Slide 3:

We work on the development, the plasticity and the production of trees according to endogeneous and exogeneous constraints. These studies are lead at different observation scales, from trees populations to cells organization.

In the context, the light microscopic images are one of supports used to observe and understand the wood behavior.

Slide 4:

At the macroscopic scale, the wood is globally composed of three regions: the pith, the bark and the conductance area. The central part contains the rings, continuous indicators of the plant growth.

At the microscopic scale, this area is composed of different kinds of cells, with particular arrangements. There are these aspects that botanists study.

Slide 5:

In the conductance area, the cells are only produced from initial cells located between the bark and the pith. The new cells push away their sisters. This mechanism explains the linear arrangement of the cells, either towards the pith, either towards the bark.

Slide 6:

The cells are spindle-shaped; they are oriented according to the longitudinal or transversal axes of the plant. The cells are specialized and more or less easily identifiable. For example,

The tracheids are oriented in the longitudinal axis; they assume supporting and conductive functions

In opposite, the rays are oriented in the transversal axis. They assume a storage function

The cell files correspond to tracheid alignments. They are studied on transversal sections of wood.

Slide 7:

The scientific questions concern the regularity and the arrangements of cellular types. Here, cells identification and counting are basic works. A statistical approach on a sufficient number of samples is crucial for such an issue.

Slide 8:

The main difficulty is the strong variability of the numerical images to process. This variability depends of three factors: the study object, the preparation and the acquisition

Even if the conifers present globally the same visual aspects, there are several sensitive differences. For example, the punctuation frequency, the size or the shape of cells.

Several preparations are used in regard to the habitudes or to the available means. It is actually impossible to uniform the preparation protocol.

To summarize, the wood can be pumiced or cut. In the first case, the images are enough accurate but not too contrasted. A local blur appears in the non-planar zones. In the second case, we use thin slices which are cleaned up with alcohol bath and are contrasted by coloration. Different colorants can be used, but the principle is the same. The slice is immerged during a given time in a staining bath and the colorant is more or less fixed by the structures. The contrast is not uniform.

The magnification used for the observation of the structures depends of the couple lens / captor, in other terms of the acquisition materials.

Slide 9:

An additional difficulty is the behavior of the structures. In the conifers, the difference between the winter wood and the summer wood is very marked.

The cell is composed of two parts: the lumen and the wall. The lumen is the light central area, the wall the dark external area.

In the winter wood, the cells present thick walls and small lumens. In summer wood, the walls are thinner and the lumens bigger. In extreme cases, the lumen disappears in the winter wood, or the lumen becomes darkest than the wall.

The visual properties of the structures change.

Slide 10:

The files are defined as the cell alignments. They are here drawn by experts.

Notice the presence of partial files and intrusions. The intrusions are false files, composed of cells coming from lower or higher planes.

Slide 11:

The processing is defined by the succession of three steps.

The first step is the segmentation: it allows extracting all the cells of the image.

The second step is the classification of the extracted structures in order to identify the tracheids cells

The third and last step is the files recognition which allows identifying consistent alignments of tracheids.

The process is run from the color image. It produces a files image and some numerical results. The files are given with pertinence coefficient evaluated in regard to their length and to the cells classification.

The results gives geometrical parameters for each cell of each file.

Slide 12:

To segment the image, we apply successively several filters to suppress noise, and we extract then the cells limit by watershed. More precisely, we use first a mean shift filter to reduce the gaussian noise and make uniform colors. Secondly, we apply a median filter to suppress the remaining punctual noise. Thirdly, the image is smoothed by a Gaussian blur, classically applied before the watershed. The watershed defines catchment basins whose the limits are similar to the cell borders. On anatomical point, there is a space between the adjacent cells. The limits given by the watershed represent this space.

Slide 13:

Mean shift filtering is a data clustering algorithm commonly used in image processing. For each pixel of a color image, the set of neighboring pixels is determined within a spatial radius and a defined color distance. For this set of neighbor pixels, the new spatial center and the new color mean value are calculated. These calculated mean values will serve as the new center for the next iteration. The described procedure will be iterated until the spatial and the color mean stops changing. At the end of the iteration, the final mean color will be assigned to the starting position of that iteration.

This filter is used for edge-preserving smoothing. It reduces the Gaussian noise without destroy the frontiers.

