An Intelligent Flower Analyzing System for Medicinal Plants

Ranga Rodrigo
School of Electrical and
Computer Engineering
University of Moratuwa
Email: ranga@uom.lk

Kalani Samarawickrame
Informatics Institute
of Technology
Email: ksamarawickrema@gmail.com

Sheron Mindya
Informatics Institute
of Technology
Email: smindya@gmail.com

***Abstract*—The natural sciences have been transformed with the incorporation of advanced technologies, and the current technological wave of change is also revolutionizing biological science. With the wide use of herbal medicines in traditional medical systems, the demand for medicinal plants has increased enormously. Though the current trend has made the use of medicinal plants more popular, we still require better methods to distinguish between medicinal plants and plants which do not possess medicinal value. Manual methods of plant recognition rely on a plants flower information. However, these manual processes of flower recognition are not straightforward for lay persons unless expert guidance is provided. Motivated by this reasoning, an intelligent flower analyzing system has been developed to recognize medicinal plants. Various tests have been performed on 160 images taken from four types of flowers, to recognize flowers based on their colour and shape features. The experiment was carried out using Support Vector Machine (SVM) and Principal Component Analysis (PCA) methodologies respectively for colour and shape extraction. Acquiring an average of 65% recognition rate implies that the applicability PCA and SVM in the specified domain is a valuable step forward.**

I. Introduction

The demand for medicinal plants has increased over the past two decades as herbal medicines are being used widely in traditional medical systems in developed and developing countries [4]. Traditional herbal medicines such as ayurveda, homeopathy, naturopathy, unani are used in traditional medicine systems such as in China, Africa and Latin America. However, although the current trend has drawn our attention towards medicinal plants, people who do not have an in-depth knowledge in biology lack awareness in distinguishing an authentic medicinal plant from an adulterant. Wrong identification of medicinal plants is one of the factors that make herbal remedy unsafe. [1] Owing to the ignorance of the exact identities of plants used in ayurvedic practice, many exotics are being used mistakenly or as substitutes in the absence of the plants originally recommended. Plants are classified most often on reproductive (flower) characteristics as opposed to vegetative (leaf, stem) characteristics. This has become customary since reproductive parts (petals, sepals, stamens and pistils) remain relatively unchanged over diverse environments, whereas vegetative parts tend to change depending on the environment in which they grow. In addition, flower recognition done with the use of flower guides is not straightforward for an amateur [5]. Identifying a flower using a field guide or key without expert guidance is also a time-consuming task. Furthermore, the fact that some of the flowers being relatively similar and different examples of the same flower differ in colour and shape implies that the recognition by laypersons or pattern recognition systems is not straightforward. Flower recognition becomes cumbersome when using keys as they require answers for a series of questions in order to recognize flowers. These features often relate to the internal structure of the flower which is in most cases visible only when it is dissected [9]. Different approaches have been proposed concerned with recognition of characters, human faces, cancer from chest X-ray, etc. However, outfield living objects such as fish, animals, and plants have not attracted much attention [6]. Moreover, the large similarities between flowers results in an extra challenge in flower recognition over categories such as bikes, cars and cats. In addition, flowers are non-rigid objects that can deform in many ways, and consequently there is also a large variation within classes [9]. Different approaches have been proposed for flower recognition. These approaches are based on either colour or shape information associated with the flower image. Researchers have presented an interactive extraction system for flower recognition based on colour [3] and conversely another group has proposed a model-based interactive flower recognition system based on the shape of flowers [17]. Based on [11] researchers have also proposed a method of extracting flower regions based on Intelligent Scissors. However, these models are highly user dependant as the systems require the user to adjust the model according to the shape of flower in order to obtain more accurate results. The automatic boundary extraction method proposed by Saitoh et al. [10], solves the aforementioned issue to some extend thus outperforms the Intelligent Scissors approach. Moreover the research shows that using colour or shape feature alone, is insufficient for flower recognition. In this paper, a novel approach has been introduced by the authors to incorporate colour information with shape of flowers. The technique proposed in this paper for flower recognition is composed of the following steps. Initially, the flower recognition system has been described briefly. The process of shape recognition based on Principal Component Analysis(PCA) is described in the following section followed by the use of Support Vector Machine for classification using colour information. Afterwards an outline of the experiment is carried out and the test results have been stated. Finally the paper is concluded by stating the findings and the overall outcome of the proposed approach.

