
Mémoire de Projet de fin d'études

PATTERN DE DÉVELOPPEMENT CHEZ LA CYANOBACTÉRIE ANABAENA
PCC7120



ÉQUIPE PROJET

M. Housseem GARCI

M. Thierry NIBERT-SIBERT

DÉPARTEMENT INFORMATIQUE

SITE DE LUMINY

Tuteur École :

M. Mouloud ADEL

22 février 2016

Liste de diffusion & de Modification

Liste de diffusion

Nom – Prenom	Rôles
Housseem Garci	Equipe projet – Relecture et correction
Thierry Nibert-siber	Equipe projet – *
Mouloud ADEL	Tuteur – enseignant chercheur
Hailin CHEN	Mandataire – doctorant
Cheig-Cai ZHANG	Mandataire – enseignant chercheur
Jean-Marc Boi	Responsable de formation –

Liste des Modification

Versions	Correction apporter
V 1.0	Première diffusion

Résumé

Nous avons effectué nos 6 semaines de projet de fin d'études (PFE) de dernière année du Master images et systèmes (M2 IS IN) à Aix-Marseille Université.

L'objectif principale de ce projet est de mettre en œuvre un suivi hiérarchique de la vie d'une cellule.

Ce mémoire a pour but de vous présenter les algorithmes que nous avons mis en place afin d'automatiser la chaîne d'analyse de lignée cellulaire, ainsi que les résultats et les analyses critiques de nos algorithmes.

Glossaire

2-OG oxo-glutarate. 3, 8

CFP protéine fluorescente cyan. 3, 10

CO2 dioxyde de carbone. 3, 7

csv feuille excel. 3, 14

dioxyde de carbone Molécule.. 3, 7

feuille excel format de tableau open sources. 3, 14

FRET énergie fluoescence transmie. 3, 10

hétérocyste hétérocyste. 3

hétérocyste cellule fixatrice d'azote, transparente, à paroi épaisse, qui se forme dans les filaments de certaines cyanobactéries. La paroi est épaisse pour limiter la diffusion de l'oxygène, car la nitrogénase (complexe fixant l'azote) y est sensible. 3

IHM interface homme machine. 3

imagerie médical l'imagerie numérique appliquer dans le milieu médical. 3

imagerie médical imagerie médical. 3

imagerie numérique domaine de l'informatique graphique.. 3

imagerie numérique imagerie numérique. 3

interface homme machine utilitaire graphique ou non permettant d'interagir avec une machine. 3

Matlab Matrix laboratory. 3

Matrix laboratory Une application et un langage de programmation. utiliser pour du prototypage dans le monde de la recherche.. 3

N Nitrate. 3

ND2 Nikon NIS-Elements. 3, 10

Nikon NIS-Elements Cette extension est associé a un format d'images acquis par "Nikon NIS-Elements" NIS-Elements est un outil pour les niveaux les plus avancés de recherche qui donnent à l'utilisateur un contrôle complet de toutes les fonctions du microscope et de la caméra. Pour enregistrer des séquences d'images avec une qualité sans perte, NIS-Elements a désigné le format ND2. Le format ND2 utilise la compression JPEG-2000, et peut être également compressé ou décompressé au format TIFF par exemple.. 3, 10

Nitrate Molécule chimique que l'on retrouvent sous la forme de nitroglycérine, nitrate, acide aminé.. 3

Open Computer Vision Bibliothèque de traitement d'image.. 3

OpenCV Open Computer Vision. 3

protéine fluorescente cyan cyan florecent proteine, molécule fluorencente cyan.. 3, 10

protéine fluorescente jaune Yellow florecent proteine, proteine fluorencente jaune.. 3, 4, 10

reconstruction géodésique reconstruction géodésique. 3

reconstruction géodésique filtrage morphologique, opération morphologique consistant à rendre les opérations non expansive.. 3

ROI Régions d'intérêt. 3, 8, 13

Régions d'intérêt Region of interest, termes le plus souvent utiliser en imagerie médical pour designer la zones d'un image comportant les information utile.. 3, 4, 8

squeletisation,squelette un squelette et le résultat d'un grand nombre d'amincissement homotopique sur une objet 2D ou 3D.. 3, 4

squelette squeletisation,squelette. 3

Tag(ged) Image File Format Le format TIFF est un ancien format graphique, permettant de stocker des images bitmap de taille importante, sans perte de qualité.

Il permet de stocker des images en noir et blanc, en couleurs réelles jusqu'à 32 bits par pixels ainsi que des images indexées, faisant usage d'une palette de couleurs.. 3, 4, 10

TIFF Tag(ged) Image File Format. 3, 10

YFP protéine fluorescente jaune. 3, 10

énergie fluorescente transmie Florecent Energie Transfert. 3, 10

Remerciements

Avant tout développement sur cette expérience professionnelle, il apparaît opportun de commencer ce mémoire par des remerciements, à ceux qui nous ont beaucoup appris et aider au cours de ces 6 semaines, et même à ceux qui ont eu la gentillesse de faire de ce projet un moment très profitable.

Aussi , nous tenons à remercier tout l'équipe enseignante du Master images & systèmes.
Et tout particulièrement nous remercions M. Mouloud Adel, ainsi que M. Jean-marc Boi pour nous avoir permis d'effectuer ce projet de fin d'étude.
Nous remercions aussi M Hailin Chen ainsi que M Cheig-cai Zhang, pour la patience et la confiance qu'ils nous ont accordé.
Enfin, nous remercions également nos collègues et amis de la promotion qui nous ont côtoyé tout au long de l'année.

Table des matières

Résumé	2
Glossaire	3
Remerciements	5
Introduction	7
1 Détection et extraction des données cellulaires	9
1.1 Fonctionnement générale	9
1.2 Algorithme de détection des cellules	10
1.2.1 Étiquetage(watershed)	12
1.2.2 Extraction Des Régions d'intérêt (ROI)	13
1.2.3 Projection des cellules	13
1.2.4 Extraction et sauvegarde des données statistiques	14
2 Analyse des Résultats	15
Conclusion	18

Introduction

Le contexte

La formation du Master Images et Systèmes spécialité Imagerie Numérique de 2nd année se complète par un projet de fin d'études. Celui-ci constitue une étape obligatoire pour l'obtention du diplôme. Dans ce cadre, nous avons travaillé sur le sujet suivant : Pattern de développement chez la cyanobactérie anabaena PCC7120.

