



Disposition des feuilles sur la vitre d'acquisition

La disposition des feuilles sur la vitre du scanner¹ est primordiale pour obtenir une labellisation conforme aux attentes de l'utilisateur.

La labellisation permet d'individualiser chacune des feuilles présentes sur l'image numérique ; elle est censée les ordonnant selon l'ordre naturel² de lecture, c'est-à-dire en parcourant l'image de gauche à droite et de haut en bas.

Plan

Principe de disposition des feuilles	2
Exemples	3
Définition des lignes.....	5

Les exemples illustrent les résultats obtenus sur quelques configurations, en insistant sur les écueils à éviter.

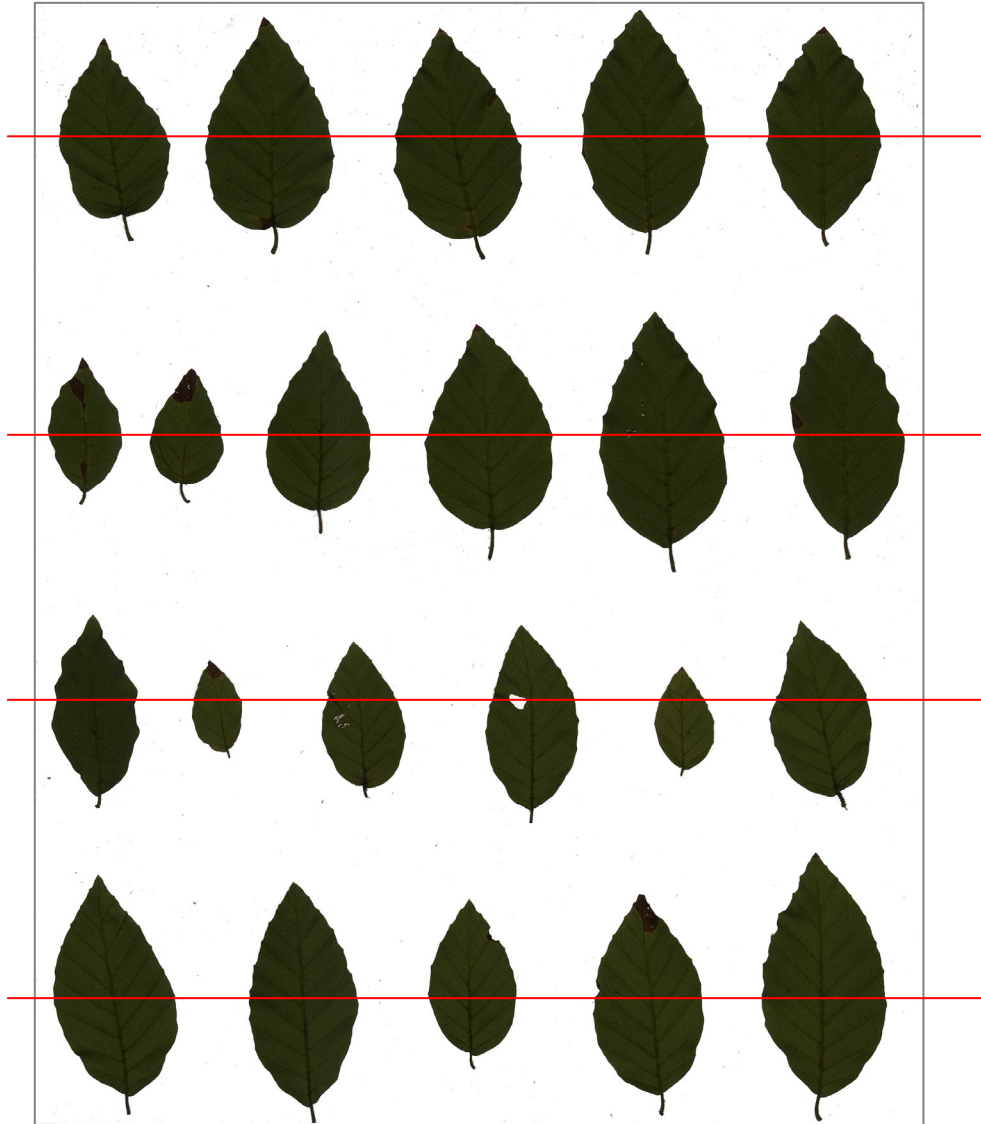
¹ Le scanner utilisé est un Epson Perfection V700 Photo est équipé d'un **dispositif de rétroéclairage** : deux barrettes de photodiodes, disposées en vis-à-vis de part et d'autre de la zone de numérisation, se déplacent de manière synchrone ; ce dispositif, développé à l'origine pour la numérisation de diapositives et de transparents, permet de limiter l'effet « ombre portée » des objets non plans, en particulier des feuilles fraîches.

