

**Titre:** Segmentation automatique des images de fibres d'ADN pour la  
Title: quantification du stress réplicatif

**Auteur:** Pierre Ghesquiere  
Author:

**Date:** 2018

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Ghesquiere, P. (2018). Segmentation automatique des images de fibres d'ADN  
Citation: pour la quantification du stress réplicatif [Mémoire de maîtrise, École  
Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/3739/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/3739/>  
PolyPublie URL:

**Directeurs de  
recherche:** Farida Cheriet, & Santiago Costantino  
Advisors:

**Programme:** Génie biomédical  
Program:

UNIVERSITÉ DE MONTRÉAL

SEGMENTATION AUTOMATIQUE DES IMAGES DE FIBRES D'ADN POUR LA  
QUANTIFICATION DU STRESS RÉPLICATIF

PIERRE GHESQUIERE

INSTITUT DE GÉNIE BIOMÉDICAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE BIOMÉDICAL)

DÉCEMBRE 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

~~SEGMENTATION~~ AUTOMATIQUE DES IMAGES DE FIBRES D'ADN POUR LA  
QUANTIFICATION DU STRESS RÉPLICATIF

présenté par : Ghesquiere Pierre

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. LESAGE Frédéric, Ph. D., président

Mme CHERIET Farida, Ph. D., membre et directrice de recherche

M. COSTANTINO Santiago, Ph. D., membre et codirecteur de recherche

M. WURTELE Hugo, Ph. D., membre

## DÉDICACE

*À toi,*

## REMERCIEMENTS

Tout d'abord, merci à tous ceux que je n'aurais pas dû oublier dans cette maladroite liste de remerciements.

Merci beaucoup à Santiago Costantino, mon co-directeur de recherche pour son accueil, sa confiance, ses encouragements indispensables pendant mes périodes de découragement, ses conseils essentiels, sa sincérité, son humour, son énergie et les soirées jeux de société.

Merci à Farida Chériet, ma directrice de recherche pour son accueil très chaleureux avec Fantin Girard, pour ses conseils précieux et intelligents, pour sa capacité à mettre les gens à l'aise, à booster la motivation, et pour les couscous.

Merci à Joannie Roy pour son accueil dès le tout premier jour, son attention naturelle à ce que tout le monde se sentent bien dans le laboratoire, son exemple de rigueur et la soirée "Reine des Neiges" avec Roxanne, Mira et François.

Merci à Diane Sayah pour ses conseils précieux, pour les pauses quotidiennes nécessaires, pour tout le temps passé à m'expliquer les différentes pathologies oculaires, pour nos taquineries indispensables et stimulantes.

Merci à Javier Mazzaferri pour tout le temps passé à débbugger mon code, pour ses engagements multiples et exemplaires, pour m'avoir fait venir à la chorale, pour son courage.

Merci à Abdelhamid Elsherbiny, Emile Fortier, Mary McQuaid, Hugo Wurtele, Elliot Drobetsky pour leurs conseils précieux, leur soutien. Merci de m'avoir partagé votre passion pour une branche de la biologie que j'ignorais jusqu'alors.

Merci à Mira Jabbour, la meilleure voisine de bureau, pour son intelligence, son écoute, son humour et les moments à discuter tous les deux des derniers potins.

Merci à Loïc Binan pour son accueil, sa sympathie, sa grandeur, son chaleureux charisme naturel.

Merci à Charles Bélanger-Nzakimuena pour la séance de relaxation, pour les discussions passionnées sur le cerveau, l'Amour et l'hypnose.

Merci à Nicolas Desjardins-Lecavalier pour les formidables soirées musicales, sa bonne humeur et son remarquable accent russe.

Merci à Maxime Gagnon pour sa fidélité, sa bonne humeur et sa sincérité.

Merci à Gabriel Lepetit-Aimon, pour les soirées à m'expliquer ou me ré-expliquer les principes

de base des réseaux de neurones, le mini-hackathon pour segmenter les brins d'ADN, pour son amitié, ses beautés multiples et pardon pour mes accès d'orgueil.

Merci à Clément Playout pour les discussions enflammées sur Kundera, la gauche mélanchonienne, le véganisme et tout le reste.

Merci à tous mes voisins du département d'ophtalmologie de l'Hôpital Maisonneuve-Rosemont pour leur chaleur humaine.

Merci à toute l'équipe de recherche du LIV4D et particulièrement à Philippe Debanné pour son efficacité et sa grande gentillesse.

## RÉSUMÉ

La réplication de l'ADN est un processus complexe géré par une multitude d'interactions moléculaires permettant une transmission précise de l'information génétique de la cellule mère vers les cellules filles. Parmi les facteurs pouvant porter atteinte à la fidélité de ce processus, on trouve le *Stress Réplicatif*. Il s'agit de l'ensemble des phénomènes entraînant le ralentissement voire l'arrêt anormal des fourches de réplication. S'il n'est pas maîtrisé, le stress réplcatif peut causer des ruptures du double brin d'ADN ce qui peut avoir des conséquences graves sur la stabilité du génome, la survie de la cellule et conduire au développement de cancers, de maladies neurodégénératives ou d'autres troubles du développement.

Il existe plusieurs techniques d'imagerie de l'ADN par fluorescence permettant l'évaluation de la progression des fourches de réplication au niveau moléculaire. Ces techniques reposent sur l'incorporation d'analogues de nucléotides tels que chloro- (CldU), iodo- (IdU), ou bromo-deoxyuridine (BrdU) dans le double brin en cours de réplication. L'expérience la plus classique repose sur l'incorporation successive de deux types d'analogues de nucléotides (IdU et CldU) dans le milieu cellulaire. Une fois ces nucléotides exogènes intégrés dans le double brin de l'ADN répliqué, on lyse les cellules et on répartit l'ADN sur une lame de microscope. Les brins contenant les nucléotides exogènes peuvent être imagés par immunofluorescence. L'image obtenue est constituée de deux couleurs correspondant à chacun des deux types d'analogues de nucléotides. La mesure des longueurs de chaque section fluorescente permet la quantification de la vitesse de progression des fourches de réplication et donc l'évaluation des effets du stress réplcatif.

La mesure de la longueur des fibres fluorescentes d'ADN est généralement réalisée manuellement. Cette opération, en plus d'être longue et fastidieuse, peut être sujette à des variations inter- et intra- opérateurs provenant principalement de différences dans le choix des fibres. La détection des fibres d'ADN est difficile car ces dernières sont souvent fragmentées en plusieurs morceaux espacés et peuvent s'enchevêtrer en agrégats. De plus, les fibres sont parfois difficiles à distinguer du bruit en arrière plan causé par les liaisons non-spécifiques des anticorps fluorescents.

Malgré la profusion des algorithmes de segmentation de structures curvilignes (vaisseaux sanguins, réseaux neuronaux, routes, fissures sur béton...), très peu de travaux sont dédiés au traitement des images de fibres d'ADN. Nous avons mis au point un algorithme intitulé ADFA (Automated DNA Fiber Analysis) permettant la segmentation automatique des fibres d'ADN ainsi que la mesure de leur longueur respective. Cet algorithme se divise en trois parties :

- (i) Une extraction des objets de l'image par analyse des contours. Notre méthode de segmentation des contours se basera sur des techniques classiques d'analyse du gradient de l'image (Marr-Hildreth et de Canny).
- (ii) Un prolongement des objets adjacents afin de fusionner les fibres fragmentées. Nous avons développé une méthode de suivi (tracking) basée sur l'orientation et la continuité des objets adjacents.
- (iii) Une détermination du type d'analogue de nucléotide par comparaison des couleurs. Pour ce faire, nous analyserons les deux canaux (vert et rouge) de l'image le long de chaque fibre.

Notre algorithme a été testé sur un grand nombre d'images de qualité variable et acquises à partir de différents contextes de stress réplicatif. La comparaison entre ADFA et plusieurs opérateurs humains montre une forte adéquation entre les deux approches à la fois à l'échelle de chaque fibre et à l'échelle plus globale de l'image. La comparaison d'échantillons soumis ou non soumis à un stress réplicatif a aussi permis de valider les performances de notre algorithme. Enfin, nous avons étudié l'impact du temps d'incubation du second analogue de nucléotide sur les résultats de l'algorithme.

Notre algorithme est particulièrement efficace sur des images contenant des fibres d'ADN relativement courtes et peu fractionnées. En revanche, notre méthode de suivi montre des limites lorsqu'il s'agit de fusionner correctement de longues fibres fortement fragmentées et superposées à d'autres brins. Afin d'optimiser les performances d'ADFA, nous recommandons des temps d'incubation courts (20 à 30 minutes) pour chaque analogue de nucléotide dans le but d'obtenir des fibres courtes. Nous recommandons aussi de favoriser la dilution des brins sur la lame de microscope afin d'éviter la formation d'agrégats de fibres difficiles à distinguer.

ADFA est disponible en libre accès et a pour vocation de servir de référence pour la mesure des brins d'ADN afin de pallier les problèmes de variabilités inter-opérateurs.



## ABSTRACT

DNA replication is tightly regulated by a great number of molecular interactions that ensure accurate transmission of genetic information to daughter cells. *Replicative Stress* refers to all the processes undermining the fidelity of DNA replication by slowing down or stalling DNA replication forks. Indeed, stalled replication forks may “collapse” into highly-genotoxic double strand breaks (DSB) which engender chromosomal rearrangements and genomic instability. Thus, replicative stress can constitute a critical determinant in both cancer development and treatment. Replicative stress is also implicated in the molecular pathogenesis of aging and neurodegenerative disease, as well as developmental disorders.

Several fluorescence imaging techniques enable the evaluation of replication forks progression at the level of individual DNA molecules. Those techniques rely on the incorporation of exogene nucleotide analogs in nascent DNA at replication forks in living cells. In a typical experiment, sequential incorporation of two nucleotide analogs, e.g., IdU and CldU, is performed. Following cell lysis and spreading of DNA on microscopy slides, DNA molecules are then imaged by immunofluorescence. The obtained image is made up of two colors corresponding to each one of the two nucleotide analogs. Measurement of the respective lengths of these labeled stretches of DNA permits quantification of replication fork progression.

Evaluation of DNA fiber length is generally performed manually. This procedure is laborious and subject to inter- and intra-user variability stemming in part from unintended bias in the choice of fibers to be measured. DNA fiber extraction is difficult because strands are often fragmented in lots of subparts and can be tangled in clusters. Moreover, the extraction of fibers can be difficult when the background is noised by non specific staining.

Despite the large number of segmentation algorithms dedicated to curvilinear structures (blood vessels, neural networks, roads, concrete tracks...), few studies address the treatment of DNA fiber images. We developed an algorithm called ADFA (Automated DNA Fiber Analysis) which automatically segments DNA fibers and measures their respective length. Our approach can be divided into three parts:

1. Object extraction by a robust contour detection. Our contour segmentation method relies on two classical gradient analyses (Marr and Hildreth, 1980; Canny, 1986)
2. Fusion of adjacent fragmented fibers by analysing their continuity. We developed a tracking approach based on the orientation and the continuity of adjacent fibers.

3. Detection of the nucleotide analog label (IdU or CldU). To do so, we analyse the color profile on both channels (green and red) along each fiber.

ADFA was tested on a database of different images of varying quality, signal to noise ratio, or fiber length which were acquired from two different microscopes. The comparison between ADFA and manual segmentations shows a high correlation both at the scale of the fiber and at the scale of the image. Moreover, we validate our algorithm by comparing samples submitted to replicative stress and controls. Finally, we studied the impact of the incubation time of the second nucleotide analog pulse.

The performances of our algorithm are optimised for images containing relatively short and not fragmented DNA fibers. Our tracking methods may be limited when connecting highly split fibers superimposed to other strands. Therefore, we recommend to reduce the incubation time of each nucleotide analog to about 20-30 minutes in order to obtain short fibers. We also recommend to foster the dilution of fibers on the slide to reduce clustering of fluorescent DNA molecules.

