Fruit Disease Recognition and Automatic Classification using MSVM with Multiple Features

A. S. M. Shafi
Department of Computer Science and Engineering
Mawlana Bhashani Science and Technology University
Tangail, 1900, Bangladesh

Md. Bayazid Rahman
Department of Computer Science and Engineering
Mawlana Bhashani Science and Technology University
Tangail, 1900, Bangladesh

Mohammad Motiur Rahman
Department of Computer Science and Engineering
Mawlana Bhashani Science and Technology University
Tangail, 1900, Bangladesh

ABSTRACT
Image processing and machine learning play an important role in fruit disease identification and classification by means of image segmentation and pattern recognition. Traditional fault detection in the fruit surface is carried out manually by means of human inspection which is very time consuming and laborious. In this paper we have proposed a method for fruit disease identification using segmentation techniques and use a supervised learning technique for classifying images based on data analyzed from RGB colored images. Three types of common apple diseases are taken into considerations in this paper. The experimental results demonstrate that the proposed approach is promising and effective by showing the classification accuracy which has achieved more than 94% using several features.

Keywords
Image segmentation, filtering, global thresholding, feature extraction, supervised classification

1. INTRODUCTION
Digital images are the most important way of representing information in these days. Extracting the information and data from digital images is one of the most important tasks. With these extracted information, they can be used for many tasks where decision has to be made. Before in earlier age digital images are used for reporting about the quality in agricultural science. It was quietly impossible to predict and make decision only analyzing the images. But now in this era this problem has come to an end. With the evolution of advancement of computer and microelectronics, image processing and digital image analysis is not a hard issue at all. The right and appropriate way to detect and identify fruit disease is direct observation of fruits by experts. But it can't be possible for all the time for the lack of transportation and many other reasons. In developing countries there are hardly any fruit disease expert is found. If the diseased fruit is not removed, it can be catastrophe for an entire farm and can worsen the current situation. Usually fruit disease can be seen on the surface of the fruit. If the pesticide is used more than standard amount for the treatment of fruit disease, it can hamper ground water at danger level and the toxic level of agricultural product will go on. The cost of pesticide is higher, so that uses of this dangerous chemical must be minimized. Fruit disease is one of the main causes of economic loss in the field of harvesting. As an example, farmers could make money like 11-million- dollar profit of the diseased soybeans named as soybean rust if it was fresh. [1].

As example, there are some common diseases of apples like scab, apple scab, apple blotch, apple rot, black pox and so on. Apple scab, apple blotch and apple rot are disease for test case. Apple scab shows as brown or gray spots. It is like deformed skin and flesh. Apple scab is not responsible for rotting fruit but it can crack apple in a large scale. Apple blotch is a disease that is only seen in dark. Apple blotch is only seen on the colored apple like green and light apple. Our favorable approach which is mentioned in the paper is by following these steps. Global thresholding technique is measured for defected segmentation in the first step. In the second step art features are collected and extracted from the segmented image. And at the last images classified into Multi-class Support Vector Machine. The rest of the paper is as follows, Section 2, we do work on related works. Section 3, using global thresholding based segmentation we discuss about fruit disease recognition. After that we represent many types of features from the segmented images. Finally we classify the images based on extracted features. Section 4 is for experimental result. Finally section 5 is for drawing conclusions.

2. LITERATURE REVIEW
Many activities are being showed on the area of image categorization. Fruit segmentation of fruits are done only using threshold approach on the area of defect segmentation [2]. Threshold method which is globally adapted (modified version of Otsu's approach) to segment rotten contamination defects on apples are presented [3]. Classification based methods are attempted for partitioning pixels into various classes by using various classification methods. Most of the researchers use Bayesian classification methods. [4][5] where they compare pixels with pre-calculated model. After examining they classified them as they are defected or healthy. Those classification are not useful and benefited which are unsupervised. They are not benefited because the lack of target values in the processing of learning. Leeman, Magein and Destain used this approach in 1998 [6] for segmentation of defect. Pal and Pal provide a detailed review on different image segmentation technique [7]. Among huge amount of existing segmentation process, many people may have used unsupervised clustering methods. As an example, the basis of image region merging of image segmentation is correlate of agglomerative clustering [8]. Graph-cut method as like minimal cut and normalized cut characterize the problem. The problem of characterize based on graph theoretic way [9]. But there is some problem on processing this method. The main problem for this type of methods is mainly known as the validity of problem. It is hard to analyze how the numbers of cluster of any image will be decided. This problem is still unresolved. Most of the techniques need the terminating criterion from the user.