² L'ordre naturel de lecture de l'image est notamment utilisé pour associer des paramètres provenant de protocoles de mesure complémentaires : par exemple, le poids et la surface des feuilles... la pesée est effectuée feuille à feuille par un opérateur, les surfaces sont évaluées de manière globale par un outil automatique. Il est donc primordial de garantir la mise en correspondance des différentes valeurs.



Principe de disposition des feuilles

Les feuilles d'une même ligne sont disposées sur la vitre de façon à ce que leurs centres soient (à peu près) alignés suivant une horizontale ; si possible, bien séparer les lignes



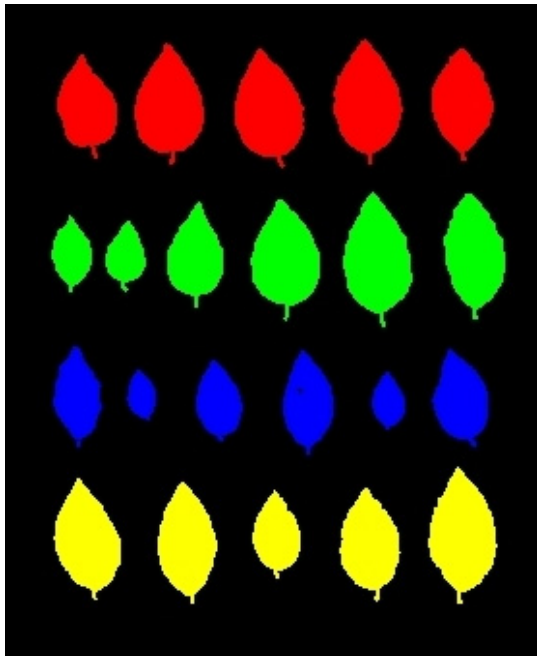
Les feuilles sont posées sur la vitre, ligne par ligne, de la gauche vers la droite, leur face supérieure orientée vers la vitre de numérisation.

Point important : il est préférable de choisir une orientation identique³ pour l'ensemble des feuilles, par exemple orienter la direction principale de la feuille selon la verticale de la vitre.

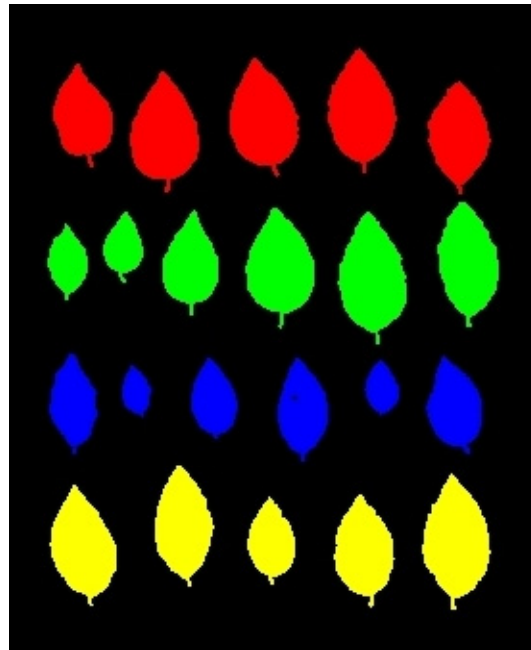
³ La numérisation, et par la même la précision des résultats numériques, est sensible aux variations d'orientations. Pour plus d'informations, se reporter au document « Biais et Incertitudes ».