ADFA is freely available as an open-source software. It might be used as a reference tool to solve inter-intra user variability.

## TABLE DES MATIÈRES

DÉDICACE . . . . .	iii
REMERCIEMENTS . . . . .	iv
RÉSUMÉ . . . . .	vi
ABSTRACT . . . . .	viii
TABLE DES MATIÈRES . . . . .	x
LISTE DES TABLEAUX . . . . .	xii
LISTE DES FIGURES . . . . .	xiii
LISTE DES ANNEXES . . . . .	xvii
CHAPITRE 1 INTRODUCTION . . . . .	1
1.1 Structure de l'ADN . . . . .	1
1.2 Réplication de l'ADN . . . . .	2
1.3 Notion de Stress Réplicatif . . . . .	5
1.4 Structure du mémoire . . . . .	7
CHAPITRE 2 REVUE DE LITTÉRATURE . . . . .	9
2.1 Imagerie de l'ADN pour la détection du stress réplicatif . . . . .	9
2.1.1 Étalement des fibres d'ADN (DNA spreading) . . . . .	9
2.1.2 Peignage de l'ADN (DNA combing) . . . . .	11
2.2 Segmentation de structures curvilignes . . . . .	12
2.2.1 Définition des Structures Curvilignes . . . . .	12
2.2.2 Détection de Structures Curvilignes par Analyse du Gradient de l'Image	13
2.2.3 Détection de Structures Curvilignes par Analyse de la Hessienne de l'image . . . . .	17
2.2.4 Détection de Structures Curvilignes par Vote de Tenseur (Tensor Voting)	18
2.2.5 Détection de Structures Curvilignes par Score d'Orientatation . . . . .	19
2.2.6 Détection de Structures Curvilignes par l'algorithme du chemin minimal	19
2.2.7 Détection de Structures Curvilignes par Opérations Morphologiques .	20
2.2.8 Détection de Structures Curvilignes par Apprentissage Machine . . . .	22

2.2.9	Discussion . . . . .	23
2.3	Objectifs de recherche . . . . .	26
CHAPITRE 3 ARTICLE 1 : AN OPEN-SOURCE IMPLEMENTATION FOR FULLY AUTOMATED QUANTIFICATION OF DNA FIBERS . . . . .		27
3.1	Introduction . . . . .	27
3.2	Results . . . . .	29
3.2.1	Inter-user variability upon manual quantification of DNA fibers on im- munofluorescence images . . . . .	29
3.2.2	Comparing ADFA vs manual quantifications . . . . .	30
3.2.3	Performance of ADFA in biologically-relevant experimental conditions	32
3.3	Materials and Methods . . . . .	36
3.3.1	DNA Fiber- Assay . . . . .	36
3.3.2	ADFA algorithm: method . . . . .	36
3.4	Discussion . . . . .	45
CHAPITRE 4 DISCUSSION GÉNÉRALE . . . . .		47
CHAPITRE 5 CONCLUSION ET RECOMMANDATIONS . . . . .		50
5.1	Synthèse des travaux . . . . .	50
5.2	Améliorations futures . . . . .	50
RÉFÉRENCES . . . . .		51
ANNEXES . . . . .		64

# LISTE DES TABLEAUX

Table 3.1	Inter-user variability: manual quantification of CldU- and IdU-labeled DNA fibers by three experienced users. The results of two independent experiments are presented. The last 3 columns show the P-value when comparing distribution of ratios $r$ ( $r = \frac{\text{CldU length}}{\text{IdU length}}$ ) between users. . . .	30
Table 3.2	Score $s$ with respect to $\theta_{Max}$ . The lower $\theta_{Max}$ is, the higher the angular score $s$ is. . . . .	41
Table 3.3	List of Experimentally Optimized Parameter. All these parameters are set as multiplication factors of the PSF of the imaging system . . . .	45
Table A.1	Base de données . . . . .	68

## LISTE DES FIGURES

Figure 1.1	Structure de l'ADN : A- Molécule d'ADN structurée en double hélice (reproduit d'après : Genomics Education Programme). B- Structure d'un nucléotide. C- Architecture de la double hélice d'ADN . . . . .	3
Figure 1.2	Réplication de l'ADN . . . . .	4
Figure 1.3	G-quadruplexe: A- G-quartet B- Plusieurs topologies de superposition des G-quartet (reproduit d'après Harris LM, Merrick CJ (2015) G-Quadruplexes in Pathogens: A Common Route to Virulence Control ? PLoS Pathog) . . . . .	6
Figure 2.1	Images des fibres d'ADN obtenues par : A- Technique d'étalement des fibres d'ADN (DNA spreading) : Les analogues de nucléotides IdU et CldU apparaissant respectivement en vert et en rouge ont été incubés pendant 20 minutes chacun. B-Technique de peignage de l'ADN (DNA combing) : Les brins d'ADN sont rectilignes et parallèles. . . . .	10
Figure 2.2	Exemples de structures curvilignes : . . . . .	14
Figure 2.3	Exemples d'opérations morphologiques : L'image de gauche contient deux carrés de 3x3 pixels séparés par 2 pixels. Après dilatation par un élément structurant de forme carrée (3x3 pixels), on obtient l'image du milieu, un rectangle de 5x10 pixels. Après érosion avec le même élément structurant, on obtient l'image de droite, un rectangle de 3x8 pixels . . . . .	21
Figure 2.4	Exemple de structures curvilignes avec interruptions sur une image de brin d'ADN : Le brin rouge traversant l'image contient de multiples discontinuités (trous). . . . .	25

Figure 3.1	Inter-user variability. Three experienced users segmented the same set of 6 images obtained from two samples. A- Example of one fluorescent image from the second sample. B- Illustration of the different segmentations performed by three experienced users. Yellow fibers were segmented by the 3 users, orange fibers by only 2 users and red fibers by only 1 user. C- Ratio (CldU/IdU) distribution for each user (sample 1). The distribution of the 3rd user shows a significative difference ( $p < 10^{-2}$ between User 1 and User 3, $p < 10^{-4}$ between User 2 and User 2, Mann Whitney test). D-Ratio (CldU/IdU) distribution for each user (sample 2). The distribution of User 3 shows a significative difference with User 2 ( $p < 10^{-4}$ ) . . . . .	31
Figure 3.2	General Framework of ADFA. After a preprocessing step in which fiber -roughly segmented with the Edge detection method- are enhanced with respect to the background, the Point Spread Function (PSF) of the imaging system is calculated. The PSF aims at tuning spatial parameters (convolution filter size, maximum splicing distance, etc.) Then, a better segmentation is obtained on the enhanced image with tuned parameters. Unexploitable clusters of fibers are detected by measuring local fiber density. A Fiber Splicing algorithm connects nearby segments that belong to the same fiber and deletes fibers passing through unusable zones. Finally, a color label (e.g. red or green) is assigned to differentiate CldU vs IdU signals. This color assignment is based on the analysis of the intensity of both IdU and CldU channels.	32
Figure 3.3	Comparison of ADFA vs human users. A- Correlation plot for 98 images quantified by ADFA and manual users. B- Distribution of $C_{opi,opj}^{green}$ (normalized difference length for green portions) for each pair of operators. C- Distribution of $C_{opi,opj}^{red}$ (normalized difference length of red portions) for each pair of operators. D- Distribution of $\Delta ratio_{opi,opj}$ (difference of ratios) for each pair of operators . . . . .	33

Figure 3.4	Comparison between control and HU-treated samples. A, B- HU is incorporated during the second pulse. Orange and Blue violins are respectively the measures of our algorithm (ADFA) and a manual user. Experiments in A and B only differ in the incubation time of CldU (90 min vs 60 min). ***: $p < 10^{-31}$ (Mann Whitney test). N: number of quantified fibers. C- Comparison between OV1946 cells overexpressing RPA vs GFP. Cells are incubated in HU-containing medium for 3 hours after the CldU pulse. Distributions of ratios of the HU-treated sample normalized by the median of the control sample are displayed. ADFA quantification is in orange, manual quantification in blue. ***: $p < 10^{-15}$ (Mann-Whitney test). N: number of quantified fibers. . . . .	34
Figure 3.5	Variation of the incubation time of CldU. A- Algorithm quantification for CldU incubation time varying from 10 to 90 minutes. N: number of quantified fibers. B- Manual quantification for CldU incubation time varying from 10 to 90 minutes. N: number of quantified fibers. . . . .	35
Figure 3.6	From the raw image to the final segmentation. A- Raw image [ $I_1$ : red channel (CldU), $I_2$ : green channel (IdU) ]. B- Color enhancement after the preprocessing step [ $I_{1N}$ , $I_{2N}$ ]. C- Superposition of $BW_{DNA2}$ in white (segmentation of DNA fibers with the tuned edge detection method) and unexploitable high density areas in red obtained by thresholding the local white pixel normalized density $dN$ . D- Zoom of the yellow rectangle in C. DNA Fibers are fragmented. An example of blob and an example of strand are respectively flagged by a red full line arrow and a blue dotted arrow. E- $SKEL_{DNA}$ : Result of the splicing method. F- Zoom of the green rectangle in E. Fragmented fibers have been spliced. G- Color assignment: analysis of the red and green channels of the color enhanced image (B) . . . . .	39
Figure 3.7	Edge detection method. A- $I_{gray_2}$ : combination of the normalized intensity channel : $I_{gray_2} = \frac{I_{1N} + I_{2N}}{2}$ . B- $I_{LoG}$ : First edge detection [Laplacian of Gaussian (LoG)]. C- $I_{Canny}$ : Second edge detection. D- $BW_{DNA2}$ : Combination of $I_{LoG}$ and $I_{Canny}$ : Objects of $I_{LoG}$ that intersect a white pixel of $I_{Canny}$ are kept. . . . .	42



Figure 3.8	Splicing parameters. A- Connection between 2 strands. $\theta_{tan1}$ and $\theta_{tan2}$ are the orientation of the tangents at each endpoint (EP1, EP2) for each strand. $\theta_1$ and $\theta_2$ are the angles between the tangents and the connection segment [EP1, EP2] in yellow. $\theta_{Max}$ is the maximum of $\Delta\theta_{tan}$ , $\theta_1$ , $\theta_2$ . B- Connection between a strand and a blob. $\theta_{Max}$ is the angle between the tangent and the connection segment [EP1,C] in yellow. . . . .	43
Figure 3.9	Graph of Strands and Blobs. A - Image of 3 strands (in blue) and one blob (in orange) B— Objects whose endpoints are separated by less than dMax are linked by an edge (s,D). The angular score s is an integer between 1 and 5. The parameter D is the distance between the endpoints. . . . .	44
Figure A.1	BScan d'un sujet sain. Les interfaces des différentes couches rétinienne et choroïdiennes sont représentées : RVI (violet), RPEa (jaune), RPEp (orange) et CSI (bleu) . . . . .	65
Figure A.2	Découpage d'une carte d'épaisseur en 9 sous-régions . . . . .	70
Figure A.3	Classification globale. DMLA : Dégénérescence Maculaire Liée à l'Âge; Glaucom.: Glaucome; HTO: HyperTension Oculaire; Bird.:Uvéite de type Birdshot . . . . .	72
Figure A.4	Classification binaire . . . . .	72

## LISTE DES ANNEXES

Annexe A	CLASSIFICATION AUTOMATIQUE DE PATHOLOGIES OCULAIRES A PARTIR DE L'ÉTUDE DES CARTES D'ÉPAISSEUR DE LA CHOROÏDE ET DE LA RÉTINE SUR DES IMAGES OCT DU FOND DE L'OEIL	64
----------	--	----

## CHAPITRE 1 INTRODUCTION

Chaque cellule de notre organisme transmet l'intégralité de son information génétique lors de sa division en deux cellules filles. La cellule mère doit donc recopier fidèlement les 6 milliards de nucléotides constituant son ADN avant de se diviser. L'efficacité de la *réplication de l'ADN* est cruciale car une simple erreur ou mutation peut avoir des conséquences importantes sur l'organisme et/ou sa progéniture.

