Types of Leaf Classification using Machine Learning

Prabhakar TS[#1], Siddharth B[#2], Mandara KR[# 3]

#1(Assistant Professor, Information Science and Engineering Mandya, Karnataka, India)
#2(Student, Information Science and Engineering Mandya, Karnataka, India)
#3(Student, Information Science and Engineering Mandya, Karnataka, India)

Abstract

Plants act like the cornerstone of all life on the planet and an indispensable resource for the Human living. Identification of leaf plays significant role in agricultural area whereas biologist can work the application of its medical usage. Different characteristics present in the leaves will classify them to different species and their different applications. The modern technology like machine learning has being used to build a model which identifies different types of leaf. This proposed model can be used at schools for classifying the leaf images where students can get to know about different kinds of leaf and their respective names. The proposed model is simple and have results in of high efficient system. It works with SVM classifier as backbone with combination of BOF and SURF feature. A multiclass Support Vector Machine (SVM) Classifier is then built by using the features of BOF and SURF features as input to the model. The system uses the customized leaf image data set. The experimental results exhibit that our proposed method is having highly efficiency in the process of identification of different category the leaf belongs.

Keywords

SURF Algorithm, Bag-of-features (BOF), SVM, Interest Points, Customised Leaf dataset, Training, Testing, Visual Word Generation.

I. INTRODUCTION

Classification of leaf in the digital world is very vital and taken the power of computer to different dimension. Plant leaf classification is always been a challenging task for the research perspective. Several approaches are presented in [1].Traditional method of identification of leaves by their knowledge and by just glancing over the leaf, has to be automated. The risk of identification of wrong leaf for wrong disease has taken away many life. The modern technologies made a foundation for all the biologist and computer scientists to evolve a machine which can be utilized for plant leaf classification or recognition.

Image processing, computer vision, AI and machine learning; which are the key research areas have many application to be developed and to be invented in the future. Researchers have shown that leaf shape and its textural features are the vital components on which they can be classified into different classes. The leaves of the plant are the vital source for classification of the plant according to its biological uses. Hence, many researchers have taken this features as the fundamental concept that is leaf shape and texture. In this paper, we have proposed a model which identifies plant leaf using leaf texture and shapes. The proposed model has two modules namely, 1) Training module and 2) Testing module. The model shown in the fig. 1, has two stages that are 1) Feature extraction and 2) classification. The feature extraction procedure is build using the fusion of two different texture techniques, bag-of-features (BOF) and Speeded up robust feature (SURF). The main inspiration of using Bag-of-features and Speeded up robust feature, is that this combination inherits the advantages of BOF, which are aggregates local components to form a histogram for the customised leaf image data set. The SURF feature is an effective one for the plant leaf texture and shape. First, we extract the image features using the SURF algorithm. The interest area points can be found using a SURF feature detector which also provides excellent scale-invariant features. Then, K-means clustering technique will run iteratively to form the k-mean clusters.

Each cluster will have a mean point which represent the whole cluster of that respective groups. After that, those mean points represent a particular type of leaf. Further, the cluster points are used to build a dictionary. Dictionary is build using the BOF concept in which normalization is done to the histogram obtained from the cluster points. Next, the histogram is obtained using SVM technique which uses the leaf texture and shape has its foundation. Later the two histograms are overlapped and concatenated, as a result the final vector is acquired. The final outcome is used by the multi-class classifier machine which works on SVM algorithm to classify the input leaf to its respective type of group.

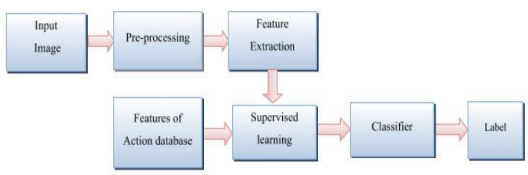

Fig .1 Flow Diagram of Proposed System

II. PROPOSED SYSTEM

The proposed system mainly consists of two phases, namely 1) Training and 2) Testing.

