Sketch of an Automatic Image Based Pollen Detection System

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

[1] i3mainz, Fachhochschule Mainz, Lucy-Hillebrand-Strasse 2,
55128 Mainz, Germany.
{gildardo.lozano, matthias.uhler, frank.boochs}@fh-mainz.de
[2] Le2i, Université de Bourgogne, B.P. 47870,
21078 Dijon Cedex, France.
{yannick.benezeth, frank.marzani}@u-bourgogne.fr

Abstract: We present the sketch of an automatic alternative method for identifying and quantifying airborne pollen grains under the framework of a project called Personalized Pollen Profiling and Geospatial Mapping. The first stage of the detection system locates and segments potential pollen particles while reducing the dataset size for convenient handling and storing. In the second stage, the proposed classification scheme successively tests different pollen characteristics until the taxon is uniquely identified. Morphological and optical properties of the pollen and image local features are selected based on the study of strengths and weaknesses of the state-of-the-art methods. This scheme allows the system to compute suitable feature vectors to discriminate similar taxa only when necessary, reducing the computational cost.

Keywords: Pollen detection, biological particle characterization, allergenic patient profile, pattern classification, palynology, image processing.

1 Introduction

The fraction of people in the world that suffer from allergic rhinitis is considerably high; estimations indicate from 10% to 20% [1]. For diagnosis purposes, it is necessary to know the sort of airborne pollen taxa and the amount the sufferer was exposed to. Moreover, in order to avoid contact to risky environments, it is also beneficial to forecast pollen concentration by geographic regions. For instance, the allergen distribution in Europe changes every year [2]. This makes the creation of prediction models more difficult due to requirements of more precise data. Thus, the ability of measuring accurately airborne pollen concentration in the environment has become an important purpose for palynology.

Traditional counting methods consist of trapping airborne samples and counting particles manually. This is time consuming, involves costly labor and requires long-trained palynologists. Additionally, it is highly susceptible to human error due to fatigue or inexperience. These factors limit the processing of huge volumes of airborne samples and the necessary information to perform analysis with geographical accuracy in timely manner.

The advent of computer vision had brought plausible possibilities of improving counting methods to palynology. Microscopic object identification is becoming faster and more robust. The ideal system needs to be able to estimate the pollen concentration accurately in time and to affordably gather information from multiple geographic points.

2 Previous Work

A practical and reliable automatic pollen detection system remains still unsolved. Some efforts in this direction are the ASTHMA project that achieved a recognition rate of 77% for 30 taxa and The University of Vigo with a recognition rate of 86% for 3 taxa [3][4].

The Massey University has been working on the development of a semi-automatic pollen detection system for more than 20 years. Results show a recognition rate up to 89% for 19 taxa [5]. The OMNIBUS project experimented on confocal and multiple-layer images with a precision rate of 98.5% and a recall rate of 86.5% on 33 pollen taxa [6]. These two projects have developed prototypes that are already working in the field.

The main approach to tackle the detection problem is the estimation of feature vector that best describes pollen grains. It has been tested shape features [4], color [7], local invariant features [6][8], texture [5], detection of specific pollen structural characteristics [3] and the combination of some of them.

Multispectral reflectance of pollen images has been barely studied. Fourier-Transform Infra-Red (FT-IR) patterns have confirmed that spectral reflectance of some pollen varies from taxon to taxon [9][10].

A quantitative comparison of the aforementioned performances is not always possible because of the variety of used methods and of the studied pollen taxa, with different levels of difficulty.

3 Scheme of the 3P-GM Project

The present work aims to sketch an automatic image-based pollen detection system, under the framework of the Personalized Pollen Profiling and Geospatial Mapping project (3P-GM). The goal of the 3P-GM project is to create an allergic patient profile based on reliable information of most allergenic pollen taxa concentration in the vicinity of the patients, measured at multiple points from personal mobile devices.

