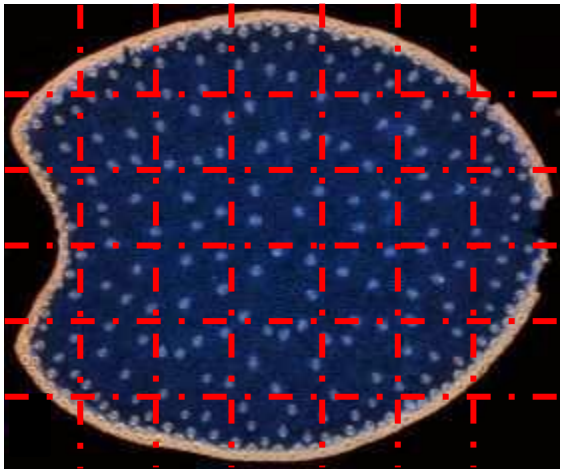
Multi-scale k-means clustering of multispectral images

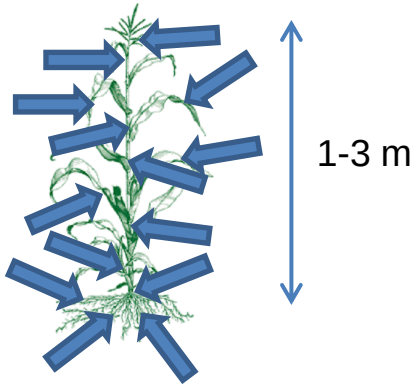
Corcel Mathias[1], Barron Cécile[2], Guillon Fabienne[1], Devaux Marie-Françoise[1]
1) UR 1268 BIA, INRA Nantes, France
2) UMR 1208 IATE, INRA Montpellier, France

Chimiométrie XVII – Namur - 19/01/2016

Large biological variability

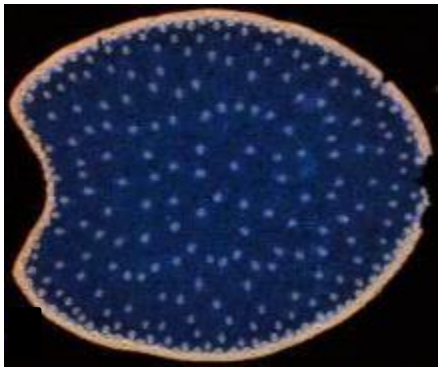
- Many images are acquired for plant caracterization
  - To form a mosaic image representing the cross-section
  - To observe multiple sections of the same sample
  - To observe multiple samples





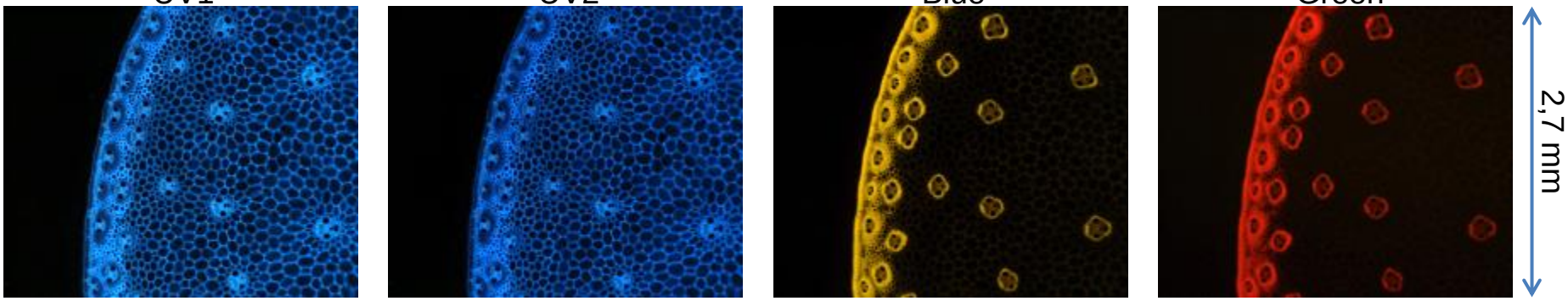
- Large data
  - Mosaic image with large number of pixels (~20M pixels / mosaic image without background)
  - + number of image per sample
  - + number of samples