A. TRAINING

In the training phase, the system is trained with a set of leaves image, where each image of a leaf is pre-processed. Next, we extract SURF (Speeded up robust feature) features from the pre-processed image. Histogram plot provides a description of the number of pixels available in the given color ranges. Once the SURF features are extracted we create an array of SURF features of all the image dataset. After creating the array we apply K-mean clustering for the array of SURF features to create Bag of the visual dictionary. Create a normalized histogram for each leaf image with reference to Dictionary by using SURF features and store it as a training feature along with their class labels. Next, we need to train the machine using SVM (Support vector machine) with the help of training features and classes. In machine learning, support vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. The system will be trained and ready to accept the input.

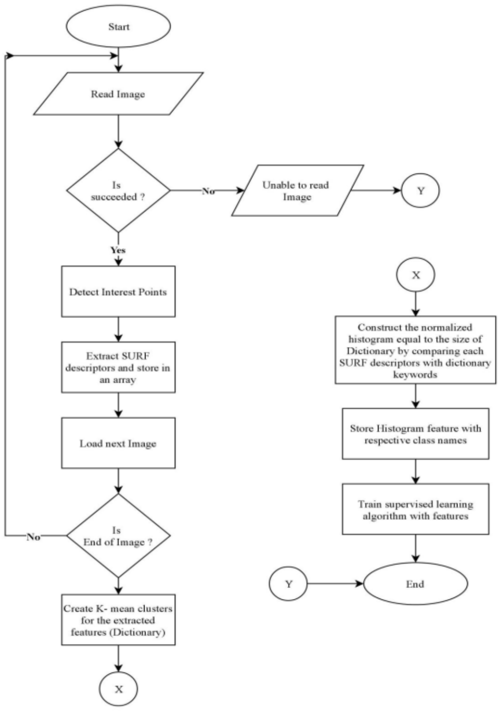

Fig.2 Flow Diagram of Training a Classifier

B. TESTING

In this phase, the user will input the leaf image and the image will go through pre-processing. Once the pre-processing is done we extract the features of the leaf. After extracting the features of the leaf, classification is done by the SVM. SVM classifier will extract the test feature vector histograms for each leaf. It then tests the extracted feature to the trained supervised algorithm to predict the class. Once the testing process is done, a message or a brief description of the leaf will be displayed on the screen.

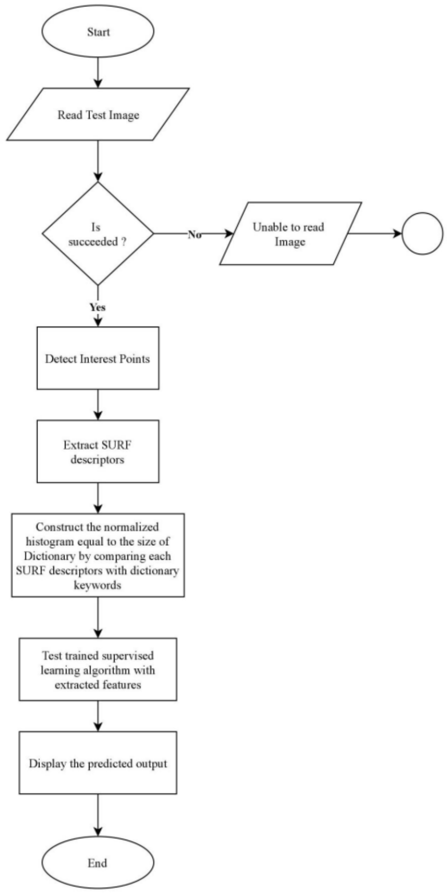

Fig.3 Flow Diagram of Testing a Classifier

III. METHODOLOGY

A. BOF Approach:

In short time, BOF is getting the lime light for its computer vision efficiency for its simplicity. BOF model is used in Modern information retrieval and natural language processing, and artificial neural networks. Images are the virtual data in the real world which can't be represented as a discrete word. Bag-of-features are extracted as follows:

1. Detect the area of interest points on the given input leaf data set.
2. Local features key points are identified around the leaf images.
3. Use K-mean clustering on the obtained interest points.
4. Build a dictionary for the cluster points by the unsupervised learning process.
5. Form a histogram for the dictionary build and do normalization on it and identify to which type the leaf belongs to.

