Fruit Disease Categorization based on Color, Texture and Shape Features

Ranjit K. N.
Department of Computer Science
& Engineering
HKBK Engineering College,
Bangaluru, Karnataka, India

Naveen C.
Department of Computer Science
& Engineering
HKBK Engineering College,
Bangaluru, Karnataka, India

Chethan H. K.
Department of Computer Science
& Engineering
Maharaja Institute of Technology,
Mysuru, Karnataka, India

ABSTRACT
Nowadays digitization and automation of machine in agriculture field plays prominent role. In this paper, we have proposed method to classify fruit as diseased and non-diseased. Firstly, we used K means clustering method for segmentation of diseased regions. Later, we used to extract shape, color and texture features on segmented diseased regions. We have collected fruit diseased images from internet to create dataset and totally we have collect 2500 images from 10 fruit classes. We have conducted extensive experimentation using Artificial Neural Network and results shows that proposed method gives better performance compared to SVM and KNN.

Keywords
Fruit Disease, Color, Shape, Texture, Categorization.

1. INTRODUCTION
Fruit nutrition is very rich and there are many effects [1]. For example, the fruit of the fiber for the pectin material beneficial defecation, and fiber ingredients can also promote the body's metabolic function. On the other hand, the fruit also contains natural pigments [2], which can effectively prevent cancer. At present, the post-processing of our fruit is by manual sorting, which inevitably arise from various problems, such as the workers' long-term monotonous repetition of work to make them fatigue, resulting in reduced accuracy of the test, the exception of workers on the classification of the standard accuracy of the test. Therefore, the automation of fruit classification [3] is an urgent need to improve worker productivity, improve classification accuracy. If the fruit is placed at room temperature for more than the specified date, the color of the fruit will change, and then the fruit will be infected. It is harmful to human body if they eat this infected fruit. Therefore, it is important to design a system for automatic detection of fruit diseases [4].

In representing the conception for human brain, images are the most basic method in physical classification of foodstuff and agricultural industry. Factors affecting fruits and vegetables can be quantified visually which is laborious, expensive and is easily effected by physical factors, including inconsistent evaluation and subjective results. The market prices are determined by such inspections and, also, the ''best-if-used-before date''. The trained human investigators have done the quality inspection by feeling and seeing. This method is significantly inconsistent, fickle and decisions are seldom same among investigators. In this type of environment, the analysis of fruits and vegetables for several aspect criterions is a continual task; machine vision systems are best befitted for conventional analysis and quality assurance. In agriculture, computer vision system and image processing is readily growing research area which is a significant analyzing technique for pre to post harvesting of crops. Fig. 1shows the number of research papers published year wise (to the best of our knowledge). From this graph, it can be easily seen the trending this research field. The data and information in agriculture mainly originates from photographic images but mathematically to estimate or process photographic data, it is challenging. Therefore, digital image pro-cessing technology helps to process images and attempt an extension for their analysis. Image processing has various applications in the field of agricultural like identification of land [5], evaluation of nitrogen recognition plant [6], recognition of pest infected areas [7], automatic classification and detection of plant disease from shape, texture and color [8]. As, information science is rapidly growing, computer vision based pattern recognition and image processing are matured technique for safety and quality analysis of several agricultural applications. Computer vision technology corresponds the effect of the human vision in inspecting quality of fruits and vegetables by electronically perceiving an image, interpret and recognize the characters and an information is provided for the quality grading and sorting machine. Various research papers have been published [9]-[11], some of them focus on particular fruit in quality analysis while others are centered on particular techniques. A detailed summarization of quality analysis of fruits and vegetables is not available. Thus, the objective of this review paper is to give comparable survey of computer vision and image processing techniques in the food industry and also to review various segmentation, image features and image descriptors in the literature and quality analysis of fruits and vegetables on the basis of color, shape, size and texture and the type of disease present. Also, the principal components, basic theories and corresponding analysis and processing methods are reported.

