Automatic identification and characterization of radial files in light microscopy images of wood
Guilhem Brunel, Philippe Borianne, Gérard Subsol, Marc Jaeger, Yves Caraglio

To cite this version:

HAL Id: hal-01076468
https://hal-sde.archives-ouvertes.fr/hal-01076468
Submitted on 15 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Automatic identification and characterization of radial files in light microscopy images of wood

Guilhem Brunel\textsuperscript{1,2,*}, Philippe Borianne\textsuperscript{2}, Gérard Subsol\textsuperscript{3}, Marc Jaeger\textsuperscript{2} and Yves Caraglio\textsuperscript{2}

\textsuperscript{1}CIRAD – UMR A51 AMAP, France, \textsuperscript{2}University of Montpellier 2, France and \textsuperscript{3}LIRMM – CNRS, University of Montpellier 2, France

\* For correspondence. E-mail guilhem.brunel@cirad.fr

INTRODUCTION

Tree development is the result of both primary growth, involving the elongation and branching of the axes, and secondary growth, consisting of their thickening over time. Many approaches have been developed to study the structural and functional aspects of plants, particularly when used to address issues of carbon sequestration and wood energy. These approaches, however, whether examining secondary growth and its inter- or intra-specific variations, or the relationships between primary and secondary growth, are often based on only fragmentary studies due to the fact that data acquisition is often costly (Rossi \textit{et al}., 2009).

The description of primary growth can be more or less expensive thanks to retrospective approaches (Barthélémy and Caraglio, 2007; Krause \textit{et al}., 2010) while dendrochronological studies are generally performed only on portions of the ligneous plane (coring). Here, the more detailed the observation level, the shorter the part of the ligneous plane described by the data. Secondary growth originates from cambial activity, the source of the wood (xylem), and thus from the different cells that make up the xylem (ligneous plane).

The biological typing (Lachaud \textit{et al}., 1999) of cells can be approached by examining their shape, size and spatial distribution, independent of their organization. The findings can then be used to determine the anatomical changes that characterize xylem production and differentiation phases (Thibeault-Martel \textit{et al}., 2008; Gue´don \textit{et al}., 2013) without truly explaining wood production or growth mechanisms.

Indeed, the mechanisms of cambial growth can be understood only by studying cell pattern rhythmicity and cell disruption or modification in space and time. A spatio-evolutionary perspective must therefore be introduced to discuss these issues. Environmental fluctuations and their effects on the differentiation of wood elements (originating from the divisions of cambial cells) require production to be monitored in specific cellular organizations. Here, for the sake of simplicity, we consider two organizations: first, the growth ring that reflects cell production over a given time (Heinrich, 2007) and, secondly, the cell file that reflects the activity of the initial cell over time (Rozenberg \textit{et al}., 2004).

Nicolini \textit{et al}., (2003) highlighted the importance of secondary growth organization by studying cell motifs, spatial rhythmicity and the variability of successive growth rings to characterize plant space occupation strategies. This approach does not, however, address the mechanisms of cell differentiation or the construction of cell motifs and structures, which are time-dependent mechanisms. The precise, dynamic monitoring of cambial...
production at the initial cells (Rossi et al., 2006; Rathgeber et al., 2011) is technically limited to a small portion of the ligneous plane. Thus, the study of cell files, which result from cambial production and its local and overall fluctuations, is a promising approach to understanding the establishment, differentiation and temporal rhythmicity of cells (Gindl, 2001).

Biological questions concerning interactions between the development and growth of trees under environmental constraints (Moell and Donaldson, 2001) may be addressed by: (1) determining the contribution made to conduction and storage by the different cell types in the ligneous plane; (2) determining the variability of cell characteristics (vascularization elements, fibres, ray cells, vertical parenchyma); and (3) breaking down this variability by identifying that which can be attributed to ontogeny and that which can be attributed to an environmental response.

To interpret and compare internal observations with observations of external morphology, various parts of the plant’s architecture must be studied simultaneously (Barthélémy and Caraglio, 2007). In particular, observations should distinguish between the constitutive anatomical features of the plant (i.e. trunk, branch) over time, and generate a sufficiently large sample for the use of statistical tools.

