

Orchid Classification, Disease Identification And Healthiness Prediction System

K. W. V Sanjaya, H. M. S. S Vijesekara, I. M. A. C. Wickramasinghe, C. R. J. Amalraj

Abstract: Floriculture has become one of Sri Lanka's major foreign exchange ventures and it has grown substantially during the last few years. Currently, we can find three major types of growers in floriculture. They are Large Commercial Ventures, Middle Level growers and Village Level growers. Both Middle Level and Village level growers usually go for low cost cultivation with minimum advanced techniques, sticking to conventional methods. Orchid cultivation is more pleasurable and profitable than any other floriculture ventures. As the orchid cultivation is so pleasurable we can introduce another group of growers who cultivate orchid in their home gardens for making their home gardens beautiful. But the problem is that most of these growers may not have the knowledge to identify the specie of the plants as there are a number of similar looking plants which are in different species. And also they may not have the knowledge about the orchid diseases. Because of that they may not be able to get the maximum outcome from their cultivations. So the aim of our project is to address the above mentioned issues by introducing a system which can identify orchid species & diseases and predict the healthiness of the orchid plants. The only input to this system is an image of an orchid leaf and the system will provide the orchid specie name, diseases if there any, healthiness of the orchid plant and suggestions to overcome the issues associated with the orchid plant as the output. We identify the orchid species and diseases by extracting the features of orchid plant leaf in the input image using image processing technics and with the use of data mining technics we predict the healthiness of the orchid plant. So, this system will be a great help for the people who love to grow orchids but don't have knowledge about the orchid species and diseases. And also they will be able to find the healthiness of their orchid plants.

I. INTRODUCTION

In 1970, Floriculture started as an industry in Sri Lanka. But cultivation of flowers has existed in Sri Lanka for ages because flowers were used for various religious and cultural functions. Floriculture has currently become one of Sri Lanka's major foreign exchange ventures as it has grown substantially during the previous few years. Floriculture industry in Sri Lanka contains of three categories of producers or growers. They are large commercial ventures for export (cultivate under greenhouse conditions, poly tunnels or netting), middle level growers who are catering to the local market (cultivate under natural shading or use locally available materials such as coir fiber mats, dried and woven coconut palm leaves or ropes to provide shade requirements of a particular crop) and village level growers who may sell their products to either of two categories mentioned above. The orchid family is one of the largest flowering plant families in the world with 21,950 – 26,049 currently accepted species found in 880 genera. The orchid has an incredible range of variety in shape, color and size and valued for cut flower production and as potted flowers. Orchid cultivation is more pleasurable and profitable than any other floriculture ventures. There are two groups of people who grow orchids. One is growing orchids for pleasure of making their home gardens beautiful and the other is growing orchids for commercial use, mainly for exportation. To grow orchids, it is really necessary to identify the specie of the orchid. Because various orchid species anticipate different growing conditions such as light level, fertilizer, humidity, water frequency, temperature and winter dormancy.

Like any other crop cultivations, orchid cultivation also having number of diseases which are caused by virus, fungi, bacteria, insects & pests. Common diseases that can be found in each orchid species are fungal and bacterial diseases. And if we consider virus diseases, same virus can produce more than one disease in different orchid species. The most common fungal diseases that you can find in orchid plants are sooty mold, rust, metal blight and black rot. The main cause for these fungal attacks is high humidity without sufficient air movements. And insects and pests also spread the fungus. And the bacterial attacks like brown rot are caused due to high cold and/or high humidity. These bacteria can also be carried by insects or pests crawling from an infected orchid plant to another orchid plant. If you found that one of orchid plant has a virus in your cultivation you should immediately remove it from the cultivation and burn it because viruses are spread by infected tissues of a plant or by pests. Viruses are not usually fatal, but it is very much incurable and it will make the orchid plant weaker and weaker over time.

