occurrence of the specified pixel pairs.
Energy	Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
Homogeneity	Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal

IV. DEFECT SEGMENTATION

Image segmentation using k-means algorithm is quite useful for the image analysis. An important goal of image segmentation is to separate the object and background clear regardless the image has blur boundary. Defect segmentation of fruits can be seen as an instance of image segmentation in which number of segmentation is not clearly known. Figure 4 shows the framework for the fruits defect segmentation. The basic aim of the proposed approach is to segment colors automatically using the K-means clustering technique and l*a*b* color space. The introduced framework of defect segmentation operates in six steps as follows

Step 1. Read the input image of defected fruits.

Step 2. Transform Image from RGB to L*a*b* Color Space. We have used L*a*b* color space because it consists of a luminosity layer in 'L*' channel and two chromaticity layer in 'a*' and 'b*' channels. Using L*a*b* color space is computationally efficient because all of the color information is present in the 'a*' and 'b*' layers only.

Step 3. Classify Colors using K-Means Clustering in 'a*b*' Space. To measure the difference between two colors, Euclidean distance metric is used.

Step 4. Label Each Pixel in the Image from the Results of K-Means. For every pixel in our input, K-means computes an index corresponding to a cluster. Every pixel of the image will be labeled with its cluster index.

Step 5. Generate Images that Segment the Input Image by Color. We have to separate the pixels in image by color using pixel labels, which will result different images based on the number of clusters.

Programmatically determine the index of each cluster containing the defected part of the fruit because K-means does not return the same cluster index value every time. But we can do this using the center value of clusters, which contains the mean value of 'a*' and 'b*' for each cluster.

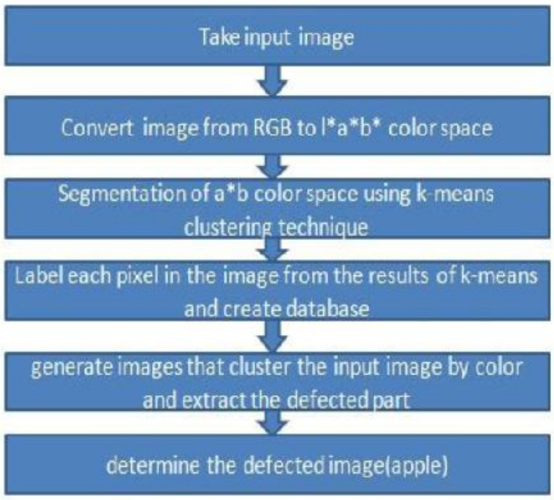

Fig.4 .Flowchart of defect segmentation.

V. ANALYSIS OF RESULT

To demonstrate the performance of the proposed approach apples are taken as a case study. The introduced method is evaluated on the defected apples. Defected apples are used for the defect segmentation.

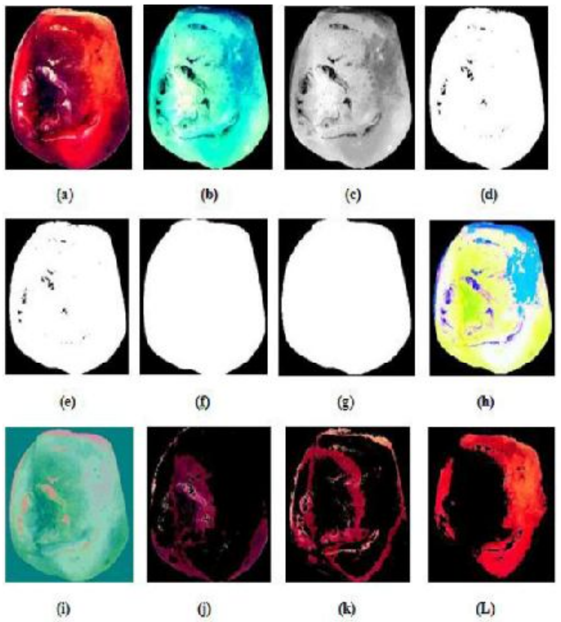

Fig.5.(a) Input image (b) Background subtracted image (c) background subtracted gray image (d) Background subtracted binary image (e) Preprocessing step I (f) Preprocessing step II (g) Preprocessing step III (h) HSV image (i) LAB image (j) Cluster number I (k) Cluster number II (L) Cluster number III.

