



Development of a semi-automatic system for pollen recognition

Alain Boucher^{1,*}, Pablo J. Hidalgo², Monique Thonnat¹, Jordina Belmonte³, Carmen Galan², Pierre Bonton⁴ & Régis Tomczak⁴

¹INRIA, Sophia-Antipolis, 2004 route des Lucioles, B.P. 93, F-06902 Sophia-Antipolis Cedex, France;

²Department of Plant Biology, University of Córdoba. Campus Universitario de Rabanales, 14071-Córdoba, Spain; ³Unit of Botany, Autonomous University of Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Spain;

⁴LASMEA, UMR 6602 du CNRS, Blaise Pascal University, F-63117 Aubière Cedex, France (*author for correspondence: Phone: +33 492387657; Fax: +33 492387939; E-mail: Alain.Boucher@sophia.inria.fr)

Received 2 February, 2001; accepted 17 June 2002

Key words: aerobiology, automatic pollen counting, automatic pollen recognition, colour image analysis, 3D study, light microscopy, pollen morphology

Abstract

A semi-automatic system for pollen recognition is studied for the european project ASTHMA. The goal of such a system is to provide accurate pollen concentration measurements. This information can be used as well by the palynologists, the clinicians or a forecast system to predict pollen dispersion. At first, our emphasis has been put on Cupressaceae, *Olea*, Poaceae and Urticaceae pollen types. 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 grain is analyzed for recognition. To accomplish the recognition, it is necessary to work on 3D images and to use detailed palynological knowledge. This knowledge describes the pollen types according to their main visible characteristics 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. The preliminary results show the recognition of some pollen types, like Urticaceae or Poaceae or some groups of pollen types, like reticulate group.

1. Introduction

This study is developed in the frame of an European Project named ASTHMA (Advanced System of Tele-detection for Healthcare Management of Asthma). The ultimate objective of this project is to provide near real time accurate information on aeroallergens and air quality to the sensitive users on an individual basis and at specific locations to help them in optimising their medication and improve their quality of life. The main goal of this study is to build a prototype system for pollen recognition to be used as a part of the pre-operational integrated forecast system of the ASTHMA project.

Classically, airborne pollen analysis is done by technicians who search, locate, recognize and count

the pollen grains encountered. This pollen analysis is a time-consuming process in which the technician spends many hours at the light microscope collecting the pollen data on a paper sheet. Later, these data have to be translated into a digital database for further processing and analysing them. Recently, the application of computer science to pollen count facilitates the reading process by means of automatic counters (Bennett, 1990), instead of classical mechanical counters. With these systems the resulting data of the human count are directly typed to a computer (Eisner and Sprague, 1987). Recently, Massot et al. (2000) used a voice recognition system to avoid the routine handling of the data. Nevertheless, the contribution of a technician during the pollen searching, detection and recognition can not be avoided until now.

Other studies try to differentiate aerobiological spores by image analysis (Benyon et al., 1999) or identify pollen texture by neural networks (Li and Flenley, 1999). However, an automated system able to detect and recognize pollen grains from a slide has not been described previously to our knowledge.

In this paper, the current work for the development of a semi-automatic system for pollen recognition is presented. This system can be seen as two complementary modules: slide analysis and pollen recognition. The novelty of this system is the use of 3D image processing and palynological information for the identification of pollen grains. The idea of a 3D image processing is not new (Erhardt et al., 1985) and it seems to be a good candidate instead of the classic statistical analysis in 2D.

