CHAPTER 1
INTRODUCTION

In this Chapter, we introduce the area of computer vision, images of medicinal plants, generic literature survey, motivation, problem definition, contributions made and organization of the thesis.

1.1 COMPUTER VISION

Image recognition is the basic task in the areas of computer vision and pattern recognition. The field of computer vision is concerned with the automated processing of any images from the real world to extract characteristics through computers and interpret information on a real time and user requirements basis. Today, computers are playing key role in our daily lives. Most significant developments in computer hardware and software have contributed to the development of many real world research applications, namely, finger print recognition, digital signature recognition, automatic medical diagnosis and treatment, optical character recognition, document processing, video processing and plant image recognition etc. In addition to these, we also find many commercial applications, namely, commerce, sports, entertainment, business, education, food science, medical science, industrial and automation. An effort has also been carried in forestry, farming, botany and in many more allied fields. The crop growth and health analysis, soil fertility, pest or disease detection, weed detection and post-harvest operations are the
most common applications. The satellite image processing has been widely employed for complex scene analysis, weather and rain forecasting. Ultimately, all the applications exhibit machine intelligence in the respective fields.

Figure 1.1: (a) Medicinal plants parts used as extracts of medicine

Figure 1.1: (b) Herbal plant parts and spices

Ayurveda, the ancient medical science of India, is accepted as the oldest scientific medical system, with a long record of clinical treatments. The origin of Ayurveda was about more than 5000 years ago during vedic times. In India the ayurvedic treatment is popularly known as Indian System of Medicine (ISM). It is envisaged that 80 percent of the world population are still relying on traditional system.
It is also known fact that allopathic medicines are also the origin of chemical constituents of plants. Kerala is the first state in our country which separates the traditional ayurveda into three, such as Indian System of Medicine(ISM), Homoeopathy and Ayurveda Medical Education. The Government of Kerala, is one of the main department, where an excellent effort is put to serve the people with ayurveda. Figure 1.1(a) thru Figure 1.1(e) gives a typical environment where medicinal plants extracts are prepared and stored. It also reveals that the parts of the plants are used in the preparing different forms of herbal products, like herbs, decotion, oil, paste, powder.

Figure 1.1: (c) People searching for the collection of medicinal plants
Figure 1.1: (d) Ancient way of preparation of herbal medicines

Figure 1.1: (e) Ayurved herbs and products

In recent years, an automatic content extraction, analysis and retrieval for medicinal plants based on visual trait studies has become more important than ever before. This is because manual identification and processing of herbal drugs suffers from two major
issues, namely, the scarcity of experts and subjectivity arising with the individuals in the field of Ayurveda. Therefore, the Ayurveda practitioners and biologists are in need of efficient computer software to automatically extract and analyze significant content of images of medicinal plants. Hence, in order to supplement the human visual system with machine vision, medicinal plant recognition is considered a research issue.

The tribes and rural people in India have enormous knowledge of medicinal plants and their use in treating different diseases. These people prepare the medicines from the extracts of parts of medicinal plants, such as leaves, bark, roots, flowers and fruits. The healing properties of many herbal medicines have been recognized in many ancient cultures. In modern medicines, plants also occupy a significant place as raw material for some important drugs and in naturopathy. As the days passed by, the knowledge of medicinal plants has become more necessary for many. But this tacit knowledge dies with the experts. For the purpose of conservation, domestication and sustainable utilization of medicinal plants there is a dire need for automation and development of computer vision system for the community. Hence, an automatic identification of medicinal plants with their visual content is the topic of this research.

All the Ayurveda medications and therapies are available in the form of oil, decoction, powder and paste of leaves. The Ayurveda treatment does not have any side effects. The majority of the
medicines are prepared by the experienced saints and experts. But the information pertaining to species of medicinal plants and geographic location are usually not shared with others. Hence, the authors have realized this drawback and attempted at automation of activities in Ayurveda through medicinal plant identification and classification.

