



UNIVERSITÉ DE CAEN

PROJET ANNUEL

Hooke

Quantification spatio-temporelle de cellules

Maxence Busson

tutoré par
Youssef CHAHIR

22 septembre 2016

Résumé

Le projet avait pour but d'analyser des images de cellules en mouvement et d'en extraire leur nombre. Par la suite, il fallait suivre ces cellules tout au long de leur déplacement. Ce logiciel devrait à terme remplacer un logiciel déjà existant qui ne répond qu'imparfaitement aux besoins.

J'ai donc créé un logiciel reprenant en partie les fonctionnalités antérieures avec des améliorations : chargement des fichiers, débruitages d'images...

Cette expérience m'a permis de confronter mes connaissances théoriques avec la réalité du terrain. Je me suis familiarisé avec l'utilisation du logiciel Qt et de la bibliothèque OpenCV. J'ai appris à gérer des problèmes informatiques variés (bruitage, mouvement ...).

Plus généralement, j'ai bénéficié d'une expérience enrichissante permettant d'appréhender concrètement le monde de l'imagerie et, au-delà de l'objectif initial de ce projet, j'ai eu l'occasion d'élargir mon horizon culturel en découvrant l'imagerie médicale.

Mots-clés : Informatique - Qt - Hooke - Imagerie médicale

Table des matières

1	Introduction	2
2	Problématique et étude de l'existant	3
2.1	Problématique	3
2.2	Etude de l'existant	3
2.2.1	L'existant	3
2.2.2	Inconvénients	4
3	Conception	5
3.1	Logiciel utilisé	5
3.2	Projet	5
3.2.1	Corpus d'images	5
3.2.2	Méthode utilisée	7
3.2.3	Finalité	8
4	Difficultés rencontrées	9
5	Compétences requises, acquises et Conclusion	10

Chapitre 1

Introduction

Dans le cadre du master 2 IMALANG, le programme prévoit un projet annuel sur divers sujets. Celui-ci a pour objectifs d'imposer une problématique et de développer des solutions concrètes, de confronter ses acquis théoriques avec les besoins de terrain, de développer un esprit de réflexion, d'analyse et de synthèse qui devra se concrétiser à travers la rédaction du rapport de stage.

Pour ma part, j'ai choisi le sujet : "Quantification spatio-temporelle de cellules". Ce projet fut proposé par Youssef Chahir. Ce projet a pour but de compter et de suivre des cellules dans leur déplacement via une succession d'images acquises grâce à un microscope.

Dans ce rapport, je m'attacherai tout d'abord à la présentation de la problématique et de l'existant. puis à la description de la base d'images. On s'attachera ensuite à l'observation des outils utilisés, le logiciel final, les problèmes rencontrés et les améliorations futures. Avant de conclure, je consacrerai un chapitre aux compétences requises et acquises lors de ce projet.

Chapitre 2

Problématique et étude de l'existant

2.1 Problématique

Lorsque les biologistes doivent observer des images de cellules, ils cherchent à extraire des cellules. Dans ces images, il faut donc faire le tri entre les cellules et les débris, les cellules peuvent aussi se superposer. Ce travail peut très vite devenir fastidieux, il est donc essentiel de fournir un outil de traitement automatique aux biologistes. Le projet a pour but de fournir cet outil.

Il doit compter et traquer les cellules sur ces images. A terme, il devra catégoriser chaque cellule et pourquoi pas déterminer leur état (cellule morte, cancéreuse, saine...) et fournir des informations sur le temps (divisions/ destruction cellulaire).

2.2 Etude de l'existant

2.2.1 L'existant

Au début du projet on m'a confié les codes sources d'un logiciel déjà existant. Ce programme fut développé par Mme. Helen Maméri alors étudiante en master DNR2I et amélioré par M. Youssef Chahir, enseignant chercheur à l'Université de Caen. Le logiciel est une boîte à outils qui contenait les options suivantes :

- Charger une image
- Créer l'histogramme
- Créer l'image en niveaux de gris
- Créer l'image en Teinte Saturation Valeur (HSV)
- Créer l'image binarisée
- Créer les contours de formes

Puis s'en suit les options pour la morphologie :

- Erosion
- Dilatation
- Ouverture
- Fermeture
- Ligne de partage des eaux (watershed)

Pour finir les dernières options se concentre sur l'analyse :

- K-Means
- Composantes connexes
- Analyse

De ce produit en découle un logiciel de base efficace pour le traitement d'images. Cependant, notre but étant de reconnaître la forme d'une cellule, de la compter et de la traquer nous n'avons pas besoin de tous ces outils.

