Classification of Herbs Plant Diseases via Hierarchical Dynamic Artificial Neural Network after Image Removal using Kernel Regression Framework

Lili N.A, F. Khalid, N.M. Borhan

Abstract— When herbs plants has disease, they can display a range of symptoms such as colored spots, or streaks that can occur on the leaves, stems, and seeds of the plant. These visual symptoms continuously change their color, shape and size as the disease progresses. Once the image of a target is captured digitally, a myriad of image processing algorithms can be used to extract features from it. The usefulness of each of these features will depend on the particular patterns to be highlighted in the image. A key point in the implementation of optimal classifiers is the selection of features that characterize the image. Basically, in this study, image processing and pattern classification are going to be used to implement a machine vision system that could identify and classify the visual symptoms of herb plants diseases. The image processing is divided into four stages: Image Pre-Processing to remove image noises (Fixed-Valued Impulse Noise, Random-Valued Impulse Noise and Gaussian Noise), Image Segmentation to identify regions in the image that were likely to qualify as diseased region, Image Feature Extraction and Selection to extract and select important image features and Image Classification to classify the image into different herbs diseases classes. This paper is to propose an unsupervised diseases pattern recognition and classification algorithm that is based on a modified Hierarchical Dynamic Artificial Neural Network which provides an adjustable sensitivity-specificity herbs diseases detection and classification from the analysis of noise-free colored herbs images. It is also to proposed diseases treatment algorithm that is capable to provide a suitable treatment and control for each identified herbs diseases.

Index Terms— HDNN, Bayesian algorithm, Fixed-Valued Impulse Noise, Random-Valued Impulse Noise and Gaussian noise

1 Introduction

erb is any plant that is used to alleviate unwanted symptoms or boost overall health. It can seem overwhelming when looking at all the herbs there are to choose from to grow. According to Herba Malaysia, a simple way to narrow down is to categorize the herbs based on their uses. There are Culinary (i.e, 'Lengkuas' – 'Alpinia Galanga' etc), Medical (i.e, 'Misai Kucing' – 'Orthosiphon Aristatu' etc) and Ornamental (i.e, 'Bunga Tanjung' – 'Mimusops Elengi' etc) herbs. Hence, herbs belong to many different plant families and can be susceptible to many types of diseases. There are diseases specific to some herbs, like "Fusarium Wilt" of Amaranthus sp, and there are other diseases that are common to many herbs such as "Powdery Mildew". Some common diseases that occur on many herbs include the diseases caused by Fungi ("Damping-off", "Root and Crown Rot", "Botrytis Blight" etc), Nematodes ("Foliar Nematodes" etc), Physiological/ Environmental factors ("Cold Water Injury", "Soluble Salt Injury" etc) and Insect Problems ("Aphids", "Fourlined Plant Bug", "Mites", "Spittlebugs" etc).

As discussed in [1], the rate of spread of disease depends on current crop conditions and susceptibility to infection. When herbs plants become diseased, they can display a range of symptoms such as colored spots, or streaks that can occur on the leaves, stems, and seeds of the plant. These visual symptoms continuously change their color, shape and size as the disease progresses. Therefore, there is a need for systems that can help herbs producers and farmers, particularly in remote areas, to identify early symptoms of herbs plant disease by means of analyses of digital images of crop samples. Identifying the disease is the first step to correcting sick herbs plants. Once the disease is known, the problem usually can be cured. The current electronic devices for capturing images have been developed to a point where there is little or no difference between the target and its digital counterpart. The success of machine learning for image pattern recognition also suggests applications in the area of identification of plant diseases.