Le *stress réplicatif* est le ralentissement anormal voire l'arrêt des fourches de réplication. Si ce phénomène n'est pas résolu, il peut conduire à la rupture du double brin d'ADN et porter atteinte à la stabilité du génome. La génotoxicité induite par le stress réplicatif est un des facteurs de cancers, de maladies neurodégénératives ainsi que de troubles du développement.

Afin d'étudier la vitesse de progression des fourches de réplication, plusieurs techniques d'imagerie fluorescente existent. Ces techniques reposent sur l'incorporation d'analogues de nucléotides dans le milieu cellulaire. Une fois ces nucléotides intégrés dans le double brin d'ADN, les cellules sont lysées, l'ADN est transférée sur une lame de microscope puis les brins contenant les analogues de nucléotides sont imagés par immunofluorescence. La mesure de la longueur des brins fluorescents permet de quantifier le stress réplicatif.

Aujourd'hui, les longueurs de ces brins sont mesurées manuellement en utilisant des outils simples de manipulation d'image (eg. ImageJ). En plus d'être fastidieuse et longue, une telle quantification pose de nombreux problèmes de reproductibilité inter et intra opérateur. A notre connaissance, il n'existe qu'un seul algorithme permettant la segmentation automatique des brins d'ADN. Il repose sur l'extraction de structure curviligne par l'analyse de la hessienne de l'image. Cependant, cet algorithme payant n'a pas été rendu disponible à la communauté scientifique ce qui n'a pas permis une validation publique de ses performances.

L'objectif principal de notre étude est de proposer un logiciel en accès-libre permettant la segmentation automatique des brins d'ADN et la mesure de leur longueur respective.

Dans la suite de ce chapitre, nous introduirons les connaissances de base en biologie moléculaire qui permettront de mieux comprendre les principes de l'imagerie de l'ADN. A la fin du chapitre, nous présenterons la structure de la suite du mémoire.

### 1.1 Structure de l'ADN

L'acide désoxyribonucléique (ADN) est une macromolécule biologique présente dans les cellules de tous les organismes vivants. Il s'agit d'une double hélice de deux brins enroulés l'un

autour de l'autre (Figure 1.1 -A). Chaque brin est un polymère formé par la succession de blocs élémentaires appelés nucléotides.

Un nucléotide est constitué de 3 parties (Figure 1.1 B) : (i) Un sucre à cinq atomes de carbone (numérotés 1', 2', ..., 5') appelé désoxyribose. (ii) Un groupe phosphate relié au carbone 5' du désoxyribose. (iii) une base azotée aussi appelée base nucléique liée au carbone 1' du désoxyribose. Les deux premiers éléments sont identiques pour tous les nucléotides. En revanche, il existe quatre types de base nucléique : L'Adénine (A), Thymine (T), la Guanine (G) et la Cytosine (C). La liaison entre deux nucléotides A et B d'un même brin d'ADN se fait entre le groupe hydroxyl situé sur le carbone 3' du désoxyribose du nucléotide A et le groupe phosphate du nucléotide B (Figure 1.1 C). C'est l'ordre dans lequel se succèdent ces bases au sein de la molécule d'ADN qui constitue l'information génétique.

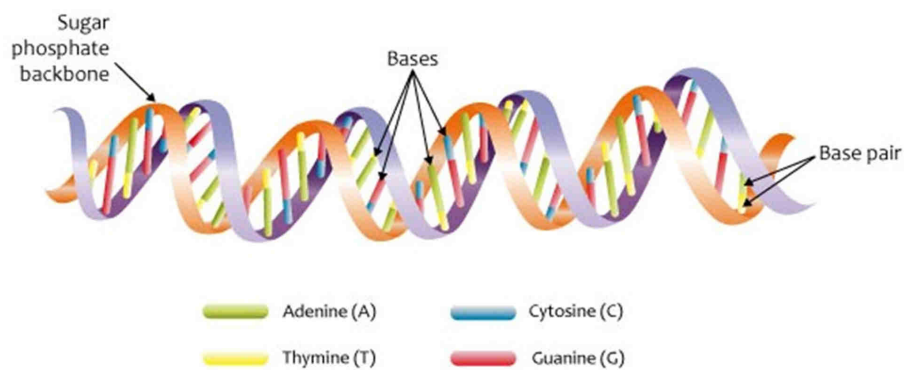
Les bases azotées de la molécule d'ADN sont situées dans la partie interne de la double hélice alors que le désoxyribose et le groupe phosphate sont situés à l'extérieure (Figure 1.1 A,C). L'appariement entre les deux hélices se fait donc entre les bases nucléiques des deux brins par des liaisons hydrogènes ce qui entraîne la formation de paires de base. Il existe deux types de paires de base : (i) L'appariement entre une Adénine (A) et une Thymine (T) par deux liaisons hydrogènes et (ii) l'appariement entre une Cytosine (C) et une Guanine (G) par trois liaisons hydrogènes. Ces deux types d'appariement canonique sont appelés paires de base de "Watson-Crick". En raison de cette complémentarité (A vs T et C vs G), l'information génétique portée par l'un des brins de la double hélice d'ADN est également portée à l'identique sur l'autre brin. Cette complémentarité entre bases joue un rôle important dans le mécanisme de réplication de l'ADN.

## 1.2 Réplication de l'ADN

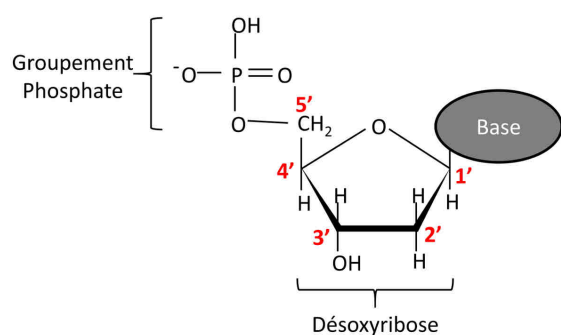
Durant la division cellulaire, la cellule mère se divise en deux cellules filles. Lors de cette étape, toute l'information génétique de la cellule mère se transmet aux deux cellules filles. Cela nécessite que l'ADN de la cellule mère soit dupliqué en amont de la division. Cette phase de synthèse (phase S) de l'ADN dans la cellule mère prend le nom de *réplication* de l'ADN. Ce processus complexe fait intervenir de nombreuses réactions moléculaires que nous résumons dans cette section.

L'*ADN polymérase* est une enzyme jouant un rôle crucial dans la réplication de l'ADN. Chaque enzyme d'ADN polymérase se fixe sur un des deux brins d'ADN et synthétise le brin complémentaire à partir des désoxyribonucléosides triphosphates présentes dans la cellule. Cette synthèse ne peut se faire que dans le sens  $5' \rightarrow 3'$  du brin (ces 2 chiffres correspon-

A-



B-



C-

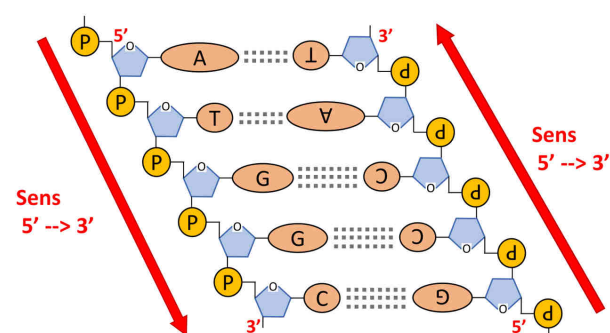


Figure 1.1 Structure de l'ADN : A- Molécule d'ADN structurée en double hélice (reproduit d'après : Genomics Education Programme). B- Structure d'un nucléotide. C- Architecture de la double hélice d'ADN

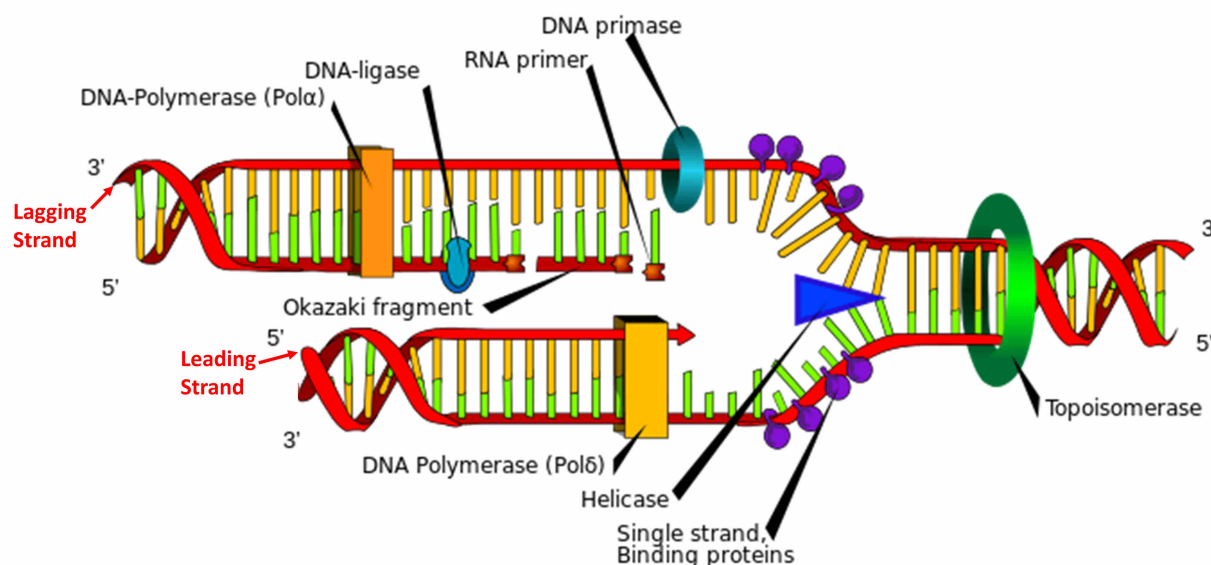


Figure 1.2 Réplication de l'ADN

dent aux numéros des carbones du désoxyribose) comme indiqué sur la figure 1.1-C. L'ADN polymérase synthétise un nouveau brin à partir d'un brin matrice préexistant : on dit que la réplication de l'ADN est un processus *semi-conservatif*. Comme les deux brins matrices d'ADN sont enroulés dans des sens opposés (de manière anti-parallèle), la synthèse de l'ADN le long de chaque brin se fait aussi dans des sens opposés. Le *brin direct* (or *leading strand*) est le brin complémentaire du brin parental orienté 3' vers 5' (le « brin direct » est donc orienté 5' vers 3'). Le *brin indirect* est le brin complémentaire du brin parental orienté 5' vers 3' (le « brin indirect » est donc orienté 3' vers 5') . Contrairement au brin direct, le brin indirect est synthétisé de manière discontinue sous forme de fragments d'Okazaki (figure 1.2).

Pour que l'ADN polymérase puisse se fixer sur un des deux brins de l'ADN, il faut que la double hélice se désenroule (ie. que les deux brins complémentaires se séparent). Comme la liaison entre les nucléotides complémentaires est très stable, cette ouverture nécessite l'intervention d'une protéine appelée *ADN hélicase*. Une fois fixée sur un brin d'ADN, l'hélicase produit un travail mécanique en hydrolysant de l'Adénosine TriPhosphate (ATP) ce qui permet son déplacement le long du brin ainsi que le désenroulement de la double hélice. On appelle *fourche de réplication* (replication fork), la zone de la molécule d'ADN où l'hélicase est en train de séparer le double brin. Pour éviter que les simples brins s'apparient

à nouveau, des protéines appelées *protéines SSB* (Single-Strand Binding protein) ou protéine de réplication A (RPA) chez les eukaryotes se fixent sur les simples brins.