Les tests effectués sur ce logiciel nous ont permis de conclure que le programme permet de reconnaître une cellule seulement sur un corpus d'images non-réaliste (dessin d'artiste ou image au microscope modifiée). De ce fait, les problématiques liés aux débris, aux cellules superposées ou autres phénomènes posant un quelconque problème d'analyse n'était pas traité.

2.2.2 Inconvénients

Le logiciel fournit par M. Chahir et Mme. Maméri est une bonne base de commencement de projet étant donné le travail déjà fournit. Cependant, il ne répond pas aux demandes de notre projet quand on se place dans un cadre réaliste. De ce fait nous pouvons reprendre certains outils et améliorer les analyses afin de prendre en compte les biais. En recréant un nouveau logiciel inspiré de celui-ci nous pourrions alors améliorer grandement les algorithmes et enlever le superflus.

Chapitre 3

Conception

3.1 Logiciel utilisé

Le projet précédent avait utilisé un logiciel en particulier, Qt Creator. Ce logiciel est un environnement de développement intégré (IDE) faisant partie du framework Qt. C'est un outils orienté pour la programmation en C++.

Etant donné que le code était déjà fait sous Qt Creator et que cet outils est extrêmement pratique pour développer rapidement une interface graphique efficace j'ai repris ce logiciel ainsi que le C++ comme langage. De plus, C++ étant un langage que je maîtrisais déjà, c'était une aubaine pour moi.

3.2 Projet

Lors du commencement du projet j'ai décidé de recréer un programme fortement inspiré de l'ancien. Je nomma alors ce programme Hooke en hommage à Robert Hooke grand scientifique tout particulièrement en biologie où on lui doit la première description d'une cellule biologique faite à partir de l'observation de végétaux. Le ton était alors donné pour notre application, l'analyse de cellule.

3.2.1 Corpus d'images

Pour qu'une analyse soit efficace et donc pour que l'algorithme qui analyse soit efficace il faut tester sur des cas concrets. Les images sur lesquels Mme. Maméri faisait les tests étaient des illustrations et non pas des images de cellules au microscope. De ce fait, les résultats en étaient influencés, si bien que les algorithmes ne fonctionnaient pas sur des images réalistes.

Il nous fallait alors trouver une base d'images pour nous aider dans nos tests. Cependant, nous avions une contrainte supplémentaire. Etant donné que nous faisons du tracking de cellule (analyse de mouvement d'une cellule) il nous fallait absolument une vidéo et non une image. C'est là où les problèmes ont commencé à émerger. On m'avait bien fournit une petite vidéo de quelques images permettant de commencer les analyses. Le problème était que cette vidéo était fortement bruitée et l'analyse des cellules était quasiment impossible. En effet, le fond et les cellules se fusionnaient par moment ce qui rendait difficile de faire la

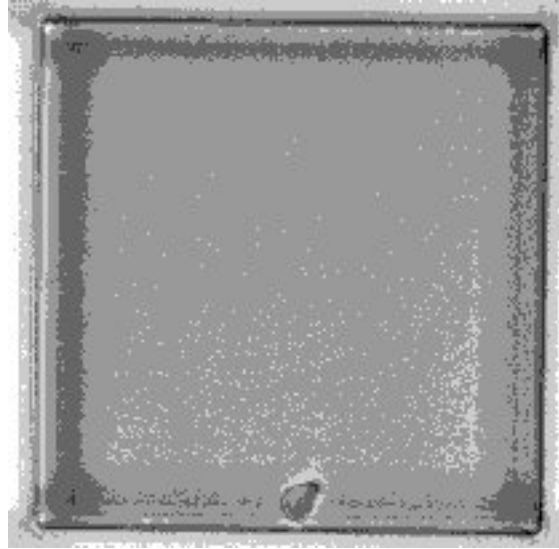


FIGURE 3.1 – Image de l’ancien corpus fortement bruitée.

différence entre les deux. La première partie de mon projet fut donc d’essayer de débruiter cette vidéo afin d’y extraire les cellules, les compter et les traquer.