2. Studied pollen types

Four pollen types have been selected for this study corresponding to the frequent and highly allergenic pollen types in the Mediterranean area, i.e. Cupressaceae, *Olea*, *Parietaria* and Poaceae pollen types. Cupressaceae pollen, inaperturate, with gemmae irregularly distributed in the exine and with a thick intine, belongs to cultivated and wild trees and shrubs widely distributed in the Mediterranean region. *Parietaria* and Poaceae are fundamentally herbs, abundant in most Mediterranean urban and wild environments. *Parietaria* pollen is distinctly small, psilate and triporate, (heptaporate in one species), with visible vestibulum under each pore. Poaceae pollen is medium to large in size, with exine ornamentation from psilate to verrucate and monoporate, with a well defined circular annulus around the pore. *Olea* pollen belongs to olive trees, widely cultivated in the Mediterranean countries. *Olea* pollen type is medium size, presents a coarse reticulum and is tricolpate (tricolporoidate). Besides these four pollen types, a total of 10 similar and also frequent pollen types were included in the image reference database. *Populus* pollen type was included to avoid possible confusions with the other inaperturate and winter pollinated type, i.e. the Cupressaceae type. Other reticulate types were considered in order to avoid confusions with *Olea* type, i.e. *Ligustrum*, *Fraxinus*, *Phillyrea*, Brassicaceae and *Salix* types. *Celtis* and *Coriaria* types (both porate) were included in order to avoid confusions with the monoporate pollen of Poaceae type. And finally, *Morus* and *Broussonetia* pollen

types to avoid confusions with the small pollen of *Parietaria* type. The system works with usual aerobiological slides coloured with fuchsin. A standardization in the fuchsin concentration was done in order to adjust the best colouring for the detection and recognition of the pollen grains. The best concentration was established in 4 μg of fuchsin /100 ml of colouring medium.

3. First module: Image acquisition

The first module of the system performs the global slide analysis that has to isolate the pollen grains on the slide and do their digitization in three dimensions (Tomczak et al., 1998). A workstation was designed for both automatic and manual handling and reading of the slides (Figure 1). The hardware of the system is composed of an optical light transmitted microscope equipped with a x60 lens (ZEISS Axiolab), a 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 was developed to allow the technician to easily operate on the system.

Two problems arise when one wants to extract information about pollen grains from image data in an automated way. First, autonomous image acquisition in microscopy requires to adjust sharpness in real time before acquiring image data. To achieve this, an automated focusing algorithm was conceived. It is based on a sharpness criterion computed from image data and a maximum criterion searching strategy. It allows the computation of the best focusing position for a given sample, from a few 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 stained with fuchsin to colour the pollen grains in pink. This is necessary to differentiate pollen grains from other particles on the slide. However, variations of coloration intensity among the pollen types are important and some other airborne particles are also sensitive to the colorant. Simple segmentation techniques (for instance, techniques only based on chrominance analysis) are not efficient enough to localise and isolate pollen grains. To solve this problem, a localisation algorithm based on a split and merge scheme with markovian relaxation was conceived. It includes three steps: colour coding

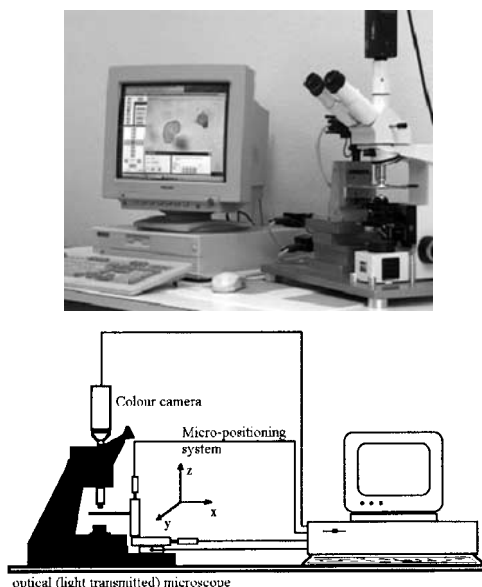


Figure 1. Slide analysis workstation.

following the best colour axes as computed by a principal component analysis on the RGB image (Noriega, 1996); split-and-merge segmentation with markovian relaxation; detection and extraction of pollen grains by chrominance and luminance analysis (Tomczak et al., 2000). Figure 2 shows an example of detection and extraction of the pollen grains. The localisation rate is estimated to be over 90% of the encountered pollen grains on the slides, which is higher than other methods like the one developed by France et al. (1997) which obtains a localisation rate of 80%. Once the pollen grain is found, the last step is to digitize it as a sequence of 100 colour images showing the grain at different focus (with a step of 0.5 microns – see Figure 3). This set of images allows the identification using 3D characteristics.

