THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Random Detection of Anthracnose of Mango by Differential Clustering

Murali Krishnan
PG Scholar, M.E Communication Systems
Bannari Amman Institute of Technology
Dr. M. G. Sumithra
Department of Electronics and Communication
Bannari Amman Institute of Technology

Abstract:
Mango anthracnose is an infection caused by the ubiquitous fungus Colletotrichum gloeosporioides. It is the anamorphic stage (asexual stage of the pathogenic fungus).C. gloeosporiodesis responsible for many diseases, also referred to as "anthracnose," on many tropical fruits including banana, avocado, papaya, coffee, passion fruit, and others. Ripe fruits affected by anthracnose develop sunken, prominent, dark brown to black decay spots before or after picking. Fruits may drop from trees prematurely.The fruit spots can and usually do coalesce and can eventually penetrate deep into the fruit, resulting in extensive fruit rotting. Most green fruit infections remain latent and largely invisible until ripening.So detection of the infected areas of the plant leads to early analysis and elimination of disease spread and saves the plant from disease spread. Proposed scheme uses a novel approach to detect the infected areas of the plant by segmenting the plant leaves by clustering. K-means clustering is performed for different cluster centres to obtain different cluster groups of the region of interest (ROI).

Key words: *anthracnose, Colletotrichum gloeosporioides, pathogenic fungus ,segmentation, clustering, k-means clustering, cluster centres, Region of interest (ROI)*

1. Introduction

Anthracnose is one of the most serious diseases affecting mangoes. The most devastating effects of anthracnose occur in areas where it rains during the mango flowering , ripening, fruit set stages.

Mango anthracnose is caused by the fungus Colletotrichum gloeosporioides varminor (also known by the name of its perfect stage Glomerella cingulata varminor)[15]. Spore production by this fungus is favored by wet or humid weather. The dispersal of these spores is particularly favored by rain and wind. This enables spread of the disease over relatively short distances. In areas where rain is prevalent during flowering and fruit set, anthracnose can cause destruction of the inflorescences and infection and drop of young fruit. This can obviously lead to serious losses. Many methods such as use of fungicides and artificial preservatives have been adopted to prevent the infection. Large scale use of these harmful repellants cause adverse effect on the consumers.

In the proposed scheme, use of image processing techniques to detect the affected area provides a cost effective approach for eradication of the diseased areas of the fruit. Clustering approach is adapted to segment the region of interest (ROI) and to perform disease detection. Based on a comparative study using Matlab, k-means clustering [4], [5] proved to be highly effective in finding the diseased areas of the fruit with a privilege for the user to select his own areas for clustering.

2. Prior and Related Work

2.1. Image Processing

Many antibodies and repellents' have been used or eradicating the fungal infection. anthracnose causes mango flowers to appear withered and blackened, and therefore, reduced fruit set results. Control this disease by spraying a copper-based fungicide during flowering and continue throughout the growing season. But these methods proved inefficient for a long term repetitive use resulting in negative impact on the plant growth. Earlier in 2012-13 scientists have tried to use Neural network classifiers for identifying and grading based on percentage of affected area and neural network classifier is used to classify normal and anthracnose affected on fruits. But the network requires huge amount of training data and run time. A random, quick and efficient algorithm is proposed in this work. After assigning random cluster centers the images obtained are iterated for forming cluster groups which specify a special cluster head. An algorithm for segmenting the ROI based on k-means clustering is proposed in this article.

2.2. Segmentation

Segmentation is the classifications of an image into meaningful data for easy analysis of the image. Some of the existing methods for segmentation are thresholding, region growing and clustering. Thresholding is the simplest method of image processing. From an RGB image converted to the corresponding gray level intensity image [1], we can partition the image by binary values, 1 and 0. The region above the threshold may be assigned 1 and that below the threshold may be assigned zero. This histogram approach cannot be relayed upon for effective classification of the image information as the binary approach of classification limits the representation of image segments and further reduces proper detection of the required area. Fig.2 shows segmented fruit image by thresholding with less accuracy in the detection of infected area. The region above threshold includes plant areas less affected with infections and may lead to misinterpretation of the affected area. Region growing is another approach which can be used to extract the plant features. But the disadvantage of this approach is that it is a split and merge algorithm which focuses on selecting pixels with approximately similar intensity values and comparing the adjacent areas to fill them with average intensity which lacks clarity for the required image information.

