

A persistent homology approach for pollens net structures recognition

P.E. Augusseau¹, M. Jacob¹, A. Bac², K. Anupama³, S. Prasad³

¹Master 2, Polytech Marseille Engineering School

²LIS Marseille

³French Institute of Pondicherry, India

Résumé

This paper investigates a new approach, based on persistence and persistent homology, for pollens classification and recognition from 2D high resolution images. We show that it is possible to identify front and back pores of pollens from a combination of thickness-breadth measures of holes (computed by persistent homology) and an original grayscale persistence defined in the present work.

Mots-clés : Topology, persistence, persistent homology, holes.

1. Introduction

Palynology studies, modern and fossil palynomorphs, such as pollen or spores. It is deeply related to different scientific fields, including geology, botany, paleontology. In the later case, palynology can help with modeling past vegetation and understand the changes which occur in this era. However, most studies (classification, characterisation) are still carried out manually by the lack of efficient automatic analysis tools. Pollens assessments are tedious and time consuming tasks, requiring high quality in palynology.

Recent developments in computer vision are paving the way for the automation of pollens detection and recognition in images. But these works are still partial and fail (among other things) to characterise pollen structures. As pollens are cavernous structures essentially characterised by their holes, the present work intends to investigate a topological approach for pollen analysis. Besides a general geometrical analysis, our method is based on algebraic topology (persistent homology) to detect holes of pollens. Starting from the thickness-breadth measure of holes defined by Gonzalez-Lorenzo and al. ([GLBMR16]), we introduce an original notion of grayscale persistence. Holes are thus identified by a combination of imaging and topological features.

The next section describes a brief state of the art both in palynology and computational algebraic topology. Then in

section 3 we describe data and section 4 gives a brief introduction to algebraic topology. In section 5, we describe our approach and we describe our experiments in section 6. Last, we conclude on our work and perspectives in section 8.

2. State of the art

The first microscopic observations of pollen date back to the 17th century. From the 20th century onwards, with the diversification of chemical preparation processes to extract pollen, as well as the development of light microscopes (LM) and electron scanning microscopes (SEM), palynology underwent a great expansion. Palynologists' studies focus mainly on the morphological characterization of pollens in order to classify them.

Classification is usually done using pattern recognition methods. Algorithms describe the pollen with different variables which can then be sorted by multivariate classifiers. Morphological and geometrical descriptors are the most common choices ([ZFH*04], [CHD*06]). They provide measures of the pollen such as its perimeter, concavity, convexity or circularity. Other approaches investigated texture descriptors in order to compare the "ornamentation" of pollens (see for instance [MLM*13]).

Finally, some studies add other intrinsic characteristics such as pores or colpi to enrich the classification ([CHD*06]). The extraction of these characteristics is done by different image analysis techniques (template matching method, Hough transforms). In this paper, we present a new

method for detecting pollen holes based on homological persistence in order to better characterize pollens with net structure (composed of lots of holes). Our works are preliminary as they provide a new descriptor that should be further integrated into the classification.

3. A brief introduction to pollens structures and pollens data

Automatic pollen recognition from images is still an open problem. Indeed, the appearance of pollens largely varies from one specimen to another. In spite of many attempts, classification and recognition approaches based on the analysis of images remain unsatisfactory. One reason for these failures is that actually, pollens are mostly characterised by the structure of their "emptiness", of their holes. And such structures can hardly be captured by imaging techniques.

Pollens are grains of roughly ovoid shape containing specific hollow structures. Their general configuration is:

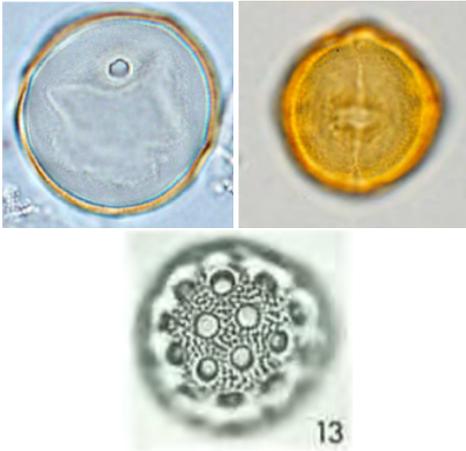


Figure 1: Illustration of the pollen structures : Pores structures (upper left), colporus structures (upper right), net structures (lower)

In the present article, we focus exclusively on net structures (Fig.1 (lower)). Our data are taken from the pollen database of the French Institute of Pondichery and more precisely focus on pollens of the plant family "Amaranthaceae".

Pollens images acquired by optical microscopes provide high resolution colour 2D data. Once cropped to isolate pollens, images typically have a 600×600 resolution. As pollens are light-permeable, focus can be set on different "planes". Fig.2 shows two different focalisations on the same pollen: hence images of a single pollen can be extremely varied (which is a major challenge for pollen analysis).

