Automatic Identification of Cell Files in Light Microscopic Images of Conifer Wood

P. Kennel[1], G. Subsol[2], M. Guéroult[3] and P. Borianne[1]

[1] Cirad – UMR Amap , Montpellier, F-34000 France

e-mail: pol.kennel@cirad.fr, philippe.borianne@cirad.fr

[2] LIRMM, CNRS / Université Montpellier 2, France

e-mail: gerard.subsol@lirmm.fr

[3] Inra – UMR Amap , Montpellier, F-34000 France

e-mail: Michael.gueroult@cirad.fr

Abstract—In this paper, we present an automatic method to recognize cell files in light microscopic images of conifer wood. This original method is composed of three steps: the segmentation step which extracts some anatomical structures from the image (based on the watershed algorithm), the classification step which identifies the interesting cells in these structures (by using CART method), and the cell files recognition step. Some preliminary results obtained on several species of conifers are presented and analyzed.

Keywords—Image processing, wood microscopic images, cell segmentation, file identification.

I. PRESENTATION OF THE CONTEXT

A. The biological issue

Within the context of analyzing the development of a conifer tree, the production and the differentiation of wood cells stemming from radial or fusiform initial cells are relevant of variations of the growth context, either from an endogenous (fluxes regulation or liber/wood exchanges) or exogenous (water resources or light conditions) point of view. Cell identification and counting are good methods to show the regularities of cell's types (cambial growth for the initial cells, differentiation for the produced cells). Applying statistical methods on a sufficient number of samples is crucial for such an issue.

B. The microscopic wood anatomy

At cell scale, the internal wood structure of conifers (Fig. 1) is composed of three kinds of structures:

- Tracheids: these cells grow more often in the longitudinal axis of the plant. They are supporting (mechanical function) and conductive (sap conduction) tissues. These cells are aligned along files whose geometrical characteristics give important information about the tree growth.
- Rays: these tissues are made of parenchyma cells which have a storage function.
- Secretory ducts: they are made of secretory cells. Either in longitudinal or in transversal axis, these ducts (with a variable diameter) conduct products of the secondary metabolism of the plant.
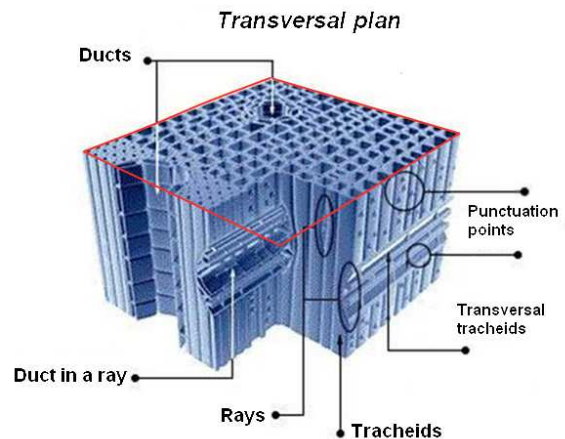

Fig. 1: Microscopic anatomy of conifer wood

In a classical procedure, the cell files, corresponding to tracheid alignments, are studied in transversal wood sections (at the level of the trunk or of branches).

C. Microscopic images and cell files

The cell files are studied on microscopic images of transversal sections of wood, where the different anatomical structures are easily visible as shown in Fig. 2.
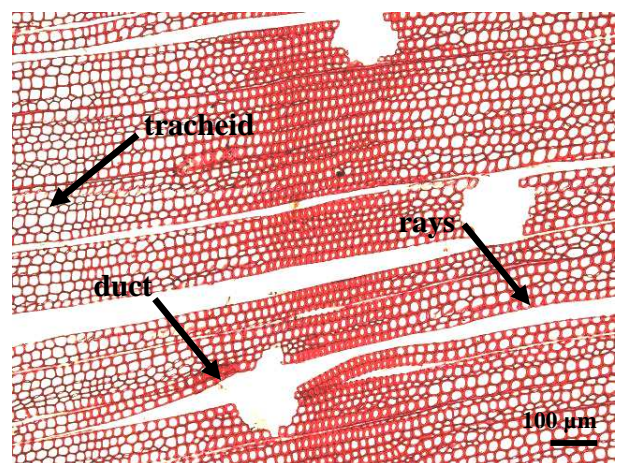

Fig. 2: Light microscopic image of *Pinus kesiya*

For statistical analysis of the wood growth, a great number of samplings must be processed. But counting and typing visually and manually the different wood structures is a very slow and tedious task. Thus wood specialists need a computer assisted method. This paper describes an algorithm for this purpose and presents some preliminary results on real data.