2.3. Clustering Methods

Segmentation approaches based on clustering has many advantages over other approaches as it provides an efficient classification of image information and can be implemented in many fields of human interest such as aviation, military and medical fields. The implementation of segmentation on agriculture has aroused the interest of many scholars for it paves an easy to implement and effective method for detecting various pathogens and it is harmless due to low consumption of artificial pesticides and herbicides.

The various clustering approaches existing are c-means, fuzzy C-means, expectation maximisation algorithm and K-means clustering.

Relevant to segmentation in real time plant images obtained at regular intervals , a brief discussion on Fuzzy c means and k-means clustering is depicted in the following sessions.

Fuzzy C-Means Clustering is a soft version of K-means, where each data point has a fuzzy degree of belonging to each cluster. Gaussian mixture models trained with expectation-maximization algorithm (EM algorithm) maintains probabilistic assignments to clusters, instead of deterministic assignments, and multivariate Gaussian distributions instead of means. Several methods have been proposed to choose better starting clusters. One recent proposal is k-means++.

The filtering algorithm uses kd-trees to speed up each k-means step. Some methods attempt to speed up each k-means step using corsets or the triangle inequality. Escape local optima by swapping points between clusters. The Spherical k-means clustering algorithm is suitable for directional data.

Even though fuzzy system is useful in effective classification, it is not adaptable for colour clustering and for plants with multiple infections. From a theoretical analysis the final clustered image has the unpredictability of being a pest or an infection. So colour clustering based on k-means is more efficient in agricultural crop segmentation.

The two key features of k-means which make it efficient are often regarded as its biggest drawbacks:

Euclidean distance is used as a metric and variance is used as a measure of cluster scatter. The number of clusters k is an input parameter: an inappropriate choice of k may yield poor results. That is why, when performing k-means, it is important to run diagnostic checks for determining the number of clusters in the data set. Convergence to a local minimum may produce counterintuitive results.

A key limitation of k-means [5] is its cluster model. The concept is based on spherical clusters that are separable in a way so that the mean value converges towards the cluster centre. The clusters are expected to be of similar size, so that the assignment to the nearest cluster centre is the correct assignment. When for example applying k-means with a value of $K=3$ onto the well-known Iris flower data set, the result often fails to separate the three Iris species contained in the data set. With K=2, the two visible clusters (one containing two species) will be discovered, whereas with K=3 one of the two clusters will be split into two even parts. In fact, K=3 is more appropriate for this data set, despite the data set containing 3 *classes*. As with any other clustering algorithm, the k-means result relies on the data set to satisfy the assumptions made by the clustering algorithms. It works well on some data sets, while failing on others.

The Gaussian models used by the Expectation-maximization algorithm (which can be seen as a generalization of k-means) are more flexible here by having both variances and co variances. The EM result is thus able to accommodate clusters of variable size much better than k-means as well as correlated clusters (not in this example).

Fuzzy C means Clustering	K-means clustering
Fuzzy C-means clustering is a soft version of K-mean, where each data point has a fuzzy degree of freedom belonging to each cluster. It is based on a logical approach to select the ROI for classification.	The two key features of k-means which make it efficient are often regarded as its biggest drawbacks: Euclidian distance is used as a measure of cluster scatter. The number of clusters k is an input parameter: an inappropriate choice of k may yield poor results. That is why, when performing k-means, it is important to run diagnostic checks for determining the number of clusters in the data set. Convergence to a local minimum may produce counterintuitive results.
Expectation Maximisation Algorithm (EM algorithm) Gaussian mixture models trained with Expectation maximisation algorithm (EM algorithm) maintains probabilistic approach to clusters, instead of deterministic assignments, and multivariate Gaussian distributions instead of means. Several methods have been proposed to choose better starting clusters. One recent proposal is k-means++.	A key limitation of k-means is its cluster model. The concept is based on spherical clusters that are separable in a way so that the mean value converges towards the cluster centre.