1.1.1 Medicinal plant images

Medicinal plants help mankind and others in one or the other way. In India, medicinal plants are used in food. But, due to environmental deterioration and lack of awareness, many rare plant species are on verge of extinction. In India, Himalaya is the rich source of medicinal plant species. These species have got medicinal importance. The medicinal plants are divided into herbs, shrubs and trees. The healthcare company called, Himalaya Drug Company, Bangalore, India, exports Ayurveda medicines since 1930. The information is available on http://www.himalayahealthcare.com. According to report of the Task Force on conservation and sustainable use of medicinal plants, the Government of India, planning commission-2000, it is observed that the shift to Ayurveda medicines earns revenue to India by way of foreign exchange. Hence, the task of medicinal plant identification and classification has become very important in view of involving people in the domestication of these plants.
The general taxonomy provides important information for the identification and classification of medicinal plants. All the information about the plant like, locality of collection, date, morphological characters of plant like leaf and flower morphology, branching, venation, inflorescence, fruit and seed character are obtained. The Ayurveda researchers and botanists require such information for identification. Instead of manual system, a database of medicinal plants is created and retrieval methodologies based on text and visual information becomes useful. However, the design of a machine vision system to recognize, classify and retrieve medicinal plants is a challenging task due to the factors like illumination levels, weather conditions, shadows, occluded regions and cluttered background. Hence, the authors have chosen to help people living in forests, rural areas and those who practice Ayurveda medicines. In the present work, the authors have attempted to design an efficient approach for the identification and classification of images of Indian medicinal plants through digital image processing techniques. Finally, an image search engine for database of medicinal plants is developed for retrieval based on text and visual information. The Figure 1.2 gives the samples of images of medicinal plants with their scientific names. The different leaves with varying shapes and margins with their botanical terminologies are given in Figure 1.3. The different types of leaves considered in the work are given in Appendix C.
Figure 1.2: Sample medicinal plant images (a) Catharanthus roseus (b) Vigna unguiculata (c) Capsicum annum (d) Aloe barbadensis (e) Saptachakra (f) Calotropis gigantea (g) Acacia catechu (h) Azadirachita indica (i) Carica papaya

Figure 1.3: Sample leaves images with different Base and Apex angle (a) Acute-Elliptic (b) Acute–Serrate (c) Obtuse-Obovate (d) Obtusetip-Obovate (e) Obtuse-Serrate (f) Obtuse– Acuminate (g) Obtuse-Circular (h) Circular-Entire (i) Obtuse-Serrate (j) Obtuse-lobed (k) Obtuse- Oblong (l) Obtuse-Entire (m) Wide Obtuse-Cordate (n) Wide Obtuse-Entire (o) Wide Obtuse- Lobed (p) Wide Obtuse- Lobed (q) Wide Obtuse-Obcordate (r) Wide Obtuse
1.2 GENERAL LITERATURE SURVEY

The authors have carried out literature survey to know the state-of-the-art applications of computer vision and digital image processing techniques in the real world, more specifically connected with plant recognition. Following is the gist of the works cited in the literature.

N.Valliammal and S.N.Geethalakshmi., have proposed a computerized living plant recognition system based on leaf vein information. The part of leaf image is segmented from canopy using Preferential Image Segmentation algorithm (PIS). The PIS algorithm works on the principle of curve evolution model. The planar curve matching technique is adopted for matching the images. The author has also developed an approach for leaf vein extraction by combining threshold with H-maxima transformation [96].

N.Valliammal and S.N.Geethalakshmi., have used Preferential Image Segmentation (PIS) method sing mathematical morphologies for automatic recognition of leaves and flowers. The image is encoded as small blocks of tree shapes representing hierarchical tree with each leaf node corresponding to part of the image. The shape and curve matching is adopted for feature comparison and recognition [95].

Abdul kadir et. al., have proposed a method using combined features such as, polar Fourier transform, color moments and vein features to retrieve images of leaves. The method is very useful for the recognition of foliage plants. The system has been tested on Flavia and
Folia leaf data sets. The retrieval accuracy of 93.13% and 90.13% are observed for Flavia and Folia data sets respectively [7].

Mahmood R. Golzarian and Ross A. Frick have developed a method for classification of images of three grasses, namely, wheat, ryegrass and brome grass species at early growth stages. A combination of color, texture and shape features is used. The features are reduced to three descriptors using Principal Component Analysis. Three components are able to distinguish three grasses with a classification accuracy of 85% and 88% for ryegrass and brome grass respectively. The study helps for weed management [85].

Mahmood R. Golzarian has investigated an adaptive learning for segmentation of plant images into plant and non-plant regions. The Kohenen’s self organizing map (SOM) algorithm is deployed for segmentation of plant images. Nine color features of three color space models are used as features. The method worked well even in the presence of noise [86].

