Analysis of Relevant Features for Pollen Classification

Gildardo Lozano-Vega[1,2], Yannick Benezeth[2], Franck Marzani[2], and Frank Boochs[1]

[1] i3mainz, Fachhochschule Mainz, Lucy-Hillebrand-Strasse 2, 55128 Mainz, Germany
gildardo.lozano@fh-mainz.de
[2] Le2i, Université de Bourgogne, B.P. 47870, 21078 Dijon Cedex, France

Abstract. The correct classification of airborne pollen is relevant for medical treatment of allergies, and the regular manual process is costly and time consuming. Aiming at automatic processing, we propose a set of relevant image-based features for the recognition of top allergenic pollen taxa. The foundation of our proposal is the testing and evaluation of features that can properly describe pollen in terms of shape, texture, size and apertures. In this regard, a new flexible aperture detector is incorporated to the tests. The selected set is demonstrated to overcome the intra-class variance and inter-class similarity in a SVM classification scheme with a performance comparable to the state of the art procedures.

Keywords: pattern recognition, feature extraction, feature evaluation, apertures, palynology.

1 Introduction

The correct estimation of airborne pollen concentration is important for the prevention and treatment of allergies. Traditional methods require manual and specialized labor, which is expensive, time consuming and susceptible to inconsistency. Commonly, collected airborne particles are analyzed manually under a brightfield microscope in order to count the frequency of different pollen taxa. With the introduction of computer vision techniques, pollen counting aspires to become automatic, faster and more accurate, which would enable a more frequent and broader analysis of samples.

Common strategies for image-based pollen classification follow a typical image classification process: image digitization, preprocessing, segmentation, pollen description, and pattern recognition. In all the cases, pollen description is based on different types of metrics and representations of the pollen image, known in the literature as *features*. Previous strategies are differentiated mainly by the type of employed features. Chen *et al.* found a relevant combination of seven general shape features, four aperture-colpus features, and a statistical gray-level feature to recognize *Birch, Grass* and *Mugwort* pollen with 97.2% of accuracy with a Linear Normal Classifier (LNC) and forward feature selection [1]. Unfortunately,

L. Iliadis et al. (Eds.): AIAI 2014, IFIP AICT 436, pp. 395–404, 2014.

the aperture and colpus detectors are type-specific and are not easily usable for other taxa.

Additionally to general shape and aperture features, Boucher *et al.* used color features for the classification of 30 pollen taxa [2]. With the same deficiency as Chen *et al.* for the aperture detection, they achieved an accuracy of 77%. The average of 11.6 images/taxon in the dataset looks small for capturing all the variations. Interestingly, they used a knowledge-based classification instead of typical statistical methods. A good performance just employing five general shape features was achieved by Rodríguez Damián *et al.* in the classification of three types of *Urticaceae* pollen with 85.6% of accuracy with a minimum distance classifier [3]. Employing only Haralick measures, Li *et al.* classified four pollen taxa with 100% of accuracy using the Fisher's linear discriminant method [4]. However, the tested taxa does not appear to have strong inter-class similarity which leaves the performance unknown on typical allergenic pollen. The aforementioned results suggest that the synergic contribution of features from different descriptive foundations could be the key for a robust and accurate classification.

Atypically, Ranzato *et al.* used local invariant features without segmentation with accuracy of 78.2% on eight pollen taxa, and employed a Bayesian classifier with a Gaussian mixture probability density function [5]. Similarly, Ronnenberger *et al.* developed 3D features from the Haar integration framework, achieving 98.5% of precision with a Support Vector Machine (SVM) and the Radial Basis Function (RBF) as the transformation kernel, for 33 taxa after the rejection of uncertain results [6]. Although very impressive, unfortunately the method requires 3D volumetric scans of the particle, which looks impractical for real-time applications.

Our proposed solution takes advantage of the strength of diverse feature groups in order to provide robustness and accuracy. The rest of the paper is organized as follows: in section 2, the proposed strategy for description and classification is explained in detail. In section 3, experiments and results are presented. Finally, in section 4, conclusions of the proposed method and the following steps are stated.

2 Strategy for Pollen Description and Classification

2.1 Outline of the Proposed Solution

We propose the application of selected features for the classification of the five most important allergenic pollen taxa in Germany: *Alder, Birch, Hazel, Mugwort* and *Grass*. Starting with microscopic 2D images, the solution involves the segmentation of the pollen from the background and the description of the particle by different types of features. Our contribution is a robust set of descriptive features, which range from general to more specific, and it is able to overcome the most important problems in pollen recognition: the inter-class similarity and the intra-class variation. Robustness is provided by employing features related to different pollen characteristics such as shape, size, texture and apertures. Moreover, an unified flexible aperture detector is employed for first time in pollen classification instead of multiple fixed algorithms. Finally, a relevant set of features is