4. Knowledge acquisition

General and specific palynological knowledge is necessary for the development of the recognition module. General knowledge of palynology like view, ornamentation, apertures, etc. is considered in a first step of the data supply. Table 1 shows the characteristics considered in the general knowledge acquisition for the four main pollen types. Not only is detailed palynological information taken into account but also other characteristics such as the flowering time.

This general knowledge is needed but not sufficient for the recognition of pollen grains based on images. For example, the notion of a pore can lead to various interpretations for an image processing system. It can be seen differently depending on the pollen types, the grain orientation, the grain quality, etc. So, one needs to analyze different examples for each pollen type to extract some more specific knowledge that can be adapted for an automatic system.

To bridge the gap between general knowledge and specific needs for an automatic system, a software tool was designed to interchange information between palynologists and computer scientists (see Figure 4). Using this tool, the experts in palynology can annotate the relevant specific characteristics at each level of the 3D sequence. The annotations refer to more practical information used for recognition.

5. Second module: Pollen recognition

Once a pollen grain has been digitized, the next step is to recognize its type. To achieve this, different image processing techniques can be applied, which include various levels of knowledge from the application (Crevier and Lepage, 1997). The identification of the pollen grain type is based on palynological knowledge (see previous section). It can be achieved by following two steps: compute global measures on the central image of the grain (2D) and search for specific pollen characteristics on the complete sequence (3D). The main differences between these two steps lie in the level of knowledge needed and in the processing time. Global measures can be computed quickly without any knowledge about the pollen grain. All global measures are computed to guide the selection of a few possible specific characteristics that will be tested. The search of these characteristics demand more time to compute and are specific to one or some pollen types.

5.1 Global measures of the grain

The strategy for recognition first isolates the grain on the central image using colour histogram thresholding techniques. Then some global measures are computed on the grain (see Table 2). These measures are similar to what have been used in other applications to recognize different kind of objects, like fungal spores (Benyon et al., 1999) or planktic foraminifera (Yu et al., 1996).

These measures are analyzed to give first estimations about the possible types of the grain. This can

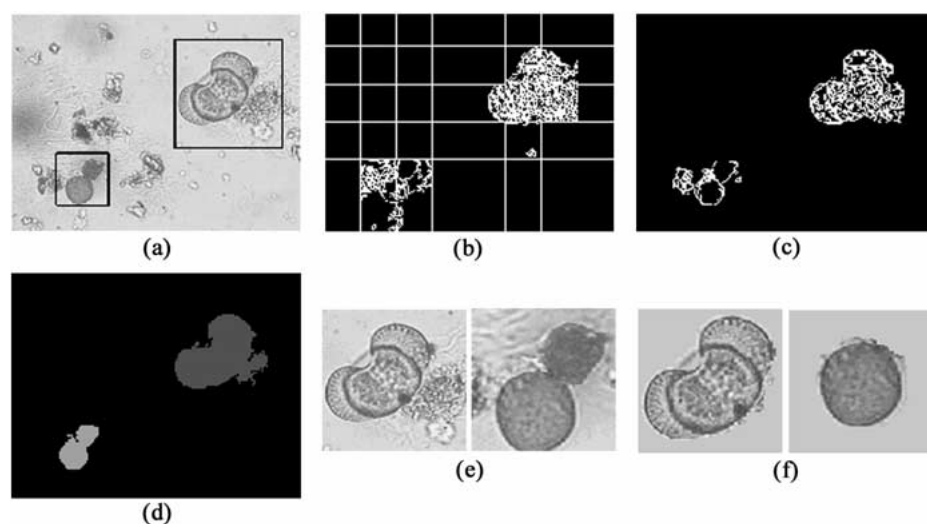


Figure 2. Detection and extraction of pollen grains from a slide. (a) Original image and computed areas of interest. (b) Splitting result. (c) Merging result. (d) Interpretation result. (e) Extracted images from areas of interest. (f) Post-processed images of extracted pollen grains.