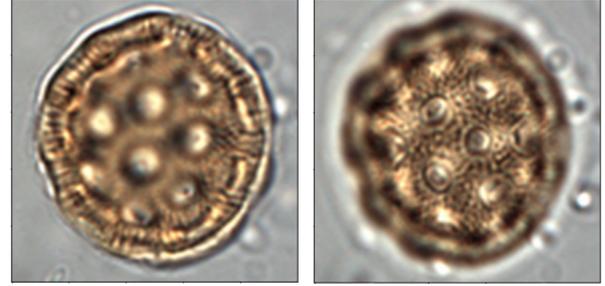


Figure 2: Input images from *Achyranthes Aspera*, first focus (left image), second focus (right image)

4. Computational homology and persistent homology

4.1. A brief overview of algebraic topology and computational homology

Topology studies the shape of objects in their totality/globality. More precisely, topology classifies objects up to a continuous deformation, that is without taking into account their geometry. In this context, algebraic topology intends to characterize the topology of spaces by means of algebraic tools. More precisely, it associates to any geometric object an algebraic object (generally a group) so that properties of this group provide information on the topology of the initial object. This concept contrasts with a geometric approach of spaces and shapes. Indeed, in Euclidean geometry two "objects" are equivalent if they have the same geometric characteristics, in other words, if we can transform one into the other by a rotation, translation ... From a topological point of view, two objects are equivalent if they differ up to a continuous deformation, and thus, topology is actually strongly related to "holes" of the object.

More precisely, algebraic topology operates on discrete structures encoding adjacency (in the present paper, we use cubical complexes which are, roughly speaking, sets of vertices (0-cells), edges (1-cells), squares (2-cells), cubes (3-cells)... plus closure properties). Algebraic topology associates *chain complexes* to these geometric objects. A chain complex C is a sequence:

$$\cdots C_{p+1} \xrightarrow{\partial_{p+1}} C_p \xrightarrow{\partial_p} C_{p-1} \cdots C_2 \xrightarrow{\partial_2} C_1 \rightarrow 0$$

where C_p is a vector space built as linear combinations of p -cells with coefficients taken in a field (or ring) A and ∂_p is a linear application called "boundary application" computing the boundary of cells. Intuitively, the boundary of an edge is the difference between both of its vertices, the boundary of a square is the sum (with appropriate signs) of its neighboring edges... A key property of the boundary operator is:

$$\forall p \in \mathbb{N} \quad \partial_p \partial_{p+1} = 0$$

and hence $\text{Im}(\partial_{p+1}) \subseteq \ker(\partial_p)$. As a consequence, we can define the following group quotient:

$$H_q(C) = \ker(\partial_q) / \text{Im}(\partial_{q+1})$$

$H_q(C)$ is called the q -th homological group of C . It can always be decomposed as $H_q(C) \simeq Z^{\beta_q} \oplus T_q(C)$ (where $T_q(C)$ is called the “torsion group” and is a sum of cyclic groups which is actually 0 up to dimension 3). Coefficient β_q is called the q -th Betti number of C . It has a strong geometrical meaning as it gives the number of holes of dimension q of the object C . Fig.3 illustrates this intuition for an object containing holes of dimension 0, 1 and 2.

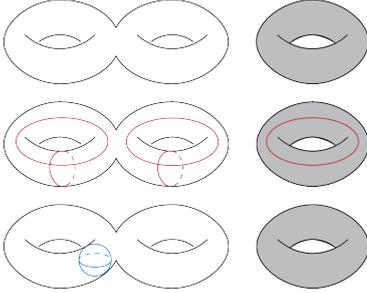


Figure 3: Illustration of Betti numbers on a hollow double torus and a full torus: (up) $\beta_0 = 2$, (middle) $\beta_1 = 5$, (down) $\beta_2 = 1$.

β_0 gives the number of connected components (ie. the number of “clusters” of contactless objects), β_1 the number of tunnels from which we deduce the 1-dimension holes of an object and finally β_2 counts the number of cavity (hence assessing whether the object is hollow or full).

4.2. Persistent homology and measuring holes

Persistent homology is a recent branch of algebraic topology. Inspired by Morse theory, it introduces a notion of “scale” of topological holes thanks. Given a filtration of a complex K (simplicial, cubical...): $K_1 \subseteq K_2 \subseteq \dots \subseteq K_n = K$, persistent homology follows the dynamics of holes. As i increases from 1 to n , holes appear and disappear (as they merge with other holes). Hence, one can associate to each hole a pair of indices corresponding to its birth and death. Such a pair is called a *persistance interval*.

In [GL16] and [GLBMR16], Aldo Gonzalez-Lorenzo and al. Use persistent homology to define geometric measures of topological holes. They first define a filtration “capturing” the geometry of the object based on its signed distance transform. The resulting filtration is:

$$K_{-m} = \emptyset \subseteq K_{-(m-1)} \subseteq \dots \subseteq K_{-1} \subseteq K_0 = K \\ \subseteq K_1 \subseteq \dots \subseteq K_n = B$$

where K_{-1}, \dots, K_{-m} are successive erosions of K , K_1, \dots, K_n are successive dilatations of K and B is the

bounding box of the object. Persistence intervals (i, j) with $i < 0$ and $j > 0$ correspond to holes of K and actually provide two (independent) measures of this hole: $-i$ give the *thickness* of the hole and j give its *breadth*. Fig.4 illustrates these measures. Breadth and thickness are clearly independent as they provide an information about the difficulty, respectively, to fill or break the hole.