Whereas stationary pollen monitors are capable of sampling just a single location to approximate data for extensive regions, the proposed individual mobile pollen trap captures airborne particles in the vicinity of the patient; and therefore with a closer relationship to the concentration of the personal exposure. The device is also able to gather information of the environmental conditions, geo-position and sampling time. Together with the symptomatic framework, this information will bring the analysis of pollen concentration to a new level. For example, it can be analyzed by geo-statistical scientists to improve pollen distribution models and forecasting or by doctors for more precise and personalized medication.

The 3P-GM project demands the processing of a large volume of data within limited time. Besides, the number of the mobile pollen traps is equal to the number of collaborating patients. All the samples from a region are collected in a single laboratory for digitization. The pollen detection system needs to process all the samples within few hours. Additionally, data must be available during the whole pollination calendar and therefore sampling slides have to be replaced frequently. Once the concentration is estimated, it becomes accessible to collaborating patient’s doctors and the involved research community. A representation of the system is shown in Fig.1.

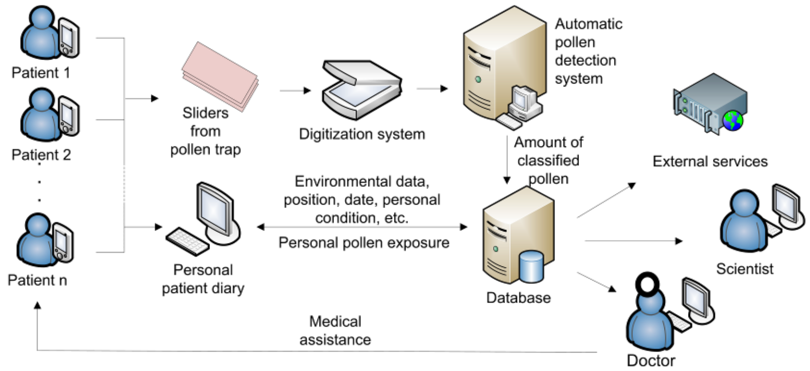

Fig. 1. Representation of the 3P-GM system. Samples from multiple patients are gathered in the laboratory and digitized for identifying and quantifying pollen. This information together with the personal diary and sampling conditions data is accessible to specialists.

4 Automatic Pollen Detection System

The aim of the pollen detection system is to optimally adapt processing to the real characteristics of the pollen, which are most suitable and reliable for robust image-based detection. However, some considerations must be taken on the selection of techniques due to limitations of available technical resources.

Our proposal is based on classical image processing schemes which have been employed in previous successful detection systems. Basically, the system consists of image digitization, pollen-grain localization and segmentation, pollen characterization according to discriminating features and classification of taxa. In this section, the selection of techniques for each component of the system is discussed centered on the analysis of their strengths and weaknesses.

4.1 Image Digitization

The expected robustness requires that enough information can be gathered from the images of the particles in order to reliably characterize pollen. The 3P-GM project focuses on five of most allergenic and frequent pollen taxa in Germany: hazel, alder, birch, sweet grass, and mugwort [11]. Average size of these taxa is from 18-40 microns [12]. The most important characteristics of the grain necessary to identify the taxon: symmetry, apertures, ornamentation, color and morphology of internal structures. In consequence, the imaging system must not only have enough resolution for preserving these characteristics but also be fast enough to process bulks of samples.

The proposed digitization consist of a virtual slide system (VSS) comprising a transmitted light microscope (TLM) with the capability of fluorescence and multiple-layer images. Compared to other microscopic systems, a TLM is less expensive and simpler to operate which makes TLM's able to distribute to several laboratories [13]. The VSS can be controlled automatically to cope with batch processing and multiple-layer imaging with minimal human intervention and within the required time and precision. With a typical magnification in a VSS/TLM of 40x; each pixel in the image captures 0.171microns of the specimen, with a camera sensor pixel size of 6.84 microns and analyzed taxa are visualized in the range of 105 to 233 pixels, which is adequate to represent most of the aforementioned pollen characteristics.

The input is a sampling slide consisting of airborne particles on a sticky medium of size 65mm x 15mm. Although it seems to be a small area, much information can be extracted from it. Indeed, the gathered amount of raw data is such huge that it is not possible to handle it all in an agile manner. Considering only ten 24bit -RGB layers and a fluorescence image with 3 channels of 8 bits, a single sample slide is digitized using one Terabyte. Data reduction is necessary because the data size increases as the amount of samples does.

