

EVALUATION OF PLANT HISTOLOGY BY AUTOMATIC CLUSTERING BASED ON INDIVIDUAL CELL MORPHOLOGICAL FEATURES

FLORENCE GUILLEMIN, MARIE-FRANÇOISE DEVAUX AND FABIENNE GUILLON

Unité de Recherche sur les Polysaccharides, leurs Organisations et leurs Interactions, INRA, BP 71627, 44316 Nantes cedex 01, France

e-mail: devaux@nantes.inra.fr, guillon@nantes.inra.fr

(Accepted January 20, 2004)

ABSTRACT

A procedure has been developed for the automatic clustering of plant cells observed by confocal microscopy. The contribution of cell morphological features to reveal histological regions has been investigated. Several adjacent images were acquired to visualise a representative region of the sample and a mosaic image was built. The cell size and shape and the cell wall thickness were quantified. The extracted features were used to automatically classify the cells into morphological groups. The technique made it possible to split the cell population into 8 groups mainly corresponding to histological regions of beet root.

Keywords: cell clustering, cell morphology, cell wall thickness, confocal microscopy, mosaic image.

INTRODUCTION

Different cell types are found in plants, each one assuming a particular function for plant physiology: carbohydrate or lipid storage, sap conduction, plant support. Cell types are generally observed from sections at the microscopic scale and identified from the individual cell morphology, namely size and shape, cell wall thickness, as well as from cell arrangement (Esau, 1977). In addition to such histological description, the microscopic structure is largely paid some attention to study the relationships between the spatial organisation and macroscopic properties (Jackman and Stanley, 1995; Travis *et al.*, 1996; De Smedt *et al.*, 1998; Guines *et al.*, 2003). Such an approach has been developed for wood characterisation in relation to drying properties (Perré, 1997). In the food domain, there is also an increasing demand to determine the contribution of the physical structure to the end-use texture or sensory properties (Gao *et al.*, 1999; Zghal *et al.*, 2002). Plant histology is generally observed by microscopic techniques. Plant cells, especially storage cells, can be very large, and only a few cells are visible on a single image (Gray *et al.*, 1999; Konstankiewicz *et al.*, 2002). The quantification of histology is not straightforward and few techniques have been reported (Travis *et al.*, 1996; Guines *et al.*, 2003).

The objective of the present work was to identify histological regions in plant material by the automatic clustering of cells from individual morphological

features measured in microscopic images. In a first step, we focused on 2D quantification. Confocal microscopy was chosen in order to minimise sample preparation and to obtain thin optical sections from thick specimens. Several adjacent images were acquired and a mosaic image was reconstructed to observe a representative area of the sample. Cell morphology was quantified by considering both cell and cell walls. A clustering procedure has been developed to automatically classify cells into groups on the basis of their morphological features. Labelling the cells according to their groups made it possible to localize the groups in the mosaic image. The relevance of the clustering was assessed by expert interpretation of the groups in regard with the known histology of the sample.

Sugar beet root has been studied as an example of plant of industrial interest. Sugar beet root is the storage organ of sugar within which several kinds of tissues are found : storage parenchyma where sugar is stored, xylem and phloem that form the food conducting tissue, cambium where cellular divisions occur.

MATERIAL AND METHODS

SAMPLES

INRA Estrées-Mons (80, France) provided roots from the variety Roberta. Hand-cut sections were collected in the medium region of the fresh root and

halfway between the central core and the external epidermis. Hand cut sectioning was carried out in order to preserve cell morphology, to minimise sample preparation and to obtain sufficiently large samples to observe the different cell types. Samples were about 1 cm large and 300 μm thick and were stained using acridin orange (0.02% in 0.1M phosphate buffer, pH 7) in order to make cell walls fluorescent.

Image acquisition

Images were acquired using a confocal laser scanning microscope (Zeiss, LSM 410). Though the present work focused on 2D information, 3D images were acquired for further 3D image processing. The excitation wavelength was 488 nm and the light emitted over 515 nm was collected using a long pass filter. The $\times 40$ lens was used to observe together small and large cells. Under these conditions, the 512×512 pixel image was 319.4 μm large and the axial resolution was 0.8 μm . As such an area was not sufficient to observe the different cell types and their relative arrangement on the same view, several adjacent images were acquired. The principle of image acquisition is given in Fig. 1. Successive lateral scans were achieved in the x-y directions so that adjacent images partly overlapped (Fig. 1a). From 25 to 30 2D images were scanned by step of 1 μm for each 3D sequence (Fig. 1b). The number of 2D images in one 3D sequence depended on the fluorescence decreasing according to the depth of the optical plane. The shape of the sample in contact with the slide was not perfectly planar mainly due to hand cutting (Fig. 1b). In order to avoid acquiring completely black images, the starting of the axial z scan was adjusted for each 3D sequence. In the following, individual 2D images were called optical images and the stacks of 2D images were called 3D sequences. The optical images were coded using 256 grey level values.