II. CELL FILE PROCESSING

The problem can be divided in two aspects:

- Segmenting and individualizing some regions of interest, namely the tracheid cells.
- Recognizing and characterizing the alignment of tracheids.

Cell segmentation is a classical problem in (bio)-medical imaging and this issue is largely handled in literature [2, 11]. Most of the time, a combination of various standard image processing algorithms like edge detection, filtering, thresholding, mathematical morphology is used. Many commercial (e.g., Visilog [13]) and open source (e.g., CellProfiler [18]) image processing softwares have been designed for cellular structures segmentation, tracking or counting.

However, much less work has been done on wood cell identification. In [3], wood confocal microscopy images are segmented by histogram thresholding, but this quite simple method will not be robust enough with respect to different contrast levels given by other acquisition systems as light microscopy. On the same type of image, [1] proposes to use a watershed algorithm as a first segmentation step which allows one to select some cell candidates. The result is then improved by merging some adjacent regions. This method could be transposed to light microscopy images. In [6], the Fast Fourrier Transform (FFT) is used to reduce the two-dimensional image data to one-dimensional data, from which lumen tracheid shape can be evaluated. In [12], some morphological operators are used to classify some plant cells in a microscopic image of a cereal plant.

In [7], an approach using image processing and data classification to characterize different anatomic structures in wood is presented. The method seems enough efficient to be developed for cell files recognition. Nevertheless, the segmentation step is based on a simple greyscale thresholding and will fail in the case of our images where the color dynamic (which is related to cell maturity) is very variable.

At last, a specific commercial software, WinCell [14], is available for wood cell analysis but is based on quite simple algorithms and does not include a cell file recognition functionality.

We present in section 3 a new method developed for cell files detection on color conifer image, inspired by some of the previously quoted work. Some results and discussion elements are given in section 4. Section 5 concludes and opens to considered outlooks.

III. MATERIAL AND METHODS

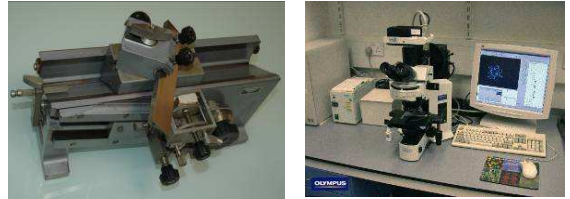### A. Data preparation and acquisition

Fig. 3: Preparation and acquisition

A wood microtome (see Fig. 3 left) is used to obtain thin wood slices. The wood slices are then kept in alcohol and immersed in a safranin staining bath. Alcohol dehydrates cells and the safranin fixes itself on the lignin compounds of cell wall. This ensures a visual contrast between the cell inner content (the lumen) and the cell wall.

Wood lamellas are then placed under an optical microscope Olympus BX51 linked to a LCD camera Olympus DP71 (see Fig. 3 right). During the acquisition, the operator chooses the zoom level and can tune some parameters as color balance, contrast or depth of field in order to get accurate digital images.

It is noticed that the overall protocol is manual. This implies that a non negligible part of the variability (e.g. the cutting plane orientation) or the defects (e.g. the non-planarity of the slice or the non-uniformity of the staining) which are observed in images, mainly depends on the operator experience.