La séparation des doubles brins commence à des endroits prédéfinis appelés *origines de réplication*. Ces zones caractéristiques permettent l'initialisation du désenroulement de la double hélice. Une cohorte de protéines dites *initiatrices* se fixent sur l'origine de réplication, et enroulent l'ADN autour d'elles ce qui forme un complexe protéine-ADN. Ce complexe se lie à une hélicase qui peut ainsi s'introduire sur un brin d'ADN initiant le désenroulement de l'hélice et la séparation du double brin d'ADN.

Le désenroulement de la double hélice n'est pas suffisant pour que l'ADN polymérase puisse initier la synthèse du nouveau brin. En effet, l'ADN polymérase ne peut pas synthétiser le nouveau brin sans la présence d'une terminaison OH-3' libre. C'est pour cela qu'une *amorce* de quelques nucléotides d'ARN est synthétisée par l'*ADN primase* en amont. Une seule amorce est nécessaire pour le brin direct car sa synthèse par l'ADN polymérase est continue. En revanche, des amorces sont synthétisées régulièrement le long du brin indirect et sont prolongées par d'autres ADN polymérases : ce sont les fragments d'Okazaki. Une fois la synthèse de l'ADN terminée, les amorces d'ARN sont éliminées puis remplacées par des bases d'ADN par l'enzyme ADN polymérase. Les différents fragments d'Okazaki sont reliés par des liaisons phosphodiester grâce à l'*ADN ligase*.

### 1.3 Notion de Stress Réplicatif

Le *Stress Réplicatif* se définit par le ralentissement ou l'arrêt de la progression des fourches de réplication. La compréhension des processus biochimiques activant le stress réplicatif est le sujet de nombreuses recherches (Zeman and Cimprich, 2014). Une des réponses biochimiques majeures provoquant le stress-réplicatif fait intervenir la kinase ATR (Ataxia telangiectasia and Rad3 related). Lorsque l'hélicase continue de désenrouler l'ADN alors que la polymérase est stoppée, il y a formation d'une portion d'ADN simple brin (DNAsb). La persistance de DNAsb, liée à la protéine de réplication A et adjacente à la fourche de réplication stoppée provoque la cascade de réponse entraînant le stress réplicatif par ATR (Byun et al., 2005). Cette cascade biochimique provoque l'inhibition du cycle cellulaire. Cela permet à la cellule de préserver ses ressources et de bénéficier de temps additionnels pour finir la réplication de l'ADN. L'ATR a aussi des effets locaux de stabilisation, de stimulation des fourches de réplication stoppées ainsi que d'inhibition des origines de réplication non activées (Zeman and Cimprich, 2014).

Les sources du déclenchement de stress réplicatif sont nombreuses. Nous résumons ici en

quelques lignes les causes principales.

- Des lésions de l'ADN peuvent perturber la progression de la fourche de réplication. Il existe plusieurs types de lésions de l'ADN comme l'appariement de deux bases non-complémentaires, l'absence de nucléotides sur un des deux brins ou l'absence de liaisons phosphodiester entre deux nucléotides le long du brin. Ces lésions peuvent être dues à des "sous-produits" du métabolisme cellulaire, aux rayons UV, ou à des agents mutagènes (Ciccia and Elledge, 2010).

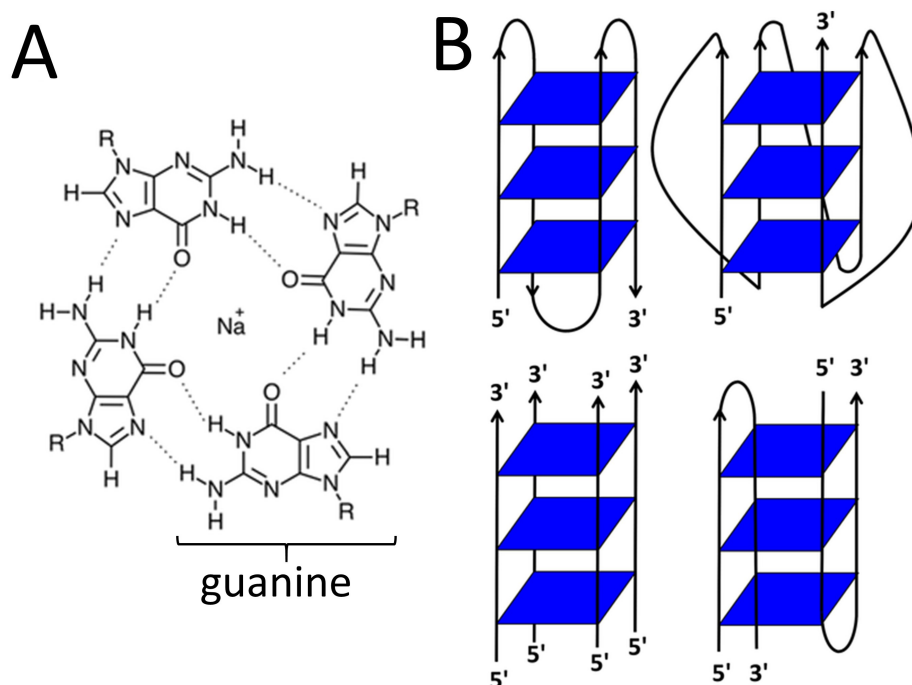


Figure 1.3 G-quadruplexe: A- G-quartet B- Plusieurs topologies de superposition des G-quartet (reproduit d'après Harris LM, Merrick CJ (2015) G-Quadruplexes in Pathogens: A Common Route to Virulence Control ? PLoS Pathog)

- Des modifications de la structure tridimensionnelle locale (ou structure secondaire) de l'ADN peut aussi être à l'origine de stress. Les G-quadruplexes sont des structures secondaires pouvant se former dans des régions de l'ADN marquées par la forte présence de la base Guanine (G). Quatre résidus de Guanine peuvent s'apparier pour former un G-quartet (voir figure 1.3-A). L'empilement d'au moins deux quartets intercalés par un cation monovalent stabilisant la structure (Na<sup>+</sup> dans la figure 1.3-A) constitue le G-quadruplexe (Figure 1.3-B). Ces structures secondaires sont la cause de ralentissement



de la réplication, de cassure du double brin et de mutation par délétion sur le site du quadruplexe (Bochman et al., 2012; Paeschke et al., 2013).

- Les processus de réplication et de transcription de l'ADN peuvent interférer. La *transcription* de l'ADN est la copie de l'information d'un gène en un brin d'ARN messager. Le mécanisme de la transcription est assez proche de celui de la réplication : l'ARN polymérase synthétise le brin d'ARN en circulant le long d'un brin d'ADN. Il est donc possible que les complexes de réplication et de transcription de l'ADN entrent en collision (Liu et al., 1993). Le brin d'ARN en cours de synthèse peut se lier à l'ADN derrière le complexe de transcription ce qui forme des R-loop, une structure à trois brins contenant un double brin hybride ARN:ADN et un ADNsb. Ces structures complexes peuvent aussi interférer avec la réplication (Gan et al., 2011).
- La limitation du nombre de composants utilisés pendant la réplication peut ralentir la fourche de réplication et induire du stress réplicatif. Par exemple, la dérégulation de la fabrication de désoxyribonucléoside triphosphate (dNTP) par les ribonucléotides réductases est un facteur de stress réplicatif (Poli et al., 2012).
- La surexpression de certains oncogènes viennent perturber la réplication. Par exemple, les oncogènes Ras et Myc stimulent les origines de réplication ce qui peut épuiser les sources de nucléotides ainsi que provoquer des collisions avec des complexes de transcription (Cottini et al., 2013).

L'arrêt des fourches de réplication et l'altération des processus de réplication ou de transcription peut causer des mutations ainsi que des maladies comme le cancer, des maladies neurodégénératives, le vieillissement prématuré, le retard de croissance ou d'autres troubles du développement comme le nanisme primordial (Negri et al., 2010; Ruzankina et al., 2007).

Le moyen le plus fiable pour détecter un stress réplicatif est de mesurer la progression des fourches de réplication en rendant fluorescent les brins synthétisés. Deux techniques d'imagerie utilisent ce principe : la technique de *propagation des fibres d'ADN* (*DNA spreading*) et le *peignage de l'ADN* (*DNA combing*). Nous détaillerons ces techniques dans le chapitre 2.

## 1.4 Structure du mémoire

La suite de ce mémoire sera organisée de la manière suivante :

Le chapitre 2 sera une revue de littérature développant d'une part les techniques d'imagerie

de l'ADN et d'autre part les techniques de traitements d'images permettant d'extraire des structures curvilignes.

Le chapitre 3 est constitué de notre article accepté avec révisions mineures dans le journal *DNA Repair*. Cet article détaille la méthodologie utilisée pour segmenter les brins d'ADN et valider les performances de l'algorithme.

Le chapitre 4 est une discussion générale revenant sur les résultats importants, les avantages et les limites de notre démarche.

Un dernier chapitre de conclusion fera la synthèse des travaux et discutera des améliorations futures envisagées.

En annexe, nous présenterons brièvement le second thème abordé durant la maîtrise. Ce travail, indépendant du reste du mémoire, est le développement d'un algorithme de diagnostic automatique de maladies oculaires.

## CHAPITRE 2 REVUE DE LITTÉRATURE

Cette revue de littérature se divise en deux parties : Dans la première, nous détaillons les techniques d'imagerie de l'ADN. Dans une seconde partie, nous ferons un tour d'horizon des algorithmes de traitement d'images capables d'extraire des structures curvilignes (ie. objets linéaires et filiformes).

### 2.1 Imagerie de l'ADN pour la détection du stress réplicatif

Les techniques d'imagerie de l'ADN reposent sur l'incorporation de nucléotides exogènes dans le brin d'ADN pendant sa réplication. Dans le passé, des précurseurs radioactifs (3H-Thymidine) étaient intégrés dans le brin d'ADN et imagés par autoradiographie (Salic and Mitchison, 2008). En raison de sa radioactivité, de sa faible résolution et de son faible rapport signal sur bruit, cette technique est désormais supplantée par des approches faisant intervenir des nucléotides exogènes non radioactifs que l'on marque par des anticorps spécifiques. Nous détaillons dans ce paragraphe les deux méthodes les plus couramment utilisées pour visualiser les brins d'ADN : la technique d'étalement des fibres d'ADN (DNA spreading) et la technique de peignage de l'ADN (DNA combing).

#### 2.1.1 Étalement des fibres d'ADN (DNA spreading)

Cette technique (Nieminuszczy et al., 2016) permet de colorer d'une certaine couleur les brins d'ADN synthétisés pendant un laps de temps  $T_1$  et d'une autre couleur ceux synthétisés pendant un laps de temps  $T_2$  (Figure 2.1-A). Les intervalles de temps  $T_1$  et  $T_2$  sont consécutifs. Un exemple classique d'expérience est de comparer la distance parcourue par chaque fourche de réplication dans un contexte de stress (intervalle  $T_2$ ) par rapport à une référence (intervalle  $T_1$ ).

