International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 3, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Defect Detection in Alphonso using Statistical Method and Principal Component Analysis: A Non-Destructive Approach

Sandeep S. Musale, Pradeep M. Patil

Abstract – **Natural image analysis uses textural property of the surface. Texture is defined as a spatial arrangement of local intensity attributes that are correlated within areas of visual scene corresponding to surface regions. Texture exhibits some sort of periodicity of the basic pattern of Spongy Tissue in alphonso mango. This leads to use textural property to identify different patterns of Spongy Tissue in alphonso for detection of defects in alphonso mango. Visual assessment of texture made by human is time consuming and inspection made by human does not achieve a high degree of accuracy and preciseness. Automated visual inspection of the textural pattern improves the accuracy and preciseness during detection of defects in alphonso mango. To develop an optimized algorithm using a non contact mechanism which will detect the defective alphonso mangoes happen to be a challenging task. In this paper we have proposed use of subspace analysis technique for extraction of textural features that identifies Spongy Tissue in alphonso mango successfully. This paper presents a methodology that combines the principal component analysis (PCA) to locate the defect in alphonso mango effectively with moment of image segment as statistical feature and Fuzzy C-Means as a data clustering technique used for classification. The proposed algorithm performance is checked on the generated database [1] that is easily available for the researchers working on the said area. Experimental results computed using the proposed algorithm has been validated manually with the cut sections of the alphonso mangoes.**

Keywords – **Alphonso Mangoes, Digital X-Ray Imaging, Non-Destructive Testing, Principal Component Analysis, Spongy Tissue, Textural Features.**

I. Introduction

Horticultural crops play an important role in the economy of the nation, as there are crops suitable for almost all the agro-climatic zones of the country. India has accelerated in total annual production of horticultural crops touching over 149 million tone during 1999-2000. Today, India has emerged as the second largest producer of the fruits (46 million tone) and vegetables (91 million tone), contributing nearly 10% and 14% respectively, of the world production. The horticulture crops cover about 8% of the total area contributing about 20% of the gross agricultural output in the country. India also enjoys the crown of being number one producer of mango and banana in the world. The overall productivity of the fruit and vegetable is 12 and 15.2 tones / hec, respectively. Though the country is the second largest producer of the fruits and vegetables in the world, per capita consumption of vegetables and fruits of our country for over one billion populations is very low. Fresh fruits and vegetables have the lion's share of exports followed by processed one.

Quality assurance in agricultural products is being more crucial with the liberalized international trade system and globalization for capturing and retaining the market. In order to strengthen the competitiveness in terms of export, advanced post-harvest technology adoption is essential. The export of fruits should lead to increase in farming income. Substandard and inadequate post harvest technology and management system is a major bottleneck to the expansion of trade for India. India is the second largest producer of fruits in the world. However, India's fruit production only contributes 1% to the export earnings from agricultural products. Also, India is the largest producer of mango fruit in the world. Amongst mangoes, Alphonso mango is the leading commercial variety produced in Maharashtra State recognized nationally as well as internationally. But in recent years this variety is facing problems due to the pests, diseases and disorders like Spongy Tissue. The peculiarity of this disorder is that the external symptoms of the affected fruits are not apparent from outside either at the ripe stage or at the time of picking. The tissue affected is visible only when the ripe fruit is cut into two halves. The overall loss due this disorder is about 30%, which amounts to a loss of nearly Rs 135 million. Traditionally, chemical and biochemical methods are available to investigate compositional and metabolic differences between the healthy and the damaged tissue. The occurrence and intensity of the disorders depend upon factors like those related to location, climate and cultivars. The symptoms are manifested at the final stage of fruit growth and maturation. Padda et al [2] by using Canonical Discriminant Analysis found that the best tools to assess changes in mango fruit during ripening were firmness, followed by flesh value and total soluble solids content. Subedi et al [3] showed that flesh value well correlated with maturity of fruit. Color of flesh was found to be consistent in various cultivars and although it is a destructive measurement, it is used as a maturity index in several producing regions [4]. Thus, online identification of defects in fruits is highly complex and most challenging because they require real time solution. So, there is an immediate necessity to develop a system which can detect the presence of Spongy Tissue in the mangoes non-destructively prior to export of the mangoes. This paper is an effort towards the direction. In this paper we have developed an algorithm using subspace analysis technique, which can detect the presence of spongy tissue from stored X-ray images. The textural property of the acquired X-ray images of alphonso mangoes shows the presence of the defects or disorders. Principal Component Analysis is a popular technique used to derive a starting set of features.