The acquisition software Olympus Cell^A produces 24-bit color images with a dimension of around 2,000×1,500 pixels.
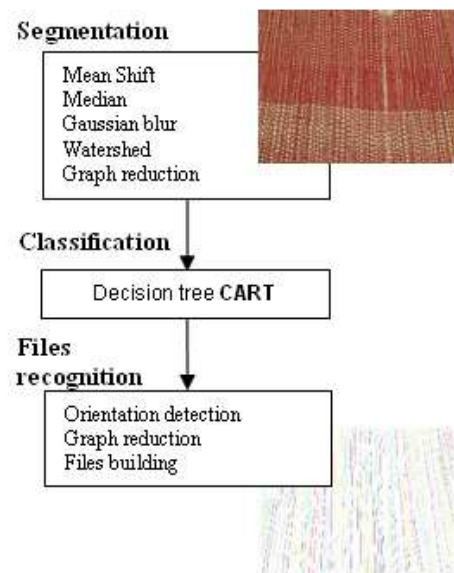
B. Method overview

Fig. 4: Overview of the cell file identification algorithm.

The method can be divided in three steps (see Fig. 4):

- The image segmentation which allows to extract some anatomical structures.

- The classification of the extracted structures in order to identify tracheid cells.

- The cell files recognition by identifying consistent alignments of tracheids.

The different steps of the process are now briefly described and some details on the implementation are given in section IV.

C. Filtering and segmentation step

The filtering step aims to improve the image legibility by removing noise and reducing the local variations of intensity. This is performed by applying successively:

- A meanShift filter [8] which reduces the Gaussian noise. In practice it homogenizes the colors by reducing the intensity dynamics.

- A median filter which removes the residual noise considered as impulsional.

- A gaussian blurring that to reduces the level of detail of the image to improve the segmentation step.

The segmentation step aims at producing a partition of the image in cell regions (see Fig. 5). This is performed by applying a watershed algorithm [5]. The image is first converted in grey levels and is considered as a topographical relief where "crest lines" are calculated. Shed basins are thus obtained and give the regions of the partition.

Once the segmentation performed, we define a dual subdivision by building a graph where the nodes represent the regions and the edges the adjacency relationship (see in Fig. 5). The watershed algorithm is known to induce an over-segmentation (this is visible in Fig. 5 where an edge divide the central cell into two regions). We reduce this over-segmentation by analyzing the intensity profiles along the edges of the regions and along the edges of the graph. For this purpose, we analyze the intensity histogram of the region in order to detect two peaks corresponding to the lumen (lighter) and the wall (darker) part. If the profile is not enough contrasted (i.e. traversing successively a lumen, then a wall, then another lumen) we merge the two adjacent regions and then the two nodes of the graph by deleting the edge. In the case of Fig. 5, this post-processing will allow to merge the two central regions (and then the two nodes of the graph).

Fig. 5: Image of *Pinus nigra:* in white, the subdivision obtained by the watershed algorithm on an image of *Pinus nigra*; in black, the dual representation given by the adjacency graph
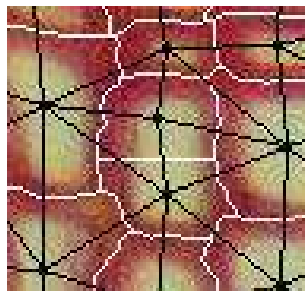
D. Classification step

The subdivision resulting from the segmentation stage is composed of regions which delineate a wood anatomical structure like tracheids, rays or vessels (see § I.B). We now focus on the characterization of the content of those regions in order to identify them and find the tracheids which compose the cellular files.

For each region, a signature vector is computed composed on many parameters based on the global geometry (region area), local geometry (lumen area, wall area), the topology (number of neighbors, mean distance to the neighbors), the intensity histograms (mean, variance, minimum, maximum of the interior of the region, of the lumen, of the wall or along region boundary) and the intensity profiles (along the graph edges).

Based on this signature vector, we want then to class the region in "tracheid" or "non tracheid". As shown in [9], many classification methods (clustering, K nearest neighbors, Bayesian classifier, decision trees…) have been proposed.
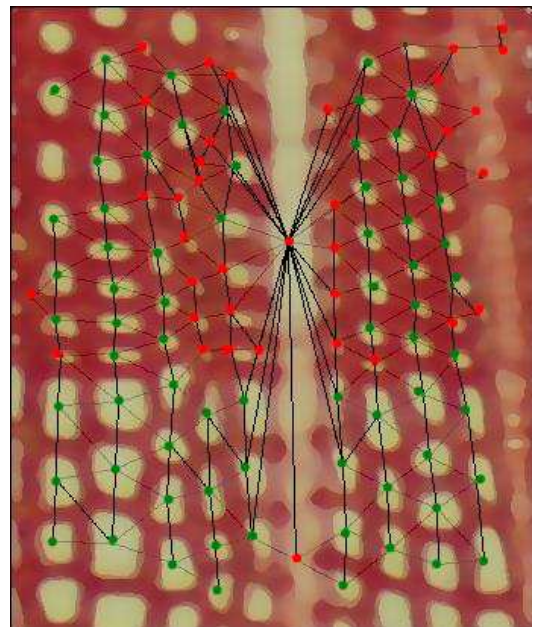