4.2 Localization and Segmentation

The scheme depicted in Fig. 2 proposes to reduce data size in several steps before being actually processed by the detection system. The VSS scans 10% of the statistically representative region of the sample slide. Then, a localization algorithm detects all particles within a size range on a down-sampled image and extracts stacks of individual snippets from the original images (multiple-layers + fluorescence). This discards most of the empty background and reduces data to 5%. Storing all the resulting particles is important because they are useful for performance benchmarking of different approaches. The final dataset size is of about 5 Gigabytes. In the next step, after validating and segmenting, most of the non-pollen particles are discarded and the dataset is reduced again to 10% with a final estimated total size of 500 Megabytes.

The reduction of data not only avoids the need of large volumes of storage but also free the pollen detection system of performing complex operations over the whole sample image. The data reduction algorithms can be also implemented in a FPGA that considerably improves the processing time.

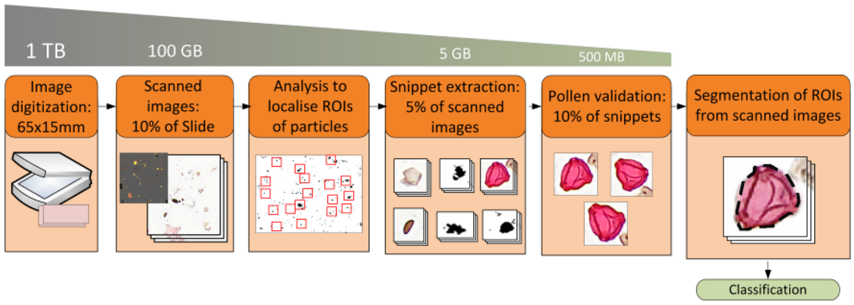

Fig. 2. Input stage of the pollen detection system. Starting with the sample digitization (*left*), snippets of single segmented pollen grains are created. After some size-reduction steps, the data size is about 0.5% of the original scanned image.

Once the pollen detection system receives the dataset with all the particles, it filters the snippets that potentially contain pollen grains based on typical size, color and roundness. In this context, the Hough transform could be employed to detect circular objects [4][6][7]. If more than one particle is inside of a snippet; it is split in several individual snippets. Pollen snippets will be input to the system and the remaining non-pollen will be saved in the data storage.

The next step is to segment the contour of the particle on the stack of layers in detail. The segmentation increases the accuracy of feature computation by considering only pixels belonging to the particle. Due to the volumetric data, a multiple-layer segmentation approach must be regarded. Energy minimization techniques are proposed because of their successful performance in similar tasks [4][6][14]. Additionally, the near-spherical shape of the pollen is taken into account to model the constraints of the segmentation along multiple layers which is especially helpful with overlapped pollen.

4.3 Characterization and Classification

The extraction of particular information that describes pollen grains is critical for the success of the detection system. The internal structure, apertures, texture from ornamentation, exine and color of a pollen grain are particular characteristics that can describe a pollen grain. For instance, 3 apertures, the exine thickness, internal structure and texture patterns can be observed in the pollen grain in Fig. 3. Local features can be also computed at pixel level. Spectral reflectance is a characteristic that is not visible to the naked eye and require special imaging techniques. In order to achieve a robust system, the information must be discriminant enough to distinguish uniquely different pollen taxa and discard non-pollen grains as well as being insensitive enough not to take intra-class variability into consideration. There exist no single feature that can characterize pollen taxa with the aforementioned conditions completely; therefore an optimal mix must be chosen.

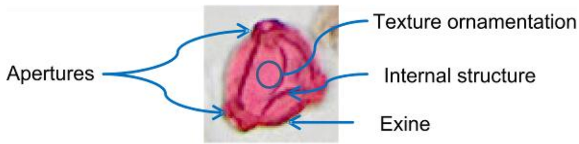

Fig. 3. Important visual characteristics found on a pollen grain. Unknown taxa.