Table 1: Comparison of different clustering algorithms

3. Proposed Method

3.1. Description of Algorithm

The algorithm is very similar to Forgy's algorithm [13]. Besides the data, input to the algorithm consists of k, the number of clusters to be developed. Forgy's algorithm is iterative, but k-means algorithm makes only two passes through the data set.

- Begin with k cluster centres, each consisting of one of the first k samples. For each of the remaining n-k samples, find the centroid nearest it. Put the sample in the cluster identified with this nearest centroid. After each sample is assigned, recompute the centroid of the altered cluster.
- Go through the data a second time. For each sample, find the centroid nearest it. Put the sample in the cluster identified with this nearest centroid. (during this step, do not recompute any centroid)

Addition of certain features in the existing k means algorithm improves the detection of the interested region effectively with minimum chance of faulty clustering. The first step in k-means clustering is the initialisation of cluster centres. Common methods for initialisation include randomly chosen starts or using hierarchical clustering to obtain k initial centres [1], [2], [5]. The initialisation steps can be explained as follows.

- Convert n×p image matrix X to n×(p-1) matrix Z, where each row Z_i of Z is the polar representation of the corresponding row $\left(X_i \in S^p \right)$ of X.
- For each column Z, find the pair of neighbouring points with the largest angular distance between them and rotate Z such that these neighbours have the largest linear distance between them.
- One dimensional matrix for k-means is initialised with greatest value integer obtained from $(K (P-2)^{1/(p-2)}$ equi-spaced quantities [3],[4].

3.2. Applying masking to K-means algorithm

For a given k and initial cluster centres {μ_k; k=1…k}, the general strategy is to position the datasets into k clusters, then to iterate the cluster mean directions until convergence [9]. The exact algorithm can be explained as follows.

- Given k initialising cluster mean directions μ_1, $\mu2,…\mu_k$, find the two closest mean directions for each observation X_i ; i= 1,2…n.
- Classify the groups by C_{1i} and C_{2i} respectively. Assign the update equation

$$V_k^- = (n_k- 1)^2 – n_k^2 ||X_k||^2 -1 \quad \text{and} \tag{1}$$
$$V_k^+ = (n_k+ 1)^2 – n_k^2 ||X_k||^2 -1 \tag{2}$$

All clusters are in the live image set at this stage.

- The live set is updated to find optimum convergence [2].
- Optimum transfer stage: For each X_i, i= 1,2…n, we calculate the maximum reduction in the objective function. By replacing the live function ζ_i with another class, maximum reduction can be obtained as
-

$$\boxed{\begin{aligned} W_i &= \left(n_{ki} +1 \right)\left(V_k^+ - 2n_{ki} X_k^{-1} X_i \right) - \\ &\quad \left(n_{\varsigma i} -1 \right)\left(V_{\varsigma i}^- + 2n_{\varsigma i} X_{\varsigma i}^{-1} X_i \right) \end{aligned}} \tag{3}$$

If Wi > 0, then the only quantity to be updated is $C_{2i} = K_i$.

- Quick transfer stage includes swapping and the objective function and the change in the objective function can be calculated as

$$Obj_k = \sum_{k=1}^{k} n_k \left(1 - \|X_k\|\right) \qquad\qquad (4)$$

providing a quick way of obtaining final value.

The exact extraction of the diseased areas of the fruit can be detected by masking the clustered sample containing the plant region and then subtracting it from the acquired image. The modified algorithm developed using k-means clustering can be discussed with the experimental results obtained from a diseased fruit.

4. Experimental Results and Discussion
- MATLAB Implementation Results
The proposed scheme can be explained briefly with the results obtained after segmentation as follows.

