I J M I
International Journal of Machine Intelligence
ISSN: 0975–2927 & E-ISSN: 0975–9166, Volume 3, Issue 2, 2011, pp-36-44
Available online at http://www.bioinfo.in/contents.php?id=31

A HYBRID INTELLIGENT SYSTEM FOR AUTOMATED POMEGRANATE DISEASE DETECTION AND GRADING

SANNAKKI S.S.[1*], RAJPUROHIT V.S.[2], NARGUND V.B.[3], ARUN KUMAR R.[4] AND YALLUR P.S.[5]

[1,2,4,5] Department of Computer Science and Engineering, Gogte Institute of Technology, Belgaum, India
[3]Department of Plant Pathology, University of Agricultural Sciences, Dharwad, India
Corresponding author. Email: sannakkisanjeev@yahoo.co.in, vijaysr2k@yahoo.com[2], nargund56@gmail.com[3]

Received: August 10, 2011; Accepted: August 17, 2011

Abstract - This paper proposes an image processing methodology to address one of the core issues of plant pathology i.e. disease identification and its grading. The proposed system is an efficient module that identifies various diseases of pomegranate plant and also determines the stage in which the disease is. The system employs various image processing and machine learning techniques. At first, the captured images are processed for enhancement. Then image segmentation is carried out to get target regions (disease spots). Later, image features such as shape, color and texture are extracted for the disease spots. These resultant features are then given as input to disease classifier to appropriately identify and grade the diseases. Finally, based on the stage of the disease, the treatment advisory module can be prepared by seeking agricultural experts, there by helping the farmers.

Keywords- Image processing, soft computing, machine learning, disease classification, disease grading.

INTRODUCTION

Sole area that serves the food needs of the entire human race is the Agriculture sector. Research in agriculture is aimed towards increase of productivity and food quality at reduced expenditure and with increased profit [1]. In the past few years new trends have emerged in the agricultural sector. Due to the manifestation and developments in the fields of sensor networks, robotics, GPS technology, communication systems etc, *precision agriculture* started emerging [2]. Precision agriculture concentrates on providing the means for observing, assessing and controlling agricultural practices. It also takes into account the pre- and post-production aspects of agricultural enterprises. The objectives of precision agriculture are profit maximization, agricultural input rationalization and environmental damage reduction, by adjusting the agricultural practices to the site demands. The challenge of the precision approach is to equip the farmer with adequate and affordable information and control technology.

Plant disease is one of the crucial causes that reduces quantity and degrades quality of the agricultural products. Disease is impairment to the normal state of the plant that modifies or interrupts its vital functions such as photosynthesis, transpiration, pollination, fertilization, germination

etc. The emergence of plant diseases has become more common now a days, as factors such as climate and environmental conditions are more unsettled than ever [2]. Plant diseases are usually caused by fungi, bacteria and viruses. Also there are other diseases which are caused by adverse environmental conditions. There are numerous characteristics and behaviors of such plant diseases in which many of them are merely distinguishable. The ability of disease diagnosis in earlier stage is an important task. Hence an intelligent decision support system for Prevention and Control of plant diseases is needed. This system uses some high-tech and practical technology, such as fuzzy logic, neural networks, support vector machines and such other soft computing techniques to appropriately detect and diagnose the plant diseases.

Technological advancement is gradually finding its applications in the field of agriculture [3]. The information and communication technology (ICT) application is going to be implemented as a solution in improving the status of the agricultural sector [4]. The idea of integrating ICT with agriculture sector motivates the development of an automated system for pomegranate disease classification and its grading.

Pomegranate *(Punica granatum),*so called *"fruit of paradise"* is one of the major fruit crops of arid region. It is Popular in Eastern as well as Western parts of the world. The fruit is grown for its attractive, juicy, sweet-acidic and fully luscious grains called 'Arils' [6]. The fruits are mainly used for dessert purposes. In India it is cultivated over the area of about 63,000 ha, and its production is about 5 lakh tons/annum. Important varieties cultivated are Ganesh, Dholka, Seedless(Bedana), Bhagwa, Araktha. Fig. (1) shows three varieties of pomegranate fruit.

