



Research Manuscript Title

**CRUSTOSE USING SHAPE FEATURES AND COLOR HISTOGRAM WITH K-
NEAREST NEIGHBOUR CLASSIFIERS**

P. Keerthana¹, B.G. Geetha², P. Kanmani³

*PG Scholar¹, Professor², Assistant Professor³,
(1,2,3) Department, Department of Computer Engineering²,
(1,2,3) K.S. Rangasamy College of Technology, Tiruchengode, Tamilnadu, India.*

Corresponding author E-Mail-ID: Keerthibe13@gmail.com

March – 2017

www.istpublications.com

Crustose Using Shape Features and Color Histogram with K-Nearest Neighbour Classifiers

P. Keerthana¹, B.G. Geetha², P. Kanmani³

*PG Scholar¹, Professor², Assistant Professor³,
(1,2,3) Department, Department of Computer Engineering²,
(1,2,3) K.S. Rangasamy College of Technology, Tiruchengode, Tamilnadu, India.*

Corresponding author E-Mail-ID: Keerthibe13@gmail.com

ABSTRACT

Automated System for Turmeric leaf Recognition can be used to classify turmeric leaf into appropriate taxonomies. These data are used by botanists, industry person, food engineer and physicians. In this work, system capable of identifying various turmeric leaves using images has been developed. Mobile application was developed to allow user to take more pictures of turmeric leaves and upload them to server. The server run pre-processing, feature extraction technique in the image before pattern matcher compare the information from this image with the one in database. The different features that are extracted are mean, median, standard deviation, mode, skewness and color histogram. A k-Nearest Neighbour cluster classifier was implemented and tested on 650 turmeric leaves belongs to 32 different genus of various turmeric leaves. An accuracy of 84% was obtained. The system was further enhanced using information obtained from color histograms which increase the accuracy to 88%. Furthermore, system is simple to use and highly scalable.

Keywords: pattern recognition; shape features; color histogram; k-nearest neighbour.

1. INTRODUCTION

Recognition of turmeric leaves is not a simple task for botanists. For machines however, the same speaks to an immense and complex computational exertion. Humans can undoubtedly recognize distinctive objects, determine their sizes, shapes, composition and hues and comprehend the relationships between them using their senses. It is usual and necessary to perform image processing techniques to extract visual information and compare them to an existing set of data[1]. Recognition system is that object of same kind will share some similar visual properties which can be captured and thereby allow the system to be feasible. Evaluate the effectiveness of the method used while using a dataset with a fair amount of turmeric leaves per genus[2]. This system allows the addition of new genus to the database without much effort.

2. RELATED WORK

Satti et. al. described a recognition system that used color and shape information to produce accuracy of 94% with artificial neural network and 86% with k-Nearest Neighbour classifier on the flavia dataset. Chaki et al. proposed a new method characterizing and recognizing turmeric leaf using combination of shape and texture

features. Filter was used to model the texture of turmeric leaf and shape was captured using curvelet transform coefficient with invariant moments[3]. System tested using two neural classifiers: neuro-fuzzy controller and feed-forward back-propagation multi-layered. Best accuracy obtained was 88% for 950 turmeric leaf images consisting of 32 different species. Easy to compare because each one uses a different dataset. Larese et al described classification of three legume species. Using images of turmeric leaf that were cleared using chemical process and increase the accuracy to 89% at expense of time and cost [4].

Table 1. Summary of Related Works

Authors	Features	Classifiers	Accuracy	Dataset	Training	Testing	Species	Samples
Arun et al	Texture	SVM, kNN	95%	250	175	75	5	50
Uluturk et al	Shape	PNN	93%	1280	1120	160	32	40
Zhang et al	Shape	MLLDR	94%	750	500	250	50	15
Kadir et al	Shape, Texture	PNN	96%	6900	5700	1200	60	95
Wang et al	Contours, Shape	kNN	97%	1125	750	375	15	75
Chaki et al	Texture	ANN	88%	930	620	310	31	30
Prata et al	Color, Texture	PDA	90%	1120	725	400	20	70

An automated system for the recognition of medicinal leaves was developed by Arun et al. They used GTSDM, grey texture and LBP operators as features for recognition. Without using any pre-processing technique, an accuracy of 95% was obtained using a dataset of 250 different images from ten species.

