Image Anal Stereol 2001;20(Suppl 1):527-532

COLOUR IMAGE IN 2D AND 3D MICROSCOPY FOR THE AUTOMATION OF POLLEN RATE MEASUREMENT

PIERRE BONTON[1], ALAIN BOUCHER[2], MONIQUE THONNAT[2], REGIS TOMCZAK[1], PABLO J HIDALGO[3], JORDINA BELMONTE[4] AND CARMEN GALAN[3]

[1]LASMEA, UMR 6602 du CNRS, Blaise Pascal University, F-63117 Aubière Cedex, France, [2]INRIA, Sophia-Antipolis, 2004 route des Lucioles, B.P. 93, F-06902 Sophia-Antipolis Cedex, France, [3]Department of Plant Biology, University of Córdoba. Campus Universitario de Rabanales, 14071-Córdoba, Spain, [4]Unit of Botany, Autonomous University of Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Spain

e-mail: bonton@lasmea.univ-bpclermont.fr, {Alain.Boucher,Monique.Thonnat}@sophia.inria.fr, bv2hifep@uco.es, jordina.belmonte@uab.es

ABSTRACT

Pollen monitoring is of great importance for the prevention of allergy. As this activity still is largely carried out by humans, there is an ever increasing interest in the automation of pollen monitoring, with the goal of reducing monitoring time in order to plan more efficient treatments. In this context, an original device based on computer vision is developped. In this paper, the colour segmentation techniques implemented on a hardware architecture are presented. The goal of such a system is to provide accurate measurement of pollen concentration. This information can be used as well by palynologists, clinicians or by a forecast system to predict pollen dispersion. The system is composed of two modules: pollen grain extraction and pollen grain recognition. In the first module, the pollen grains are observed in light microscopy and are extracted automatically from a pollen slide coloured with fuchsin and digitized in 3D. In the second module, the pollen grains are analyzed for recognition. To accomplish recognition, it is necessary to work on 3D images and to use deep palynological knowledge. This knowledge describes the pollen types according to their main visible characteristerics and to those which are important for recognition. Some pollen structures are identified, like the pore with annulus in Poaceae, the reticulum in Olea and similar pollen types or the cytoplasm in Cupressaceae. Preliminary results show correct recognition of some pollen types, like Urticaceae or Poaceae, and some groups of pollen types, like reticulate group.

Keywords: colour image processing, markovian image segmentation, pollen identification, transmitted light microscopy.

INTRODUCTION

Automatic recognition of pollen grains is a relatively new application in computer vision. There have been studies trying to differentiate aerobiological spores by image analysis (Benyon *et al*., 1999) or identify pollen texture by neural networks (Li and Flenley, 1999). Recently, work has been presented on pollen recognition using 2D statistical classification (Jones, 2000) or using 3D gray scale invariants with confocal microscopy (Ronneberger, 2000).

The original aspects of our approach to pollen recognition are the combination of statistic-based and knowledge-based techniques, the use of 3D and colour information, and the use of external information regarding to the origin of the grain (sampling date and location).

FIRST MODULE: hardware and image analysis for pollen grain extraction

The first module analyses the pollen slide and extracts the pollen grains. In this section, the hardware of the system first is briefly presented. Then, the software called global slide analysis is presented. This software performs the automation of the 3D image acquisition and the isolation of the pollen grains on the slide using a two dimensions algorithm (Tomczak, 2000).

The input samples are translucent slides which represent daily harvests (Stillman, 1996; Soldevilla, 97). A workstation for both automatic and manual handling and reading of the slides has been designed (Fig. 1). The hardware of the system includes an optical transmitted light microscope equipped with a 40× lens (ZEISS Axiolab), a mono CCD colour camera (SONY XC711) with a framegrabber card (MATROX Meteor RGB) for image acquisition, and a micro-positioning device (PHYSIK INSTRUMENTE) to shift the slide under the microscope. These components are driven by a PC computer. A graphic interface enables the technician to easily operate the system. The semi-automatic pollen counting software (Fig. 2) is implemented on this workstation.