(a)Bhagva (b)ganesh (C)Araktha

Fig1- Varieties of pomegranate

Based on size and color, pomegranate fruits are graded as follows:

Super size- in which fruits are free from spots and individual fruit weight is more than750grams.

King size in- which fruits are attractive and individual fruit weight is 500-700 grams.

Queen size- in which fruits are attractive, red and individual fruit weight is 400-500 grams.

Prince size- in which fruits are attractive, red and individual fruit weight is 300-400 grams.

Unfortunately there are no organized marketing systems for pomegranate. Usually farmers dispose their produce to contractors who will later transport to far off markets. There is scope for exporting Indian pomegranates to Bangladesh, Bahrain, Canada, Germany, United Kingdom, Japan, Kuwait, Sri Lanka, Omen, Pakistan, Qatar, Saudi Arabia, Singapore, Switzerland, U.A.E. and U.S.A.

Diseases and insect pests are the major problems that threaten pomegranate cultivation. These require careful diagnosis and timely handling to protect the crops from heavy losses [6]. In pomegranate plant, diseases can be found in various parts such as fruit, stem and leaves. Major diseases that affect pomegranate fruit are bacterial blight (*Xanthomonas axonopodis pv punicae*), antracnose (*Colletotrichum gloeosporoides)* and wilt complex (*ceratocystis fimbriata*). Image samples of these diseases are shown in Fig. (2).

(a)Wilt complex

(b)Anthracnose

(c)Scab

(d)Bacterial blight on fruit

(e)Bacterial blight on leaf (f)Bacterial blight on stem

Fig 2- Various diseases affecting pomegranate

Bacterial blight is the most severe disease of the pomegranate. The disease symptoms can be initially found on stem part which gradually pervades to leaves and then to fruits. On stem, the disease starts as brown to black spot around the nodes. In advance stages of nodal infection girdling and cracking of nodes lead to break down of branches. On leaves, the disease starts with small, irregular, water soaked spots that are 2 to 5 mm in size with necrotic centre of pin head size. Spots are translucent against light. Later, these spots turn light to dark brown and are surrounded by prominent water soaked margins. Numerous spots may coalesce to form bigger patches. Severely infected leaves may drop off. On fruits brown to black spots appear on pericap with L,Y and star shaped cracks. The disease spreads as the bacterium survives on the tree as well as in diseased fallen leaves. High temperature and high relative humidity favors the disease. The disease spreads to healthy plants through wind splashed rains and in new area through infected cuttings.

RELATED WORK

Image processing a form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or, a set of characteristics or parameters related to the image.

Digital image processing is the use of computer algorithms to perform image processing on digital images [5]. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. The general approach of image processing is shown in the Fig. (3).

Image Capture Image Enhance & analysis

Fig 3- General approach for image processing

In agricultural sector, numerous computerized tools and sensors like temperature sensors, humidity sensors, leaf wetness sensors etc have been developed to help farmers to monitor the proper growth of their crops. But plant diseases and weeds pose a great threat in degrading quality and quantity of the agricultural products. Hence, special attention has been geared up towards the detection of early stages of the diseases. In the area of disease control, most research has been focused on the treatment and control of weeds, and few studies have been focused on the automatic identification of diseases of various plants as described below.

Reference [4] presents a new technique for sorting and automated grading of 4 types of fruits viz. apple, banana, orange and mango and one type of vegetable i.e. carrot. The system can be used to overcome the problems of manual grading. The system combines three processes (feature extraction, sorting and grading) without human intervention. The methodology begins with capturing the fruit's image using a regular digital camera and loading into the MATLAB environment. The captured image is then processed by applying various image processing techniques. Shape features such as area, MajorAxisLength, MinorAxisLength and perimeter are extracted. These feature values are then fed as input to Support Vector Machine (SVM) to sort the fruit based on its shape and size. Finally, with the help of Fuzzy Logic (FL), the fruit's grade is determined based on the extracted features and the determined fruit type. The results are very promising for three of the five chosen fruits. Similar shape-size

fruit misclassifications can be reduced by using additional features such as color and texture.

In reference [6] Fuzzy Logic was applied as a decision making system to grade apples. The same set of apples was graded by both a human expert and a FL system designed for grading purpose. Five quality features i.e, color, defect, shape, weight and size were measured using various apparatus. The number of apples used was determined based on the availability of apples with quality features of the 3 quality groups (bad, medium and good). A total of 181 golden delicious apples were graded first by a human expert and then by the proposed FL approach. Grading results obtained from FL showed 89% general agreement with the results from the human expert, providing good flexibility in reflecting the expert's expectations and grading standards into the results.