However, these biological questions and their associated characterizations are highly constrained due to cumbersome protocols and the difficulties inherent in acquiring large datasets. Indeed, a precise understanding of growth mechanisms and their modelling has mostly been achieved by searching for invariants in large observation samples, thus restoring intra- and interspecies variability.

Progress made in the preparation of complete cross-sections of axes for ligneous plane observation from a macroscopic to a microscopic level without using histological sections (polishing), combined with modern image acquisition equipment (stage for programmed movements, vibratome, etc.), today provide access to quality information on broad areas of the ligneous plane. This technology, however, is hampered by its limited ability to manage and process data. Acquiring a slice 10 cm in diameter at a 200 magnification requires a mosaic of about 1000 images. On average, each image contains 500 cells and the manual evaluation of each image requires about 40 min using an image editor (ImageJ, Photoshop). Cellular arrangements are identified by studying several successive rings in several growth units, and here the number of images required is too large for manual processing. Usually, counts are made on three files per section (Rossi et al., 2006) to save time and reduce costs.

Automating the study of cell files would make it easier to establish tighter links to the functional and ecological aspects of the species by having a better appreciation of the range of phenomena variability. Wood provides a continuous record of all the developmental changes undergone by a tree (ontogeny, the least known) and a record over time of its environment (dendrochronology and dendroclimatology, the most studied). One particular concern with automated methods is the need to assess and quantify the reliability of their results. This means establishing reliability indicators for the results produced.

The automated identification of cell structures is one of the new challenges to be met in studies of the structural biology of plants (Quelhas et al., 2011), and requires a multidisciplinary approach. In matters of (bio-)imaging, cell segmentation is a problem that has been widely discussed in the literature (Baggett et al., 2005; Fourcaud et al., 2008). For example, Park and Keller (2001) combined four conventional image analysis approaches to segment cells: watershed algorithm (Vincent and Soille, 1991), snakes (McInerney and Terzopoulos, 1990), multi-resolution analysis (Jeacocke and Lovell, 1994) and dynamic programming. In the specific context of wood, cell segmentation is generally combined with the determination of cell type, i.e. its biological characterization. For example, Wang et al. (2008) and Marcuzzo et al. (2009) sorted cells using the support vector machines (SVM) classification. This is a supervised incremental training method where the major challenge is to create the training data set that is sufficiently representative to recognize the intrinsic variability of individuals.

In studies of cellular organization, certain authors have used geometric models based on topology. This also provides a description of the cell’s surrounding environment. For example, the studies of Jones and Bishop (1996) and Kennel et al. (2010) were based on a graph that was orientated so as to extract cell files from images of gymnosperms. More specifically, cells are extracted by applying the watershed algorithm to the converted greyscale image. The adjacency graph of the cells is then built from the basin diagram. Cell typing obtained by the supervised classification method (CART; Breiman et al., 1984) can thus be used to extract tracheid alignments from the graph. This method is sensitive to the reinforcement training used to classify biological types.

From a technical perspective, the searches we conducted failed to find any software solutions suitable for the automatic identification of cell files. Specialized commercially available tools such as WinCell (Hitz et al., 2008) are intended for use in the analysis of wood cells, but do not recognize or characterize cellular organizations. Users have only limited or no possibility of improving their functions. By contrast, commercial platforms such as Visilog (Travis et al., 1996) offer a rich environment, but are not sufficiently specialized, because although it is possible to create macro functions (by assembling and configuring basic ones), it is impossible to add new basic functions.

Open source platforms such as ImageJ (Clair et al., 2007) are an interesting option as their source codes may be enriched by special functions, and specialized macros may be developed. At the same time, they facilitate method sharing and dissemination. It is thus possible to develop new functions for automated processing. We thus preferred this solution for our technical development.

In this paper we describe a cell file detection method that is based solely on cell geometry (size and shape), not on biological type, thus avoiding the bias and uncertainty inherent to the creation of reinforcement learning representing biological variability. Reliability indices are employed to determine the accuracy of the results produced for both the geometrical parameters characterizing the cell and its components (lumen and wall), and the identification of cellular alignments. Results are presented from application of the process to real images of wood slices, obtained by a plugin implementation in ImageJ (Schneider et al. 2012), a Java-based image analysis freeware.

**MATERIALS AND METHODS**

**Study species**

For the purposes of our study we examined histological sections and blocks of gymnosperms, specifically *Pinus canariensis,*
P. nigra and Abies alba. Results were extrapolated to angiosperms by studying histological sections of Khaya ivorensis.

