Automatic Identification of Flower DiseasesUsing Artificial Neural Networks

Getahun Tigistu
Department of Computer Science
Arba Minch University
Arba Minch, Ethiopia
gettigistu@gmail.com

Yaregal Assabie
Department of Computer Science
Addis Ababa University
Addis Ababa, Ethiopia
yaregal.assabie@aau.edu.et

***Abstract*—The floral industry has increasingly become one of the most important sectors for export earnings, especially in developing countries. However, during the cultivation process there may be a number of challenges that affect it, one of which is flower disease. This paper presents an automatic identification of of flower dieases based on image processing techniques. In view of this, normal and diseased flower images are acquired to create a knowledge base where images are pre-processed and segmented to identify the region of interest. Texture features of images are extracted using Gabor feature extraction, from which we computed seven different measures of dispersion and central tendency with the purpose of reducing the dimensionality of features. Then, an artificial neural network is trained with seven input features extracted from individual images and eight output nodes representing eight classes of diseases considered in this work. Unkown samples of flower images are then tested based on the training model and we achieved an average accuracy of 83.3% in the identification of the flower diseases.**

Keywords—Flower Disease Identification; Artificial Neural Networks; Gabor Feature Extraction

I. INTRODUCTION

Floriculture as an industry is reported to have begun in the late 19th century in the United Kingdom. Historically, the industry has been dominated by the developed world where major producers and consumers exist. Over the past few decades, however, a paradigm shift in the floral industry has been observed when flower growers moved to the developing countries of Africa, South America and Asia due to better climatic conditions and cheaper production costs [7, 8]. In general, the floral industry has achieved significant growth worldwide where it has increasingly become a major source of revenue for some countries around the world. The industry has grown from less than $3 billion in the 1950s to more than $100 billion in 2003 in global trade volume [8]. Ethiopia and Kenya are the biggest exporters of flowers in Africa, accounting for substantial percentage of export earnings for both countries [7].

In spite of the lucrative market, the cultivation of flowers is facing a number of risks. One of the major risks is disease that reduces the quality and quantity of flower production. A plant becomes diseased when it is continuously disturbed by some causal agent that results in an abnormal physiological process that disrupts the plant's normal structure, growth, function, or other activities [10, 14]. Flowers are one of the plant species that are attacked by different types of diseases. To identify disease of a flower, it is usually necessary to look at the flower closely; examine the flower, leaves, stem and sometimes the roots; and do some detective work to determine possible causes [13]. This identification of flower disease can be done through experts in this field. However, getting experts for the identification and investigation of flower diseases is difficult (expensive) for flower growers since experts are required to continuously monitor flowers. Moreover, the problem becomes worse due to the fact that environmental conditions are not conducive for experts due to chemicals and green houses used in the cultivation process of flowers [4]. Thus, the development of a system that can automatically detect flower diseases is of paramount importance to the floral industry as it helps to overcome the shortage of experts and alerts growers so that they use the biological pest control mechanisms timely.

The general field of automatic plant disease identification and agricultural product classification has been studied for decades. Most research and development works have focused on plant leaf disease identification [1, 2, 15], fruit/food grading [6, 16] and weed detection [4, 13]. Although the subject of identifying plant diseases using digital image processing has been studied for at least 30 years, the advances achieved so far have not been proportional to the attention it received from researchers [3]. Previous works have been generally characterized by too specific methods and strict operation conditions. Many of the methods proposed deal not only with limited species of plant, but also at a specific growth stage in order for the algorithm to be effective. Furthermore, images are captured under certain very strict conditions like lighting, angle of capture, distant between objects and capture device [1, 3].Due to such limitations with respect to the applicability of the previously proposed methods, it is difficult to directly adopt the technology to the floral industry. To our best knowledge, there is no system that automatically detects flower diseases and growers are still employing manual processes to detect and identify flower diseases, leading to a high cost of production. In this work, we propose a generic system that automatically identifies flower diseases using artificial neural networks providing a prediction mechanism for unknown samples of flowers as well.