A hybrid intelligent system has been used in grape leaf disease detection from color imagery in reference [7]. The system consists of three main parts, a) Grape leaf color segmentation b) Grape leaf disease segmentation and c) Analysis and classification of diseases. All the divisions used various image processing techniques. A *Self Organizing Feature Map* together with a back propagation neural network was deployed to recognize colors of the grape leaf. Then, the grape leaf disease segmentation was performed using modified self organizing feature map along with genetic algorithms for optimization. Finally the resulting segmented image was filtered by *Gabor wavelet* which allows the system to analyze leaf disease color features more efficiently. SVMs are then applied to classify types of grape leaf diseases to distinguish whether it is a scab disease, rust disease or no disease. The proposed system shows desirable results which can be further developed for any agricultural product analysis/inspection system.

Detection of rice diseases early and accurately, using an application of image processing techniques and SVM has been presented in [8]. In this paper, initially Rice disease spots were segmented by the otsu method. Edges of the disease spots were important to identify the disease; hence to retain the disease spots image threshold values have been changed. Four shape features (rectangularity, compactness, elongation and roundness) and 60 texture features (contrast, uniformity, entropy, inverse difference and linearity correlation) of disease spot from 4 orientation angles (0º, 45º, 90º, 135º) and 3 spatial (H.S.V) were extracted. An accurate and automatic method based on image processing and SVM was developed to detect three rice diseases including rice bacterial leaf blight (RBLB), rice sheath blight (RSB) and rice blast (RB). The results showed that SVM could effectively detect and classify these disease spots to an accuracy of 97.2%.

In reference [9], a new method of recognizing cucumber leaf disease based on computer image processing and SVM was studied to improve recognition accuracy and efficiency. For detecting diseases of the plants early and accurately, the authors have used image processing and classification techniques. At first, diseased images are acquired using cameras or scanners and vector median filter was used to remove the noises. Secondly, disease spots were segmented from the background using statistic pattern recognition. Later color, shape and texture features were extracted. Finally, classification method, named support vector machine was used to classify diseased images. The experimental results of recognition of cucumber disease by SVM showed that correct recognition rate of both using shape feature and texture feature is higher than that of shape feature only. Cucumber disease was recognized more correct and faster based on color feature. Experimental results also indicated that the classification performance by SVM is better than that of neural networks.

Determination of Size and Ripeness of a Banana has been discussed in [10]. This paper emphasizes on the classification of bananas according to its size, shape, texture, color. Three varieties of banana considered for the classification purpose were Mas, Berangan and Cavendish. In terms of size, a banana was classified as extra large (XL), large (L), medium (M) and small (S). Along with the classification, six methodologies of edge detection were discussed, tested and compared in order to get the best method of edge detection. Canny's method was proved to be

the best one for edge detection among the 6 methodologies. The key idea is the edge detection and the color changes that help in determining the quality of the banana. The area, perimeter, length and thickness of a banana were also determined. Discussion was focused on ripeness percentage. The ripeness percentage was calculated based on the color changes such as when ripe, the fruit turns yellow. The percentage of ripeness can be determined by evaluating the individual pixels of the image.

A study in reference [2] describes an image-processing based method that identifies the visual symptoms of plant diseases from the analysis of colored images. The algorithm starts by converting the RGB image into the H, I3a and I3b color transformations. The transformed image is then segmented by analyzing the intensity distribution. Once the image is segmented, the extracted region is post-processed to remove those pixels that are not considered part of the target region. This procedure is accomplished by analyzing the neighborhood of each pixel and the gradient of change between them. To test the accuracy of the proposed algorithm, manually segmented image are compared with those segmented automatically. The strength of the algorithm is the ability to identify the correct target (diseased region) in images even with different range of intensities distribution.

METHODOLOGY

The proposed methodology aims to model disease detection/classification and a promising disease grading system for pomegranate plant. The system makes use of various image processing, soft computing and machine learning techniques. The proposed work is divided into four steps: (1) image acquisition (2) image pre-processing (3) image post-processing (4) disease classification and grading. A flowchart of the complete process is as depicted in Fig. (4).