Any detailed study of cell typology requires an examination on multiple planes. In our study, we therefore considered only three biological cell types: tracheids, which provide support and are the main cells present in the radial cut; vessels; and rays. For each cell we calculated its circularity, height, wall thickness, and surface area of the cell and its lumen.

Preparation protocol and microscopy

Two preparation methods were evaluated to verify the general applicability of the processing and analytical techniques used.

The conventional preparation method for the study of cell files consists of creating a mosaic of histological sections from wood cores. In practice, this approach does not guarantee the juxtaposition of samples or tissue integrity. It is also relatively time-intensive as it takes approx. 1 h to prepare a mosaic of ten stained cores. In practice, this approach does not guarantee the juxtaposition from basins generated by the watershed algorithm.

The image analysis process followed a three-step sequence, starting with cell identification, then cell geometrical characterization and finally file detection. Cell typing and a determination of the reliability of the results obtained are detailed below.

Each image processing step was itself divided into several tasks, as described below.

Cell segmentation. This first step aimed to identify the different cells in the original image. It was divided into three tasks: image noise reduction preprocessing, contrast enhancement by filtering and colour to greyscale conversion, and cell identification from basins generated by the watershed algorithm.

Image noise reduction. At the magnifications used, the images acquired showed ‘salt and pepper’-type noise, caused by heat from the lamp. This noise was attenuated using a median filter applied on a 5 × 5-pixel neighbourhood.
**Contrast enhancement.** The micrographs showed alternating light areas, corresponding to the lumen, and dark areas, corresponding to the wall. To enhance the contrast between different areas in the image and highlight the walls and lumens, we used a Difference of Gaussian (DoG) operator (Haddad and Akansu, 1991). This filter acts as a band-pass, only letting through the frequency range corresponding to the lumen and filtering out the noise present at higher frequencies. The first image was built from a slightly blurred Gaussian filter application with smoothing factor being selected close to average wall size. The second image was built from a heavily blurred Gaussian filter application with smoothing factor tenfold that used for the first image. The heavily blurred image was then subtracted from the slightly blurred image. These operations were performed on color images, with negative values being set to zero. This process boosts the intensity of the lumen while substantially reducing wall intensities.

**From colour to greyscale images.** The choice of color system is an important aspect of color image processing. Several studies have shown that the 'best system' depends on image contents (Busin et al., 2009). For our study, we preferred the RGB system in which colors correspond to the wavelengths that stimulate the three cones of the human eye. This system can define all the hues but not all the saturations.

However, if the watershed algorithm is to be applied, a color to greyscale image conversion is required, given that our intention was to avoid basing automated processing on human perception. Instead to use the relative luminance formula, which give to greyscale image conversion is required, given that our intention was to use the relative luminance formula, which gives a weighted average of the three color components, we preferred to apply a simple average of the three color channels.

**Cell extraction and identification.** Cell extraction from the grayscale image was based on the watershed algorithm. This powerful algorithm (Vincent and Soille, 1991) considers the image as a landscape, with altitude is given by grey values. Here, the lower points in the topographic relief, corresponding to dark regions, act as catchment basins (as if a drop of water were to fall onto the topography) and are separated by a watershed, represented by the lighter pixels.

These catchment basins are thus obtained from a partition of the image. The crest lines constitute the intercellular border, corresponding to the middle lamella (Fig. 1C). The watershed algorithm is known to produce an exaggerated partitioning, resulting in lines that abnormally intersect cells. This phenomenon is due to the fact that each local minimum in the image produces a potential catchment basin. Typically, two methods are used to reduce or even eliminate the exaggerated partitioning: a priori filtering of the local minima values (Gilmore and Kelley, 1995) used to initialize the watershed, or a posteriori merging of similar adjacent basins in terms of both mean and minimal intensity (Beucher, 2012). Neither method guarantees perfect cleaning of the watershed lines. Here, we propose a method based on the biological observation that the lumen of the cell is empty. Hence, if a watershed line crossing a lumen shows a characteristic intensity profile whose overall maximum is greater than a threshold experimentally set from average lumen values, then it will be deleted. Conversely, a watershed line near the middle lamella contains only values below the low-intensity threshold (Brunel et al., 2012).

