ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Classification and Recognition of Herbal Leaf Using SVM Algorithm.

A Maheswari[1], N Bharathi[2], P Neelamegam[3]*, and T Gayathridevi[4].

[1]School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India.
[2]School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India.
[3]School of Electrical and Electronics Engineering,SASTRA University,Thanjavur, Tamil Nadu, India.
[4]Dept of ECE, SRC,SASTRA UNIVERSITY,Kumbakonam, Tamil Nadu, India.

ABSTRACT

Ayurveda is an ancient holistic healing system, which is nearly 5000 years old that uses natural herbal leaves to heal various diseases. Herbal leaves have been widely utilized by the industries like food, medical and cosmetic. But most of the herbal plants are endangered. So, there is a need to identify and regrow them for the use of future generations. So it is very indispensable to develop an automatic system to identify species correctly. This paper proposes an algorithm for recognition of herb leaf through image processing and classification technique using Support Vector Machine (SVM) classifier. The scale invariant feature transform (SIFT) method is used for the feature extraction in leaves. The work is carried out using 5 species of leaves in MATLAB environment. SVM classifier is trained with 273 samples and tested with 129 samples of herbal and non-herbal leaves are presented in this paper. The output showed that the system reached the recognition rate with 92.25% accuracy. The system is convenient to use and cost effective for the botanist, Scientist and Researcher to identify herbal leaves.
Keywords: Herbal leaves recognition, classification, Image processing, SIFT features, SVM classifier.

***Corresponding author**

INTRODUCTION

India has a rich history in medicinal system. In particular, the Ayurveda medicinal system followed is based on herbal leaves. Ayurveda has been considered as an alternative form of allopathic medicine in the world. A large number of industries have been emerged especially in herbal leaves and hence there is a need to recognize different kind of herbs. Apart from this most of the medicinal herbs are at the risk of extinction without proper monitoring. An important aspect is that the medicinal herbal plants should be protected by periodically recording any herb species in detail. Hence knowing the particular type and use of herbal leaves would be very difficult since manual identification is prone to error and limited within the human knowledge. With the increasing utilization of computer and artificial intelligence made it so easier for classification of herbs compare to the traditional methods followed. Generally the herbal plants are collected through human expert recognition from dense forest. But this approach is hard and ineffective. It also consumes more energy and time in case of large and dense forest. Therefore, by the application of image processing and recognition techniques an efficient herbal leaf identification system can be achieved.

Conventional method like Lucid, Uconn and Cal Floraare used for the plant recognition and data management for the recognized plant. The shortcomings of these systems are it uses grayscale inputs to recognize plant species and it does not support image processing techniques [1, 2,3]. Probabilistic Neural Network (PNN) was trained by with 32 species of leaves for recognition and the experiment results showed the accuracy of 90% [4]. A novel approach using Wavelet Transform (WT) and Gaussian interpolation has been proposed for the leaf recognition based the Skeleton feature. The result showed that the recognition rates must be significantly improved [5].Thai Herb Leaf Image Recognition System (THLIRS) involved 32 species of Thai herbs for the recognition system. It achieved precision rate of 93.29 for trained dataset and 0 for the untrained dataset [6].Recently, automatic approach for plant recognition is employed using computer vision and pattern recognition techniques [7]. Automatic techniques have certain limits by assuming prior knowledge of the leaf. Several approaches have been proposed to effectively classify and recognize herbal leaves. But most efforts in this field are not robust to some changes like rotation and illumination of images. In this work, an efficient and novel Scale Invariant Feature Transform (SIFT) algorithm for herbal recognition is proposed which has a strong adaption to misalignment and affine transformation. An affine transformation preserves co-linearity relation between points and ratios of distances along a line [8, 9] like changes of light, rotation, etc. By using SIFT herbal leaf features are extracted from the interested point and it is used for reliable matching between different views of herbal leaf images. The extracting the invariant features and classification of the leaf images is categorized into herbal and non-herbal leaves. Finally, two sets of database contain 404 images of leaf images for herbal and non-herbal was employed by a SVM classifier to identify the image correctly.