L'origine du projet

Ce projet se place dans le cadre d'une étude de recherche sur le phénomène de différenciation de cellules souches afin d'une part de prédire quelles cellules vont se spécialiser en hétérocyste en mesurant le nombre de molécules présentes dans une cellule, et d'autre part de pouvoir suivre leur lignée cellulaire au cours du temps.

La cyanobactérie anabaena

Les cellules souches appartiennent à une variété d'algue appelé cyanobactérie anabaen PCC7120, utilisée pour la production d'hydrogènes. Elle a la propriété de se nourrir du dioxyde de carbone (CO₂) présent dans son milieu grâce à la photosynthèse, et de produire les protéines qui lui donnent une couleur verte.

La différenciation et la spécialisation

La différenciation des cellules se produit lors d'une carence en Nitrate(N) dans le milieu où vit la cyanobactérie, en le synthétisant grâce aux cellules hétérocystes (voir figure 1)

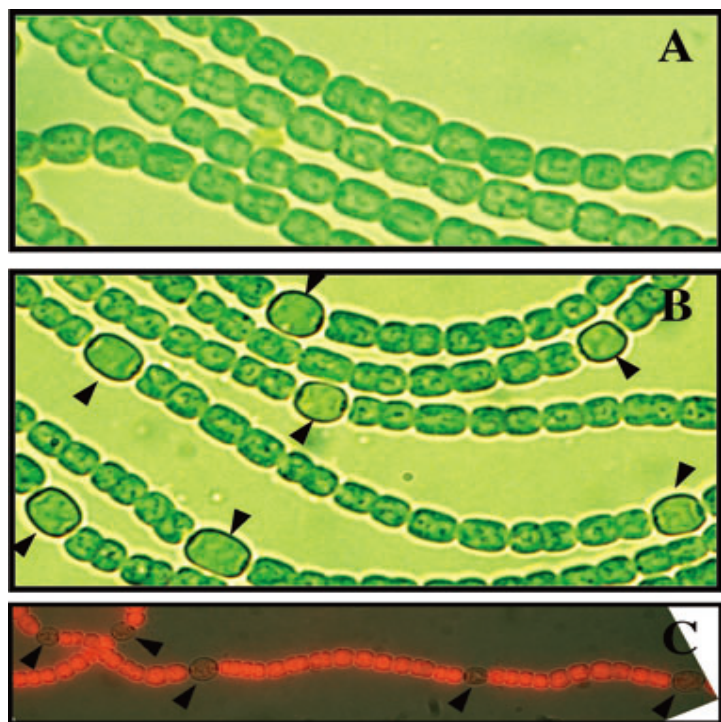


FIGURE 1 – filament avant(A) et après(B) privation en nitrogène.

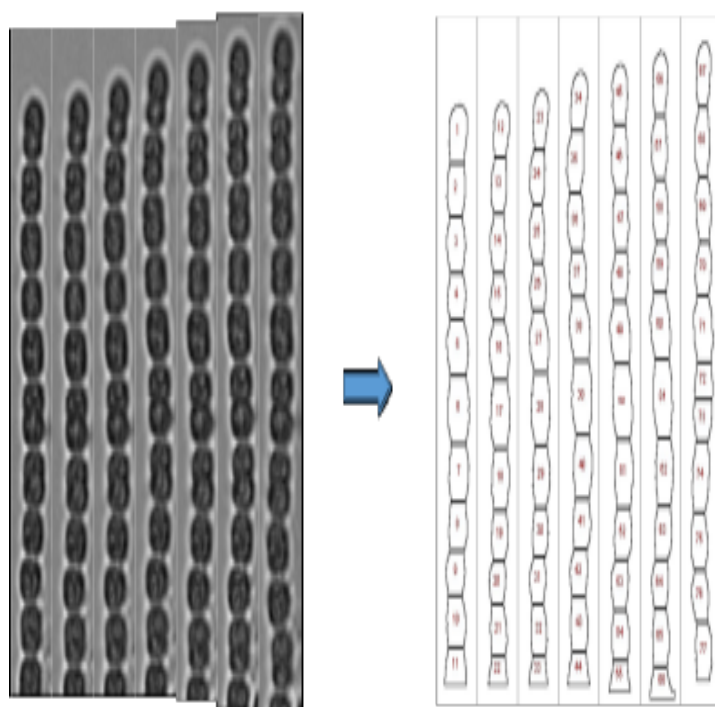


FIGURE 2 – Filament étiqueté.

Les hypothèses, objectifs et le processus d'acquisition des images

En se basant sur l'hypothèse que le niveau de oxo-glutarate (2-OG) varie dans les cellules hétérocystes permet de prédire quelles cellules se diviseront. L'objet est de mesurer cette variation de façons non destructives, pour pouvoir suivre l'évolution d'une cellule, tout en observant l'évolution de la population cellulaire. Afin de savoir si une cellule se différencie, on quantifie le niveau de 2-OG (voir figure 3) présent au sein de la population, pour ce faire on injecte une protéine fluorescente, qui rayonne sur une longueur d'onde unique

Notre missions

Pour des raisons de délais nous avons décidé d'automatiser la première partie du processus d'analyse qui consiste à extraire les ROI qui sont les cellules dans un filament de façon automatique 2, puis d'extraire pour chaque ROI son intensité lumineuse.

Ce traitement ne concerne que l'image de l'état initial de la cyanobactérie et le suivi d'une séquence complète.

Nous vous présentons dans ce mémoire l'algorithme que nous avons mis en place ainsi que les raisons de nos choix techniques.

