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

[1]P. Keerthana, [2]B.G. Geetha, [3]P. Kanmani

[1]PG Student, Department of Computer Engineering, K.S. Rangasamy College of Technology, Tiruchengode [1]

[2]Head of Department, Department of Computer Engineering, K.S. Rangasamy College of Technology, Tiruchengode [2]

[3]Assistant Professor, Department of Computer Engineering, K.S. Rangasamy College of Technology, Tiruchengode [3]

Abstract: Automated System for Lichen Recognition can be used to classify lichen into appropriate taxonomies. These data are used by botanists, industry person, food engineer and physicians. In this work, system capable of identifying various lichens using images has been developed. Mobile application was developed to allow user to take more pictures of lichens 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 Neighbor cluster classifier was implemented and tested on 650 lichens belongs to 32 different genus of various lichens. An accuracy of 84% was obtained. The system was further enhanced using information obtained from color histograms which increase the accuracy to 88%. Furthermore, our system is simple to use and highly scalable.

Keyword: Pattern Recognition, Shape Features, Color Histogram, k-Nearest Neighbor.

1. INTRODUCTION:

Recognition of lichens 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. 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 lichens per genus. This system allows the addition of new genus to the database without much effort.

2. RELATED WORKS:

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 lichen using combination of shape and texture features. Filter was used to model the texture of lichen and shape was captured using curvelet transform coefficient with invariant moments. System tested using two neural classifiers: neuro-fuzzy controller and feed-forward back-propogation multi-layered. Best accuracy obtained was 88% for 950 lichen 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. Obtained a relative high accuracy of 85% using PDA approach. Using images of lichen that were cleared using chemical process and increase the accuracy to 89% at expense of time and cost.

3. METHODOLOGY:

The proposed systems are divided into two phases: they are Client and Server phases. The client side is software that allows a user to upload the picture of lichen 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 lichen images with newly acquired one and uses a k-Nearest Neighbour algorithm to find out matcher.

3.1 Dataset:

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

3.2 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, image restoration and image enhancement. Pre-processing refers to initial processing of lichen image to perform the smoothening, filtering, noise removal applied for improvement of the quality lichen image.

Figure.1. Original Image

3.2.1 Grey Scaling:

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

Figure 2. Grey Scale Image

3.2.2 Thresholding:

Threshold operation is performed to obtain the binary image. Grey scale image is converted into binary image. The binary image is inverted to represent the background as black. Image will represent in black and white.

3.2.3 Edge Extraction:

The contours in the lichen 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.

3.2.4 Edge Filtering:

Filter is a nonlinear smoothing which is used to remove impulsive noise and reduce blurring edges of lichen. Filter will make comparison with two pixel and decide the better pixel and replace the old pixel with new one by the mean value.

3.3 Feature Extraction:

3.3.1 Morphological Characteristics:

The morphological characters of a lichen 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 prunia on disc, shape of disc (convex or concave) are necessary characters to observe. Such structures are called unorganized ascocarp or fruiting body. They also note color of surface, presence of pores, presence or absence of rihizines, color, distribution, and branching abundance.

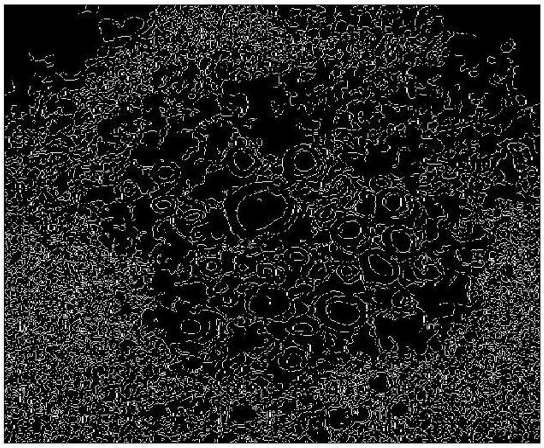
3.3.2 Color Histogram:

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

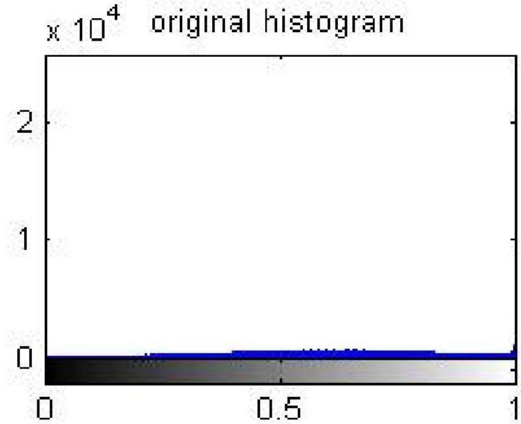
3.3.3 K-Nearest Neighbour Classifier:

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

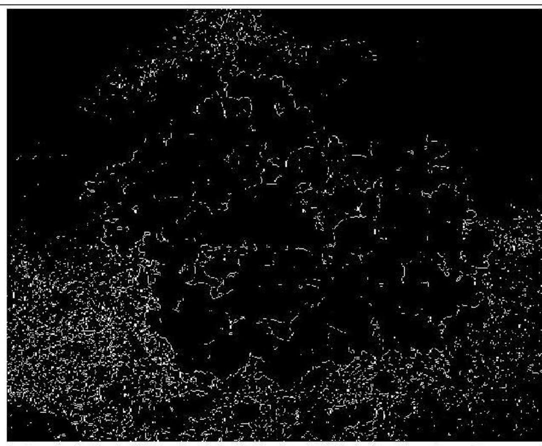
Stage 1: The new values for ratio are normalized. New lichen is compared to each lichen in the training set one by one. The sum of Euclidean distances between the new lichen and those in database are calculated. The three closest results are obtained. Each ratios are used as a feature in KNN classifier.