Thus, we have proposed to investigate the defects in the alphonso mangoes by using statistical method to locate the defect effectively along with the Principal Component Analysis [5, 6] to lower the computational complexity. Moment [7] of image segment is used as statistical feature and Fuzzy C-Means [8-10] is a data clustering technique used for classification. Fuzzy clustering is the partitioning of a collection of data into fuzzy subsets or clusters based on similarities between the data. The validity test of the developed algorithm has been performed manually with the cut sections of the alphonso mangoes available in the database [1].

II. Statistical Feature Extraction using Moments

In order to extract statistical features of the alphonso mango image, the X-ray image is initially divided into four equal parts for computing textural features based on proposed statistical method. Region moment representation interprets a normalized gray-level texture function as probability density function of a 2D variable. Properties of this random variable can be described using statistical characteristic moments [7]. Moments can be used for binary or gray-level region description. A moment of an order $(p+q)$ is dependent on scaling, translation and rotation. Translational invariance can be achieved by using the central moments given by,

$$\mu_{pq} = \sum_{i=-\omega}^{\omega} \sum_{j=-\omega}^{\omega} (i - x_c)^p \, (j - y_c)^q \, f(i,j) \quad (1)$$

where,

x_c, y_c are the coordinates of region's centre of gravity (centroid), defined as,

$$x_c = \frac{m_{10}}{m_{00}}$$

$$y_c = \frac{m_{01}}{m_{00}}$$

Second order moments focus on the inertia matrix can be defined as,

$$C = \frac{1}{\mu_{00}} \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix} \mu_{pq} \quad (2)$$

Shape attributes like elongation (ϵ) and compactness (k) have been calculated for each region. Both attributes are translation, scaling and rotation invariant.

$$\epsilon = \frac{\lambda_2}{\lambda_1} \quad (3)$$

$$k = \frac{\mu_{00}}{4\pi\sqrt{\lambda_1 \lambda_2}} \quad (4)$$

where,

λ_1 and λ_2 are eigen values.

The image is divided till two shape attributes elongation and compactness is same, at least for 75% of the regions.

III. Principal Component Analysis

Principal Component Analysis is a way of identifying patterns in data, and it expresses the data to highlight their similarities and differences. PCA is a variable reduction procedure. In reality, the number of components extracted in a principal component analysis is equal to the number of observed variables to be analyzed. In most analyses, however, only the first few components account for meaningful amounts of variance, so in subsequent analysis only these components are retained, interpreted, and used. In our work only first two components would account for a meaningful amount of variance. Therefore only first two components have been retained for interpretation.

IV. Feature Extraction

The feature vectors computed for defect detection in alphonso mango are the features derived from the Eigen values of the covariance matrix of each region. The first component extracted in PCA accounts for a maximal amount of total variance in the observed variables. The second component accounts for a maximal amount of variance in the data set and it will be uncorrelated with the first component. In the proposed method PC_1 is first principal component and PC_2 is the second principal component. Finally to receive feature vector, similarity indexes s_x^p and s_y^p are calculated for each p^{th} region of a textural pattern, which are defined as,

$$S^p{}_x = \sum_{j=1}^{L} \sum_{i=1}^{W} ((PC_1{}^{pi} - PC_1{}^{ji})^2 + (PC_2{}^{pi} - PC_1{}^{ji})^2)^{1/2} \quad (5)$$