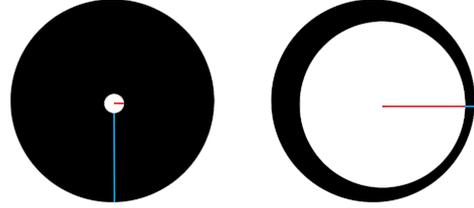


Figure 4: Illustration of the TB-measure of two holes: thickness (red) and breadth (blue).

These measures provide even more precise information about holes as the computation of persistence entails to locate them. Gonzalez-Lorenzo and al. Thus define thickness/breadth balls and prove the stability of such a representation even in the presence of noise. Fig.5 illustrates these balls for a well known 3D-model.

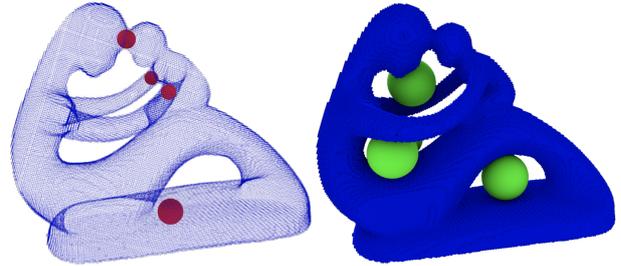


Figure 5: Illustration of TB-balls on a 3D model: (left) breadth balls, (right) thickness balls.

The present work investigates the possibility to use such measures to characterise net structures from images of pollens.

5. Our approach

This section describes our approach as summarized in Fig6.

As illustrated in section 3, pollen images can be extremely diversified and pollens appear on changing backgrounds. Hence, before extracting the pollen net structure, we first preprocess images in order to enhance contrasts, extract contours and remove the background. On the one hand, this preprocess is a basis for calculating geometrical characteristics of the pollen (barycentre, maximum and minimum extension). Such data are actual important characteristics for

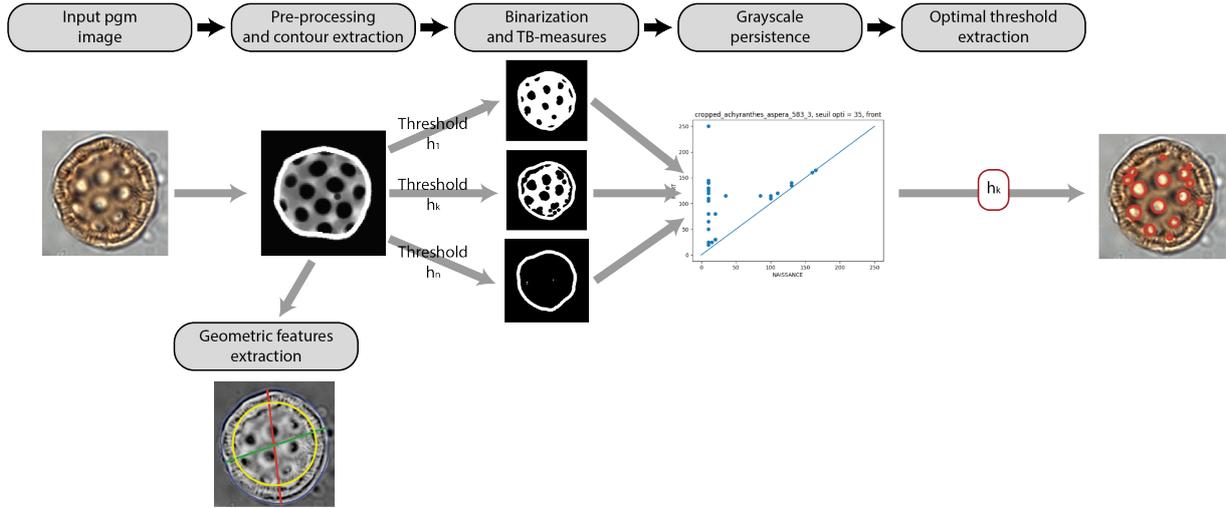


Figure 6: Overview of the approach

Palynologist and their measure “by hand” on images is an arduous task. On the other hand, this preprocess is used to prepare the data to be fed in the TB-holes algorithm [GL16].

Then, for each grayscale value h , we binarize the image, thus creating a “black and white” image, ie. a binary object for which we can compute homological information. For each such binary image, the algorithm then extracts holes and associate TB-balls.

In view of the diversity of images, choosing an optimal grayscale value h is a delicate issue and “traditional” approaches based on histogram computations all failed. However, we show that the sequence of binarized images with thresholds ranging from 0 to 255 forms a filtration. As a consequence, it is possible to compute the persistence of each hole through this filtration. In our case the persistence of a hole is the range of gray levels where the hole is detected in the sequence of binary images. Therefore, the larger the persistence interval, the more relevant the hole. We call “grayscale persistence” this step. Let us eventually point out that, in order to combine this grayscale persistence information with the geometric information about TB-balls, we chose to compute the grayscale 1-dimension persistence of only some of the holes (geometrically significant) and hence introduce an original persistence algorithm (taking advantage of TB-balls).

Finally by computing the corresponding grayscale persistent diagram, we predict the best threshold value h_{opt} to extract holes (see section 5.3).

