Identification of Infected Pomegranates using Color Texture Feature Analysis

Meenakshi M. Pawar
Professor
Department of Electronics and Tele-Communication
SVERI'S College of Engineering Pandharpur, Maharashtra, India

Sanman Bhusari
Under Graduate Student
Department of Electronics and Tele-Communication
SVERI'S College of Engineering Pandharpur, Maharashtra, India

Akshay Gundewar
Under Graduate Student
Department of Electronics and Tele-Communication
SVERI'S College of Engineering Pandharpur, Maharashtra, India

ABSTRACT
In this study, a new approach is used to automatically detect the infected pomegranates. In the development of automatic grading and sorting system for pomegranate, critical part is detection of infection. Color texture feature analysis is used for detection of surface defects on pomegranates. Acquired image is initially cropped and then transformed into HSI color space, which is further used for generating SGDM matrix. Total 18 texture features were computed for hue (H), saturation (S) and intensity (I) images from each cropped samples. Best features were used as an input to Support Vector Machine (SVM) classifier and tests were performed to identify best classification model. Out of selected texture features, features showing optimal results were cluster shade (99.8835%), product moment (99.8835%) and mean intensity (99.8059%).

Keywords
Pomegranate, disease detection, machine vision, color co-occurrence method, texture features, SVM.

1. INTRODUCTION
The pomegranate (Punica granatum) is a fruit-bearing deciduous shrub or small tree that grows to between five and eight metres tall and is best suited to climates where winters are cool and summers are hot. The pomegranate is thought to have been first cultivated 5 to 6,000 years ago and is native to the regions from Iran through to north India. It is now widely cultivated throughout Eastern Europe, Asia and the USA, the main areas of world production being in India, Iran, Spain and California. Pomegranates can be consumed as fresh fruit or used in fruit juices, teas, pharmaceutical and medicinal products and in dyes or as decoration. There are several hundred different varieties of pomegranate recognised in Iran alone and even more globally, some of the cultivars that have the greatest impact are 'Moller', 'Ahmar', 'Bhagawa', 'Hicaznar' and 'Dente di cavallo'. Globally, it is estimated that total production amounts to around 2,000,000 tonnes, of which India produces approximately 50% in the states of Maharashtra and Andhra Pradesh. Iran is the second largest, producing around 35% of global production. Spain produces around 2.5% and the USA has around 10,000ha under production.

This study centres at developing a method to detect infected pomegranate using color texture features. The various steps involved for development of database of features are summarized in figure 1 below:

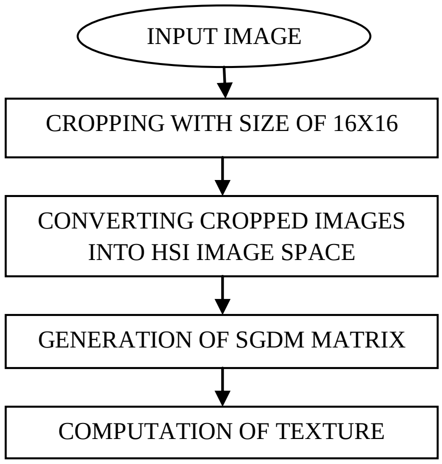

Fig 1: Process of developing the database of features.