Pour pouvoir repérer les séquences d'ADN synthétisées pendant un laps de temps  $T$ , on incorpore des analogues de nucléotides pendant ce laps de temps dans le milieu de culture des cellules. Un analogue de nucléotide est un nucléotide exogène dont la structure est légèrement différente d'un nucléotide endogène mais pouvant tout de même être inclus dans le brin d'ADN lors de la réplication. Les deux types d'analogues de nucléotides couramment utilisés pour l'imagerie de l'ADN sont IdU (5-Iodo-2'-deoxyuridine) et CldU (5-Chloro-2'-deoxyuridine). Ces deux analogues sont inclus dans les brins d'ADN par l'ADN polymérase durant la réplication. Ils sont ensuite liés à des anticorps spécifiques et imagés par microscopie

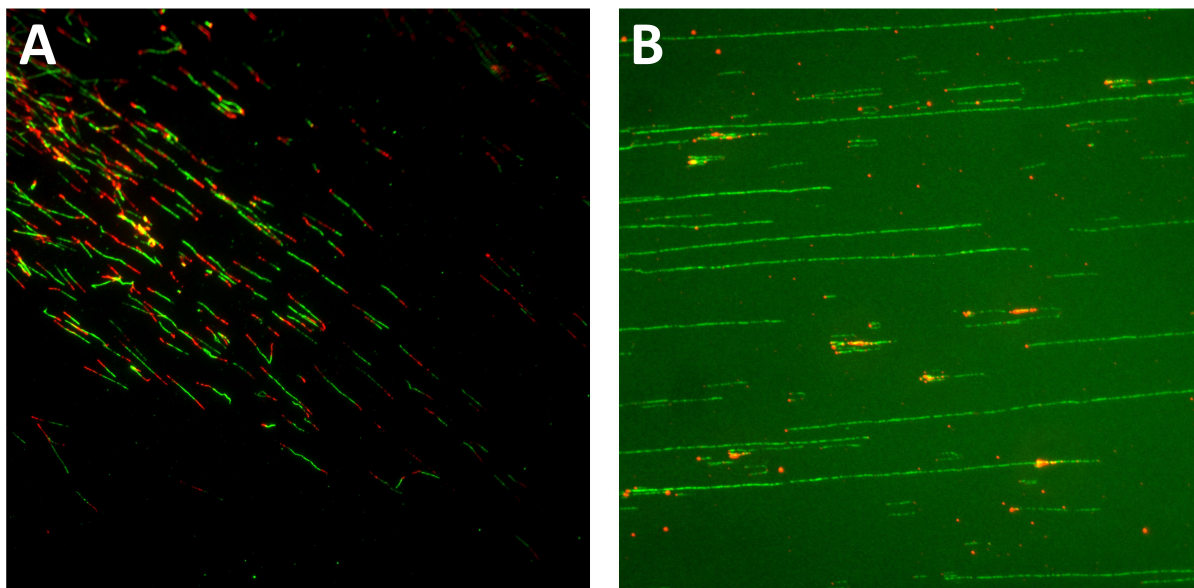


Figure 2.1 Images des fibres d'ADN obtenues par : A- Technique d'étalement des fibres d'ADN (DNA spreading) : Les analogues de nucléotides IdU et CldU apparaissant respectivement en vert et en rouge ont été incubés pendant 20 minutes chacun. B-Technique de peignage de l'ADN (DNA combing) : Les brins d'ADN sont rectilignes et parallèles.

à fluorescence.

Dans la suite, nous détaillons de manière plus précise les différentes étapes de la technique d' "étalement des fibres d'ADN" (DNA spreading) :

- **Marquage pulsé des cellules**

On incube des cellules avec les nucléotides IdU pendant un laps de temps  $T_1$  puis avec des nucléotides CldU pendant un laps de temps  $T_2$ . Après chacun de ces deux *marquages pulsés* (pulse labelling), le milieu est rincé par une solution tampon (PBS) afin de se débarrasser du surplus d'IdU et de CldU. La solution de PBS utilisée après  $T_2$  est glacée afin de stopper l'activité cellulaire. Ensuite les cellules sont récoltées et replacées dans une solution de PBS glacée avec une concentration cellulaire optimale dépendant de la lignée étudiée.

- **Transfert de l'ADN sur la lame de microscope**

Une goutte de la solution est posée sur une lame de microscope. Ensuite, les cellules sont lysées et la lame est inclinée de manière à ce que la goutte s'y répande. Ensuite, l'ADN

est dénaturée (ie. elle perd sa structure géométrique et les deux brins se séparent) par une solution d'acide chlorhydrique puis nettoyée avec une solution de PBS.

- **Coloration des analogues de nucléotides par immunofluorescence**

L'immunofluorescence est une technique qui permet de lier une molécule d'intérêt à un fluorochrome par l'intermédiaire d'un ou plusieurs anticorps. Ici, l'immunofluorescence est indirecte car elle fait intervenir deux types d'anticorps : (i) Des anticorps primaires monoclonaux vont se lier de manière spécifique aux analogues de nucléotides. (ii) Des anticorps secondaires polyclonaux (marqués par un fluorochrome) sont dirigés contre les anticorps primaires.

Avant d'incorporer les anticorps, il est nécessaire de saturer les sites de fixations non spécifiques de notre préparation en y incorporant une solution de "blocage" [e.g. Albumine de Sérum Bovin (ASB)]. Ensuite, les lames sont incubées pendant deux heures avec deux anticorps primaires : des anticorps anti-BrdU de rat réagissant contre CldU et des anticorps anti-BrdU de souris agissant sur IdU. Après un nouveau rinçage avec une solution de PBS, les anticorps secondaires fluorescents sont injectés : anticorps Alexa-594 anti-rat et anticorps Alexa-488 anti-souris. Les longueurs d'ondes maximales d'émission sont respectivement 617nm (rouge) and 519nm (vert). Finalement, on obtient par microscopie une image bicolore des brins d'ADN. Les sections rouges et en vertes correspondent respectivement aux brins synthétisés pendant les laps de temps  $T_1$  et  $T_2$ .

### 2.1.2 Peignage de l'ADN (DNA combing)

Le peignage de l'ADN (Bensimon et al., 1994, 1995) est similaire à la technique précédente mais permet l'alignement rectiligne et parallèle de tous les brins d'ADN.

Le *peignage de l'ADN* (*DNA combing*) aligne les brins d'ADN sur une surface hydrophobe. Pour préparer la surface, une plaque de verre est débarrassée de ses impuretés par traitement plasma afin de la rendre chimiquement réactive. La lame activée est ensuite enduite de silane grâce à un traitement dans une chambre saturée en vapeur de silane. Les cellules subissent le même marquage pulsé que pour l'expérience d'étalement des fibres d'ADN (cf. section 2.1.1). Les protéinases et détergents présents dans le bloc d'agarose vont lyser les cellules et dégrader les protéines. Cela a pour effet de libérer les protéines liées à l'ADN dans la solution. La lame de microscope est ensuite plongée verticalement dans la solution. Les molécules d'ADN s'attachent au groupe vinyle de la couche de silane. En enlevant lentement la lame de la solution, une force antiparallèle au mouvement aligne les molécules d'ADN le

long de la lame de microscope. L'ADN est ensuite immobilisé en déshydratant la surface et peut être observée par microscopie à fluorescence (Figure 2.1-B).

## 2.2 Segmentation de structures curvilignes

En imagerie, le terme *segmentation* est l'action de délimiter une zone d'intérêt dans une image. En d'autres termes, segmenter un objet consiste à isoler les pixels appartenant à cet objet par rapport au reste de l'image. La segmentation des brins d'ADN nécessite de détecter des objets curvilignes, c'est à dire des objets allongés, filiformes. Parmi les objets curvilignes les plus étudiés en imagerie, on trouve les vaisseaux sanguins cardiaques et rétinien, les membranes neuronales, les fissures dans des surfaces en béton (poutres, chaussée...) ou encore le réseau routier. De nombreux algorithmes de segmentation permettent d'extraire des structures curvilignes. Dans cette partie, nous nous proposons d'en faire la revue.

### 2.2.1 Définition des Structures Curvilignes

Bien que largement utilisé dans la littérature, le terme "structure curviligne" a rarement été clairement défini. A notre connaissance, Bibiloni et al. (2016) sont les seuls à proposer une définition plus formelle :

Une Structure Curviligne est une région de pixels dans une image qui remplit plusieurs critères géométrique (G) et photométrique (P) :

- G1. Les pixels de la structure curviligne doivent être globalement connectés.
- G2. La région doit être fine à travers un long chemin
- G3. La variation de largeur le long de la région doit être progressive
- G4. La structure globale des structures curvilignes peut prendre des formes variées : arbres binaires (e.g. réseau vasculaire, cf. figure 2.2.1.A), séries de segments non connectés (e.g. fibres d'ADN, cf. figure 2.1.A), etc.
- G5. La courbure des structures curvilignes peut varier selon les caractéristiques des objets. Par exemple, certains objets peuvent être très rectilignes, d'autres peuvent être légèrement courbés et d'autres encore peuvent être hautement tortueux.
- G6. Les structures curvilignes peuvent présenter un certain nombre de bifurcations. Une bifurcation se définit comme la division d'un chemin en trois branches. Il s'agit de la position où les trois branches se rejoignent.

- G7. Certaines structures peuvent se superposer sur l'image. On appelle intersection, le lieu de superposition d'au moins deux chemins donnant naissance à un carrefour d'au moins quatre voies.
- P1. Les pixels des structures curvilignes doivent avoir une intensité "significativement différente" par rapport à l'arrière plan.
- P2. La variation de l'intensité le long du chemin doit être progressive.
- P3. Le profil en travers - les valeurs d'intensité transverses à la direction du chemin - doit suivre une distribution spécifique. Dans les image 2D, ce profil est représenté comme une fonction unidimensionnelle tandis que pour les images 3D, le profil est une surface 2D.

Des exemples d'images de structures curvilignes sont représentés sur la figure 2.2.1.

### 2.2.2 Détection de Structures Curvilignes par Analyse du Gradient de l'Image

L'étude du gradient de l'image permet de détecter les zones présentant un contraste comme les contours des objets par exemple. Dans cette section, nous présentons quatre algorithmes de segmentation utilisant le gradient de l'image. Nous nous attarderons plus spécifiquement sur les algorithmes de Marr and Hildreth (1980) et de Canny (1986) car ils seront utilisés dans le chapitre 3 de ce mémoire.

#### Définition du gradient de l'image

Soit  $I(x, y)$ , une image à un seul canal avec  $x$ , l'indice de la colonne et  $y$ , l'indice de la ligne. Le gradient  $\vec{G}$  est défini en tout pixel  $(x, y)$  de l'image par

$$\vec{G}(x, y) = \begin{pmatrix} \frac{\partial I}{\partial x}(x, y) \\ \frac{\partial I}{\partial y}(x, y) \end{pmatrix} \approx \begin{pmatrix} I(x+1, y) - I(x, y) \\ I(x, y+1) - I(x, y) \end{pmatrix} \quad (2.1)$$

Il s'agit d'un vecteur qui représente le taux d'accroissement local de l'intensité de l'image dans les deux directions  $x$  et  $y$ .

Le gradient est utile pour formaliser la notion de *contour* dans une image. Un contour est défini comme le lieu des maximums de  $\|\vec{G}\|$  dans la direction de  $\vec{G}$ . Nous verrons dans les deux paragraphes suivants comment les algorithmes de Marr and Hildreth (1980) ou de Canny (1986) mettent en pratique cette définition pour extraire les contours des objets d'une image.