Molécule 2-OG 
Oxo-Glutarate

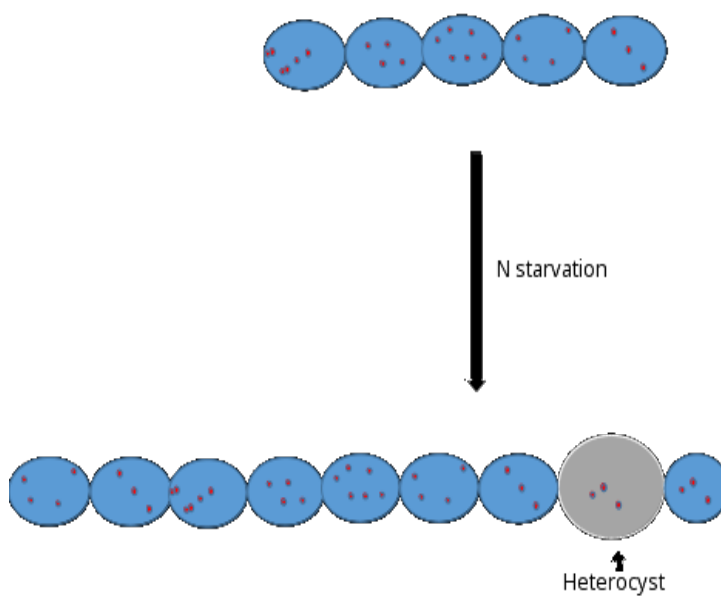
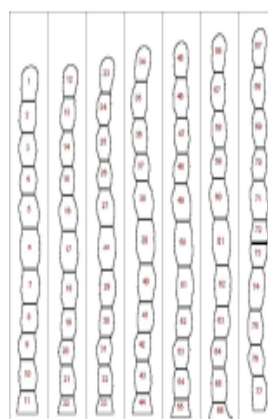


FIGURE 3 – variation du niveau de 2-OG dans un cellule Hétérocyste.



	A	B	C	D	E
	FRET	CFP	YFP		
1	4018.99	4029.32	1821.01		
2	4494.05	4478.75	1923.42		
3	4685.35	4685.18	1966.65		
4	5286.3	5249.39	2095.32		
5	6052.51	6033.51	2478.56		
6	6068.24	6035.51	2260.61		
7	4335.5	4343.19	1845.79		
8	5273.51	5302.44	2075.88		
9	4974.51	4963.08	2006.26		
10	4469.09	4480.06	1874.25		
11	4297.36	4299.02	1888.74		
12	3968.12	3940.57	1824.43		
13	4412.96	4334.81	1929.75		
14	4618.77	4534.82	1956.9		
15	5168.1	5080.54	2091.85		
16	6790.14	6624.22	2458.41		
17	5996.91	5862	2261.41		
18	4297.07	4207.99	1848.97		
19	5221.01	5100.34	2067.18		
20	4947.18	4840.21	2005.94		
21	4424.33	4358.72	1867.81		
22	4520.22	4434.19	1881.35		
23	3921.58	3865.14	1817.67		
24	4364.99	4276.62	1916.5		
25	4534.34	4416.79	1940.81		
26	5027.32	4959.35	2058.19		
27	6575.86	6425.52	2409.05		
28	5866.58	5731.79	2217.78		
29	4236.11	4132.21	1837.21		
30	5108.16	5009.55	2049.27		
31	4853.1	4753.95	1990.79		
32	4400.39	4306.91	1887.4		
33	4436.06	4357.81	1872.21		
18	5007.10	4990	1805.43		

FIGURE 4 – cellules étiquetées avec leurs YFP CFP et FRET exportés dans un fichier csv.

1 Détection et extraction des données cellulaires

1.1 Fonctionnement générale

Schéma fonctionnel

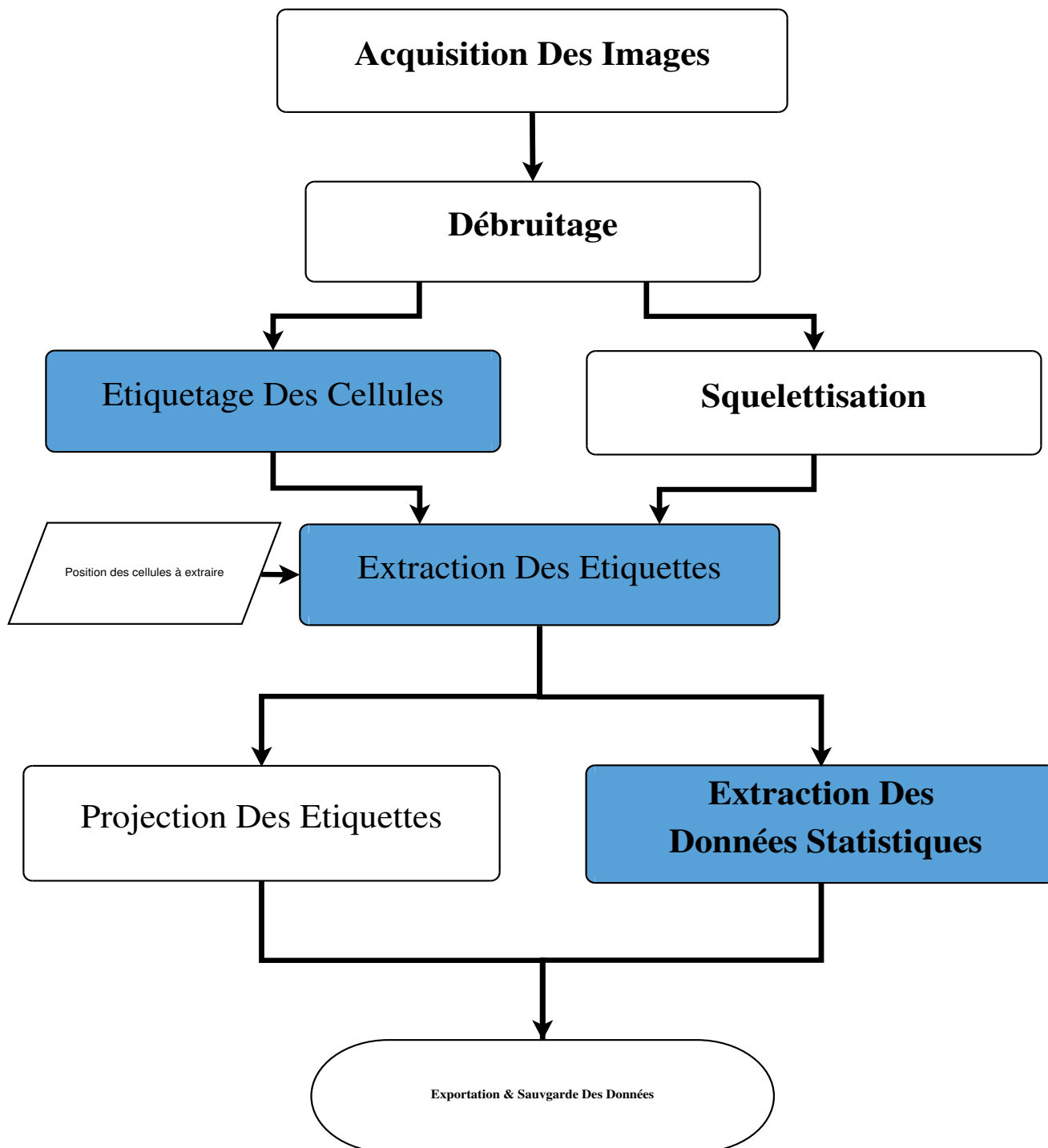


FIGURE 5 – Algorithme : Processus général.