2. LITERATURE REVIEW
Thomas J. Burks et al (2009) [1] demonstrated that color imaging and texture feature analysis could be used for classifying citrus peel diseases under the controlled laboratory lighting conditions. The present work is an extension of that research, providing a feasibility analysis of the technology in classification of infected pomegranates. Edwards and Sweet (1986) [2] used reflectance spectra of the entire citrus plant for estimating the damage caused due to citrus blight. Hetal Patel et al (2011) [3] designed the algorithm aiming at calculation of different weights for features like intensity, color, orientation and edge of the test image. S.Arivazhagan et al (2010) [4] used computer vision strategy to recognize a fruit rely on four basic features which characterize the object on the basis of intensity, color, shape and texture. R. Pydipati et al (2006) [5] used the color co-occurrence method to determine whether texture based hue, saturation, and intensity color features in conjunction with statistical classification algorithms could be used to identify diseased and normal citrus leaves under laboratory conditions. Shearer Scott (1990) [6] proposed a dissertation for the development of new color-texture analysis method to characterize and identify canopy sections of nursery plant cultivators. Yousef Al Ohali (2010) [7] studied the performance of a back propagation neural network classifier and tested the accuracy of the system on preselected date samples. Czeslaw Puchalski et al (2008) [8] developed a system for identifying surface defects on apple and successfully obtained 96% classification accuracy. H. Al-Hiary et al (2011) [9] experimentally evaluated a software solution for automatic detection and classification of plant leaf

diseases. A. Camargo, J.S. Smith(2009) [14] reported reports a machine vision system for the identification of the visual symptoms of plant diseases, from coloured images.In this study, features were then used as inputs to a Support Vector Machine (SVM) classifier and tests were performed to identify the best classification model. Lanlan Wu et al (2010) [10] investigated the support vector machine, as a classifier tool to identify the weeds in corn fields at early growth stage. Li Daoliang et al (2012) [11] shown that texture-related features such as co-occurrence matrices might be used as effective discriminators for high resolution remote sensing images. Xing-yuan Wang et al (2012) [12] presented an effective color image retrieval method based on texture, which used the color co-occurrence matrix to extract the texture feature and measure the similarity of two color images. Ryusuke Nosakaet al (2012) [13] suggested a new image feature based on spatial co-occurrence within micro patterns, where each micro pattern is presented by a Local Binary Pattern (LBP). Qinghua Guo, Maggi Kelly, Catherine H. Graham [15] developed an automated classification system of pizza sauce spread using colour vision and support vector machine (SVM).

3. COLOR CO-OCCURRENCE MATRIX

Thomas J. Burks et al (2009) [1] used the color co-occurrence method for citrus peel fruit classification. Pydipati et al. (2006) [5] utilized the color co-occurrence method to extract various textural features from the color RGB images of citrus leaves. There are two main analysis methods for calculation of texture viz.

1) Structural Approach
2) Statistical Approach

Statistical approach, which is used here, is a quantitative measure of arrangement of intensities in a region. Statistical methods use second order statistics to describe the relationships between pixels within the region by constructing Spatial Gray-level Dependency Matrices (SGDM). A SGDM matrix is the joint probability occurrence of gray levels 'i' and 'j' for two pixels with a defined spatial relationship in an image. Distance'd' and angle 'θ' are used to define the spatial relationship. If the texture is coarse and distance'd' is small compared to the size of the texture elements, the pairs of points at distance d should have similar gray levels. In turn, if the texture is fine and distance d is comparable to the texture size, then the gray levels of points separated by distance d should often be quite distinct, so that the values in the SGDM matrix should be disperse uniformly. Thus, texture directionality can be analyzed by examining spread measures of SGDM matrices created at various distances'd'. Extraction of a numerous features is possible using the SGDM matrices generated in the above manner.

The test image is then cropped such that around 2000 cropped images of size 16 x 16 each are formed. These RGB images are converted into HSI color space representation. Then each pixel map is used to generate a color co-occurrence matrix, resulting in three CCM matrices, one for each of the H, S and I pixel maps. These matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance and orientation from any pixel, given that pixel has a second particular gray level. For a position operator p, we can define a matrix 'Pij' that counts the number of times a pixel with grey-level i occurs at position p from a pixel with grey-level j. For example, if we have four distinct grey-levels 0, 1, 2 and 3, then one possible SGDM matrix P (i, j, 1, 0) is given below as shown:

$$I(x,y)=\begin{matrix}1&0&3&0\\2&2&0&3\\3&2&0&2\\1&3&2&3\end{matrix}\qquad P=\begin{matrix}0&1&3&3\\1&0&0&1\\3&0&2&3\\3&1&3&0\end{matrix}$$