This paper explains an efficient way of classification and finding of herb leaf based on structural classification such as SIFT feature extraction and SVM classifier. The proposed system flow diagram is shown in figure 1.

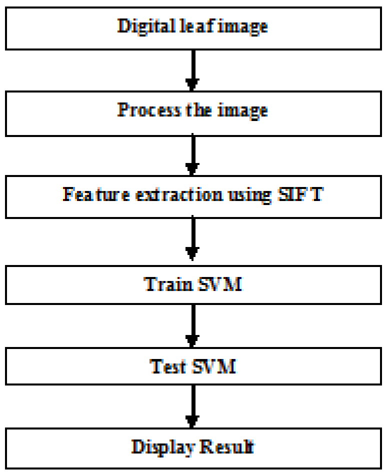

Figure 1: Flow diagram

ISSN: 0975-8585

SYSTEM CONCEPTUAL ARCHITECTURE

Image Acquisition

This work aims to develop a computerized intelligence system for recognizing herbal leaf among a trained dataset. To achieve this, SIFT algorithm and SVM classifier is employed. In this work 5 species of leaf were processed. The 404 sample leaf images were used for model training and testing. The images were sized to 512 X 512 pixels with 100 dpi resolution in order to reduce the complexity of large sized images. The Conceptual Architecture of system is shown in figure 2.

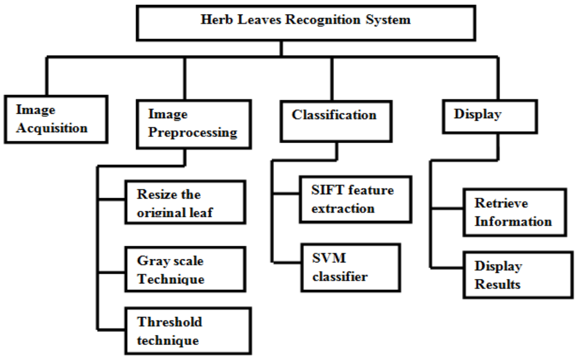

Figure 2: Conceptual Architecture of system

SIFT feature extraction

David Lowe proposed a SIFT mathematical algorithm for feature extraction [10].It includes extracting the interest point features from images used for the matching of same images in different views.A staged filtering approach is used for the identification of features. The extracted features are invariant to any changes in scale, affine transformation and illumination. It achieves more accurate verification and orientation determination compared to the approaches based on indexing. The extracted SIFT features from the images is stored and new image is compared with it for the identification.The following are the four-stage algorithm used for the feature extraction.

Scale-space extrema detection

This first stage of computation involves the searching of image location andoverall scales under different views of image object. Scale space is defined as finding the stable feature using a continuous function of scale to find.It is efficiently implemented by the mathematical approach using a difference-of-Gaussian function. It gives the interest points which are invariant to orientation and scale.
First a scale space is given by equation

$$A(x, y, \sigma) = B(x, y, \sigma) * C(x, y)$$

where the convolution operator is represented by *, the variable-scale Gaussian is defined by B(x, y, σ) and the input image is defined by C(x, y). Difference of Gaussian function is convolved with input image to find the stable features. It is calculated with difference of two nearby scales. The Gaussian image is down-sampled with factor of 2 for each octave repeated.

$$D(x, y, \sigma) = A(x, y, \kappa\sigma) - A(x, y, \sigma)$$

Compare each point to its 8 neighbors at the same scale and additionally with 9 neighbors at neighboring scales. It gives the local maxima and minima of D(x, y, σ).The pixel is a candidate key point for local maximum or minimum.

Key point localization and filtering

For each candidate location and scale determination a detailed model is fixed. Key points are chosen based on stability measurement. This step attempts to eliminate key points which are located on edges or the contrast between point its neighbors and is low.