**After cleaning the watershed.** An adjacency graph of the basins was used to describe the neighbouring relationships between the different basins. This is a simple graph conventionally defined by a set of vertices and a set of edges. More precisely, each edge of the graph connects the geometric centres of two neighbouring basins, i.e. incidents at the same crest line.

**Geometric properties at the cell scale.** Each cell was then individualized. Cells are composed of two anatomical structures: the wall and the lumen. On the original color image, these structures are automatically divided into two classes by an unsupervised two-means classification algorithm (Forgy, 1965), also equivalent to Otsu’s thresholding operator. The bright pixels class corresponds to the lumen and the dark pixels class to the wall. Surface areas are calculated simply by counting the pixels in each class.

**Cell geometrical properties at larger scales: the global cell arrangement orientation.** When considering a basin in the adjacency graph, it can be seen that the cells are arranged in a staggered manner, and the edges of the graph are orientated in three directions. At the full graph scale, the most frequent orientation corresponds to the cellular alignments. This is defined as the main direction. We will use the method originally described by Jones and Bishof (1996) and reused by Brunel et al. (2012), studying the distribution (via a histogram) of the angle that each edge of the adjacency graph forms with the horizontal. In theory, due to the arrangement of the cells, the distribution should be trimodal and show at least one marked amplitude mode. The mode of the greatest amplitude indicates the most represented orientation in the adjacency graph, i.e. the main orientation.
orientation of the files (Fig. 2). This mode is retrieved by searching the histogram maximum amplitude.

**File detection**

The process used to identify cell files in an image is based on the following assumptions: (1) files are pair-wise alignments of similar cells (in terms of size and shape) and (2) cell alignments are independent of the orientation of the image.

File identification is based on a double scale constructive approach, applied to the adjacency graph. The first step builds cell alignments, i.e. candidate cell files under spatial and geometric constraints. The second step manages the case of alignment recoveries. Finally, the resulting isolated alignments are linked to build cell file fusions.

(1) **Building alignments by applying spatial and geometric constraints.** This task extracts the longer rectilinear straight paths of ‘geometrically similar’ vertices from the adjacency graph. In other words, it detects a vertex sequence along which the surface of the underlying basin varies gradually. These alignments are constructed step by step, by successive aggregations of vertices. More specifically, it means finding which vertex w of the graph must be added at the end of the v line under construction to complete it.

The geometrical similarity constraint. The candidate most geometrically similar to vertex $v_i$ is retained. The geometrical comparison is performed based on their mutual underlying basins. Let $GS_{pq}$ be the geometric similarity index between two vertices $p$ and $q$ that is given by the normalized difference between the surface areas $S_p$ and $S_q$:

$$GS_{pq} = \frac{|S_p - S_q|}{(S_p + S_q)}$$

The more the index tends toward 0, the more similar the cell surface areas. This index is particularly well suited to softwoods in which the shape and surface area of the tracheid varies very little. For angiosperms, the index is a good indicator of rough size ruptures while permitting continuous and progressive surface area variations.

In practice, if the smallest $GS_{pq}$ value is lower than 0.5, the candidate is assigned to the current alignment. If the geometric similarity index of the vertices is greater than 0.5, the candidate is
reduced, the geodesic distance \( k \) is incremented and the process is repeated. The neighbourhood explored is progressively enlarged to a geodesic width of 5, at which point the construction of the alignment stops. This neighbourhood width was an experimental limit set arbitrarily by the wood anatomists involved in the project.

Before validating the alignment as a cell file section, a feedback control test is applied as the construction process should be reversible.

This feedback uses the main orientation of the file as a reference to reduce file drift. When a vertex \( w \) is chosen, a check is performed that the reciprocal is true, i.e. that we obtain current vertex \( v \) when falling back from vertex \( w \) using the main orientation of the files as a directional reference. If reciprocity is not confirmed, the construction of the file stops. If it is confirmed, the vertices located on the shortest path linking \( v \) to \( w \) are added to the cell alignment by the Dijkstra algorithm (Cormen et al., 2001). The criteria employed are described in the next section. Finally, vertex \( w \) becomes the new extremity of the alignment and the search process is repeated until the alignment construction stops.

Once stopped, the detected alignment builds a candidate cell file. The two next steps (recovery management and fusion) then concern the cell file level.