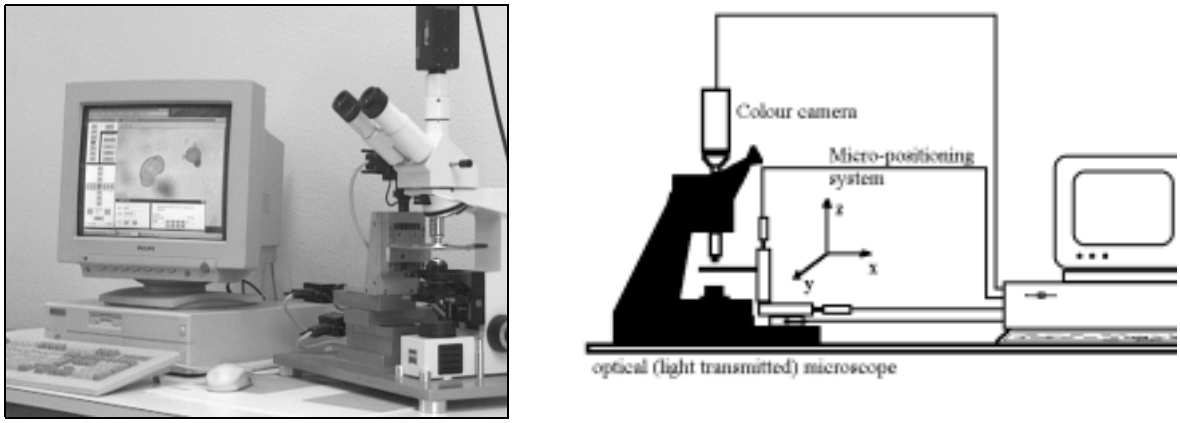

Fig. 1. *Slide analysis workstation.*

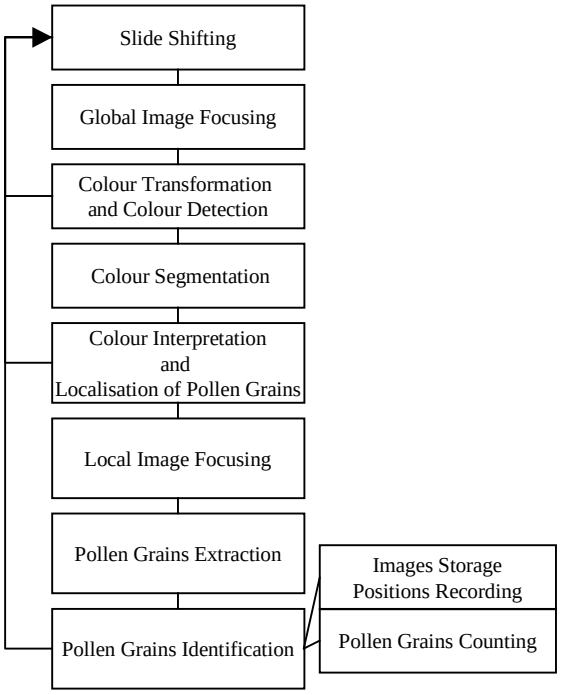

Fig. 2. *Semi-automatic pollen counting software.*

The system needs to extract some information about pollen grains from image data. To achieve this, two problems have been solved. First, autonomous image acquisition in microscopy requires to adjust sharpness in real time before acquiring image data. Therefore, an image automated focusing algorithm has been conceived. It is based on a sharpness criterion computed from image data and on a maximum criterion searching strategy (Tomczak, 1998). It allows the system to compute the best focusing position for a given sample from a small number of measuring positions in real time. Once the image has been focused, the second problem is the detection of pollen grains in the scene. The slides are currently coloured in pink with fushine. However, the variation of coloration among the pollen types is important and some other airborne particles are also sensitive to the colorant. For this

reason, simple segmentation techniques (for instance, techniques only based on chrominance analysis) are not efficient enough to localise and isolate the pollen grains. To resolve this problem, a localisation algorithm based on a split and merge scheme with markovian relaxation has been conceived. It consists of three stages: colour coding (Noriega, 1996), segmentation and interpretation (Rouquet,1998) and detection and extraction of pollen grains (Tomczak, 2000).

In Fig. 3, an example is shown for detection and extraction of the pollen grains from a RGB image. The localisation rate is estimated to be over 90% of the encountered pollen grains on the slides. This rate may be increased if a more precise dye dosing is used in the preparation of slides. However, it is better than the method proposed by France (1997) which performs localisation of 80% of the pollen grains from grey level images.