2. RELATED WORK
Wu [12] adopted a four-step to classify the type of fruits. Firstly, this paper used split-and-merge method to remove the background of the image. Secondly, extracting 79 features from a fruit image. Thirdly, in this paper, the authors used PCA to reduce the dimension to 14. Finally, three kinds of multi-class SVMs and kernels were chosen to classify. The results showed that the "Max-Wins-Voting SVM + Gaussian Radial Basis kernel" performed best. Wang [13] employed a novel feature extraction method-wavelet entropy (WE). In this paper, 3-level decomposition was chosen to extract 10 features from each color channel for each fruit image, and a total of 30 features were extracted. The time spent on the training of 30 features is still relatively large, so it is necessary to reduce the dimension. After that, authors chose FSCABC-FNN and BBO-FNN as classifier. The accuracy of those two methods is the same, but the latter calculation time less than the former 5 seconds on average. Ji [14] was based on the

improvement of the paper [12], and the biggest difference between the two papers is the choice of the classifier. In this paper, authors proposed "FSCABC + FNN", afterwards, compared with four methods, it is found that the method has the highest accuracy, 89.1%. Wu [15] proposed "PCA+BBO-FNN" method, and it is based on the paper [13] in order to improve the classification performance. By comparison, it found that the two methods were same in classification accuracy, but the average compute time of BBO-FNN was significantly less than FSCABC-FNN. Wang [16] demonstrated that HPA's performance was better than other hybridizations, and SLFN has achieved good results in terms of classification. Therefore, Lu [17] proposed a novel "HPA + SLFN" method. At the same time, the method was compared with other five existing methods. The Kuang [18] used the proposed weighted score-level fusion to classify the type of fruits according to different single feature and complementary features. The result showed that "color + shape + edgeLBP + HOG + LBP" achieved the higher accuracy. Afterwards, this paper compared with other four state-of-the-art methods, and got a conclusion that WSL-LW (our method) achieved 90.7% which was higher than other methods.

3. PROPOSED METHOD

Place Tables/Figures/Images in text as close to the In this section we proposed method to classify diseased fruits and non-diseased fruits using texture, color and shape features more detail about features are presented in feature extraction section. Before feature extraction we proposed to use K-means clustering for segmentation.

3.1 Defect Segmentation

In this paper, K-means clustering technique is used for the defect segmentation similar to Dubey and Jalal (2014a). Images are partitioned into four clusters in which one or more cluster contains only infected region of the fruit. The K-means clustering algorithms classify the objects (pixels in our problem) into K number of classes based on a set of features. The categorization is carried out by minimizing the sum of squares of distances between the data objects and the corresponding cluster. Algorithm for the K-Means image segmentation –

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

In this experiment, squared Euclidean distance is used for the K-means clustering. We use L*a*b* color space because the color information in the L*a*b* color space is stored in only two channels (i.e. a* and b* components), and it causes reduced processing time for the defect segmentation. In this experiment input images are partitioned into four segments. From the empirical observations it is found that using 3 or 4 cluster yields good segmentation results. Figure 1 depicts some defect segmentation results using the Kmean clustering technique.

Figure 1: Depicts some defect segmentation results using the K-Means clustering technique.

3.2 Feature Extraction

The aim of this stage is to extract the attributes or characteristics that describe an image. The categorization accuracy are mainly depends upon feature extraction stage. So the presented work investigates using two methods for extracting images features which are shape and color features and Scale Invariant Feature Transform (SIFT). Since, color consider as an significant feature for image representation due to the color is invariance with respect to image translation, scaling, and rotation [19]. Therefore, the first feature extraction method uses color and shape characteristics to generate the feature vector for each fruit image in the dataset. The used color moments to describe the images are color variance, color mean, color kurtosis, and color skewness [20], [21]. The shape features is described using Eccentricity, Centroid, and Euler Number features [22]. Eccentricity computes the aspect ratio of the distance of major axis to the distance of minor axis. It is calculated by minimum bounding rectangle method or principal axes method. The shape centroid defines the centroid position of image which is fixed in relative to the shape. The image Euler number defines relation between the connecting parts number and the holes number on image shape. Euler number is calculated by subtract the shape holes number from the contiguous parts number. [22]. The second method generates the feature vector uses the Scale Invariant Feature Transform (SIFT) algorithm [13]–[15]. It is an algorithm for image features extraction which is invariant to image rotation, scaling, and translation and partially invariant to affine projection and illumination changes. SIFT contains four main steps namely: scale-space extreme detection, keypoint localization, orientation assignment and keypoint descriptor [23]. The scale-space extreme detection step identifies the points of potential interest using differenceof-Gaussian function (DOG). In the keypoint localization, a model is fit to define location and scale for each candidate location. The selected Keypoints are determined based on their stability measures. In the orientation assignment step, orientations are allocated for each keypoint location according to the local image gradient directions. Then the operations are

1: Build the image Gaussian pyramid L(m, n, σ) using the following equations 1, 2, and 3.

$$G(m,n,\sigma) = \frac{1}{2\Pi\sigma^2} exp^{\frac{-(m^2+n^2)}{2\sigma^2}}, \qquad (1)$$

$$L(m,n,\sigma) = G(m,n,\sigma) * I(m,n), \qquad (2)$$

$$D(m,n,\sigma) = L(m,n,k\sigma) - L(m,n,\sigma), \qquad (3)$$

Where σ is the scale parameter, G(m, n, σ) is Gaussian filter, I(m, n) is smoothing filter, L(m, n, σ) is Gaussian pyramid, and D(m, n, σ) is difference of Gaussian (DoG).

