A NON-DESTRUCTIVE TECHNIQUE BASED ON TEXTURE ANALYSIS FOR IDENTIFICATION OF DEFECTIVE ALPHONSO MANGOES USING SUBSPACE ANALYSIS TECHNIQUES

MUSALE S.S.¹* AND PATIL P.M.²

¹Sinhgad College of Engineering, Pune- 411 004, MS, India.
²RMD Sinhgad Technical Institutes Campus, Pune- 411 058, MS, India.
*Corresponding Author: Email- sandeepmusale@yahoo.co.in

Received: May 18, 2014; Accepted: June 16, 2014

Abstract- In many circumstances, texture is the only information that can be used in natural image analysis because it is an important property of the surface that characterizes its nature. Texture is defined as a spatial arrangement of local (gray level) intensity attributes which are correlated within areas of visual scene corresponding to surface regions. An image region has a constant texture if sets of its local properties in that region are constant. Thus texture analysis has received considerable attention in the field of image analysis and pattern recognition. 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. In the literature, the researchers worldwide have been working in various texture analysis algorithms for different applications like detection, recognition, classification, segmentation, clustering etc. Many algorithms suffer from low sensitive detection, difficult back ground adaption and high memory requirement. Problems and limitations associated with the available techniques have been reported by many studies. Each has some drawback under all lighting conditions and no one has used a robust, reliable algorithm for detection of spongy tissue in alphonso mango under real life test environment. To develop an optimized algorithm using a non contact mechanism which will detect the defective alphonso mangoes happen to be a challenging task. The objective of the proposed research work is to obtain computationally cost effective and noncontact solution that achieve better recognition rate under various conditions in consultation with the agriculture scientist. In this paper we have proposed use of subspace analysis techniques for extraction of textural features that identifies Spongy Tissue in alphonso mango successfully. Performance of the proposed algorithm is carried out on the alphonso database [1]. This paper presents a methodology that combines the principal component analysis (PCA) and fisher linear discriminant analysis (FLD) applied for defect detection in alphonso mangoes. Performance of the proposed algorithm has been compared with that of PCA and FLD applied individually for both healthy and defective alphonso mangoes. Experimental results computed using the proposed integrated algorithm, individual PCA and individual FLD have been validated manually with the cut sections of the alphonso mangoes available in the database. It is observed that the proposed method shows significant improvement for both defective as well as healthy alphonso mangoes over PCA and FLD applied individually. The feature vectors extracted by using integrated PCA-FLD method have 99% of the highest discriminant power.

Keywords- Alphonso mangoes, Spongy Tissue, digital X-Ray imaging, Non-Destructive testing, subspace analysis, principal component analysis, fisher linear discriminant analysis, textural features

Citation: Musale S.S. and Patil P.M. (2014) A Non-destructive Technique Based on Texture Analysis for Identification of Defective Alphonso Mangoes Using Subspace Analysis Techniques. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 6, Issue 1, pp.-388-392.

Copyright: Copyright©2014 Musale S.S. and Patil P.M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

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. Padma, 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 techniques, which can detect the presence of spongy tissue from stored X-ray images. The textural property of the acquired X-ray images shows the presence of the defects or disorders. Principal Component Analysis is a popular technique used to derive a starting set of features. Application of PCA for defect detection that uses textural properties, does not consider the classification aspect since it is based on the optimal representation criterion in terms of mean-square error. Talukder, et al [5] pointed out that PCA alone (as an unsupervised method) is not necessarily good at drawing distinctions between patterns. To improve the PCA standalone classification performance, one needs to combine further this optimal representation criterion with some discrimination criterion. One widely used discrimination criterion is the Fisher linear discriminant (FLD), or linear discriminant analysis (LDA) [6], which defines a projection that makes the within-class scatter small and the between-class scatter large. As a result, FLD derives compact and well-separated clusters. As the original image space is high dimensional, most of these methods apply PCA first for dimensionality reduction. Subsequent FLD transformation is then used to build the most discriminating features (MDF) space for classification. For good generalization FLD requires large training sample size. One possible remedy for this drawback is to improve FLD’s generalization performance by balancing the need for adequate signal representation and subsequent classification performance using sensitivity analysis on the spectral range of the within-class eigen values. The second well-known approach coming under the name of fisher face is insensitive to large variation in the conditions already enumerated above. It uses both PCA and FLD. It is worth stressing that by maximizing the ratio of between-scatter matrix and within-scatter matrix, FLD produces well separated classes even in the subspace having low dimensions under large variation in lighting conditions.