Accuracy of 72% by using 64 features are derived from shape only developed by Amin et al. DHNG for pattern recognition and K-Nearest Neighboring for pattern classification. Database consisting of 1600 images from different species. Database of 70% are used as training set and rest as testing set[5].

Several features are extracted such as mean, median, Skewness and Intensity are extracted. Framework uses a ten-fold cross validation technique and 92% accuracy was obtained. Using Principal Component Analysis as a feature selection method improves the accuracy about 2%.

3. METHODOLOGY

The proposed system is divided into two phases: Client and Server phases. The client side is a software that allow a user to upload the picture of a turmeric leaf into the server. An overview of system describes as follows.

The server uses shapes and color information to compare the information extracted from the database of the turmeric leaf images with newly acquired one and uses a k-Nearest Neighbour algorithm to find out matcher[6].

A. Dataset

Pictures of turmeric leafs were taken from nearby locations. The database consists of 20 different pictures for each genus, for 32 different turmeric leaf genus. The picture have been taken in daylight with a smartphone camera having a resolution of 1980*1024.

B. Pre-Processing Steps

Pre-processing involves a collection of techniques that are used to improve quality and visual appearance of an image. Pre-processing operations are image reconstruction and image enhancement. Pre-processing refers to initial processing of turmeric leaf image to perform the smoothening, filtering, noise removal applied for improvement of the quality turmeric leaf image[7].



Figure. 1 Original Image



Figure. 2 Grey Scale Image

1) Grey Scaling:

The image is converted into grey scale since proposed system needs shape information of turmeric leafs. Converts RGB images to grayscale that eliminate hue and saturation information retaining the luminance.

2) Thresholding:

Threshold operation is performed to obtain the binary image. Grey scale image are converted to binary turmeric leaf image. The binary image are represented in the background as black. Turmeric leaf Image will represent in black and white.

3) Edge Extraction:

The contours in the turmeric leaf images were extracted. Edge detection is a technique for finding the boundaries of objects. It works by detecting brightness. Used for segmentation and data extract in image processing and machine vision.

4) Edge Filtering:

Filter is a nonlinear smoothing which remove noise and reduce edges of turmeric leaf. Filter will make comparison with two pixel and decide the better pixel and replace the old pixel with new one by the mean value. Using histogram for the filtered image, analyzing color which present in the turmeric leaf is easy for finding various features of turmeric leaf. Each turmeric leaf have a different color it will be analysed for the further classification[8]. Pores present in the turmeric leaf are of different colors. It takes each pixel in the turmeric leaf image are evaluated by its neighbors to decide whether or not it is representatives of its surroundings. Replacing the pixel value with median of other pixel value that replaces with median of those values.

C. Feature Extraction

1) Morphological Characteristics:

The morphological characters of a turmeric leaf specimen are studied under dissection or stereomicroscope. Type of thallus, its shape and size can be learnt from this examination. In apothecia, shape (round, stretched and lirellate), size, mode of attachment, color and texture of apothecia margin with disc, presence or absence of pruina on disc, shape of disc (convex or concave) are necessary characters to observe. Such structures are called unorganized ascocarp or fruiting body. Color of surface, presence of Pores, presence or absence of rihizines, color, distribution, and branching abundance are noted[9].

2) Mean:

Mean is the "average" where add all the numbers and then divide by the numbers. The dimension becomes 1 while the sizes of all other dimensions remain the same.

3) Standard deviation:

Output pixel contains the standard deviation of neighborhood around the corresponding pixel for input image A. A can have any dimension. The output image A1 is the same size as the input image A.

4) Color Histogram:

A color histogram is compute for a cropped part of the turmeric leaf image since if the whole image is used, white spaces surrounding the turmeric leaf would affect the histogram. To crop the central part of the turmeric leaf Image, the length and width of the bounding box are used.