$$S^p{}_y = \sum_{j=1}^{L} \sum_{k=1}^{C} ((PC_1{}^{pk} - PC_1{}^{jk})^2 + (PC_2{}^{pk} - PC_1{}^{jk})^2)^{1/2} \quad (6)$$

where,

$p = 1, \ldots L, p \neq j,$

W is the number of rows in the analyzed rectangular regions,

L is the number of regions of the X-ray image of alphonso mango,

PC_r^{ji} is the r^{th} PC calculated for i^{th} row and j^{th} rectangular region,

C is the number of columns in the analyzed rectangular regions,

PC_r^{jk} is the r^{th} PC calculated for k^{th} column and j^{th} rectangular region.

Similarity indexes s_x^p and s_y^p have been used as data for clustering process. s_x^p and s_y^p are smaller for similar rectangular region and higher when difference between them is larger. The similarity factor is calculated by adding the similarity indexes s_x^p and s_y^p as,

$$\text{Similarity factor} = s_x^p + s_y^p \quad (7)$$

Intensity information of each region is extracted from plotting histogram of each region. Histogram of region 1 and region 2 are shown in Figure 1. Histogram Comparison of each region is done by finding the difference of each histogram with all the regions as,

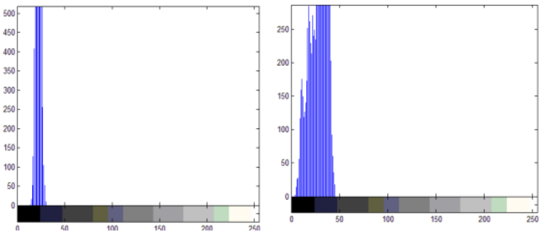
$$\text{Difference histogram} = \sum_{i=1}^{k} H_i - H_j \quad (8)$$

where,

k is number of regions,

H_i and H_j is i^{th} and j^{th} region histogram.

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 3, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209



(a) (b)

Fig.1 (a) Histogram of region 1 and (b) Histogram of region 2.

The sum of histogram differences for each region is calculated. The region having maximum histogram difference is considered as defective region.

V. Fuzzy C-Means Clustering

In the proposed algorithm of defect detection in alphonso mangoes, fuzzy C-means data clustering technique has been used for classification. Fuzzy clustering is the partitioning of a collection of data into fuzzy subsets or clusters based on similarities between the data. Fuzzy C-means [8-10] is a data clustering technique wherein each data point belongs to a cluster to some degree which is specified by a membership grade. A method is provided here that shows how to group data points that populate some multidimensional space into a specific number of different clusters. The fuzzy C-means (FCM) algorithm is one of the most widely used methods in fuzzy clustering. The input data for the Fuzzy C-means clustering are Difference histogram and Similarity factor.

The proposed algorithm of clustering for defect detection in alphonso mangoes comprises of the following steps:

Step 1: Initialize the fuzzy partition matrix $U = [u_{ij}]$ matrix, $U^{(0)}$

Step 2: At k-step: calculate the centers vectors

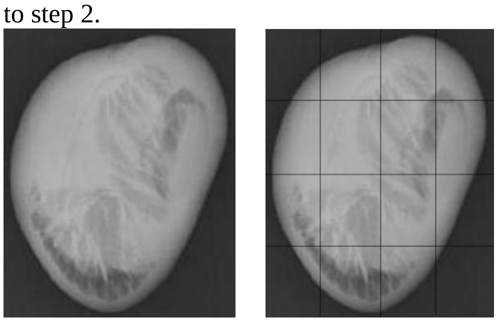
$$C^{(k)} = [c_j] \text{ with } U^k$$

Step 3: $$C_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

Step 4: Update $U^{(k)}, U^{(k+1)}$ using,

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i-c_j\|}{\|x_i-c_k\|}\right)^{\frac{2}{m-1}}}$$

Step 5: If $\|U^{(k+1} - U^{(k)}\| < \varepsilon$ then STOP; otherwise return to step 2.