(2) **Managing recoveries.** During construction, some vertices may be assigned to multiple files due to the presence of biological organizations (vessels and channels) that disrupt the staggered arrangement. The assignment of these vertices to the most optimal file is solved by the shortest path search algorithm in the adjacency graph. The Euclidean distance between the geometrical centres of the watershed basin is used as a criterion. This distance sums the angle and shape considered in file building. The greater the angular deviation, the greater the Euclidean distance. The greater the size difference between successive cells, the greater the Euclidean distance. The shortest path in terms of Euclidean distance corresponds to the best alignment. As above, we use the Dijkstra algorithm here, and the corresponding vertices are then removed from the other files, which will be fractionated.

(3) **Linking isolated cell file sections.** Due to the presence of intrusions, cell tears or limitations of the watershed algorithm, the file identification process may cause isolated file sections to appear (Fig. 4). Here, merging can be used to concentrate several parts of the files using simple topological rules to establish whole files, based on file adjacency (Brunel et al., 2012). Let us assume that two files are adjacent when two of their cells are adjacent.

The merging process is then based on two further assumptions: (1) each file crosses the image from one side to the other and (2) the files do not intersect. A file section does not fulfil the first assumption, i.e. one or both tip(s) do(es) not match with the underlying image border.

From those assumptions, we deduce a property resulting from file adjacency.

Given two file sections \( T_2 \) and \( T_2' \) adjacent to files \( F_1 \) and \( F_3 \), it is highly likely that \( T_2 \) and \( T_2' \) are two parts of a single \( F_2 \) file. In other words, \( T_2 \) and \( T_2' \) are located between \( F_1 \) and \( F_2 \).

Note that, under the exposed assumptions, restricted cases of file merging concern only sections between two lines crossing the image from side to side. In other cases, it was decided to leave the situation at the expert’s discretion and evaluation. This method creates files or sections of files with high confidence (Fig. 5, right). Note that isolated cells or artefacts (due to segmentation errors or to poor image resolution preventing an acceptable segmentation) cannot be part of any file and no attempt is made to connect them to a file. Complex cases are left to the discretion of the user.

**Typing**

Cell typing is the ultimate step as used to classify the different cells present (fibrous, tracheids, vessels, rays, etc.). It is based on the geometric and densitometric characterization of the catchment basins associated with the cells’ adjacency graph vertices. Unlike Kennel et al. (2010), we decided to discard supervised classification due to the difficulties encountered in providing reinforcement training that is both sufficiently complete and representative of biological variability.

A decision tree (Fig. 6) was established with wood anatomists and is based on the following observations that are sequentially tested:

1. The perimeter of the cells, denoted \( P \), can be used to differentiate ‘large’ structures from generally ‘small’ cells.
2. The circularity coefficient (Zunic and Hirota, 2008) is given by a weighted ratio between the surface area and the squared perimeter \( -4\pi \text{area/perimeter}^2 \). Circularities range between 0.0 and 1.0: the circularity of a circle is 1, and is far less than for a starfish footprint. Values may not be valid for very small particles due to the definition of perimeter and area in a square grid.

The thresholds used in the decision tree are re-evaluated for each image by automated classification of their numerical values. The classifier used is an implementation of two-means clustering. This conventional data mining method consists of

![Figure 4](https://academic.oup.com/aob/article-abstract/114/4/829/2769059/1)

**FIG. 4.** Linking isolated sections. (A) Result of linkage processing without connection. (B) Result with connection. One colour is used per file.
Fig. 5. Left. The native image; the images show different the preparations (sanded wood and stained histology sections), species (*Pinus*, fir, *Pinus*, mahogany) and clades (gymnosperm and angiosperm) that were processed. Right. Automatic identification of cell files; random colouring is used to enhance the visibility of the identified files (these colours should not to be confused with the reliability index colour).
allocating \( n \) observations in two clusters to minimize the intra- and maximize inter-class standard deviation. The threshold is therefore assigned by the median value between the upper bound of the weakest group and the lower bound of the strongest group.

Beyond the identification of alignments and biological typing, the anatomist requires a set of numerical results characterizing the shape, size and nature of more or less complex biological structures (walls, lumen, cells, files, etc.). These characteristics are defined by parameters that are automatically evaluated. In the context of mass processing, it is advantageous to be able to qualify the accuracy of these evaluations. A certainty index is thus assigned to each one calculated, whether it focuses on the cells, their components or their alignments.