The remaining part of this paper is organized as follows. The types of common flower diseases are discussed in Section II. In Section III, we present the architecture of proposed system. Experimental results are reported in Section IV. Section V presents the conclusion and discusses future works.

II. FLOWER DISEASES

As diseases of a plant are inevitable, detecting disease plays a major role in the industry. Flower diseases are caused by fungi, bacteria and pests. Many flower diseases produce symptoms which are the main indicators in the diagnosis. Some of the causes of flower diseases are powdery mildew, aphid, Japanese beetles, rosette, goldenrod-soldier, crown gall, rust, soldier beetles, gray mold, botrytis blight and black spot [10]. Some of the diseases took their names from the agents that cause the diseases [14]. Figure 1 shows normal and diseased rose flowers along with the pests that cause diseases.

Fig. 1. Samples of normal and diseased rose flowers along with pests.

III. THE PROPOSED SYSTEM

In this work, identification of flower diseases is achieved by making use of characteristic features of flowers. The features of diseased and normal flowers are extracted to train the system. A knowledge base is created as a result of the training process. The knowledge base is later used during the indentification of unknown samples of flower diseases. Figure 2 shows the architecture of the proposed system.

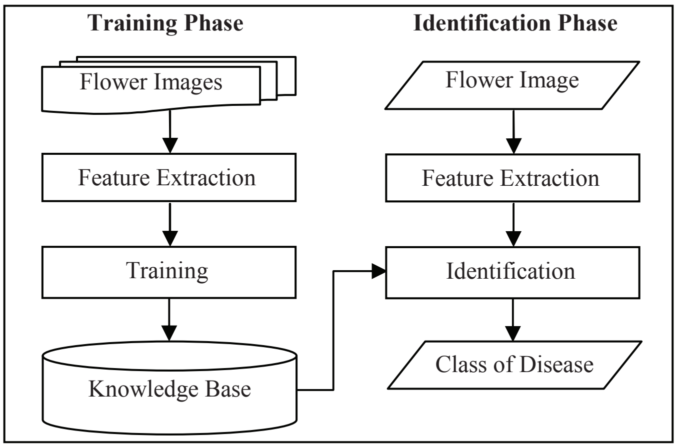

Fig. 2. Architecture of the proposed system.

A. Feature Extraction

Images have different characteristics like shape, texture and color features. In this work, texture features are used to represent images for the identification of the normal and diseased flowers. We have used texture features because of their role in differentiating various types of flower diseases. Although flowers are known to change colors and shapes due to diseases, we did not use color and shape features because there are diseases that change the color of one flower into a similar color of another flower. Furthermore, flowers at different growing stages how different colors and shapes. Thus, color and shape features may be used if the system is to be applied only for one type of flower. To select regions of interest, we applied noise removal and background subtraction pre-processing techniques followed by an extended version of Otsu's image segmentation method [11]. Figure 3 shows the result of flower image segmentation.

Fig. 3. Segmentation of flower image using Otsu's method.

Texture features from flower images are extracted using a family of two-dimensional Gabor filter defined as follows [9].

$$\begin{cases} W_{(x,y,\theta,\lambda,\varphi,\sigma,\gamma)} = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \\ x' = xcos(\theta) + ysin(\theta) \\ y' = -xsin(\theta) + ycos(\theta) \end{cases} \quad (1)$$

where (x, y) specifies the position of a light impulse in the visual field and μ, φ, γ, λ, and σ are parameters of the filter whose values are given in TABLE I.