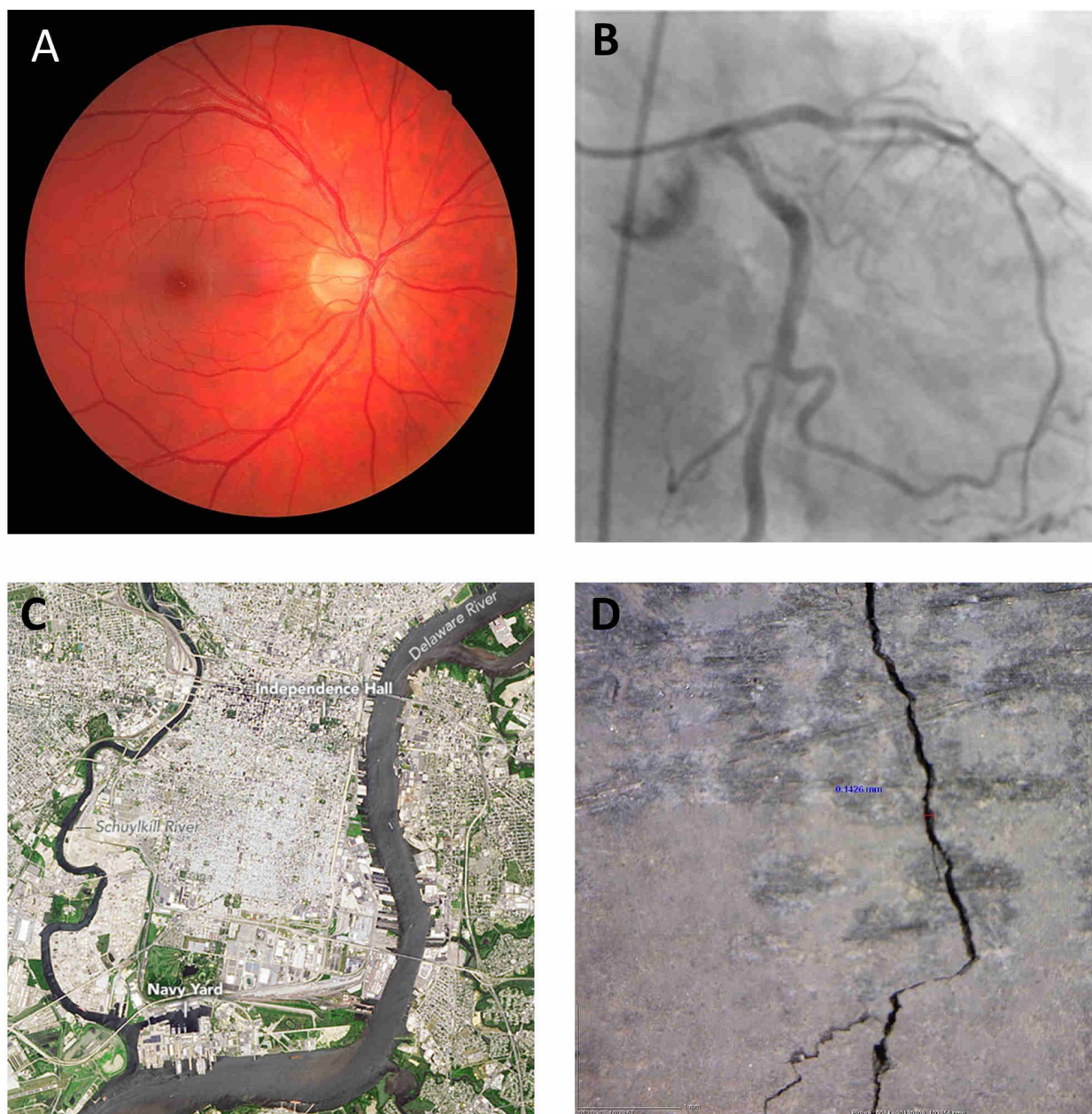


Figure 2.2 Exemples de structures curvilignes :

A- Réseau vasculaire sur une image du fond de l'oeil B- Artère coronaire issue d'une angiographie coronarienne. C- Fleuve Delaware et rivière Schuylkill en bordure de Philadelphie (image satellite), D- Fissure sur béton (reproduit avec la permission de Dr. Vladimir Ronin / EMC Cement)

### Algorithme de Marr-Hildreth

Les contours sont les pixels où le gradient passe par un maximum et donc où le laplacien (dérivées secondes) s'annule. L'algorithme de Marr and Hildreth (1980) est le suivant. Dans



un premier temps, on calcule le laplacien de l'image :  $\Delta I = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} \\ \frac{\partial^2 I}{\partial y^2} \end{pmatrix}$  En pratique, ce laplacien est calculée en convoluant l'image avec le laplacien d'une gaussienne (la gaussienne servant à débruiter et lisser l'image). On note  $L(x, y)$ , la norme du laplacien au pixel  $(x, y)$ . Dans l'approche de Marr-Hildreth, deux conditions sont nécessaires pour qu'un pixel soit considéré comme faisant partie d'un contour :

- (i)  $L$  passe par zéro selon la direction x ( ie.  $L(x + 1, y) * L(x, y) < 0$  )  
OU  $L$  passe par zéro selon la direction y ( ie.  $L(x, y + 1) * L(x, y) < 0$  )
- (ii) Le gradient de  $L$  selon la direction du passage par zéro est supérieur à un seuil  $T$  :  
 $L(x + 1, y) - L(x, y) > T$  OU  $L(x, y + 1) - L(x, y) > T$ .

Le seuil  $T$  permet de filtrer les contours à faible gradient souvent causés par le grain bruité de l'image. Cependant, l'absence total de filtre (ie. lorsque  $T=0$ ) est un cas particulier intéressant : Certes, de nombreux contours faux positifs liés au bruit de l'image sont extraits, mais les contours obtenus sont généralement fermés (ie. les contours sont des lignes qui bouclent sur elles-mêmes). Cette astuce que l'on utilisera dans le chapitre 3 de ce mémoire, permet de passer de l'extraction de contours à l'extraction d'objets.

L'algorithme de Marr-Hildreth possède les inconvénients suivants :

- (i) La décision de l'appartenance d'un pixel à un contour ne se fait que sur des paramètres très locaux (laplacien du pixel, gradient du laplacien). La prise en compte du contour dans sa globalité pourrait permettre un meilleur filtrage des artefacts liés au bruit de l'image.
- (ii) Le seuil unique  $T$  filtre les contours à faible gradient, même quand ces derniers sont le prolongement des contours à fort gradient.

L'algorithme de Canny permet de répondre à ces limitations en utilisant un seuillage par hystérésis sur la norme du gradient.

## Algorithme de Canny

L'algorithme de Canny se décompose de la manière suivante :

1. Lissage de l'image par un filtre gaussien

2. **Suppression des non-maxima locaux :** On détecte les pixels dont la norme du gradient est maximale dans la direction du gradient. Tous les autres pixels sont mis à zéros. On obtient ainsi une image binaire  $M_1$  : pixels maximum (1-pixels), pixels non-maximum (0-pixels).
3. **Seuillage par hystérésis :** Un seuil bas  $T_b$  et un seuil haut  $T_h$  sont appliqués sur les 1-pixels de  $M_1$  de la manière suivante : Tout d'abord, on met à zéros tous les 1-pixels de  $M_1$  dont la norme du gradient est inférieure à un seuil bas  $T_b$  ( $\|\vec{G}(x, y)\| < T_b$ ). Ensuite, on met à zéro tous les 1-pixels de  $M_1$  qui ne sont pas reliés à un 1-pixel dont la norme du gradient est supérieure au seuil haut ( $T_h$ ). Grâce à ce seuillage par hystérésis, on peut à la fois supprimer le bruit de l'image tout en conservant les contours à faible gradient s'ils sont reliés à des contours à fort gradient. Cependant, ce seuillage ne garanti pas de segmenter des contours fermés (contrairement à Marr-Hildreth avec  $T=0$ )

Zhao et al. (2002) se sont appuyés sur l'algorithme de Canny pour segmenter des routes à partir d'images satellite. Malgré le seuillage par hystérésis, les contours peuvent être fragmentés. Leur orientation est donc analysée afin de les prolonger s'ils présentent une continuité.

Sukkaew et al. (2005) comparent plusieurs filtres de contours dont Marr-Hildreth et Canny pour segmenter les vaisseaux sanguins sur des images de fond d'oeil. Bien que Marr-Hildreth semblerait légèrement plus efficace, les performances de ces filtres restent largement limités par la non homogénéité de la taille et du contraste des vaisseaux.

### Flux Orientés Optimaux (Optimally Oriented Flux)

Law and Chung (2008) ont développé une technique devenue assez populaire permettant de segmenter des structures curvilignes à partir de l'étude du flux de gradient entrant et sortant d'une sphère autour de chaque voxel. En trouvant l'axe optimal sur lequel le gradient de l'image est projeté, il est possible de segmenter des formes allongées. Law and Chung (2008) compare cette technique à l'étude de la hessienne de l'image (voir section 2.2.3) sur dix images tridimensionnelles synthétiques et un volume d'angiographie coronarienne par IRM. Il semblerait que les Flux Orientés Optimaux soient plus robustes au bruit et permettent de distinguer des objets proches et à faible contraste.

Turetken et al. (2013) étend l'approche de Law et al. en étudiant le flux de gradient à travers des objets dont la forme est plus complexe qu'une simple sphère. Cette amélioration est validée sur 14 images issues de micrographies confocales ou à champ large du cerveau de la souris.

De nombreuses études (Benmansour and Cohen, 2011; Rigamonti and Lepetit, 2012) et (Becker et al., 2013) intègrent des métriques issues de l’approche par flux de gradient pour améliorer les performances de leur algorithme (voir ci-après).

### Autres descripteurs dérivés du gradient

Moreno and Smedby (2015) proposent deux descripteurs dérivés du gradient de l’image. Le premier résulte de l’analyse des valeurs propres et des vecteurs propres d’un tenseur construit à partir du gradient. Cette approche s’inspire de celle de l’analyse de la hessienne de l’image (voir section 2.2.3). Le second descripteur se rapproche de l’analyse par Flux Orienté Optimaux. Il est construit à partir de l’analyse harmonique sphérique de la distribution du gradient sur une sphère autour de chaque pixel. Rivest-Henault and Cheriet (2013) avaient proposé un descripteur très similaire basé sur l’analyse harmonique sphérique non plus du gradient mais des dérivées secondes de l’image à plusieurs échelles. Nous verrons dans la partie suivante que les métriques issues de la dérivée seconde de l’image sont en effet très utilisés pour extraire des structures curvilignes.

### 2.2.3 Détection de Structures Curvilignes par Analyse de la Hessienne de l’image

L’étude de la matrice hessienne permet de détecter des structures curvilignes dans une image. La matrice hessienne d’un pixel  $(x,y)$  se définit à partir des dérivées secondes selon la formule suivante :

$$H = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{pmatrix} \quad (2.2)$$

(En pratique,  $H$  se calcule en convoluant l’image avec les quatre dérivées secondes de la gaussienne :  $H = \begin{pmatrix} \frac{\partial^2 G}{\partial x^2} & \frac{\partial^2 G}{\partial x \partial y} \\ \frac{\partial^2 G}{\partial y \partial x} & \frac{\partial^2 G}{\partial y^2} \end{pmatrix} * I$

On note  $\lambda_1$  et  $\lambda_2$  les valeurs propres associées aux vecteurs propres unitaires  $\mathbf{u}_1$  et  $\mathbf{u}_2$ . L’étude des vecteurs propres et des valeurs propres de la matrice hessienne pour un pixel  $P$  nous donne des informations sur le profil d’intensité au voisinage de  $P$ . Un pixel appartenant à une structure curviligne sera caractérisé par  $\lambda_1$  très faible et  $\lambda_2$  grand. Les vecteurs propres  $\mathbf{u}_1$  et  $\mathbf{u}_2$  indiquent respectivement la tangente et la normale à la structure curviligne.

A partir des valeurs propres, Lorenz et al. (1997), Sato et al. (1997) et Frangi et al. (1998) proposent différentes métriques permettant de mesurer la probabilité qu’un pixel appartienne à une structure curviligne. Les métriques issues de la hessienne sont très populaires et ont été adaptées à différentes modalités d’imagerie dont l’angiographie par rayon X (Li et al.,

2017), l’Imagerie par Résonance Magnétique (Ballerini et al., 2018) ou l’imagerie ultrasonore (Hennersperger et al., 2015).

Les métriques dérivées des valeurs propres de la hessienne comportent des limites. En effet, elles sont assez peu homogène entre des structures de diamètres différents et elles peinent à détecter les parties des objets peu curvilignes telles que les croisements ou les bifurcations.