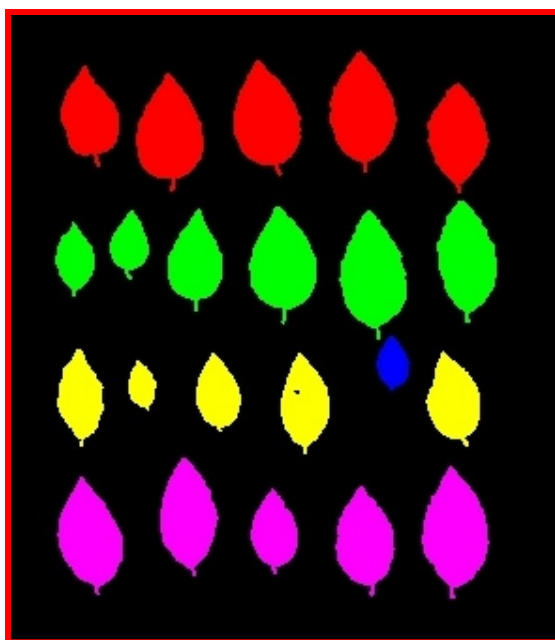
Exemples



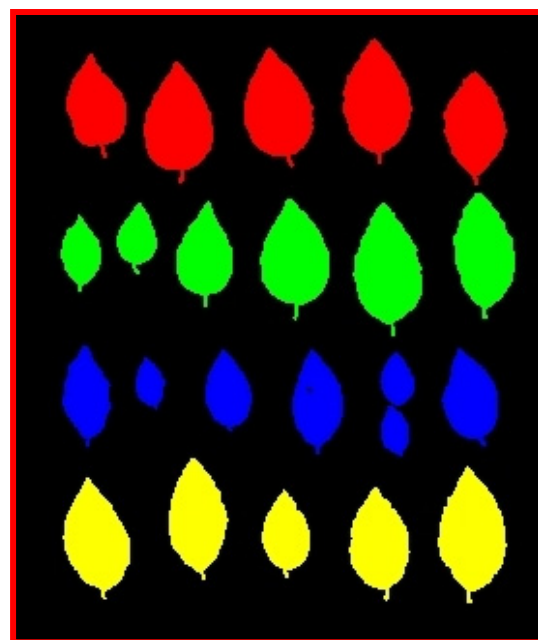
Définition automatique des lignes sur l'image précédente



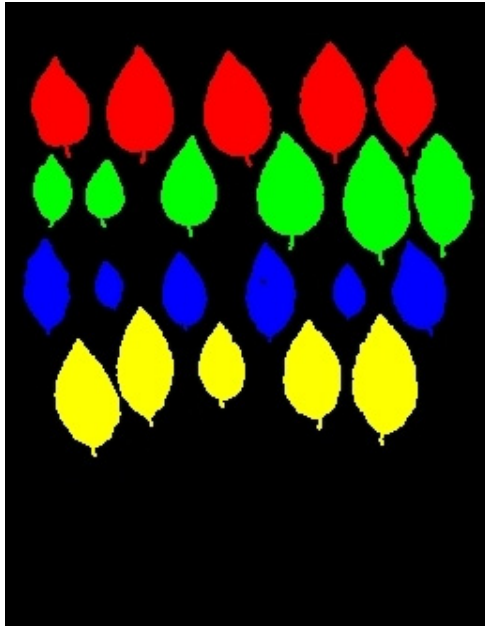
L'image initiale a subi une légère perturbation : pas de modification dans le résultat d'identification des lignes