Mosaic image

A 2D mosaic image was extracted from the 3D sequences (Fig. 1b). As the microscope was not equipped with a motorised stage, the shift between 2 sequences had to be calculated in the x (or y) and z direction. Axial z shift was taken into account because of the different starting points of z scan between the 3D sequences (Fig. 1b).

The mosaicking procedure was developed within the matlab environment (<http://www.mathworks.com>) and is described in Fig. 2. Lateral and axial shifts were sequentially searched for. The maximum correlation coefficient between the grey levels of the overlapping regions was chosen as quality criterion. The algorithm progressed following the different steps (Fig. 2):

- The operator chooses a starting optical image i in the first 3D sequence.
- The homologous i^{th} optical image of the next 3D sequence is considered first.
- Lateral shifts from 1 to n are tested and the one for which the correlation coefficient between the overlapping regions is maximum is selected. The lateral shift research stops at step n after the correlation coefficient has decreased five times consecutively. The first estimation of the lateral shift is $l = n - 5$.
- The lateral shift is fixed to l .
- The axial shift is searched for by testing the evolution of the correlation coefficient for the optical images over and under the current optical image. Again, the research stops when a maximum is found.
- The axial shift is fixed to j , optical image number for which the correlation coefficient is maximum.
- The lateral shift is adjusted by searching a new maximum before and after the preceding value l for images i of the first sequence and j of the second one.
- The procedure is repeated from step 4 to step 7 until neither the lateral shift nor the axial shift are modified.
- The two optical images are joined by giving the overlapping region the grey level values of the first image.
- If another image is to be combined, the shift research goes on from step 3 between the last combined image and the i^{th} optical image of the new 3D sequence.
- The final result is a 2D mosaic image.

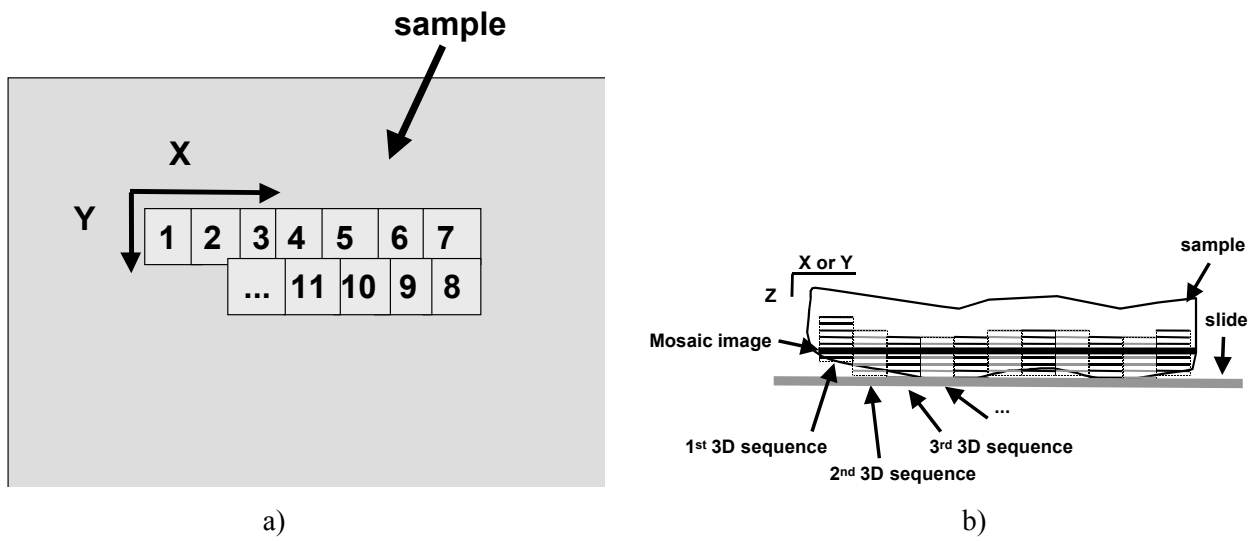


Fig. 1. Acquisition of overlapping images in the x-y plane a) and x-z plane b).