(a) (b)

Fig.2. Execution procedure of the algorithm (a) Input image (b) segmented input image.

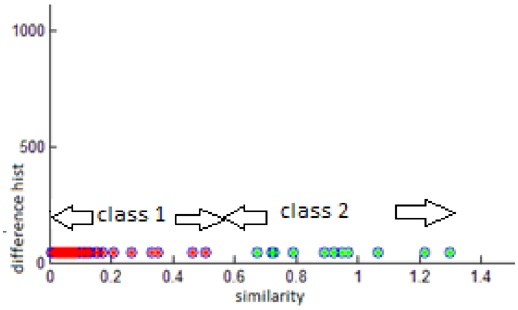
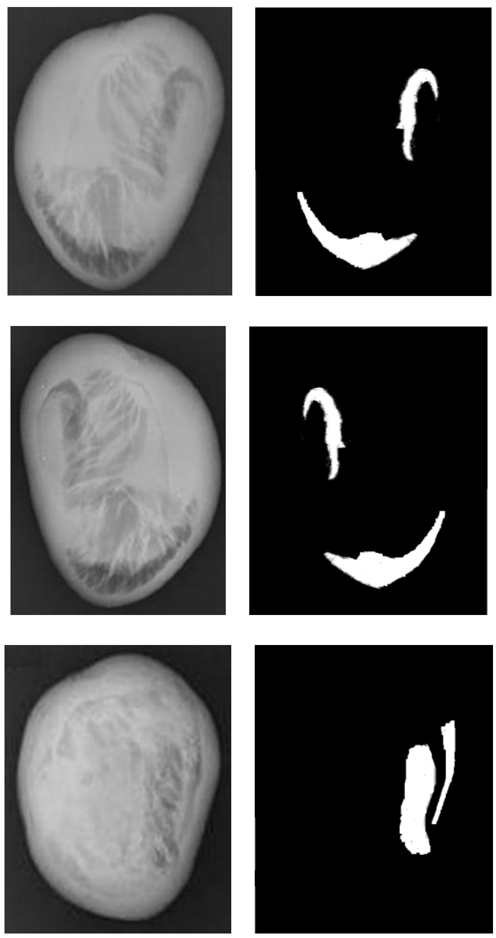


Fig.3. Fuzzy C mean Clusters.

In order to explain the execution procedure of the proposed algorithm on the input image as shown in Figure 2 (a). The input image is initially segmented in to various parts as shown in Figure 2 (b). After segmenting the X-ray image of alphonso mango under consideration, fuzzy C-means clustering algorithm has been used. The resultant data contains two clusters, defective and healthy portions, as shown in Figure 3. First cluster represents healthy tissue and second represents Spongy Tissue defects in the sample alphonso mango under consideration. Cluster containing lesser number of elements is recognized as a cluster that represents the defect in the alphonso mango.

VI. Results and Discussions



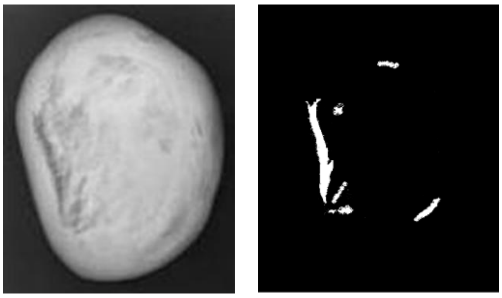


Fig.4. X-ray images of defective alphonso from database and detected defects

The proposed algorithm for detection of defects in the alphonso mango is a non-destructive technique that combines PCA to locate the defect effectively with moment as statistical feature and Fuzzy C-Means as a data clustering technique. Performance of the algorithm is carried out on the generated database of healthy and defective alphonso mangoes [1] that is easily available for the researchers working on the said area. Experimental results computed using the algorithm has been validated manually with the cut sections (available in the database) of few samples of defective and healthy alphonso mango images.