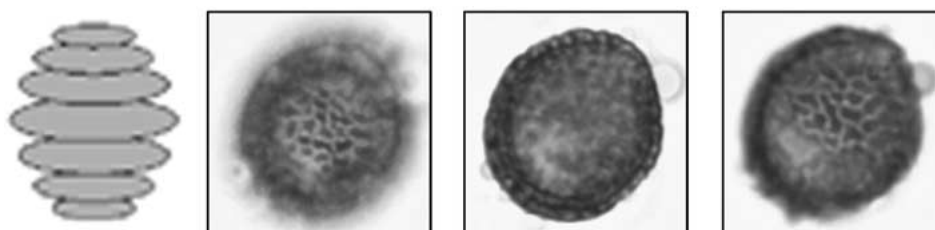


Figure 3. Image digitization in three dimensions. (a) For each pollen grain, a sequence of 100 colour images is taken, showing the grain at different focus (with a step of 0.5 microns). (b–d) Images at different focus of an *Olea* grain, showing different details needed for its identification.

Table 1. Abstract of the general characteristics of the main pollen types considered in this study. This information is a part of the general knowledge on palynology.

Pollen type	Cupressaceae	Olea	Poaceae	Parietaria
Apertures	Inaperturate	Tricolporoidate	Monoporate	Triporate
Size (m: medium, s: small, b: big)	m	s-m	s-m-b	s
Polarity	Apolar or heteropolar	Isopolar	Heteropolar	Isopolar or apolar
Symmetry	Radial	Radial	Radial	Radial
Shape	e: circular	e: circular-elliptic	e: circular-elliptic	e: circular
(p: polar view; e: equatorial view)	p: circular	p: circular	p: circular-elliptic	p: circular
Exine ornamentation	Microgemmate	ReticulateScabrate to verrugate	Psilate	
Exine thickness	1 μm	2–3.5 μm	1–1.5 μm	<1 μm
Flowering period	All the year	April–July	All the year	All the year
(more frequent)	(September–June)		(spring–summer)	(February–July)

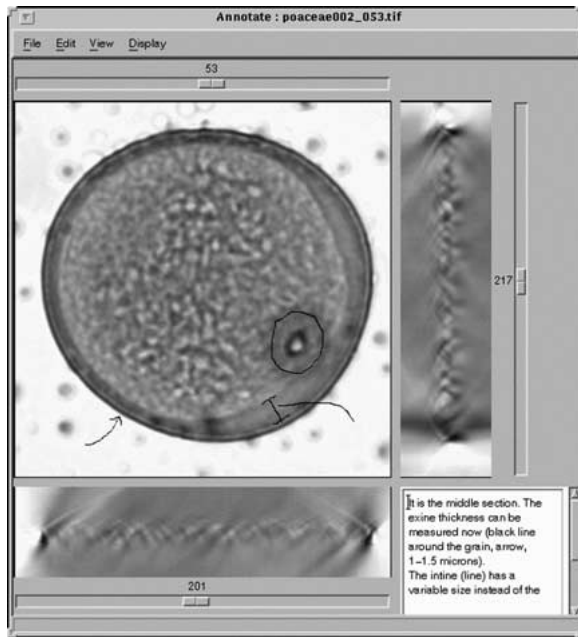


Figure 4. Example of the tool developed for the specific palynological knowledge acquisition in order to exchange information on pollen characteristics. The expert in palynology can look into the sequence like through a microscope and annotate the relevant specific characteristics for the recognition of the grain at each level (Poaceae here).

Table 2. Description of the computed feature measurements. These measures are done on the global pollen grain to initialize a first type estimation and also on the different characteristics extracted from the images to validate them.

Measure	Formula
Mean colour (M)	For red, green and blue colours
Area (A)	Number of pixels
Perimeter (P)	Perimeter length
Compactness (C)	$C = \frac{P^2}{4\pi A}$
Equivalent circular diameter (D)	$D = \sqrt{\frac{4A}{\pi}}$
Moments of inertia (m)	$m_{ij} = \sum (x - \bar{x})^i (y - \bar{y})^j$
Eccentricity (E)	$E = \frac{m_{20} + m_{02} + \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}{m_{20} + m_{02} - \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}$
Convex hull area (CHA)	Area of convex hull
Convex hull perimeter (CHP)	Perimeter length of convex hull
Concavity (CC)	$CC = CHA - A$
Convexity (CV)	$CV = \frac{CHP}{P}$
Solidity (S)	$S = \frac{A}{CHA}$

help not only to guide the system in its choices, but also to identify quickly some special cases (broken Cupressaceae grains for example can be identified easily by their non-circular shape). Moreover, the information of the sampling date and location, if available, can be used to compute these first estimations, by reducing the list of possible pollen types.