Orientation assignment

Each key point location is assigned with single or more than one orientation to make them invariant to rotation. It is done based on gradient directions of the local image. The images are transformed to the assigned orientation, scale, and location for each feature and further operations are carried out using these transformed images only. Hence it provides invarianceto transformations. Suppose for a key point, A is the image with closest scale. For each image gradient magnitude and orientation can be computed using equation given below respectively.

$$\text{Gradient vector} = \begin{bmatrix} A(x+1, y) - A(x-1, y) \\ A(x, y+1) - A(x, y-1) \end{bmatrix}$$

$$g(x, y) = \sqrt{(A(x+1, y) - A(x-1, y))^2 + (A(x, y+1) - A(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((A(x, y+1) - A(x, y-1))/(A(x+1, y) - A(x-1, y)))$$

In next stage a gradient histogram (36 bins) is created. A keypoint with that orientation is created using any local peak within 80% of the highest peak.

Key point descriptor

The image gradients can be measured at the each chosen scale in the interested region around every key point. Then it is transformed into a representation and also it allows for certain levels of change in illumination and local shape distortion. The descriptor is calculated at the final step to achieve invariant to remaining variations. The presence of the image is confirmed when there are at least three keys with low residual on the model parameters. The image has dozens of SIFT keys with substantial levels of occlusion. The reliability is retained in high level. Whereas there are some basic differences between leaf like shape, size, boundary, color, etc, SIFT features are different in two classes and these features could be used for herbal leaf classification.

CLASSIFICATION USING SVM

Support Vector Machines (SVM)is a supervised learning techniqueusually preferred in classification and recognition because of their promising performance. Based on the selection of parameters and kernel design the performance can be improved. SVM multiclass classification can also solve real world problems efficiently with the well-defined theoretical model for multiclass data set. SVM can maximize the geometric margin at the same time it can minimize the empirical classification error.

SVM classifier

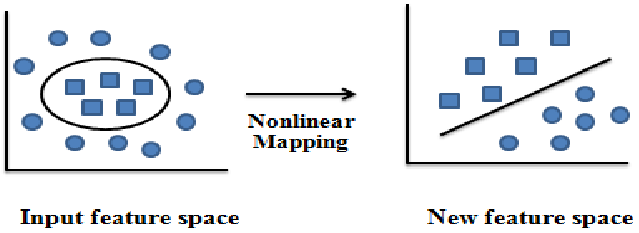

Figure 3: SVM Classifier.

SVM classifier maps the input vectors into a new high-dimensional feature space. The nonlinear mapping gives the solution for two class classification problem.It constructs an optimahyper planeto separate the two classes by finding the largest margin. The two classes are represented by positive and negative class [11]. The figure 3 shows the two class classification.

SVM is inspired from statistical learning theory and based on concept of structural risk minimization. For a training set $\{(s_i, c_i) \mid s_i \in R^p, c_i \in \{-1,1\}\}_{i=1}^{n}$ where c_i indicates the class to which sample s_i is belongs. Each $s_i$is a p-dimensional real vector. The problem is finding the optimal hyper plane that classifies the samples in two classes. It is defined byequation

$$f(x) = \sum_{i=1}^{l} (c_i \alpha_i \mathrm{k}(\mathrm{s}, s_i) + \mathrm{b}$$

Where constants are represented by α and b and kernel function is represented by $k(s, s_i)$ and the class label of sample s is determined by f(x). For linear SVM the kernel function is dot product of two N-dimensional vectors equation

$$k(s_i, c_i) = s_i \cdot c_i$$

For nonlinear SVM the kernel function projects the samples to a Euclidean feature space of higher dimensional M $\psi : R^N \rightarrow F^M, >> N$ and constructs a hyper plane in F. In this case, kernel function defined as equation, where Ψ is the nonlinear projection function.

$$k(s, c_i) = \psi(s) \cdot \psi(s_i)$$

EXPERIMENTAL RESULTS

Experiments modules are developed using MATLAB R2013a, which runs in the environment Windows7.Five speciesof samples are taken for the experiment, whose digital images are obtained by means of camera. Table 1 shows the species type and numbers of leaves images for each species.

