RESEARCH ARTICLE OPEN ACCESS

Identification and Classification of Fruit Diseases

Ranjit K N[1] Chethan H K[2] Naveena C [3]
[1.] *Dept of CS & E, Maharaja Institute of Technology, Mysore, Karnataka*
[2.] *Dept of CS & E, Maharaja Institute of Technology, Mysore, Karnataka*
[3.] *Dept of CS & E, SJBIT,Bangalore, Karnataka*

ABSTRACT
Diseases in fruit cause major problem in agricultural industry and also causes economic loss.The diseases in fruits reduce the yield and also deteriorate the variety and its withdraw from the cultivation.So, earlier detection of symptoms of fruit disease is required.In this paper a solution for identification and classification of fruit diseases is proposed and experimentally validated.Ten different fruits has been selected with few of their diseases . The image processing based proposed approach is composed of following main steps,first step is segmentation using K-means and C-Means clustering algorithms,Second step is conducted a performance evaluation of segmentation algorithm by measuring the parameters such as Measure of overlapping (MOL), Measure of under-segmentation (MUS), Measure of over segmentation (MOS), Dice similarity measure (DSM), Error-rate (ER). The segmentation performance is calculated based on the comparison between the manually segmented ground truth G and segmentation result S generated by the image segmentation approach. After segmentation features are extracted using GLCM. The k Nearest Neighbours Algorithm classifier is used to classify the diseases in fruits .The collected database consists of 34 classes with 243 images .The experiment has been conducted on database of 34 classes having training samples of 30,50,70 percent of database .The result of classification has relatively higher accuracy in all cases when segmented using K-Means than C-Means clustering algorithms

Keywords: *K-MeansClustering, Performance measures,Gray Level Co-ocurrence Matrix.*

I. INTRODUCTION

India is the agricultural land where it stands as a second largest producers of fruits by producing over 44.04 million tons of fruits with an area of 3.72 million hectares and 10% of world fruit production is been accounted by India[1]. Variety of fruits are grown in India in which apple, banana, citrus, grape, mango, guava are the major ones. Fruit industry contributes about 20% of the nation's growth. But due to lack of maintenance ,improper cultivation of fruits and manual inspection there has been a decrease in production of quality of fruits.

Farmers find difficulty in finding the fruits affected with disease which result in loss of revenue to the farmers and the nation. The types of diseases on fruits determine the quantity, quality and stability of yield. The diseases in fruits reduce the yield and also deteriorate the variety and its withdrawal from the cultivation .Early disease detection can facilitate the control of fruit disease through proper management approaches such as fungicide applications, pesticide application and disease specific chemical applications can improve production. Visual perception by expert people was the classical approach for detection of diseases in fruits which are very expensive.

Fruits with different diseases as shown in Fig. 1. cause major problem in economy and production in agriculture industry worldwide manually identifying diseases in the fruits is used by the experts till date, but it is expensive for farmer to consult an expert , so automatically detect the symptoms of the fruit disease is required[2] .Diseases in fruit can cause major losses in yield and quality appeared in harvesting . majority of the works performing segmentation by K-Means clustering ([3],[4]) and C-Means clustering methods have been proposed in [5].performance evaluation of segmentation algorithm by measuring the parameters such as Measure ofoverlapping (MOL), Measure of under-segmentation (MUS), Measure of oversegmentation(MOS), Dice similarity measure (DSM), Error-rate (ER),Precision (P) and Recall (R) is conducted [6].Features are extracted using GLCM ([7],[8]) and diseases are classified using K Nearest Neighbours Algorithm [9].

We have proposed and experimentally validated the significance of using clustering technique for the disease segmentation and K Nearest Neighbours Algorithm as a classifier for the detection and classification of fruit diseases. In order to validate the proposed approach, we have considered 34 classes of diseases of 10 different fruits.

(a) (b)

Figure 1. Common fruit diseases : (a)bitter rot of apple, (b) Anthracnose of mango

II. THE PROPOSED APPROACH

The steps of the proposed approach are shown in the Fig. 2.Fruit disease classification problem require, precise imagesegmentation otherwise the features of the noninfected region will dominate over the features of the infectedregion. In this approach K-Means and C-Means based imagesegmentationis preferred to detect the region of interest which is the infected part only. After segmentation, features are extractedfrom the segmented image of the fruit. Finally, training andclassification are performed by KNN classifier

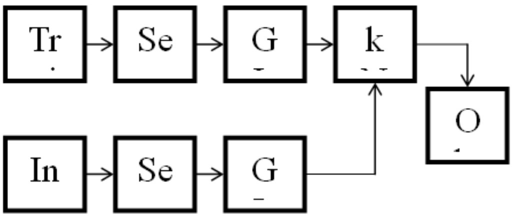

Figure 2. Block diagram of proposed system

A. Image segmentation

In K-Means clustering images are partitioned into four clusters in which one cluster contains the majority of the diseased part of the image. K-Means clustering method is used for the image segmentation. K-Means clustering algorithm [3] was developed by J. MacQueen (1967). The k-means clustering algorithms classify the objects into K number of classes based on a set of features.