TABLE I. PARAMETERS USED FOR GABOR FEATURE EXTRACTION

Parameter	Symbol	Values
Orientation	θ	$\left\{0,\frac{\pi}{8},\frac{2\pi}{8},\frac{3\pi}{8},\frac{4\pi}{8},\frac{5\pi}{8},\frac{6\pi}{8},\frac{7\pi}{8}\right\}$
Wavelength	λ	$\{4,4\sqrt{2},8,8\sqrt{2},16\}$
Phase	φ	$\left\{0,\frac{\pi}{2}\right\}$
Gaussian radius	σ	$\sigma=\lambda$
Aspect ratio	γ	1

Since we have used five spatial frequencies and eight distinct orientations, we obtain a total of forty distinct responses for a single filtering operation. This would be computationally complex when we apply the filter for the entire image. Thus, in order to minimize the computational complexity, we reduce the dimensionalities of the resultant feature vectors representing an image using a descriptive statistical data representation technique. Based on this statistical data representation technique, feature vectors are changed into seven different measures of dispersion and central tendency. For data elements $\boldsymbol{x}_i$ (i=1, 2, 3, …, n; where n is the total number of data elements), the selected measures of central tendency and dispersions are calculated as follows.

Mean: The arithmetic mean $\bar{x}$ of a data set is calculated as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} xi \quad (2)$$

Square of Deviations: The square of deviation $DEVSQ$ is computed as

$$DEVSQ(x_1 \dots x_n) = \sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (3)$$

Absolute deviation: The absolute deviation $Adev$ is computed as

$$Adev(x_1 \dots x_n) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \quad (4)$$

Variance: The variance Var is computed as

$$Var(x_1 \dots x_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (5)$$

Standard Deviation: the standard deviation Sdv is calculated as the square root of variance.

$$Sdv(x_1 \dots . x_n) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (6)$$

Skewness: characterizes the degree of asymmetry of a distribution around its mean.

$$Skew(x_1 \dots x_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^3 \quad (7)$$

Kurtosis: represents non-dimensional quantities that are used to measure the relative peakedness or flatness of a distribution.

$$Kurt(x_1 \dots x_n) = \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^4\right\} - 3 \quad (8)$$

By computing these measures of central tendency and dispersions, we hugely reduce the dimensionality of feature vectors while maintaining most relevant features used for the classification purpose.

B. *Training*

To train the system with diseased flower images, first diseased flowers images were labeled with the names of the respective disease categories. Normal flower images were also labeled as normal. Then, each labeled image is pre-processed and texture features are extracted from the image as described in the previous section. Since we have a labeled data, we use a supervised machine learning method to identify the class of the image. In this work, an artificial neural network with feed forward multilayer perceptron architecture is used for training because of its favorable properties that make it an excellent choice for object classification. The most important of these properties are generalization, expandability, representing multiple samples, and memory saving.

In feed forward multi-layer perceptron architecture, the neural networks have distinct input, output and hidden layers where the output from one layer of neurons feed forward into the next layer of neurons. There are never any backward connections, and connections never skip a layer [38, 39]. Typically, the layers are fully connected, meaning that all units at one layer are connected with all units at the next layer. A feed forward multilayer neural network consists of a layer of input units, one or more of hidden units, and one output layer of units.

In our work, the seven measures of central tendency and dispersions representing texture features of flowers are used as input values (nodes) and disease categories are used as output values (nodes). Seven different types of flower diseases were included in this work, which leads to have eight nodes (including one node for normal flower) in the output layer. The number of hidden layers is determined to be 4 using the following equation which is recommended to give better classification results [12].

$$z_h = \frac{4d^2+3}{d^2-8} \quad (9)$$

where z_h is number of hidden layers and d is number of input layer nodes. Figure 4 shows the multilayer perceptron network model used for training and classification of flower diseases where the numbers 1,2,…,8 in the output layer represent aphid, Japanese beetles, rosette, golden rod-rose, mossy-rose, gall wasp, normal, and soldier beetles, respectively. Since we have eight groups of flower images, we have used eight bit binary number for each group in which case each bit refers whether that feature belongs to a group represented at that bit position.