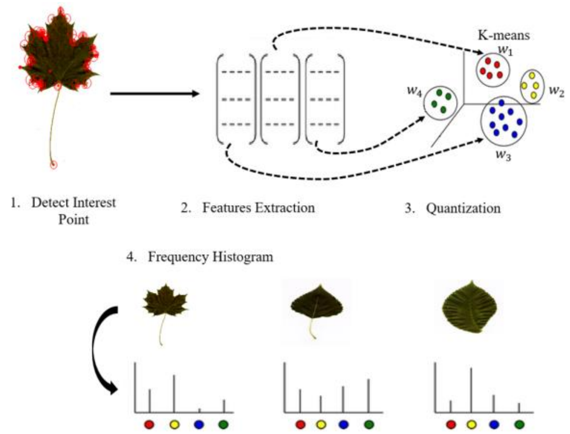

Fig.4. Stages of bag-of-feature construction

B. Interest Point Detection:

The initial stage is the process is to identify the key interest points. Harris Corners are not scale-invariant or rotation-invariant interest point detectors [12]. Many scale-invariant point detectors are developed to identify the interest points effectively.

We use Speeded-up-robust-feature for identification of local points. Hessian matrix is seen as excellent performance in the field of computational time and the accuracy it gives. Thus, SURF use the Hessian matrix. SURF is said to be more firm in the performance and operation. It is having high accuracy and less time consuming.

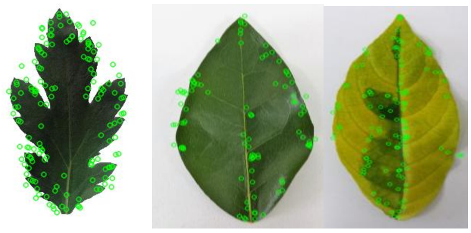

Fig.5 Interest Points Detection on Leaves

C. *Speeded Up Robust Feature (SURF)*:

In computer vision, speeded up robust features (SURF) is a patented local feature detector and descriptor. It can be used for tasks such as object recognition, image registration, classification or 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a pre-computed integral image. SURF descriptors have been used to locate and recognize objects, people or faces, to reconstruct 3D scenes, to track objects and to extract points of interest.

***D. Visual Word Generation/Vector Quantization*:**

Once the SURF feature extracts the interest points it summarises up by forming the BOF vector. K-mean is one of the unsupervised leaning process which is used to build a dictionary for the interest points obtained from the SURF feature.

The process starts with initialising the K-mean clusters; and all interest-points will be vector quantized opposed against these clusters. The clusters are found by iterative process to identify the interest points. The cluster points are obtained by using the mathematical distance formula. Then, Mean is calculated for all the points to identify the one mean point which will represent the whole cluster of that respective class. Eventually, to construct the final dictionary normalisation is done to the BOF model. Dictionary is built of each class which represent each type of leaf. Histogram is built per image. The built histogram has one types of class for each bin in the dictionary.

E. SVM:

After the dictionary is built the proposed model is trained. The trained model works with the SVM classifier for plant identification. SVM is used when it has many classes. SVM is a type of classification technique which defines boundaries for each classes.

IV. CUSTOMISED LEAF DATASET

All images taken are in the jpg. Format with a dimension of 1200x1800. The datasets are divided into 80% for training and 20% are used for testing.

Fig. 6 Customised Leaf Dataset

IV. EXPERIMENTAL RESULTS

In our proposed model the result is established using the technique discussed above for classifying the testing dataset from the customised dataset. The proposed system consist two parts. First part is training and the second is for testing. For, calculating the results we have taken 300 examples per class for a training set and the reaming 30 images are used for testing.

	BARBABOS CHERRY	INDIAN BALSAM	BETEL	BITTER ORANGE	BRAHMI	CARICATURE	CHINESE GUAVA	CARISSA	COFFEE	CORIANDER	SWEET POTATO	PAPAYA	MINT	EGG PLANT	THEVETIA
BARBABOS CHERRY	5														
INDIAN BALSAM		5													
BETEL			5												
BITTER ORANGE				5											
BRAHMI					5										
CARICATURE						4									
CHINESE GUAVA							5								
CARISSA								5							
COFFEE									5						
CORIANDER										5					
SWEET POTATO											5				
PAPAYA												5			
MINT													5		
EGG PLANT														4	
THEVETIA															5

Fig. 7 Confusion matrices

The obtained perfection of our proposed model is 97.3%.