Algorithm for the K-Means image segmentation –
Step 1. Read input image.
Step 2. Transform image from RGB to L*a*b* color space.
Step 3. Classify colors using K-Means clustering in 'a*b*' space.
Step 4. Label each pixel in the image from the results of K-Means.
Step 5. Generate images that segment the image by color.
Step 6. Select the segment containing disease.

C-means clustering method is used for the image segmentation.This Segmentation method is based on a basic region growing method and uses membership grades' ofpixels to classify pixels into appropriate segment[5].

BEGIN
Assumptions: Image transformed into feature space, number of clusters *c*,stop condition ε, fuzziness parameter *m*.
Step 1: Cluster image in feature space, with next conditions: number ofclusters is *c*, fuzziness index is *m*and stop condition is ε.
Step 2: Repeat for each pixel a_{ij} of image *I* .
Step 2.1: Find out, into which cluster C_l belongs pixel a_{ij} at most.
Step 2.2: Find out, whether in the closest surroundings of pixel a_{ij} exists segment R_k, which points belong to same clusterC_l
Step 2.3: If such segment R_k exists, than pixel a_{ij} add to segment R_k ,else create new segment R_n and add pixel a_{ij} to new segment R_n.
Step 3: Merge all segments, which belong to one cluster and are neighbors.
Step 4: Arrange borders of all segments.
END

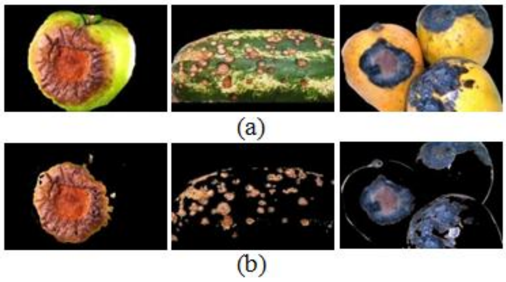

Figure 2 Image segmentation using K-Means clustering algorithm (a) Images before segmentation, (b) Images after segmentation

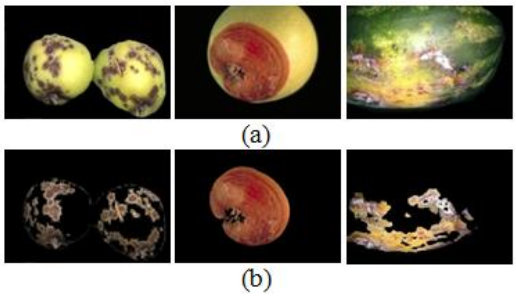

Figure 3 Image segmentation using C-Means clustering algorithm (a) Images before segmentation, (b) Images after segmentation

Fig. 4 and Fig.5 depicts image segmentation results using the K-Means and C-Means clustering technique

B. Performance Evaluation of Segmentation

Measuring the performance of a segmentation algorithm is necessary for mainly two reasons first is based on a performance metric, good parameter settings can be found for a segmentation algorithm and second is performance of different segmentation approaches can be compared. The region based measures are used when the size and location measurement of the area of the object is essential to detect the disease and is the objective of the segmentation. Therefore, region based measures are used to evaluate the segmentation algorithm[19] using (Equation. 1 to 5)

Measure of Overlap (Jaccard Similarity Measure)

$$MOL=\frac{S\cap G}{S\cup G} \quad (1)$$

where

MOL=Measure of overlap,S=Segmented area,G=Ground truth area

Measure of under Segmentation

$$MUS= \frac{|U|}{|G|} \quad (2)$$

where

U=Unsegmented lesion area

G=Ground truth area

$U= |G/(S \cap G)|$

Where

S= Segmented area

Measure of over Segmentation

$$MOS=\frac{|V|}{|S|} \quad (3)$$

Where

V =Segmented non lesion area

S= Segmented lesion area

$V= |S/(S \cap G)|$

Where

G =Ground truth lesion area

Dice Similarity Measure (DSM)

$$DSM=\frac{2\times |A(C)\cap A(C_r)|}{|A(C)+A(C_r)|} \quad (4)$$

Error Rate (ER)

$$ER=\frac{|A(C)\oplus A(C_r)|}{|A(C)+A(C_r)|} \quad (5)$$

Performance Evaluation of Segmentation is applied on K-means and C-Means clustering algorithms and results are noted.