Faisal Ahmed et. al., have investigated the use of Support Vector Machine (SVM) and Bayesian classifier as machine learning algorithms for the effective classification of crops and weeds in digital images. From the performance comparison, it is reported that SVM classifier has outperformed Bayesian classifier. Young plants that did not mutually overlap with other plants are used in the study [41].
Jinhai Cai et al., have developed a novel image segmentation approach for plant analysis using machine learning. The proposed algorithm uses the difference between the images to obtain initial estimation of background and then to refine the estimation using machine learning and statistical pattern recognition. It is reported that the proposed algorithm has achieved promising performance in terms of accuracy and speed [67].

Minggang Du and Xianfeng Wang, have presented a plant recognition method based on leaves images using Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). The Histogram Oriented Gradients (HOG) are used for the representation of leaf shape with different orientation and edges. The combination of PCA and LDA has provided an effective leaf image classification of varying shapes [94].

B. Sathyabama et al., has presented Content Based Leaf Image Retrieval (CBLIR) for e-commerce application. The Log-Gabor wavelet and Scale Invariant Feature Transform (SIFT) are deployed for leaf image texture and shape features respectively. The retrieval accuracy of 97.5% is observed [13].

Meng Lamei et al., have proposed classification of plants based on bark texture using texture features and genetic programming. It is reported that only five texture features are sufficient for the
recognition accuracy of plant species and have given good classification results [90].

Liying Zheng, et. al., have proposed a hybrid method of combining the Mean Shift (MS) with the Fisher Linear Discriminant (FLD) for improving the crop image segmentation. The experimentation is carried out on soya bean and weed images. The segmentation accuracy of 95% is observed with MS-FLD. The method is found to be more robust than any cluster-index based methods [82].

Xiaosong Wang et. al., have developed a method for tree image segmentation from complex background. Based on the visual characteristics differences of the tree and the surrounding objects, the trees from different backgrounds are separated into a single set of tree image pixels. The tree image is segmented based on color and texture features of gray level co-occurrence matrix [136].

Ulrich Weiss et. al., have contributed a method for recognition of plant species by robots for weed detection and nursery plant detection. The plant species are distinguished using 3D LIDAR sensor and supervised learning. Different learning methods like, logistic regression functions, support vector machines and neural networks are found to be suitable. The average classification accuracy on six types of plant species, twenty images of each class is found to be 98% [127].
Hanife Kebapci et al., have presented a content-based image retrieval system for house plant images. The plant image is segmented using max-flow min-cut technique. The color, shape and texture features are extracted and used for retrieval. The accuracy of 73% is obtained for 132 image subset from top-15 results [48].

R.S. Sabeenian and V. Palanisamy have developed an automated visual system for discriminating crop and weed from field image. The crop is segmented based on excess green method. The texture features are obtained using wavelet and gray-level co-occurrence matrix. A new method of Multi-Resolution Combined Statistical and spatial Frequency (MRCSF) is used for classification [107].

Eric J. Pauwels et al., have presented an architecture for designing a web service, which allows the user to upload the leaf images for image-based identification. The leaf image is segmented from the background and characterized using shape features. The nearest neighbour classifier is adopted for classification. But, it is reported that the limitation of the work in giving metadata of plant species. The focus of the work can be extended by including new and generalized features to develop an efficient search engine [39].

Jiangming Kan et al., have developed computer vision techniques to measure the diameters of trunk and branches of the standing tree. The template matching is used to locate the calibration stick and the principle of automatically calculating the diameters of the trunk and
branches are explained. The results are compared with manually measured parameters [68].

Meyer and Camargo Neto, have provided an advanced color vegetation index for crop detection. The results reported are satisfactory. Jiazhi Pan and Yong [66] have presented an approach for identifying different plants through digital images of leaves in farming. The image processing and neural network is used for the recognition of leaf plant image leaf blocks. Various shape features and artificial neural network classifier are adopted. The efficiency is around 80%. The proposed method when incorporated in machines assists farming saving labor charges and increasing the crop productivity [91].

Yovel Y et. al., have presented a new method for Plant classification from bat-like echolocation signals. In this work, a plant is considered as a three-dimensional array of leaves emitted bat call. The plants are classified based on signals from a database of plant echoes that are created by plants with a frequency-modulated bat-like ultrasonic pulse. The algorithm uses the spectrogram of a single echo from which it uses features that are accessible to bats. The Support Vector Machine (SVM) learning is used to automatically extract suitable linear combinations of time and frequency cues from the spectrograms. The classification is reported to be high [139].