A set of herbal and non-herbal leaf images of five species type containing 404 images are used as the train set and test set. The name of species, number of images used for the training and testing is shown in the Table 2. Besides, 273 images were chosen as the training set, including 45 Circinatum images, 57 Garryana images, 49 Glabrum images, 67 Kelloggii images and 55 Macrophyllum images. Similarly, 129 images were chosen as the testing set, including 21 Circinatum images, 27 Garryana images, 24 Glabrum images, 30 Kelloggii images and 27 Macrophyllum images.

Table 1: Species type and Number of leaves

Species name	Number of leaf images
Circinatum	66
Garryana	84
Glabrum	75
Kelloggii	97
Macrophyllum	82

Table 2: Species type and Number of training and testing leaves

Leaf species Name	No. of Training samples	No. of Test samples
Circinatum	45	21
Garryana	57	27
Glabrum	49	24
Kelloggii	67	30
Macrophyllum	55	27

Pre-processing Module

This will pre-process all the training images to make optimal decision. Give the current directory location of training image in input text field and input the directory name for current training data set. Processed image directory will be shown after processing of training data. The figure 4 shows the Pre-processing stage.

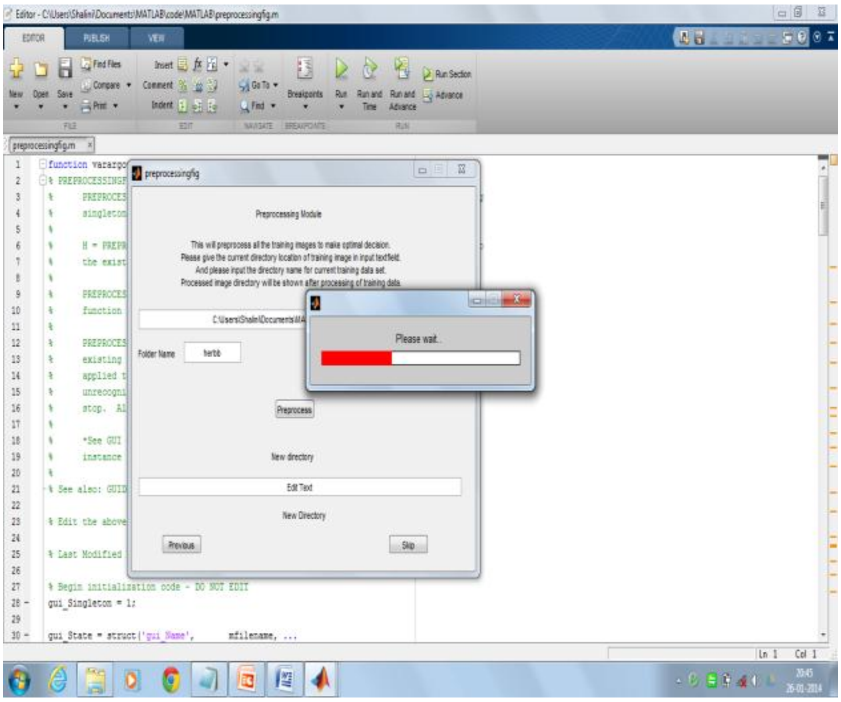

Figure 4: Pre-processing of training images

Generate meta data module

This module will generate meta data for a category of image for classification purpose. Generate meta data once for each category. Specify the metadata file name in text field. The figure 5 shows the Generate meta data stage.