II. System Description

A database of 160 flower images from four different types of flowers has been used in the system. These images have been obtained under various conditions. All pictures have been captured using a digital camera having a constant dark background, lighting condition and a constant distance from the camera to the flower. These are 75 by 50 pixels. From the 160 images, 120 have been taken for the training database and the rest have been used as test images. When the system is given an image of an unknown flower, it extracts colour and shape information and selects the respective flower from the training database. Afterwards, the particular medicinal plant details along with an image of the medicinal plant are provided by the system. With the use of these images the colour and shape information were obtained as illustrated in the next sections.

III. PCA Approach for Shape Extraction

In flower recognition, the flower image is transformed into a small set of characteristics, called eigenfaces, which are the principal components of the initial training set of face images. Similarly, it can be defined as the mean squared distance between the data points and their projections [2].This is called the eigenface method. It is a well known template matching approach where the feature values transformed by principal component analysis are used for recognition [6]. In order to process eigenfaces calculation, each two dimensional (NxN) image in a set of training images, is converted into a column vector of dimension NxN. Subsequently, all columns are joined in order to compute the average image flower vector. This is followed by obtaining mean flower and co-variance matrix to generate new eigenvectors [7]. Eigenvectors are a set of features which together characterize the variation between images. Each image location contributes more or less to each eigenvector [14]. Thus the eigenvectors can be displayed as ghostly faces of flower images. Nevertheless, aforementioned vector contains information on each pixel after projecting the image into eigenface components. When a new image is given, it is also represented by its vector of weights. Hence, recognition of the test image is done by locating the image in the database which has the closest weight to the test image [16]. An image vector will contain w*h pixels of information when width and height contain w and h number of pixels respectively. However, prior to shape extraction the images were enhanced by compressing images and filtering noise in order to reduce execution time and preserve edges of the flowers respectively.

IV. Recognition using Euclidean Distance

For an unknown flower image, the closest matching train image is selected with the use of Euclidean Distance. The distance from the test image to each eigenface which is obtained subtracting the average flower image vector from each training image is calculated using Euclidean Distance. Thus it gives the nearest image from the training data set which matches with that of the given flower image. Based on the calculated minimum Euclidean distance, the relevant flower index from the training database is selected. Afterwards, the matched flower image along with the relevant medicinal plant and the description is provided by the proposed flower analyzing system.