One next step that can be done without any a priori knowledge is the selection of the images of interest to process. The system does not need to analyze all the 100 images of the digitized sequence. Only 5 to 10 images of interest can be enough to identify the pollen type, but which images depend on the pollen grain. The selection of the images of interest is done using SML operator (Sum Modified Laplacian) which provides local measures of the quality of image focus (Nayar and Nakagawa, 1994).

One value is plotted per image to indicate its clearness. The local maximum peaks of the curve for the complete sequence indicate the images of interest, representing the clearest images with details on focus and high contrast images with colour variations.

5.2 Type-specific characteristics

Using the first estimations, the system now looks and tests for more specific pollen characteristics. Different algorithms are developed to identify a single characteristic with different appearances. For example, a pore is seen differently from the polar view than the equatorial view. Two different algorithms are needed to detect both cases. Also a region of interest, where the characteristic can be found, can be defined for an algorithm (exine, inside the grain, near the thickest intine, ...).

An algorithm which detects a specific characteristic must include the following elements:

- *Which images will be processed?* The algorithm must select one or several images where the characteristic can be found. For a pore, all the images of interest will be tested, because this feature may appear at different position on the grain. For a reticulum, the images located near the border of the grain when analysing the optical section and in the center when analysing the upper surface (see Figure 5) will be tested, because clear views of the reticulum can be found there.
- *Which image processing techniques will be used?* Depending on the characteristic, various techniques can be used (thresholding, Laplacian of Gaussian, region segmentation, colour analysis, etc.) (Pal and Pal, 1993). Using these techniques,

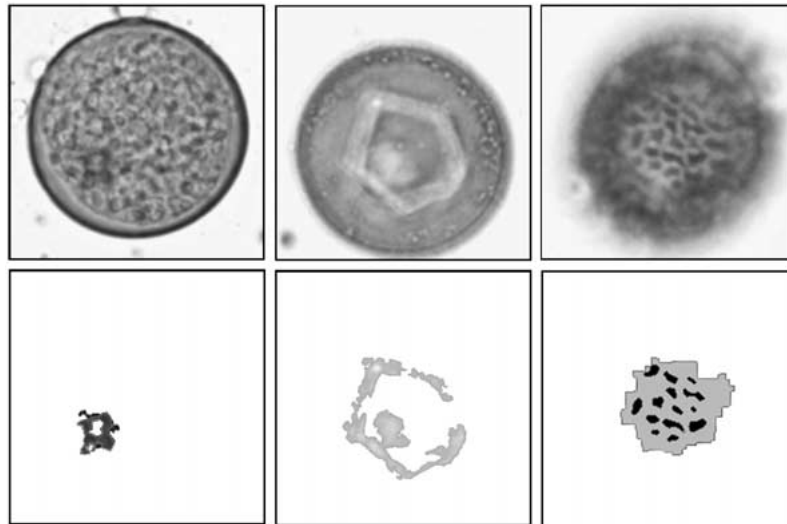


Figure 5. Some examples of characteristic recognition with (from left to right) the pore of Poaceae (annulus), the cytoplasm of Cupressaceae and the reticulum of *Olea* (reticulated zone and lumina extraction).

the characteristics are defined with their colour (dark or bright), their size, their shape and so on.

- *How to validate the detection of the characteristics?* The validation methods are designed to avoid false detections and include comparison between detection on several images of the sequence (3D validation), comparison of the regions with some previously learned parameters of the region (measures of Table 2 can be computed locally on the detection result) and detection of complementary characteristics (annulus and pore for example).

Figure 5 shows some examples of recognition of specific characteristics, with the identification of the pore of Poaceae, the cytoplasm of Cupressaceae and the reticulum of *Olea*. The detection (or not) of such characteristics and the robustness of the given algorithm are used to update the possible type estimations. This process iterates until the pollen type can be clearly identified or until no more characteristics can be tested. In the last case, not only is the highest estimation kept to identify the grain, but also some other possible types can be given to the system operator together with the corresponding probability parameters.