Kamarul Hawari Ghazali et. al., have addressed the importance of color image processing for weed recognition. The combination of color
and texture feature has been used for classification of narrow and broad type of weeds. The performance of classification is compared for using Fast Fourier transform (FFT) and Gray-Level Co-occurrence Matrices (GLCM) features [75].

Sooyoung Kim has discussed a method for identifying tree species to characterize forest biological attributes. It provides a mechanism for understanding the tree species identification through the structural and spectral characteristics derived from LIDAR data. The spectral intensity information collected from airborne LIDAR data is used for differentiating tree species from other materials. The crown structure of tree species from Washington Park Arboretum is predicted based on LIDAR intensity. The classification accuracy for broadleaved and coniferous species is better using leaf-off data than using leaf-on data while in terms of the structure analysis, Leaf-on data resulted good accuracy than leaf-off data [117].

Wajid Ali, has presented a robust method for tree detection method based on classification and distance measurement from base of the tree and the tire of vehicle based on pixel ratio and the length of tire as a reference. The approach gives an automated navigation in forest environment for tree detection. The tree image is segmented into trunk and background objects using color and texture features. It is reported that segmentation of objects from forest images into tree trunk and background objects is a challenging task due to high
variations in illumination, effect of different color shades, non-homogenous bark texture, shadows and foreshortening [129].

Mersad Jelacic, has used unsupervised-learning methods for plant identification. Different existing clustering methods are performed on sugar-beet plants and weed plants without knowing what samples are in the clusters. They have implemented a mobile robot guided by a camera system to capture pictures of row of cultivated sugar-beet plants. Another camera system is taking pictures of plants, including both weed and crop. The plants are segmented and features are extracted for each plant. A classifier system is then trained to identify the plants. Features are used from three different datasets and tested for classification accuracy [87].

Ashok Samal et. al., have used three textural features obtained from the co-occurrence matrix. The nearest-neighbor method is used to discriminate between trees and their backgrounds. Three of the tree species namely, Japanese yew, Hicks yew and eastern white pine are considered in the study [12].

Lars Linsen has constructed computer graphics models for describing the fractal branching structure of trees typically to exploit the modularity of tree structures. The models are based on local production rules, which are applied iteratively and simultaneously to create a complex branching system. Tree growth is visualized for certain tree of particular Species of London. The tree height, diameter-at-breast-height, crown height, crown diameter and leaf area are
measured manually. The comparison between these parameters and local parameters of computer generated model is reported [81].

F. Cointault and B. Chopinet, have studied an analysis of color texture features in wheat head counting for yield prediction. The background objects such as soil, waste, sick leaves, safe leaves and stem are segmented from field under natural weather condition. The texture features are obtained using gray level co-occurrence matrix. The recognition accuracy in the range of 73%-85% is obtained [40].

Castillo P. and Melin P., have proposed a new approach for plant monitoring and diagnostics using fuzzy logic and fractal concepts. A set of fuzzy rules is used to represent the knowledge base The combination of fuzzy fractal approach has given efficient results for growth observation [21].

Helmut Dalitz and Jurgen Homeier, have provided a system in which biodiversity information of plants is made available to people through internet based systems. A system called Visual Plants is generated. The system provides digitized images of plants and their parts with all labeled information for identification of individual plants. The system is used on-site for the education of students and for taxonomic ecological research [53].

Zheru Chi et. al., have presented an approach for plant species recognition using bark texture features. The multiple narrow band signal of Gabor filters are used for characterizing plant bark. The
Gabor filter banks of high frequency band are trained by classifier [144].

D.S. Shrestha and B.L. Steward, have given a machine vision based techniques for measuring the growth stage of corn plant using video images. The vegetation part of corn plant is segmented using a truncated elliptical decision surface. The plant count is obtained by measuring total number plant pixels and their medians. The results are compared with manual expert and errors are reported [30].

D.G. Sena Jr et. al., have identified damaged maize plants by the fall armyworm using digital color images. Images of damaged and non-damaged maize plants are taken in eight different stages and in three different light intensities. The images are processed to create binary images, where the leaves are segmented from the other pixels. Then, images are divided into blocks and classified as damaged and non-damaged [34].