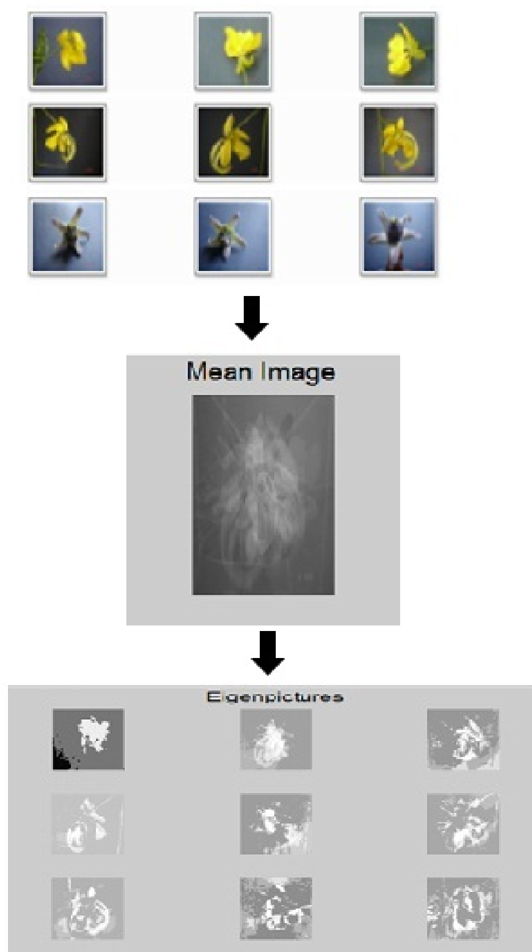

Fig. 1. Generating Eigenpictures

V. Machine Learning for Recognition using color

Support Vector Machines are based on the results of statistical learning theory carried out by Vapnik. SVM maps feature vectors into a higher dimensional space and classify data using linear algebra by employing a kernel function. Then an optimal hyper plane that fits into the training data is created. In a linear classification the margin between the separating hyper plane and the nearest feature vectors from both classes is maximal. The feature vectors closest to the hyper plane are called support vectors. SVM has evolved from sound theory to implementation and experiments while Neural Networks has followed a more heuristic path, from application and extensive experimentation to theory. Moreover it is stated that SVM has achieved practical learning benchmarks in digit recognition, computer vision and text categorization [15]. Thus SVM had been adopted as the machine learning technique in order to classify medicinal plants based on flowers using the color feature.

The color feature plays a foremost role when classifying a flower image. According to [13] , a color space with a separation of the luminance and chrominance information tends to provide better results than a color space with these information mixed (as RGB). Luminance and chrominance are separate in the LAB color space which is more effective for this task. a and b dimensions of Lab color space had been applied in order to extract color information. L has not been considered since it refers to the lightness in the image. Dimension a denotes the position between red/magenta and green where as dimension b denotes its position between yellow and blue.

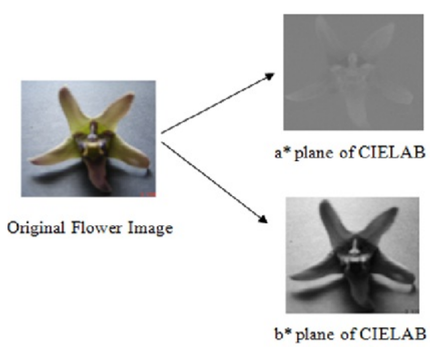

Fig. 2. Seperation of a and b color spaces

Initially a and b color planes have been extracted from a given image and combined before fed into the SVM classifier. At that moment it is also labeled with the relevant plant name. A train data set will be created from the described mechanism. Hence, when a test flower image is given, a and b color features would be extracted and combined in order to be used for testing. A linear classification would be performed and whether the flower falls into a particular category would be decided based on which side of the hyperplane, the test image lies.

VI. Evaluation

The system had been evaluated by considering classification performance using both shape and color features for 4 kinds of medicinal plants. As illustrated in the tables below, testing had been done for 10 flowers from each kind.Confusion Matrix had been drawn in order to denote how many flowers of a particular category had been identified correctly and how many flowers had been identified falsely. Analysing the data in Table 1 it could be seen that Shape Extraction using PCA has given an overall accuracy of 77.5%. According to Table2 classification of flowers using color information using SVM has given an overall accuracy of 82.5%. Based on Table2 when using SVM three instances had been identified where the Wara flower had not fallen into any of the 4 categories. The system had also been tested combining the above mentioned technologies. Thus the results of Table 3 shows that SVM and PCA technologies combined has given an overall accuracy of 65%. In this mechanism an output was given only if both PCA and SVM identified the flower to be in the same category. Accuracy can be enhanced by capturing images with a higher resolution and increasing the number of training and testing images.

TABLE I
Confusion Matrix for classification of flowers using PCA

	Clitoria	Ranawara	Ahala	Wara
Clitoria	8			
Ranawara		8	4	
Ahala	2		6	1
Wara		2		9
Accuracy	80%	80%	60%	90%

TABLE II
Confusion Matrix for Classification using SVM

	Clitoria	Ranawara	Ahala	Wara
Clitoria	10			
Ranawara		6		
Ahala		3	10	
Wara		1		7
Accuracy	100%	60%	100%	70%

TABLE III
Confusion Matrix for classification of flowers using PCA and SVM

	Clitoria	Ranawara	Ahala	Wara
Clitoria	8			
Ranawara		6		
Ahala			6	
Wara				6
Accuracy	80%	60%	60%	60%

VII. Conclusion and Discussion

This application is highly important in the development of natural sciences especially in plant taxonomy related studies as the manual recognition done using books is time consuming and the use of internet to search for information would not be accurate all the time. Furthermore at the exportation of flora, this kind of system would be useful in order to prevent bio piracy as it would be straightforward and practical than expecting the lab to test samples of each and every flora that is been exported.