Sowmya and Sheelarani have used the soft computing techniques on segmenting color image [10]. The soft computing techniques they had used mainly were competitive

neural network and also possibilistic C means algorithm (PCM). Fuzzy set and Fuzzy logic techniques were used for solving segmentation problem by researchers. Fan et al proposed automatic image segmentation by integrating color edge detection and seeded region growing [11]. Adams and Bischof [12] proposed another method by using seeded region growing. Another image segmentation method in the image [13] was proposed by Shih and Cheng. In where these segmentation methods were based on standard deviation in the neighbor. Concept of background subtraction for clustering k-means have been used by authors [14] [15]. A fruit may have been infected with more than one disease at a single time. So more than two clusters must be used to segment and inspect the infected part with background and fruit.

3. FRUIT DISEASE RECOGNITION

The framework of the proposed approach is shown in Figure 1. For the fruit disease recognition problem, exact defect segmentation is required; otherwise the features of the non-infected region will dominate over the features of the infected region. In this approach global thresholding based segmentation is performed to detect the region of interest, i.e., infected part only. In the second step, features are extracted from the segmented images that are being used for the training and store in a feature database. Then we train multi class support vector machine (MSVM) with the feature stored in the feature database. Finally any input image can be classified into one of the classes using feature derive from the segmented part of the input image and trained MSVM. Gaussian filters are used for image preprocessing because they have a property that their support in the time domain is equal to their support in the frequency domain.
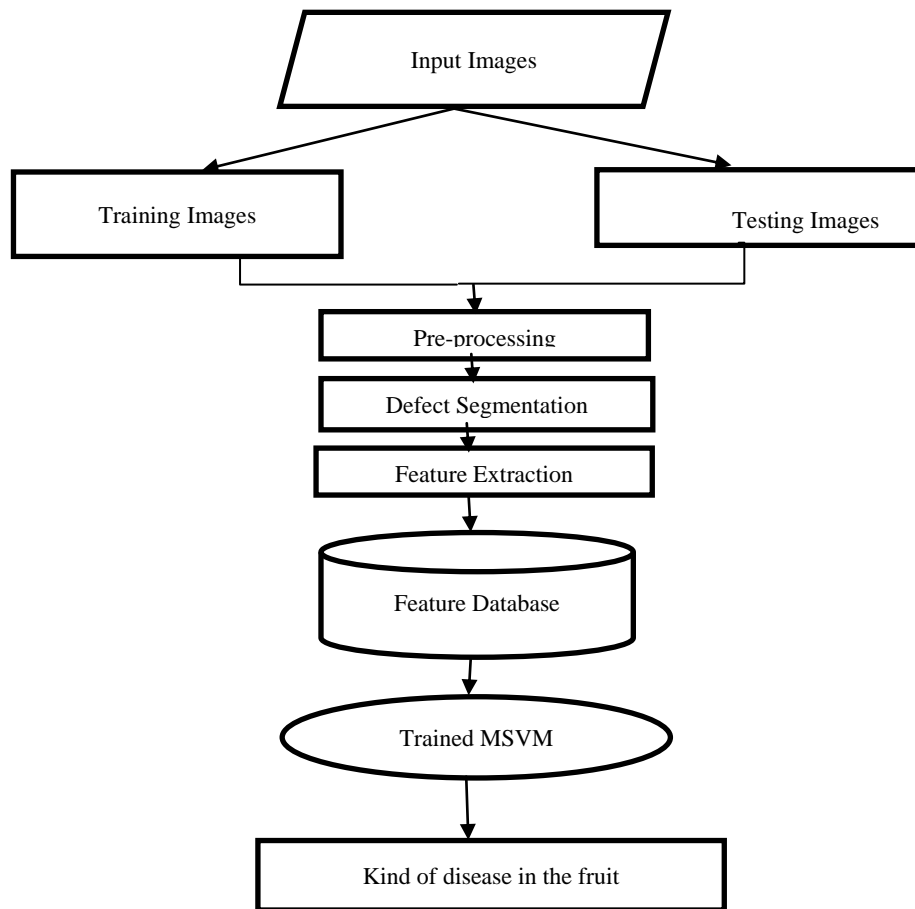

Fig. 1. Framework of the proposed approach

3.1 Defect Segmentation

Global Thresholding is used to segment the infected portion of the input images which requires the following steps

Step 1 Select an initial estimate for T.