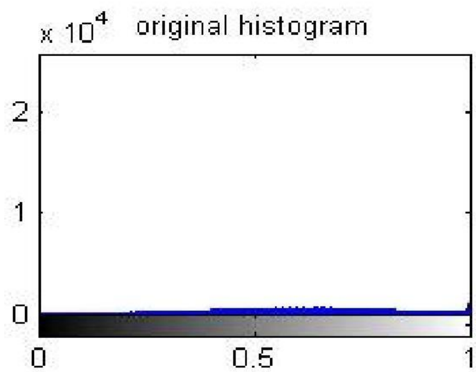
Canny Image



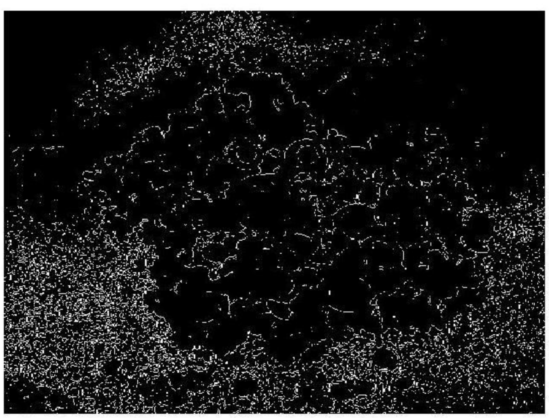
Canny Histogram



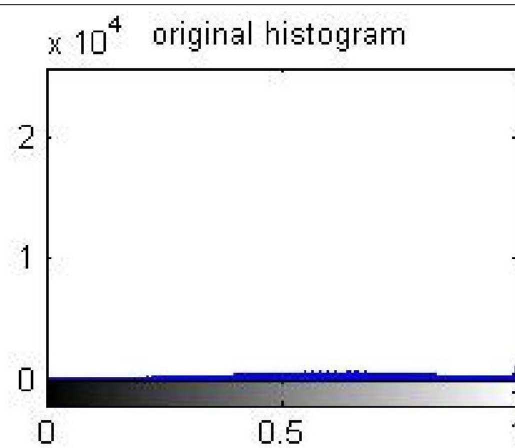
Prewit Image



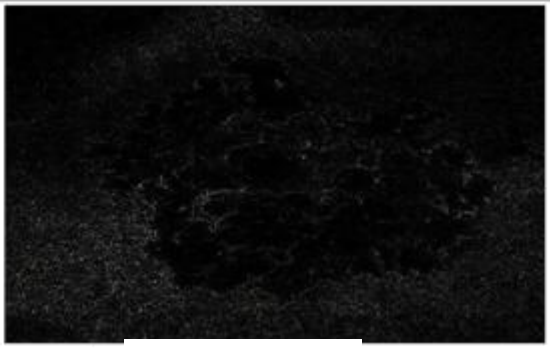
Prewit Histogram



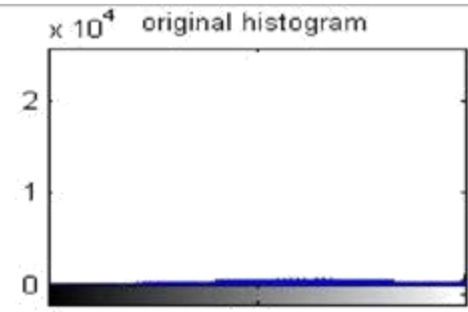
Sobel Image



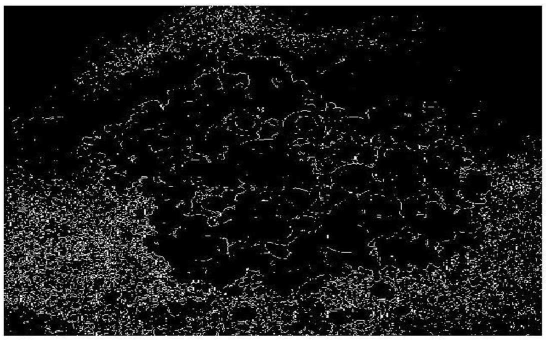
Sobel Histogram



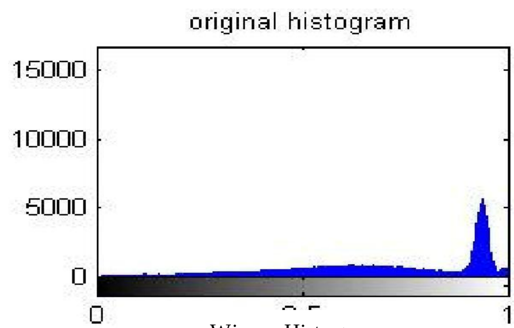
Laplacian Image



Laplacian Histogram



Wiener Image



Wiener Histogram

Stage 2: If the result set from stage 1 consists of different genus, the color histogram of new lichen is compared to those from result set. To analyse 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.