2: Calculate the Hessian matrix.

3: After that, calculate the determinant of the Hessian matrix as shown in the equation 4 and eliminate the weak keypoints.

$$Det(H) = I_{mm}(m,\sigma)I_{nn}(m,\sigma) - (I_{mn}(m,\sigma))^2 \quad (4)$$

4: Calculate the gradient magnitude and orientation as in equations 5 and 6.

$$\begin{aligned} Mag(m,n) = & \ ((I(m+1,n) - I(m-1,n))^2 \\ & + (I(m,n+1) \\ & - I(m,n-1))^2)^{1/2} \quad (5) \end{aligned}$$

$$\theta(m,n) = \tan^{-1}(\frac{I(m,n+1) - I(m,n-1)}{I(m+1,n) - I(m-1,n)}) \quad (6)$$

5: Apply the sparse coding feature based on SIFT descriptors as in equations 7 and 8.

$$\min \sum_{i=1}^{S}(\|m_i - \sum_{j=1}^{Z} a_i^{(j)}\phi^{(j)}\|^2 + L) \quad (7)$$

$$L = \lambda \sum_{j=1}^{Z} |a_i^{(j)}| \quad (8)$$

Where m_i is the SIFT descriptors feature, a_j is mostly zero (sparse), ϕ is the basis of sparse coding, λ is the weights vector.

Algorithm 1: SIFT Feature Extraction Algorithm.

3.4 Shape Feature Extraction
A noise free image is obtained and shape features are calculated. The procedure to calculate the shape features from image of fruits is given in our methodology. Different shape features are calculated like area, perimeter, major-axis length and minor-axis length.

3.5 Color Feature Extraction
The images of fruit are used to extract different color features. Hence the RGB image is converted into different color like HSV, HIS etc. its mean is calculated for each color space. Similarly corresponding values are calculated for Apple, orange and strawberry also.

3.6 Texture Features Extraction
Texture features are calculated for various fruit images. For texture feature calculation RGB image is converted into a Gray scale image. GLCM matrix is determined using this grayscale image and Energy, Entropy, Contrast, Variance, Correlation, Homogeneity, Dissimilarity, Max probability, Sum of squares, Sum average, Sum variance, Sum entropy, D V, D E, M CC, IMC1, IMC2, I D, I DN, IDMN, CS, CP, 22 texture features are extracted.

4. RESULTS AND DISCUSSION
In this section we describe the categorization results on our dataset. In our dataset we considered 10 types of diseased fruit classes namely Apple, Mango, Plum, Pear, Papaya, Banana, Orange, Pomegranate, Citrus limetta, Sapota. From each class we have collected 250 samples and totally 2500 images are

used for experimentation. For classification we have used Artificial Neural Network.

Figure 2: Our database sample fruit images.

In Table 1. shows that how the proposed method gives better results and we have presented result for both after segmentation(AS) and before segmentation(BS), the results of AS is better compared to BS for average categorization for different training samples.

Table 1. Proposed method performance evaluation on categorization for Before Segmentation (BS) and After Segmentation (AS).

Proposed Method		SVM classifier		KNN classifier	
Training Samples	BS	AS	BS	AS	BS	AS
80	0.89	0.95	0.76	0.86	0.62	0.72
50	0.76	0.88	0.58	0.72	0.58	0.67
20	0.60	0.71	0.52	0.58	0.48	0.53

5. CONCLUSION
In this paper, we have proposed method to classify fruit as diseased and non-diseased. Firstly, we used K means clustering method for segmentation of diseased regions. Later, we used to extract shape, color and texture features on segmented diseased regions. We have conducted extensive experimentation and results shows that proposed method gives better performance compared to traditional classifiers.