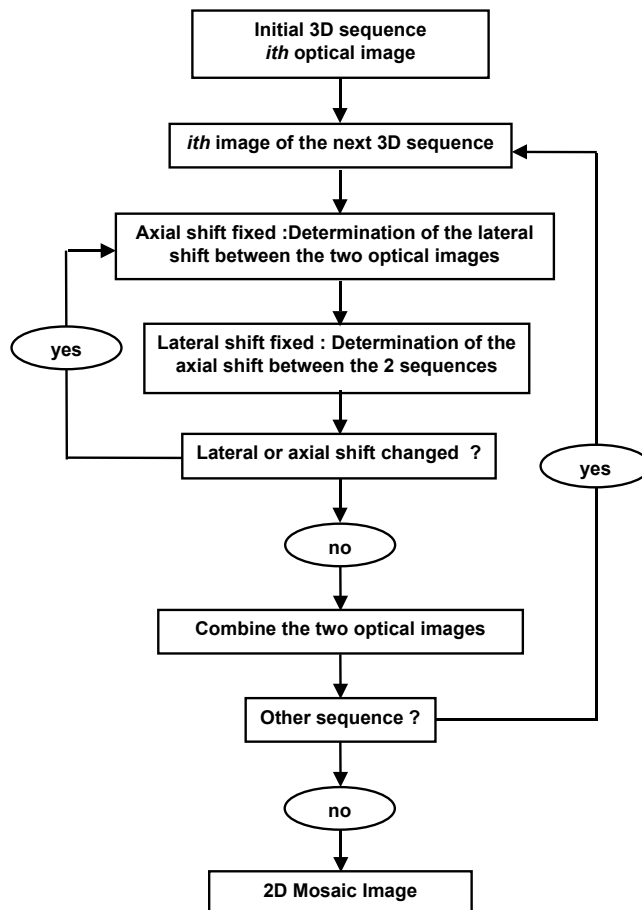


Fig. 2. Main steps of the construction of the mosaic image.

Image segmentation

A standardisation of the grey levels was achieved in order to apply a single threshold. The grey level variations over the image were considered as a multiplicative shading effect (Tomažević *et al.*, 2002). This shading was estimated by dilating the image with a squared structuring element of size 11×11 pixels and by filling the remaining holes, holes being defined as areas of dark pixels surrounded by lighter pixels (Soille, 1999). Dividing the original image by the calculated «shading» image and adjusting the grey levels between 0 and 255 achieved the normalisation. A single threshold of 50 was visually chosen leading to the segmentation of the cells. Cells with an area lower than 100 pixels were eliminated as well as those intersecting the image border. The cell label image and the cell wall binary image were computed.

MEASURES

Individual features were extracted for each segmented cells. The matlab morphological parameters : *Area*, *Length* and *Width* along the major axes, *Area of the Convex Hull*, were measured. From these parameters, an *elongation* factor was computed as the *Width* to *Length* ratio, and a *solidity* index estimated as the ratio of the *Area* to the *Convex Hull Area*. This

index makes it possible to reveal concavities in an object shape.

A procedure has been developed to estimate the *Cell Wall Thickness* (Fig. 3). For this purpose, both the cell wall and cell label images were considered. Each cell was analysed separately in order to extract its individual cell wall. The cell was first dilated using a disk structuring element of 31 pixel diameter within its extended bounding box. The surrounding cell walls were extracted as the intersection of the dilated cell and the cell wall binary image. The skeleton of the local cell wall was computed and spur pixels were removed. The distance function was then applied to the local cell wall image using the Euclidean distance metric. The intersection between the distance image and the skeleton made it possible to extract the distance values of the skeleton. The *Cell Wall Thickness* was computed as twice the mean of the distance values observed along the skeleton. In the example given in Fig. 3, the value was 8.2 pixels corresponding to 5.1 μm .

The result of feature extraction was a data table containing the 4 variables : *Area* as size parameter, *Elongation* and *Solidity* as shape parameters and *Cell Wall Thickness* measured for each cell.

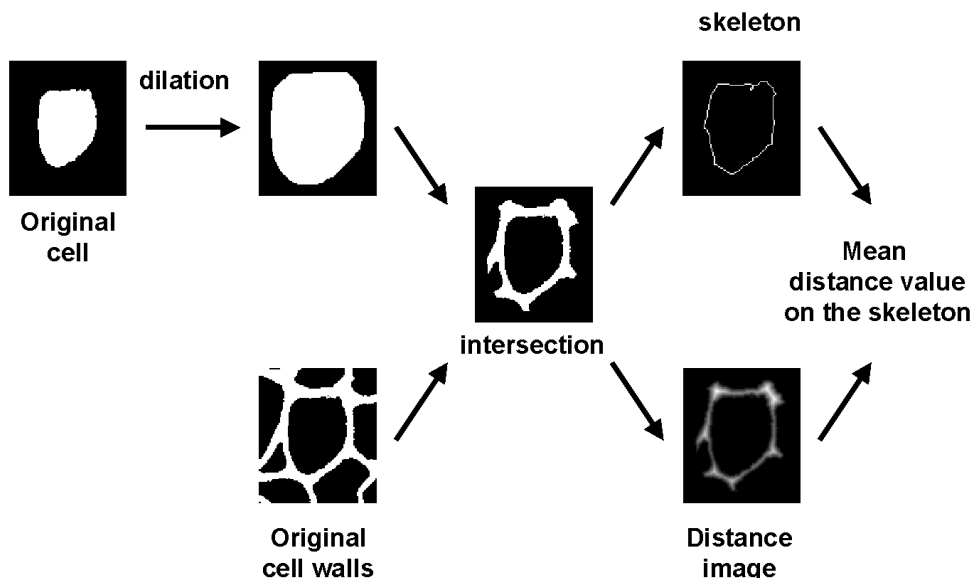


Fig. 3. *Extraction of cell wall thickness.*

Cluster analysis

An automatic clustering procedure, based on the 4 morphological features, has been developed to reveal cell groups without any prior information on searched groups.

