Pattern Recognition 47 (2014) 158–168

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Automatic classification of legumes using leaf vein image features

Mónica G. Larese [a,*], Rafael Namías [a], Roque M. Craviotto [b], Miriam R. Arango [b], Carina Gallo [b], Pablo M. Granitto [a]

[a] CIFASIS, French Argentine International Center for Information and Systems Sciences, UAM (France)/UNR-CONICET (Argentina), Bv. 27 de Febrero 210 Bis, 2000 Rosario, Argentina
[b] Oliveros Experimental Station, National Institute of Agricultural Technology (INTA), Ruta Nacional 11 km 353, 2206 Oliveros, Santa Fe, Argentina

ARTICLE INFO

Available online 21 June 2013

Keywords:
Leaf vein features
Leaf vein images
Legume classification
Leaf vein analysis
Unconstrained hit-or-miss transform

ABSTRACT

In this paper, a procedure for segmenting and classifying scanned legume leaves based only on the analysis of their veins is proposed (leaf shape, size, texture and color are discarded). Three legume species are studied, namely soybean, red and white beans. The leaf images are acquired using a standard scanner. The segmentation is performed using the unconstrained hit-or-miss transform and adaptive thresholding. Several morphological features are computed on the segmented venation, and classified using four alternative classifiers, namely support vector machines (linear and Gaussian kernels), penalized discriminant analysis and random forests. The performance is compared to the one obtained with cleared leaves images, which require a more expensive, time consuming and delicate procedure of acquisition. The results are encouraging, showing that the proposed approach is an effective and more economic alternative solution which outperforms the manual expert's recognition.

1. Introduction

The automatic analysis of leaf images aimed at plant classification or plant image retrieval has been addressed by many researchers in the recent literature. Several approaches have been proposed, including leaf shape [12,1,4,7,20,5], color information [11,15] and leaf texture analysis [8,2].

Although all these approaches are valid, they are not useful when dealing with species having similar leaf size, color, shape and texture features. For example, such is the case with individuals from different varieties of the same species, which have no clear visual differences in the previously mentioned leaf characteristics. Recently, some authors [6,14,22] stated that leaf venation properties may be of high importance to perform plant recognition. This hypothesis is also supported by recent studies [17,18] which show correlations between venation networks and leaf properties (for example, drought and damage tolerance). Under these assumptions, it is feasible to think that the particular physiological characteristics of the plants are reflected in their leaf veins, even when the leaves have similar appearance.

In this work an automatic procedure exclusively based on the analysis of leaf vein morphological features is proposed for plant recognition. Leaf shape, texture, color and size are discarded. Leaf vein segmentation is performed resorting to the unconstrained hit-or-miss transform (UHMT) [19] and adaptive image thresholding applied to the gray scale leaf images. The UHMT is a mathematical morphology operator similar to template matching. It allows to extract all the pixels having a certain foreground and background neighboring configuration.

Simple morphological features are measured on the segmented veins, and four different state-of-the-art classifiers are compared to perform plant identification, namely support vector machines (SVM) [23] with linear and Gaussian kernels, penalized discriminant analysis (PDA) [10] and random forests (RF) [3].

The whole procedure was used to recognize three classes of legumes, namely soybean (*Glycine max* (*L*) *Merr*), red and white beans (*Phaseolus vulgaris*). Red and white beans belong to the same species, presenting similar leaves except for their vein color, which is dark for the red bean. However, color is not taken into account in this paper. Only vein morphological features are considered on gray scale images.

We report the quantitative performance of the whole procedure, discussing the classification accuracies per class achieved by the automatic classifiers and the advantages of the proposed methodology. The procedure was developed searching also for simplicity and low cost. For this reason, the leaf images were acquired using a standard scanner, without any staining procedure. The results were compared to the performance achieved by human experts.

The proposed approach was also compared to the more sophisticated methodology of analyzing digital photographs of

* Corresponding author. Tel.: +54 341 4237248x303; fax: +54 341 4237248x301.
E-mail addresses: larese@cifasis-conicet.gov.ar, mnclar@yahoo.com.ar (M.G. Larese), namias@cifasis-conicet.gov.ar (R. Namías), rcraviotto@correo.inta.gov.ar (R.M. Craviotto), marango@correo.inta.gov.ar (M.R. Arango), cgallo@correo.inta.gov.ar (C. Gallo), granitto@cifasis-conicet.gov.ar (P.M. Granitto).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.06.012

cleared leaves images. This alternative provides with enhanced high contrast leaf veins and higher orders of visible veins, but it is a much more expensive and time demanding procedure given the chemical staining process applied to the leaves.

The rest of the paper is organized as follows. In Sections 2.1 and 2.2 the proposed vein segmentation procedure is explained. The morphological measures computed on the segmented veins are summarized in Section 2.3. The employed classification algorithms are briefly described in Section 2.4. In Section 3, we describe the leaf images datasets and discuss the obtained results. Finally, some conclusions and future work are presented in Section 4.

2. Materials and methods

2.1. Unconstrained hit-or-miss transform (UHMT)

The UHMT is an extension of the hit-or-miss transform (HMT) for gray scale images [19]. It extracts all the pixels matching a certain foreground and background neighboring configuration. A composite structuring element **B** is employed, which is a disjoint set formed by one structuring element that specifies the foreground configuration, B_{fg}, and one structuring element for the background setting, B_{bg}. The origin of the composite structuring element matches the foreground.