H T Sogaard and H J Olsen, have discussed computer vision methods for detection and localization of crop rows for small-grain crops images. A color video camera and computer are used for obtaining clips of field images. Instead of segmentation, the centers of gravity for row segments of the plant image are obtained. The orientation and the lateral position of the centre lines of the rows are obtained by weighted linear regression [46].

M.J. Aitkenhead et. al., have developed a visual method for discriminating crop seedlings from weeds for automatic weed
detection and weed control system in agriculture. The carrot plants are recognized from rue grass and flatten weed using morphological characters of leaf shape. The method is robust to variation in plant size. The self organizing neural network has given more than 75% accuracy without the plant description [84].

Paula J. Peper and E. Gregory McPherson have estimated the leaf area of trees using four methods. The methods are color digital image processing, LAI-2000 plant canopy analyzer, the CI-100 Digital Plant Canopy Imager and a logarithmic regression equation. The functional and physiological processes of urban forests are estimated based on these four methods. The result is better for trees without having much background [100].

Timothy W. Hindman, has described a fuzzy logic approach for segmentation and classification of images of plants. Digital images of grass, bare soil, corn stalks and wheat straw are processed to obtain two parameters such as average color values and constant indices. Rules based on individual color values yielded better classification results than rules based on color contrast indices. The classification accuracy is better for two-class than four class using color indices. The system is able to correctly classify over 90% of all plant regions, as plants and background regions [125].

J. Hemming and T. Rath, have provided a digital image analysis for identification of weeds in crops. The cabbage and carrots images of
plants from open fields are used for the study. The morphological and color features were extracted and fuzzy logic classifier is adopted. The average classification accuracy of 51-95% is reported depending on the growth stage. The approach is found to be efficient than any other plant identification system [62].

C.C. Yang et al., have designed an ANN based weed recognition system for precision spraying of herbicides in agricultural fields. A backpropagation neural network is developed for distinguishing young corn plants from weeds. The color indices of pixels are used as features. The corn plants and weeds are classified with an accuracy of 100% and 80% respectively [16].

George E. Meyer et. al., have studied a machine vision system based on classical image processing techniques for plant detection and identification. It is reported that key to successful detection and identification of plants as different species types is the segmentation of plants form the background. From the segmentation, an edge image is obtained, which contains shape feature information of canopy. The RGB formats are found to be very efficient for segmentation. The textural features of the plant are also obtained using gray image co-occurrence matrices [44].

N.D. Tillett et. al., have demonstrated a computerized spot treatment for cauliflower plant image analysis in the field. A charge coupled device is located at the vehicle and images of crop rows are captured. The segmentation is performed using image intensity, size and
geometric information to pick up crop plants that requires treatment [124].

A.J. Pérez et. al., have developed an image processing technique to generate weed maps. The broad leaf weeds are detected and separated from cereal crops under natural fields. The color and shape features are used for distinguishing crop, weed and soil [4].

Woebbecke et. al., have used shape feature analysis for discriminating between monocots and dicots. The plants used in the experiment are grown individually in pots and color information is used to separate target plants from the soil and residue background. It is reported that canopy shape feature do not change for dicots. But for plants with few leaves there is no assurance that any particular shape continues to work as the plant grows [132].

Marchant and J. A. Brivot R. have worked with infrared images instead of color images, which use only shifts in gray level for segmentation of plants in images. The height is used as an index of forest site quality (few researchers have developed). Machine vision application [88].

S. A. Shearer and R.G.Holmes., have developed a method for identifying plants based on color and texture characterization of canopy sections. Color co-occurrence matrices are computed in HSV color space. Thirty three color texture features are used by a discriminant analysis model to identify plants. Overall classification
accuracy of 91% is achieved for identifying seven common cultivars of nursery stock [111].

D. E. Guyer et al., have developed a machine vision system using image processing algorithms for identifying the juvenile stages of corn, soybeans, tomatoes, Johnson grass, Jimson weed, velvet leaf, giant foxtail and lamb squarters. Plants are grown in containers and images are obtained in a laboratory by setting with the camera and light source positioned directly above the plants. Spatial parameters of the plants are computed and used in a classification and identification [31].