Experimental results for defect detection in alphonso mangoes using the proposed algorithm are shown in Figure 4 and Figure 5. Experimental results show that the proposed method can detect defects correctly. To find out the efficiency of the proposed method, classification rate C_r have been calculated as,

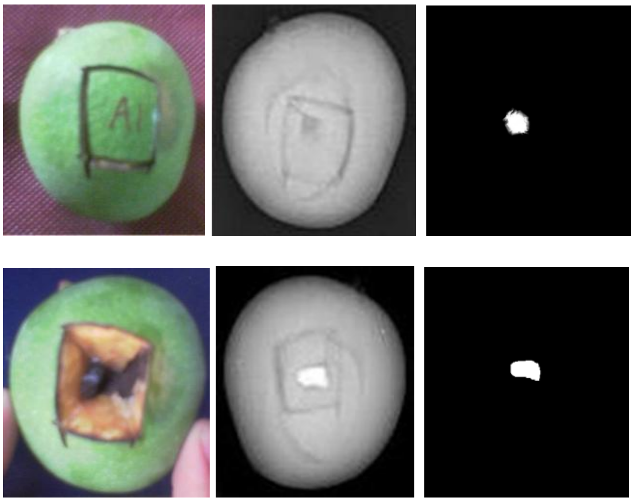
$$C_r = \frac{(N_c + N_d)}{N_t} * 100\ \% \qquad (9)$$

where,

N_c is the number of segmented regions containing healthy tissue,

N_d is the number of segmented regions having defects,

N_t is the total number of segmented regions.



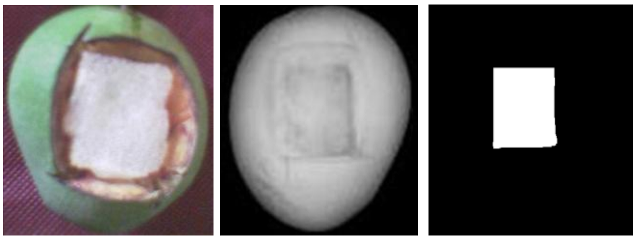


Fig.5. Manually created defects in alphonso mangoes, their X-ray images and detected defects.

The efficiency graph is plotted against classification rate versus number of clusters as shown in Figure 6.

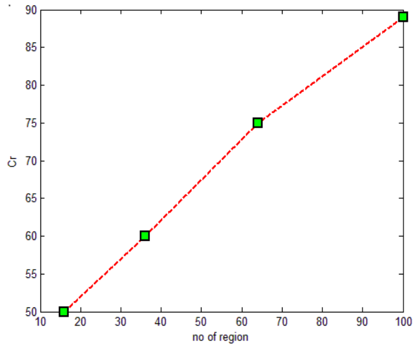


Fig.6. Efficiency graph.

VII. Conclusion

The proposed method for defect detection in alphonso mango combines a concept of PCA and histogram. The number of regions depends on size of the defect. Classification rate increases as the number of segmented region increases and after 89% of classification rate the efficiency of the proposed method remains constant. This method does not require training collection. The principal components preserved the most energy of the original data, and provided good performance in recognizing obviously separated classes. The first principal component of PCA presents the highest discriminant power among other principal components.

The step and mask sizes affects performance of the algorithm. The step size should be less than or equal to the size of mask in order to cover all the pixels from the original image. Step size and window size of the mask can be tuned in such a way that size of mask should be large enough to cover the defect at least more than two times in horizontal as well as vertical scanning for selected step size and mask size should be small enough to reduce the width of patch at abrupt changes in orientations in order to detect exact region boundaries. The number of clusters used for the classification plays a vital role in final segmentation results.