Histological analysis through multispectral imaging



- Histological analysis of plant tissues
  - Maize (*Zea mays)* stems
  - Many phenolic components present in cell walls
    - Hydroxycinnamic acids : ferulic, p-coumaric …
    - Lignins with various types of monomers
- These components are autofluorescent
- Explore the variability of the different tissues using their autofluorescence
  - Multispectral fluorescence imaging using 4 different excitation conditions of UV or visible light

UV1 UV2 Blue Green



Objective

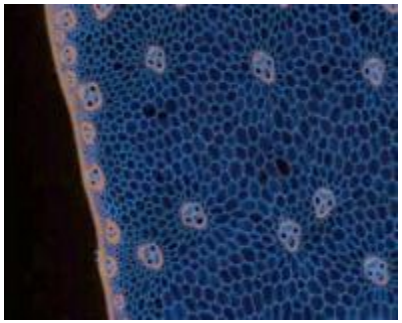
- Use methods to classify a large number of pixels using all spatial variability with unsupervized classification

- In order to explore the variability

- Using data with large population but small number of variables
  - No a priori knowledge on autofluorescence for most tissues
  - Not possible to select pixels or regions of interest

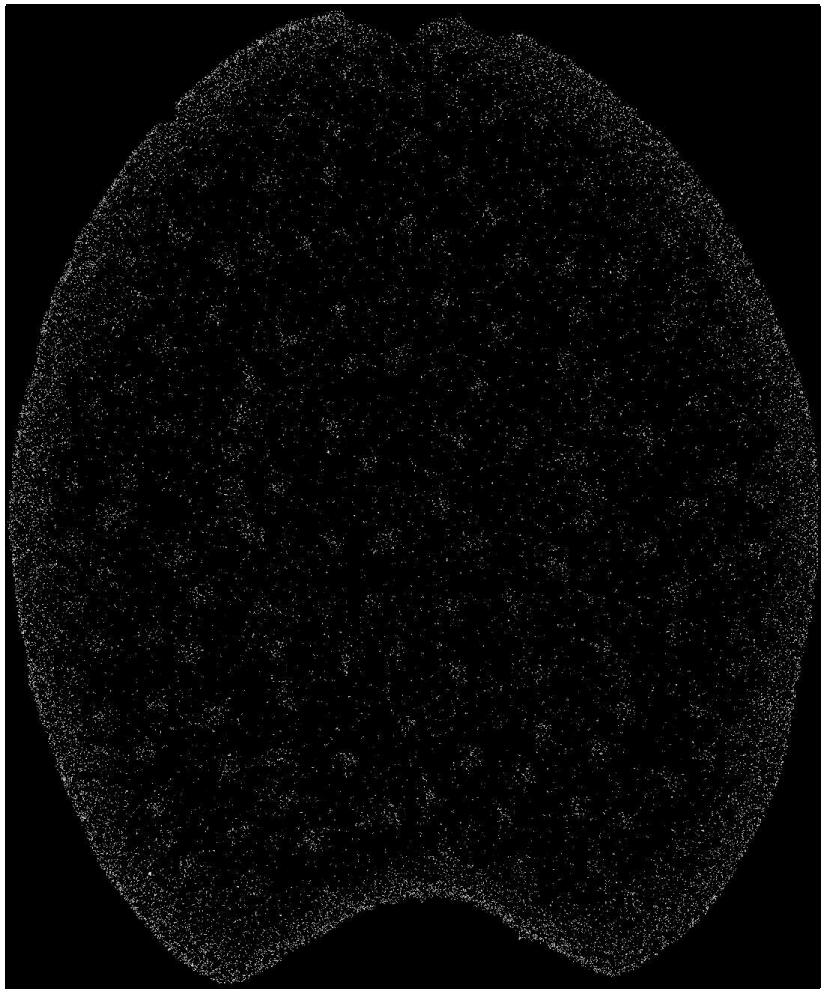
Pyramids in image processing

- Multi-scale representation on an image
- The image is reduced in each level
- Simple pyramid : Mean pyramids with halved resolution each level
  - Each new pixel is the mean of 4 pixels