II. PROBLEM IN BRIEF

Most of the growers from both commercial growers and the growers, who grow orchids for pleasure, do not have the exact knowledge in orchid cultivation. And knowledge in different orchid species and the orchid diseases is really important when growing orchids. If someone grows orchids without knowledge in above mentioned areas, then they won't be able to get the maximum outcome of the cultivation. This will mainly affect for the commercial producers. And different species of orchid expect different growing conditions such as lighting level, temperature, humidity, fertilizer, water frequency and winter dormancy. Therefore, In order to provide the appropriate growing conditions, it is necessary to know the correct specie of the orchid plant. There are various ways of classify the specie of the orchid. The easiest way is to identify the plant by its flower. So someone can look at the size, shape, pattern and color of the bloom. But, if the orchid is not in bloom how can someone find the orchid specie. Then it will be really difficult to identify the specie of the orchid. But you can narrow it down by looking at the characteristic of leaves

- *K.W.V Sanjaya, H.M.S.S Vijesekara*
- *I.M.A.C Wickramasinghe., C.R.J. Amalraj.*
- *Faculty of Information Technology, University of Moratuwa, Katubedda, Sri Lanka.*

such as width, height & thickness and the other characteristics of the plant such as vegetative characteristics (monopodial or sympodial).

III. RELATED WORK

There are several similar approaches for classifying crops and identifying crop diseases, but following approaches are very much similar to our approach.

- *Leafsnap: A Computer Vision System for Automatic Plant Species Identification*

This is a mobile application for identifying plant species using automatic visual recognition. And it helps users to identify tree species from photographs of their leaves. Leafsnap can identify 184 trees in the Northeastern United States. This application shows how computer vision can be used to significantly simplify the plant species identification problem. Their system requires a single leaf specimen is photographed on a solid light colored background. First, they classify whether the image is a valid, to decide whether it's worth the processing further, with the aid of a binary leaf/non-leaf classifier. But, they have to inform the user on how to take a right image. Then they segment the image centered on colors. It means they segment the images by estimating foreground and background color distributions and using these two independently classify each pixel. And also they are saying that their color-based segmentation has several advantages compared to other approaches. According to them segmenting leaf images by the shape of the leaf is highly complex because some species of leaves are compound (consisting of small leaflets) and some others found grouped into clusters. So it is difficult to use edge-based methods or region-based methods that bias towards compact shapes. Then they are doing the initial segmentation via Expectation-Maximization. They have experimented with different color spaces and noted that both the saturation and value of the HSV space were consistently useful to distinguish leaf pixels from the background. After doing the initial segmentation, they are removing the Stem of the leaf because they need to standardize the shape of the image. And they are using multiscale curvature measures in order to represent the shape of the leaf effectively. And then they are computing histograms of curvature over scale. And finally species are identified using the Nearest Neighbor search using the features extracted from the input image as a query.

- *Plant species identification using digital morphometric*

In this project they have used two-dimensional outline shape of a leaf or petal for extracting features using their leaf analyzing method. And they are studying the structure of the vein network and also the characters of the leaf margin. The outline shape has received by far the most attention when applying computational techniques to botanical image processing. The first step is a segment leaf shape based on image segmentation analysis. They used Elliptic Fourier descriptors to identify the leaf shape. The leaf shape is examined in the frequency domain, rather than the spatial domain. A set number of Fourier harmonics are calculated for the outline. This is a useful method for

helping to explain shape variation is to reconstruct the shape for some "average" descriptor, and then to create reconstructions from this descriptor as it is modified along the first few principal components. Second method was Contour Signatures which used for identifying shape is a sequence of values calculated at points taken around a leaf's outline, beginning at some start point, and tracing the outline in either a clockwise or anti-clockwise direction. When it comes to identify the vein structure in leaf, there are a wide variety of methods have been applied to the extraction of the vein networks, even though debatably with limited success thus far. Some of the best vein extraction results were achieved by using a combined threshold and neural network approach. By these techniques, they have extracted image shape and data in leaf. And also it can be used to identify the deceases in leaf.