Malgré tous mes efforts de recherche et d’algorithmie je ne pu trouver une solution à ce bruit. J’entama alors une deuxième étape, celle de rechercher un corpus de vidéo de cellules en mouvement et de bonne qualité. Pour se faire, je me tourna d’abords vers le CHU de Caen. Je leur demanda de bien vouloir me fournir quelques vidéos de cellules pour les besoins de ma recherche et de mon projet. J’ai eu la forte déception de ne recevoir que des réponses négatives à mes demandes. Il faut bien comprendre une chose, il est plus aisé de trouver des images de cellules que des vidéos. Les bases de données d’images que l’on trouve sur le Web nous permette de trouver des images mais, pas des vidéos. Je me suis donc tourné vers les concours qu’organise certaines organisation. Le but du concours est de fournir un algorithme pour analyser les cellules dans certains cas spécifiques. Ils mettent alors une base d’images ou de vidéos de cellules et c’est celà qui m’intéressé. Suite à une inscription et un échange de mail avec les organisateur leur faisant part de mes revendication et de ma non participation au concours. Ils refusèrent de m’accorder le droit d’utiliser leur base d’image et me souhaitèrent bonne chance.

Devant cette problématique je décida de changer le cahier des charges. Ainsi donc nous nous limiterions donc à la simple analyse de cellule et comptage, laissant de côté le tracking. Ce changement me permettait alors de ne pas travailler sur une vidéo mais, sur une image. Ainsi je contournais le problème avec un grand regret. Le tracking de cellule est possible et est simple à réaliser, cependant à cause d’un corpus de vidéo vide nous ne pouvons faire les analyses. J’ai songé pendant un moment à prendre une image et à me déplacer dedans afin de faire une pseudo-vidéo. Cependant, cette astuce n’a pour effet que de produire des déplacement de caméra et non de cellules ce qui induit des translations quasi-constantes et donc une mauvaise analyse à terme.

3.2.2 Méthode utilisée

Avant d'aborder les méthodes utilisées pour reconnaître les cellules et les compter, revenons sur les enjeux. Nous voulons extraire d'une photo des cellules, dans notre cas ce seront des cellules banales, c'est-à-dire plus ou moins ronde avec un noyau à leur centre. Il faut donc séparer ces cellules du fond (background) et des débris. Dans certains cas les cellules se superposent ce qui complique la tâche pour les différencier. Ensuite il faut les compter et donc leur affilier un label.

Première étape enlever le fond. Pour se faire nous utilisons un principe assez simple fondé sur le fait qu'une cellule ressort par rapport au fond (ce qui n'était pas le cas avec le précédent corpus), grâce à ce principe nous utilisons une binarization de la photo. La binarization va faire en sorte que les couleurs qui se rapprochent de la couleur du fond vont y être assimilés et celles se rapprochant des cellules se rapprocheront des cellules. Ainsi on sépare en deux nos photos.