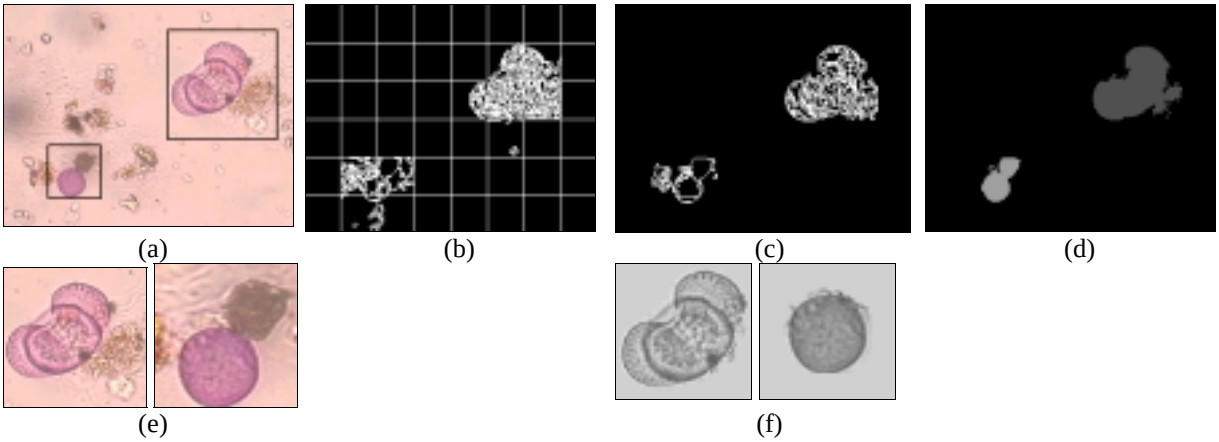

Fig. 3. *Detection and extraction of pollen grains. (a) RGB image (+computed areas of interest). (b) Splitting result. (c) Merging result. (d) Interpretation result. (e) Extracted images from areas of interest. (f) Post-processed grey level images of grains.*

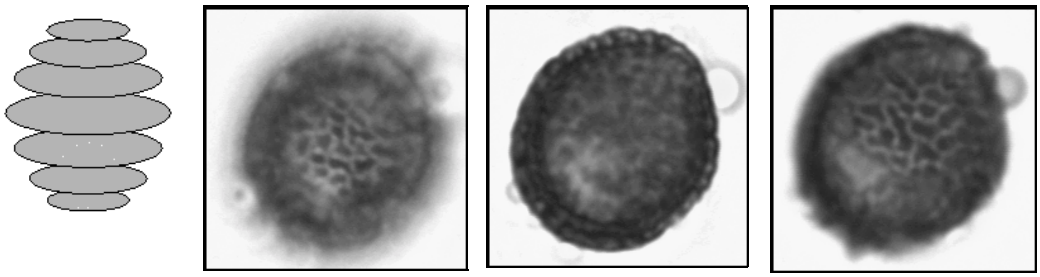

Fig. 4. *Image digitisation in three dimensions. (a) For each pollen grain, a sequence of 100 colour images is taken, showing the grain at different focus (with a step of 0.5 microns). (b-d) Images at different focus of an Olea grain, showing different details needed for its identification.*

Once the central image of a pollen grain is detected, the last step is the acquisition of the whole grain in three dimensions. To achieve this, the system automatically digitize the grain as a sequence of 100 colour images showing the grain at different focus (with a step of 0.5 microns - see Fig. 4). This set of images allows to perform the identification using 3D characteristics.

SECOND MODULE: pollen grain identification

From a sequence of 100 images representing the pollen grain at different focus levels, the next step is to recognize its type. The identification of the pollen grain type is done using two kinds of information:

- Global measures and statistics computed on the central image of the grain
- Type-specific characteristics searched on selected images of the sequence.

The main difficulties for recognition are linked with the particular appearance of pollen grains in

our images. The pollen grains are seen as 3D translucent objects, which are almost spherical, with sizes varying mostly from 20 to 80 microns. They are observed using an optical microscope, as described in the previous section, which can focus only partially on the grains, introducing blur in the digitized images (see Fig. 4).

The first step of recognition is to compute hypotheses of the possible types of an unknown grain. These hypotheses will then be used to guide the next processing steps. The grain is segmented from the central image of the sequence using automatic thresholding techniques based on colour histogram (for example k-means method applied on RGB histograms) and some morphological operations. Some global measures are then computed on the grain. These measures are classical pattern recognition features, as mean colour, size, perimeter, compactness, eccentricity, moments of inertia, convex hull area, concavity, convexity, etc. Such choice has already been used in other applications such as fungal spores differentiation (Benyon *et al*., 1999) or planktic foraminifera identification (Yu *et al*., 1996).