V. FUTURE SCOPE

The outcomes in this research are based on that involve only sample datasets. It is necessary that additional datasets should be considered for the evaluation of different classification problems as the information growth in the recent technology is growing and data are by nature dynamic. Hence, further classification of the entire system needs to be implemented right from scratch since the results from the old process have become obsolete-contained. Causal Productions has used its best efforts to ensure that the templates have the same appearance. We can develop a real-time system by using the technique of Tensor flow, which is having a very efficiency for the dynamic working of the model. By collecting datasets of different classes we can develop the model to identify many types of leaves with still more accuracy.

VI. CONCLUSION

In this paper, types of leaf identification are presented based on SVM (Support Vector Machine) classifier with BOF and SURF feature extraction. The data set consist of 500 different class of leaf image. After the training, the BOF will create a dictionary. The proposed model will use this dictionary for testing of input leaf image given by comparing the histogram of BOF with the dictionary already obtained from the training process. The experiment has been carried out on different types of leaf images to perform the type of leaf classification. It is concluded that the proposed method effectively recognizes the leaf. To improve the recognition rate in the classification of leaves, different features, different classifiers and the different size of the leaf can be used as the dataset.

REFERENCES

[1] Khairnar, Khushal, and Rahul Dagade. ”Disease Detection and Diagnosis on Plant using Image Processing–A Review.” International Journal of Computer Applications 108, no. 13 (2014): 36-38.

[2] Narayanan, Barath Narayanan, Ouboti Djaneye-Boundjou, and Temesguen M. Kebede. ”Performance analysis of machine learning and pattern recognition algorithms for malware classification.” In Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National, pp. 338-342. IEEE, 2016.

[3] Ali, Redha A., and Russell C. Hardie. ”Recursive non-local means filter for video denoising.” EURASIP Journal on Image and Video Processing 2017, no. 1 (2017): 29.

[4] Almahdi, Redha, and Russell C. Hardie. ”Recursive non-local means filter for video denoising with Poisson-Gaussian noise.” In Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National, pp. 318-322. IEEE, 2016.

[5] Sakai, N., S. Yonekawa, A. Matsuzaki, and H. Morishima. ”Two-dimensional image analysis of the shape of rice and its application to separating varieties.” Journal of Food Engineering 27, no. 4 (1996): 397-407.

[6] Neto, Jo˜ao Camargo, George E. Meyer, David D. Jones, and Ashok K. Samal. ”Plant species identification using Elliptic Fourier leaf sharp analysis.” Computers and electronics in agriculture 50, no. 2 (2006):121-134.

[7] Zhang, Shanwen, Yingke Lei, Tianbao Dong, and Xiao-Ping Zhang.”Label propagation based supervised locality projection analysis for plant leaf classification.” Pattern Recognition 46, no. 7 (2013): 1891-1897.

[8] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. ”Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.” In null, pp. 2169-2178. IEEE, 2006.

[9] Wang, Jinjun, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong. ”Locality-constrained linear coding for image classification.”In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 3360-3367. IEEE, 2010.

[10] Harris, Chris, and Mike Stephens. ”A combined corner and edge detector.” In Alvey vision conference, vol. 15, no. 50, pp. 10-5244. 1988.

[11] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. ”Surf: Speeded up robust features.” In European conference on computer vision, pp. 404-417. Springer, Berlin, Heidelberg, 2006.

[12] Kadir, Timor, and Michael Brady. ”Saliency, scale and image description.”International Journal of Computer Vision 45, no. 2 (2001): 83-105.

[13] Mikolajczyk, Krystian, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. ”A comparison of affine region detectors.” International journal of computer vision 65, no. 1-2 (2005): 43-72.

[14] Bosch, Anna, Xavier Mu˜noz, and Robert Mart´ı. ”Which is the best way to organize/classify images by content?.” Image and vision computing 25, no. 6 (2007): 778-791.