1.2 Algorithme de détection des cellules

L'utilisation d'images multi-canaux aux formats Tag(ged) Image File Format (TIFF) et Nikon NIS-Elements (ND2) nous on été imposés pour les raisons suivantes :

- Matériel d'acquisition utilisé dans le laboratoire pour la capture des images sous microscopie.
- Support de la haute qualité d'images ce qui facilite le traitement automatique et l'étiquetage.
- Intégration d'une structure de données multi-couches ou multi-canaux.

Nos Images sont composées de 5 canaux :

Canal 1 image source en niveaux de gris

Canal 2 image des cellules en énergie fluorescente transmise (FRET).

Canal 3 image des cellules en protéine fluorescente cyan (CFP).

Canal 4 image des cellules en protéine fluorescente jaune (YFP).

Canal 5 image des cellules en FRET+YFP.

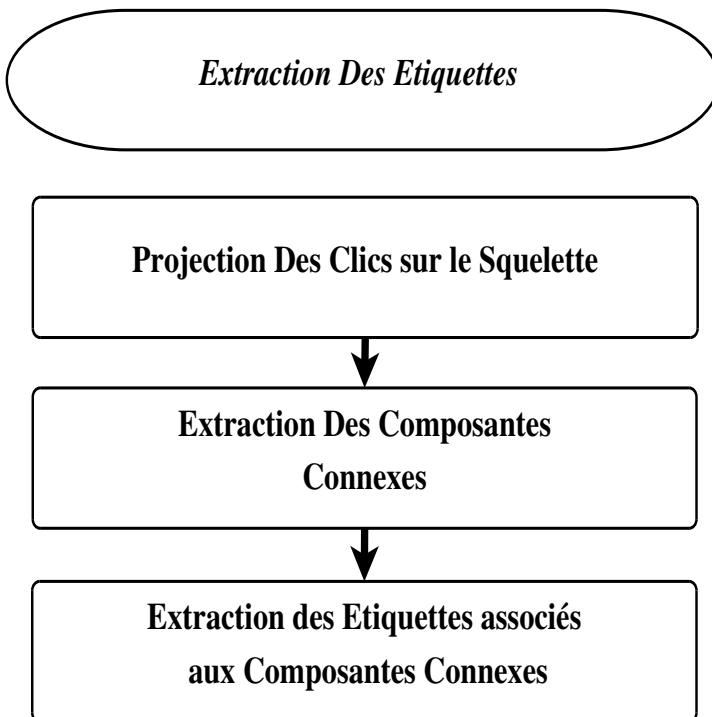


FIGURE 6 – Algorithme : Extraction de la population cible.

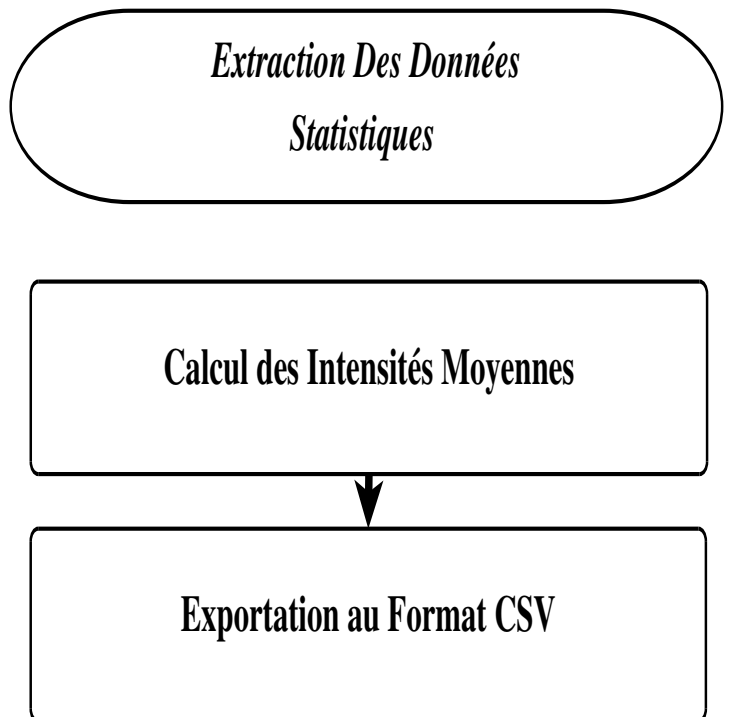


FIGURE 7 – Algorithme : Export des données statistiques de chaque cellule.

Filtrage Morphologique

Après acquisition des images, au premier lieu, on effectue une étape de pré traitement pour préparer les cellules à l'étiquetage.

Cette étape consiste à dé-bruiter l'image en utilisant des opérateurs de morphologie mathématique au niveaux de gris.

Nous avons essayé plusieurs méthodes de filtrages : le filtrage linéaire gaussien et le filtrage morphologique (Top-Hat Filter) nous avons obtenu des images dé-bruités mais on a remarqué une importante perte d'informations au niveau de zones d'intérêt (lignée de cellules) ce qui nous a amené à appliquer un filtrage morphologique basé sur les opérateurs d'ouverture et de la fermeture par reconstruction.

La reconstruction géodésique (voir figure 8) est un filtre morphologique.

L'ouverture par reconstruction élimine les composantes connexes plus petits que l'élément structurant qui n'appartiennent pas à l'ouvert sans modifier les autres.

la fermeture par reconstruction bouche les trous plus petits que l'élément structurant

les ouvertures et fermetures par reconstruction éliminent les structures tout en préservant les contours des cellules ce qui justifie notre choix pour ce filtrage.