Fig.5.(a) shows the input image of defected apple. The given input image is segmented in which background subtraction is done which results in clustering to increasing computational efficiency and speed of the process. The background subtracted image is as shown in Fig.5.(a),(b) and (c).But the problem in background subtracted binary image is that it produces the black spot due light. To eliminate black spots, filtering using morphing of an image is used. Fig.5.(e),(f) and (g)shows preprocessing steps I,II,III respectively. The subtracted image is then segmented into four clusters using K-means clustering algorithm.

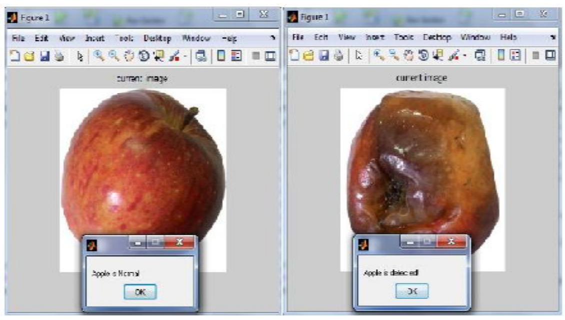

Fig.6. Experimental results.

Fig.6 shows the experimental results .The experimental result suggests that the implemented method for defect segmentation is robust and effective. It can accurately segment the defected part with the fruit region, background and analyze the defected as well as normal apple.

CONCLUSION

In this paper, the texture and color feature extraction Algorithm is implemented. K-means clustering algorithm is use for segmentation. Gray Level Co-occurrence Matrix is used for Feature analysis and feature extraction.

The system that has been developed is able to recognize all the test fruit images which are being selected by user or system tester from the fruit selection menu on the system There are some future works should be implement on the Fruits Quality Evaluation System in order to improve and enhance the functionality and flexibility of the system for more widely usage. The system should be improved by extending its functions to process and recognize more variety of different fruit images. Besides that, the texture based analysis technique could be combined with the existing three features analysis technique on the system in order to gain better discerning of different fruit images. The defected and normal apples are separated successfully.

REFERENCES

[1] Van Huy Pham. Byung Ryong Lee, “An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm”, Vietnam J Comput Sci (2015) 2:25–33 DOI 10.1007/s40595-014-0028-3

[2] J.Ramprabhu, S.Nandhini, “Embedded Based System for the Fruit Quality Management Using PIC Micro Controller” International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 4 Issue 1 January 2015, Page No. 10051-10056

[3] Ghobad Morad , Mousa Shamsi, Mohammad H. Sedaaghi, Setareh Moradi, “Using Statistical Histogram Based EM Algorithm for Apple Defect Detection” American Journal of Signal Processing 2012, 2(2): 10-14 DOI: 10.5923/j.ajsp.20120202.02

[4] Rohan Kandwal,Ashok Kumar ,Sanjay Bhargava,’ Existing Image Segmentation Techniques’, ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 4, Issue 4, April 2014

[5] A. M. Khan, Ravi. S , ‘Image Segmentation Methods: A Comparative Study’, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013

[6] Ms.Rupali S.Jadhav, PROF. S.S.Patil , “A Fruit Quality Management System Based On Image Processing” IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 8, Issue 6 (Nov. - Dec. 2013)

[7] Dr. Vilas D. Sadegaonkar, Kiran H. Wagh, “Improving Quality of Apple Using Computer Vision & Image Processing Based Grading System” International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438.

[8] Taun Anh Pham, “Optimization of Texture Feature Extraction Algorithm”, Computer engineering Mekelweg 4 2628 CD Delft the Netherlands http://ce.et.tudelft.nl/2010