The literature survey has reveals that the computers are used in many automation tasks related to plant domain. Still there are enormous applications connected to plant image recognition. We have observed plant identification and classification in various fields such as, agriculture, weed classification, plant growth analysis, horticulture, forestry biomass prediction and vegetable recognition. Some works are also reported on plant species recognition based on their parts such as leaves, flowers and bark. But, the plant identification with respect to full image is very much scarce. Hence, from the literature survey, to the best of our knowledge, it has been observed that no considerable work has been cited in the literature on the development of machine vision systems for identification, classification and retrieval of medicinal plants images in the Indian context. The automatic identification of medicinal plants with their
relevant information such as plant names in different languages and medicinal usage is very much necessary in the present era. The methodologies help the development of a search engine for the database of Ayurveda medicinal plants in information retrieval through plant snap. This is the motivation for taking up this work.

1.3 MOTIVATION AND PROBLEM DEFINITION

In order to popularize alternate system of medicine Ayurveda in this world, it is necessary that medicinal plants to be identified by common man. This system of medicine is being practiced since vedic period. The vedic literature has abundant information on preparation of medicines and treatment procedure for certain diseases like cancer, aids and the like. Since medicinal plants are basic to this system of medicine, the authors thought of taking leverage of technology in identifying and classifying the plants. This being the strong motivation for the work undertaken, the problem called Domestication, Identification and Classification of Medicinal Plants in the Indian Context is defined with specific objectives.

1.4 OBJECTIVES FULFILLED

The following are the objectives fulfilled as part of the research work undertaken.

- A text database for medicinal plants is created with information on each of the 500 medicinal plants. An algorithm is developed for retrieval of text information that contains description of plant features such as height, shape and color of leaf, flower,
seed and roots. This is an important contribution as there are no standard databases available.

- Suitability of segmentation techniques for separating images of medicinal plants from the cluttered background is proposed.

- Feature sets for the images of medicinal plants, namely, herbs, shrubs and trees are defined for the purpose of identification and classification. The feature extraction algorithms are devised.

- The artificial neural network parameters are determined for the purpose of classification of medicinal plants. The classification algorithms are devised.

- The Content Based Image Retrieval (CBIR) techniques, keeping precision and recall in the background, are tested for the database of images of medicinal plants. The retrieval algorithms are developed. An image based information retrieval is also developed in which information of the medicinal plants is reported based on the visual features of the plant and their corresponding parts.

1.5 CONTRIBUTION MADE IN THE WORK

The main contributions in this thesis includes the development of different, but interrelated modules, namely, database design, identification, classification and content based information retrieval of images of medicinal plants. The schematic block diagram of the proposed methodology is shown in Figure 1.4.
In India the medicinal plants are identified by their names and their characteristics. But there is no standard database of Indian medicinal plants available. For proper identification of medicinal plants, automation is necessary in the field of Ayurveda. Hence, an effort is made to create a database of medicinal plants to assist the people working with these plants. The database includes systematic description of 500 plants, their geographic distribution, details of different parts, their medicinal value, botanical and vernacular names and their images. The database has been stored in Datastore-1. We have developed a methodology for text information retrieval from the database of Indian medicinal plants. The taxonomical terminologies related to plants are given as query to the database. The information about the query medicinal plant is provided.

![Figure 1.4: Schematic Block diagram of the proposed methodology](image)

Since the images of the medicinal plants are important in their recognition in the real world, in addition to the text information the
images are also placed in the database. The images of different medicinal plants are captured using the digital camera, sony 3300 with 14 Megapixels from eight angles of view, under fixed distance and normal light conditions. The imaging distance for herbs, shrubs and trees images changes due to variation in height. The herbs and shrubs are captured from a distance of 2meters and trees are captured from a distance of 12meters. The images acquired from camera are of size 2048 X 1536 pixels. The images are resized to 256 X 256 pixels to reduce the computational and storage overheads during feature extraction and training. The images of the nearby available medicinal plants and their leaves are stored in the database.

The medicinal plants obtained from natural outdoor are not devoid of complex background. The plant images obtained contains cluttered backgrounds, namely, soil, stone, building, river, sky, weed and pot. There is a need to segment the plants from the background. The existing segmentation methods, suitable for images of medicinal plants are studied and their performances are obtained. The segmentation techniques based on namely, color, K-means clustering technique, level sets, threshold and watershed transform are considered in the work. The level set method is found to be more appropriate for the segmentation of images of medicinal plants.