From a database containing 350 reference pollen grains of 30 different types, the system has learnt the covariance matrices representing the different types regarding to their most descriptive measures. The Mahalanobis distance, regarding these measures, is computed between an unknown grain and the existing types. For example, for an unknown grain, one can obtain the following sorted list of possible types with their respective distances: Cupressaceae (2.23), *Coriaria* (2.63), *Platanus* (6.27), *Alnus* (6.69), Brassicaceae (6.86), ...

This list of possible types will be used further to select the characteristics the system will search to prove or to invalidate the initial hypotheses.

Using the leave-one-out technique (Lachenbruch, 1968) on the previous database of digitized pollen grains containing only global measures, some intermediate results of classification around 67% of well-recognized grains are obtained. These are only intermediate results that lead to the conclusion that one needs to look further and include more information in the recognition process, like domain-dependant characteristics.

The second step of recognition is to look for specific pollen characteristics in 3D. Different pollen types can have different characteristics already used by human experts to identify the pollen grains (cytoplasm, pores, reticulum, granules, ...). Such characteristics can be located at different places on the 3D grain and can appear differently depending on the orientation of the grain under the microscope.

Depending on the first hypotheses made about the possible type of an unknown pollen grain, some type-specific characteristics will be tested in order to improve the initial estimations.

The general algorithm for testing a given characteristic for a specific type is:

- 2D segmentation of several selected images
- 3D validation combining all segmentation results.

The recognition system does not need to analyze all 100 images of the digitized sequences to find a characteristic. Only 5 to 10 images of interest may be enough to validate or not the presence of a characteristic. To find those images of interest, two methods are possible. First, the sequence can be sampled to extract n images with a given step. Second, the sequence can be analyzed globally to find the most meaningful images (in terms of clear content and blur). This last method is performed using the SML operator (Sum Modified Laplacian) which provides local measures of the quality of image focus (Nayar and Nakagawa, 1994). Computing this operator for each image of the sequence can identify the clearest images, containing picks and high contrast details with stronger colour variations. Both methods of selection for images of interest can be used, depending on the characteristic that is aimed. On the selected images, it is also possible to select some regions of interest where the characteristic can be found.

Various segmentation algorithms are used to detect the characteristics (automatic thresholding, Laplacian of Gaussian, ...) (Pal and Pal, 1993). The goal here is not necessarily a perfect segmentation, but a sufficently correct one to validate or not the presence of the characteristics. To accomplish the validation of the different segmentations, the same kind of features already used (see previous section)

for the first estimations can be computed again. In addition, other features like the spatial position of the segmented regions and their overlap are computed. The same method of learnt covariances is used for validation, so the result of this is again a list of sorted possible types, which can be combined with the current hypotheses for updating them.

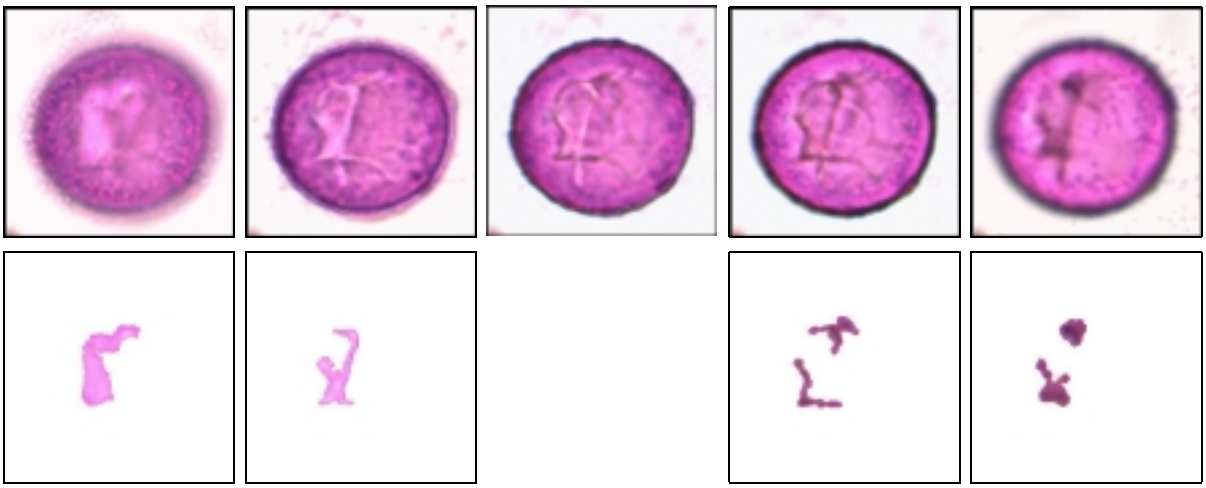

Fig. 5. *Example of type-specific characteristic recognition with the Cupressaceae cytoplasm. 2D segmentations of some selected images around the central images are combined to validate the presence or not of the cytoplasm.*