Principle of classification using pyramids

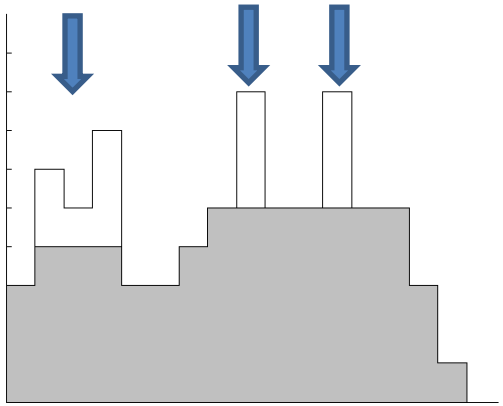
- 1) Start at the level with the smallest resolution
- 2) Classify pixels into *k* clusters
  - Using k-means with Euclidian distance on Principal components
- 3) Select *n* pixels in each cluster
  - Random selection
- 4) Expand the selected pixels in the next level
- 5) Repeat from step 2 until the last level of the pyramid

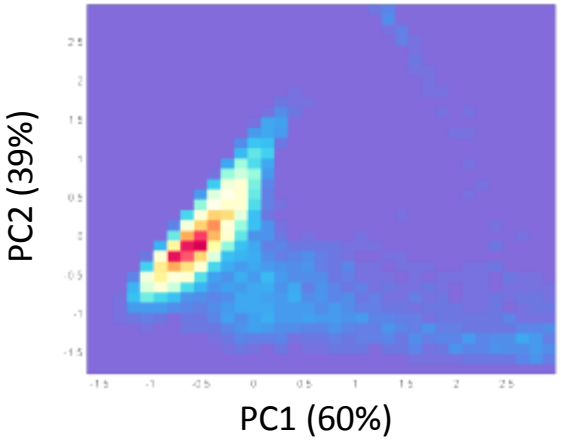


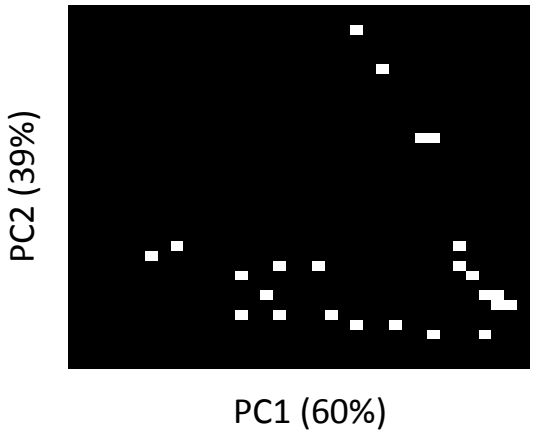
Selection of the pixels in each level

K-means initialisation : Number and position

- Histogram (+ Smoothing) of the Score plot Principal Components 1 and 2
- H-maxima transform
- Gives initial positions for clusters on components 1 and 2
- Positions on other components set to 0