In this section, we describe in details each step of our approach. The whole code has been developed with Python (with the SciKit-Image library) excepted thickness-breadth measure computation (which has been developed in C++ by Aldo Gonzalez-Lorenzo).

5.1. Preprocessing of images

As described in section 3, our work focuses on images of pollens containing net structures and we investigate the relevance of algebraic topology to assess such structures. However, such measures can only be defined on cubical complexes (and hence on binary objects: images in 2D or sets of voxels in 3D). Hence we first have to convert images into black and white PGM images (which will be the input of TB-measures computation). In addition, as we intend to extract cavernous structures of these pollens, we must define a pre-processing of images that highlights such holes in the images and removes the background.

In this context, we set up the following pre-processing pipeline of pollen images:

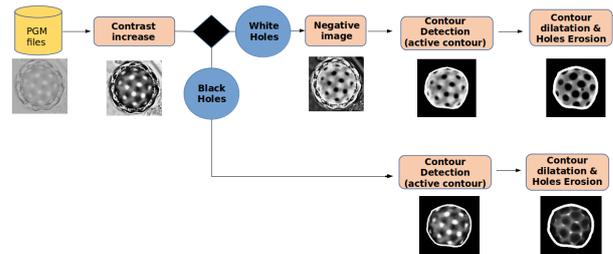


Figure 7: Illustration of pre-processing steps.

Most images obtained from microscopes are very “flat”, hence we first increase the contrast of images. Then, as we will independently compute black and white holes, if necessary, we compute the negative image for further processing. We then extract holes boundaries using an active contour approach (SciKit-Image library). An active contour model

is based on a two-dimensional curve, actually a polyline formed by a sequence of points. The curve (which is close in our case) is placed in the area of interest in the image and an energy is defined to simultaneously evaluate the “distance” between the curve and areas of interest of the image, the smoothness of the curve, it’s stretching... Several equations describe the evolution of the curve (ie. the displacement of its points) towards a curve of minimum energy. Once extracted, this contour is first used to clean the background and perform a 1 pixel erosion of holes to strengthen them.

Then, on the one hand, as described in the next section, we extract from this contour geometrical characteristics of pollens (barycentre, largest and smallest size and corresponding directions). On the other hand, the contour is used, prior to any further topological extraction, to “close” the shape before assessing its holes.

5.2. Geometrical features extraction

As stated earlier, we use the contour extracted in the previous step to compute geometrical descriptors characterizing the boundary of pollens. More precisely, as illustrated on Fig.8 several geometrical values play a major role for palynologists: barycenter of pollens contour, maximum and minimum axis of pollens and its extension along these axes. These simple measures play a major role in pollens classification.

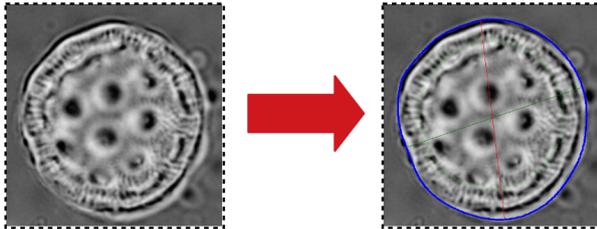


Figure 8: Overview of axes find on *Achyranthes aspera*

Technically, we start from the contour extracted in the previous step (ie. a sequence of points forming a closed polyline), to compute the barycenter of the contour line. Then, by assessing lines going through this barycenter, we extract the maximum and minimum extension directions and sizes.

5.3. Computing holes of pollens images

5.3.1. Binarization and TB extraction

After closing the shape of the pollen (see section 5.1) we create a set of binary images using a threshold h ranging from 10 to 240 with a step of 5 (fig.9). More precisely, given a threshold h and a pixel p of value $\text{val}(p)$:

$$\text{val}(p) \leftarrow \begin{cases} 255 \text{ (white)} & \text{if } \text{val}(p) > h \\ 0 \text{ (black)} & \text{otherwise} \end{cases}$$

For each binary image, we then use the TB-measures algorithm to extract holes of each binary image together with their thickness/breadth balls (ie. measures of holes and center/radius of thickness/breadth balls). Fig.9 illustrates results obtained for various thresholds.

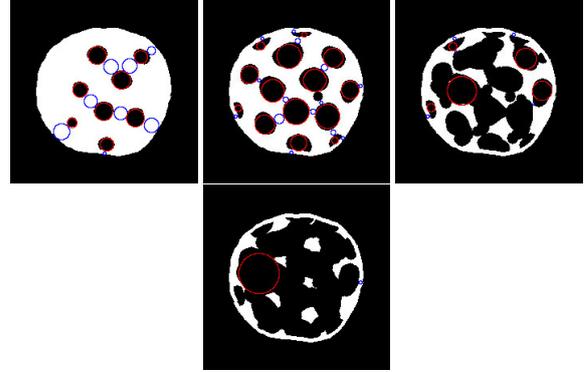


Figure 9: Binarization of a *Atriplex hortensis* 4359 pollen image with various thresholds, (from left to right : threshold 10, 100, 130, 160). TB-measures are displayed for each image through red circles (breath) and blue circles (thickness).