Malgré ces limites, Wang et al. (2011) ont conçu un outil de segmentation des brins d’ADN en s’appuyant sur la Hessienne. C’est à notre connaissance, la seule étude proposant de mesurer automatiquement la longueur des brins d’ADN. Tout d’abord, le score de curvilinéarité développé par Frangi et al. (1998) est calculé pour chaque pixel de l’image. Ensuite, les brins d’ADN sont reconstitués par marche aléatoire (random walk) : Le point de départ est un pixel fortement curviligne autour duquel on recherche un autre pixel très curviligne dans une région distante de rayon  $R = 6$  pixels orientée dans la direction du vecteur propre tangent  $\mathbf{u}_1$ . Ce suivi (tracking) s’arrête lorsque le score de curvilinéarité des pixels de la zone investiguée passe en dessous d’un seuil  $T$  ( $T = 0.01$ ). La détection des analogues de nucléotides (IdU ou CldU) se fait en analysant les canaux vert et rouge dans un voisinage de chaque pixel retenu. La méthode a été validée sur 10 images en comparant les fibres segmentées automatiquement avec une référence obtenue manuellement. Cette approche comporte plusieurs limites : Tout d’abord, le score de curvilinéarité de Frangi et al. (1998) est souvent trop élevé dans le prolongement des fibres ce qui ne permet pas de les délimiter précisément. De plus, cette méthode ne prend pas en compte les fibres qui se croisent ce qui restreint grandement l’analyse. Par ailleurs, tous les paramètres de segmentation (eg.  $T$  et  $R$ ) sont codés "en dur" ce qui ne permet pas une flexibilité de l’algorithme en fonction d’éventuelles variations de la structure et de l’intensité de l’image. Enfin, la méthode a été validée sur seulement une dizaine d’images et la comparaison entre l’algorithme et l’opérateur manuel est assez restreinte.

#### 2.2.4 Détection de Structures Curvilignes par Vote de Tenseur (Tensor Voting)

L’approche par *vote de tensor* (ou *tensor voting*) peut être employée pour segmenter les structures curvilignes dans une image. Développée par Medioni et al. (2000), cette technique prend inspiration de la *théorie gestaltiste* de la perception de la forme. Selon cette théorie, la perception d’objets repose sur l’extraction de formes globales et non pas sur la simple juxtaposition des parties. L’approche par tensor voting détecte les objets d’une image par leur capacité à s’intégrer à une forme plus globale. Cette méthode prend donc en compte des critères de proximité et de continuité. Le processus commence par initialiser les directions de chaque pixel. Ces directions sont stockées sous la forme d’un tenseur du second ordre similaire

à la hessienne (cf.section 2.2.3). Ces orientations locales sont ensuite raffinées et propagées chez les pixels voisins lors de l'étape dite de "vote". Il est ainsi possible de reconnecter les objets appartenant à une même structure et de calculer un score appelé "saillance structurelle" (*structural saliency*) pour chaque pixel. Ce score évalue la capacité du pixel à se distinguer en s'intégrant dans une forme globale.

La technique de tensor voting a été appliquée à l'extraction de vaisseaux sanguins de l'image du fond de l'oeil (Park et al., 2007; Goatman et al., 2011; Christodoulidis et al., 2016) et à la segmentation des jonctions d'ancrage de cellules sur des images de microscopie à fluorescence (Loss et al., 2011).

### 2.2.5 Détection de Structures Curvilignes par Score d'Orientation

Les *Scores d'Orientation* (*Orientation Score*) sont une méthode de détection de structures curvilignes s'attaquant particulièrement aux problèmes d'intersections de structures. Les méthodes précédemment présentées ne sont pas capable de discriminer deux structures qui se croisent. Parfois même, les croisements empêchent la détection localement car le modèle utilisé ne prend pas en compte la superposition d'objets (eg. étude de la hessienne).

L'approche par Score d'Orientation développée par Duits et al. (2007) permet de résoudre ce problème. Le Score d'Orientation d'une image est une fonction qui pour chaque pixel  $(x,y)$  et chaque orientation  $\theta \in [0, 2\pi]$  assigne un score  $S$  ( $S \in \mathbb{R}$  ou  $\mathbb{C}$ ). Ce score décrit la probabilité que le pixel  $(x,y)$  s'inscrive dans une forme orienté par  $\theta$ . Le traitement de l'image s'effectue dans l'espace du Score d'Orientation : l'ensemble  $SE(2)$  (*eucledian motion group*). Par inversibilité, il est ensuite possible de revenir dans le domaine de l'image.

La plupart des études utilisant le score d'orientation ont été appliquées à la segmentation du réseau vasculaire à partir d'images du fond de l'oeil. Les méthodes de construction et de traitement du score d'orientation sont différentes: Bekkers et al. (2014) construit le score d'orientation à partir d'ondelettes (*cake and gabor wavelets*). Zhang et al. (2016) applique les propriétés de la matrice hessienne dans le domaine du score d'orientation. Abbasi-Sureshjani et al. (2018) complexifie l'espace  $SE(2)$  en un espace Riemannien à cinq dimensions en ajoutant des caractéristiques d'intensité et de courbure.

### 2.2.6 Détection de Structures Curvilignes par l'algorithme du chemin minimal

L'algorithme du chemin minimal a été étudié dans le cadre de la segmentation de structures curvilignes et notamment sur des images de vaisseaux sanguins. Il s'agit de trouver un trajet entre deux points tout en minimisant une fonction d'énergie définie comme la somme des

coûts des pixels le long du chemin parcouru. L'algorithme de Dijkstra ou la méthode de Fast Marching sont utilisés pour résoudre ce problème d'optimisation.

Li and Yezzi (2006) ont adapté l'algorithme du chemin minimum à la segmentation de structures tubulaires sur des images 3D. Il s'agit de transposer le problème dans un espace en 4 dimensions : pour chaque point du trajet, on définit une sphère caractérisée par sa position mais aussi par son rayon. Il est ainsi possible d'extraire la surface des vaisseaux sanguins. La fonction coût est ici définie à partir de l'intensité et du gradient de l'image. Benmansour and Cohen (2011) reprennent la méthode de Li and Yezzi (2006) en complexifiant la fonction coût. Cette dernière devient une métrique orientée selon la direction du vaisseau et basée sur les Flux Orientés Optimaux (voir section 2.2.2).

L'inconvénient des deux approches précédentes est qu'il faut définir manuellement un point de départ et un point d'arrivée pour segmenter un chemin. Kaul et al. (2012) s'affranchissent de cette contrainte en proposant une approche où seul un point source est nécessaire. Cet algorithme a été validé à la fois sur des images de fissure de béton (pavement crack) et de vaisseaux sanguins rétiniens.

L'algorithme du chemin minimum a le défaut de favoriser les chemins courts car ils ont un faible coût. Chen et al. (2016) s'attaque à ce problème en ajoutant une étape de *backtracking* lors de la recherche du chemin : L'information collectée depuis les points atteints vers le point de départ permet d'améliorer les performances de la segmentation. L'approche est validée sur des images de fissures de béton, et de vaisseaux sanguins issus d'angiographies coronariennes.

Plus récemment, Liao et al. (2018) améliorent le calcul de la fonction coût. Cette dernière est calculée dynamiquement durant la propagation du chemin alors que précédemment elle était calculée en amont. Cette fonction coût prend en compte la courbure de la trajectoire ainsi que la tangente du front d'onde de l'approche Fast Marching. L'algorithme a été testé sur un grand nombre d'images : 40 images de fond d'oeil, 10 images synthétiques de réseau vasculaire avec 30 niveaux de bruits différents, 40 angiographies par résonance magnétique du cerveau et 10 images de Google Earth pour la segmentation de rivières.

### 2.2.7 Détection de Structures Curvilignes par Opérations Morphologiques

Les opérateurs morphologiques sont des outils utilisés pour modifier la forme des objets d'une image. Un noyau convolutif appelé *élément structurant* se déplace sur tous les pixels de l'image. Pour chaque configuration, on assigne au pixel central le maximum (dilatation) ou le minimum (érosion) de l'intensité des pixels contenu dans le noyau. Un exemple de

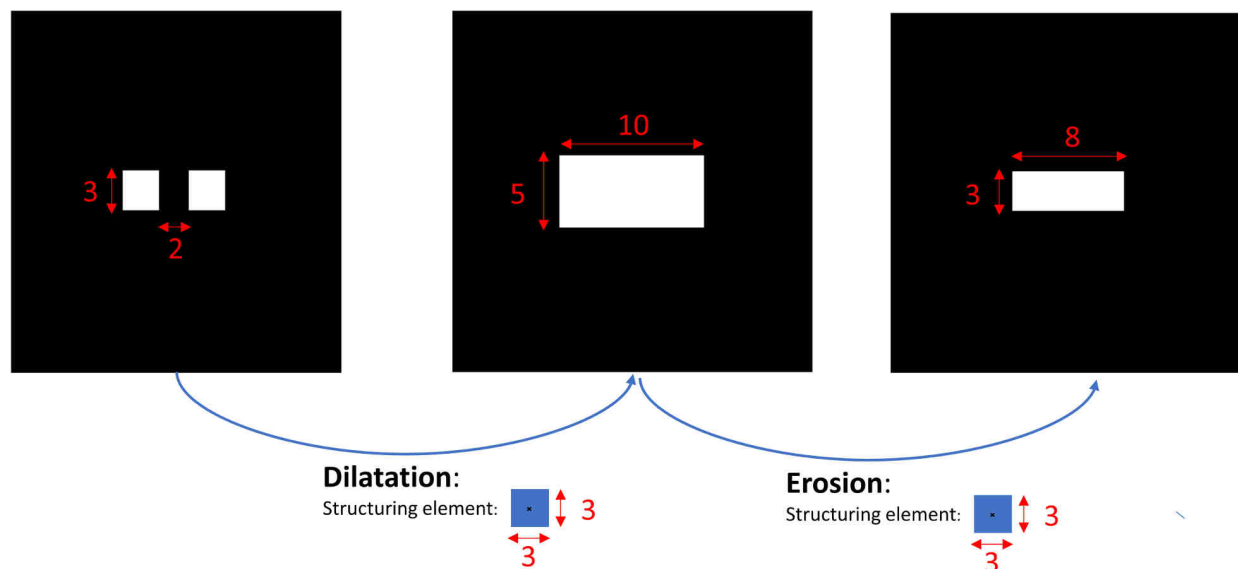


Figure 2.3 Exemples d'opérations morphologiques : L'image de gauche contient deux carrés de 3x3 pixels séparés par 2 pixels. Après dilatation par un élément structurant de forme carrée (3x3 pixels), on obtient l'image du milieu, un rectangle de 5x10 pixels. Après érosion avec le même élément structurant, on obtient l'image de droite, un rectangle de 3x8 pixels

dilatation et d'érosion est représenté sur la figure 2.3. La forme et l'orientation de l'élément structurant ainsi que la succession des opérations morphologiques permettent de filtrer les objets présents dans l'image. De nombreuses études (Thackray and Nelson, 1993; Soille et al., 1996; Zana and Klein, 2001; Tankyevych et al., 2008; Bouraoui et al., 2010) proposent d'extraire les structures curvilignes à partir d'opérations morphologiques.

Le travail récent proposé par Merveille et al. (2018) se distingue de ceux de ses prédécesseurs par l'introduction de nouveaux opérateurs morphologiques : les *opérateurs de chemin* (*path operators*). Ces opérations morphologiques ne se font plus avec un élément structurant fixe mais avec une famille de forme adaptative permettant une plus grande flexibilité géométrique. A partir de ces opérateurs, un score de curvilinéarité et un descripteur d'orientation peuvent être calculés. Cet algorithme intitulé RORPO (Ranking the Orientation Responses of Path Operators) a été testé sur 10 images synthétiques et 1 angiographie coronarienne CT en le comparant à la méthode de Frangi et al. (1998) et de Flux Orienté Optimaux (Law and Chung, 2008).

### 2.2.8 Détection de Structures Curvilignes par Apprentissage Machine

Les méthodes précédentes extraient de nombreux descripteurs de structures curvilignes (Flux Orientés Optimaux, valeurs propres de la Hessienne, courbure, orientation etc.). Il serait intéressant de tester ces descripteurs dans le cadre d'un algorithme d'apprentissage machine. Dans cette section, nous présentons les travaux utilisant des algorithmes de classification automatique pour extraire des structures curvilignes.