Step 2 Segment the image using T. This will produce two groups of pixels: G_1 consisting of all pixels with gray level values > T and G_2 consisting of pixels with gray level values < = T.

Step 3 Compute the average gray level values μ_1 and μ_2 for the pixels in regions G_1 and G_2.

Step 4 Compute a new threshold value.

Step 5 T = 0.5 $(\mu_1 + \mu_2)$

Step 6 Repeat steps 2 through 4 until the difference between the value of T in successive iterations is smaller than a predefined parameter ΔT.
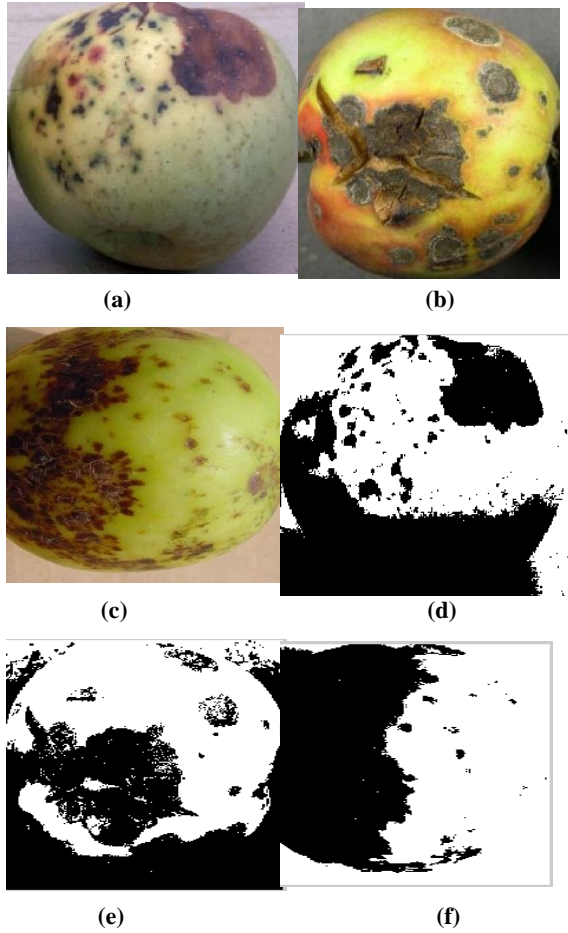
Fig. 2. Sample images of (a) Apple Rot (b) Apple Scab (c) Apple Blotch and defect segmented images using global thresholding (d) Apple Rot defect (e) Apple Scab defect (f) Apple Blotch defect

3.2 Feature Extraction

In general image categorization relies on combinations of statistical, structural and spectral features. Statistical approaches describe the objects using global and local descriptors such as mean, variance, and entropy. Structural approaches represent the object's appearance using well-known primitives such as patches of important parts of the object. Finally, spectral approaches describe the objects using some spectral space representation such as Fourier spectrum [16].

3.2.1 Gray Level Co-occurrence Matrix (GLCM)

Four important features contrast, energy (Angular Second Moment), correlation and the homogeneity are selected for implementation.

3.2.1.1 Contrast

Contrast returns a measure of the intensity contrast between a pixel and its neighbor over the whole image.

Contrast = $\sum_{i,j} |i - j|^2 p(i,j)$

Range = [0 (size (GLCM, 1) - 1) ^ 2]

Contrast is 0 for a constant image.

3.2.1.2 Energy (Angular Second Moment)

Angular Second Moment is also known as Uniformity or Energy. It is the sum of squares of entries in the GLCM Angular Second Moment measures the image homogeneity.

Energy = $\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p_{ij}^2$

Where i, j are the spatial coordinates of the function p (i, j), Ng is gray tone

3.2.1.3 Correlation

Correlation measures the linear dependency of grey levels of neighboring pixels.

Correlation = $\frac{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i,j)P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$

3.2.1.4 Homogeneity

Homogeneity returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Homogeneity = $\sum_{i,j} \frac{p(i,j)}{1+|i-j|}$

3.2.2 Entropy

Image entropy is a quantity which is used to describe the amount of information which must be coded for by a compression algorithm.

3.2.3 Global Color Histogram (GCH)

A GCH is a set of ordered values, for each distinct color, representing the probability of a pixel being of that color. Uniform normalization and quantization are used to avoid scaling bias and to reduce the number of distinct colors [16].