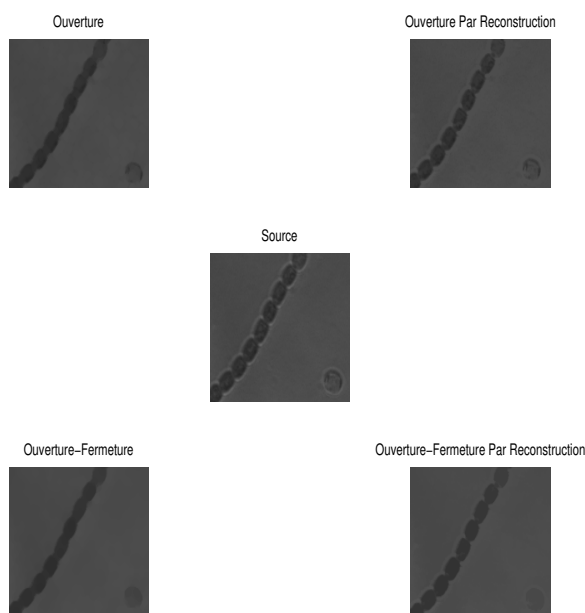


FIGURE 8 – Exemples : Filtrage morphologique avec et sans reconstruction.

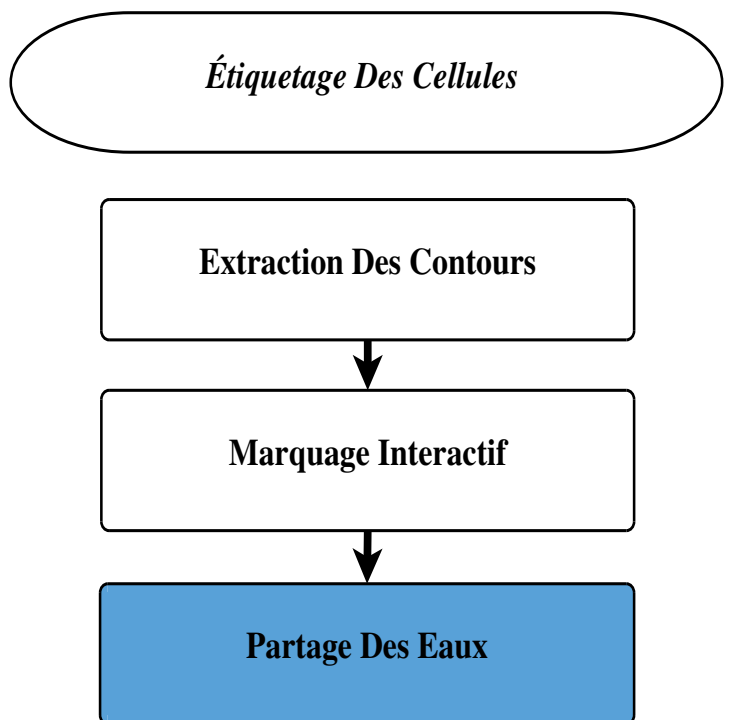


FIGURE 9 – Algorithme : Étiquetage des cellules.

1.2.1 Étiquetage(watershed)

La segmentation des images de cellules de la cyanobactérie *Anabaena* repose sur le fait de pouvoir donner une étiquette à chaque cellule du filament. On a utilisé la fameuse méthode de partage des eaux.

la segmentation par ligne de partage des eaux (watershed [9]) fait partie de la famille de méthodes de segmentation d'images issues de la morphologie mathématique qui considèrent une image à niveaux de gris comme un relief topographique, dont on simule l'inondation.

Avant d'appliquer cette approche, il faut extraire les contours des cellules ce qui nous mène à calculer le gradient de l'image à l'aide du filtre Sobel.

Le principe de ce filtre est que l'opérateur calcule le gradient de l'intensité de chaque pixel. Le gradient indique la direction de la plus forte variation du clair au sombre, ainsi que le taux de changement dans cette direction.

On connaît alors les points de changement soudain de la luminosité, correspondant probablement à des contours, ainsi que l'orientation de ces bords.

Définitions :

- minimum local : c'est le point ou plateau d'où on ne peut pas atteindre un point plus bas sans être obligé de remonter.
- bassin versant : c'est la zone d'influence d'un minimum local. Une goutte d'eau s'écoulant dans le bassin versant arrive au minimum local.
- ligne de partage des eaux : C'est la ligne séparant deux bassins versants de cette ligne, une goutte d'eau peut s'écouler vers au moins 2 minima locaux distincts.

Principe de l'approche Technique de l'immersion (figure 10 et 11) :

- On perce chaque minimum local de la surface.
- On inonde la surface à partir des minima locaux, l'eau montant à vitesse constante et uniforme dans les bassins versants.
- Quand les eaux issues de 2 minima différents se rencontrent, on monte une digue pour qu'elles ne se mélangent pas.
- A la fin de l'immersion, l'ensemble des digues constituent la ligne de partage des eaux.

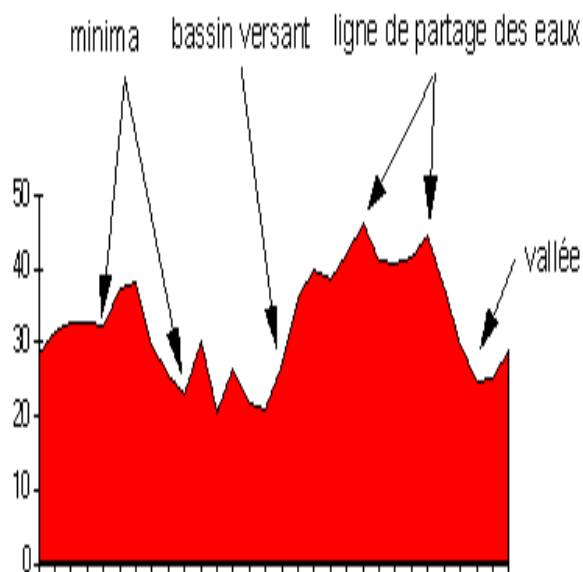


FIGURE 10 – Ligne de partage des eaux : Histogramme.