Once the image of a target is captured digitally, a myriad of image processing algorithms can be used to extract features from it. The usefulness of each of these features will depend on the particular patterns to be highlighted in the image. Patterns are particular features of an image. They should be invariant to translation, rotation and scale, if potential misjudgments are to be avoided. The advantage of image classification by feature assessment is that patterns remain identical if preliminary conditions are changed. For example, pictures might be captured in conditions of different light intensity, or with different distances between camera lens and target. A key point in the implementation of optimal classifiers is the selection of features that characterize the image. Basically, in this study, image processing and pattern classification are going to be used to implement a machine vision system that could identify and classify the visual symptoms of herb plants diseases. The image processing is divided into four stages: Image Pre-Processing to remove image noises (Fixed-Valued Impulse Noise, Random-Valued Impulse Noise and Gaussian Noise), Image Segmentation to identify regions in the image that were likely to qualify as diseased region, Image Feature Extraction and Selection to extract and select important image features and Image Classification to classify the image into different herbs diseases classes.

2 Previous Work

Using image-processing for disease detection has become popular in medicine, because it is a rapid and reliable way to assess a patient's condition. This practice is particularly useful when the distance between practitioner and patient prevent direct consultations. The advantage in using computerize process is the high level of accuracy in the diagnosis and prognosis, as well as the significant reduction of costs in comparison to the traditional method of face-to-face diagnosis [2][2]. In agriculture, numerous image processing based computerized tools have been developed to help farmers to monitor the proper growth of their crops.

Special attention has been put towards the latest stages of growth, that is, when the crop is near harvesting. For example, at the time of harvesting, some computer tools are used to discriminate between plants and other objects present in the field. In the case of machines that uproot weeds, they have to discriminate between plants and weeds, whilst in the case of machines that harvest; they have to differentiate one crop from the other. Previous studies [3][4][6][7][8][9] have identified the challenges and have successfully produced systems that address them. The requirements for reduced production costs, the needs of organic agriculture and increase of diseases have been the driving forces for improving the quality and quantity of food production.

Thus, 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. Automatic plant disease identification by visual inspection can be of great benefit to those users who have little or no information about the crop they are growing. To take hold of this great benefit, the researchers [10] [11] have concentrated on detecting the visual symptoms of plant diseases from analysis of colored images by proposing adaptive image processing based method. The processing algorithm developed starts by converting the RGB image of the diseased plant or leaf into the H, 13a and 13b color transformation. Features then were extracted from each segmented region and used as inputs to a classifier, which is Support Vector Machine (SVM) classifier. The strength of this algorithm was the ability to identify the correct target (or diseased region) in images with different range of intensities distribution. Even though this algorithm works well in identifying the diseased region in leaves images with different range of intensities distribution, it still need proper modification to scale the approach to other part of plants such as stems, seeds etc. As we know the ranges of disease visual symptoms of plant leaves are different with other parts of plant. In term of multiple classes' classification that involves more than two classes; the researchers should use a number of methods to deal with this common scenario such as one-against-one method.

Another approach for detecting the visual symptoms of plant diseases was developed by Narvankar et al [12]. This research investigated the potential of soft X-ray imaging to detect fungal infection in wheat. Healthy wheat kernels and infected kernels were scanned using a soft X-ray imaging system and algorithms were developed to extract the image features and for classification. They used varieties of classifier such as statistical discriminant classifiers (linear, quadratic and Mahalanobis) and Back-Propagation Neural Network (BPNN) classifier. From their experimentation, Linear Discriminant Classifier gave better results than other statistical classifiers and neural network classifier in identifying fungal infection in wheat. Their study indicates that soft X-Ray imaging holds promise to discriminate healthy and fungal-infected wheat kernels. In their experimentation, BPNN classifier did not improve the classification accuracy, however from the results stated in their paper, only some cases this classifier gave false errors compared to statistical classifiers. Therefore, they cannot just simply conclude that this classifier is not good in classifying the infected wheat kernels images.

Other work has demonstrated the value of image processing in the detection of plant disease [13]. The researchers tried to detect and identify the type of external defects in citrus fruits including mandarins and oranges using multispectral computer vision for citrus fruits classification. The approach of segmentation of the images are considered to obtain a collection of unlabelled regions; healthy and defects skin. They found that the combination of Visible RGB Color images, Non-Visible Near-Infrared Reflectance (NIR) and Non-Visible Ultraviolet-induced Fluorescence (UVFL) color images was more effective in the detection of most defects than the same systems used separately. The Visible RGB Color Images were processed using Region-Growing algorithm and Hue, Saturation, Lightness (HSL), while both NIR and UVFL Images using run-length code. As some external defects of citrus have a similar color, it is difficult to distinguish between different types of defects only by color feature. Therefore, they proceeded by combining the color and morphological features. The morphological analysis of defects aimed to describe the shape of the defects using Fourier Analysis. Then, the defect classification is done by using Non-Linear Bayesian discriminant analysis. Overall, the system proposed successfully discriminate between 11 defects in 86% of cases. However, the algorithm developed is considered success if the order is UVFL, NIR and RGB and the processing all the images is still time consuming.