3.2.4 Local Binary Pattern (LBP)

Given a pixel in the input image, LBP is computed by comparing it with its neighbors [17]:

$\text{LBP}_{N,R} = \sum_{n=0}^{n-1} S(v_n - v_c) 2^n, s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$

3.3 Classification

The new testing images are classified on the basis of feature database of training images using MSVM. Multiclass Support Vector Machine is a set of Support Vector Machines (SVMs) for the training and classification.

3.4 Proposed MSVM Algorithm

Inputs:
- L, a learner (training algorithm for binary classifiers)
- samples X
- labels y where $y_i \in \{1, \dots K\}$ is the label for the sample X_i

Output:
- a list of classifiers f_k for $k \in \{1, \dots, K\}$

Procedure:

- For each k in $\{1, \dots, K\}$:
    - Construct a new label vector $y_i = 1$ where $y_i = k$, 0 (or −1) elsewhere
    - Apply L to X, y to obtain f_k

Making decisions means applying all classifiers to an unseen sample x and predicting the label k for which the corresponding classifier reports the highest confidence score: $y = \arg\max_{k \in 1 \dots k} f_k(x)$

4. EXPERIMENTAL RESULTS

To illustrate the proposed approach, a data set composed of four different categories: Apple Blotch (33), Apple rot (40), Apple scab (37), and Normal Apple (50) are used. The total number of apple fruit images is 160.
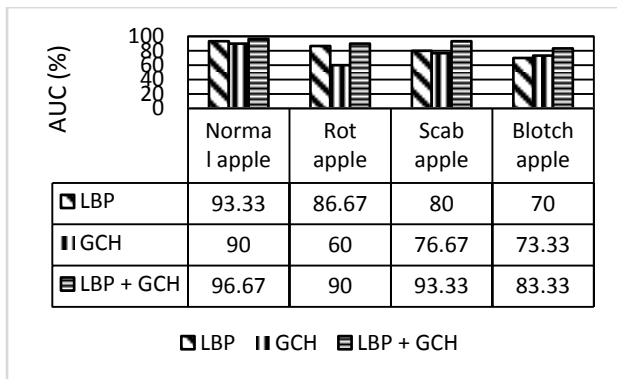
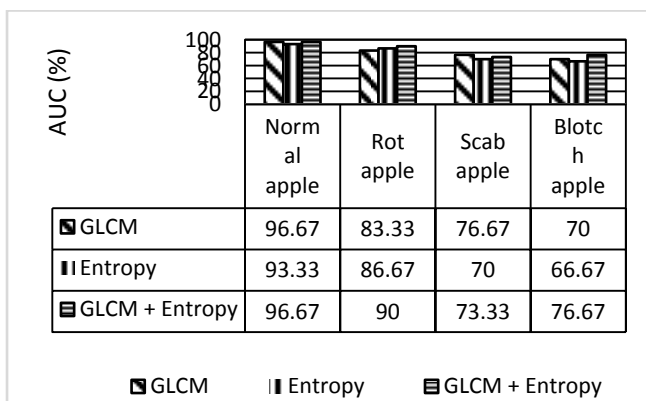**Fig. 3. AUCs obtained for defect classification using structural features**

	Normal apple	Rot apple	Scab apple	Blotch apple
LBP	93.33	86.67	80	70
GCH	90	60	76.67	73.33
LBP + GCH	96.67	90	93.33	83.33

LBP   GCH   LBP + GCH

Fig. 4. AUCs obtained for defect classification using statistical features

	Normal apple	Rot apple	Scab apple	Blotch apple
GLCM	96.67	83.33	76.67	70
Entropy	93.33	86.67	70	66.67
GLCM + Entropy	96.67	90	73.33	76.67

GLCM   Entropy   GLCM + Entropy
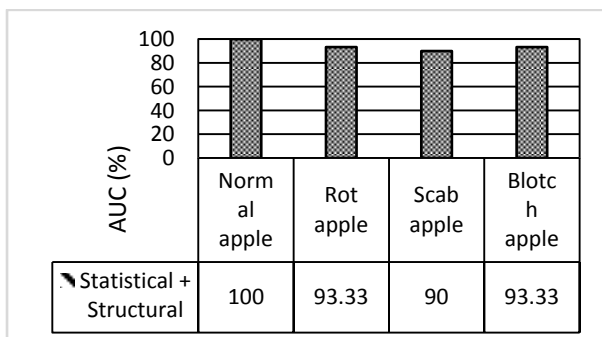
Fig. 5. AUCs obtained for defect classification using statistical in combined with structural features