In this paper, a generic approach was presented in order to recognize medicinal plants using its flower. The shape and color features of a flower had been taken into consideration when determining the plant. The results of using Support Vector Machine for color classifying a flower using color information indicate the optimality of the technique for the task. It also shows that incorporating shape information has also contributed for the high accuracy rates. In order to improve the recognition rate of 65%, leaf recognition could also be incorporated as discussed in [8].

It is believed that a system like this would contribute to the Sri Lankan ayurvedic treatments. At the same time since the solution has a generic approach where it could be trained to recognize any kind of flower and not restricted to medicinal plant flowers, the system could be used in many applicable areas where flower recognition is needed.

REFERENCES

[1] D Bagozzi. (2004, Feb) Medicinal plants guidelines to promote patient safety and plant conservation for a US$ 60 billion industry. [Online]. http://www.who.int/mediacentre/news/notes/2004/np3/en/

[2] C Bishop , Pattern Recognition and Machine Learning. Singapore: Springer Science+Business Media., 2006.

[3] M Das, R Manmatha, and E. M Riseman, "Indexing flower patent images using domain knowledge," Intelligent Systems and their Applications, vol. 14, pp. 24-33, Sep/Oct 1999.

[4] D. Deshpande, A Handbook of Medicinal Herbs. India: Agrobios, 2006.

[5] A. Evans , J. Sikorski , P. Thomas , J. Zou, and G. Nagy ,"Computer assisted visual interactive recognition (CAVIAR) technology," in 2005 IEEE International Conference on Electro Information Technology, NY, 2005, pp. 6-6.

[6] K Hyun-Chul, K Daijin , and S. Y Bang, "Face recognition using the mixture-of-eigenfaces method," Pattern Recognition Letters, vol. 23, no. 13, pp. 1549-1558, November 2002.

[7] A S Khan and L K Alizai, "Introduction to Face Detection Using Eigenfaces," in ICET '06. International Conference on Emerging Technologies, 2006, pp. 128 - 132.

[8] N Kumar, P N Belhumeur, A Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, Joo V. B. Soares, "Leafsnap: A Computer Vision System for Automatic Plant Species Identification", ECCV Proceedings of the 12th European conference on Computer Vision, 2012, vol.2, pp. 502 - 516.

[9] M Nilsback and A Zisserman, "A Visual Vocabulary for Flower Classification," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp. 1447-1454.

[10] T Saitoh,K Aoki, and Kaneko, "Automatic recognition of blooming flowers ," in Proceedings of the 17th International Conference on Pattern Recognition , 2004, pp. 27-30.

[11] T Saitoh and T Kaneko,"Automatic recognition of wild flowers ," in 15th International Conference on Pattern Recognition, 2000, pp. 507 - 510 vol.2.

[12] M Storer , M Roth Peter , M Urschler, and B Horst , "Fast-Robust PCA," in Proceedings of the 16th Scandinavian Conference on Image Analysis , 2009, pp. 430-439.

[13] L Torres, J Y Reutter, and L Lorente , "The importance of the color information in face recognition," in International Conference on Image Processing, Kobe, 1999, pp. 627 - 631.

[14] M A Turk and A P Pentland, "Face recognition using eigenfaces," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition , Maui, HI, 1991, pp. 586 - 591.

[15] L Wang, "Support Vector Machines: Theory and ApplicationsSeries," Studies in Fuzziness and Soft Computing, vol. 177, 2005.

[16] W Zhao, R Chellappa, P J Phillips, and A Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys (CSUR), vol. Volume 35, no. Issue 4, pp. 399 - 458, December 2003.

[17] J Zou and G Nagy, "Evaluation of model-based interactive flower recognition," in Proceedings of the 17th International Conference on Pattern Recognition, 2004, pp. 311 - 314 Vol.2.