Fig. 6: The identified nodes on a *Pinus nigra image*.

We opted for the classical CART [10] (*Classification And Regression Tree*) method which is a supervised approach with a training step. It is easy to implement and it gives a decision tree which is understandable by the biologists. A training set composed of representative images of about ten conifer species: *Chamaecypavis obtusa, Chamaecypavis thyoides, Pinus banksiana, Pinus kesiya, Pinus longifolia, Pinus monticola, Pinus radiata, Pinus nigr* and *Pinus caribensis* were selected. For all these images, the regions found by the segmentation step were manually annotated as "tracheid" or "non tracheid". The CART method then builds a decision tree from the training data set and generates a classification model, as in the Fig. 7, fully comprehensible for biologists.

Fig. 6 shows the results of the regions classification; in green, the nodes identified as "tracheid cell"; in red, the "non tracheid".
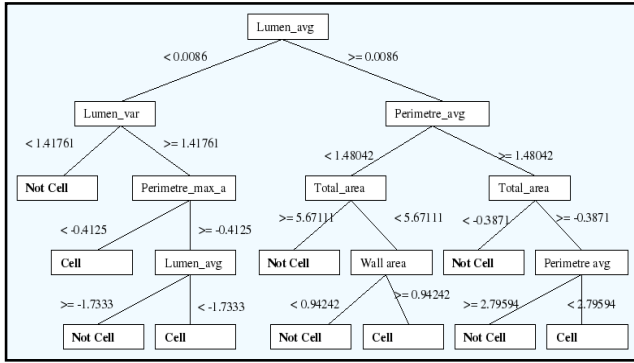
Fig. 7: The CART classification tree used to label regions in "tracheid cell" or "non tracheid"

The generalization accuracy assessed by a 10-fold cross-validation test is more than 95%.

E. Recognition step

This last step consists in using the graph with the "tracheid" nodes automatically classified as being "tracheid" to recognize the files.

1) File orientation analysis

Firstly, we focus on the edge orientation in the graph. The processed images correspond to areas which are small enough to consider that all the files are roughly oriented along the same direction. This main direction can be found by analyzing the angle distribution formed by the edges of the graph with a reference axis (in our case, the horizontal one) [7].
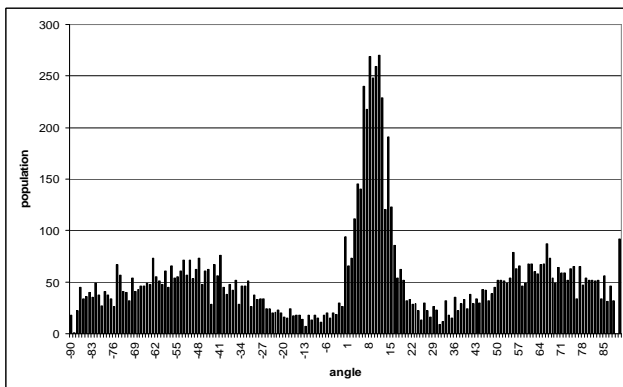

Fig. 8: Angle distribution of the edges in a *Pinus kesiya image (cf. Fig. 2).*

For example, we can observe in the histogram of Fig. 8 a pick which is on the 11° class. This will give the principal direction of the files as one can visually check in Fig. 2.

The adjacency graph is then simplified by removing all the edges which orientation is too far (up to a tolerance tuned by the operator) from the principal direction.