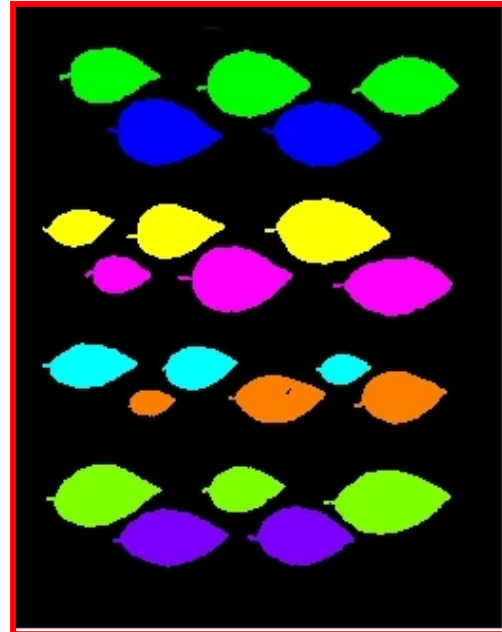
L'image initiale a subi une perturbation entraînant une identification erronée des lignes. La feuille affichée en bleu est trop haute. Le principe de labellisation décrit au § suivant est mis en défaut.



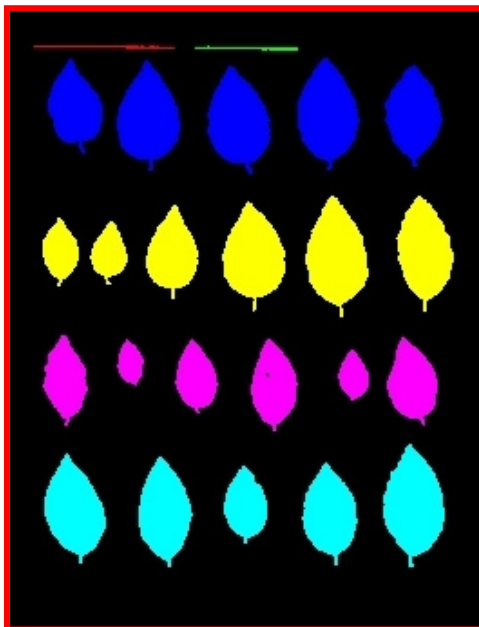
Eviter de disposer des feuilles sur deux rangs au sein de la même ligne. Ici dans la ligne n°3, représentée en bleu, deux feuilles sont alignées selon la même verticale. Dans cette configuration, la feuille inférieure sera labellisée avant la feuille supérieure



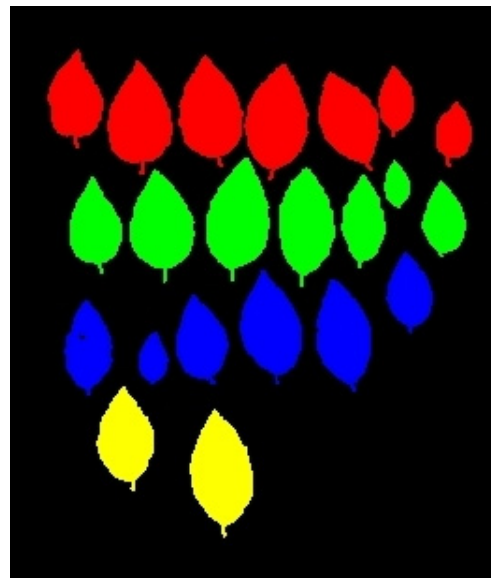
Possibilité de disposer les feuilles en quinconce pour gagner de la place. Si l'alignement des centres est respecté, l'identification des lignes se fait correctement.



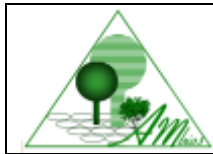
La lecture sinusoïdale n'est pas possible. Chaque sinusoïde est décomposée en deux lignes. Dans ce cas, il faudra réordonner l'ensemble des résultats.



Attention aux structures parasites provoquant un décalage dans la numérotation des lignes, et donc dans la labellisation des composantes connexes. Ici, les « lignes » rouge et verte correspondent à l'ombre du bord de la vitre utilisée pour aplanir les feuilles.