The UHMT is defined as

$$UHMT_{\mathbf{B}}(Y)(y) = \max\left\{\varepsilon_{B_{fg}}(Y)(y) - \delta_{B_{bg}}(Y)(y), 0\right\}, \quad (1)$$

where Y is a gray scale image with set of pixels y and **B** is a composite structuring element. It can be computed as the difference between an erosion with B_{fg}, $\varepsilon_{B_{fg}}(Y)(y)$, and a dilation with B_{bg}, $\delta_{B_{bg}}(Y)(y)$, if $\delta_{B_{bg}}(Y)(y) < \varepsilon_{B_{fg}}(Y)(y)$. Otherwise it equals 0.

2.2. Vein segmentation

The color information was removed by converting the RGB images to grayscale. The color information is discarded since there is interest in detecting vein patterns associated to vein morphology only.

The binary masks for the leaves were obtained via thresholding (automatic iterative threshold selection [21]), holes filling using morphological reconstruction [19] and removal of all the connected components except the largest one.

In order to segment the veins in the scanned images, the UHMTs on five different sized versions of the images, namely at 100%, 90%, 80%, 70% and 60%, were computed. Each version is intended to highlight a different level of vein detail. Then, each resulting UHMT was resized back to its original size and added to obtain the combined UHMT, which highlights both small and large visible veins simultaneously. For this purpose, four composite structuring elements (foreground and background configurations) were used aimed at detecting leaf veins in four directions (vertical, horizontal, +45° and −45°). These structuring elements are shown in Fig. 1. After that, the contrast of the combined UHMT was enhanced and then binarized by means of a standard adaptive thresholding algorithm. All the connected components with less than 20 pixels were removed.

For the cleared images, the veins are already highlighted due to the staining procedure. For this reason, the segmentation was performed by simply applying adaptive histogram equalization followed by standard adaptive thresholding.

2.3. Vein measurements

In order to measure vein and areole features without the influence of the leaf shape, a central patch was extracted from each segmented scanned and cleared leaf, respectively. Ideally, we would like to work with the entire leaf venation network. However, in practice we cannot achieve this since we want to discard the leaf shape contour influence in order to analyze exclusively vein features. For this reason, in this paper we chose to extract a patch located at the center of each leaf, which we consider significant in order to capture primary and secondary order veins features, with a size big enough to include higher order veins. Another significant point of analysis could have been the union between the leaf blade and the petiole, or the leaf apex, but the vein characteristics at both locations are very much influenced by the leaf shape, so they were discarded. However, if the patch is too big we risk to touch the leaf contour and include it unintentionally. With these requirements in mind, we selected 100×100 pixel-sized patches for scanned leaves since this selection accomplished the goal, and 400×400 pixel-sized patches for cleared leaves, since the resolution of the latter is approximately four times higher than the former's. All the traits were computed on these patches, and the same traits were computed for the scanned and cleared leaves.

LEAF GUI measures [16] were adapted to extract a set of features of interest for the veins and areoles. For the particular problem of leaf classification, the individual vein/areole measures computed by LEAF GUI are not suitable. For this reason, the median, minimum and maximum measure values were computed for the veins and areoles where it was appropriate. An extra measurement not available in LEAF GUI, namely the vein orientation, was also considered in this paper. Altogether, 52 measures were computed for each leaf patch. In Appendix A, the explanation and computing procedure for each measure is provided. Further information can be found in the work by Price et al. [16].

However, we found that 17 out of these 52 measures had a near zero variance across the leaves and therefore were discarded for

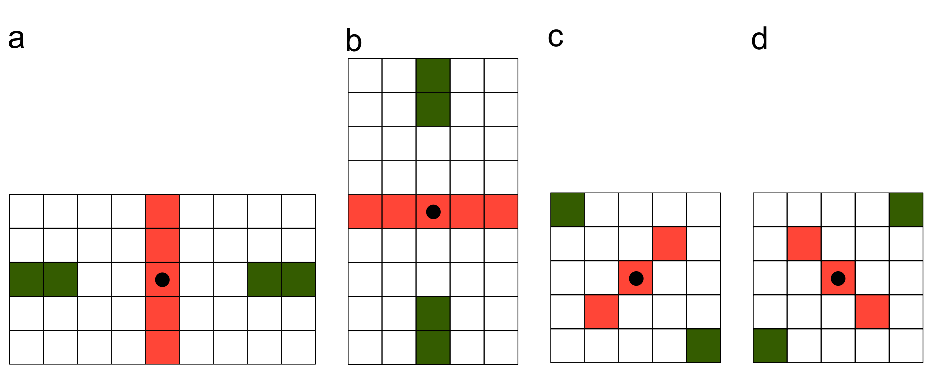

Fig. 1. Flat composite structuring elements used for the UHMT to detect veins in four directions: (a) Vertical. (b) Horizontal. (c) +45°. (d) −45°. Foreground and background pixel configurations are shown in red and green, respectively. The center of the composite structuring element is marked with a black dot. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)