Generally, all researchers in [10][11][12][13] as well as others developed their automatic detection of plant diseases by analysis of digital images that involve four stages; image preprocessing, image segmentation, image feature extraction and selection, and image classification. The task of image preprocessing is to enhance the image and to reduce noise without destroying the important image features for diagnosis. Noise makes the visual observation and interpretation difficult. Therefore, removing noise without destroying important features for diagnosis is critical. Some noise reduction techniques only work well on additive noise and logarithmic compression are often employed to convert multiplicative noise into additive noise. In addition, Fixed-Valued Impulse Noise, Random-Valued Impulse Noise and Gaussian Noise are also capable to damage the images during acquisition and transmission processes. Up to now, most of the researchers failed to take proper care of these kinds of situation. Most of them are aimed at removing either Fixed-Valued Impulse Noise (Salt-and-Pepper Impulse Noise) or Random-Valued Impulse Noise (Uniform Impulse Noise) or Gaussian noise. As far as we know that an image can be corrupted by more than one type of noise.

In the literature, a large number of algorithms have been proposed to remove impulse noise while preserving image details. One of the most popular and robust nonlinear filter is the Standard Median Filter (SMF), which exploits the rank-order information of pixel intensities within a filtering window and replaces the center pixel with the median value. The noise removal and filtering algorithm [14] ensures that most of the Fixed-Valued Impulse Noise can be detected even at a high noise level provided that the window size is large enough. But, it increased the computation complexity especially at high density impulse noise. In [15], a two-phase method for detecting and removing Random-Valued Impulse Noise is proposed. The first phase is the detection of noise by Adaptive Center-Weighted Median Filter (ACWMF), while the second phase is the restoration of the noisy image by Variation method. The proposed algorithm can effectively remove the impulse noise with a wide range of noise density and produce better results in terms of the qualitative and quantitative measures of the images even at noise density as high as 50%. However, they have certain limitations as they are sensitive to the size and shape of the filter window. If the window size is too large, over smoothing will occur. If the window size is too small, the smoothing capability of the filter will decrease and the noise cannot be reduced effectively.

The weaknesses of noise removal algorithms developed by previous researchers are resolved in [16]. The proposed approach is based on Interval-Valued Fuzzy Sets (IVFS) entropy application. IVFS makes it possible to take into account the total uncertainty inherent to image processing, and particularly noise removal is considered. The proposed approach effectively combines image histogram information with the spatial information about pixels of different gray levels by using an IVFS multi-thresholding technique. Like other filtering techniques based on image thresholding , this proposed technique is simple and computationally efficient. The main advantage of the proposed technique is to restrict the number of thresholds or parameters which have to be tuned. Although IVFS filter was originally designed to remove impulse noises, it also effectively reduces Gaussian noise. However, for higher level of Gaussian noise the proposed filter (like most of the existing filters) does not produce really good images from a perceptual viewpoint. This is due to the fact that histogram information is flawed. They should focus on Gaussian filter (using IVFS) in their future works.

Of course the solution developed in [16], work well for removing Fixed-Valued Impulse Noise and Gaussian noise, but the situation changes dramatically if both Fixed-Valued Impulse Noise and Gaussian noise are present in the same image, a problem which commonly found in practice. It is hard to distinguish a Gaussian corrupted pixel which happens to have a high error from an impulse corrupted pixel. Then, the researchers in [17] addressed the problem by estimating the probability that a certain pixel is affected by any two kinds of noise. They Execute Iteratively Reweighted Norm (IRN) or any other method on the input image to yield predictions of the original pixel values and compute the corresponding predicted errors. Train the noise model by the Expectation-Maximization (EM) algorithm and perform a preliminary Kernel Regression to assign weights to the input pixels in the construction of the output image. The proposed algorithm can effectively remove the impulse noise with a wide range of noise density and produce better results in terms of the qualitative and quantitative measures of the images even at noise density as high as 90%.