- *Detection of weeds using image processing and clustering*

They first get two different images at the same time (red and infrared). Then they have removed the background and unwanted object like stone, soil. And with the use of binarization techniques, with a gray level threshold that segments the foreground objects from the background. The segmentation step identifies single foreground objects as objects. The approach uses bi-spectral images, which allow a good separation between plants and background. Then the system will store all the information in a central database. In here, prototypes have to be defined for the analysis and a classifier can be trained with this prototype information. They used three prototypes EPPO code, BBCH code, third is an attribute which describes special cases that may occur due to the segmentation. They are using supervised classifier that uses the training data of the prototypes is used to assign classes to the objects. Clustering was used here to group plants with similar shapes. In a second step class can be set for these automatically derived clusters and prototypes can be identified. The benefit of the approach is, classes with similar features can be identified. The number of weeds, separated for each species, was counted manually from the images and compared to the results of the image classification. Additionally a manual field sampling was done using a frame to count the weed densities for each species. The images position and manual sampling points are not exactly the same, but differ up to two meters. The main advantage of this project is they have used unsupervised clustering for separation of shapes of objects using datasets and also it is also fully automatic approach the reason is they have used selected prototypes for selected weed and crop plants.

- *Early pest identification in agricultural crops using image processing techniques*

In this approach they firstly retrieve images which are included pest in leaf/crops. Especially pest size is small as such insects (dimensions are about 2mm) is a real challenge for this project. They have used Scanner to get high resolution images to overcome these problems. For eradicating background from the object they have used region based and edge based subtraction method. And

then extract features such like color, shape and shape descriptors. For segmentation pest from leaf they have used SOBEL OPERATER for edge detection. Then they have calculated the area of pest in leaf used simple formula ($\text{Percent infection} = (\text{Infected area} \div \text{total area}) \times 100$). From this result, they have calculated the total infection on leaf which in turn gives the information about intensity of pest's infection on leaves. And the last step is calculating the size of the pest in the image. For calculating pest size they have calculated are in pest (shape blob) in leaf and it gives a number of pests in the image. The problem is their approach is they have cut the pest affected leaf area and send to the image processing analyze part. For the classification they have not used any neural network or clustering pest in the image. So they can't identify what are the types of pests in leaf. This is the main drawbacks in this project.

IV. OUR APPROACH

To solve the above-discussed problem we came up with the solution of an Orchid Classification, Disease Identification and Healthiness Prediction System. And our whole system depends on an image of an orchid leaf. So, before examine the patterns the system first need to bring all the images to a certain level where the patterns are much more visible for the usage without noisy and unwanted data then we can identify diseases and extract the certain features like area features color features with the use of image processing techniques, that can be used for create the classification model with the use of data mining techniques. With this classification model system can predict the orchid species and the healthiness of the orchid plants. Finally our system provides comments and suggestions about the orchid plants based on the predictions.

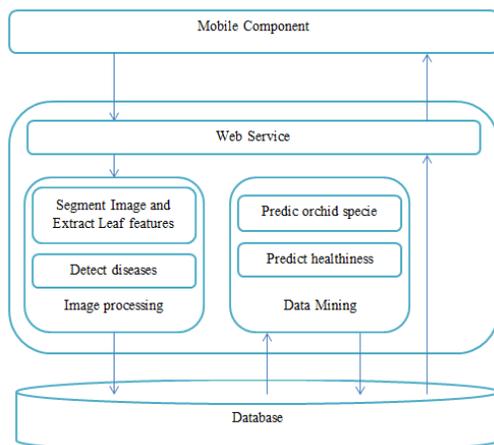


FIGURE 1: TOP LEVEL ARCHITECTURE OF THE SYSTEM

This research project discussed lot of researching in image processing and data mining for figure outing what are the most efficient and accurate ways of using the novel techniques to provide a better solution. We started our research based project with a master plan. So according to our plan we divided the whole project in to below major parts,

- Capture image using a smart phone camera and send it to a central server through a web service for further processing.
- Use image processing techniques to classify the orchid leaf in the input image and identify the orchid diseases (If there any) associated with that orchid leaf with the image by extracting the geometric and color features of the orchid leaf.
- Use data mining techniques to predict the specie of the orchid plant using extracted orchid leaf features from the image and to predict the healthiness of the orchid plant.