Slide 14:

Any color image is considered as a topographic surface: for each point, the color is used to define an altitude. When we flood this surface from its minima and prevent the merging of the waters coming from different sources, we partition the image into two different sets: the catchment basins and the watershed lines.

For our image, the watershed lines represent the boundaries between the different cells.

Slide 15:

The watershed produced often an over-segmentation. This is traduced by the presence of edges in the lumen of the cells. Here we detect the edges which cross the lumen by study of their profile of densities. The cell boundaries stay in the wall: the densities of the points of these edges have to stay low.

The edges are defined from the watershed subdivision by the study of the adjacent basins. A point incident to at least three basins is a vertex. An edge is composed of points joining to vertices.

Slide 16:

Once the segmentation performed, we define a dual subdivision by building a graph where the nodes represent the basins and the edges the adjacency relationship. The nodes are finally the cells of the image. They are typed to indicate if the represented cell is a tracheid or not.

Slide 17:

The typing is get from a classification and regression tree. It is supervised method which proposes a decision algorithm from a training set. We constituted the training set of representative images from ten conifers species. For each conifer, we labeled about one hundred cells, tracheids and non tracheids. Geometry, topology and density of each cell are described by different parameters.

The classification method leads to a decision tree which gives the discriminative parameters and the best thresholds. The cross-validation has bee realized on a second training set. The cross-validation test is superior to 95%.

Slide 18:

The cells alignments are defined from the edges alignments in the adjacency graph. The principle consists to follow the graph edges in regard to the main direction.

The main direction is defined from the study of the distribution of angles between the edges and the horizontal. In conifers, the cells are arranged in staggered-row: the edges perpendicular to the main direction are distributed in two directions. There is so only one direction which is more represented than the other.

Slide 19:

The files recognition consist to follow the main direction: from the last edge of the file, we look for the incident edge which will give the best alignment.

A threshold is used to avoid the file takes a bad direction. The geometry of the files depends on the topology of the adjacency graph. When the process leads to a bad segmentation, the graph is incomplete. Some edges between cells visually adjacency can be absent.

So we prefer built short aligned files instead long curved files.

These short alignments are joined from their geometrical distance. This distance will be calculated from orthogonal regression lines. The joining is validated by a reciprocity condition: *the best candidate of a given short alignment also has this last as best candidate.*

Slide 20:

The orthogonal regression gives the least squares regression equation where the direction is fixed. This method allows fitting an affine straight line onto the coordinates of the nodes of a given section.

The director vector of the regression line is fixed by the main direction. All the regression lines are so parallel. The issue consists to evaluate the best value of the parameter c in the affine expression of the regression line. We have to minimize this function, a and b being given.

Slide 21:

The reciprocity condition limits the inconsistent joins: the best candidate of a given section also has to have this last as best candidate. When this condition is not assumed, the join is not automatically made.

In the case, the join stays possible, but only onto the responsibility of the expert.

Slide 22:

The visual representation of the identified cell files allows controlling the relevance of the results. A good part of files are correctly detected, and can be used in statistical analysis.

Cells files can be qualified by computing some feature like the ratio of the file length to the length of the longest file or the ratio of the number of tracheids to the total cells number of the file.

Slide 23:

In order to access the quality of the recognition, we compare the result of the automatic method with some reference files manually defined. The results are globally

Slide 24:

Some limits have been identified: segmentation default, recognition default and join default.

The main issue stays the segmentation of the image. A bad segmentation leads to a bad recognition. The recognition problems are linked to the behavior of anatomical structures, by example when the lumen becomes darkest than the wall. The presence of intrusions leads so to bad files.

Slide 25:

As perspectives, two points will be studied.

First point: the place of the typing of cells. In the recognition step, the usage of the typing has been limited to the initialization of the process. We do not use the labeling cells in the recognition process because it was too restrictive.

Second point: the extension to the hardwood species which present different anatomical structures. The cell topology hypothesis has to be adapted to new features: for example, vessels are big cells which distort the adjacent cells.

Slide 26:

In conclusion, we have a beginning of automated method with encouraging results.

But several points have to be taken up again: for example, the mean shift is too long and not easy to set. Considering the strong variability of the images, the using of a supervised method of classification could be the weak point, the learning cost seemed high.