(1) Image acquisition – Every image processing application always begins with image acquisition. The images can be captured using a regular digital camera with minimum 6 megapixels of resolution for better quality; maintaining an equal distance, equal angle and equal illumination to the object with uniform background. All the images should be saved in the same format such as JPEG, TIF, BMP, PNG etc.

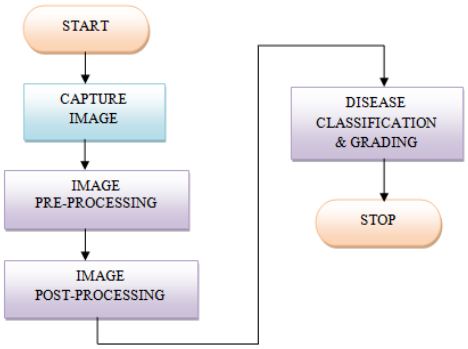

Fig 4. Flowchart for disease detection and grading in pomegranate

(2)Image pre-processing – image pre-processing can be defined as a technique in which the data from an image are digitized and various mathematical operations are applied to the data. Pre-processing creates an enhanced image that is more useful or pleasing to a human observer. Pre-processing uses various techniques like image resize, filtering, segmentation, morphological operations etc. In most of the image processing applications the initial captured images are resized to a fixed resolution to utilize the storage capacity or to reduce the computational burden. Since images may be captured from the fields it will be unavoidable that some due drops, insects' excretes and dust might appear on the captured images. In image processing these are treated as image noises. They must be removed or weakened before any further image analysis. Filters like Gaussian, median, Linear, Low pass, High pass, Laplacian filters etc can be used to remove the image noise. Once the image has been enhanced, the next process is to extract region of interest in the image i.e. diseased portion of the image. This can be achieved by image segmentation.

Image segmentation refers to the process of partitioning the digital image into multiple segments

to change the representation of an image into something that is more meaningful and easier to analyze i.e. to identify regions in the image that are likely to qualify as diseased portions. The level to which the partitioning is carried depends on the problem being solved i.e. segmentation should stop when the objects of interest in an application have been isolated [5]. There are various techniques for image segmentation such as clustering methods, compression-based methods, histogram-based methods, region growing methods etc.

(3) Image post-processing – once the image has been enhanced and segmented, there may be some stabs, empty holes etc remained in the images. Hence to remove these noises morphological operations, region filling can be applied. Further, the interested part can be extracted and its features can be analyzed. For an image, a feature can be defined as the "interest" part in an image. The name feature is often used in the pattern recognition literature to denote a descriptor. The desirable property for a feature detector is *repeatability*; i.e., whether or not the same feature will be detected in different images. Features play a fundamental role in classification. In image processing, image features usually include color, shape and texture features.

(4) Disease classification and grading – Once the features are extracted, next step is to find to which disease class the query image belongs to and in which stage the disease is. Once the disease stage is identified, appropriate treatment advisory can be provided by seeking the help from agricultural experts so that the disease can be prevented from further spreading. For this purpose we can make use of the different machine learning techniques.

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. Various approaches to machine learning are described in the next section. Selection of a particular technique depends on the problem being solved.

MACHINE LEARNING APPROACHES

Decision tree learning - uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value.

Association rule learning - Association rule learning is a method for discovering interesting relations between variables in large databases.

Artificial neural networks - An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

Genetic programming - Genetic programming (GP) is an evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that perform a user-defined task. It is a specialization of genetic algorithms (GA) where each individual is a computer program. It is a machine learning technique used to optimize a population of computer programs according to a fitness landscape determined by a program's ability to perform a given computational task.

Inductive logic programming - Inductive logic programming (ILP) is an approach to rule learning using logic programming as a uniform representation for examples, background knowledge, and hypotheses. Given an encoding of the known background knowledge and a set of examples represented as a logical database of facts, an ILP system will derive a hypothesized logic program which entails all the positive and none of the negative examples.

Clustering - Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. Clustering is a

method of unsupervised learning, and a common technique for statistical data analysis.

Bayesian networks - A Bayesian network, belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph (DAG).