Statistical methods use order statistics to model the relationships between pixels within the region by constructing Spatial Gray-level Dependency Matrices (SGDM's). A SGDM matrix is the joint probability occurrence of gray levels 'i' and 'j' for two pixels with a defined spatial relationship in an image. The SGDMs are represented by the function P (i, j, d, Θ) where 'i' represents the gray level of the location (x, y) in the image I(x, y), and j represents the gray level of the pixel at a distance d from location (x, y) at an orientation angle of Θ. The nearest neighbour mask is exemplified in figure 2, where the reference pixel is shown as an asterisk.

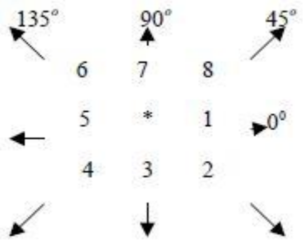

Fig 2: Nearest neighbour mask for calculating spatial Gray-level dependence matrix (SGDM)

4. TEXTURE FEATURE CALCULATION

Texture is one of the characteristics that segments images into regions of interest (ROI) and classifies those regions. It gives information about the spatial organization of the intensities in an image. In general, texture can be defined as a characteristic of image that can provide a higher-order description of the image and includes information about the spatial distribution of tonal variations or gray tones. Texture involves the spatial distribution of gray levels. Thus, two-dimensional histograms or co-occurrence matrices are reasonable texture analysis tools. An image can have one or more textures. These features are useful in carrying out differentiation algorithms for detection purpose ahead.

For differentiating infected pomegranate from the uninfected ones the following features are reckoned for the components H, S and I:

1) Cluster Shade: It is measures of the lack of symmetry which shows skewness of the matrix.

$$F1 = \sum_{i,j=1}^{Ng-1} (i - P_x + j - P_y)^3 P(i,j) \qquad (1)$$

2) Sum Entropy: It measures the randomness of the elements of the matrix, when all elements of the matrix are maximally random entropy has its highest value.

$$F2 = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} P(i,j) \ln P(i,j) \qquad (2)$$

3) Uniformity: It is the measure of uniform pattern in the image.

F3 = $$\sum_{i=1}^{Ng-1}\sum_{j=1}^{Ng-1}[P(i,j)]^2 \tag{3}$$

4) Mean Intensity: It is the measure of image brightness derived from the co-occurrence matrix.

F4 = $$\sum_{i=0}^{Ng-1} iP_x(i) \tag{4}$$

5) Product Moment: It is analogous to the co-variance of the intensity of co-occurrence matrix.

F5 = $$\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}(i-F_4)(j-F_4)P(i,j) \tag{5}$$

6) Inverse Difference Moment: It returns the measures of closeness of the distribution of SGDM elements to the SGDM diagonal.

F6 = $$\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}\frac{P(i,j)}{1+(i-j)^2} \tag{6}$$

Where, P (i, j) is the SGDM matrix and Ng is total number of gray levels considered for generation of SGDM matrix.

5. SUPPORT VECTOR MACHINE

A support vector machine (SVM) is based on a statistical learning theory showing various advantages over existing soft computing methods. As compared with the soft computing methods generally used for various applications, SVM gives better performance. Lanlan Wu et al (2010) [10] achieved classification accuracy ranging 92.31 to 100% for different feature selections using SVM. It is used in case of finite sample data. It aims at acquiring the worthy solutions on the ground of present data rather than the optimal value for infinite samples.

If the original problem is stated in a finite dimensional space and the sets required to separate out are not linearly separable in that space then it can be mapped into a higher dimensional space, so as to make separation between the sets more distinguishable. This peculiar quality of SVM shows its good potentiality of abstraction. As its ramification is independent of sample dimension, it solves the problem of dimension disaster. This method is successfully implemented in face recognition, machine fault detection and in handwritten digit recognition for tablets, PDA's and other electronic devices.