V. DESIGN OF THE SYSTEM

Our proposed expert system mainly consists of following components;

- A. Image processing component.
- B. Data mining component

A. Image processing component

Image processing is the main part of the design process in our system. Initially it is required to enhance the input image before applying image processing techniques to identifying the orchid leaf object in order to perform the image processing process. We have used Gaussian smoothing process in order to remove noise from the image and also we have used histogram equalization method to enhance the contrast of the image. Then we are using algorithms to segment the image, detect diseases and extract features from the enhanced image. Then we will be done with the image segmentation of the image. After extracting features and detecting diseases, data will be saved in to a database.

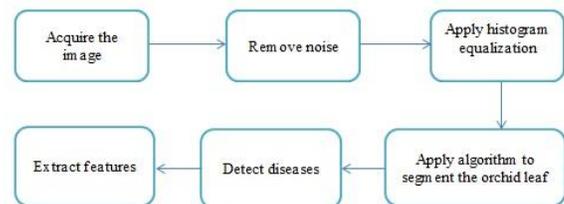


FIGURE 2: PROCESSES OF THE IMAGE PROCESSING COMPONENT

B. Data mining component

The features obtained from the image processing unit will be examined within this component. Before that, we need to develop two separate classification models for predicting species and predicting the healthiness with the data extracted from our orchid leaf images data set. Here we have used Naïve Bayes' algorithm as the data mining classification algorithm. Finally, with the use of these two data mining classification models, our system will predict the orchid specie and the healthiness of the orchid plant.

VI. IMPLEMENTATION

The implementation of the system can be described module

A. Image processing component

When it comes to the implementation of the image processing component, we selected MATLAB as our development tool because it provides almost all the image processing functionalities that we need to develop the

image processing component. Since we are using the smart phone camera for inputting images to our system high percentage of noise can be associated with the images. So we have to consider that matter before starting the implementation of the image processing component. The main component of our proposed solution is the image processing component. This component performs several operations in order to extract the essential geometric features and the color features of the orchid leaf in the input image for identifying the specie of the orchid plant and diagnose the orchid diseases associated with the orchid leaf. But, before performing those operations, we need to enhance the input image because the input image may have several noises associates with it. So that we have used noise removal techniques and image enhancement techniques such as filters and histogram equalization techniques because if we do not remove those noises we can't expect an accurate output from our feature extraction algorithms. In this project we are using the Gaussian operator and histogram equalization method for removing noise and enhancing the image. After enhancing the input image, then we are segmenting the image in order to acquire the orchid leaf object by removing the background noise in the image. Removing background of the image was not an easy task. So, what we did was we kept a white color sheet behind the orchid leaf when we are capturing the image of the orchid leaf. It made our life easy because by doing that we could easily overcome the background noise removal issue. We developed our own algorithm for extracting the features of the orchid leaf because existing algorithms did not cater our requirements. In our algorithm, the input RGB image is converted into the HSV (Hue, Saturation and Value) model which is based on the artists' concept of Tint, Shade and the Tone respectively. The reason for selecting the HSV color model was it is very much similar to the perception of the human eye. Then we applied thresholds for Hue, Saturation and Value color bands separately. We manually selected these threshold values by observing the histograms of three color bands for segmenting the green and yellow objects.

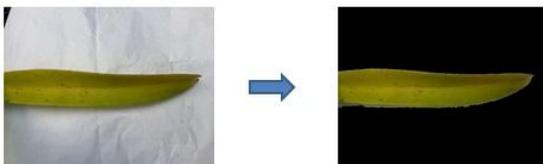


FIGURE 3: SEGMENTING THE ORCHID LEAF IMAGE

After segmenting the image then the system will extract following features from the orchid leaf object.