In this paper a methodology that combines the principal component analysis (PCA) and fisher linear discriminant analysis (FLD) to aid in the criteria development for selecting spectral bands for multi-spectral imaging applications has been proposed. This method and the individual PCA and FLD methods have been applied for defect detection in alphonso mangoes. The defect detection performance of the proposed method has been compared with PCA and FLD. Performance of the proposed algorithm is verified with the alphonso mango database [1] which is easily available for researchers in this area after sending requisition to the authors through mail. The database consists of X-ray images of healthy alphonso mangoes, artificially created defects in the mangoes as well as naturally affected spongy tissue mangoes collected using digital X-ray machine. In the captured images, the spongy tissue affected portion or artificially defect created portion of the mangoes shows dark gray levels, whereas unaffected part or healthy portion shows a uniform light grey levels. The database also consists of the digital photographs of all the alphonso mangoes along with the cut sections of the alphonso mangoes may be required for manual validation of the automated results obtained with the help of proposed algorithms. Defects in alphonso mangoes have been detected successfully using the proposed method.

**Feature Extraction with Proposed PCA-FLD Method**

PCA is a transform that uses orthogonal axes for dimensionality reduction by performing an eigen decomposition of the spectral covariance matrix of the data. Let \( C_t \) represent the covariance matrix of \( N \times N \) dimension and computed by,

\[
C_t = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T
\]

where,

- \( C_t \) represents covariance matrix,
- \( x_i \) : the \( i \)-th random sample with dimension \( N \),
- \( \mu \) : represents mean vector of the set of pixels,
- \( N \) : the sample size.

In order to maximize the energy of the projected samples, the eigen-decomposition transform is used and computed as,

\[
E_{pca} = \phi_m^T C_t \phi_m
\]

where,

- \( E_{pca} \) : represents the energy of the samples projected,
- \( \phi_m \) : is one of the transform vectors used to project the data samples, \( m = 1, ..., N \),
- \( C_t \) : the covariance matrix,
- \( \phi_m^T \) : the transpose of the transform vectors used to project the data samples.
A set of eigenvectors is orthogonal and each eigenvector corresponds to a unique combination of bands. The residual mean square error between transformed data and original data is minimized by performing eigen-decomposition of the covariance matrix. The first eigenvector corresponding to the largest eigen value and provides largest variance of the data. Similarly, the first two eigenvectors explain more variance of the data than any other pair of original bands. Magnitude of eigen value indicates the energy residing in the data along the direction parallel to the corresponding eigenvector. The larger the eigen value, higher is the energy it represents. Therefore, to reduce the original N-dimensional data to a lesser dimension M, one can project the original data to the M eigenvectors corresponding to the largest of the M eigen values. Considering the first few significant compositions, this transform results in a lower-dimensional multivariate feature vector. This feature vector still preserves most of the energy in the original higher-dimensional system.

FLD is an effective class-specific method which projects the scatter of data to make them more reliable for classification. Two matrices are introduced in the FLD namely, the between-class scatter matrix, denoted as $S_b$ and the within-class scatter matrix $S_w$. The between-class scatter matrix $S_b$ is defined as [7],

$$S_b = \sum_{i=1}^{C} | x_i | (\mu_i - \mu)(\mu_i - \mu)^T$$

(3)

The within-class scatter matrix is defined as [7],

$$S_w = \sum_{i=1}^{C} \sum_{x_j \in x_i} (x_k - \mu)(\mu_k - \mu)^T$$

(4)

where,

- $\mu$: represents the mean vector of $i^{th}$ class,
- $\mu$: the mean of total samples,
- $C$: the number of classes,
- $x_i$: represents the $i^{th}$ class and $| x_i |$: represents number of samples in $i^{th}$ class.

The optimal data projection, denoted as $E_{bld}$, can be obtained by selecting the transform-vector $\Phi_b$, that maximizes the ratio of the projected between-class samples to the projected within-class samples and is defined as,

$$E_{bld} = \sum_{m=1}^{M} \phi_m S_b \phi_m$$

(5)

PCA does not guarantee the feature class separation ability of the selected band, whereas, the FLD method is effective in class segmentation, sensitive to noise and may not cover enough energy from the original data.

To design a set of transformation vector-basis providing supervised classification information in a better way that simultaneously preserves enough energy from the original data cube, a method is proposed that combines (1) to (4) in order to construct an evaluation equation called integrated PCA-FLD method. A weight factor K introduced to adjust the degree of classification that preserves the desired energy. The constructed evaluation equation is given as,

$$E_{v1} = \sum_{m=1}^{M} \frac{(KC_i + (1-k)S_b)|\phi_m|^2}{\phi_m^T(KI_i + (1-k)S_w)\phi_m}$$

(6)

where,

- $E_{v1}$: represents the constructed evaluation,
- $K$: [0, 1],
- $C_i$: the covariance matrix,
- $S_b$: the between-class scatter matrix,
- $S_w$: the within-class scatter matrix
- $I_i$: the identity matrix.