Acknowledgment

The authors are thankful to the “Board of College and University Development (BCUD), Pune University, Pune” for supporting this non-commercial research entitled “Detection and Analysis of Defects in Agro Products,” Online Proposal Number: 13ENG001578 under the

"Research Proposals under University Research Grant Scheme for the year 2013 – 2015." The authors are also thankful to the anonymous reviewers for their valuable suggestions and comments in improving the quality of the paper.

REFERENCES

[1] Sandeep M. Musale and Pradeep M. Patil, "Database development of defective and healthy alphonso mangoes," in proceedings of International Conference on Food, Agriculture and Biology (FAB-2014), to be held at Kuala Lumpur (Malaysia), June 11-12, 2014.
[2] Padda S.M., do Amarante C. V. T., Garcia R. M., Slaughter D. C., and Mitcham E. M, "Methods to analyze physico-chemical changes during mango ripening: A multivariate approach," Postharvest Biology and Technology, vol 62, pp 267-274, 2011.
[3] Subedi P.P., Walsh K.B. and Owens G, "Prediction of mango eating quality at harvest using short-wave near infrared spectrometry," Postharvest Biology and Technology, vol 43, pp 326-334, 2007.
[4] Sivakumar D, Jiang Y and Yahia E.M, "Maintaining mango fruit quality during the export chain," Food Research International, vol 44, pp 1254-1263, 2011.
[5] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp 71-86, 1991.
[6] B. Josso, D. Burton and M. Lalor, "Texture orientation and anisotropy calculation by Fourier transform and Principal Component Analysis," International Journal of Mechanical Systems and Signal Processing, Vol. 19, pp 1152–1161, 2005.
[7] J. Flusser, "Moment Invariants in Image Analysis," Proceedings of World Academy of Science, Engineering and Technology, Vol. 11, pp 196-201, 2006.
[8] D. Kim, K. Lee, and D. Lee, "A novel initialization scheme for the fuzzy c-means algorithm for colour clustering," Pattern Recognition Letters, Vol. 25, No. 2, pp 227-237, 2004.
[9] W. Cai, S. Chen, D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," Pattern Recognition, Vol. 40, No.3, pp 825-838, 2007.
[10] J. Fan, M. Han, J. Wang, "Single point iterative weighted fuzzy C-means clustering algorithm for remote sensing image segmentation," Pattern Recognition, Vol. 42, No. 11, pp 2527-2540, 2009.

AUTHOR'S PROFILE

Sandeep S. Musale

(M'12) received his B. E. (Electronics& Telecommunication) degree in 2004 from University of Pune, Pune, (India) and M. E. (Electronics) degree in 2008 from University of Pune, Pune, (India). From 2004 to 2014 he worked as Lecturer and Assistant Professor in department of Electronics Engineering at PVG's College of Engineering & Technology, Pune, (India). He is member of various professional bodies like ISTE and IEEE. His research area is Defects detection in Agro Products using non-contact techniques.

Dr. Pradeep Mitharam Patil

(M'04) received his B. E. (Electronics) degree in 1988 from Amravati University, Amravati, (India) and M. E. (Electronics) degree in 1992 from Marathwada University, Aurangabad, (India). He received his Ph.D. degree in Electronics and Computer Engineering in 2004 at Swami Ramanand Teerth Marathwada University, (India). From 1988 to 2011 he worked as Lecturer and Assistant Professor and Professor in department of Electronics Engineering at various engineering colleges in Pune University, (India). Presently he is working as Dean, RMD Sinhgad School of Engineering and Director of RMD Sinhgad Technical Institutes Campus, Warje, Pune, (India). He is member of various professional bodies like IE, ISTE, IEEE and Fellow of IETE. He has been recognized as a PhD guide by various Universities in the state of Maharashtra (India). His research areas include pattern recognition, neural networks, fuzzy neural networks, power electronics and agro-electronics. His work has been published in various international and national journals and conferences including IEEE, Science Direct and Elsevier.