- Shape based features

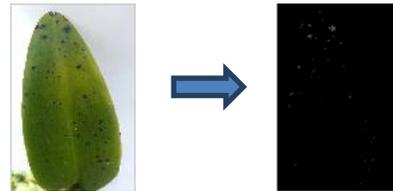
Geometric features are the features that commonly used for recognizing leaves. Slimness, Roundness and the fullness are the features that we use in our system.

- Color features

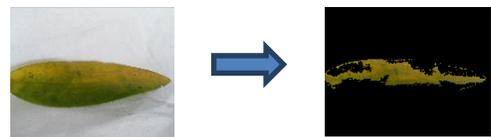
Color moments represent color features to characterize an image. Those features are mean, skewness, standard deviation, and kurtosis. For HSV color space, the three

features are extracted from each plane H, S, and V. They are mean Hue, mean Saturation and mean Value. And the other important task of the image processing component is detecting orchid diseases. For that system uses the same algorithm with different threshold values. Although there are number of orchid diseases, currently the system can detect only three diseases. They are,

- Phyllosticta



- Yellowing orchid leaves



- Black rot

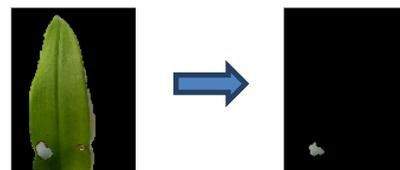


FIGURE 4: SEGMENTING THE ORCHID LEAF IMAGE

After identifying the first two diseases we are calculating the affected percentages of those diseases. And those results can be used for providing more information of the diseases. Finally, the all the results from the image processing component will be written in to a text file. And the service in the server machine will read those text files and will calculate the elongation, roundness and the fullness from the extracted features. Then all the data will be saved in the database.

B. Data Mining component

The analytical techniques used to identify patterns have a long history. In order to make predictions such as predicting the orchid specie and predicting the healthiness of the orchid plant using provided images, we can use data mining techniques such as decision trees and clustering methods. Here, in our solution we use decision trees as our prediction method. Firstly we have to develop two models for orchid species and orchid healthiness using a dataset which we extracted using the image processing module. Then we are using those models to do the predictions with the use of J48 algorithm. We have developed this component as a service. So the component will read the database and will update the records in the database with the predicted results.

VII. EVALUATION

Evaluation of the uniqueness of our system was the first thing that we had to do. So we prepared following table to

provide a better understanding of the uniqueness of our system.

	A	B	C	D	E
Reliability	High	Medium	Low	Medium	High
Speed	Low	High	Medium	Medium	Medium
User Friendliness	Medium	Medium	Low	High	High
Scale	Medium	Low	Medium	Low	Medium
Output detail	Not Clear	Very clear	Clear	Clear	Very Clear
Used method(s)	Image processing	Image processing	Image processing, Clustering	Image processing	Image Processing, Data Mining

TABLE 1: COMPARISON BETWEEN SIMILAR APPROACHES

- A- Leafsnap: A Computer Vision System for Automatic Plant Species Identification
- B- Plant species identification using digital morphometric
- C- Detection of weeds using image processing and clustering
- D- Early pest identification in agricultural crops using image processing techniques
- E- Our System

And, next we wanted to find the accuracy of the threshold based image segmentation algorithm. Therefore, in order to do that we used 250 image set. And we ended up with 212 correctly segmented images and 38 incorrectly segmented images as the result. So, we could achieve 80% plus accuracy through our segmentation algorithm. When it comes to the data mining module we found that decision trees are the most suitable method for doing predictions for a system like ours. But we found several algorithms which uses decision trees. Among them, J48 and Naive Bayes are the most commonly used algorithms for numerical data. So we decided to select the best algorithm from those two algorithms. And we compared the percentages of correctly classified instances, incorrectly classified instances, relative absolute error and root relative squared error by creating a comparison table between those two algorithms.