Table.2 Filtered Image with Histogram

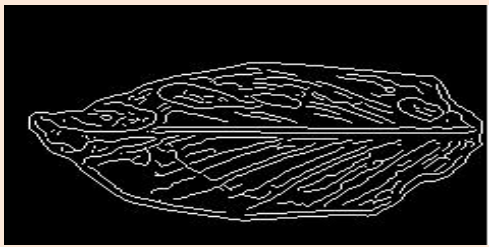
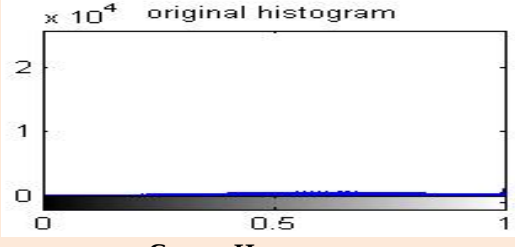
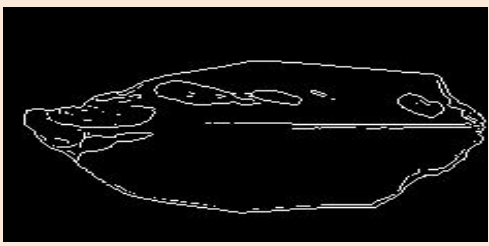
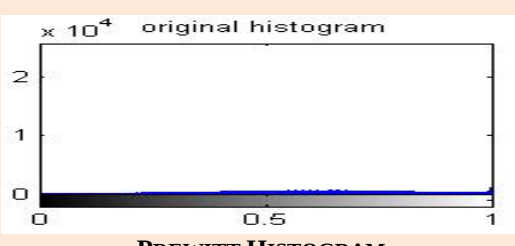
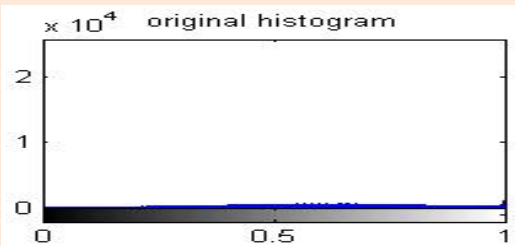
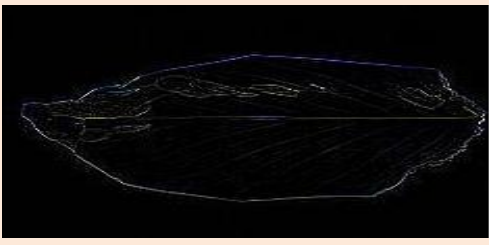
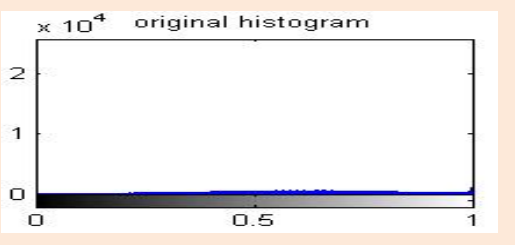
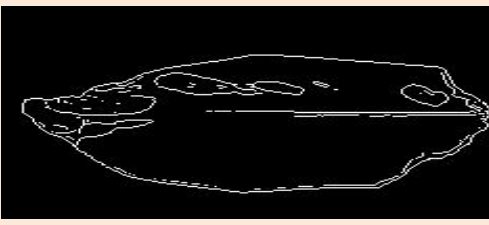
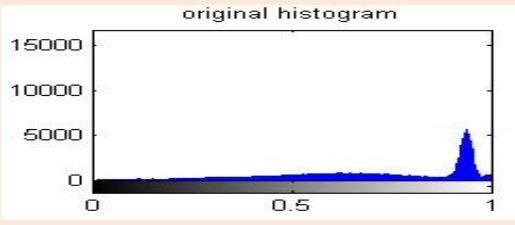
 CANNY IMAGE	 CANNY HISTOGRAM
 PREWITT IMAGE	 PREWITT HISTOGRAM
 SOBEL IMAGE	 SOBEL HISTOGRAM
 LAPLACIAN IMAGE	 LAPLACIAN HISTOGRAM
 WIENER IMAGE	 WIENER HISTOGRAM

Table.3 Feature Extraction from various Turmeric leaf Image

Images	Eccentricity	Entropy	Correlation	Mode	Contrast	Solidity	Convexity
Image 1	187.50	75.95	201	255	12,10,13	-0.698	5.0352e3
Image 2	171.58	69.02	161	255	74,71,57	-0.081	4.7648e3
Image 3	158.66	78.88	158	255	84,104,160	-0.108	3585.810
Image 4	159.09	74.09	150	255	107,118,108	0.070	5490.635
Image 5	163.96	69.29	159	255	74,169,61	0.010	4801.743

5) K-Nearest Neighbour Classifier:

All the ratios are normalized to a value between 0 and 1 before any comparison is made[10].