Fig. 5 shows an example of detection of a characteristic with the cytoplasm of the Cupressaceae pollen type (cypress tree). The cytoplasm is more visible for this type than for others. It is located in the center of the grain, without precise shape, appearing bright for images above the center and dark for images below the center. So the algorithm for detection uses 5 to 7 images, equally distributed around the central image, and looks for bright or dark regions in the center, depending on the location of the image (above or below the central image). The resulting bright and dark regions are compared using several features (shape, colour, size and overlap) for validation.

DISCUSSION AND CONCLUSION

Using this algorithm, the resulting candidate types are different from the candidate types given by global measures. This is a key point for the success of identification. For example, using global measures the similar types of the Cupressaceae type (see Fig. 5) are Plantago, Platanus or Populus. By detecting the cytoplasm, the similar types are Poaceae, Salix and Parietaria, which are different types (not only by their names, but also in appearance). When combining the two lists, it can be expected that the Cupressaceae hypothesis will be enforced. This strategy is used by iterating on several measures and characteristics until no possible confusion remains (or until no other characteristic can be tested).

The recognition system is currently being integrated. The preliminary results of classification using 2D global measures and very few 3D type-specific characteristics for some pollen types shows the recognition of 73% of the pollen grains (database of 350 pollen grains of 30 different types), compared to 67% using only global measures. We aim to improve this result by integrating other characteristics to the system. One goal is to include more characteristics to ensure a level of redundancy in the process of recognition to cope with possible partial occlusions of the grains by dust or other particles.

REFERENCES

Benyon FHL, Jones AS, Tovey ER, Stone G (2000). Differentiation of allergenic fungal spores by image analysis, with application to aerobiological counts. Aerobiologia 15:211-23.

France I, Duller AWG, Lamb HF, Duller GAT (1997). A comparative study of model based and neural network based approaches to automatic pollen identification. British Machine Vision Conference 1:340-9.

Galan Soldevilla C (1997). The use of the Hirst volumetric trap: Operation, adhesive coatings, drum preparation, slide mounting, site location. 3nd European Course in Basic Aerobiology.

Jones AS (2000). Image analysis applied for aerobiology. 2nd European Symposium on Aerobiology. p.2. Vienna (Austria).

Lachenbruch PA, Mickey RM (1968). Estimation of error rates in discriminant analysis. Technometrics 10:1-11.

Li P, Flenley JR (1999). Pollen texture identification using neural networks. Grana 38:59-64.

Nayar SK, Nakagawa Y (1994). Shape from focus. IEEE Trans. Patt Anal and Machine Intell 16:824-31.

Pal NR, Pal SK (1993). A review on image segmentation. Patt Recog 26(9):1277-94.

Noriega LA (1996). A feature-based approach to the problem of colour image segmentation. ICAC'96 1:145-57.

Ronneberger O (2000). Automated pollen recognition using gray scale invariants on 3D volume image data. 2nd European Symposium on Aerobiology. p.3. Vienna (Austria).

Rouquet C, Bonton P, Tomczak R (1998). A comparative study of unsupervised region segmentation strategies by Markov Random Fields. Traitement du Signal 15(1):39-55.

Stillman EC, Flenley JR (1996). The needs and prospects for automation in Palynology. Quaternary Sciences Reviews 15(1):1-5.

Tomczak R, Bonton P (1998). Survey of an image automated focusing algorithm. RFIA'98 2:347-56.

Tomczak R, Rouquet C, Bonton P (2000). Colour image segmentation in microscopy: application to the automation of pollen rates measurement. CGIP'2000, first international conference on Color in Graphics and Image Processing, October 1-4, Saint-Etienne, France.

Yu S, Saint-Marc P, Thonnat M, Berthod M(1996). Feasability study of automatic identification of planktic foraminifera by computer vision. J of Foramineferal Research 26(2):113-23.