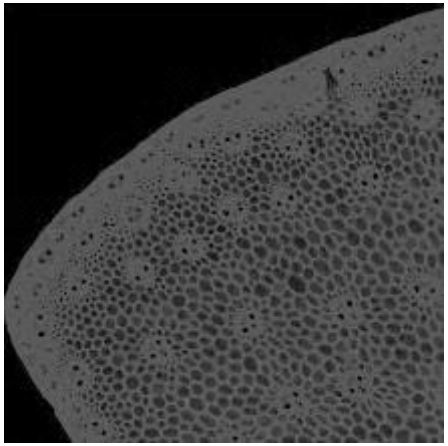


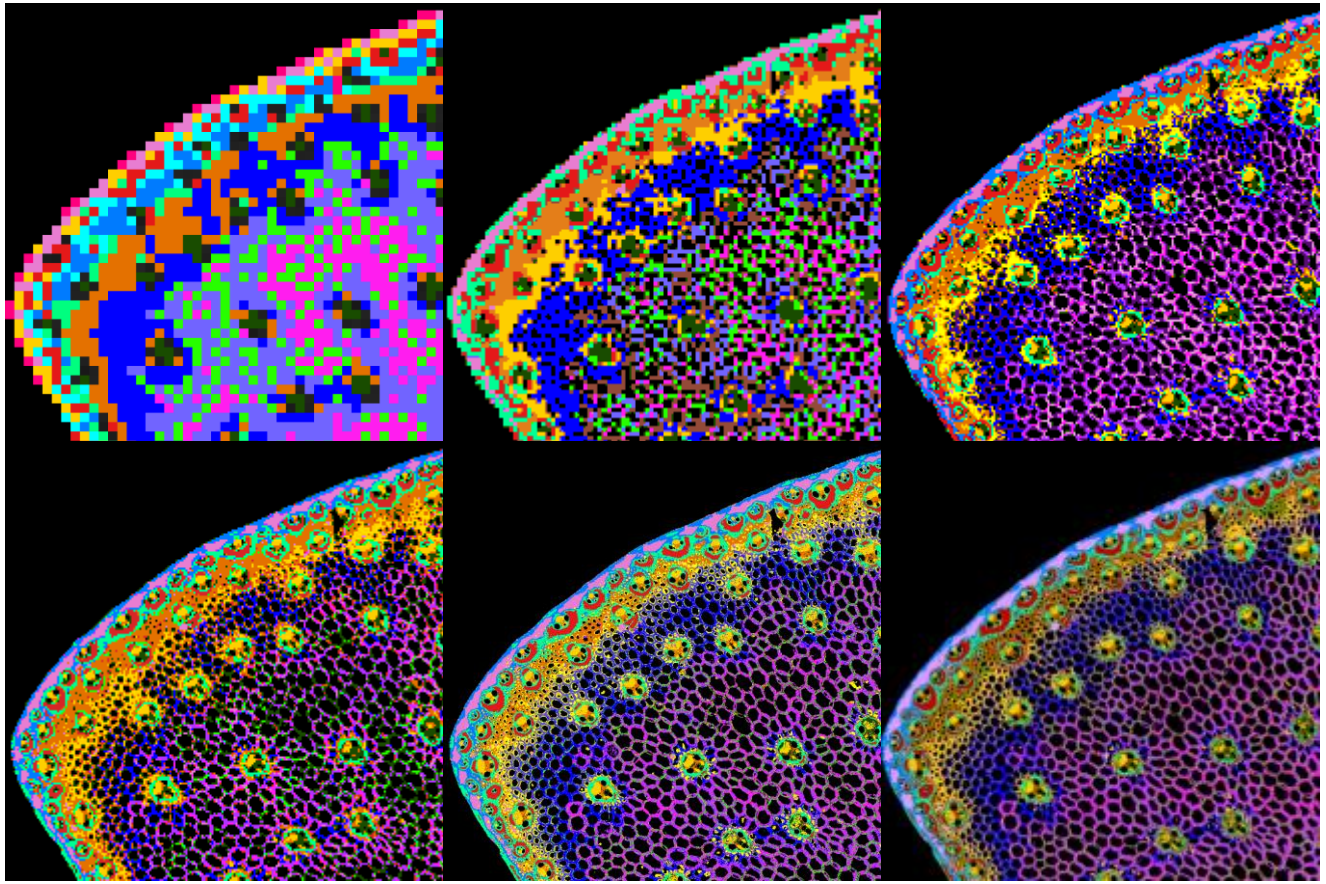




Method used on an image

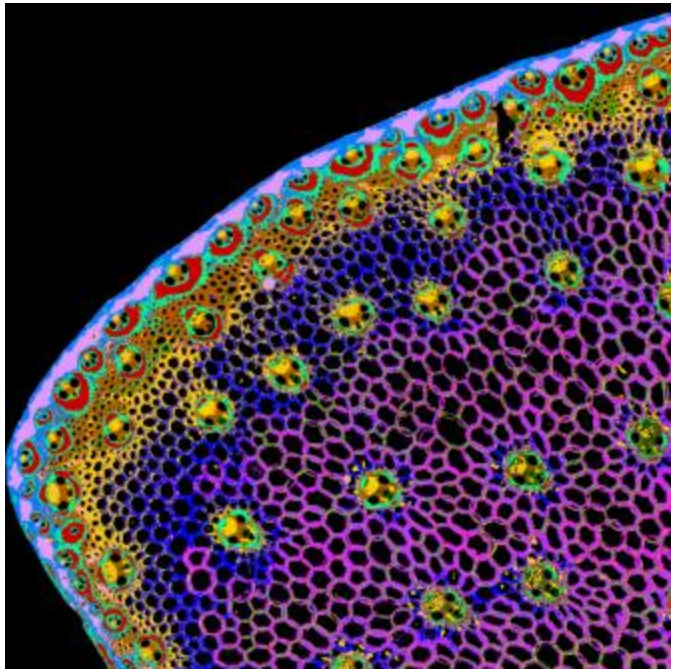
- Start at level 6
  - Resolution 1/32
- 15 clusters found
- 2000 pixels per cluster selected
  - Constant



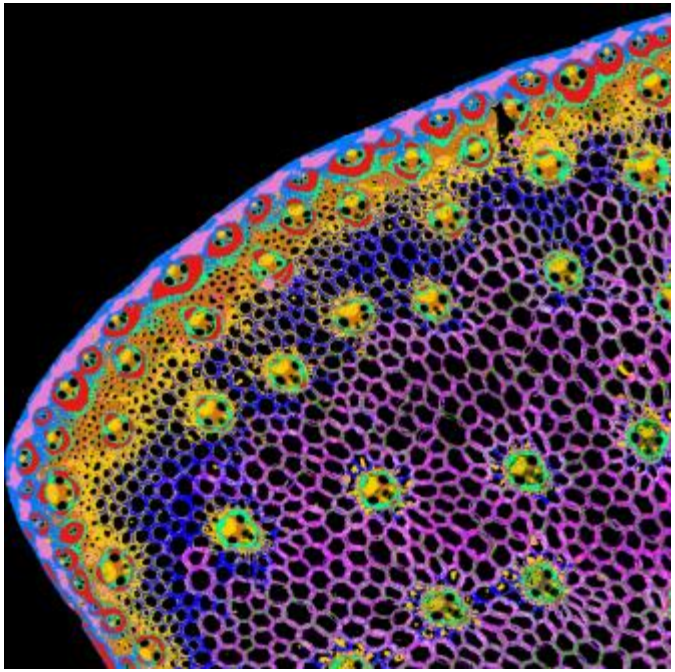


Comparison of classification with and without pyramids

Using pyramids
(with the same parameters as before)



Without pyramids



- The clusters obtained are largely similar

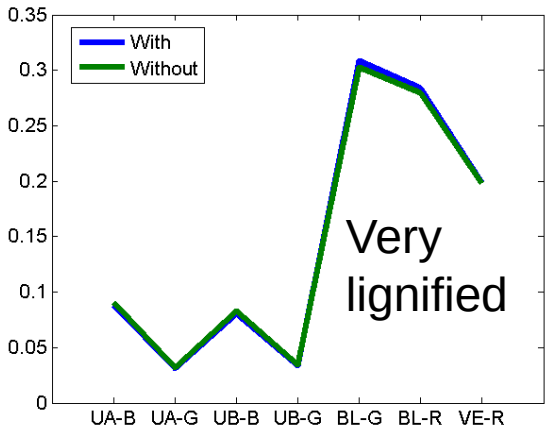
Comparison of classification with and without pyramids