	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	Relative absolute error (%)	Root relative squared error (%)
J48	80.7365	19.2635	77.1962	88.1254
Naïve Bayes	79.0368	20.9632	75.8541	75.8541

TABLE 2: COMPARISON BETWEEN PREDICTION ALGORITHMS WITH ORCHID SPECIE DATA

So we used J48 as the prediction algorithm as it provides a higher accuracy for orchid specie dataset. And next we compared the same two algorithms with color features dataset. In there also J48 algorithm provided the higher accuracy.

	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	Relative absolute error (%)	Root relative squared error (%)
J48	70.7365	29.2635	15.8926	40.7533
Naïve Bayes	69.0368	30.9632	75.8541	75.8541

TABLE 3: COMPARISON BETWEEN PREDICTION ALGORITHMS WITH ORCHID HEALTHINESS DATA

And also we tested the accuracy of the prediction results over the size of the dataset. Following figures will give you a very clear idea about the results.

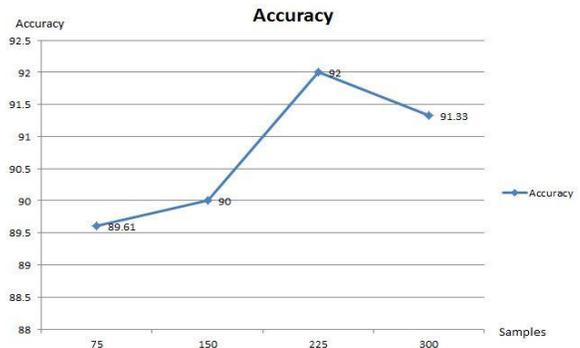


FIGURE 5: VARIATION OF THE ACCURACY WITH SIZE OF THE ORCHID SPECIE DATASET

According to the above figure, we can see an improvement of the accuracy of the prediction results when increasing the size of the dataset.

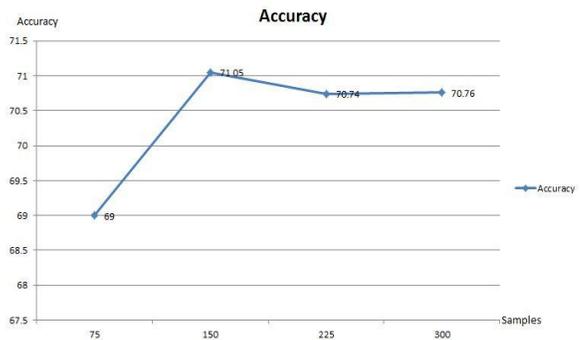


FIGURE 6: VARIATION OF THE ACCURACY WITH SIZE OF THE ORCHID COLOR FEATURES DATASET

VIII. CONCLUSION

Our system can segment the orchid leaf object from the input image with 84% accuracy. So we can extract the features of the leaf image with higher accuracy. In our solution, the extracted geometric features of the leaf image will be used to predict the orchid specie and the color features will be used to predict the healthiness of the orchid plant. According to our comparison between the classification algorithms which are using most commonly, we found that the J48 has the best accuracy than other algorithms. So we decided to use J48 algorithm for prediction purposes. And by using that algorithm we could predict the orchid specie with 91% accuracy and

healthiness of the orchid plant with 70% accuracy. We have proven that our system has above mentioned accuracies associated with the modules in our system. Therefore, we can say that the reliability of our system is high.

IX. LIMITATIONS

There are several limitations associated with our system. A main limitation is that the system can identify only three orchid species and only three orchid diseases. And when you are capturing the orchid leaf image you have to keep a white color sheet behind the orchid leaf in order to avoid capturing noisy objects in the background. You should not use any light effects when capturing images. The distance between the smart phone lens and the orchid leaf should be 10 – 15 cm. currently; the mobile application can be only used by the android smart phone users.

X. FURTHER WORK

As a further development to the system we are targeting to expand the multi-platform capability through mobile support. Also the computer vision algorithms will be accelerated with GPU support. As many mobile devices are coming with high performance GPU devices we can combine above both advancements together with mobile platform support. Mobile platform is leading technologies in modern day. So we are targeting release Android mobile platform and IOS compatibility in the near future. Also with the usage and the demand of the system we will expand the number of diseases and the number of orchid species which are to be recognized by the system in to a considerable amount. And also to attract and help local community more we are planning to enhance the local language support for the system with Sinhala and Tamil languages. And much more user friendly user interface for the mobile application will make it easy to use for the users. Therefore we hope to improve the user interface of the mobile application also.