4. RESULTS:

The method of testing used is, every photo of the lichen in the database as input image to the system, compare it to all the other lichen and calculate the percentage accuracy of the system. This technique has the advantage of testing all the lichens in the database rather than small percentage. Every time the system applies the matcher to lichen, it will create a record in CSV file with the actual genus name. Particular, 100% of accuracy for ten different types of lichens. The overall accuracy at the first stage was 85%. Color Histogram matching operation maintained in the results from first stage and accuracy rises to88%respectively.

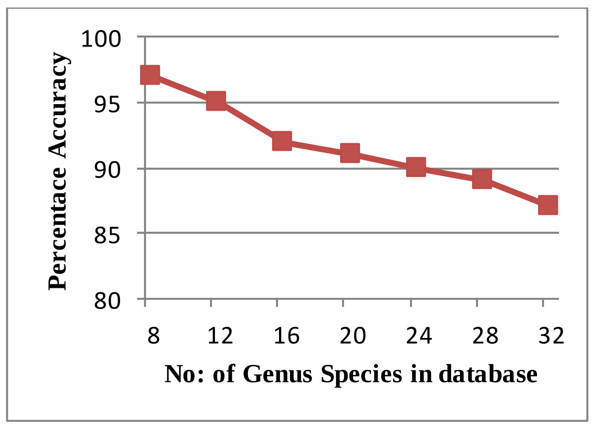


Figure 3: Effect of increasing number of genus species on accuracy

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 lichens 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 lichens that is added to database after the first 10 lichens.

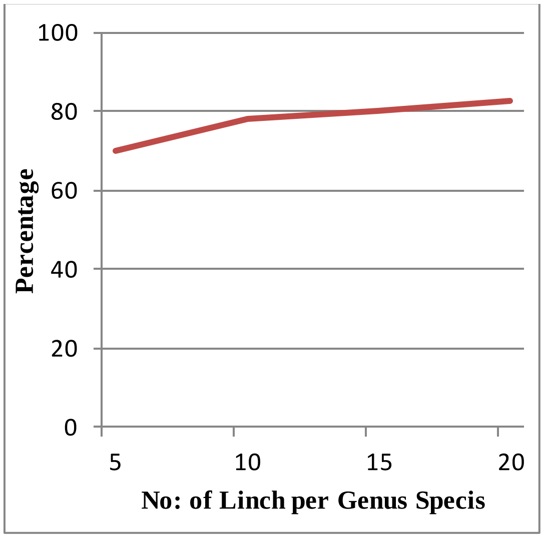


Figure 4: Effect of increasing number of lichens on accuracy

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 lichens. 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 lichens had not sufficiently tackled.

5. CONCLUSION:

Demonstrated an approach to classify genus into appropriate species using images of their lichens. A high resolution camera was used to take pictures of 32 different species of genus. For each genus, 20 different lichen images are captured. The images are pre-processed and numbers of features are extracted. Each lichen image are compared with other lichen 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 lichens 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 In SPIE Conference on Storage and Retrieval for Image and Video Databases III, volume 2420, pp. 381-392, Feb. 1995.
2. O. Miljkovi´c. Image Pre-processing Tool. Kragujevac J. Math. 32 pp. 97-107, 2009.
3. T. Cervinka, and I. Provazn´ık. 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. G. Chen and T. D. Bui, Invariant Fourier-wavelet descriptor for pattern recognition, Pattern Recognition, vol. 32, pp. 1083_1088, 1999.
5. M. Khalil and M. Bayoumi, A dyadic wavelet an invariant function for 2D shape recognition, IEEE Trans.
6. Yuan Tian Multiple Classifier Combination for Recognition of Wheat Leaf Diseases Intelligent Automation and Soft Computing, Vol. 15, No. X, pp. 1-10, 2009
7. Xiaoyi Song, Yongjie Li, WufanChen, "A Textural Feature Based Image Retrieval Algorithm", in Proc. of 4th International conference on Natural Computation, Oct. 2008.
8. Andrea Baraldi, FlavioParmiggiani, "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
9. S.Sindhuja, "Gray scale Image Analysis Using Morphological Filtering", International Journal of Innovations in Scientific and Engineering Research, Vol.1, no.1, pp. 13-18, JAN 2014.
10. Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu – Xuan Wang Yi – Fan Chang[2007] A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network, IEEE 7th International Symposium on Signal Processing and Information Technology.
11. Abdul Kadir Experiments Of Zernike Moments for Leaf Identification, Journal of Theoretical and Applied Information Technology, Vol. 41 No.1, and 15 July 2012
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.