A key issue is thus to choose the right threshold for hole detection. Unfortunately, classical approaches based on histogram analysis or Otsu (local or global) thresholds, do not give stable results. Provided the range of images of pollens obtained by optical microscopes (colors, contrasts, sharpness, focal...), our experiments proved that a correct threshold does not depend on the distribution of pixel colors but on the contrast around holes.

In order to solve this problem, we developed a notion of TB-grayscale persistence described in the next section.

5.3.2. TB-grayscale persistence

As illustrated in Fig.9, the location of holes in binarizations with increasing threshold is not random. On the contrary, it follows a clear organization. By construction of our pre-processing, each set of binary images starts with an image with just a white disk with a black background (threshold 0). Then, as the threshold increases, new holes appear in the white disk and holes eventually merge. Holes eventually totally fill the initial white disk when the threshold reaches the highest gray of the image.

To state it formally, given two gray-level thresholds $i < j$ and I_i and I_j the corresponding binary images, we easily show that $I_i \subseteq I_j$. As a consequence, the sequence of binary images forms a filtration. Hence, it is possible to compute a grayscale persistence diagram of holes through this filtration.

However, our filtration contains a nested sequence of images (not cubical complexes). Moreover, the holes we wish

to extract should be both large enough (geometrically) and with a sufficient contrast with respect to the background. Taking advantage of TB-disks previously extracted, we can estimate a size and location for each 1-hole (respectively the center and radius of its breadth ball). Starting from these disks, introduce a dedicated algorithm which discards smallest holes and computes the persistence of others. This algorithm is illustrated by the pipeline of Fig.10.

We will illustrate this pipeline through the example represented Fig.11.

Let us assume that three holes appear in the red frame at threshold i . Let us now describe how the TB-grayscale persistence algorithm handles these holes. As holes appeared at threshold i , holes are "new" and no comparison is required. All three holes are labeled with new labels and their birth is set to i (by default their death is set to 250 which is the maximum threshold). Let c_k^i be the coordinates of the center of the k th hole breadth-ball (actually the center of the largest disk inscribed in the hole k (red circles)) and let r_k^i denote the radius of this inscribed circle.

Nota: Fig.11 also exhibits blue circles which are the "thickness"-balls of these holes (cf. section 4.2).

When moving to the next threshold ($i + 1$), persistence consists in "following holes" of the previous step to find out whether they are still "alive" (and therefore update their labels) or if they are "dead" (ie. merged with another hole) and therefore update their death.

Let us assess one by one each hole of Fig.11. Hole $h_{k_0}^i$ corresponds to $h_{j_0}^{i+1}$ as their breadth-balls overlap (ie. the distance between their respective centers is less than the sum of their radius). On the contrary $h_{k_1}^i$ does not match directly any hole of step $i + 1$ by such a breadth-ball overlaps. We thus determine its possible matching by incrementally computing the connected component of its breadth-ball in the complementary of the image (here, we find that $h_{j_1}^{i+1}$ overlaps the connected component of $h_{k_1}^i$). Last, $h_{k_2}^i$ directly matches $h_{j_1}^{i+1}$ by the overlap of their breadth-balls.

In case several holes of step i (here $h_{k_1}^i$ and $h_{k_2}^i$) match a single hole of step $i + 1$ (here $h_{j_1}^{i+1}$). In this case, persistence states that "the oldest survives whereas the newest dies" (in our context, we rephrase this rule as follows to avoid any ambiguity: "the oldest (or if both have the same age, the largest) survives whereas the newest dies").

In our case, $h_{k_1}^i$ and $h_{k_2}^i$ were born at the same time, hence $h_{k_2}^i$ "wins" by its size. As a consequence, $h_{k_1}^i$ dies at time $i + 1$ whereas $h_{k_2}^i$ survives. Finally, if new holes were born in time $i + 1$, we would have added them.

Last, let us clarify the values of thresholds. Contrary to persistent homology algorithms, we compute the persistence of 1-holes using breadth-balls. Therefore, there is no need

to decompose the filtration into elementary "1-cell" adjunction steps and hence successive gray-levels only need to be strictly increasing. Hence by "threshold i ", we just mean "threshold θ_i " at step i (with $0 \leq \theta_1 < \dots < \theta_i < \theta_{i+1} < \dots < \theta_N \leq 255$).

5.3.3. Optimal threshold prediction

The set of holes of successive binary images is now labeled so that we can follow them throughout the 10-250 threshold range. Hence our TB-grayscale persistence associates to each hole in this filtration an interval $[b, d]$ with its respective "birth" and "death" dates. We thus obtain a persistence diagram in which each point represents one hole in this filtration.

A point $[b, d]$ of this persistence diagram measures the persistence or "lifetime" of a hole (given by $d - b$). Holes of low lifetimes (ie. points located near the diagonal of the persistence diagram) actually have very low contrast. Such holes are mostly noise.

Therefore, as illustrated in Fig.12, we first apply a filtering on the persistence diagram to remove points close to the diagonal. This filtering is represented by the orange axis.

On the contrary, most relevant holes (red circle) are points located as far as possible from the diagonal. The optimal threshold θ_{opt} is then computed such that the corresponding binary image contains as many relevant holes as possible (which should thus be "already born" and "not yet dead"). Classically in persistence, it is the value θ such that the upper-left quadrant of (θ, θ) contains as many points as possible.