The procedure combined a hierarchical clustering and k-means clustering steps (Filtzmoser *et al.*, 1999). Divisive hierarchical clustering aims at partitioning the initial population into an increasing number of groups while the objective of the k-means method is to find a partition into k groups, k being fixed (Lebart *et al.*, 1995). In the present work, the techniques were implemented as follow (Fig. 4). At the beginning step, all cells belonged to the same group. The population was split into two groups using the k-means technique with $k = 2$. The next hierarchical step consisted in choosing one group to be further divided and in initialising the two new groups. The k-means method was then applied with $k = 3$ by considering the two new groups and the remaining one to calculate the 3 final groups (Fig. 4). The procedure was repeated to find a 4 group partition, 5 group ... The interest of the k-means steps applied by considering all the current groups, and not only the one in division, is to avoid freezing the frontier between groups during the hierarchical divisive steps.

Several criteria had to be defined to implement the procedure: choice of the group to be divided and for the k-means technique: initialisation of the new

groups and distance criterion to assign cells to groups. Principal Component Analyses were performed to determine the group to be divided and to find the initial group centres. Principal Component Analysis is a data analysis technique that reveals the structure within observations and variables (Jolliffe, 1986). It was applied on the data table constituted by the cell morphological parameters. At the first step, all the cells were considered and for the next steps, the groups were analysed separately. A large percentage of variance calculated for the first principal component was considered as representative of heterogeneous observations. At a given step k , the percentage of variance of the first principal component was therefore assessed for each of the k groups and the group with maximum value was considered for division (Fig. 4). The first Principal Component was also used to initialise the new groups as being the cells for which positive or negative scores were respectively obtained (Fig. 4). In the k-means method, the minimum Euclidian distance was chosen to affect cells to groups and the procedure stopped when all the groups were stable.

The splitting procedure was repeated until the number of groups was found satisfactory. At each step, a label image was assessed on which the cells belonging to the same group are given the same colour. The operator chose the final number of groups by visual examination of these images. The average values of the 4 morphological features were assessed for each group.

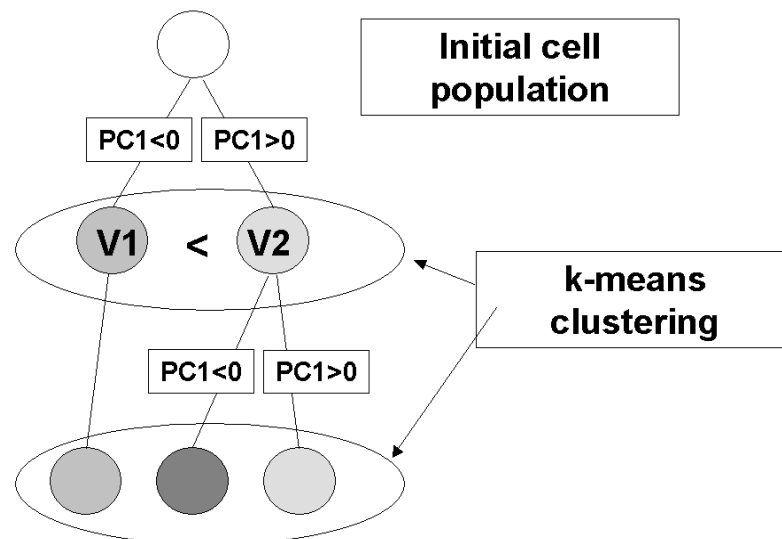


Fig. 4. Clustering principle. PC1: principal component 1. V1 and V2: percentage of variance of principal components 1 for groups 1 and 2. The group to be divided is the one with the maximum percentage of variance. Initialisation of the division of the groups is based on the positive or negative values of the first principal component scores. At each step, the groups are adjusted by k-means clustering.

RESULT

MOSAIC IMAGE

Thirteen 3D sequences were acquired to visualise a complete vascular bundle region of the beet root. Two adjacent rows of 3D sequences were necessary. The images were combined to extract a 2D mosaic image. An example of image combination is given in Fig. 5 for two 3D sequences. Optical image 10 in the first 3D sequence was chosen as starting image (Fig. 5a). The

homologous image in the adjacent 3D sequence was first considered to estimate the lateral shift (Fig. 5b). The maximum correlation coefficient was 0.53 for a lateral shift of 17 pixels (Fig. 5c, step 1). This 17-pixel shift being fixed, optical image 14 lead to a maximum of 0.92 for the correlation coefficient (Fig. 5d, step 2). Step 3 (Fig. 5e) showed that the best lateral shift value was again 17 when considering optical planes 10 of the first 3D sequence and 14 of the second one. The final values were therefore 17 pixels for the lateral shift and +4 for the axial shift.