	Normal apple	Rot apple	Scab apple	Blotch apple
Statistical + Structural	100	93.33	90	93.33

5. CONCLUSION AND FUTURE WORK

An image processing based experiment along with machine learning is proposed for the automatic detection and classification of fruit diseases from images using statistical and structural features. Based on the experiment we have achieved more than 94%. In this paper only single type of fruit disease can be detected from a single image and different background color of an image sometimes affect the classification result. In future we will improve our work so that it will be capable of working with complex background and can detect various types of diseases in a single fruit.

6. REFERENCES

[1]. Roberts, M. J., Schimmelpfennig, D., Ashley, E., Livingston, M., Ash, M., & Vasavada, U. (2006). *The Value of Plant Disease Early- Warning Systems* (No. 18). Economic Research Service, United States Department of Agriculture.

[2]. Li, Q., Wang, M., & Gu, W. (2002, November). Computer Vision Based System for Apple Surface Defect Detection. *Computers and Electronics in Agriculture*, 36, 215-223.

[3]. Kim, M. S., Lefcourt, A. M., Chen, Y. R., & Tao, Y. (2005). Automated Detection of Fecal Contamination of Apples Based on Multispectral Fluorescence Image Fusion. *Journal of food engineering*, 71, 85-91.

[4]. Kleynen, O., Leemans, V., & Destain, M. F. (2005). Development of a Multi-Spectral Vision System for the Detection of Defects on Apples. *Journal of Food Engineering*, 69, 41-49.

[5]. Leemans, V., Magein, H., & Destain, M. F. (1999). Defect Segmentation on 'Jonagold' Apples using Color Vision and a Bayesian Classification Method. *Computers and Electronics in Agriculture*, 23, 43-53.

[6]. Leemans, V., Magein, H., & Destain, M. F. (1998). Defect Segmentation on 'Golden Delicious' Apples by using Color Machine Vision. *Computers and Electronics in Agricultur*e, 20, 117-130.

[7]. N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, pp. 1227-1294, 1993.

[8]. H. Frigui and R. Krishnapuram, "Clustering by competitive agglomeration," *Pattern Recognition*, vol. 30, no. 7, pp. 1109-1119, 1997.

[9]. Y. Boykov, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision* (IJCV), vol. 70, no. 2. pp. 109–131, 2006.

[10]. B. Sowmya and B. Sheelarani, "Colour image segmentation using soft computing techniques," *International Journal of Soft Computing Applications*, vol. 4, pp. 69-80, 2009.

[11]. J. F. David, K. Y. Yau, and A. K. Elmagarmid, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Transactions On Image Processing* (TIP), vol. 10, no. 10, pp. 1454-1466, 2001.

[12]. R. Adams and L. Bischof, "Seeded region growing," *IEEE Transaction on pattern analysis and machine intelligence* (PAMI), vo1. 6, no. 6, pp. 641-647, 1994.

[13]. F. Y. Shih and S Cheng, "Automatic seeded region growing for color image segmentation," *Image and Vision Computing* (IVC), vol. 23, no. 10, pp. 877-886, 2005.

[14]. S. R. Dubey and A. S. Jalal, "Robust Approach for Fruit and Vegetable Classification", *Procedia Engineering*, vol. 38, pp. 3449 – 3453, 2012.

[15]. S. R. Dubey and A. S. Jalal, "Species and Variety Detection of Fruits and Vegetables from Images", *International Journal of Applied Pattern Recognition* (IJAPR), vol. 1, no. 1, pp. 108 – 126, 2013.

[16]. Gonzalez, R., Woods, R., 2007. *Digital Image Processing, 3rd edition*. Prentice-Hall.

[17]. T. Ojala, M. Pietik äinen and T. T. Mäenpä ä , Multiresolution gray-scale and rotation invariant texture classifi cation with local binary pattern, *IEEE T. Pattern Anal.* 24 (2002), 971 – 987.

[18]. Dubey, S.R. and Jalal, A.S. (2014) 'Fruit disease recognition using improved sum and difference histogram from images', *Int. J. Applied Pattern Recognition*, Vol. 1, No. 2, pp.199–220.