6. Experiments

As described in section 3, our work focuses on pollens of the family "Amaranthaceae". More precisely, we studied three species in this family : *Achyranthes Aspera*, *Sueda Maritima*, *Sueda Monoica*. For the first one, *Achyranthes Aspera*, we used two images of the same pollen taken with two different focus (see Fig.2). We exploit only one image of both remaining species.

6.1. Extraction of geometrical characteristics

The first part of our pipeline proved quite efficient on the pollen images. We extracted correctly the pollen shape for each image, which allowed us to calculate their geometrical characteristics. However, it should be noted that depending on whether we work with the negative image or not, geometrical values vary slightly (see Fig.13).

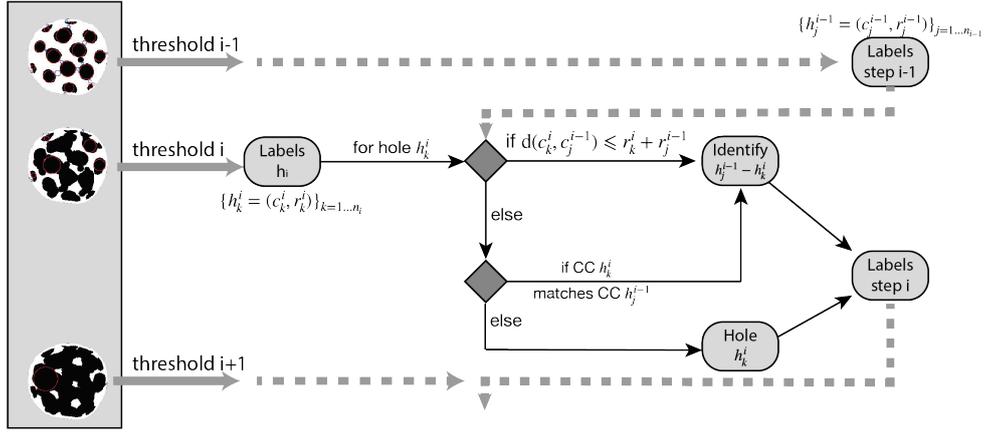


Figure 10: Grayscale persistence computation. At step/threshold i , the image contains n_i holes $h_k^i = (c_k^i, r_k^i)$ where c_k^i and r_k^i respectively denote the center and radius of the breadth-ball of the k th hole.

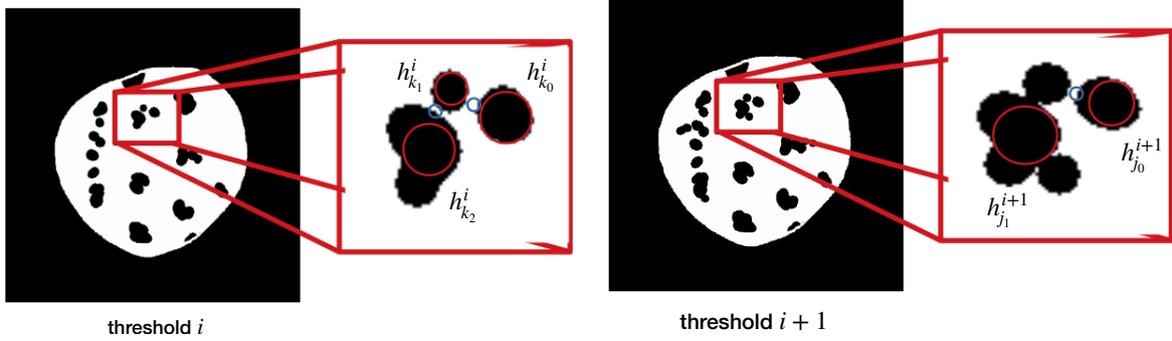


Figure 11: Illustration of TB-grayscale persistence computation: (up) threshold i , (down) threshold $i + 1$.

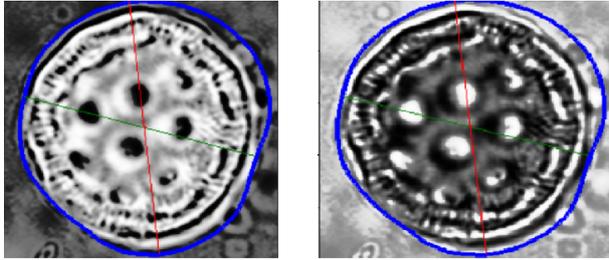


Figure 13: Geometrical characteristics from the first focus of *Achyranthes Aspera*. The left image is the reverse image use for detecting 'front' holes. The blue circle is the pollen shape. Red line is the maximum axis. The green line is the minimum axis.

6.2. Extraction of pollen net structure

The second part of our work is much more complicated to evaluate. In the next section (see section 6.3), we discuss about the solution that we used to estimate the performance

of the algorithm. But in this section, we first illustrate our results by displaying on the input image the holes found with the optimal threshold.

The focus used for the picture seems to have an influence on the performance of the detection. The second focus for the pollen *Achyranthes Aspera* generates lots of details in the center of the pollens and blur the pollen's contours (see Fig.15). These overloads of information, disturb the detection of 'back' holes and the result seems to be unusable.

