ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Synergy of geometric and roughness measures for herbal leaf classification

S.Sruthi*, K.V.Subathra, S.Md.Mansoor Roomi, G.S.Sasi Rekha
Department of ECE, Thiagarajar College of Engineering, Madurai.
***Corresponding author: E-Mail: sruthistce@gmail.com**

ABSTRACT

In the current scenario, it is really hard to identify the type of leaf or its class by having a mere look at the leaf. Since people are not expertise in this field, an algorithm is required for doing such classification. This algorithm also paves a way for the future generations to have sound knowledge on herbal leaves. These technologies can also be applied widely in industries, medicine and agriculture. The required portion of the input image is segmented. Features of the input images are extracted by using Center-Symmetric Local Binary Pattern (CS-LBP) and Scale-invariant feature transform (SIFT) descriptor. Scale-invariant feature transform has ability to handle many complicacies and finally K –Nearest Neighbor Classifier is employed to identify leaves automatically for Herbal Leaf classification.

KEY WORDS: SIFT, CSLBP, KNN classifier.

1. INTRODUCTION

Digital image processing is the use of the algorithms for many different domains such as image enhancement, compression, analysis, mapping, geo-referencing, etc. Now-a-days, numerous methods for computer vision and techniques of recognising the pattern have been applied towards automated procedures of plant recognition. In Authentication of Leaf image using Image processing technique, the edge features are obtained by using Sobel. Although this method is simple it has a major disadvantage of Minimum samples in the database and used PCA which gave only 89.2% accuracy. Classification of Leaves using shape, colour, and texture features employs Shape, vein and texture features using Probabilistic Neural networks (PNN). The major drawback is the colour of the leaves change due to change in climatic conditions. So, no proper classification based on colour features. As the life of the leaf increases, they become dry so colour cannot be used as a main feature for classification of the leaves. When compared to previous techniques, the proposal of extracting the Key point descriptor and Roughness (Texture) features of the leaf is effective. The ultimate aim of the work is to develop a Leaf recognition algorithm based on key points obtained using the descriptors. The outcome ensures that this method is the simplest and efficient one. 9 different types of herbal leaves were considered.

2. PROPOSED METHOD

Pattern recognition is a very important field within computer vision, and the target of pattern recognition is to classify or recognize the patterns based on extracted features from them. The pattern recognition involves three process: a) Pre-processing, b) Feature Extraction, c) Classification. In Pre-processing, Segmented image is obtained by transforming the RGB values in the image to greyscale values. In feature extraction the Key point required for the classification is extracted. Finally, the leaf is classified into its required class.

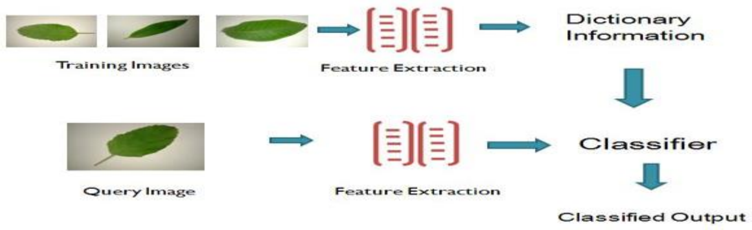

Fig.1. Flow of proposed methodology

Image pre-processing: The input leaf image is acquired by digital cameras. These images are of 800 x 600 resolutions with JPEG format. A grayscale image is obtained from the input RGB image. (1) Represents the equation to change RGB value of a pixel into its gray scale value.

$$BW = 0.2989\,R + 0.5870\,G + 0.1140\,B \qquad (1)$$

Where BW represents the gray scale value and R, G, B corresponds to three color channels of the input image respectively.

Feature Extraction: In the feature extraction step, the task is to describe the regions based on chosen representation, e.g. a region may be depicted by its boundary which is collective properties of features such as color, texture, interest points etc. In this paper, features are extracted using two descriptors namely SIFT and CSLBP. This type of hybrid feature gives better accuracy when compared single feature. All the interest points are obtained using SIFT and Surface roughness by CSLBP.