As shown in figure 3 the two sets of sample points are represented by green and red color, H is the hyper plane or separator, H1 and H2 are planes running parallel to the hyper plane and pass through the sample points closest to the hyper plane. As there are many such hyper planes which can show divergence between the two sample sets, a hyper plane which has maximum distance from both H1 and H2 is chosen. The distance between H1 and H2 is called as margin. The hyper plane with largest margin is called as maximum-margin hyper plane and the linear classifier it defines is addressed as maximum-margin classifier. The more margin a hyper plane has, less is the error in optimizing the separation between sample points.

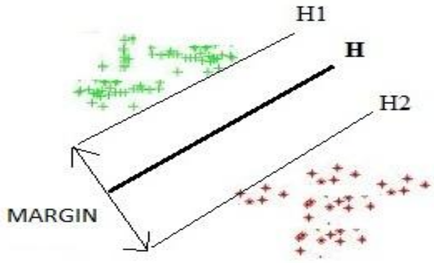

Fig 3: Distinguishing sample point using SVM

This case study uses the Gaussian Radial Basis Function Kernel for the classification. For the nonlinear case, we project the original space into a higher dimension space in which the SVM can construct an optimal separating hyper-plane as explained above. Let us consider a function ϕ (x), such that F = $\{\phi(x): x \in X\}$ is the feature point corresponding to the data item x. For SVM, the kernel function is represented as

$$K(x, z) = (\phi(x), \phi(z)), \forall x, z \in X$$

replacing the inner product (x, z).

6. RESULTS AND DISCUSSION

For the purpose classification, it is often anticipated that the linear separabilty of the mapped samples is enhanced in the kernel feature space so that applying traditional linear algorithms in this space could result in better performance compared to those obtained in the original input space. If an inappropriate kernel is chosen, the classification performance of kernel-based methods can be even worse than that of their linear counterparts. Li Daoliang et al (2012) [11] verified that SVM classification systems leads to the successful discrimination of targets when fed with appropriate information. Therefore, selecting a proper kernel with good class separability plays a vital role in kernel-based classification algorithms. Here the Gaussian basis function performs well to do the task. The contour plots for features cluster shade and mean intensity are shown in figure 4 and figure 5 below.

