BIOSYSTEMS ENGINEERING 170 (2018) 72–84

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Leaf recognition using contour unwrapping and apex alignment with tuned random subspace method

Sruti Das Choudhury [a,*], Jin-Gang Yu [b], Ashok Samal [a]

[a] Department of Computer Science and Engineering, University of Nebraska-Lincoln, USA

[b] School of Automation Science and Engineering, South China University of Technology, China

ARTICLE INFO

Article history:
Received 3 October 2017
Received in revised form
9 March 2018
Accepted 4 April 2018

Keywords:
Leaf recognition
Generalised Procrustes analysis
Random subspace method

The variation in scale, translation and rotation pose the main challenges to automatic leaf recognition. This paper introduces an automatic leaf recognition method which uses generalised Procrustes analysis (GPA) to mutually align all leaf contours of each of the known classes with respect to scale, translation, rotation and reflection. A mean contour is computed as a representative of each known class. A test leaf is subjected to ordinary Procrustes analysis to be aligned with the mean contour with respect to scale, translation, rotation and reflection. However, experimental analyses show that in the cases where the leaf contours are significantly rotated with respect to each other, generalised Procrustes analysis fails to correctly align. To overcome this, we introduce a novel leaf apex detection algorithm based on Newton's divided method of interpolation and second order differentiation for critical point analysis. The 2-dimensional GPA-transformed contours are unwrapped by computing the distances between the contour points and the centre-of-mass of the contour starting from the leaf apex in an anticlockwise direction to generate a 1-dimensional distance signal. Principal component analysis is used for dimensionality reduction and linear discriminant analysis is used to achieve optimal class separability. The paper extends the use of random subspace method as an ensemble classifier in leaf recognition to exploit the high dimensionality of the feature space for improved identification by avoiding overlearning. Experimental analyses using two publicly available datasets demonstrate the effectiveness of the proposed method.

1. Introduction

The importance of plants in both natural, managed and urban ecosystems is well documented. The environmental conditions, i.e., soil, water, fertiliser, temperature, etc., impact the growth of plants. In natural ecosystem, identification of plants is an important task for resource managers to ensure the diversity of species and assess the health of the ecosystem. Automated plant recognition is also important in identifying endangered plant species (Kalyoncu & Toygar, 2015). In a managed ecosystem, identification of weeds is a critical

* Corresponding author.
E-mail addresses: S.D.Choudhury@unl.edu (S.D. Choudhury), Jingangyu@scut.edu.cn (J.-G. Yu), samal@cse.unl.edu (A. Samal).

https://doi.org/10.1016/j.biosystemseng.2018.04.001

element to improving crop productivity. In an urban ecosystem, it is important to maintain an inventory of different species of plants to grow them properly in order to improve the quality of life of citizens. Building accurate knowledge about the identity of the plants is essential for a sustainable development of agriculture and conservation of biodiversity. Thus, identification of plants is an important problem with widespread applications in applied botany, e.g., agriculture, forestry and nature conservation, and for general public awareness. The shape and size of leaves are defining characteristics of plants and are often used to assist in the identification of plants. Therefore, an automated plant identification method based on computer vision has the potential to successfully replace the current tedious and manual process. In addition to assisting resource managers, the application in mobile computing platforms can be used to educate the general public about the identity of plants in real-time without requiring their presence in the field.