### Reliability

Intuitively, the computed measures will be more or less close to the true value depending on the cumulative errors stemming from image quality, biological configuration and algorithmic approximations. It is therefore important to assign each result a reliability estimation.

A file \( f \) must be sufficiently long to be significant; its length \( L_f \) must be greater than threshold \( L \). \( L \) is defined from the length \( L_f \) distribution analysis. For each image the threshold is re-evaluated to reflect the biological characteristics inherent to each of them as a file must be composed of similar cells in terms of shape, size and appearance. For the sake of simplicity, these three criteria are reduced to the single height of cells defined from all points in the normal direction of the file. This simplification stems from the maturation mechanism of tracheids, for which cell variations only concern cell extension, not thickening. Clearly, this mechanism cannot be applied to broadleaved trees, but in practice this single criterion appears to be sufficient to characterize the continuous variation observed in the cell files of broadleaved trees. As a result, the heights \( H_j \) and \( H_{j+1} \) of consecutive cells in file \( f \) are compared. The reliability coefficient \( R_f \) of file \( f \) is described by a product of standardized terms ranging from 0 to 1.

\[
R_f = \left( 1 - \max \left( \frac{L - L_f}{L}, 0 \right) \right) \prod_{j=0}^{n-1} \left( 1 - \frac{|H_j - H_{j+1}|}{H_j + H_{j+1}} \right)
\]

The closer the coefficient to 1, the more reliable file \( f \).

By analogy, we can determine reliability coefficients for the geometrical parameters that characterize file cells. The expression for this coefficient will depend on the nature of each parameter and more specifically on the factors influencing that parameter. The surface area of the basin is independent of local image blur because it is obtained by watershed segmentation, while the surface area of the lumen is highly sensitive to local image blur because it is obtained by an intensity classification. The reliability coefficient for the lumen is therefore directly related to blur level. Ladjal (2006) defined a blur indicator that characterizes the spread of maximum intensity amplitudes, i.e. the speed at which the signal passes from the lowest to the highest intensity. This is defined by the following expression where \( p(x) \) shows the pixel intensity of a basin and \( p'(x) \) the variance.

\[
\gamma(p) = \frac{\max_{x \in \mathbb{R}} p(x) - \min_{x \in \mathbb{R}} p(x)}{\max_{x \in \mathbb{R}} |p'(x)|} = \frac{\text{amplitude}(I)}{\max(\nabla I)}
\]

When the coefficient is high, blur is strong. The reliability coefficient for lumen surface area \( l_{p(e)} \) is defined below where threshold \( T_s \) is defined by studying lumen variation using a Gaussian blur. Surface areas above the threshold are not significant.

\[
l_{p(e)} = \begin{cases} 0 & \text{if } \gamma(p) > T_s \\ 1 - \frac{\gamma(p)}{T_s} & \text{else} \end{cases}
\]

Note that the accuracy of the surface areas and circularities are sensitive to the size of the objects measured; for example, the circularity of a discrete circle (André, 1994) is meaningful only above 10 pixel rays.

### Parameter computation

Our method produces several layers of results corresponding to different observation levels: (1) radial files are classified according to their length and their cellular self-similarity, (2) cells are characterized by geometric size, diameter, shape, etc., and (3) cell components (wall, lumen) are characterized by geometric parameters (size, thickness, diameter, etc.).

The method has been implemented in Java and integrated as a plug-in for ImageJ freeware. As an indication, the plug-in allows images consisting of 1600 × 1200 pixels to be processed in less than 20 s on a computer with an Intel Q720 processor clocked at 1.6 GHz.

### RESULTS AND DISCUSSION

The study described herein aimed to compare the results produced by our method with those in an expert database containing 12 images whose cells had been manually cropped by wood anatomists. Our results were obtained without any specific user settings (Fig. 5): all parameters were automatically re-evaluated by the application based on the intensity histogram generated for each image. The different tests conducted allowed us to evaluate the limitations of our method, namely with regard to different biological contexts, different preparation protocols, and diverse and varied acquisition conditions. The tests were essentially...
motivated by simple questions: (1) is our method suitable for both angiosperms and gymnosperms? and (2) is our method sensitive to preparation protocols, to optical blur, to luminosity levels, etc.?