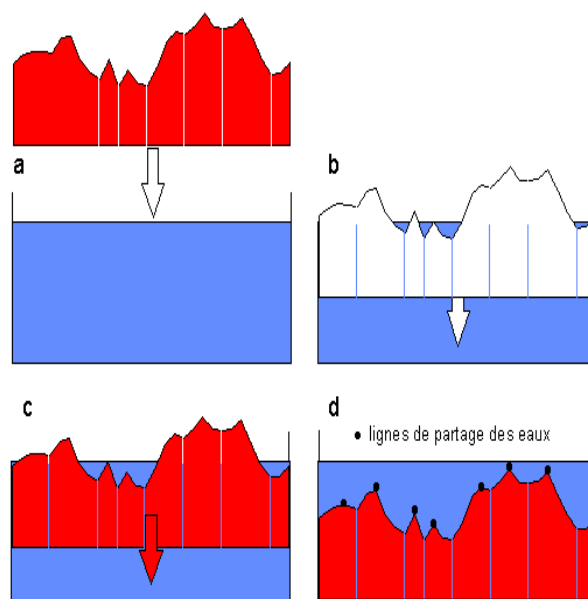


FIGURE 11 – Ligne de partage des eaux : Principe de l'immersion.

Squelettisation

Les cellules de la bactérie sont sous forme de filaments qui accroissent au cours du temps, sur nos images on peut avoir un ou plusieurs filaments au même temps d'où l'idée de calculer le squelette de chaque filament ce qui nous permettra de distinguer les cellules de chaque lignée des autres.

Nous avons remarqué que les images en FRET sont plus adaptés au calcul du squelette, en effet ces images affichent l'aspect général de tout le filament avec moins de détails pour les cellules dans le filament lui même.

La difficulté de cette étape repose sur l'extraction de chaque filament en une composante connexe pour-laquelle nous calculerons le squelette.

Pour ce faire nous avons segmenté l'image avec des opérateurs de morphologie mathématique et nous avons réussi à trouver le format et la taille de l'élément structurant adapté

1.2.2 Extraction Des ROI

Après chargement de l'image l'utilisateur doit interagir pour sélectionner le bout du filament à traiter. Cette opération se manifeste sous forme de deux clics souris :

Premier clic pour définir la position de la cellule départ. Deuxième clic pour définir la position de la cellule de la fin.

À partir de l'image du squelette nous avons extrait le squelette le plus proche de ces deux positions. Sur ce même squelette nous avons cherché les deux points les plus proches de ces deux positions. Nous avons obtenu le bout du squelette délimité par les deux clics. le parcours de ce bout de squelette nous a permis de récupérer la liste des étiquettes en intersection.

1.2.3 Projection des cellules

La projection des cellules suivant un axe horizontale à été décidé pour des raisons esthétiques et de convenance pour faciliter la comparaison hiérarchique des cellules entre deux séquences d'images (voir exemple figure 2).

Comme cette étape ne demande pas une grande précision nous avons effectué une rotation des ROI suivant l'algorithme :

Data: liste des ROI

Result: liste des ROI projeter

liste de ROI ordonnées;

while *toute les régions qui n'ont pas été visitées* **do**

 Lire les propriétés de l'angle α de la région courante ;

 Projeter sur l'axe x par rotation d'angle α en interpolation bilinéaire;

 Redimensionner de l'image projetée;

 Concaténer la nouvelle image;

end

Algorithm 1: Rotation des ROI.

cet algorithme a l'avantage d'être simple et rapide, mais reste beaucoup trop sensible à la grande variation curviligne, ce qui donne des erreurs de projection au niveau de certaines cellules. De plus la projection produit des erreurs d'interpolation, c'est pourquoi cette étape ne sert qu'à fournir à l'utilisateur un rendu visuel lui permettant d'estimer la qualité de la segmentation.

1.2.4 Extraction et sauvegarde des données statistiques

A ce stade, nous avons une liste des étiquettes que chacune représente une cellule (voir figure 2). Chaque étiquette contient toutes les informations concernant l'intensité moyenne, minimale, maximale et aussi la variance de la cellule en deux longueurs d'ondes Cyan ($425 \mu m$) et Yellow ($520 \mu m$).

Ces données sont stockées dans un fichier au format feuille excel (csv) qui sera exploité dans un temps ultérieur pour effectuer une corrélation entre les données d'une séquence et celles des séquences suivantes (figure 4 : une couleur par séquence).

2 Analyse des Résultats



FIGURE 12 – Image : Source.



FIGURE 13 – Image : Réduction du bruit.

Résultats de la segmentation L'image filmée sous microscopie (Figure 12) présente un bruit de fond ce qui complique l'extraction des ROI. Pour mieux mettre en évidence ces ROI et réduire le bruit nous avons effectué un filtrage morphologique (Figure 13) tout en conservant les contours réels des cellules dans le filament, l'extraction de ces contours s'effectuera par la suite en calculant la norme du gradient (Figure 11).

Une difficulté qui se pose dans le passage de l'image de la norme du gradient à l'image binaire (Figure 18) au premier lieu dans l'opération de seuillage et le choix du seuil et en second lieux dans le paramétrage des éléments structurants afin de fermer correctement les contours de chaque cellule.

Nous observons au niveau du filament certaines cellules en étape de division et la décision de considérer si ces cellules seront étiquetées en une ou deux étiquettes dépendra de la taille de l'élément structurant.

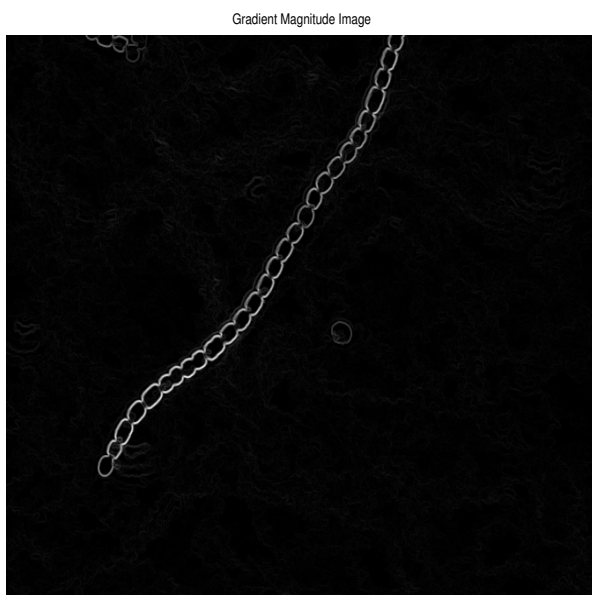


FIGURE 14 – Image : Contours extrait par gradient.