According to the theory of plant taxonomy, a plant can be identified based on its leaves and flowers (Lee & Chen, 2003). Flowers have complex 3-dimensional (3D) structures, and can only be found in the blooming seasons. Also, there are many non-flowering plants. On the other hand, leaves are present during most of the life of the plant and furthermore can be classified comparatively easily using their 2-dimensional (2D) shape characteristics and textures. Hence, they have been widely used for computer-aided automatic plant identification. While the differences in boundary shape characteristics contribute positively to the inter-class leaf discrimination, the intra-class variations in leaf shape due to rotation, scale, translation and reflection, limit reliability. This paper presents an algorithm that improves the efficiency and reliability over the state-of-the-art methods, by providing a new approach to leaf recognition based on minimising the intra-class shape variation using generalised Procrustes analysis (GPA) to mutually align all contours of a known (training) class with respect to scale, rotation, translation and reflection. A mean contour is computed as a representative for each class. The method unwraps the 2D contour starting from the leaf apex to generate a 1-dimensional (1D) distance signal corresponding to all GPA-transformed contours of a training class by computing the distances between the points on the contour and its centre-of-mass (COM). During testing, the contour of the unknown leaf is first transformed by ordinary Procrustes analysis (OPA) to match the mean contours of individual training classes and then reduced to 1D distance signal. The normalised distance signal has been used in the existing leaf recognition methods (e.g., (Wang, Chi, Feng, & Wang, 2000)) due to its scale and translation invariant properties. However, if two leaves of the same species are rotated in different directions, the distance signals thus obtained, are considerably different, resulting in reduced recognition rate. The use of GPA in the proposed method addresses limited variation in rotation, but cannot deal with the extreme situations where the leaves are rotated drastically with respect to each other, such as in opposite directions. Thus, in order to make the system more robust, the proposed method introduces a novel leaf apex detection technique based on Newton's divided method of interpolation to reliably eliminate the shape distortion due to rotation.

The feature sets comprising distance signals have very high dimensionality compared to the number of available leaves in the training classes, which leads to overfitting. We extend the use of random subspace method (RSM) (Ho, 1998) in leaf recognition for improved identification rate by avoiding overfitting. RSM is an ensemble classifier which randomly selects multiple subspaces from the feature space, and associates a classifier with each subspace. It addresses the apparent dilemma of accuracy optimisation and over-adaptation by exploiting high dimensionality (Choudhury & Tjahjadi, 2015).

The main contribution of this paper is to introduce a novel leaf recognition method based on contour unwrapping and leaf apex alignment, which successfully overcomes the shortcoming of traditional GPA to align contours with variations in scale, rotation and translation. It also introduces the application of RSM in leaf recognition in order to augment the effectiveness of GPA to minimise the intra-class shape distortions for increased accuracy by exploiting high dimensionality of the feature space. The proposed method uses Newton's divided method of interpolation for leaf apex detection and converts the leaf contour into a distance signal by contour unwrapping starting from the leaf apex.

The rest of the paper is organised as follows. Section 2 discusses related work and Section 3 presents the proposed method. Experimental results are analysed in Section 4, and Section 5 concludes the paper.

2. Related work

Many leaf classification methods are proposed in the literature. Morphological leaf parameters are used by Du, Wang, and Zhang (2007); Wu et al. (2007) to classify leaves at low computational complexity. The method proposed by Du et al. (2007) introduces move median centres hypersphere (MMCH) classifier to classify leaves based on morphological parameters, e.g., aspect ratio, rectangularity, area ratio of convexity, perimeter ratio of convexity, sphericity, circularity, eccentricity, form factor, and Hu moments. The method by Wu et al. (2007) applies PCA to orthogonalise the 12 morphological parameters derived from 5 basic geometric features, e.g., diameter, physiological length, physiological width, leaf area and perimeter, extracted from the segmented binary leaf images. A Probabilistic Neural Network (PNN) is used to classify leaves. The method requires manual entry of the start and end points of the midrib, and hence not fully automated.

Bruno, d. O. Plotze, Falvo, and d Castro (2008) uses fractal dimension computed by box-counting (Tricot, 1995) and Minkoswki's multiscale methods (Costa & Cesar, 2000) as the only morphological parameter to analyse the complexity of a leaf shape for plant identification. The method uses LDA for classification. The effectiveness of the method is demonstrated on reasonably small amount of data, i.e., fifty samples from ten species native to the Brazilian Atlantic forest and Brazilian Cerrado scrubland, and 20 samples from four different species of the genus Passiflora. A content-based image retrieval method for leaf classification is proposed by Park, Hwang, and Nam (2008). The method analyses venation pattern to categorise leaves and performs retrieval from each