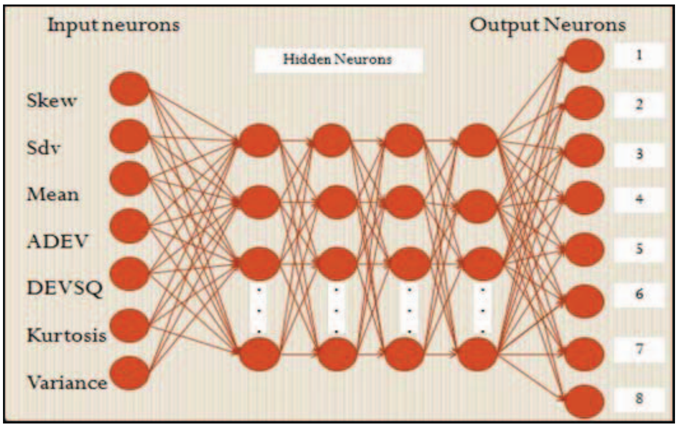

Fig. 4. Multilayer perceptron network model.

The training process finally generates a knowledge base which contains the complex relationship between various feature values of flowers and their respective disease classes. The knowledge base becomes the primary input for the decision making process.

C. *Disease Identification*

Identification of flower diseases is made by making use of the knowledge base created during the training phase. The procedures we use in this phase are similar to that of the training phase except that flower images are not labeled. Unknown flower image samples pass through pre-processing and segmentation processes. Then, the seven measures of central tendency and dispersions representing texture features of flowers discussed above are computed where the system matches against the knowledge base to predict the type of disease associated with the unknown flower image.

IV. EXPERIMENT

A. Dataset Collection

The image dataset used for training and testing the system is acquired from two flower producers in Ethiopia. During the image acquisition both normal and diseased flower images are captured by a digital camera. For this experiment, seven diseases of flowers (mentioned in Section III) were identified by producers as most prevalent and their images were taken in the same controlled environment in order to avoid external effects of sunlight and other environmental conditions. Accordingly, the distance between the flowers and the camera was approximated to 30 cm and images were taken after sunset. A total of 320 flower images are collected from which 40 of them are normal flowers and 280 diseased flowers (40 images for each disease category).

B. Training

Identification of an object using a machine learning approach has two basic phases: *training* and *testing*. In the training phase, data is repeatedly presented to the pattern recognizer, while weights are updated to obtain a desired response. Thus, we design the identifier by partitioning the image dataset into training, validation and testing data set. From the total dataset 80% was used for training, 5% was used for validation and the remaining 15% was used for testing. Since the expected output is a sequence of binary digits, a sigmoid transfer function was used in the output layer. The network was trained to output 1 in the correct disease class of the output vector and to fill the rest of the output vector with 0, as it is discussed in Section III.

During training, the connection weights of the neural network were initialized with some random values. The training samples were input to the neural network in random order and the connection weights were modified according to the error. This process was repeated until the mean squared error (MSE) fell below a predefined tolerance level or the maximum number of iteration is achieved. The validation set was used for improving generalization. During the training process the error in the validation set is monitored. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to over-fit the data, the error on the validation set typically begins to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned. In this experiment, the best validation performance was measured at 132 iterations and the validation error was 0.015137. The performance of the validation sate is shown in Fig. 5.

C. Test Result

From the flower image dataset, 48 images (15% of the total) which were not included in the training set were used to test the performance of the system. Each disease category (including normal) was represented by 6 images. Test results showed that 40 diseased flowers (83.3%) were correctly identified. Most of the errors were due to misclassification of an input to a different disease category. For example, an analysis of a confusion matrix showed that 16.7% of the flowers that were affected by Japanese-beetle were misclassified as normal roses.

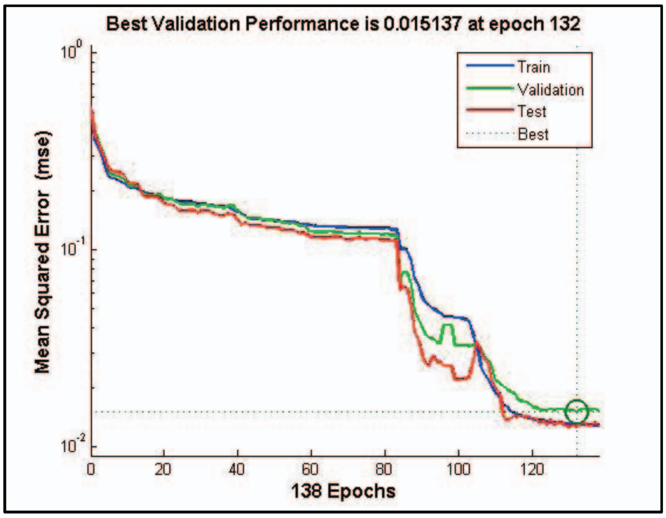

Fig. 5. Screen shot of performance measure.

