Herbal Leaves Image Clustering via K-Means

Dr. A. Anushya

Assistant Professor, Department of Computer Applications

Abstract: India is a pervasive origin for the medicinal plants which fruitages traditional medicine. Modern research in computer science is developing an automatic recognition system of plants by its leaf images. Features decide individual leaves. In this research, features are extracted from herbal leaf images and clustered by Gray Level Concurrence Matrix and K-means respectively. Both algorithms are executed on MATLAB and reached 86.96% accuracy.
Keywords: Gray Level Concurrence Matrix, K-means, Image clustering, herbal leaf images

I. INTRODUCTION

Traditional Medicine is time-tested and still caters to the health needs of the civilization and affords health care from side to side prophylactic treatment and renovation [1]. All and sundry distinguish subsidy of herbal for our life. Herbals are instantaneous remedy. Every kitchenette resides not only food stuffs, also includes some important herbals. Some of the herbals are used in our diet regularly to enhance the flavour to various dishes. Apart from enriching the taste and aroma of the dish, they are used as accepted cures for countless health welfares. These herbs can be cultivated in the backyard, terrace, and porch or in a pot also. This research concentrates on these herbal which are used daily and keep in our kitchen cabinet. They are Basil leaves, Coriander, Mint, Curry leaves, Purple fruited pea egg, Betel and Neem. At the beginning images are captured by using a mobile camera or a digital camera. A complete database is assembled by packing every meticulous information of every image. Next, algorithm extracts significant features from the images which are used for identical of the same in changed views. Identification of features consists of staged filtering methodology. The images can be categorized to their respective species based on the clustering results.

The rest of paper is catalogued as follows: Section 2 bulletins the current research done in this field. Section 3 demarcates the data source. Section 4 and Section 5 bounces the K-Means clustering for images and experimental results respectively. Section 6 concludes the paper and discusses the Future Works and Section gives the References.

II. RELATED WORKS

In this section, a meticulous review of the studies on herbal medicine image recognition and retrieval. Overall speaking, there is pint-sized study on herbal image retrieval, while there are some on herbal image recognition. Therefore, we mainly introduce the previous work on herbal image recognition, and simply review image retrieval in the computer vision community. Li [2] conducted a research considered 5 herbal medicine categories using low-level features for medicine recognition, such as shape, color and texture features. As the herbal medicine category becomes more, Tao *et al*. [3] found that color and shape features are not reliable because many herbal medicine categories have the similar color and shape, while different medicine categories have different textures. Therefore, they proposed to use various texture features from different aspects to describe the herbal images, and obtained promising recognition precision on 18 herbal medicine categories. I.Kiruba Raji et.al., [4]examined various object detection techniques for segmenting leaves based on color, shape and texture. Features like local adaptive mean color, evidence based color model, color histogram techniques were used. Boundary structure model was used to detect the leaves based on boundary descriptors of an image and Chan-Vese algorithm was used to segment the leaves from complex background. To extract leaves from texture background, edge focusing algorithm was used. From our experimentation analysis, shape is the powerful characteristics of segmenting leaf images and Chan-Vese algorithm provided better results compared to other techniques without affecting the leaf colors, texture etc. Xin Sun et.al., [5] used the Convolutional Neural Network (CNN) for Chinese herbal medicine image recognition and retrieval and practiced the softmax loss to optimize the recognition network; then for the retrieval problem, fine-tune the recognition network by adding a triplet loss to search for the most similar medicine images. To appraise this method, a public database of herbal medicine images constructed with cluttered backgrounds, which has in total 5523 images with 95 popular Chinese medicine categories. Experimental results showed that the method can achieve the average recognition precision of 71% and the average retrieval precision of 53% over all the 95 medicine categories, which are quite promising given the fact that the real world images have multiple pieces of occluded herbal and cluttered backgrounds. Roopashree S.et.al., automated identification and classification of Ayurveda leaf would be extraction of SIFT feature and implementing SVM classifier for better accuracy IJY on using Traditional Indian Medicine database. Also, portability could be achieved through different smartphone platforms.

From the exceeding literature analysis, it is absolutely amplification the helpfulness of herbal leaf image identification and classification. In the similar style, real images are collected and then images are clustered. But the unique of this study is focused on herbal leaves which are used daily life and preserved in our kitchen such as Basil leaves, Coriander, Mint, Curry leaves, Purple fruited pea egg, Betel and Neem. Data gathering and the particulars of the images are expounded in subsequent section.