**Qualitative performance**

At this stage in our studies, the performance of our method was subject to a qualitative assessment by comparing – in a fairly global manner – the results produced with those of an expert appraisal. Of course, it is always possible to compare "processing" time, number of cells detected or files identified (Table 2). For example, Table 2 shows that the automatic method detects a consistent number of cells about 80–100-fold more rapidly than the expert, and this regardless of preparation protocol, ana-
tomical configuration or magnification used, and that, overall, our method works well (88% of files detected on average) indicating slight over-estimation by the automated method.

The low shift of 0.0127 confirms the middle error of 5% on right areas. The automated method appeared to over-estimate the values (or the expert appraisal under-estimated the values). The only certainty is that the automated method is repeatable.

The major limitation to the automatic file identification method stems from image content, namely the photometric characteristics of the image and the structural biological layouts.

**Geometric precision**

We have seen that files are created only from geometric and topological rules applied to the basin adjacency graph. It is therefore important to ensure the robustness of the basin detection, i.e. the method used must be precise in cell cropping and insensitive to image blurring.

Figure 7 compares about 60 normalized areas obtained by the manual and automated methods. The areas were divided by the median area to refine the regression. The coefficient of determination tends toward 1, showing that the areas are closely correlated. The slope of the regression is weakly greater than 1, indicating slight over-estimation by the automated method.

The study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to the study could feasibly be extended to geometric shift, namely to

### Table 2. Outline of some significant results: size of processed images, number of extracted cells, CPU time on a PC with an Intel Xeon at 2.3 GHz, and total Quality Index defined as the ratio of number of files automatically identified and reconstructed over total number of lines; for the purposes of the expert analysis, the cells were manually cropped using ImageJ software

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (pixels)</th>
<th>Expert Cell numbers</th>
<th>Expert Times (s)</th>
<th>Plugin Cell numbers</th>
<th>Plugin Times (s)</th>
<th>% of files detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mahogany</td>
<td>1024 × 768</td>
<td>1302</td>
<td>2520</td>
<td>1359</td>
<td>14-3</td>
<td>86</td>
</tr>
<tr>
<td>Fir</td>
<td>1360 × 1024</td>
<td>787</td>
<td>1300</td>
<td>800</td>
<td>12-4</td>
<td>88</td>
</tr>
<tr>
<td>Black Pine 1</td>
<td>1600 × 1200</td>
<td>1794</td>
<td>2750</td>
<td>1873</td>
<td>23-2</td>
<td>92</td>
</tr>
<tr>
<td>Caribbean Pine</td>
<td>1360 × 1024</td>
<td>819</td>
<td>1500</td>
<td>828</td>
<td>11-5</td>
<td>91</td>
</tr>
<tr>
<td>Black Pine 2</td>
<td>1600 × 1200</td>
<td>1411</td>
<td>2450</td>
<td>1458</td>
<td>16-1</td>
<td>83</td>
</tr>
</tbody>
</table>

**Figure 7.** Study of basin surfaces on mahogany. The x-axis shows the normalized areas obtained by the fully manual method. The y-axis shows those obtained by the automated method. The coefficient of determination is close to 1, showing a good fit.

**Figure 8.** Comparison of watershed results on a Pinus cross-section. (A) Crest lines from a sharp image. (B) Crest lines from a blurred image. Basins outlines are virtually identical between the sharp and blurred images. The sharp image has a small additional basin corresponding to wall detachment caused by cutting. This basin was deleted in the supernumerary removal phase.
Reliability and blur

Image blurring is a source of potentially major bias when evaluating lumen areas and wall widths. As indicated before, the cell segmentation is based on a two-means clustering. This classification is sensitive to the intensity dynamics of the cellular points and is therefore modified if intensity dynamics are reduced by optical blur (Fig. 10). We observed that the surface area of the lumen for a given cell was reduced when blur increased, i.e. when intensity dynamics decreased.

The local blur indicator introduced earlier is used to estimate the error on the surface area measurements of cell constituents. The figure on the left shows the consistent behaviour of the blur estimator that was calculated from cells considered in images that were digitally disrupted by an incremental Gaussian blur.