2) Files building

File building is performed in two steps:

- File sections are identified by finding the paths, made of adjacent edges in the simplified graph. Starting from a node classified as "tracheid", nodes are successively added. More precisely, an adjacent node will be added to a given section only if it is located in the neighborhood of the section and if is roughly aligned with the two last vertices of the section.

- These file sections are then linked when they verify the three following conditions:

  o The two sections are at least separated by a given topological distance. This means that there is bounded number of nodes between them.

  o The two sections are geometrically aligned. The alignment is based on the difference between the two orthogonal regression lines computed from the nodes coordinates of the sections. Notice that the section orientation is based on the principal direction; so, all these lines are parallel.

  o Reciprocity condition: the two previous conditions allow finding eventually a best candidate for each section. But we will link both of them only if the best candidate (s2) of one section (s1) also has this section (s1) as best candidate.

F. Implementation

The method has been implemented in Java and integrated as a plug-in in the public domain Java image processing program ImageJ [15]. We use some free libraries (for example the Java Universal Network/Graph Framework [17]) in order to use efficient data structure. The classification part was performed by using Weka [16] which is an open source data-mining software.

Notice that the method is not fully automatic as it uses some thresholds (for example, in the different filtering operators, the tolerance with respect to the principal direction or the maximal topological distance for linking two file sections) which are tuned by the operator and which may depend on the image characteristics.

For the results presented below, we used the following parameters:
- Meanshift (mask radius = 3, RGB radius = 20)
- Median (mask radius = 2) / Blur (mask radius = 2)
- Orientation detection (tolerance function of principal direction = 10°, topological distance = 3).

For a 2,000×1,500 pixel image, the whole method takes around three minutes on a computer with a Intel core Duo 1066 MHz processor and 2Gb of RAM.

The definitive plug-in will be soon available under GPL license at *http://amap-dev.cirad.fr/projects/toaster/files* (use the *Ficeler* module).

IV. RESULTS AND DISCUSSION

A. Preliminary results

The visual representation of the identified cell files (see Fig. 9) in color allows controlling the relevance of the results. In particular, it emphasizes the "broken" files.
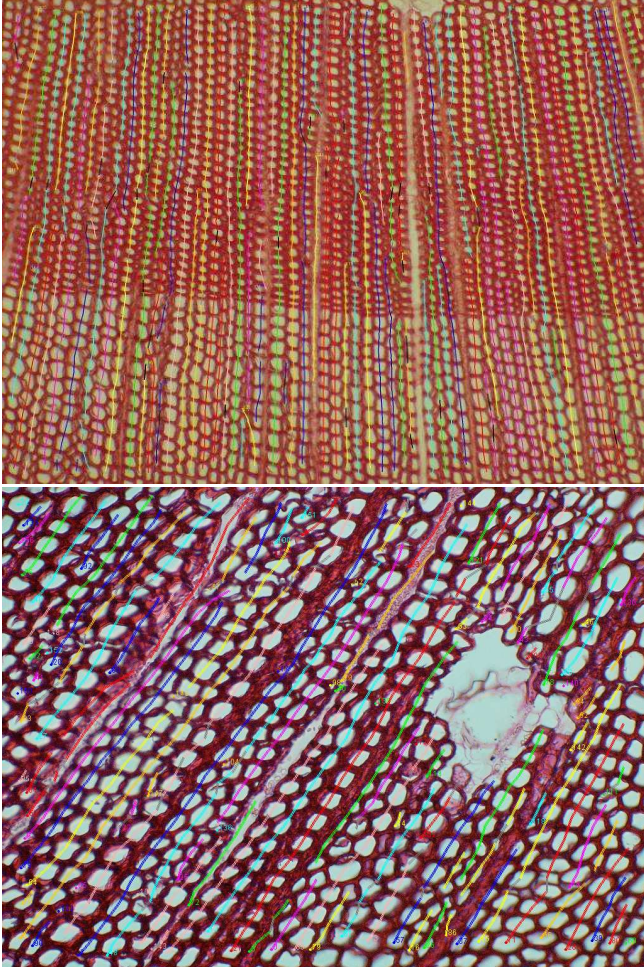

Fig. 9 Automatic identification of tracheid cell files in *Pinus nigra* (top) and *Pinus Caribensis* (bottom).