SIFT- Scale-invariant Feature Transform: "Feature" description of the object is provided by the interest points of that object. SIFT descriptors extract features which are invariant to scaling and rotation. For generating the feature, the SIFT approach takes an image and transforms it into a "huge collection of local feature vectors". The SIFT Feature Extraction process involves four stages.

Scale-Space Extrema Detection: This segment is able to find locations and scales which could be predicted from the same object under different views using 'scale-space' function. It is calculated by the function:

$$M(r, s, t) = H(r, s, t) * J(r, s) \quad (2)$$

Where * is the convolution operator, H(*r*, *s*, t) is a variable-scale Gaussian and J(*r*, *s*) is the input image. The stable key point location can be identified by using different techniques. Difference of Gaussian E(*r*, *s*, t) is the difference between two images, i.e., original one and another one that is scaled by *k*. E(*r*, *s*, t) is then given by:

$$E(r, s, t) = M(r, s, kt) - M(r, s, t) \quad (3)$$

The local minima and maxima of E(*r*, *s*, t) is estimated by comparing each point with its neighbours and from that maxima and minima values are calculated.

Key point Localisation: As the previous stage gives numerous key points, it is essential to eliminate the least significant points. This is done by identifying the points which are localised on edges or the points that are of low contrast. The Laplacian value for each and every key point is determined and depending on this value the key point is neglected. Extremum location, **Z**, is given by:

$$\mathbf{Z} = -\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial X} \quad (4)$$

A threshold is fixed and it is checked with the **Z** value. If the calculated Z value is less than the threshold value, the point is omitted. This cancels the extrema with low contrast. A small curvature is obtained orthogonally which is the difference of Gaussian functions. Then Hessian matrix of size 2x2 at that location and key point's scale is determined. If the difference of Gaussian is smaller than the ratio of largest eigenvector to smallest eigenvector and the calculated hessian matrix, the key point is rejected.

Orientation Assignment: After filtering the key points, a consistent orientation to the Key points are assigned. Depending on the properties of image, these orientations are assigned. A histogram for orientation is formed and the highest peak is determined. Then key point descriptor is employed to represent relative to the orientation, which achieves the property of rotation invariant.

Key point Descriptor: The local gradient data is used to create the key point descriptors. The local gradient information is rotated to line up with the orientation of the key point. Following that it is weighted by a Gaussian with variance of 1.5 * key point scale. Key point descriptor employs a set of 16 histograms, that are aligned in a 4x4 grid, each with 8 orientation bins, one for each of the main compass directions and one for each of the mid-points of these different directions. This result in a feature vector containing 128 elements for each interest points. These output vectors are known as SIFT keys and these are used in the classifier to identify possible objects in an image.

CSLBP- Center-Symmetric Local Binary Pattern: In LBP, each pixel is compared with the center pixel. But in CSLBP, the pixel is compared with centre-symmetric pairs of pixel. This reduces the total number of comparisons for the same set of neighbours. It is evident that there are eight neighbours for each pixel; LBP produces a value in the range [0,255], whereas for CS-LBP the range is 0-15 and it is closely related to gradient operator since it also considers difference of gray-level between pairs of pixels that are opposite to each other in a neighbourhood.

CS-LBP Descriptor: The segmented region of the leaf is the input for this descriptor. In the following experiments, the region size is fixed after normalising and the pixel value lies between 0 and 1.

Feature extraction with CS-LBP: By employing CS-LBP operator the features are extracted and the operator has parameters such as Radius R, Count of neighbouring pixels N and threshold on the gray level difference T.

Feature weighting: Each pixel of the leaf portion has got some importance in weighting based on the used feature. While comparing different weighting strategies it is found that the best choice for CS-LBP feature is uniform weighting.

Descriptor construction: Cartesian and log-polar grids were experimented in the work and it was found that the Cartesian gives better performance among both. In the experiments presented either a 3×3 (9 cells) or 4×4 (16 cells) Cartesian grids are used. For each cell a CS-LBP histogram is built. Thus, 3D histogram of CS-LBP feature locations and values constitute the descriptors.