Stage 1: The new values for ratio are normalized. New turmeric leaf is compared to each turmeric leaf in the training set one by one. The sum of Euclidean distances between the new turmeric leaf and those in database are calculated[11]. The three closest results are obtained. Each ratio is used as feature in the KNN classifier.

Stage 2: If the result set from stage 1 consists of Different genus, the color histogram of new turmeric leaf is compared to those from result set. To analyze Correlation coefficient is calculated. Value will lies between 1 and -1. Value closes to 1 indicates high Positive correlation, which means two images are similar. Closest match are calculated using KNN algorithm[12].

4. RESULTS

The method of testing used is, every photo of the turmeric leaf in the database as input image to the system, Compare it to all the other turmeric leaf and calculate the percentage accuracy of the system[13]. This techniques has the advantage of testing all the turmeric leafs in the database rather than small percentage. Every time the system applies the matcher to a turmeric leaf, it will create a record in CSV file with the actual genus name. Particular, 100% of accuracy for ten different types of turmeric leafs. The overall accuracy at the first stage was 85%. Color Histogram matching operation maintained in the results from first stage and accuracy rises to 88% respectively.

Possible to obtain a high value for classification accuracy by using a relative large number of genus species but with only a small number of sample genus per species. Demonstrate how the accuracy varies on number of genus and number of turmeric leafs. The accuracy obtained is compared with existing works. KNN expected to run faster than comparable approaches using probabilistic neural networks or support vector

machines. Expect of varying the number of genus and varying number of turmeric leaves had not sufficiently tackled.

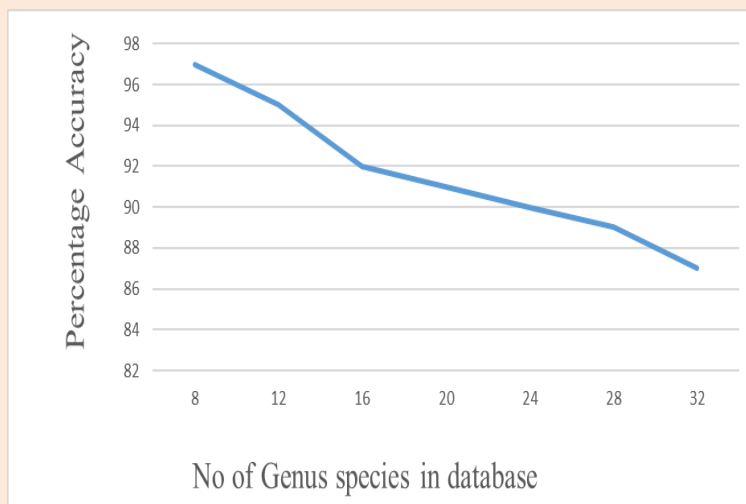


Figure. 3. Effect of increasing number of genus species on accuracy.

Notice that with only eight species, have an accuracy of 97% but this slowly drops to 87% with 32 different genus. Recognition accuracy goes down when there are more variety in dataset. The accuracy is going down very slowly and it is still high with 32 different species.

Number of turmeric leaves increases from 5 to 20, the accuracy of the KNN classifier rises from 70% to 80%. The overall accuracy (KNN + color histogram) follows a similar trend but there is an increase of approximately 1% for each additional set of 5 turmeric leaves that is added to database after the first 10 turmeric leaves.