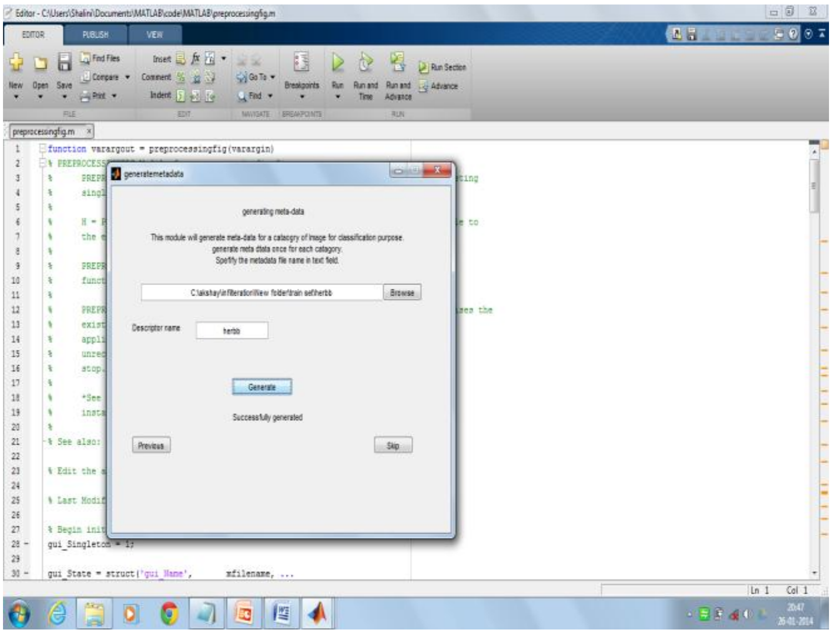

Figure 5: Generate meta data

SVM training module

This module trains the system to a particular class and stores the training element with a specified name. For the same pair of math same training element can be used. The figure 6 shows SVM training stage.

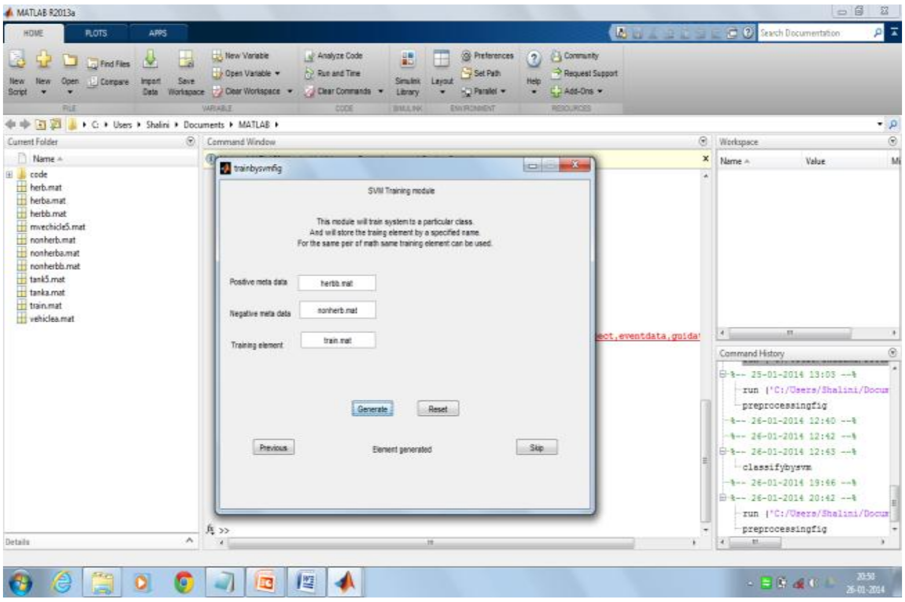

Figure 6: SVM training

SVM classification module

The developed module is able to classify the two different category of image and display the name of the particular image. Load the image in the image name field. Image of loaded file appear on the windowand then classify it. The figure 7 shows SVM classification of herbal leaf and figure 8 shows the classification of non-herbal leaf.