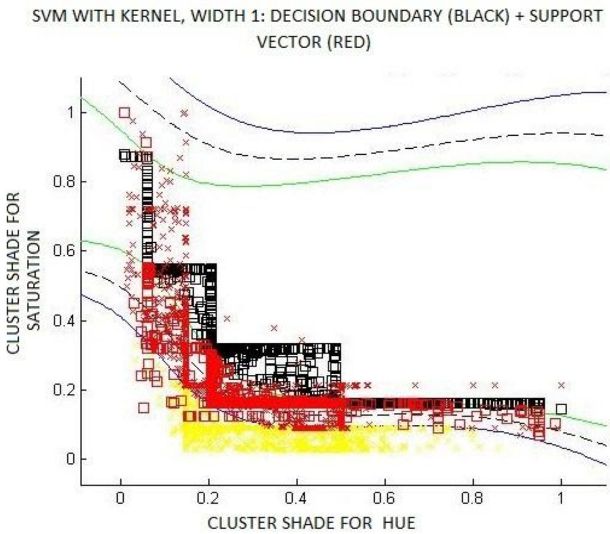

Fig 4: Contour Plot of Cluster shade

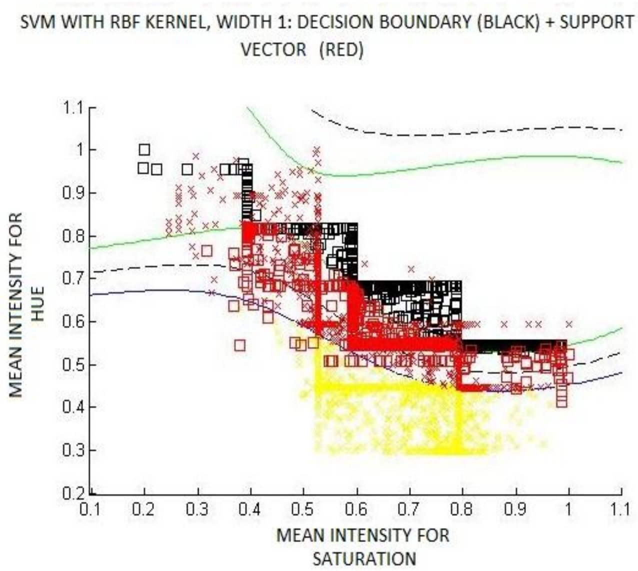

Fig 5: Contour Plot of Mean Intensity

Various features are computed from the test images for components H, S and I which are further used for generating the 3D plot of each feature independently for distinguishing the infected and uninfected pomegranate over the plot. In figure 3, red color signifies the feature points for the infected pomegranate whereas green color signifies the same for uninfected ones. Here X, Y and Z axes represent feature of image components Hue, Saturation and Intensity respectively. It can be easily observed through figure 6 that Cluster Shade (Plot 1), Mean Intensity (Plot 4) and Product Moment (Plot 5) features are best for differentiating the diseased pomegranate as the data points are clearly separable whereas Sum Entropy (Plot 2), Uniformity (Plot 3) and Inverse difference Moment (Plot 6) features do not possess separate explicit data points.

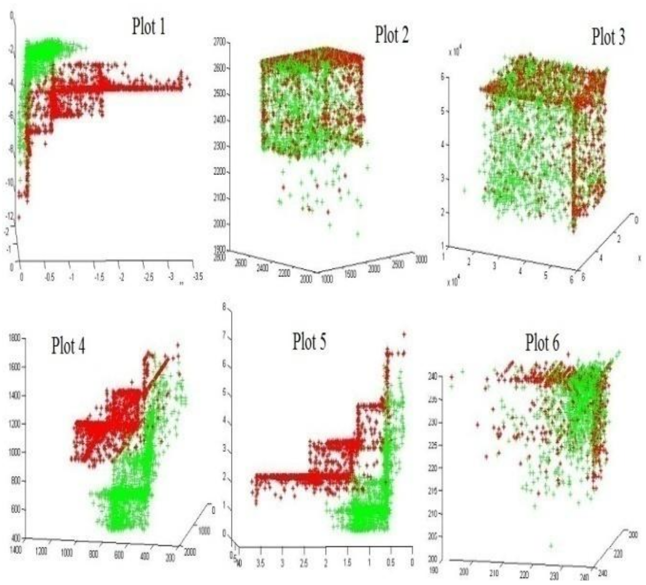

Fig 6: 3D Plots for features Plot 1-Cluster Shade, Plot 2-Sum Entropy, Plot 3-Uniformity, Plot 4-Mean Intensity, Plot 5-Product Moment & Plot 6-Inverse Difference Moment

Figure 7 shows bar diagram of success rates obtained after passing the test images through SVM training function. As it is seen that cluster shade, mean intensity and product moment show a eminent success rate whereas sum entropy, uniformity and inverse difference moment have modest success rate and hence we egest these features for final classification of images under test.

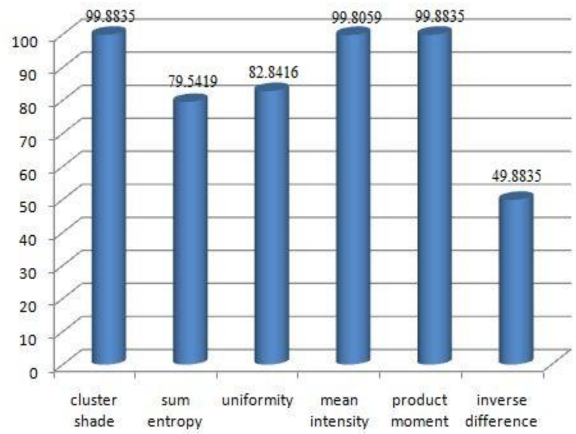

Fig 7: Success Rate percentage of Features after undergoing SVM training

Based on obtained results, features with percentage above 90% are selected to acquire higher accuracy for classifying the images. The selected features for classification show less error or misclassification. Cluster Shade and Cluster Prominence have 6, variance has 10 such misclassifications.
The final classification is accomplished by comparing the features calculated for the test images with the reference training database generated using support vector machine. The images are classified into two categories namely infected and uninfected.