Reinforcement learning - Reinforcement learning is concerned with how an agent ought to take actions in an environment so as to maximize some notion of long-term reward. Reinforcement learning algorithms attempt to find a policy that maps states of the world to the actions the agent ought to take in those states. Reinforcement learning differs from the supervised learning problem in that correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected.

Representation learning - Representation learning algorithms often attempt to preserve the information in their input but transform it in a way that makes it useful, often as a pre-processing step before performing classification or predictions, allowing to reconstruct the inputs coming from the unknown data generating distribution, while not being necessarily faithful for input configurations that are implausible under that distribution.

In this paper, SVM is recommended for Disease Identification and Fuzzy Logic is recommended for Disease Grading.

SUPPORT VECTOR MACHINES

The foundations of Support Vector Machines have been developed by Vapnik and are gaining popularity due to many attractive features, and promising empirical performance. SVMs are currently a hot topic in the machine learning community, creating a similar enthusiasm at the moment as Artificial Neural Networks used to do before. Far being, SVMs yet represent a powerful technique for general (nonlinear) classification, regression and outlier detection with an intuitive model representation. Support vector machines are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

The original SVM, can be characterized as a supervised learning algorithm capable of solving linear and non-linear classification problems. In comparison to neural networks, SVM is described as a feed-forward neural network with one hidden layer (Fig.5). The main building blocks of SVM's are structural risk minimisation originating from statistical learning theory, non-linear optimisation, duality and kernel induced feature spaces. Meanwhile, several extensions to the basic SVM have been introduced, e.g. for multi-class classification as well as regression and clustering problems, making the technique broadly applicable in the data mining area.

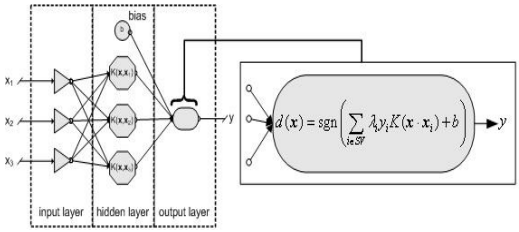

Fig 5. Architecture of SVM classifier

In the parlance of SVM literature, a predictor variable is called an *attribute*, and a transformed attribute that is used to define the hyperplane is called a *feature*. The task of choosing the most suitable representation is known as *feature selection*. A set of features that describes one case (i.e., a row of predictor values) is called a *vector*. So the goal of SVM modelling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other side of the plane. The vectors near the hyperplane are the *support vectors*. Figure 6 presents an overview of the SVM process.

FUZZY LOGIC

FL, which was first introduced by Lotfi Zadeh (1965), is used to handle uncertainty, ambiguity and vagueness. It provides a means of translating qualitative and imprecise information into quantitative (linguistic) terms [7]. Fuzzy set theory is an extension of conventional set theory that deals

with the concept of partial truth. Fuzzy logic aims to model the vagueness and ambiguity in complex systems. In many image processing applications, expert knowledge must be used for applications such as object recognition and scene analysis. Fuzzy set theory and fuzzy logic provide powerful tools to represent and process human knowledge in the form of fuzzy IF-THEN rules.

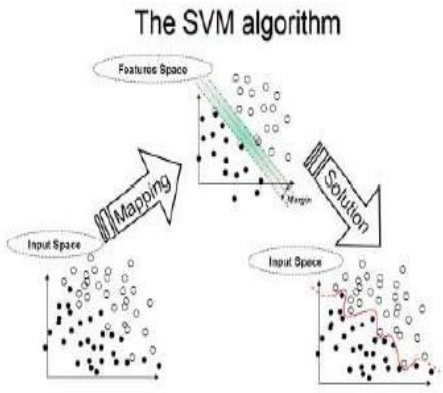

Fig 6. SVM algorithm

Over the past few decades, *fuzzy logic* has been used in a wide range of problem domains. The areas of applications are very wide: process control, management and decision making, operations research, pattern recognition and classification. In the lack of precise mathematical model which will describe behavior of the system, Fuzzy Logic is a good "weapon" to solve the problem: it allows using *logic if-then rules* to describe the system's behavior. Key reasons for using Fuzzy Logic are:

- Fuzzy logic is conceptually easy to understand
- It is flexible
- It is tolerant of imprecise data
- It can be blended with conventional control techniques
- It is based on natural language