The cell files recognition observed is encouraging. A good part of files are correctly detected, and can be used as it in statistical analysis.

Moreover, cell files can be qualified by computing some features like the ratio of the file length to the length of the longest file or the ratio of the number of regions tagged as "tracheid " to the total cells number of the file. These parameters could be used to compute a reliability index for each cell file.

In order to assess the quality of the result, we compare the result of the automatic method with some reference files which were manually defined (see Fig. 10)
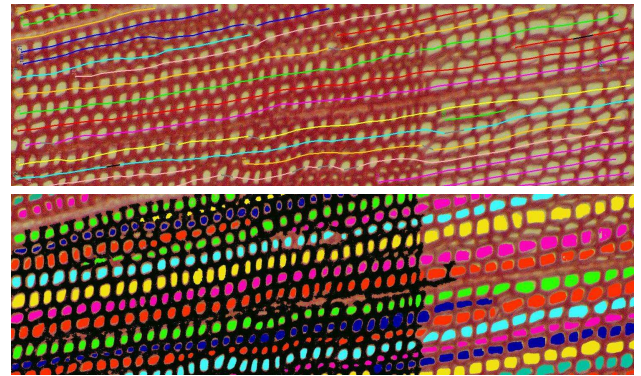

Fig. 10: Comparison of the automatic identification (top) with and an expert tracing (bottom).

B. Some limits of the method

The figure 11 illustrates the files splitting. The automatic method does not separate the two files whereas, the expert, with his knowledge, selected the inferior file to continue.
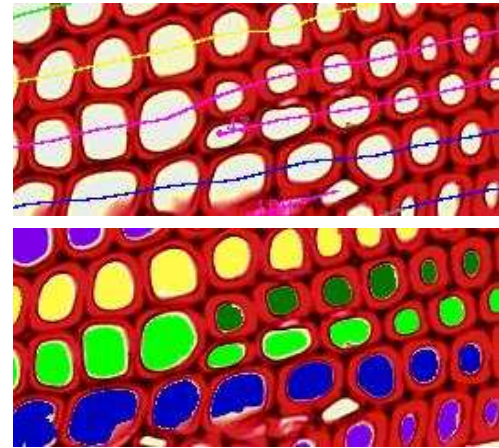

Fig. 11: The files splitting problem

Two scenarii can explain this limit:

- The peripheral growth of tissue can lead to a file splitting on ring switching and two new files may uprise from tone.

- In the case of a section which was not perfectly cut, some cells belonging to a superior or an inferior layer can be introduced in the current layer, creating an artefact file.

As we have seen, the relevance of the cell file identification depends mainly on the segmentation step of the image. For example, in Fig. 12, the algorithm detects two files. In fact, the right extremity of the blue section is an incorrect segmented region: this yields an inflection which forbids the linkage between the blue and red section as the expert does.
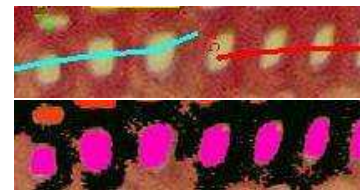

Fig. 12: The non-connection problem

Many others problems can occur. For example in Figure 9 bottom, we can see that cell files are interrupted by a duct (the white spot on the right). This will result to detect two separated sections whereas the expert will enable to connect them in a unique cell file.

V. CONCLUSION AND OUTLOOKS

A. Concluding remarks

We developed an automated method of cell file recognition in conifer species. It is composed of three parts: a segmentation step which extracts some anatomic structures in the image, a classification step which identifies the tracheid cells, and a recognition step which looks for consistent tracheid cell alignments.

The first results obtained on 10 different trees species are encouraging and we checked that on the training data set, a majority of long cell files are correctly detected.

In addition, the analysis of the qualitative and quantitative results may lead to a change of the biologists practices:

- The file discontinuities are easy to detect and the expert will have only to concentrate on these zones to connect or not the sections.

- The reliability index associated to a cell file could be used to discard some "uncertain" data in the statistical process which allows the expert to concentrate only on reliable cell files.

The solution will be soon tested on a large database in order to evaluate expert time-saving.