Even previous work was capable to remove two types of image noises that can operate on the same image, but still to this point there are no commercial computer vision systems for removing three types of noises that are present in the same image. Based on this reason, exploration of more sophisticated image noise removal techniques is a promising line of study. Another crucial step to be considered after image noise removing stage, which is still an open problem in color image analysis, is that of segmentation, that is, separating the region of interest (diseased region) from the background. Previous studies document a variety of approaches for segmenting crop images. Among them, threshold-based methods are simple and promising. The threshold techniques, which have being proposed for segmenting crop images, include dynamic thresholding strategy, fixed threshold and the method based on the entropy of a histogram. In recent years, researchers have developed some more complicated but efficient methods for crop image segmentation. For example, the image segmentation that is based on the phsysic-based reflection models [18], an algorithm for separating the images containing vegetation and soil.

As an approach for feature space analysis, mean shift is a popular low-level segmentation technique for images. But until now no one has applied this promising method to crop image segmentation. Therefore, the researchers in [19] have improved the segmentation rate of the images containing green vegetation by introducing a mean-shift procedure into the segmentation algorithm. However, their method suffered from long time running, so it is not suitable for real time applications. In the future, the researchers should include choosing some other features, such as shape and texture, to represent each pixel of an image to improve the segmentation performance, and using Bayesian Neural Network (BNN) and combine with some quick mean-shift procedure rather than the basic mean-shift procedure they have adopted in their research to speed up the segmentation algorithm.

Another stage to be considered is image classification. Classification problems have been examined in fields as diverse as biology, medicine, business, image recognition, and forensics. Developing more accurate and widely classification method has significant implications in these and many other fields. The researchers in [20] introduced Dynamic Artificial Neural Network (DAN2) as an alternative for classification problems. They examined both two-class and multi-class scenarios. The method used to build the tree for the hierarchical DAN2 used with the data sets is very simple. Based on the improvement in classification performance of the simple hierarchical Dynamic Artificial Neural Network (DAN2) model used in this research, exploration of more sophisticated tree-building techniques is also a promising line of study. Additional research is needed to examine alternative hierarchical formulations that could be beneficial to DAN2.

3 Methodology And Propose Solution

The research objective as follows:

1. To propose new image noise detection and filtering algorithm that is based on a Bayesian classification of the input pixels, which is combined with the kernel regression framework that can effectively remove a wide range of Fixed-Valued Impulse Noise, Random-Valued Impulse Noise and Gaussian Noise while preserving herb image details.
2. To propose unsupervised diseases pattern recognition and classification algorithm that is based on a modified Hierarchical Dynamic Artificial Neural Network which provides an adjustable sensitivity-specificity herbs diseases detection and classification from the analysis of noise-free colored herbs images.
3. To propose diseases treatment algorithm that are capable to provide a suitable treatment and control for each identified herbs diseases.

3.1 Expected Experimentation

We will run the experiment aas follows.

a) Select synthetic herbs diseases images.

i) TWO(2) color images of sizes 512 x 512 and 768 x 512.

ii) TWO(2) grayscale images of sizes 512 x 512 and 768 x 512.

b) Test image noise removal techniques (Adaptive Median Filter (AMF) and Convolution-Based Impulse Detector and Switching Median Filter (CD-SMF)).

c) Test image segmentation techniques (Hue Intensity Saturation (HIS) model and mean-shift procedure).

d) Test image feature extraction and selection techniques (Principal Component Analysis (PCA) and Discriminant Analysis (DA)).

e) Test image classification techniques (Support Vector Machine and Artificial Neural Networks).