The concepts of set theory are of powerful in mathematics. However, the principal notion underlying set theory that, an element can (exclusively) either belong to set or not belong to a set, makes it highly impossible to represent much of human discourse. Ordinary set-theoretic representations will require the maintenance of a crisp differentiation in a very artificial manner (Ex: high, not quite high, very high etc). Many decision making and problem solving tasks are too complex to be understood quantitatively. However, people succeed by using their own knowledge that is imprecise rather than precise. Fuzzy set theory resembles human reasoning in its use of approximate information and uncertainty to generate decisions. It was specifically designed to mathematically represent uncertainty and vagueness and provide formalized tools for dealing with the imprecision intrinsic to many problems. Since knowledge can be expressed in a more natural way by using fuzzy sets, many engineering and decision problems can be greatly simplified. Boolean logic uses sharp distinctions. It forces us to draw lines between members of a class and non-members.

The term fuzzy logic is used in two senses:

Narrow sense - fuzzy logic is a branch of fuzzy set theory, which deals (as logical systems do) with the representation and inference from knowledge. Fuzzy logic, unlike other logical systems, deals with imprecise or uncertain knowledge. In this narrow and perhaps correct sense, fuzzy logic is just one of the branches of fuzzy set theory.

Broad sense - fuzzy logic synonymously with fuzzy set theory.

CONCLUSION

In this paper, an automated system for the disease detection and grading in pomegranate plant is proposed. The system encompasses various image processing and soft computing techniques. The methodology begins with image acquisition. Captured images are enhanced and segmented with appropriate algorithms. Further, feature extraction is carried out and selected features are used as input to the disease classifier which appropriately identifies and grades the disease. Once the disease and its stage are identified accurately, a proper disease treatment advisory can be provided. This ultimately supports farmers during their daily struggle against disease outbreaks.

ACKNOWLEDGMENT

We thank Visvesvaraya Technological University, Belgaum, India for funding and providing a platform to the proposed Research work.

REFERENCES

[1] Santanu Phadik ar and Jaya Sil (2008) *Proceedings of 11th International Conference on Computer and Information Technology* (ICCIT 2008)

[2] Camargo A., Smith J.S. (2009) *Biosystems Engineering* I02, 9-21

[3] Narendra V.G., Hareesh K.S. (2010) *International Journal of Computer Applications*, 2, 1.

[4] Nur Badariah Ahmad Mustafa, Syed Khaleel Ahmed, Zaipatimah Ali, Wong Bing Yit, Aidil Azwin Zainul Abidin, Zainul Abidin Md Sharrif (2009) *IEEE International Conference on Signal and Image Processing Applications*.

[5] Rafael C.Gonzalez, Richard E.Woods, Steven L.Eddins, *Pearson Education*, 2nd Edition

[6] Üsmail Kavdir,Daniel E. Guyer (2003) *Turk J Agric*, 27, 375-382.

[7] Meunkaewjinda A., Kumsawat P., Attakitmongcol K. and Srikaew A. (2008) *Proceeding of ECTI-CON.*

[8] Qing Yao, Zexin Guan, Yingfeng Zhou Jian Tang, Yang Hu, Baojun Yang (2009) *International conference on Engineering Computation.*

[9] Tian Youw, Li Tianlai, Niu Yan *(*2008) *Congress on Signal and image processing.*

[10]Nur Badariah Ahmad Mustafa, Nurashikin Ahmad Fuad, Syed Khaleel Ahmed, Aidil Azwin, Zainul Abidin, Zaipatimah Ali, Wong Bing Yit, and Zainul Abidin Md Sharrif (2008) 978-1-4244-2328-6/08 © IEEE.

[11] Nedeljkovic I. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34.

[12] Tian You-Wen, WANG Xiao-Juan (2009) *World Congress on Software Engineering,* 978-0-7695-3570-8/09 © IEEE.

[13] Camargo A., Smith J.S. (2009) *Computers and Electronics in Agriculture,* 66 , 121-125.

[1] Aline Baggio, "*Wireless sensor networks in precision agriculture".*

[2] Benagi V.I. (2009) "*Pomegranate – identification and management of Diseases, insect Pests and Disorders".*

NOTES