C. Gray Level Co-ocurrence Matrix

The Gray Level Co-ocurrence Matrix (GLCM) method is a way of extracting the second order statistical texture features.GLCM's lower left triangular matrix is always a reflection of the upper right triangular matrix and the diagonal always contains even numbers. GLCM is dimensioned to the number of gray levels G and stores the co-occurrence probabilities gij..

Grey-Level Co-occurence Matrix can be defined as Equation 1:

$$c_{\Delta x,\Delta y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, if I(p,q) = i and I(p+\Delta x, q+\Delta y) = j\\ 0, \quad otherwise\end{cases} \quad (6)$$

Where m and n are grey levels normally m=n for symmetry GLCM

Normalized symmetry GLCM can be formed by Equation 2:

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \quad (7)$$

Where $V_{i,j}$ is the GLCM frequency and $P_{i,j}$ is the probability of cooccurence.

Some of the most common GLCM features are evaluated using (Equation.8 to 15):

Contrast:

$$\sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2 \quad (8)$$

Homogeneity:

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (9)$$

This statistic is also called as Inverse Difference Moment.

Angular Second Moment and Energy:

$$\sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (10)$$

The square root of the ASM is sometimes used as a texture measure, and is called Energy.

$$Energy=\sqrt{ASM} \quad (11)$$

Entropy:

$$\sum_{i,j=0}^{N-1} P_{i,j}\,(-\ln P_{i,j}) \quad (12)$$

Since ln(0) is undefined, assume that 0* ln(0) = 0.

Variance:

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}\,(i-\mu_i)^2 \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}\,(j-\mu_i)^2 \quad (13)$$

where

$$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j})\mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j})$$

are the means relative to ith pixel and jth pixel respectively.

Standard deviation are given by

$$\sigma_i = \sqrt{\sigma_i^2}\sigma_j = \sqrt{\sigma_j^2} \quad (14)$$

Correlation:

$$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(i-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right] \quad (15)$$

D. The k-Nearest Neighbors algorithm

In pattern recognition, the k-Nearest Neighbors algorithm is a non-parametric method used for classification. In k-Nearest Neighbors algorithm classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. If k = 1, then the object is assigned to the class of that single nearest neighbor.K nearest neighbors is a simple algorithm that stores all available cases and classifies the new cases based on a similarity measure. In the proposed method KNN classifier is used to classify the fruit diseases.

III. EXPERIMENTAL RESULT

A. Dataset preparation

The images in the dataset determine how realistic the analysis is. Database consists of 243 images of 10 different fruit type with 34 different fruit disease classification. Fruit images of Apple, banana, citrus, grape, guava, mango, papaya, peach, pomegranate, watermelon are collected. Each fruit type consists of different fruit diseases(e,g.,Bitter rot,Anthracnose).Fig. 4 Shows the few samples of images of different classes.

Figure 4 Sample images from the dataset

B. Result discussion

The collected database consists of 34 classes with 243 images .The images are segmented ,and features are extracted using GLCM .These extracted features are classified in to different fruit diseases using classifier.The experiment has been conducted on database of 34 classes having training samples of 30,50,70 percent of database .The result of classification has relatively higher accuracy in all cases when segmented using K-Means than C-Means clustering algorithms.

IV. CONCLUSION

An image processing based solution is proposed andevaluated in this paper for the identification and classification offruit diseases. The proposed approach is composed of main steps, in the first step image segmentation isperformed using K-Means and C-Means clustering technique. In the secondstep features are extracted. In the third step classification performed using KNN classifier.We haveused 34 classes of 243 images as a dataset.

Our experimental results indicate that the proposed solution can significantly identify andclassify the fruit diseases. Further work includes consideration of more dataset to improve the output of the proposed method.

REFERENCES

[1]. JagadeeshDevdasPujari,RajeshYakkundimathand AbdulmunafSyedhusainByadgi ," Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features"International Journal of Advanced Science and Technology Vol. 52, March, 2013

[2]. Manisha A. Bhange, Prof. H. A. Hingoliwala, "A Review of Image Processing for Pomegranate

[3]. Disease Detection"International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 92-94

[4]. Shiv Ram Dubey, Anand Singh Jalal,"Detection and Classification of Apple FruitDiseases using Complete Local Binary Patterns" 2012 IEEE

[5]. Shiv Ram Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta,"