Descriptor Normalization: Concatenate the histograms of the features computed for the cells to form a (4×4×16)256-dimensional vector. This provides the final descriptor. The descriptor is then normalized to unit length. By thresholding, larger descriptor elements are reduced. Each element is thresholded to be no larger than 0.2.

KNN Classifier: The dataset consists of many samples and every sample has its own attributes paved a way to construct an n dimensional vector.

$$x = (x1, x2, \ldots xn).$$

x is an independent variable in addition to that every sample has dependent variable called y whose values are completely dependent on x. The nature of y is categorical and f(x) is a scalar function with the relation y=f(x) assigned to every such vectors. For each class a set of N vectors are assigned in such a way that the flowing set is considered as the training set.

$$x(i), y(i) \text{ for } i = 1, 2, \ldots, T.$$

While testing the class of the sample to be identified is taken as x=u and the function has to be identified to calculate v=f(u). Now k samples in the training set is evaluated similarly to u based on the distance or dissimilarity measure computed using Euclidean distance is the most popular distance measure and it is defined as follows:

$$d(\mathbf{x}, \mathbf{u}) = \sqrt{\sum_{i=1}^{n} (x_i - u_i)^2}.$$

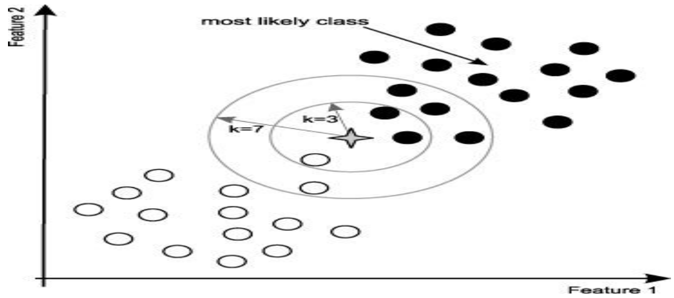

Fig.2. Classification using k-NN

The advantage with the k-NN classifier is that finding more number of neighbouring samples to u gives the large value of k which avoids over fitting problem.

3. EXPERIMENTS AND RESULTS

The main objective of the paper is to classify the different varieties of herbs by utilizing the various features extracted. As manual sorting of leaves are tedious, it would be better if it is done by automation. The solutions for these problems are discussed earlier and the results for these issues are shown below.

Herbal leaf Database: In terms of herbal leaf classification system, the details required for the classification should be given by the leaf database. Initially only 4 herbal types were considered. The images in the database are collected using an 8 mega pixel Digital camera. The images were taken under a controlled environment with required lighting and at a distance of 20 cm. The database collected for herbs are shown in fig 3.

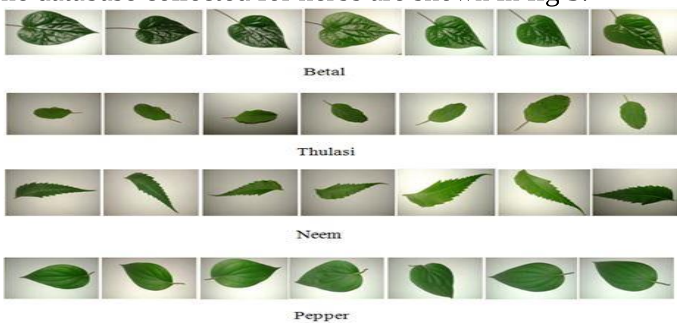

Fig.3. Database of spinach leave

SIFT Feature: Fig.4 shows the input image. This image is then pre-processed and the first step is Feature Extraction.

Fig.4. Input Image

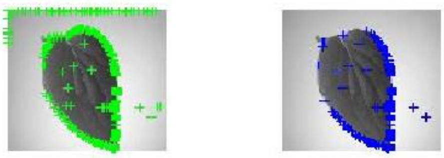

Fig.5. Orientation and descriptor detection

The leaf image is given as the input to the SIFT feature extraction. The output of the SIFT descriptor is shown in fig 5.The SIFT Feature vector is obtained is shown in fig 6, is given as the input to the classifier.