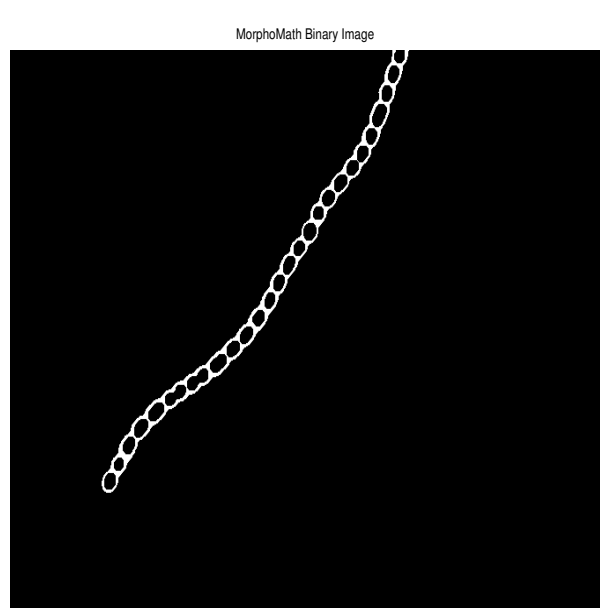


FIGURE 15 – Image : Contours binarisés.

À partir de l'image étiquetée (Figure 14) nous obtiendrons une image en espace RGB (Figure 15) et une couleur différente sera affectée à chaque étiquette.

Comme nous avons expliqué au chapitre 1.2.1, dans un premier temps nous avons essayé de calculer le squelette à partir de l'image source mais même avec un filtrage morphologique (ouverture et fermeture sans reconstruction pour éliminer les contours dans le filament) nous n'avons pas obtenu de résultats satisfaisants (squelettes incorrects ou discontinus). Comme remède à ces difficultés nous avons choisi de calculer le squelette à partir de l'image au canal associé au FRET après avoir constaté que ces images représentent tout le filament comme une seule composante connexe et les contours des cellules dans le filament lui-même sont négligeables. Avec ce canal nous avons obtenu un squelette correct (Figure 16). La (figure 17) illustre les deux clics utilisateurs projetés sur le squelette (en bleu) et la section du squelette (en vert) le tout superposé sur les étiquettes des cellules dans le filament.

A partir de la section du squelette et la liste de toutes les étiquettes nous avons pu extraire que les étiquettes concernés (Figure 19) chaque étiquette extraite est coloré différemment et directement superposée sur la cellule concernée du filament dans l'image source (Figure 20).

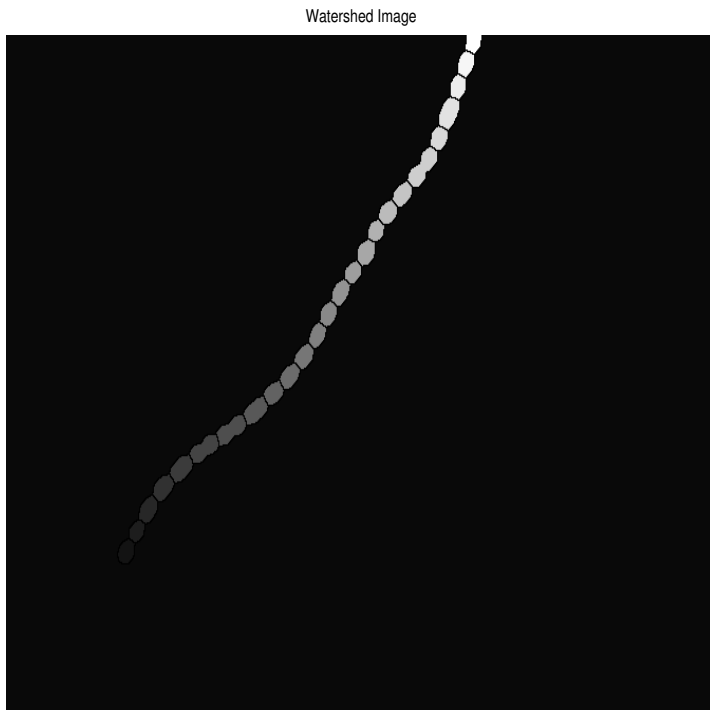


FIGURE 16 – Image : Étiquetage par Watershed.

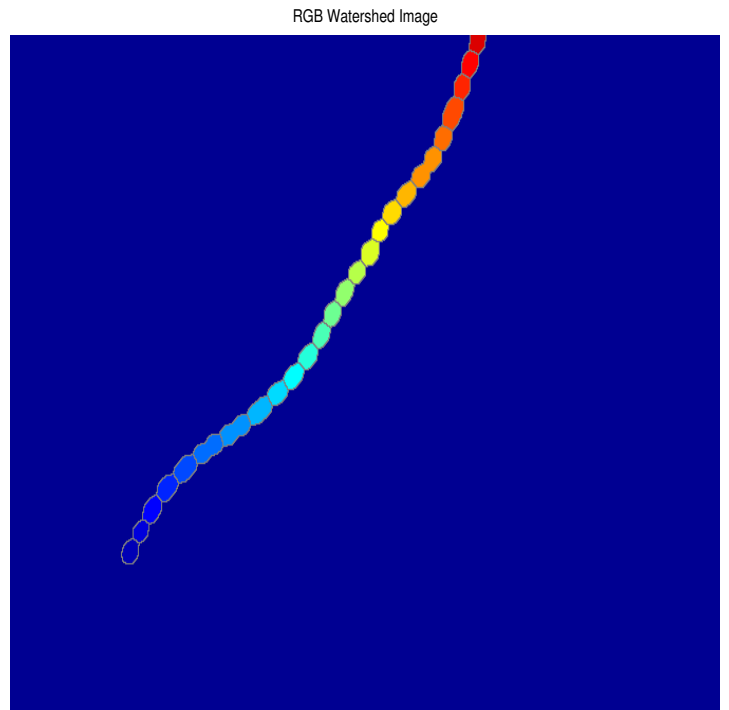


FIGURE 17 – Image : Étiquetage en couleurs.