3.2 Data Collection

a) Synthetic Image Acquisition

- Randomly obtain 100 colors healthy and infected herbs plant images from Herba Malaysia Institute Image Database with different size (i.e. 256 x 256, 512 x 512, 768 x 512 etc) and the pixel value lie in the range [0, 255].

b) Real Image Acquisition

i) Use color real herbs plant diseases images with different size (i.e. 256 x 256, 512 x 512, 768 x 512 etc) and the pixel value lie in the range [0, 255]. Acquire the images using digital camera with low and high resolution.

- Acquire 100 healthy herbs images (leaves, roots, stems, strips etc).

- Acquire 100 infected (diseased) herbs images (leaves, roots, stems, strips etc) with known disease types.
- Acquire 100 infected herbs images (leaves, roots, stems, strips etc) with unknown disease types.

ii) Use grayscale real herbs plant diseases images with different size (i.e. 256 x 256, 512 x 512, 768 x 512 etc) and the pixel value lie in the range [0, 255]. Acquire the images using scanner.

- Acquire 100 healthy herbs images (leaves, roots, stems, strips etc).
- Acquire 100 infected herbs images (leaves, roots, stems, strips etc) with known disease types.
- Acquire 100 infected herbs images (leaves, roots, stems, strips etc) with unknown disease types.

c) For all synthetic and real images (including healthy and infected), the following sequence of events occurs :

i) First, a Gaussian noise process affects all pixels. This process will typically result from the physical limitations of the image acquisition procedure such as thermal noise, photon counting noise and film grain noise. The images should be corrupted by Gaussian noise at four different levels (standard deviation (SD)) 15, 20, 25 and 30.

ii) After that, both Random-Valued impulse noise and Salt-and-Pepper impulse noise process operates on the already Gaussian corrupted image. Note that the impulse corrupted pixels do not carry any information about the original image or the previous Gaussian noise.

- For each value of the Gaussian noise SD, the possible proportions of Random-Valued impulse noise are between 0 and 0.4, in 0.05 increments.
- Then, corrupted by 'Salt' (with value 255) and 'Pepper' (with value 0) impulse noise with equal probability with the noise levels widely vary from 10% to 90% with increments of 10%.

3.3 Data Testing

a) For image noise removal algorithm evaluation :

- Restoration performances are quantitative measure by Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE). The main concept of PSNR is to make comparison of the difference between original images and resulted images. For, PSNR, higher is better, while, MSE, lower is better.
- Comparison of performance with various well-known image noise removal algorithms including Heavily Damages Image Restoration (HDIR), Iteratively Reweighted Norm (IRN), Iterative Steering Kernel Regression (ISKR) and Progressive Switching Median Filter (PSMF).

b) For herbs diseases classification evaluation :

i) Use noise-free herb diseases images.

ii) Classification performances are quantitative measure by using multi-class model classification; 'healthy', 'infected', <known disease type>', '<unknown disease type>' and 'error'.

Comparison of performance with various well-known image classification algorithms including Support Vector machine (SVM)

4 CONCLUSION

The identification of the symptoms of herb diseases by means of a machine vision system may support herb farmers and producers during their daily struggle against disease outbreaks. The digital images of crop plants that showed visual symptoms of particular herb disease will be used as an input. After the image noise removing processes, these diseased regions will be identified and segmented with the help of to-be image segmentation algorithm. Feature is extracted from each segmented region and use as inputs to a classifier. Because not all features are supposed to give the same amount of information about the target, cross-validation will be used to identify those which comprised the best classification model.

The result of this research can be divided into three. Firstly, noise removal algorithm for removal of wide range impulse noises and Gaussian noise that operating on the same image, which produces a hard damage. Secondly, machine learning system that can be used to identify the visual symptoms of herb plant diseases and this may have a particular application for farmer or crop producers in remote locations. Thirdly, the expert system that is capable to suggest the proper treatment and control of the identified diseases.

The research is hoped to contribute a significant idea to the development of technology in modernizing the agriculture in Malaysia.

REFERENCES

[1] Ellis, S. D. and Boehm, M. J. (2008). *Plants get sick too! An introduction to plant diseases*, Agriculture and Natural Resources.