The proposed method is a decision tree that successively computes different feature vectors on the test images until the correct taxon class is found as shown in Fig. 4. At each step, a feature vector is computed and particles are classified into classes of similar taxa, based on the evaluated characteristic. This process is repeated iteratively, testing a kind of feature at a time. The amount of unknown taxa is reduced each iteration until all the taxa are uniquely identified by a single class or there are no more features to test.

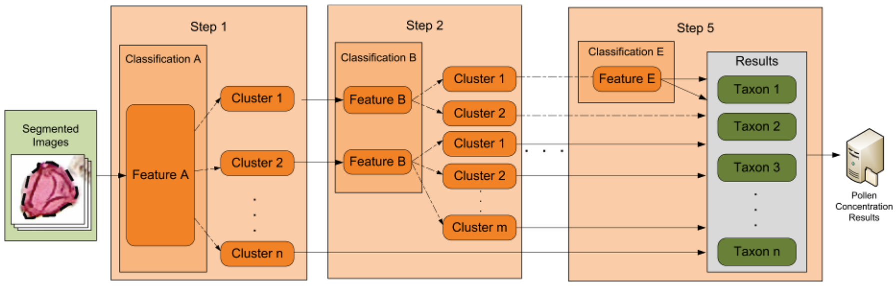

Fig. 4. Output stage of the pollen detection system. At each step, segmented images are classified by analyzing only one feature. Only classes containing more than two taxa are analyzed and classified according to the following feature. Finally, just single-taxon classes are obtained and pollen concentration can be estimated.

A similar classification scheme is proposed in [3] to take advantage of the combination of specific characteristic detectors and geometric features. With this scheme, it is possible to exploit the discriminating capabilities of each kind of feature by dividing the overall classifying system into smaller classifiers with the possibility of selecting the most suitable method in each step. For example, neural networks have been proved to accurately perform on texture feature vectors [5][15]. Minimum distance vector have shown capabilities to classify shape features [3][4]. In the case of local features, the Support Vector Machine approach was successfully employed on 3D invariants in [6].

The proposed classification scheme begins with the simplest and less time-consuming features. Only the classes that cannot be separated in the first steps are committed to further stages of more sophisticated feature analysis. This scheme efficiently employs resources by extracting complex features only when it is actually needed.

The system can have up to 5 levels of classification when all the features are computed: morphological, color, texture, local invariant features and special palynological detectors. We have chosen this order according to the computational complexity but it will be confirmed based on performance results on our pollen samples.

Geometric and second order moment features are morphological measures of the pollen shape that have shown discriminating capabilities as a result of the variation of the symmetry and apertures among taxa [3][4][5]. Color features are used due to the capability of separating pollen from debris [7]. Also, it is possible to show differences among some taxa when surveying on fluorescence images [16]. Correct calibration must be carried out in the system in order to accurately estimate this kind of feature. Different micro-ornamentation and internal structures of each taxon builds particular patterns under the microscope. Thus, texture from pollen had been employed successfully with 20x magnification images in [5] with recognition rates from 88 to 100% for 6 different taxa. This supports our system to analyze texture patterns with a magnification of 40x. Additionally, these global measures of shape, color and texture can provide some robustness, for example to rotation.

Local features computed on volumetric data were successful in [6]. These features are computational expensive because they need multiple-layer images. However, they provide the desired invariance and robustness to translation, rotation and deformation to the system.

The identification and qualification of a set of distinctive characteristics is essential for the recognition of each pollen taxa in traditional manual methods. Examples of these characteristics are apertures, ornamentation and exine as depicted in Fig. 3. Rich palynological information is available and it is not fully employed in automatic detection systems.

Similar to [3], it is proposed that identification of a set of specific characteristics can help to make a decision when a particle is too difficult to classify by means of other kind of features. The detection is taxon-specific and takes advantage of volumetric information from multiple-layer images. This is computationally expensive as the number of needed detectors increases. Therefore, this technique is only used as the last tested feature and only in cases of visually similar taxa.

Finally, when all the taxa are identified and quantified, it is possible to estimate the concentration of a specific sample. This information is loaded to the database containing the allergenic patient profile and the concerning information of the sampling conditions.