The figure 8 shows the GUI running feature calculation process for the test image.

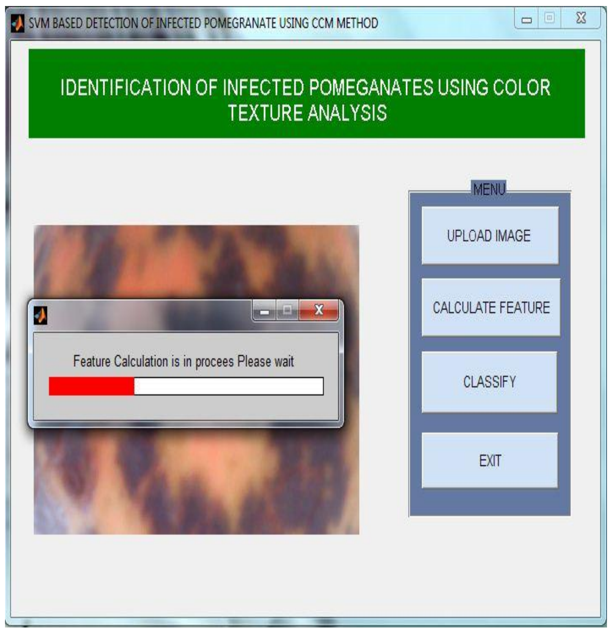

Fig 8: Graphical User Interface showing feature calculation of image under test

Figure 9 shows the result of the final classification whether the test image is infected or uninfected.

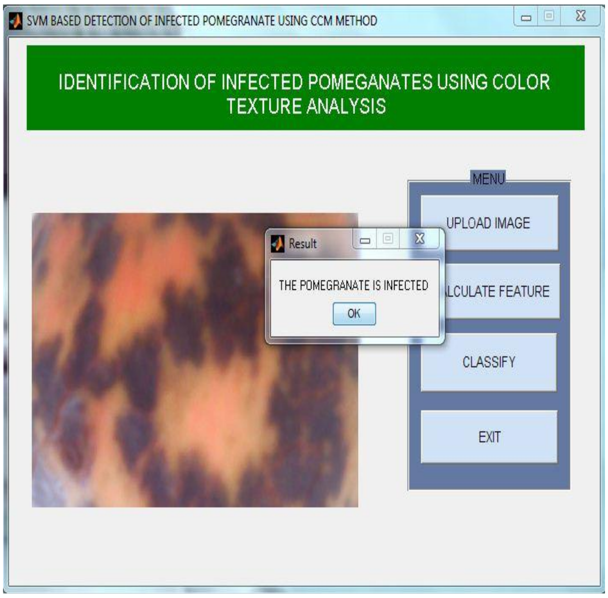

Fig 9: Graphical User Interface showing final result of classification

7. CONCLUSION

A study for early sensing of the surface infections on pomegranates using image processing was accomplished. The test image was cropped to form around 2000 samples of size 16x16. The RGB pomegranate image was then transformed into HSI image space for nullifying effects of light reflected by image. Then each pixel map was used to generate a color co-occurrence matrix, resulting in three CCM matrices, one for each of the H, S and I pixel maps.