However, results of the three others pictures are really encouraging. Using the inverse focus improve the detection of the 'front' and 'back' holes in the center of the pollen image. We can notice some noisy holes with a small radius located on the pollen contour including the picture of the first focus of *Achyranthes Aspera* as well as *Suaeda Monoica* (Figs.14,17). Such noise is due to the increase of details on pollen's contour.

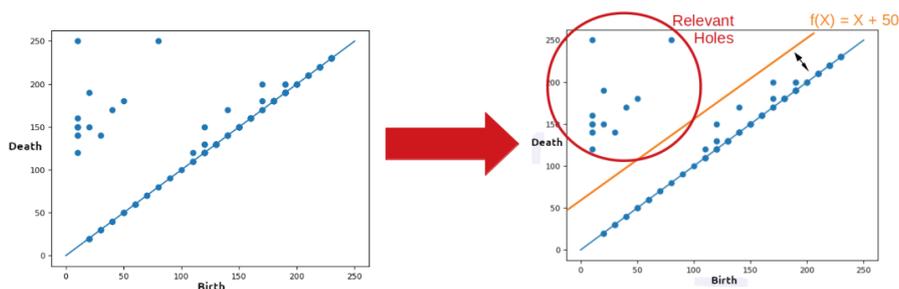


Figure 12: TB-grayscale persistence diagram and selection of the optimal threshold.

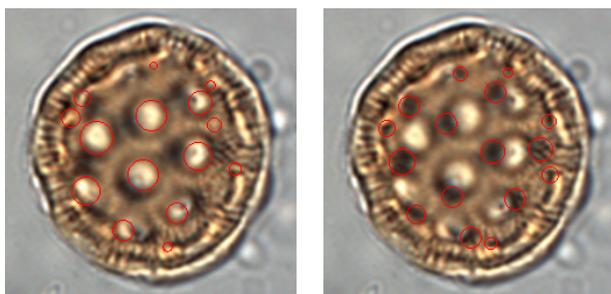


Figure 14: Holes extracted from *Achyranthes Aspera* (first focus). 'Front holes' are red circles in the left image and 'Back Holes' are those in the right image)

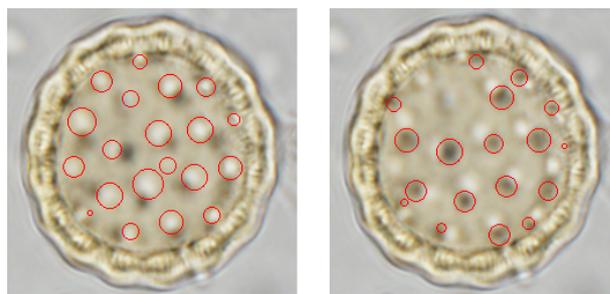


Figure 16: Holes extracted from *Suaeda Maritima*. 'Front holes' are red circles in the left image and 'Back Holes' are those in the right image)

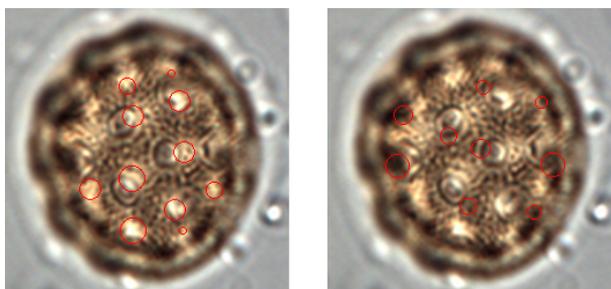


Figure 15: Holes extracted from *Achyranthes Aspera* (second focus). 'Front holes' are red circles in the left image and 'Back Holes' are those in the right image)

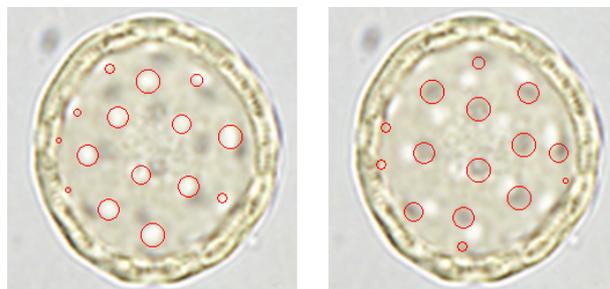


Figure 17: Holes extracted from *Suaeda Monoica*. 'Front holes' are red circles in the left image and 'Back Holes' are those in the right image)

6.3. Validation of the grayscale persistence

To estimate if the grayscale persistence gives us a relevant threshold we face a problem which was the lack of labelled data to compare with our results. We decided to implement a tool that Palynologist can be used to label those dates. It's a small GUI code in python, which allows the user to draw circle in the pollen image to label the different pollen holes in the image (Fig.18). Then the radius and the center of each circle are saved in a TXT file when the image is completely

labelised. The type of holes 'front' or 'back' is also saved in the TXT file.