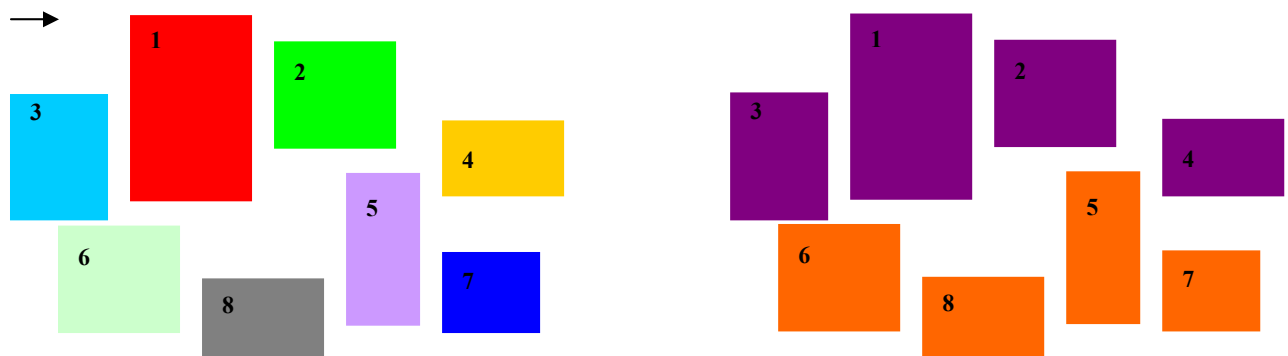


Si nécessaire, les feuilles peuvent être disposées selon des obliques (lignes bleue et jaune). Mais pour avoir un résultat conforme à ce que l'on attend, mieux vaut bien comprendre le processus de reconnaissance de lignes (voir paragraphe suivant).



Définition des lignes

Les feuilles de l'image numérique sont regroupées par ligne selon une méthode censée reproduire une lecture naturelle de la planche, c'est-à-dire de gauche à droite et de haut en bas. La méthode, baptisée Lactoxa pour Labellisation Automatique par ClusTering des ordonnées et Classification Croissante des Abscisses) repose sur deux étapes



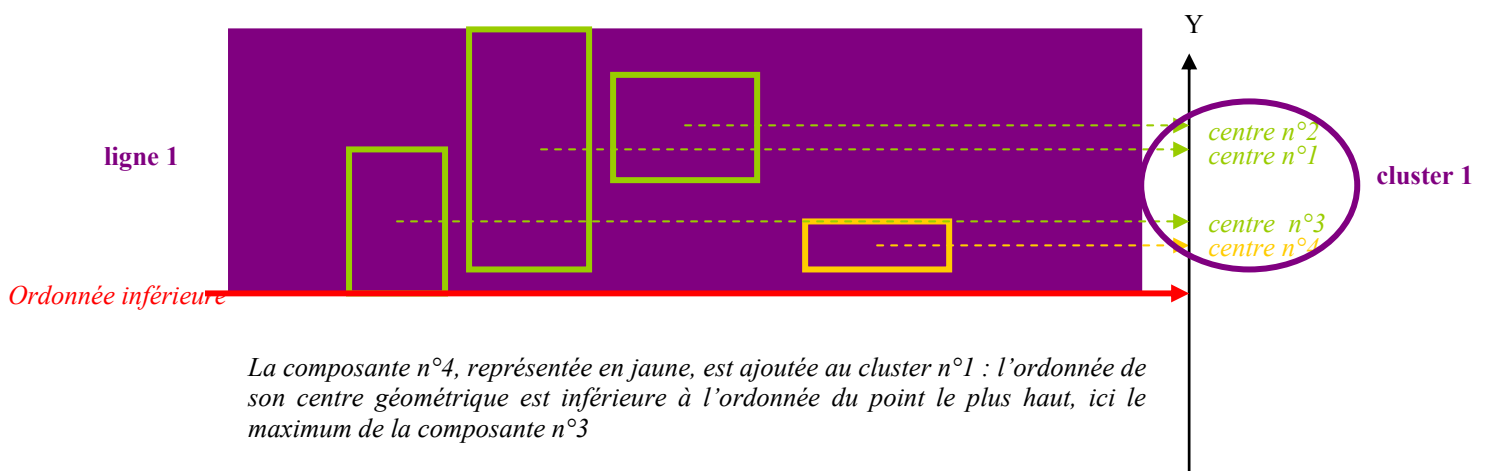
A gauche : labellisation de chaque composante connexe par balayage ligne par ligne de l'image. La flèche indique le sens et le départ du parcours