4.1. Image Acquisition
Image acquisition is simply the conversion of the image obtained from real time environment to
computer readable formats such as JPEG, PNG etc. The acquired image is shown in figure1.

Figure 1. Acquired image of mangoes affected with anthracnose

4.2. Segmentation to Extract Objects in the Cluster
After acquiring the image, clustering is done to separate the background and the foreground image. This is done by updating the live set with cluster groups of lower intensities as a group. In this step it is checked whether the Cn_i group satisfies $Wi > n$. The foreground image is thus mapped with a zero level intensity to perform further logical operations on the cluster group.

4.3. Clustering Based on Intensity Mapping
The acquired image is then separated to cluster groups based on k-means clustering. Then gray level mapping is performed to separate the image to intensity fields which helps in separation of the fruit image from the obtained image. The fruit image area to be highlighted is labelled in with the cluster index as shown in figure 3.B,4.B and 5.B.

4.4. Highlighting the Fruit Area
The pixel groups belonging to the intensity marked area alone are extracted and is shown in the figure 3C,4C and 5C and 5D to obtain the highlighted fruit area alone from the acquired image. These images are used as the reference images to subtract them from the base images .The extracted fruit image is further corrected by masking with an image matrix of similar intensity pixels. It provides a better clarity for the obtained image and aids the separation of image using distinguishable features of the fruit image.

4.5. Subtracting the Clustered Fruit Image from the Base Image
The final clustering is done by subtracting the reference image from the base image formed by subtracting the cluster with fruit details from the input image. The corresponding output contains the diseased area alone.The advantage of this algorithm is that it gives high precision with low operating time. The final clustered image showing the disease alone is shown figures 3D,4D and 5D.

The overall flow of the program can be summarised with the following steps
- Step1: Acquiring the image.
- Step2: Storing the ROI as the base image to be clustered for further operations.
- Step3: Cluster to extract useful fruit area from the ROI
- Step 4: Storing the fruit image obtained after applying the cluster field and using it as reference image.
- Step 5: Subtracting the reference image from the base image.

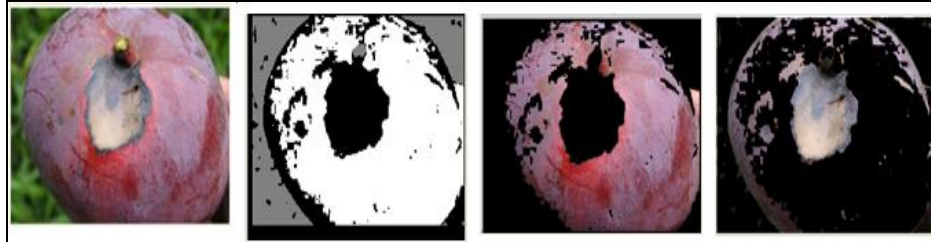*Figure 2. Acquired image of mangoes affected with anthracnose*

*Figure 3. This image shows the primary symptom of antracnose disease which is the scalped region in the input image.
A).input image B)image labelled by cluster index. C)Useful areas or plant part.
D).Area affected with anthracnose obtained after subtraction from the reference image*

*Figure 4. This image shows antracnose (dark region) affected mango .A).input image B)image labelled by cluster index.
C)Useful areas or plant part. D).Area affected with anthracnose obtained after subtraction from the reference image*
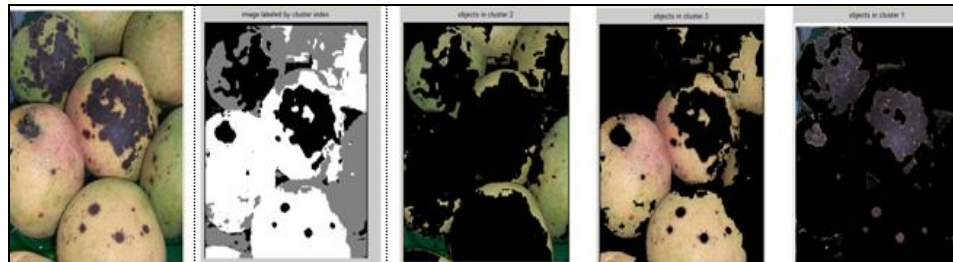