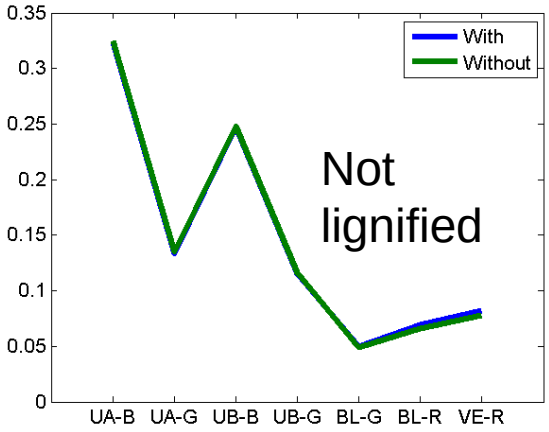
With Pyramids

Without Pyramids

Sclerenchyma

Parenchyma





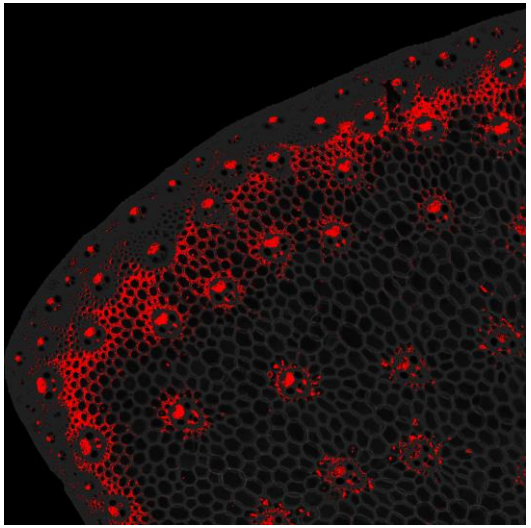
Comparison of classification with and without pyramids

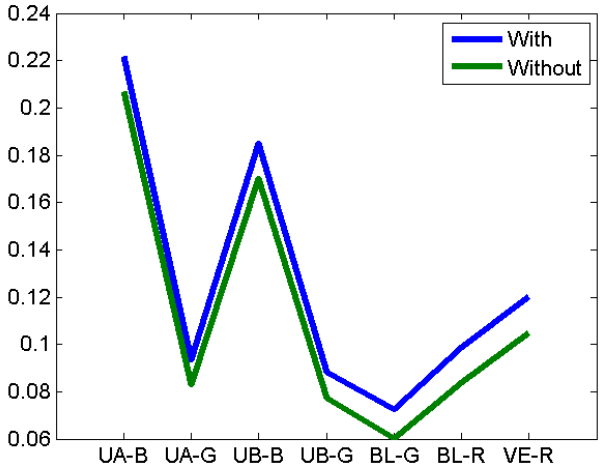
- Some clusters regroup multiple tissues (Identified by their biological role and/or morphology)
  - Their autofluorescence is too similar
  - Their react the same way to some enzymes
- And both approaches with/without pyramids have the same clusters

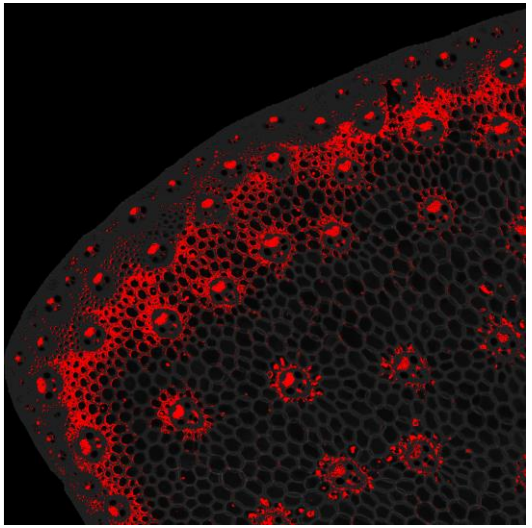
With Pyramids

Without Pyramids

Phloem + Some Parenchyma cells







Conclusion

- Simple adaptation of image analysis and chemometry : image pyramids and k-means, allow to segment large images with satisfactory results

- The number of pixels is reduced by a factor of ~1000, allowing to analyze a batch with many mosaic images

- The segmentation is done using all the initial variability and without loss during the process

Perspectives

- Test the limits of pyramids
  - Further decrease the resolution
- Improve the selection of pixels
  - Validate the repeatability of the method due to the random selection of pixels in each cluster
  - New method of selection based on the variance inside each mean pixel
- Find a solution to objectively compare unsupervized classifications
- Analyse complete collections of images

Thanks for your attention !