The generated SGDM matrix was further used for calculation of the discussed texture features. The 3D plot for each feature were plotted and examined. The features Cluster shade, Mean intensity and Product Moment show the towering separation of the data points whereas the Sum Entropy, Uniformity and Inverse Difference Moment were found to give lower separation hence omitted in final process. This analysis was verified by the result of the SVM training which showed the success rate of 99% for the selected ones. The testing process involved the feature calculation for the test image succeeded by the SVM classification.

This algorithm can be implemented for automatic grading and sorting system for quality control. This can be further extended for detecting multi diseases of pomegranate using multi-class classifier.

8. REFERENCES

[1] Dae Gwan Kim, Thomas F. Burks, Jianwei Qin, Duke M. Bulanon 2009 Classification of grapefruit peels diseases using color texture feature analysis. Int. Journal of agriculture and biological engineering, vol. 2, University of Florida, Gainesville, FL 32611-0570, USA.

[2] Edwards J G, Sweet C H. Citrus blight assessment using a Microcomputer: quantifying damage using an apple computer to solve reflectance spectra of entire trees. Florida scientist, 1986; 49 (1): 48－53.

[3] Hetal N. Patel, Dr. R.K.Jain, Dr. M.V.Joshi 2011 Fruit Detection using Improved Multiple Features based Algorithm. IJCA vol. 13, A.D.Patel Institute of Technology, New V.V.Nagar, Gujarat, India.

[4] S.Arivazhagan, R.Newlin Shebiah, S.Selva Nidhyanandhan, L.Ganesan 2010 Fruit Recognition using Color and Texture Features. JETCIS, vol. 1 no. 2, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.

[5] R. Pydipati, T.F. Burks, W.S. Lee 2006 Identification of citrus disease using color texture features and discriminant analysis. Science Direct, University of Florida, 225 Frazier-Rogers Hall, Gainesville, United States.

[6] Scott A. Shearer 1986 Plant identification using color co-occurrence matrices derived from digitized image. Trans ASAE, 1990; 33(6):2037-2044, the Ohio University.

[7] Yousef Al Ohali 2011Computer vision based date fruit grading system: Design and Implementation. Journal of King Saud University – Computer and Information Sciences 23, 29–36, King Saud University, Riyadh, Saudi Arabia.

[8] Czeslaw Puchalski, Jozef Gorzelany, Grzegorz Zagula, Gerald Brusewitz 2008 Image analysis for apple defect detection. TEKA Kom. Mot. Energ. Roln. – OL PAN, 8, 197–205, Oklahoma State University, Stillwater, USA.

[9] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh 2011 Fast and Accurate Detection and Classification of Plant Diseases. International Journal of Computer Applications (0975 – 8887), Vol. 17, no.1.

[10] Lanlan Wu and Youxian Wen 2009 Weed/corn seedling recognition by support vector machine using texture features. African Journal of Agricultural Research vol. 4 no. 9, Huazhong Agricultural University, Wuhan, 430070, P. R. China.

[11] Li Li, Chen Yingyi, Gao Hongju and Li Daoliang 2012 Automatic recognition of village in remote sensing images by support vector machine using color co-occurrence matrices. American Scientific Publishers vol. 10, Numbers 1-2, pp. 523-528.

[12] Xing-yuan Wang, Zhi-feng Chen and Jiao-jiao Yun 2012 An effective method for color image retrieval based texture. Journal of computer standards and interface vol. 34 Issue 1, January, 2012.

[13] Ryusuke Nosaka, Yasuhiro Ohkawa and Kazuhiro Fukui 2012 Feature Extraction Based on Co-occurrence of Adjacent Local Binary Patterns. Advances in image and video technology vol. 7088/2012, 82-91.

[14] A. Camargo, J.S. Smith. Image pattern classification for the identification of disease causing agents in plants. Computers and Electronics in Agriculture , Volume 66, Issue 2, May 2009, Pages 121–125

[15] Qinghua Guo, Maggi Kelly, Catherine H. Graham. Support vector machines for predicting distribution of Sudden Oak Death in California. Ecological Modelling , Volume 182, Issue 1, 25 February 2005, Pages 75–90