The eigen-decomposition becomes inaccurate, if the within-class scatter matrix $S_w$ in [Eq-5] becomes very small. [Eq-6] eliminates this problem. If $S_w$ becomes very small, the effects of $S_w$ can be ignored by adjusting the weight factor K towards 1, which means the principal components are more heavily weighted. The ratio between $S_b$ and $S_w$ dominates, if the selected value of K is small meaning more differential information between classes is taken into account. The integrated method enhances the advantages of PCA and FLD and reduces the disadvantages of both PCA and FLD simultaneously. In fact, FLD and PCA methods represent the extreme situations of [Eq-6], as K=0, it becomes FLD [Eq-5] and when K=1, the evaluation equation is equivalent to PCA method [Eq-2].

To find the transform that works equally well on representation and discrimination, one can find the set of $\Phi_b$ that maximizes [Eq-6] with K<0.5. The solution of $\Phi_b$ called the generalized eigen vector, can be obtained by equating the derivative of [Eq-6] with respect to $\Phi_b$ to zero as in [Eq-7].

$$\left\{ KC_i + (1-K)S_b \right\} \phi_m = \lambda_n \phi_m$$

(7)

where,

- $\lambda_n$: represents eigen values,
- $\Phi_b$: the corresponding generalized eigen vector.

For hyper spectral dimensionality reduction, if the original feature space is having N dimensions, then the reduced M (M<N) dimensional features can be obtained by selecting the value of M generalized eigenvectors corresponding to the M largest eigen values obtained from [Eq-7]. The value of K varies between 0 to 1 that shifts the weight between PCA and FLD. The projected samples become the best in representation and classification at the proportion represented by the value of K. Based on application in hand, an optimal proportion between PCA and FLD can be selected by searching the optimal value of K that achieves best classification rate of samples. The transformation matrix $T (N \times M)$ is formed by selecting M generalized eigenvectors as the column vectors for the matrix. The linear transform is defined as in [Eq-8].

$$Y = TX$$

(8)

where,

- $X$: represents the original feature space with given N x J dimension,
- $N$: the features, $J$: the sample size,
- $Y$: the transformed feature space with M x J dimension.

The transformed feature space consists of the new features (or bands) that can be obtained from the linear combination of all original features/bands for all pixels. Since the transformation matrix consists of M eigen vectors corresponding to the M largest eigen values, transformed feature space can provide a better classification and representation ability than any subset of M bands of the original data set. For the experimentation purpose, we have selected the value of M equal to 1. In order to extract the features, the eigen vector corresponding to the largest eigen value of [Eq-7] is
used. In this case, the transformed feature space represents a single feature vector i.e. the first principal feature vector, generated by linear combination of original \( N \) features by using the integrated PCA-FLD method. Similarly, by applying PCA or FLD method individually, one can obtain the first principal feature vector of PCA or FLD, respectively. The physical meaning of the first principal feature vector in the integrated PCA-FLD method provides the best discriminative feature compared with any individual feature in the original \( N \) feature space. A classifier is then designed to determine the category of input sample images. The mean and standard deviation of the feature vectors are calculated for spectral feature vectors that belong to one image sample. The ratio of the standard deviation to the mean is used as numerical input variable to a \( k \)-Nearest Neighbor classifier algorithm. The \( kNN \) algorithm is used to determine the \( k \) samples in the training set that are close to the \( X^{th} \) unknown sample in the testing set. As the patterns with similar attributes should be assigned to the same class, samples with closer ratios are classified as the same category. Hence, if a majority of the \( k \) Nearest Neighbors belongs to the class of ‘healthy alphonso mango’, then the sample is classified as a healthy alphonso mango, otherwise it is classified as the defective one. For experimentation purpose, we have used \( k = 15 \) in the classifier.

**Results and Discussions**

In the proposed method of defect detection in alphonso mangoes, the value of \( K \) was used to adjust the balance between the representative and classification abilities of the projection matrix, and therefore increased the flexibility of the method for Spongy Tissue detection in alphonso mangoes. For different image samples, the optimal values of \( K \) differs. An effective method for \( K \) value selection was needed.

