Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 324 – 332

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Optimizing Feature Selection using Particle Swarm Optimization and Utilizing Ventral Sides of Leaves for Plant Leaf Classification

Arun Kumar*, Vinod Patidar, Deepak Khazanchi and Poonam Saini

Sir Padampat Singhania University, Udaipur 313601, Rajasthan, India

Abstract

As the digital images produce a lot of information about the pixels, there is a need to find alternative methods to reduce the image feature dataset for faster and automatic classification of plants through digital leaf images. In the present work, the leaf image texture features have been extracted through Gabor based techniques and then subjecting them to PSO-CFS based search method for identifying the best set of features from the complete feature set and then classifying them using four classification algorithms like KNN, J48, CART and RF. Another objective of this work is to utilize the two faces available on the plant leaves (Dorsal and Ventral), instead of one (i.e. Dorsal) for classification of plants on the basis of digital leafimages and to analyse the effects on classification accuracy values for dorsal and ventral sides of leaf images.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Dorsal Side; Gabor Filter; Leaf Images; Particle Swarm Optimization; Ventral Side.

1. Introduction

Man has been enjoying the boons of nature, its flora and fauna since the ages. Due to the technical advancement and the human need for better roads, bridges and houses, there has been a reckless felling of trees and cutting of vegetation to pave the way for roads and bridges. The development on one end is leading to disappearance of flora and fauna, though essential in maintaining the ecological balance. But, at the same time, the human quest for identifying and taxonomically classifying the plants and their sub-species and then devising methods for preserving them for the future before the plant species get extinct, has been going on in scientific world since decades. The plants have been studied for their flowers, leaves, seeds and fruits. There are millions of different plant species, but many of the sub species are still unknown and would die and become extinct, before their turn comes up to know them. Therefore, there is a need for automatic plant identification and taxonomically classification methods which could speed up the process of knowing the individual plant species. The biologist and computer scientists have been playing their roles in suggesting newer methods for identifying the plant species. The computer vision methods have revolutionized the work of automatic plant classification and are based on finding suitable characteristic features from the digital images

*Corresponding author. Tel.: +09828752479.
E-mail address: arunkumarsai@gmail.com

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.079

and then suitably classifying them in to various species. As the data collected from the digital images is enormous, therefore there is a need to find subset of the data which would do the same work as that by the whole dataset.

In order to reduce the large dataset to a smaller subset, the role of feature selection algorithms is pivotal and is evolving day by day. By using the feature selection methodology, there is a drastic improvement in the average predictive classification accuracy performance results, and at the same time the time required to compute the predictive accuracy from the complete dataset is reduced due to the presence of lesser number of processing variables left in the dataset. The new dataset is partitioned into a test set and a training set model. The training data is utilized for training the classifier and the test set data is utilized for testing the overall accuracy achievable from the data.

Since plants are essential reservoir for fulfilling the basic human needs for food, shelter and other medicinal values, therefore to study them is really appealing in the present day world. In the present work, the plant leaves have been chosen as the subject of study. The plant leaves have somewhat smoother and less hairy surface on the dorsal portion of the leaves as compared to the ventral sides, which have prominent vein structures present. In this work, it is proposed to study the effects of Particle Swarm Optimization technique, a technique within the family of evolutionary optimization algorithms, to search for the best set of features for predicative classification of digital leaf image dataset. The overall process involves extracting the important texture features from the dorsal and ventral parts of the leaves of digital images and then subjecting them to feature selection and further classification process, and to study the effects of PSO based feature search on the overall classification accuracy values. The work substantiates the fact that ventral leaf images can be another valuable alternative for the discrimination of plant images on the basis of application of PSO based feature search techniques applied on the digital leaf images.

In this present work, the Section 2 explains about the techniques utilized to extract the important features from the different leaf images of different plant species and the formation of the complete dataset which has all the extracted features. After the preparation of the feature dataset, the 3rd Section describes about the adoption of the feature selection methodology to the complete dataset, to extract unique, useful and uncorrelated features and the preparation of the subset of dorsal and ventral feature sets with the application of PSO-CFS based technique over the entire dataset. The 4th part of this research work, describes about the application of classification algorithms like KNN, J48, CART and RF on two different sets of data, one representing dorsal and the other representing the ventral feature set extracted from leaf images. The 5th part represents the result analysis and comparative study with other works of similar nature.

2. Methodology Adopted

2.1 Database creation and its preprocessing

The plant leaves have two faces viz.: dorsal and ventral, therefore there is a need for paying critical attention on both the sides of the leaves, as both have independent and unique set of features. The existing leaf image databases available on the internet contain the leaf images of dorsal sides, but to achieve the objective of this work, there is a need to create an independent leaf image database with dorsal as well as ventral sides of the leaves. Therefore, 25 dorsal side and 25 ventral side leaf images were clicked for each plant species and a database of ten plant species has been created with 250 dorsal and 250 ventral sides of leaf images and 500 images in totality. A sample of such colored leaf images is shown in Fig. 1.

The 500 images captured were subjected to the process of background removal, size reduction to 256×256 and converting them to gray scale. The gray scaled images were subjected to the process of image contrast and intensity enhancement techniques and then stacked together as slices for further processing.

2.2 Extraction of texture features

The digital leaf images have been studied by using their color, geometrical shape, texture etc. for discrimination of plants. In the present work, the leaf images have been studied through the texture features present on its dorsal and ventral sides. The term texture refers to the visual features present on the objects of interest. The Fig. 2 shows the 3D surface plot of texture of the digital leaf image with dorsal (Fig. 2(a)) and ventral sides (Fig. 2(b)) of the Slice-24E from the leaf image dataset created in the present study. Here E in Slice-24E represents the enhanced image.