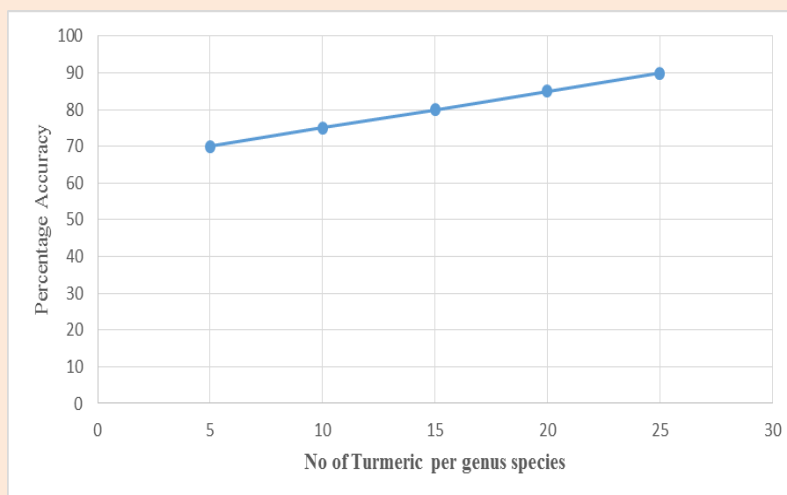


Figure.4 Effect of increasing number of turmeric leaves on accuracy

5. CONCLUSION

Demonstrated an approach to classify genus into appropriate species using images of their turmeric leaves. A high resolution camera was used to take pictures of 32 different species of genus. For each genus, 20 different turmeric leaf images are captured. The images are pre-processed and numbers of features are extracted. Each turmeric leaf image are compared with other turmeric leaf images in the database. The accuracy obtained was 84% at the first stage. The next stage of using information from color histogram in order to different more features. Now the recognition accuracy obtained was 5%. Increase number of species leads to small decrease in the accuracy but increase number of turmeric leaves beyond the threshold of fifty had no significant impact on overall accuracy. The main difficulty in work was need to take all the photos in daylight which the accuracy could be affected. In future, create a system which is robust to light variation, create elaborate dataset and include various features.

REFERENCES

- [1] M. Stricker and M. Orengo, Similarity of color images. *In SPIE Conference on Storage and Retrieval for Image and Video Databases*, volume-III, pp. 381-392, Feb. 1995.
- [2] O. Miljkovic. Image Pre-Processing Tool. *Kragujevac J. Math.* 32 pp. 97-107, 2009.
- [3] T. Cervinka, and I. Provaznk. Pre-processing for Segmentation of Computer Tomography Images. *The Faculty of Electrical Engineering and Communication Brno University of Technology, Department of Biomedical Engineering*, 2007.
- [4] M. Khalil and M. Bayoumi, A dyadic wavelet an invariant function for 2D shape recognition, *IEEE Trans.*
- [5] Yuan Tian, Multiple Classifier Combination for Recognition of Wheat Leaf Diseases Intelligent Automation and Soft Computing, Vol. 15, No. X, pp. 1-10, 2009
- [6] Xiaoyi Song, Yongjie Li, WufanChen, "A Textural Feature Based Image Retrieval Algorithm", in *Proc. of 4th International conference on Natural Computation*, Oct. 2008.
- [7] Andrea Baraldi, Flavio Parmiggiani, "An Investigation of Textural Characteristics Associated with Gray Level Co-occurrence Matrix Statistical parameters", *IEEE Trans. On Geoscience and Remote Sensing*, vol. 33, no. 2, COM-28, March 1995
- [8] G. Chen and T. D. Bui, Invariant Fourier-wavelet descriptor for pattern recognition, *Pattern Recognition*, vol. 32, pp. 1083_1088, 1999.
- [9] Abdul Kadir Experiments Of Zernike Moments for Leaf Identification, *Journal of Theoretical and Applied Information Technology*, Vol. 41, No.1, and 15 July 2012
- [10] A.H. Kulkarni et al "A Method to Classify Plants Using RBPNN for *IJLTET*", Dec 2012.
- [11] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu – Xuan Wang Yi – Fan Chang A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network, *IEEE 7th International Symposium on Signal Processing and Information Technology*, 2007
- [12] K Jalja, C. Bhagvati, B.L. Deekshatulu, A. K. Pujari, "Texture Element Feature Characterizations for CBIR", in *Proc. of Geoscience and Remote Sensing Symposium (IGARSS 05)*, Vol. 2, July 2005.
- [13] Jing Zhang, Gui-li Li, Seok-wum He, "Texture-Based Image Retrieval by Edge Detection Matching GLCM", in *Proc. of 10th Int. conference on High Performance Computing and Communications*, Sept. 2008.