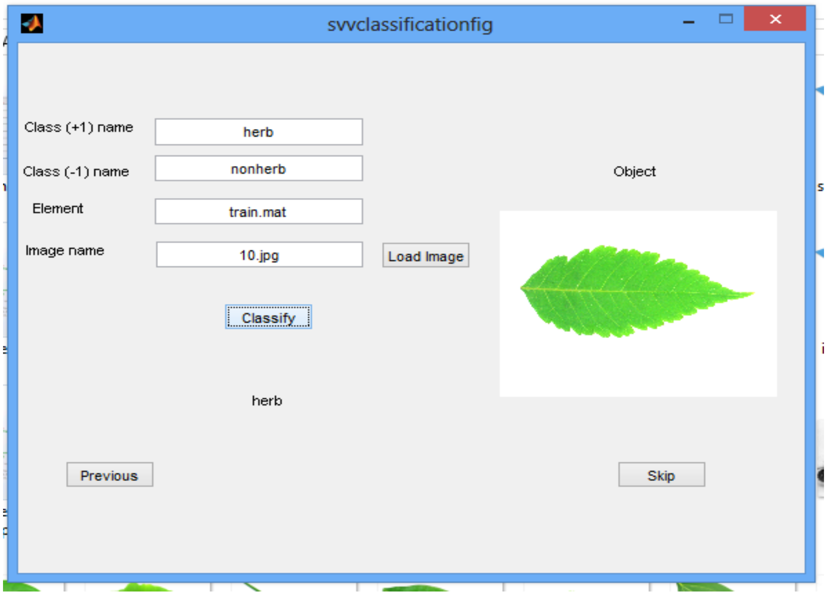

Figure 7: SVM classifications for herbal leaf

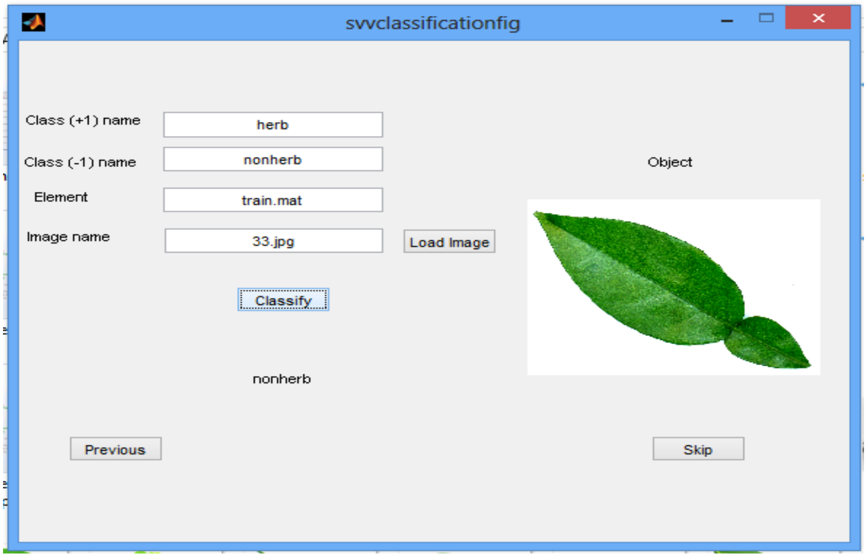

Figure 8: SVM classifications for non-herbal leaf

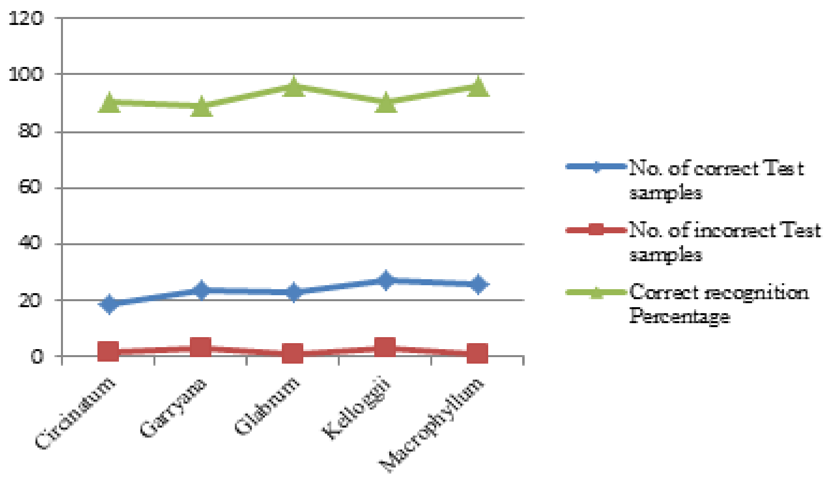

Figure 9: Comparison of accuracy percentage between the five species

DISCUSSION

A new approach based on SIFT features extraction was proposed for herbal leaf recognition in this work. The whole process of leaf classification can be implemented using leaf detection, feature extraction and classification. For examining the proposed method, the collected dataset is used. The dataset contains herbal and non-herbal images. Images were preprocessed and cropped to a fixed standard size. Then, features are extracted from all the leaf images in the dataset using SIFT algorithm. For each image leaf more frequent SIFT key points are extracted to identify a unique feature. It allows finding similar features for different image. The attributes of each key point descriptor describe that region in an orientation and scale invariant way. Ultimately, the extracted SIFT features are rendered to a SVM classifier for purpose of classification. There are distinctive differences between herbal and non-herbal leaf in shape, color, size etc. Therefore, recognition is based on these differences. In other words, differences between herbal and non-herbal leaves and the key points which are extracted from leaf are used for classifying. The novel approach is performed and the datasetis divided randomly in two parts, 70% for training and 30% for testing. The correct recognition rate for the five species of leaf is shown Table 3. The comparison of accuracy percentage between the five species is shown in the figure 9. The overall accuracy percentage obtained from performing the experiment is 92.25%.

Table 3: Correct recognition Percentage of five Species

Name of the species	No. of correct Test samples	No. of incorrect Test samples	Correct recognition Percentage
Circinatum	19	2	90.48
Garryana	24	3	88.89
Glabrum	23	1	95.83
Kelloggii	27	3	90
Macrophyllum	26	1	96.3
Total	119	10	92.25 %

CONCLUSION