CSLBP: Fig 4 is the input image for the CSLBP (Texture) Feature Extraction method. This image is processed and the CSLBP feature vector is shown in Fig 7 which will be combined with the SIFT feature and sent to the classifier for classification.

k-NN Classifier: KNN structure is trained with 50 images from each different class or type of herbs. Based on the features obtained, KNN classifier calculates the distance and determines which herb type the leaf belongs to.

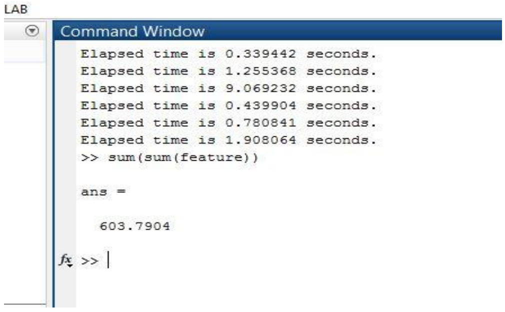

Fig.6. SIFT feature extracted

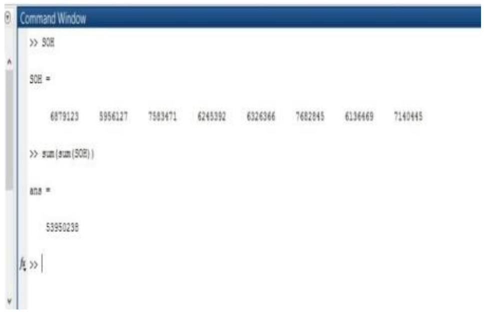

Fig.7. CS-LBP Feature

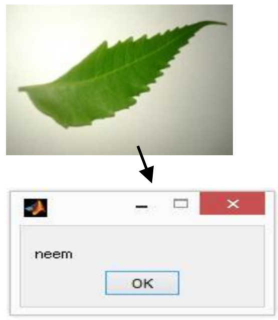

Fig.8. Classified output

4. CONCLUSION

The study ensures the necessity of classifying herbal leaves. However, as the paper work was based on a minimum sample size, it is essential to reconfirm the findings with a larger dataset. As the herbal leaf classification is a novel method of identification of plants it is believed that the results obtained are acceptable and it has the capability of being implemented in real time application. The limitation of the proposed work is it is not extendable to images that are with a varied background. The future work is to segment the leaf separately from a plant image and do the process of feature extraction and classify the leaves using a better classifier like ‘Kernel Sparse Representation Classifier’.

REFERENCES

Hanane Rami, Mohammed. Hamri, Lhoucine. Masmoudi, Objects Tracking in Images Sequence Using Center-Symmetric Local Binary Pattern (CS LBP), International Journal of Computer Applications Technology and Researcsh, 2 (5), 2013, 504 – 508.

Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang, Leaf shape based plant species recognition, Applied Mathematics and Computation, 185, 2007.

Jyotismita Chaki, Ranjan Parekh, Plant Leaf Recognition using Shape based Features and Neural Network classifiers, International Journal of Advanced Computer Science and Applications, 2 (10), 2011.

Mohamad Faizal AbJabal, Suhardi Hamid, Salehuddin Shuiband Illiasaak Ahmad, Leaf features extraction and recognition approaches to classify plant, Journal of Computer Science, 9 (10), 2013, 1295-1304.

Pande Ankita V, ShandilyaV.K, Digital Image Processing Approach for Fruit and Flower Leaf Identification and Recognition, International Journal of Engineering and Computer Science, 2 (4), 2013.

Wijesingha D, and FMMT Marikar, Automatic Detection System for the Identification of plants using Herbarium specimen images, Tropical Agricultural Research, 2012.

Ye Y, Chen C, Li C.T, Fu H, and Chi Z, A computerized plant species recognition system, in Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, 2004.