III.   DATA SOURCE

In the area of herbal medicine image recognition, no such database is open to researchers, thus it is tough to appraise dissimilar methods with the same standard. Construction of a database with herbal leaf images for research is done at first. This section first announces how to construct the database in detail. To construct the database, we select 7 common herbal leaf categories whose number varies from 25 to 50 for each category. Finally, we catch 307 herbal medicine images with 7 categories in total. Table 1 have the funds for instincts into the number of artworks used for each different species sheltered. Figure 1 shows specific images in dataset, and we can see clearly that most of these images have multiple pieces of mutually occluded herbal and various backgrounds, which are more appropriate for research and applications. The database will be available online soon. Images were taken by Samsung J7 mobile Camera. Samples of collected images are as presented in Figure 1.

Table 1: The Classes of images used in the Dataset

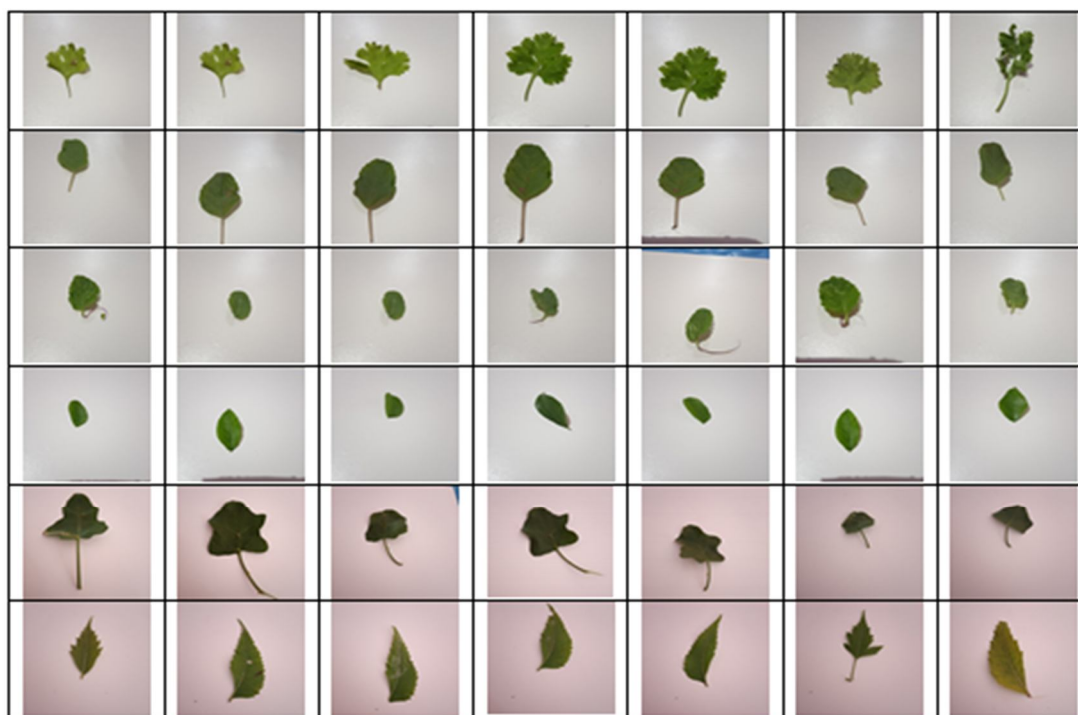| Species No.	Name of the species	No. of images
1.	Basil leaves	47
2.	Coriander	50
3.	Mint	50
4.	Curry leaves	49
5.	Purple fruited pea egg	45
6.	Betel	25
7.	Neem	41
Total		307

Figure 1: Sample of herbal leaf images in database.

A. Image Clustering Via K-Means

The self-effacing formula for image clustering principally involves of four steps such as Image acquisition, Image processing, Features are extraction and Clustering. Initially the images are acquired and a database is created. Then, Image processing is done to images in order to improve image quality. After that, desired features are extracted and selected the relevant features and input to the model. Lastly a K-means is applied to cluster the images into groups. To begin with image processing, visually examined the images that used in this experiment and classified them according to their features. The marine products having good shape, large size, high intensity, high softness and no defects were patented as of the finest class.