6. REFERENCES

[1] Safiri, S., M. Sani, et al., Fruit and Vegetable, Fat, and Sugar-Sweetened Beverage Intake Among Low-Income Mothers Living in Neighborhoods With Supplemental Nutrition Assistance ProgramEducation: Methodological Issues. Journal Of Nutrition Education And Behavior, 2017. 49(3): pp. 272-272

[2] Estevez-Santiago, R., B. Olmedilla-Alonso, et al., Lutein and zeaxanthin supplied by red/orange foods and fruits are more closely associated with macular pigment optical density than those from green vegetables in Spanish subjects. Nutrition Research, 2016. 36(11): pp. 1210-1221

[3] Seema, A. Kumar, et al. Automatic Fruit Grading and Classification System Using Computer Vision: A Review. in 2nd IEEE International Conference on Advances in Computing and Communication Engineering (ICACCE). 2015. Dehradun, INDIA: IEEE. pp. 598-603

[4] Lin, Y.F., M.Y. Chen, et al., DNP and ATP induced alteration in disease development of Phomopsis longanae Chi-inoculated longan fruit by acting on energy status and reactive oxygen species production-scavenging system. Food Chemistry, 2017. 228: pp. 497-505

[5] Erdenee, B., Ryutaro, T., Tana, G., 2010, Particular Agricultural Land CoverClassification Case Study Of Tsagaannuur, Mongolia. In: IEEE InternationalGeoscience & Remote Sensing Symposium, 3194-3197.

[6] Krishna, M., Jabert, G., 2013. Pest control in agriculture plantation using image processing. IOSR J. Electron. Commun. Eng. (IOSR-JECE) 6 (4), 68–74.

[7] Tewari, V.K., Arudra, A.K., Kumar, S.P., Pandey, V., Chandel, N.S., 2013. Estimation of plant nitrogen content using digital image processing. Int. Commission Agricu. Biosyst. Eng. 15 (2), 78–86.

[8] Patil, J.K., Kumar, R., 2011. Advances in image processing for detection of plant diseases. J. Adv. Bioinf. Appl. Res. ISSN 2 (2), 135–141.

[9] [Naik, S., Patel, B., 2017. A machine vision based fruit classification and grading: a review. Int. J. Comput. Appl. 170 (9), 22–34.

[10] Dubey, S.R., Jalal, A.S., 2015a. Application of image processing in fruits and vegetables analysis: a review. J. Intell. Syst. 24 (4), 405–424.

[11] Zhang, B., Huang, Z., Li, J., Zhao, C., Fan, S., Wu, J., Liu, C., 2014a. Principle, developments and applications of computer vision for external quality inspection of fruits and vegetables: a review. Food Res. Int., 326–343

[12] Wu, L., Classification of fruits using computer vision and a multiclass support vector machine. Sensors, 2012. 12(9): pp. 12489-12505

[13] Wang, S., Y. Zhang, et al., Fruit Classification by Wavelet-Entropy and Feedforward Neural Network Trained by Fitness-Scaled Chaotic ABC and Biogeography-Based Optimization. Entropy, 2015. 17(8): pp. 5711-5728

[14] Ji, G., Fruit classification using computer vision and feedforward neural network. Journal of Food Engineering, 2014. 143: pp. 167-177

[15] Wu, J., Fruit classification by biogeography-based optimization and feedforward neural network. Expert Systems, 2016. 33(3): pp. 239-253

[16] Wang, S., Y. Zhang, et al., Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. International Journal of Imaging Systems and Technology, 2015. 25(2): pp. 153-164

[17] Lu, S.Y., Z.H. Lu, et al. Fruit Classification by HPA-SLFN. in 8th International Conference on Wireless Communications & Signal Processing. 2016. Yangzhou, China: IEEE. pp. 11-17

[18] Kuang, H.L., L.L.H. Chan, et al., Fruit classification based on weighted score-level feature fusion. Journal of Electronic Imaging, 2016. 25(1), Article ID: 013009.

[19] [Garcia, F., J. Cervantes, et al., Fruit Classification by Extracting Color Chromaticity, Shape and Texture Features: Towards an Application for Supermarkets. IEEE Latin America Transactions, 2016. 14(7): pp. 3434-3443

[20] Zawbaa, H.M., M. Hazman, et al. Automatic fruit classification using random forest algorithm. in 14th International Conference on Hybrid Intelligent Systems (HIS). 2014. Kuwait, KUWAIT: IEEE. pp. 164-168

[21] Rocha, A., D.C. Hauagge, et al., Automatic fruit and vegetable classification from images. Computers and Electronics in Agriculture, 2010. 70(1): pp. 96-104

[22] Samajpati, B.J. and S.D. Degadwala. Hybrid Approach for Apple Fruit Diseases Detection and Classification Using Random Forest Classifier. in International Conference on Communication and Signal Processing (ICCSP). 2016. Melmaruvathur, INDIA: IEEE. pp. 1015-1019.

[23] Capizzi, G., G. Lo Sciuto, et al. Automatic Classification of Fruit Defects based on CoOccurrence Matrix and Neural Networks. in Federated Conference on Computer Science and Information Systems. 2015. New York: IEEE. pp. 861-867