5 Conclusions

Experience from previous work shows the need of considering multiple characteristics of the pollen. Some require volumetric information for accuracy. The proposed pollen detection system carefully combines successful methods that are suitable to put into practice according to their strengths and weaknesses.

The system is not fully implemented and no results are available yet. Nevertheless, the 3P-GM project is expected not only to short the distance to the goal of improving the accuracy of the pollen concentration measure, but also of the availability of the data to allergists focused on the relief of the patients. Future work involves to analyzed studied features on samples from our mobile pollen trap and find the most suitable and practicable classifier for each step to identify pollen. It is also expected to study the suitability of analyzing important taxa in France or additional high allergenic taxa such as ragweed and rye.

Acknowledgments. The authors are grateful to the Bundesministeriums für Wirtschaft und Technologie in Germany for its financial support of the project under the program Zentrales Innovationsprogramm Mittelstand ID KF2848901FR1.

References

1. Dykewicz, M. & Hamilos, D.: Rhinitis and sinusitis. The Journal of allergy and clinical immunology 125(2), 103--115 (2010)
2. D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T. & Van Cauwenberge, P.: Allergenic pollen and pollen allergy in Europe. Allergy 62(9), 976--990 (2007)
3. Boucher, A., Hidalgo, P. J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., & Tomczak, R.: Development of a semi-automatic system for pollen recognition. Aerobiologia, 18(3), 195--201, Springer Netherlands (2002)
4. Rodriguez-Damian, M., Cernadas, E., Formella, A., & González, A.: Automatic identification and classification of pollen of the urticaceae family. In: Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS 2003), pp. 38-45. (on CD) (2003)
5. Allen, G., Hodgson, B., Marsland, S., Arnold, G.: Automatic Recognition of Light Microscope Pollen Images. In: Image Vision and Computing New Zealand IVCNZ 06, pp355-360. Massey University, New Zealand (2006).
6. Ronneberger, O., Wang, Q., & Burkhardt, H.: 3D invariants with high robustness to local deformations for automated pollen recognition. In: Hamprecht, F., Schnörr, C., Jähne, B. (eds.) Proceedings of the 29th DAGM Symposium 2007 LNCS vol. 4713, pp. 425--35. Springer, Heidelberg (2007)
7. Landsmeer, S. H., Hendriks, E. a, de Weger, L. a, Reiber, J. H. C., & Stoel, B. C.: Detection of pollen grains in multifocal optical microscopy images of air samples. Microscopy research and technique 72(6), 424--430 (2009)
8. Ranzato, M., Taylor, P., House, J., Flagan, R., LeCun, Y., & Perona, P.: Automatic recognition of biological particles in microscopic images. Pattern Recognition Letters 28(1), 31--39 (2007)
9. Gottardini, Elena, Rossi, S., Cristofolini, F., & Benedetti, L.: Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. Aerobiologia 23(3), 211--219 (2007)
10. Dell'Anna, R., Lazzeri, P., Frisanco, M., Monti, F., Malvezzi Campeggi, F., Gottardini, E., & Bersani, M.: Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. Analytical and bioanalytical chemistry 394(5), 1443--52 (2009)
11. Stiftung Deutscher Polleninformationsdienst. Studien, Analysen und Veröffentlichungen, http:// www.pollenstiftung.de
12. Winkler, H., Ostrowski, R., Wilhelm, M.: Pollenbestimmungsbuch der Stiftung Deutscher Polleninformationsdienst. TAKT-Verlag, Germany (2001)
13. Wu, Q., Merchant, F. & Castleman, K.: Microscope image processing. Academic Press, USA (2008)
14. France, I., Duller, A. W. G., & Lamb, H. F.: A Comparative Study of Approaches to Automatic Pollen Identification. In: Proceedings of the British Machine Vision Conference, pp. 1--10. British Machine Vision Association (1997)
15. Li, P., Treloar, W. J., Flenley, J. R., & Empson, L.: Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. Journal of Quaternary Science 19(8), 755-762 (2004)
16. Mitsumoto, K., Yabusaki, K., & Aoyagi, H.: Classification of pollen species using autofluorescence image analysis. Journal of bioscience and bioengineering 107(1), 90-94 (2009).