REFERENCES

- [1] Jayamala K. Patil, Raj Kuma ADVANCES IN IMAGE PROCESSING FOR DETECTION OF PLANTDISEASES Bharti Vidyapeeth C.O.E. Kolhapur, Bhatati Vidyapeeth (Deemed Univ.) Pune Defence Institute of Advanced Tech., Deemed University, Girinagar, Pune.
- [2] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, Paulus Insap Santosa Leaf Classification Using Shape, Color, and Texture Features Department of Electrical Engineering, Gadjah Mada University Yogyakarta, Indonesia.
- [3] Maria Rossana C. de Leon, Eugene Rex L. Jalao A Prediction Model Framework for Crop Yield Prediction Southern Luzon State University, Lucban, Quezon, 4328, PHILIPPINES University of the Philippines Diliman, Quezon City, 1101, PHILIPPINES
- [4] Doraiswamy, Paul C., et al., "Operational Prediction of Crop Yields Using MODIS Data and Products." 2007.
- [5] SÖKEFELD M., GERHARDS R., OEBEL H., THERBURG R.-D. (2007): Image acquisition for weed detection and identification by digital image analysis. In J.V. Stafford, editor, Precision agriculture volume 6, pages 523–529, The Netherlands, 6th European Conference on Precision Agriculture (ECPA), Wageningen Academic Publishers. ISBN 978-90-8686-024-1
- [6] John D.W. Dearnaley Further advances in orchid mycorrhizal John D.W. Dearnaley Faculty of Sciences and Australian Centre for Sustainable Catchments, The University of Southern Queensland, Toowoomba 4350, Australia
- [7] Alexandre A. Bernardes, Jonathan G. Rogeri, Roberta B. Oliveira, Norian Marranghello and Aledir S. Pereira Identification of Foliar Diseases in Cotton Crop Universidade Estadual Paulista (UNESP) / Instituto de Biociências, Letras e Ciências Exatas (IBILCE), São José do Rio Preto, São Paulo, Brasil.
- [8] Ernesto Sanz, Noreen von Cramon-Taubadel & David L. Roberts Species differentiation of slipper orchids using color image analysis Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, C/Darwin, E-28049 Madrid, Spain.
- [9] Martin Weis, Roland Gerhards Detection of weeds using image processing and clustering: Department of Weed Science, University of Hohenheim, Otto-Sander-Straße 5, 70599 Stuttgart, Germany
- [10] C.-C. YANG, S.O. PRASHER, J.-A. LANDRY Recognition of weeds with image processing and their use with fuzzy logic for precision farming: Department of Agricultural and Bio systems Engineering and 2Department of Food Science and Agricultural Chemistry, Macdonald Campus of McGill University, Ste-Anne-de-Bellevue, QC, Canada H9X 3V9. Received 18 May 2000; accepted 1 November 200
- [11] Kamarul Hawari Ghazali, Mohd. Marzuki Mustafa and Aini Hussain Machine Vision System for Automatic Weeding Strategy using Image Processing Technique : Faculty of Engineering, University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia
- [12] Piron; Heijden; Destain, Weed detection in 3D images: Precision Agriculture; Oct2011, Vol. 12 Issue 5, p607
- [13] Leafsnap: A Computer Vision System for Automatic Plant Species Identification; Neeraj Kumar1, Peter N. Belhumeur2, Arijit Biswas3, David W. Jacobs3, W. John Kress4, Ida Lopez4, and Joao V.B. Soares3; 1. University of Washington, Seattle WA, 2. Columbia University, New York NY, 3. University of Maryland, College Park MD, 4. National Museum of Natural History, Smithsonian Institution, Washington DC