A droite : regroupement des composantes ligne par ligne. Ici, nous avons deux contenant chacune quatre composantes connexes

Au sujet de la seconde étape :

Les lignes sont définies par regroupement ou clustering des composantes connexes selon le postulat suivant :

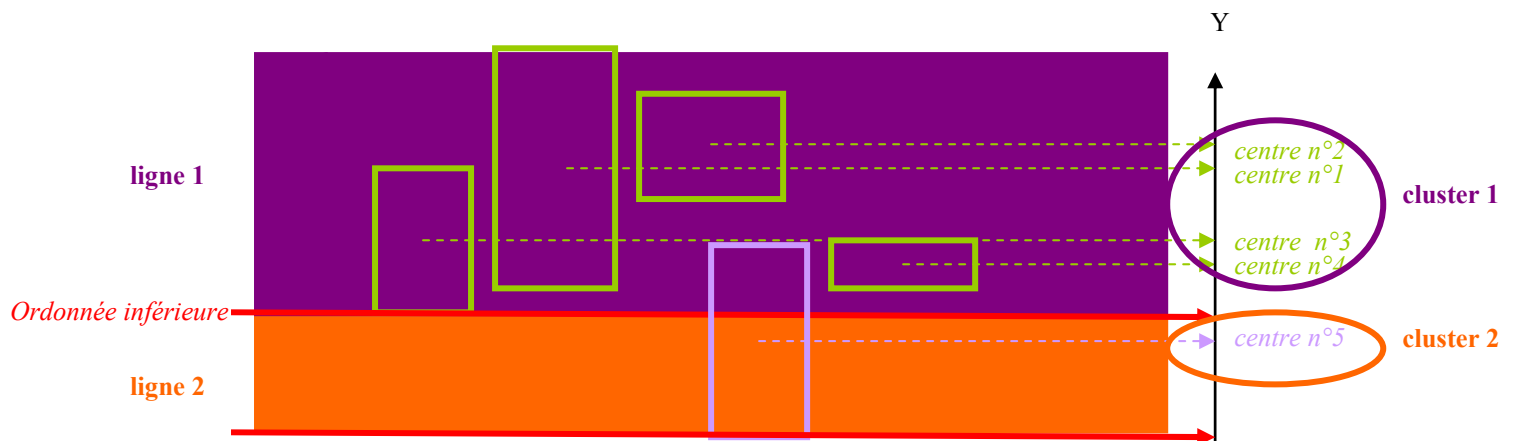
- **une nouvelle composante connexe appartient à la ligne courante si et seulement si l'ordonnée de son centre géométrique est plus basse que l'ordonnée du point inférieur de l'ensemble des composantes déjà présentes dans la ligne**



La composante n°4, représentée en jaune, est ajoutée au cluster n°1 : l'ordonnée de son centre géométrique est inférieure à l'ordonnée du point le plus haut, ici le maximum de la composante n°3



De ce fait, si le postulat n'est pas vérifié, la composante courante initialise une nouvelle ligne.



La composante n°5, représentée en mauve clair, ne peut être rajoutée au cluster n°1, l'ordonnée de son centre géométrique étant inférieure à l'ordonnée du point le plus haut, ici le maximum de la composante n°3. Un nouveau cluster est donc créé à partir de cette nouvelle composante.