This paper demonstrates a novel technique for leaf recognition based on SIFT features and SVM classifier. The proposed leaf detection technique with ability of detecting the leaf locating in white backgrounds and the SIFT features with ability of finding the interest points of the leaf describing the discriminate characteristics of herbal and non-herbal leaves, lead us to construct a system for herbal leaf recognition with high robustness in input images and accuracy of recognition output. By employing a SVM classifier the proposed method yields a recognition rate of 92.25% which is appreciable comparing to the other methods with constraints of having simple backgrounds and being time consuming.

REFERENCES

[1] Fu H, Chi Z. IEE Proceedings 2006; 153(6):881-892.

[2] Xun Liu. Index and Recognition for the Shape Contour of Plant Leaves; Intelligent Systems GCIS WRI Global congress; 2009; 436-440.

[3] Feng Chen, Yun-Feng li. A Multiresolution Segmentation method for Tree Crown Image using Wavelets; International Conference on Wavelet Analysis and Pattern Recognition; 2007; 1556-1559.

[4] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, I-Fan Chang, Qiao-Liang Xiang. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network; IEEE International Symposium on Signal Processing and Information Technology; 2007; 11-16.

[5] Xiao Gu, Ji-Xiang Du, Xiao-Feng Wang. Leaf Recognition Basedon the Combination of Wavelet Transform and Gaussian Interpolation; International Conference on Intelligent Computing; 2005; 3644: 253-262.

[6] Chomtip Pornpanomchai, Supolgaj Rimdusit, PiyawanTanasap, Chutpong Chaiyod. Kasetsart J: Nat Sci 2011; 45(3): 551-562.

[7] Pan J, He Y. Recognition of plants by leaves digital image and neural network; IEEE International Conference on Computer Science and Software Engineering; 2008; 4: 906- 910.

[8] De Villiers M. The affine invariance and line symmetries of the conics; Australian Senior Mathematics Journal; 1993; 7(2): 32-50.

[9] Yaglom Y. Geometric Transformations III; Mathematical Association of America (MAA), Washington; 1973; 9-20.

[10] Lowe David G. Object Recognition from Local Scale-Invariant Features; Proceedings of the Seventh IEEE International Conference on Computer Vision; 1999; 2:1150 - 1157.

[11] Mehdi Moradi, Purang Abolmaesumi, Robert Siemens D, Eric Sauerbrei E, Alexander Boag H, Parvin Mousavi. IEEE Trans Biomed Eng 2009; 56(9):2214 – 2224.