*Figure 5. This image shows a less severe infecton of the disease in mango .A).input image
B)image labelled by cluster index. C)Useful greeny areas of fruit. D). Useful riped Area of fruit.
E) Area affected with anthracnose obtained after subtraction from the reference images(both ripened*

5. Conclusion and Future Work
Early detection of the plant diseases helps in eradication of various ailments associated with the plant and it helps to improve the plant growth to a larger scale. The application of clustering algorithm for detecting the disease is an easy to implement and affordable method which also helps in reducing manual labour in fields. Segmentation approach can also be extended in finding diseases of plant species affected with infections due to environmental conditions and it can also be used to monitor weather conditions hazardous to plants. Compared to Fuzzy C means clustering k means proves simple and effective in determining the infected area with reduced requirement of manual cluster selection. As in major plants or trees the symptoms of infections are first noticed on the leaves and stem it can be easily detected my means of image processing before the Xylem of the plant starts circulating the infections to other parts of the trees or plants. With the help of ADSP target boards and FPGA tools we could further implement the algorithm on aerial time basis. With the help of advanced microscopic devices the host organisms can be detected by a detailed study on the behavioural and growth stages of the organism. Clustering algorithms based on constraints and certain relaxation algortithms could also be enhanced to provide accuracy to the proposed scheme. An algorithm for updating the clusters through iteration could further improve the obtained results.

6. References

1. Constrained K-means Clustering with Background Knowledge Proceedings of the Eighteenth International Conference on Machine Learning, 2001, p. 577{584. Kiri Wagsta, Seth Rogers rogers@rtna.daimlerchrysler.com Stefan Schroed.
2. An Efficient k-Means Clustering Algorithm: Analysis and Implementation Member, IEEE Tapas Kanungo, SeniorMember, IEEE, David M. Mount, Member, IEEE,Nathan S. Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, Senior
3. Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation Siddheswar Ray and Rose H. Turi School of Computer Science and Software Engineering Monash University, Wellington Road, Clayton, Victoria, 3168, Australia.
4. Constrained K-means Clustering.P.S.Bradley, K.P.Bennet, A.Demiriz.MSR-TR-2000 65
5. .Hopkins, D.L., and W.C. Adlerz. 1988. Natural hosts of Xylella fastidiosa in Florida. Plant Disease 72: 429-431.
6. Hammerschlag, R., J. Sherald, and S. Kostka. 1986. Shade tree leaf scorch. Journal of Arboriculture 12: 38-43.
7. Hearon, S.S., J.L. Sherald, and S.J. Kostka. 1980. Association of xylem-limited bacteria with elm, sycamore, and oak leaf scorch. Canadian Journal of Botany 58: 1986-1993.
8. K. Fukunaga, Introduction to Statistical Pattern Recognition, 2ndedition. New York: Academic, 1990.
9. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24,no. 5, pp. 603–619, May 2002.
10. B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: a texture classification example," in Proc. 9th Int.Conf. Computer Vision, Nice, France, 2003, pp. 456–463.
11. Pattern Recognition and image analysis., Earl Gose ,Richard Johnsonbaugh, Steve Jost ,Prentice- Hall, India. Pg-210-219
12. P. Boissard,V.Martin, and S. Moisan. A cognitive vision approach to early pest detection in greenhouse crops. Computer and Electronics in Agriculture Journal, 62(2):83–93, April 2008.
13. C.P. Wijekoon et al. / Journal of Microbiological Methods 74 (2008) 94-10
14. Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features Jagadeesh Devdas Pujari, Rajesh Yakkundimath and Abdulmunaf Syedhusain Byadgi. International Journal of Advanced Science and Technology Vol. 52, March, 2013