6. Results

The system presented is still under development, but some results are already available to show the feasibility of such a system to recognize and count pollen grains. At present, the pollen grain extraction module has been used and tested to build a database of 350 reference pollen grains (without dust or other particles on the grains) with 10 to 20 sequences per pollen types (30 pollen taxa at total). The results for the recognition module show the recognition of around 77% of the pollen grains (with 30 pollen types) on reference images using the leave-one-out validation method.

7. Discussion and conclusion

The results are encouraging and conclude the need for both steps, global measures and specific characteristics. This last part is necessary to distinguish between visually similar pollen types that have different palynological characteristics. The use of a common language between the system and the palynologist can help to understand the reasoning made by the system and to make possible use of it as a learning tool as well as a counting tool. A semi-automatic system for pollen recognition, as studied in the ASTHMA project, will not eliminate all the human work that must be done, but will reduce the most time consuming and critical task, which is the

recognition and counting process. The technician will still be needed for sample collecting and slide preparation tasks and also to solve the dubious pollen grain determinations. However, with this time reduction factor, it is also possible to increase the number of samplers in the city/country to increase the reliability of the pollen statistics.

Acknowledgements

The authors are grateful to the European Community, project ENV4-CT98-0755, and to ZAMBON Group SpA for financial support.

References

- Bennett K.D.: 1990, Pollen counting on a pocket computer. *New Phytol* **114**, 275–280.
- Benyon F.H.L., Jones A.S., Tovey E.R. and Stone G.: 1999, Differentiation of allergenic fungal spores by image analysis, with application to aerobiological counts. *Aerobiologia* **15**, 211–223.
- Crevier D. and Lepage R.: 1997, Knowledge-based image understanding systems: A survey. *Computer Vision and Image Understanding* **67**(2), 161–185.
- Eisner W.R. and Sprague A.P.: 1987, Pollen counting on the microcomputer. *Pollen et Spores* **29**, 461–470.
- Erhardt A., Zinser G., Komitowski D. and Bille J.: 1985, Reconstructing 3-D light-microscopic images by digital image processing. *Applied Optics* **24**, 194–200.
- France I., Duller A.W.G., Lamb H.F. and Duller G.A.T.: 1997, A comparative study of model based and neural network based approaches to automatic pollen identification. *British Machine Vision Conference BMVC'97* **1**, 340–349.
- Li P. and Flenley J.R.: 1999, Pollen texture identification using neural networks. *Grana* **38**, 59–64.
- Macias Garza F, Diller K.R., Bovik A.C., Aggarwal S.J. and Aggarwal J.K.: 1989, Improvement in the resolution of three-dimensional data sets collected using optical serial sectioning. *J of Microscopy* **153**, 205–221.
- Massot O., Boero E., Coldeboeuf N., Remoleur C., Thibaudon M. and Razzouk H.: 2000, The “C.Scope”: A new application to acquire and process simply and comfortably pollen count on P.C. 2nd European Symposium on Aerobiology. Vienna, Austria, p. 42.
- Nayar S.K. and Nakagawa Y.: 1994, Shape from focus. *IEEE Trans Patt Anal and Machine Intell* **16**, 824–831.
- Noriega L.A.: 1996, A feature-based approach to the problem of colour image segmentation. *ICAC'96* **1**, 145–157.
- Pal N.R. and Pal S.K.: 1993, A review on image segmentation. *Patt Recog* **26**(9), 1277–1294.
- Tomczak R. and Bonton P.: 1998, Survey of an image automated focusing algorithm, Reconnaissance des Formes et Intelligence Artificielle RFIA'98 Clermont-Ferrand. *France* **2**, 347–356.
- Tomczak R., Rouquet C. and Bonton P.: 2000, Colour image segmentation in microscopy: Application to the automation of pollen rates measurement. International Conference on Color in Graphics and Image Processing CGIP'2000, Saint-Etienne, France.
- Yu S., Saint-Marc P., Thonnat M. and Berthod M.: 1996, Feasibility study of automatic identification of planktic foraminifera by computer vision. *J of Foraminiferal Research* **26**(2), 113–123.