Therefore, the local blur indicator can be used to filter the results and retain only the sharpest cells. It is used, in particular, to create mosaic images (Brunel et al., 2013), for retention of the most reliable basins. It could also possibly provide a correction factor for measures made on areas that are more or less subjected to optical blur.

Reliability of files

The overall reliability coefficient can be used for file characterization of files, in terms of both significant expression and geometric behaviour, i.e. a ‘good’ file is a sufficiently long line in which cell size varies gradually. This coefficient acts directly to filter the files (Fig. 11). Files with a good coefficient are shown in green while those with a poor coefficient are in red. The red–orange hues indicate configurations with cuts or sharp discontinuities. This estimator is well suited to softwoods with very organized and continuous cell arrangements. In angiosperms, it may be used for a rapid visualization of vessels and adjacent files. These files are useful for analysing the endogenous geometric constraints caused by vessels on files.

This indicator is ultimately a very integrated aspect of the files identified. It is more comprehensive than the performance indices mentioned in the first part of the results, and is also more complex to explain as it goes beyond file identification reliability to actually reflect file quality. For example, some alignments, shown in yellow in the image below, appear to be ‘visually correct’ while their coefficient is rather low. They are penalized by excessively elevated surface area variations. For example, the third orange file from the left looks to be good, but small basins can be seen within the file arising from over-segmentation caused by wall detachment. Files can be filtered for statistical analysis thanks to their cell coefficients.
General conclusions

The method described herein works correctly for images that show both high contrast between walls and lumens (without local color inversion) and a visible cellular organization. Under these conditions, all the obvious lines determined by qualified experts are correctly identified with significant time savings (a typical manually expert identification required 10 h on our samples). Moreover, the certainty index allows for selective exploitation of the results for statistical studies.

CONCLUSIONS

The automated analysis of anatomical wood sections will contribute greatly to understanding secondary plant growth and development. Herein, we propose a fully automatic method for the recognition of operational cell files in mass processing.

We have also established reliability coefficients to characterize the files, their components and their measurement. These reliability coefficients are mainly used to filter results for statistical studies conducted by botanists and aim to draw their attention to specific biological configurations.

We have shown that the method is equally applicable to digital images of stained histological wood sections and images of sanded ligneous planes.

File identification is based on detection of cell alignments showing gradual geometric variations. The cells in the image are automatically cropped by watershed lines calculated from a set of judiciously filtered local minima, thus reducing the risk of exaggerated segmentation. Cell files are defined by the longest straight vertices that are ‘geometrically similar’ to the adjacency graph created from the watershed catchment basins. These alignments are constructed one after the other by successive aggregations of vertices in progressively enlarged neighbourhoods. Unlike traditional approaches, cell typing here is not based on the creation of reinforcement training, but on a decision tree established with wood anatomists.

In addition to file detection, the cells are characterized by a set of values (wall thickness, lumen surface, cell circularity, etc.) that are necessary for the statistical analysis of files and their variability. To ensure that the measures are relevant, reliability indices have been created, based on existing methods and biological reality (file length and continuity). These indices are used to select the most important files or highlight particular biological disturbances or configurations.

The percentage of unusable files is offset by the quantity of images that can be processed. Mass processing allows us to study spinal files up to the cambium over several wood growth units.

These initial results are already very interesting and confirm that the method has great potential and could be suitable for statistical studies and the detection of invariants or trends in large observation samples. Comparative studies undertaken with images annotated by experts have shown that the proposed method is time-efficient and generally insensitive to the optical blur potentially present in micrographs. A blur indicator has been introduced to study the stability of the method and in the short term could be used to correct the measurements obtained. Finally, a reliability indicator to characterize files, in terms of both significant expression and geometric behaviour, is given either to draw the attention of experts to specific biological configurations or to remove potentially erroneous data from statistical studies.

ACKNOWLEDGEMENTS

We thank the NUMEV Laboratory of Excellence (University of Montpellier 2) and the SIBAGHE Graduate School of the University of Montpellier 2 for their support. We extend special thanks to Christine Heinz and Christophe Godin for their availability and the relevant and constructive observations they provided during different interviews and meetings. Additionally, a big thank you to François Pailler and Merlin Ramel for their patience and help in preparing the different samples and optimizing technical protocols. Special thanks also to Michæl Guéroult who took the time to provide assistance and share his skills and knowledge in the use of optical microscopy.

LITERATURE CITED