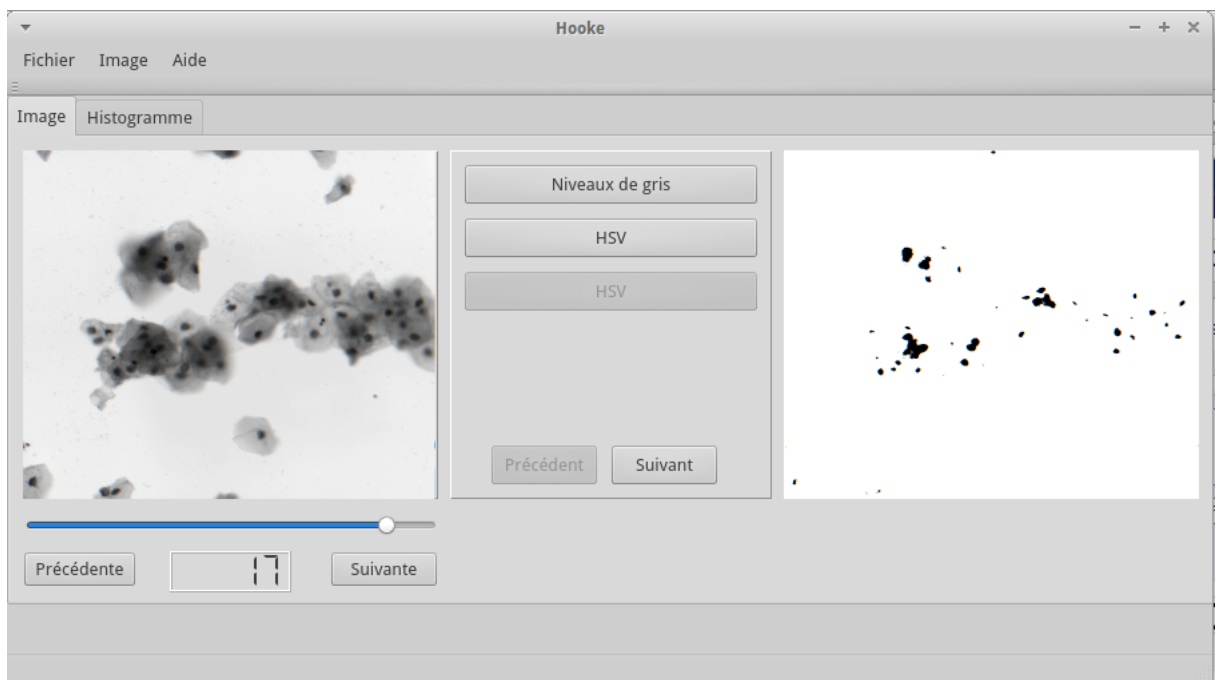


FIGURE 3.2 – Erosion et binarization de l'image pour y trouver les noyaux. On remarque aussi à gauche la superposition cas extrême.

Par la suite on va repérer les cellules, comme vu précédemment il peut y avoir des débris ou des cellules superposées. Pour éviter ces problèmes on va appliquer une érosion afin de n'avoir que le noyau des cellules. Ainsi les débris sont enlevés et les cellules superposées sont distinctes. Il n'y a qu'un seul cas où deux cellules ne peuvent pas être vu distinctement, c'est le cas où les noyaux se "touchent" sur la photo. Dans ce cas notre algorithme ne verra qu'une seule et même cellule étant donné qu'on se base sur les noyaux. Une fois cette érosion faite on compte le nombre de noyaux et on récupère l'image d'origine en appliquant ce qu'on a trouvé.

3.2.3 Finalité

Notre logiciel ne réponds pas à la contrainte d'origine qui était de traquer et de compter des cellules dans une vidéo. Cela ne dépend pas de la difficulté algorithmique mais, bien de la difficulté à trouver un corpus de données réaliste. Dans le cas de photo et de comptage de cellule notre algorithme répond en grande partie aux demandes. Il n'y a que dans les cas extrêmes (superpositions de noyaux) où les résultats ne seront pas performants.

Dans l'avenir avec une meilleur corpus (qualité des images, canaux différents autre que la couleur) on pourrait déterminer si une cellule est cancéreuse ou pas. Dans la théorie une cellule cancéreuse si sur la spectre de la luminosité elle apparaît plus sombre que des cellules saines. Ce spectre peut être donné dans l'image via un canal spécial. Ainsi ce logiciel pourrait aider les médecins à traiter plus rapidement les images de cellules pour y extraire une analyse.

Chapitre 4

Difficultés rencontrées

Lors de ce projet j'ai eu beaucoup de mal non pas sur l'algorithme, mais bien sur le fait de trouver un corpus de données viable. Le domaine médicale est un milieu assez fermé dû aux compétitions entre les laboratoires biologiques ce qui cause des tords pour la recherche dites publique ou encore comme moi dans le but de faire un simple projet sans financement.

Ce problème nous a obligé à complètement changer de cahier des charges original. Malgré cela nous avons pu nous arranger pour que seul la partis "tracking" soit retirée.

Chapitre 5

Compétences requises, acquises et Conclusion

Ce projet a requis des connaissances en C++ et en imagerie. C'était pour moi un premier projet concret dans l'imagerie ce qui m'a beaucoup plu.

Cela m'a apporté de grande connaissances en image, en Qt Creator et surtout dans le milieu médical. Un milieu passionnant et qui m'a fait apprendre beaucoup de chose dans la cancérologie notamment.

Ce fut une agréable expérience et malgré certains aspect qui ma chagrine et que je n'ai pu aborder j'ai envie de peaufiner au mieux ce logiciel prometteur.