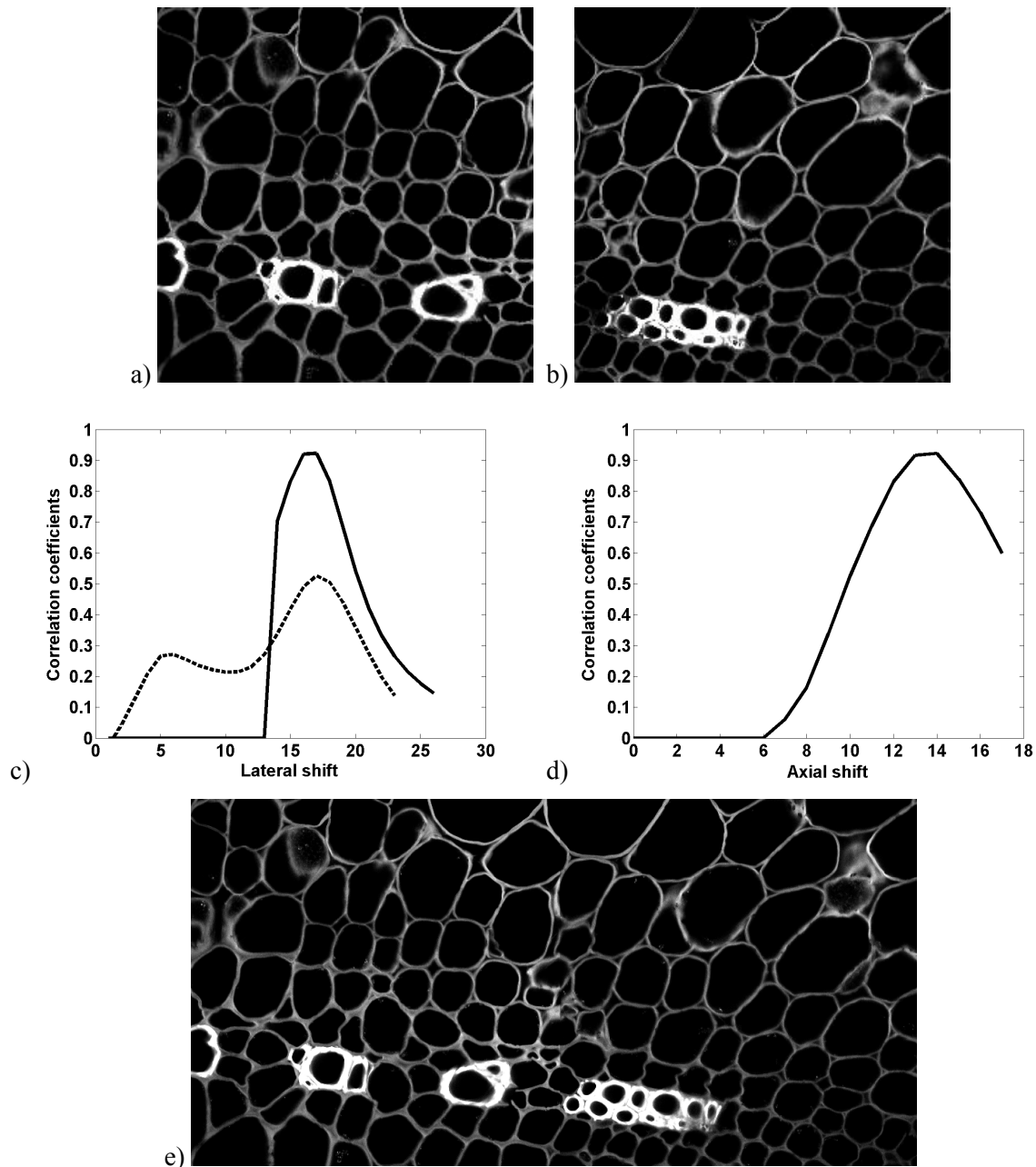


Fig. 5. Example of mosaic image construction. (a) and (b) Initial images : 10th optical plane of the 3D sequences. (c) Correlation coefficients according to the lateral shift: steps 1 (dotted line) and 3 (solid line). (d) Correlation coefficients according to the axial shift: step 2. (e) Final mosaic image obtained by combining optical image 10 of the first 3D sequence and 14 of the second one for a lateral shift value of 17.

A first 2D mosaic image was obtained from the first row of 3D sequences acquired. The last image selected was chosen to find the first starting optical image of the second row and a second mosaic image was constructed. The final mosaic image was obtained by searching the lateral shift between the two mosaic images. The resulting image is shown in Fig. 6a. The correlation coefficients ranged between 0.89 and 0.97 for 11 of the 12 combinations, the last one being 0.49 between images 11 and 12. For these 2 images, no overlapping region could be found as a few pixel lines were missing on acquisition. As no drastic deformations of the cells was observed, the images were nevertheless combined in order to visualise a larger region of the section. The correlation coefficient between the two row mosaic images was 0.78.