In this work, we have searched the values of \( K \) from 0 to 1 with a step size of 0.05. When the \( K \) value was changed, the recognition rates varied. In general, by choosing the optimal \( K \) values, both the representation and classification effects were properly weighted for the testing samples. The small set sample tests can be considered a validating process before the training and the testing process. Those image samples were called validating image samples, which were used to determine the variable parameters in the method before final application to the training and testing processes.

| Table 1: Samples used for various experiments |
|-----------------|-----------------|-----------------|-----------------|
| Experiment | Sample Type | Total number of samples | No of samples for Training Set | No of samples for Testing Set |
| 1 | Healthy and defective alphonso mangoes | 200 | 50 | 150 |
| 2 | Healthy and defective alphonso mangoes | 200 | 100 | 100 |

Performance of the proposed algorithm is carried out on the alphonso database [1]. The database consists of 1000 X-ray images of both healthy and defective alphonso mangoes. Out of these 1000 images we have randomly selected 100 images of healthy and 100 images of Spongy Tissue affected alphonso mangoes (with moderate and severe degrees of damage) for our experimentation purpose. The samples of healthy and Spongy Tissue affected alphonso mangoes used for these experiments are listed in [Table-1]. In the first experiment, out of 100 X-ray images of healthy alphonso mangoes we have selected 25 samples and out of 100 X-ray images of Spongy Tissue affected alphonso mangoes we have selected 25 samples (total 50 samples) were used for training and then tested together with the remaining 150 samples. The total 200 sample testing results with \( K \) set to 0.60 have been shown in [Fig-1]. The results indicate that the PCA method and the integrated PCA-FLD method achieved a similar better performance compared with the FLD method. The sample classes used in the first experiment were well separated hence the solution was dominated by the PCA method at classifying data since they are well distinguishable. On the other hand, since FLD maximizes the \( E_{fl} \) ratio shown in [Eq-5], which is small in this case, it becomes unreliable and noise sensitive, and leading to the poor performance in this situation.

In order to test the reliability of these algorithms, number of samples during training was increased. In this experiment, 100 samples were used randomly for training, and the remaining samples were used for testing; therefore, the recognition rates for both good and faulty textile fabric images were high. The recognition results with \( K \) value (set empirically) set to 0.20 have been illustrated in [Fig-2]. Since these two classes had very similar spectral characteristics, the differentiating information played a more important role in classification. Therefore, the FLD solution showed better performance than PCA. By using the FLD method, the defect detection rate achieved 91%, and the healthy sample detection was 98%, which were, respectively, 6% and 1% higher than those achieved with the PCA method. It is observed that in both the experiments the results of the proposed integrated PCA-FLD method achieved a similar better performance.
Conclusions
The number of regions in PCA and FLD depends on the size of the defect in alphonso mango. Classification rate increases as the number of segmented region increases. Also, the efficiency of the FLD method remains constant. This method does not require training collection.

The integrated PCA-FLD method for feature extraction when applied to defect detection and classification of alphonso mangoes achieves better recognition performance. The integration process was neither a simple combination of the two methods nor using them sequentially during the entire procedure. The method was derived based on a constructed evaluation equation that combined the representation and classification effects together. The principal components preserved the most energy of the original data, and provided good performance in recognizing obviously separated classes. The discriminant analysis contributed to classifying similar patterned classes. By properly adjusting the weight factor, the integrated PCA-FLD method was more flexible in processing with different sample patterns and the result became robust to noise. The results show that the integrated PCA-FLD method outperformed the PCA or FLD method. The reason is that both the representation information and class specific information are included in the integrated PCA-FLD feature. FLD provides the highest discriminant power of any feature vector. The feature vectors extracted by using integrated PCA-FLD method have 99% of the highest discriminant power. The first principal component of PCA presents the highest discriminant power among other principal components. Compared with FLD and integrated PCA-FLD, the first principal component has the lowest discriminant power.

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 to elevate quality of the paper.

Conflicts of Interest: None declared.

References
[1] Musale S.S. & Patil P.M. (2014) Database development of defective and healthy alphonso mangoes, International Conference on Food, Agriculture and Biology, Kuala Lumpur, Malaysia, 60-65.
[2] Padda S.M., do Amarante C.V.T., Garcia R.M., Slaughter D.C., & Mitcham E.M. (2011) Postharvest Biology and Technology, 62, 267-274.
[3] Subedi P.P., Walsh K.B. & Owens G. (2007) Postharvest Biology and Technology, 43, 326-334.
[4] Sivakumar D., Jiang Y. & Yahia E.M. (2011) Food Research International, 44, 1254-1263.
[5] Talukder A. & Casasent D. (1998) Journal of Optical Engineering, 37(3), 904-913.
[6] Fisher R. (1936) Ann. Eugenics, 7, 179-188.
[7] Belhumeur P., Hespanha J. & Kriejman D. (1997) IEEE Trans-