However, in order to analyze the relevance of our technical choices, some methods used in the segmentation and classification steps should be compared to others.

B. Future work

The proposed method has to be adapted to hardwood species which present a different anatomical structure than conifer species. In hardwood, specific cells ensure posture, and vessels provide a sap conduction of plant whereas in conifers, the tracheid cells carry out both roles.
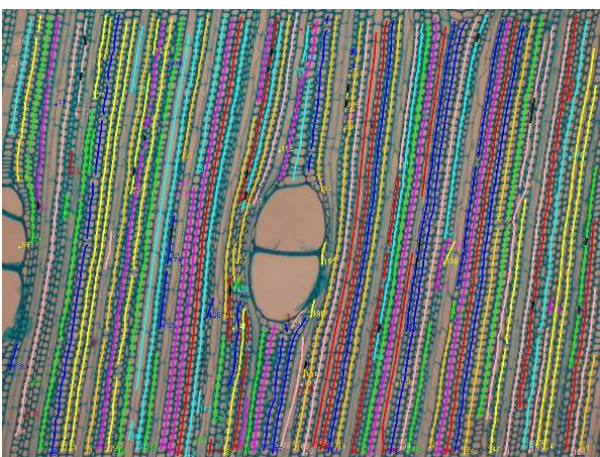

Fig. 13: Cell files identification on a hardwood species, *Meliaceae*

Cell file identification has then to deal with other difficulties like different inner cell types. The cell typology hypothesis have to be adapted to these new features: for example vessels are big cells looking like resiniferous ducts, and the adjacent cell files are distorted (see Fig. 13).

REFERENCES

[1] M. Marcuzzo, P. Quelhas, A. Campilho, A.M. Mendonça, A. Campilho, Automated Arabidopsis plant root cell segmentation based on SVM classification and region merging, *Computers in Biology and Medicine, 39 : 9, 785-793, 2009*

[2] D. Baggett, M.A. Nakaya, M. McAuliffe, T.P. Yamaguchi, S. Lockett, Whole cell segmentation in solid tissue sections, *Cytometry A. 67(2):137-43, 2005*

[3] M.K. Moël, L. Donaldson. Comparison of segmentation method for digital image analysis of confocal microscope images to measaure tracheid cell dimensions, *IAWA Journal, Vol. 22 (3), 2001: 267-288.*

[4] M. Milgram, J.P. Cocquerez, Fermeture de contour par un opérateur local, *Traitement du Signal, volume 3 – n°6, 1986, 303-311.*

[5] L. Vincent, P. Soille, "*Watershed in digital spaces, an efficient algorithm based on immersion simulation*". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol 13, n° 6, 1991.

[6] M.K. Mattias, M. Fujita, Fourier Transform Methods In Image Analysis Of Compression Wood At The Cellular Level, *IAWA Journal, 25 (3), 311–324, 2004*

[7] R. Jones, L. Bischof. A graph--based segmentation of wood micrographs.*In L. Billard and N.I. Fisher, editors, Computing Science and Statistics, 28, Sydney, pages 12-20, July 1996. Interface Foundation of North America.*

[8] D. Comaniciu, P. Meer, Mean Shift: A Robust Approach Toward Feature Space Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603-619, May, 2002*

[9] J. McQueen, Some methods for classification and analysis of multivariate observations, *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and probability, 1:281-297, 1967*

[10] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, *Belmont, California: Wadsworth, 1984.*

[11] T.W. Nattkemper, Automatic segmentation of digital micrographs: A survey, in *Proc. of 11th World Congress on Medical Informatics (MEDINFO), San Francisco, USA, 2004, AMIA/IMIA.*

[12] A.J. Travis, J. Hirstd, A. Chesson, Automatic classification of plant cells according to tissue type using anatomical features obtained by the distance transform, *Annals of Botany, 1996, vol. 78, no3, pp. 325-331.*

[13] *http://www.noesisvision.com/fr/index.html*

[14] *http://www.regentinstruments.com/*

[15] *http://rsbweb.nih.gov/ij/*

[16] *http://www.cs.waikato.ac.nz/ml/weka/*

[17] *http://jung.sourceforge.net/doc/index.html*

[18] *http://www.cellprofiler.org/*