Résultat d'extraction des données des cellules

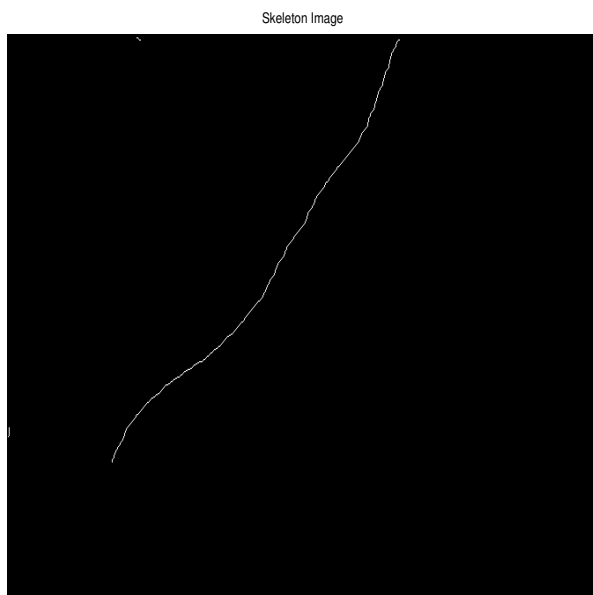


FIGURE 18 – Images : Squelette.

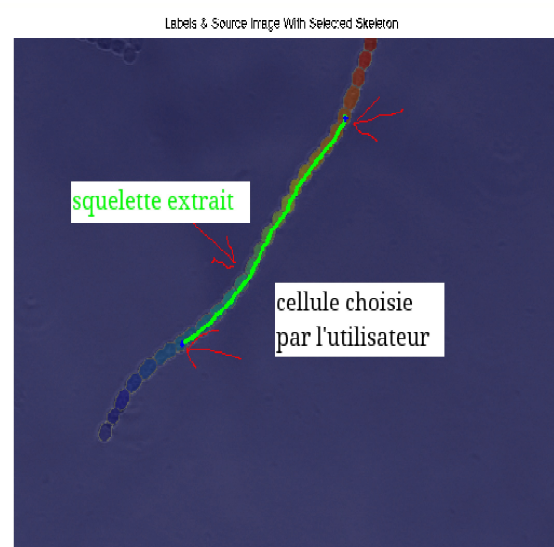


FIGURE 19 – Image : Section à extraire.

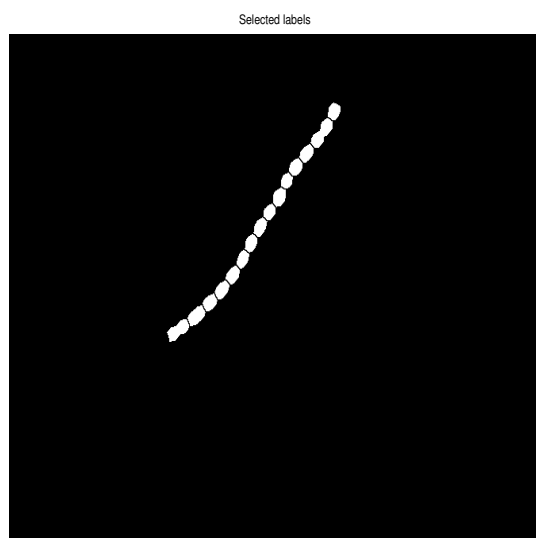


FIGURE 20 – Image : Étiquettes sélectionnées.

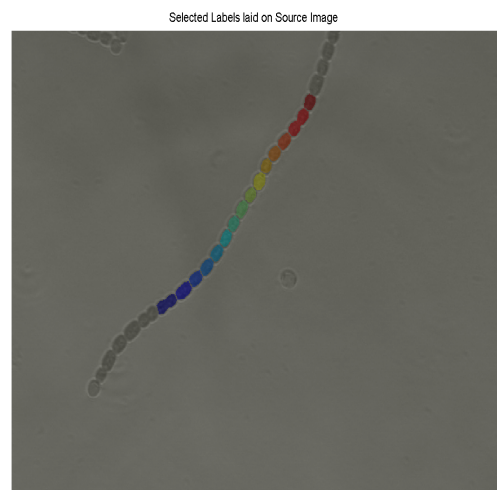


FIGURE 21 – Image : Section à extraire superposé sur l'image source.

Conclusion

Nous avons pu remarquer qu'une méthode aussi efficace que la ligne de partage des eaux peut être adaptée pour segmenter des images micro-biologiques, mais se retrouver contraint par la détermination du marqueur de forme le plus générique impossible.

La question qui se pose naturellement est le problème de l'amélioration des marqueurs de région de départ. Il résultera qu'à l'avenir nous garderons un esprit critique face aux résultats obtenus, d'autant plus qu'il est difficile de comprendre certains aspects du problème avec un échantillon de test. Établir ce qu'est une bonne segmentation ne peut pas se faire sans l'expertise des futurs utilisateurs, malgré les limites en temps et en complexité d'un projet.

À titre de conclusion, il semble intéressant de mettre en évidence les questions actuelles qui se posent sur l'avenir de l'imagerie numérique, dans le monde de la micro biologie et de savoir comment les futurs acteurs que nous sommes feront pour porter à un autre niveau toutes les problématiques qu'elles soulèvent.

Table des figures

1	filament avant(A) et après(B) privation en nitrogène.	7
2	Filament étiqueté.	7
3	variation du niveau de 2-OG dans un cellule Hétérocyste.	8
4	cellules étiquetées avec leurs YFP CFP et FRET exportés dans un fichier csv.	8
5	Algorithme : Processus général.	9
6	Algorithme : Extraction de la population cible.	10
7	Algorithme : Export des données statistiques de chaque cellule.	10
8	Exemples : Filtrage morphologique avec et sans reconstruction.	11
9	Algorithme : Étiquetage des cellules.	11
10	Ligne de partage des eaux : Histogramme.	12
11	Ligne de partage des eaux : Principe de l'immersion.	12
12	Image : Source.	15
13	Image : Réduction du bruit.	15
14	Image : Contours extrait par gradient.	15
15	Image : Contours binarisés.	15
16	Image : Étiquetage par Watershed.	16
17	Image : Étiquetage en couleurs.	16
18	Images : Squelette.	17
19	Image : Section à extraire.	17
20	Image : Étiquettes sélectionnées.	17
21	Image : Section à extraire superposé sur l'image source.	17