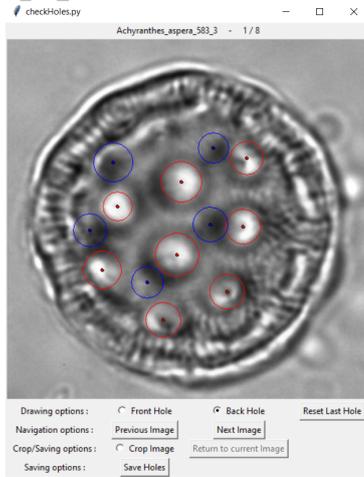


Figure 18: GUI tools - 'CheckHoles'

With these two measures we estimate the performance of holes detection of a threshold based on false positive and false negative. To estimate the number of good detections we compute the number of inclusions between holes from the TXT and the TB file. Then we subtract the number of all multiple inclusions (more than one hole from TXT or TB are included in one hole from the other).

With this number of good inclusions : $goodDetection$, the number of labelled holes in the TXT file : $goodHoles$ and the number of detected holes in the TB file : $detectHoles$, we can calculate the performance P by computing the number of false positive FP and false negative FN .

$$FP = detectHoles - goodDetection$$

$$FN = goodHoles - goodDetection$$

$$P = goodDetection / (FP + FN + goodDetection)$$

The average performance for the eight thresholds compute by the persistence was 55.63% (Tab.1). As we talk on the last session (see Section 6.2), the performance of detection of 'back' holes for the second focus for *Achyranthes Aspera* is really low with 33%. The performance for the detection of the two types of holes for *Suaeda Maritima* is more than 55%. The best performance was perform with the detection of 'back' holes from *Suaeda Monoica* with 79%.

Achyranthes Aspera, first focus	DPFH	42%
	DPBH	50%
Achyranthes Aspera, second focus	DPFH	57%
	DPBH	33%
Suaeda Monoica	DPFH	57%
	DPBH	68%
Suaeda Maritima	DPFH	59%
	DPBH	79%

Table 1: Performance of the computed threshold by the grayscale persistence. DPFH is the Detection Performance for Front Holes. DPBH is the Detection Performance for Back Holes.

7. Discussion

In this study, we tested if persistent homology can provide information about the pollen net structure. The lack of data prevents us to affirm this question. However, the present results are promising. Persistence homology used to compute TB-measures seem to work on both types of pollens holes (front and back).

Results from the section 6.3 show clearly the limits and possibilities of our approach. Fig.19 shows the labelled holes of *Sueda Maritima* compare to holes detected by the threshold of the grayscale persistence. There are some labelled holes invisible to the eye, especially on the pollen contour. These holes are really difficult to identify in the image. Palynologist labeled this area because they pre-identified the pollen image and then deduced an analysis in the image. It seems impossible to detect these holes with any image analysis approach.

Nevertheless, we can identify an area where the algorithm has a really good accuracy of holes detection. This area is the center of the pollen (Fig.19). Here the persistence homology manages to identify pollens holes and geometrical characteristics of these holes (radius, center) seems to fit well with labelled holes. We can notice that the focus used for the input image seems to impact the performance of holes detection.

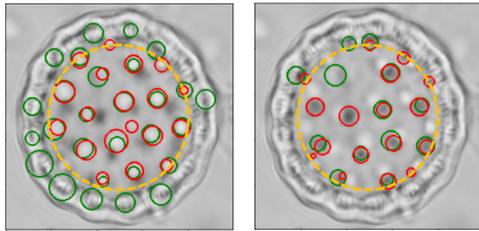


Figure 19: Images of the pollen *Sueda Maritima*. Visualisation of labelled pollen holes, green circle, from the GUI 'checkHoles' and detected pollen holes, red circle, from the threshold of the grayscale persistence. The orange dotted circle is the area where the persistence homology has good results. 'Front' holes are located on the left image and 'back' holes on the right image.

8. Conclusion

Persistence homology appears as a good solution to the extraction of pollen net structure in the image. Our approach gives good results for holes which can be detected by image analysis computation. These holes are usually located in the central area of the pollen. In this area the method can detect two types of holes (front and back) and give some geometrical characteristics of the pollen (barycentre, maximum and minimum extension).

Références

- [CHD*06] CHEN C., HENDRIKS E. A., DUIN R. P. W., REIBER J. H. C., HIEMSTRA P. S., DE WEGER L. A., STOE B. C.: Feasibility study on automated recognition of allergenic pollen : grass, birch and mugwort. *Aerobiologia-2006* (2006).
- [GL16] GONZALEZ-LORENZO A.: *Computational Homology Applied to Discrete Objects*. Theses, Aix-Marseille Université ; Universidad de Sevilla, novembre 2016.
- [GLBMR16] GONZALEZ-LORENZO A., BAC A., MARI J.-L., REAL P.: Two Measures for the Homology Groups of Binary Volumes. In *19th IAPR International Conference on Discrete Geometry for Computer Imagery (DGCI 2016)*, *Lecture Notes in Computer Science (LNCS 9647)* (Nantes, France, 2016), Springer, (Ed.), pp. 154–165.
- [MLM*13] MANDER L., LI M., MIO W., FOWLKES C. C., PUNYASENA S. W.: Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. *Proceedings of the Royal Society B: Biological Sciences*. Vol. 280, Num. 1770 (2013), 20131905.
- [ZFH*04] ZHANG Y., FOUNTAIN D., HODGSON R., FLENLEY J., GUNETILEKE S.: Towards automation of palynology 3: Pollen pattern recognition using gabor transforms and digital moments. *Journal of Quaternary Science*. Vol. 19 (12 2004), 763 – 768.