Some grey level variations can be seen over all

the area of the final mosaic image (Fig. 6a). They were caused by the fluorescence attenuation with optical depth in confocal microscopy. The optical planes in the 3D sequences ranged from 12 to 19 for the 6 top images of the mosaic image versus 0 to 9 for the 5 bottom ones. A standardisation of the grey levels was achieved in order to correct these grey level heterogeneities (Fig. 6b). The cell walls were extracted by thresholding the normalised image (Fig. 6c) and the cells were labelled (Fig. 6d). The resulting image contained 795 objects mainly composed of individual cells sections. Some cell walls were rough, as some cells were sectioned on their top or bottom resulting in viewing cell walls partly in surface. A few cells were still connected. The 4 morphological features, *Area*, *Elongation*, *Solidity* and *Cell Wall Thickness*, were computed for each of the 795 cells.

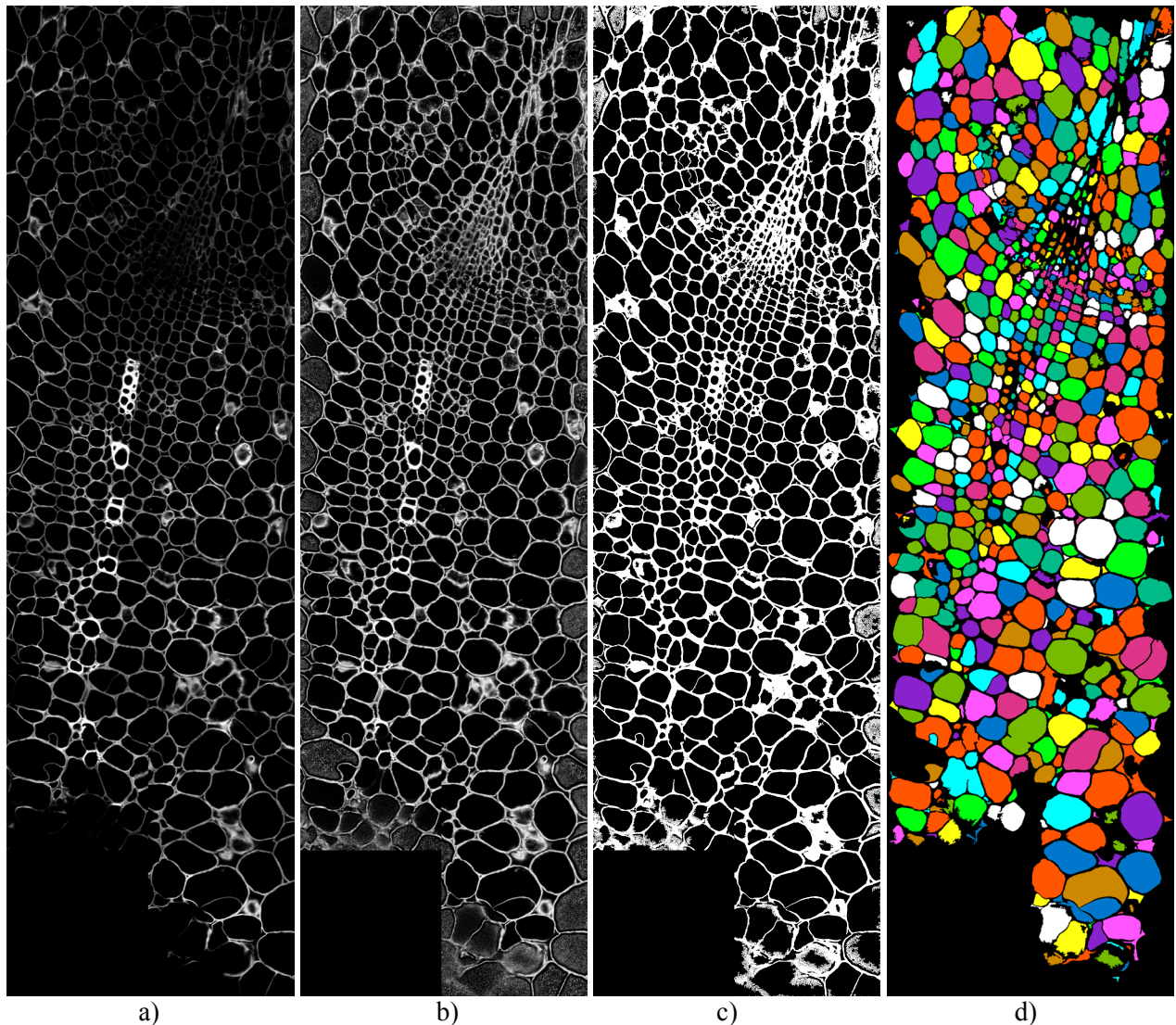


Fig. 6. Mosaic image. a) Original. Individual images were acquired along two rows from 1 (lower left) to 6 (upper left) and 7 (upper right) to 13 (lower right). b) Image corrected for grey level variations. c) Binary cell wall image. d) Label image of segmented cells.