In order to retrieve the medicinal plants from the database, it is necessary to obtain the necessary features from the image of the plants. We have considered height as one of the features in
discriminating herbs from shrubs and trees. Since the absolute height of a plant cannot be measured from images, we have developed a method where plants are distinguished based on relative dimensions of canopy and stem parts. We have devised a knowledge-based classifier with aspect ratio as feature for identification and classification of medicinal plants.

The plant classification based on height depends upon a number of factors, namely height, shape, width of canopy, cross-sectional area of stem and the like. The height and shape features of the plants are not constant and are found to vary with age, environment and soil characteristics. Hence, we have obtained new geometrical features in terms of canopy and stem parts. Due to these physiological limitations, leaf structures are used by humans beings for recognition of plants precisely. This is the motivation to take leaf characteristics as another feature to improve the classification results.

The medicinal plants exhibit heterogeneous texture in canopy (also called leafy mass) and stem parts. The medicinal plants are also recognized using color and textural arrangement of canopy. The leafy mass of a plant in turn depends on branching pattern of leaves with main trunk. We have used combination of both color and texture features for identification and classification of plants as herbs, shrubs and trees. The color histogram features, like number of valleys and peaks are observed in different color spaces. The edge and edge direction histogram uses edge patterns spreaded in canopy and stem
and gives spatial location of edges in different directions. A method based on color, edge and edge direction histograms are devised. The identification of medicinal plants is also analyzed in spectral domain. The Different shape descriptors, namely, Zernike moments, Fourier Descriptors, Generic Fourier Descriptors and Gabor texture descriptors are used. These features are scale, rotation and translation invariant. The developed methodologies are tested with Radial Basis Exact fit Neural Network (RBENN), minimum distance and Support Vector Machine (SVM) classifiers. For input image, plant, canopy and bark the features are extracted and classified into herbs, shrubs and trees and stored in Datastore-2.

Usually the Ayurveda practitioners, botanists, tribal and rural people identify a plant based on morphological characteristics of the leaves. There are two forms of leaves, simple and compound. A simple leaf has undivided blade, whereas compound leaf has divided blade, with more than one small leaflets attached to main or secondary veins. The leaves images of different medicinal plants species vary with base, apex and margin shapes. The base and tip angles are measured. In order to improve the classification accuracy, the leaf margin is also considered. Therefore, automated leaf image identification is attempted and a methodology is developed.

We have developed a methodology for content based image identification and information retrieval from the database of images of medicinal plants. Initially, Content Based Image Retrieval is adopted
using developed features and similarity measure. For the similarity measure, Euclidean distance, Absolute distance and Mean Square Error are adopted. The precision and recall rates are evaluated for different query image. It is reported that mean square error gives the more precision and recall rates for Indian medicinal plant images. From the retrieved images, the most relevant image is retrieved and given as input to Image Based Information Retrieval. For the retrieved image the text database is accessed from the Datastore-1 and detail information about the retrieved plant such as morphological details of full plant and its parts, medicinal usage are reported.

1.6 ORGANIZATION OF THESIS

The remaining part of the thesis is organized into seven chapters and the details of each chapter are given as under.

**Chapter 2** presents the design and development of a database of medicinal plants. A text based retrieval methodology for the domestication of medicinal plants is proposed.

**Chapter 3** discusses the suitability of different image segmentation techniques for images of medicinal plants. A quantitative evaluation of segmentation accuracy for canopy and stem parts are proposed.

**Chapter 4** gives the features and feature selection for images of medicinal plants. The robust features for the identification of medicinal plants are suggested.
Chapter 5 gives a methodology for identification and classification of images of medicinal plants and its parts such as leaves using different segmentation techniques. A geometrical and texture features based identification of images of medicinal plants is described. The viability of classifiers and different segmentation techniques is given in this chapter.

Chapter 6 provides the details of overall content based image retrieval of images of medicinal plants based on visual features. It presents the development of an interface for Image Based Information Retrieval system for retrieving the required species of the medicinal plants along with information based on similarity measures.

Chapter 7 gives a summary of the research work undertaken and the scope for future work leading to possible improvements in the proposed methodologies.

Appendix A gives the medicinal plants’ images, Appendix B gives leafy mass and stem images and Appendix C presents leaves’ images. Appendix D gives the color lookup table. At the end of the thesis, an exhaustive bibliography is given.