Gray Level Concurrence Matrix is used for texture features. In the GLCM number of rows and columns are same as number of gray levels of that image. GLCM is used for number of applications in image processing. Images in database are resized and convert both images into Grayscale. Then the Converted Grayscale images are transformed to binary image. In binary image, count the number of pixels concealed by the images by using GLCM feature. Next, Calculate the GLCM feature of both images and find the difference in area and apply the Canny edge detection method on grayscale images. Afterward Extract the hue, saturation and intensity a from the uncropped test image. Now average of difference in GLCM feature, edge and color histogram is obtained. The above methods are repeated for all the images in the database. As a final point, K-means clustering is applied on images [8]. The steps of K-Means algorithm works is as follows:

1) Specify number of clusters K.
2) Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3) Keep iterating until there is no change to the centroids.

IV. EXPERIMENTAL EVALUATION

Herbal leaves identification is not a soothing task by eyes. But it was very multipart process because most of the leaves have analogous morphological features such as shape, color and texture. In this section, results are pondered about the observes. The database contains 360 real images. The resolutions of images were fixed to 89 X 142 pixels for the sake of workable computational speed. GLCM and K-means have been executed in MATLAB for real marine products images. Features such as color, shape, size and texture are extracted by GLCM and clusters of images are achieved by K-Means correspondingly. The image processing steps with an image is exposed in Figure 2, Figure 3, Figure 4 and Figure 5. The number of images are grouped by K-Means bestowing to its clusters is tabularized in Table 2. This method categorizes images into nine clusters in an efficient way and 83% of accuracy is reached as shown in Table 2.

Figure 2: Original image

Figure 3: After background removed

Figure 4: Gray scale image

Figure 5: Saturated image

Table 2: Experiment analysis of K-means on images

Name of the species	No. of images in Cluster	Accuracy (%)
Basil leaves	41	87.23
Coriander	50	100
Mint	43	86.00
Curry leaves	37	75.50
Purple fruited pea egg	39	86.66
Betel	25	100
Neem	33	73.33
Overall Accuracy		86.96

V.   CONCLUSION

This research efforts the image clustering of herbal leaf images. Because the leaf images are used in our diet mostly to enrich taste as well as aroma. The real herbal leaf are collected and stored in database at first. Then GLCM is smeared on images to extract features. Finally K-means is executed and obtained clusters of herbal leaf images. These algorithms are executed in MATLAB and gained accuracy is 86.96%. In future after identifying the shapes of leaves, some of the shape representation techniques may be introduced for identifying the similarity of shapes. With these shapes, analyzing leaf vein, margins and tip and base of leaves will also be considered because families of leaves include the characteristics of vein, base and tips.

REFERENCES

[1] Nushrath Roshana Mansoor et.al., "A Literature Review on Medicinal Plants That are being Used in Traditional Medicine for the Management of the Snake Bites in Sri Lanka", Asian Plant Research Journal, Volume :1, Issue 1, Pp : 1-18, 2018.

[2] Li Z., "Chinese herbal medicine feature extraction and identification system," in Master Thesis, 2013. [Google Scholar]

[3] Tao Peng et.al., " "Construction of Traditional Chinese Medicine Database Based on ISIS Base", IEEE 2010.

[4] I.Kiruba Raji et.al., " An Analysis of Segmentation Techniques to Identify Herbal Leaves from Complex Background" Procedia Computer Science 48 (2015) 589 – 599, 2015.

[5] Xin Sun et.al., "Chinese Herbal Medicine Image Recognition and Retrieval by Convolutional Neural Network", PLoS One. 2016; 11(6): e0156327.

[6] Roopashree S.et.al., "Indian Herbal Leaf Identification using Artificial Intelligence Approach", International Journal of Science, Engineering and Technology, 2017.

[7] Aitwadkar et.al., "Identification of Indian Medicinal Plant by using Artificial Neural Network", International Research Journal of Engineering and Technology (IRJET) , Volume: 05 Issue: 04 , Apr-2018.

[8] Simone Marini et.al., , "Tracking Fish Abundance by Underwater Image Recognition", Scientific Reports", 2018.

[9] Muhammad Azfar Firdaus Azlah et.al., "Review on techniques for plant leaf classification and recognition" Computers, 2019.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)