Cluster analysis

Images in Fig. 6 clearly reveal that some regions can be visually identified. For example, small rectangular cells that were arranged along a circle arc, corresponded to the cambium area while the small and larger elongated cells forming a triangle over the cambium area corresponded to the phloem area. Xylem vessel cells were aligned and were small and circular with thick cell walls. Both individual cell features and cell arrangement contribute to identify homogeneous areas in the image. In the present paper, only the contribution of individual morphological features to identify cell groups was tested. An automatic clustering procedure has been developed to reveal groups without any prior information.

The image obtained for 8 groups was retained and is shown in Fig. 7. Some histological regions were clearly revealed from the cell clustering. Yellow cells mainly corresponded to the cambium region while the cyan ones corresponded to storage parenchyma. The interpretation of each cell group according to histology and the average values calculated for each parameter are reported in Table 1. The average values show which features were important in group determination. The storage parenchyma was characterised by its large cells. Blue cells that were medium sized, convex, circular and with thin cell walls, corresponded to vascular parenchyma. Grey cells constituted another class of parenchyma cells that were smaller and exhibited thicker cell walls. They were mainly found around the xylem vessels and also within the two other parenchyma groups. A few phloem cells were classified in this group. Within the green group were found xylem vessels, which had thick cell walls. For the other green cells, the walls were found thick because they were partly viewed in surface. Pink cells were characterised by an elongated shape and thick cell walls. They were localised in several regions of the section. The phloem area was mainly constituted by this kind of cells. They also exhibited a lower *Solidity* than the preceding groups, due to rough edges caused by a rather poor segmentation. Yellow cells were mainly found in the cambium region that was characterised by small sized cells with thin cell walls. Cambium is the region of cellular division,

which indeed result in thin cell walls and also in a concave shape. The two remaining groups were composed of small sized cells, with elongated and highly concave shapes. They were distinguished from the other groups by their cell wall thickness. The red cells could be found in the cambium region or when cells were in division. Some intercellular spaces were also classified in this group. The white group was mainly composed by intercellular spaces and by cells where walls were viewed in surface.

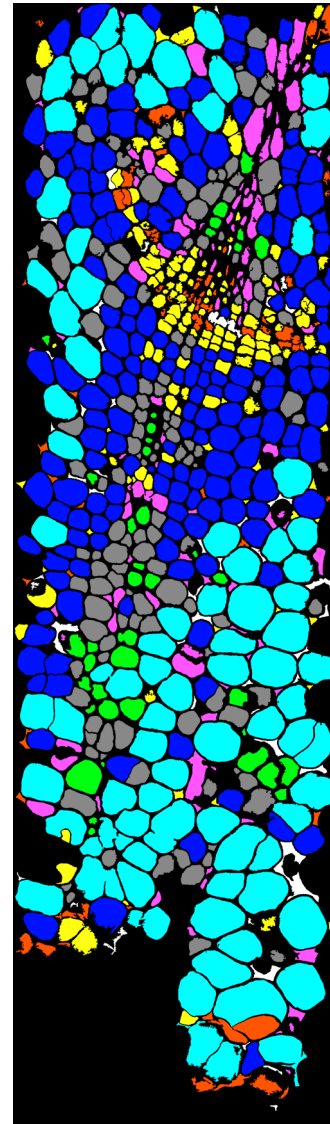


Fig. 7. *Label image for eight groups.*

Table 1. Assignment of groups to cell types and average values of Area, Elongation, Solidity and Cell Wall Thickness. Colour refers to the cell group colours in the label image Fig 7. Values were rounded to 2 digits.

Group colour	Assignment to cellular type	Area (μm^2)	Elongation	Solidity	Cell Wall Thickness (μm)
Cyan	Storage parenchyma	3600	0.73	0.94	4.5
Blue	Vascular parenchyma	1300	0.75	0.94	4.0
Grey	Touching vessel parenchyma	670	0.72	0.91	5.0
Green	Xylem vessels	600	0.75	0.90	6.5
Pink	Phloem	340	0.43	0.80	4.9
Yellow	Cambium	300	0.71	0.83	3.6
Red	Cambium, intercellular spaces	280	0.48	0.70	2.9
White	Intercellular spaces	230	0.34	0.55	4.3

DISCUSSION - CONCLUSION

The objective of the present work was to test the contribution made by individual cell features to identify histological regions in plant material. Cell morphology was defined both by the cell size and shape and by the cell wall thickness. The task required considering several steps from image acquisition to cell clustering.

Plant cells are often large requiring the acquisition of adjacent images to observe a representative area in the samples. Grape and potato cells can be as large as 150-200 μm (Gray *et al.*, 1999, Konstankiewicz *et al.*, 2002,). In tomato fruit, cells can even be larger than 500 μm long (Barret *et al.*, 1998). In the present work, 13 images were necessary to visualise one vascular bundle of beet root. A smaller magnification would have allowed viewing a larger area but would have also lead to a poor resolution concerning the cell wall thickness. Typically, cell walls were 5 to 10 pixels large using the $\times 40$ lens. Another advantage of mosaic image was to analyse cells larger than what could have been possible using single images, as large cells often intersect with image borders. The mosaic image has to be reconstructed. Few softwares and techniques have been reported concerning the mosaicking of confocal images (Karen *et al.*, 2003). In the present work, the procedure has been developed using the correlation coefficient as quality criterion to reveal the shift positions (Fig. 3). Fig. 3c shows that some false maximum could be found, depending on the local cell shape in the overlap region. As neither geometrical deformation nor rotation occurred during acquisition, only lateral and axial shifts were searched for. The method developed was

straightforward and only allowed shifts in the x-z or y-z direction. Alternative techniques to automatically assemble images are based on pyramidal decomposition (Dani and Chaudhuri, 1995). Computation time would be reduced and shifts in the x, y, z direction could be envisioned.