V. CONCLUSION AND FUTURE WORKS

We described a system that automatically identifies flower diseases which would help flower growers to apply relevant measures at the early stage of the disease. This is important when there are no or limited experts which closely monitor the flowers. Being an emerging export sector in developing economies, the system can assist the sustainability of the floral industry. Better results could be achieved by training the system with large dataset with a variety of diseases. Moreover, the performance of the system can be enhanced by applying other machine learning techniques. Thus, future work is directed at the use of hybrid systems to take advantage of the synergy effects of two or more classifiers.

REFERENCES

[1] H. Al-Hiary, S. Bani-Ahmed, M. Reyalat, M. Braik and Z. Al-Rahamneh, “Fast and Accurate Detection and classification of plant diseases”, *International Journal of Computer Applications*, Vol. 17, No. 1, 2011.

[2] S. Arivazhagan, R. Shebiah, S. Ananthi, and S. Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", *International CIGR Journal of Agricultural Engineering*, Vol. 15, No. 1, pp. 211-217, 2013.

[3] J. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases", *SpringerPlus*, 2:660, 2013.

[4] J. Bossu, Ch. Gee, G. Jones, F.Trechetet, “Wavelet transform to discriminate between crop and weed in perspective agronomic images”, *Computers and Electronics in Agriculture*, Vol. 65, No. 1, pp. 133-143, 2009.

[5] A. Coolen, R. Kühn, and P. Sollich, *Theory of neural information processing systems*. Oxford University Press, 2005.

[6] E. R. Davies, "The application of machine vision to food and agriculture: a review", *The Imaging Science Journal*, Vol. 57, No. 4, pp. 197-217,2009

[7] Ethiopian Horticulture Development Agency, *Ethiopian Horticulture Sector Statistical Bulletin*, Issue 01, 2012.

[8] The Flower Expert, "Guide on flowers and gardening", Available at http://www.theflowerexpert.com/content/flowerbusiness/floral-industry, Accessed in September, 2014.

[9] Y. Jemaa, and S. Khanfir, "Automatic local Gabor features extraction for face recognition", In *Proceedings of the First Workshop on Image Processing Theory, Tools and Applications*, pp. 1-6, 2008.

[10] J. McMaugh, *What garden pest or disease is that*? New Holland Publishing, Sydney, 2001.

[11] N. Otsu, "A threshold selection method from gray-level histograms", *Automatica*, Vol. 11, pp.23-27, 1975.

[12] K. Sheela and S. Deepa, "Review on methods to fix number of hidden neurons in neural networks." *Mathematical Problems in Engineering*, 2013.

[13] M. Siddiqi, I. Ahmad and S. Sulaiman, "Weed recognition based on erosion and dilation segmentation algorithm" In*Proceedings of International Conference on Education Technology and Computer*, *IEEE*, pp. 224-228, 2009.

[14] B. Sparks, J. Woodward and G. Wade, "Insect and Disease Management in Annual Flower Beds", University of Florida, Available at: http://leon.ifas.ufl.edu/COMHORT/Insect_and_Diseases_Management.shtml, Accessed in September, 2014.

[15] P. Tzionas, S. Papadakis and D. Manolakis. "Plant leaves classification based on morphological features and a fuzzy surface selection technique" In *Proceedings of the Fifth International Conference on Technology and Automation*, Thessaloniki, Greece, pp. 365-370, 2005.

[16] B. Woodford, N. Kasabov and C. Wearing. "Fruit image analysis using wavelets" In *Emerging Knowledge Engineering and Connectionist-Based Systems, Proceedings of the ICONIP/ANZIIS/ANNES*, pp. 88-91, 1999.