[2] Tek, F. B., Dempster, A. G. and Kale, I. (2010). *Parasite detection and identification for automated thin blood film malaria diagnosis*, Computer Vision and Image Understanding, 114(1) pp. 21-32.

[3] Dhawan, A. P. (1990). *A review on biomedical image processing and future trends*, Computer Methods and Programs in Biomedicine, 31(3-4) pp. 141-183.

[4] Bossu, J. Gée, C., Jones, G. and Truchetet, F. (2009). *Wavelet transform to discriminate between crop and weed in perspective agronomic images*, Computers and Electronics in Agriculture, 65(1), pp. 133-143.

[5] Onyango, C. M. (2003). *Segnmentation of row crop plants from weeds using color and morphology*, Computers and Electronics in Agriculture, 39 pp. 141-155.

[6] Hemming, J. and Rath. T. (2001). *Computer-vision-based weed identification under field conditions using controlled lighting*, Agricultural Engineering, 78(3) pp. 233-243.

[7] Philip, I. and Rath, T. (2002). *Improving plant discrimination in image processing by use of different color space transformations*, Computers and Electronics in Agriculture, 35(1) pp. 1-15.

[8] Chen, Y. R., Chao, K. and Moon, S. K. (2002). *Machine vision technology for agriculture applications*, Computers and Electronics in Agriculture, 36(2-3) pp. 173-191.

[9] Wang, D., Ram, M. S. and Dowell, F. E. (2002). *Classification of damaged soybean seeds using near-infrared spectroscopy*, American Society of Agricultural Engineers, 4(6) pp. 1943-1948.

[10] Camargo, A. and Smith, J. S. (2009). *An image-processing based algorithm to automatically identify plant disease visual symptoms*, Biosystems Engineering, 102(1) pp. 9-21.

[11] Camargo, A. and Smith, J. S. (2009). *Image pattern classification for the identification of disease causing agents in plants*, Computers and Electronics in Agriculture, 66(2) pp. 121-125.

[12] Narvankar, D.S., Singh, C. B., Jayas, D. S. and White, N. D. G. (2009). *Assessment of soft X-ray imaging for detection of fungal infection in wheat*, Biosystems Engineering, 103(1) pp. 49-56.

[13] Aleixos, B. N., Sanchís, J. G. and Moltó, E. (2009). *Recognition and classification of external skin damage in citrus fruits using multispectral data and morphological features*, Biosystems Engineering, 103(2), pp. 137-145.

[14] Wang, S. S. and Wu, C. H. (2009). *A new impulse detection and filtering method for removal of wide range impulse noises*, Pattern Recognition, 42(9) pp. 2194-2202.

[15] Zhang, J. (2009). *An efficient median filter based method for removing random-valued impulse noise*, Digital Signal Processing, In Press, Corrected Proof.

[16] Bigand, A. and Colot, O. (2010). *Fuzzy filter based on interval-valued fuzzy sets for image filtering*, Fuzzy Sets and Systems, 161(1), pp. 96-117.

[17] Rubio, E. L. (2010). *Restoration of images corrupted by Gaussian and uniform impulsive noise*, Pattern Recognition, 43(5) pp. 1835-1846.

[18] Pauplin, O., Caleb-Solly, P. and Smith, J. (2010). *User-centric image segmentation interactive parameter adaptation tool*, Pattern Recognition, 43(2) pp. 519-529.

[19] Zheng, L., Zhang, J. and Wang, Q. (2009). *Mean-shift-based color segmentation of images containing green vegetation*, Computers and Electronics in Agriculture, 65(1) pp. 93-98.

[20] Ghiassi, M. and Burnley, C. (2010). *Measuring effectiveness of a dynamic artificial neural network algorithm for classification problems*, Expert Systems with Applications, 37(4) pp. 3118-3128.

Lili N. A. She completed PhD from University Kebangsaan Malaysia in 2008. Currently she is Senior Lecturer at the department of Multimedia, Faculty of Computer Science and Information Technology, University Putra Malaysia. Her research interest includes video retrieval, computer game and animation.