Simple size and shape parameters were used to describe the cell morphology. Cell wall thickness was assessed for each cell using a procedure based on the skeleton and the distance function. Travis *et al.* (1996) estimated cell wall thickness by a similar procedure after a watershed segmentation. Such approaches make the extraction of cell wall thickness profile easy by collecting all values measured by the distance function along the skeleton. This would reveal local thickness variations within the cell perimeter. Considering both cell lumen and cell wall is a challenge as both features may contribute to the mechanical resistance and texture of plant material (Jackman and Stanley, 1995).

Cell groups were determined by a clustering procedure. Discriminant analysis could have been used alternatively and would have required building a calibration set (Travis *et al.*, 1996). An automatic clustering procedure has been preferred to reveal the groups that came out from cell morphology. The cell population was successively split into an increasing number of groups in a completely determined way and the procedure did not depend on random starting groups. Only the final number of groups had to be decided from the visual examinations of the label images.

In the present work, individual cell features were extracted to study plant histology. Results showed that these features were relevant to reveal much of the

group that would have been defined by an expert. The next steps are to study the contribution of cell arrangement and to take into account the 3D information visualised by confocal microscopy.

ACKNOWLEDGEMENT

Financial support by the Region des Pays de la Loire is gratefully acknowledged.

REFERENCE

- Barret DM, Garcia E, Wayne JE (1998). Textural modification of processing tomatoes. *Crit Rev Food Sci* 38:173-258.
- Dani P, Chaudhuri S (1995). Automated assembling of images: image montage preparation. *Pattern Recogn* 28:431-45.
- De Smedt V, Pauwels E, De Baerdemaeker J, Nicolaï B (1998). Microscopic observation of mealiness in apples: a quantitative approach. *Postharvest Biol Tec* 14:151-8.
- Esau K (1977). *Anatomy of seed plants*. Second edition. John Wiley & Sons.
- Filzmoser P, Baumgartner R, Moser E (1999). A hierarchical clustering method for analyzing functional MR images. *Magn Reson Imaging* 17:817-26.
- Gao X, Tan J, Shatadal P, Heymann H (1999). Evaluating expanded-food sensory properties by image analysis. *J Texture Stud* 30:291-304.
- Guines F, Julier B, Ecalle C, Huyghe C (2003). Among and within cultivar variability for histological traits of lucerne (*Medicago sativa* L.) stem. *Euphytica* 130: 293-301.
- Gray JD, Kolesik P, Høj PB, Coombe BG (1999). Confocal measurement of the three-dimensional size and shape of plant parenchyma cells in a developing fruit tissue. *Plant J* 19(2):229-36.
- Jackman RL, Stanley DW (1995). Perspectives in the textural evaluation of plant foods. *Trends Food Sci Tech* 6:187-94.
- Jolliffe IT (1986). *Principal Component Analysis*. Springer series in statistics. New York: Springer-Verlag.
- Karen P, Jirkovská M, Tomori Z, Demjénová E, Janáček J, Kubínova L (2003). Three-dimensional computer reconstruction of large tissue volumes based on composing series of high-resolution confocal images by GlueMRC and LinkMRC softwares. *Microsc Res Tech* 62:415-22.
- Konstankiewicz K, Czachor H, Gancarz M, Król A, Pawlak K, Zdunek A (2002). Cell structural parameters of potato tuber tissue. *Int Agrophysics* 16:119-27.
- Lebart L, Morineau A, Pirin M (1995). *Statistique exploratoire multidimensionnelle*. Paris: Dunod, 145-205.
- Perré P (1997). Image analysis, homogenization, numerical simulation and experiment as complementary tools to enlighten the relationship between wood anatomy and drying behavior. *Dry Technol* 15(9):2211-38.
- Soille P (1999). *Morphological Image Analysis: Principles and Applications*. Berlin: Springer-Verlag, 173-4.
- Tomažević D, Likar B, Pernuš F (2002). Comparative evaluation of retrospective shading correction methods. *J. Microsc* 208:212-23.
- Travis AJ, Hirst DJ, Chesson A (1996). Automatic classification of plant cells according to tissue type using anatomical features obtained by the distance transform. *Ann Bot* 78:325-31.
- Zghal, MC, Scanlon MG, Sapirstein HD (2002). Cellular structure of bread crumb and its influence on mechanical properties. *J Cereal Sci* 36:167-76.