Rigamonti and Lepetit (2012) entraînent plusieurs classifieurs (Forêt d'arbres aléatoires et régression logistique) à partir de deux types de descripteurs : (i) des descripteurs connus *a priori* (*handcrafted*) extraits de la Hessienne et des Flux Orientés Optimaux (ii) des descripteurs appris issus de filtres linéaires entraînés pendant la phase d'apprentissage. Annunziata et al. (2015) suivent cette même approche en intégrant des descripteurs qui prennent en compte un contexte plus large pour chaque pixel. Il s'agit d'agrandir le champ d'activation des descripteurs pour comprendre les relations entre les différentes sections de la structure curviligne. Un an plus tard, Annunziata and Trucco (2016) réduisent drastiquement le temps d'entraînement des descripteurs par une meilleure initialisation des filtres linéaires. Cette approche a été testée sur des vaisseaux sanguins du fonds de l'oeil et sur des micrographies neuronales.

Turetken et al. (2016) proposent une approche hybride entre la théorie des graphes et l'apprentissage automatique. Après avoir calculé un score de curvilinearité basé sur les flux orientés optimaux pour chaque pixel de l'image, un graphe est construit en reliant les régions voisines à forte curvilinearité. Par la suite, des classifieurs (gradient boosting decision trees) établissent les poids le long du graphe. Le meilleur sous-graphe est ensuite extrait en résolvant un problème d'optimisation en nombre entier (integer programming). Cette approche a été testée sur des modalités variées aussi bien pour la segmentation du réseau routier, de réseaux neuronaux ou d'arbres vasculaires rétinien.

Schneider et al. (2015) et Zhang et al. (2018a) appliquent un outil de détection d'objets appelé Forêt de Hough à la segmentation de vaisseaux rétinien (photos du fond d'oeil), de routes (images aériennes) et de neurones (microscopie confocale). La détection de structures curvilignes par Forêt de Hough se décompose comme suit : L'image est découpée en patches. Pour chacun des patches, les arbres de décision vont voter sur son appartenance ou non à une structure curviligne. Le vote des arbres de décision se décompose en deux variables  $C_p$  et  $v_p$ .  $C_p$  est une variable binaire décrivant l'appartenance ou non du patch à une structure curviligne.  $v_p$  est un vecteur estimant la position du patch par rapport au centre de la structure curviligne. Par accumulation des votes dans l'espace de Hough, puis par extraction des maxima locaux, il est possible d'extraire les objets d'intérêt. Les descripteurs des patches



sont obtenus par entraînement de filtres convolutifs comme pour les méthodes d'apprentissage précédent.

Enfin, depuis 3 ans, les réseaux de neurones convolutifs sont entraînés pour segmenter le réseau vasculaire rétinien (Maninis et al., 2016; Fu et al., 2016; Lahiri et al., 2017) , les artères coronaires à partir d'angiographie à rayon X (Fan et al., 2018; Nasr-Esfahani et al., 2018) ou les vaisseaux du foie (Ibragimov et al., 2017; Huang et al., 2018). Il ne s'agit plus d'entraîner quelques filtres linéaires comme pour Rigamonti and Lepetit (2012); Annunziata et al. (2015); Annunziata and Trucco (2016) mais plusieurs cascades de filtres de convolutions associées à des fonctions non linéaires. L'avantage de cette approche est d'à la fois optimiser des filtres de bas niveau (première couche du réseau) qui extraient des caractéristiques simples (eg. gradients orientés ) et des filtres de haut niveau permettant de contextualiser les informations de bas niveau (e.g. intégration des différentes parties dans une forme globale).

### 2.2.9 Discussion

#### La comparaison des différentes méthodes est difficile

Comme nous venons de le voir, les méthodes de détection de structures curvilignes sont très nombreuses et variées. Il est difficile de pouvoir juger objectivement de la qualité de chacune des méthodes présentées. Souvent, la plupart des études testent leur algorithme sur des échantillons d'images restreints et différents. La quasi-totalité des codes n'est pas en libre accès ce qui rend la comparaison d'autant plus difficile.

Certaines études proposant une analyse comparative de plusieurs algorithmes de segmentation peuvent même avoir des résultats contradictoires. C'est le cas de Phellan and Forkert (2017) et Merveille et al. (2018) qui comparent tous les deux l'approche de Frangi et al., les Flux Orientés Optimaux (Law and Chung, 2008) et l'algorithme RORPO développé par Merveille et al. Phellan and Forkert (2017) a testé ces 3 algorithmes sur 35 images avec différents niveaux de bruit (15 images synthétiques 3D de réseau vasculaire, 20 images angiographiques par temps de vol du cerveau incluant des malformations artério-veineuse pour 10 d'entre elles). Merveille et al. (2018) a testé ces algorithmes sur 10 images synthétiques avec différents niveaux de bruit et seulement 1 image d'angiographie CT du réseau coronarien. Les performances de l'approche de Frangi sont beaucoup plus faibles dans l'article de Merveille et al. (2018) que dans Phellan and Forkert (2017) pour les images synthétiques (score de Mathews de 0.19 vs 0.99). De plus, RORPO semble être robuste en présence de bruit pour Merveille et al. (2018) ce qui est contraire aux résultats de Phellan and Forkert (2017). Même si les échantillons d'images testées sont différents, on peut s'étonner de ces

contradictions.

Enfin, lorsqu’il arrive que différents algorithmes sont testés sur les mêmes bases de données, les métriques utilisées sont différentes. C’est le cas de Zhang et al. (2018b), Rigamonti and Lepetit (2012), Annunziata and Trucco (2016) et Liao et al. (2018) qui, sur la base de données DRIVE d’images du fond de l’oeil, utilisent respectivement:

- le coefficient de Dice
- le couple précision/rappel
- l’aire sous la courbe précision rappel, le F-score et l’index de Jaccard
- une métrique *ad-hoc* appelée  $Q_{inside}$

### **Certaines méthodes permettent de prolonger les interruptions des structures curvilignes**

La segmentation des structures curvilignes peut parfois présenter des trous au milieu des structures (Figure 2.4). Ces erreurs sont souvent dues à des artefacts de l’image causant une baisse de l’intensité de la structure curviligne dans ces régions.

Ces discontinuités de l’intensité sont omniprésentes sur les images de brins d’ADN (Figure 2.4). Leur présence peut s’expliquer de plusieurs manières : L’efficacité de la solution de blocage (Albumine de Sérum Bovin), l’affinité entre anticorps et antigènes, les étapes de rinçage de la préparation sont autant de paramètres expérimentaux dont l’optimisation n’est pas toujours simple. Outre ces artefacts expérimentaux, l’origine des discontinuités de la fluorescence peut aussi être expliquée par la capacité de l’ADN polymérase à sauter certaines régions d’ADN dans un contexte de stress réplcatif. L’ADN polymérase laisse derrière elle des sections d’ADN simple brin sur lesquels les anticorps fluorescents ne peuvent pas se fixer. Enfin, une dernière hypothèse d’explication de la présence de ces discontinuités est la compétition entre les nucléotides endogènes non fluorescents et les analogues de nucléotides exogènes fluorescents.

Parmi les méthodes de traitement d’image présentées précédemment, certaines prennent en compte le problème de discontinuité de l’intensité de l’image.

Joshi et al. (2011) et Forkert et al. (2012) proposent deux méthodes basées sur la théorie des graphes pour combler les discontinuités au sein de vaisseaux rétiniens et du réseau vasculaire cérébral obtenu par IRM. Joshi et al. (2011) ont utilisé l’algorithme de Dijkstra pour relier les extrémités des vaisseaux à partir d’une fonction coût basée sur l’orientation du chemin et

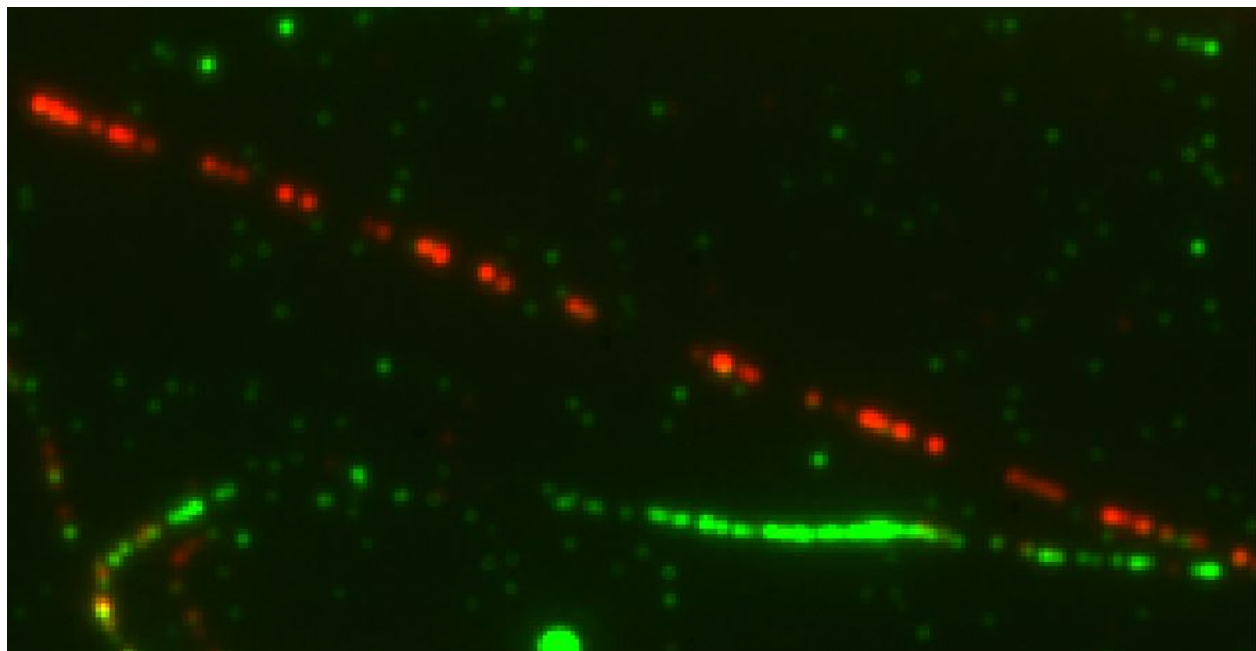


Figure 2.4 Exemple de structures curvilignes avec interruptions sur une image de brin d'ADN : Le brin rouge traversant l'image contient de multiples discontinuités (trous).

l'intensité des pixels parcourus. L'algorithme est testé sur 25 images de fond d'oeil et 81.63% des discontinuités ont été comblées avec succès. Pour Forkert et al. (2012), la fonction coût est calculée à partir du score de curvilinearité de Sato et al. (1997). Parmi les 100 discontinuités artificiellement introduites, 83 ont été correctement prolongées.

Risser et al. (2008); Loss et al. (2011); Christodoulidis et al. (2016) utilisent l'approche par tensor voting pour combler ces discontinuités. Bien que le vote par tenseur s'inspirant des théories de Gestalt a pour vocation de prolonger des structures locales dans un objet global, cette approche a une efficacité limitée lorsque des objets parasites à proximité de la discontinuité viennent perturber le vote des tenseurs.

L'approche par chemin minimal proposé par Liao et al. (2018) est aussi capable de remplir les trous dans les vaisseaux rétiniens. 99.61% des 508 trous analysés sur 10 angiographies IRM du cerveau ont été correctement complétés.

Récemment Zhang et al. (2018b) étudient de manière extensive l'efficacité des scores d'orientation pour prolonger les discontinuités. La méthode, testée sur une large base de données (109 images du fond de l'oeil et 30 images de fibres nerveuses cornéennes ) réussit à combler 95.27% des 40 457 trous dont la taille varie de 7x7 à 39x39 pixels.

## 2.3 Objectifs de recherche

Comme nous venons de le voir, il existe de nombreuses méthodes de segmentation des structures curvilignes. Cependant, il est difficile d'estimer lesquelles de ces méthodes seraient performantes sur les images de brin d'ADN car cette modalité, peu étudiée en traitement de l'image, se distingue par de multiples discontinuités au sein des structures curvilignes. A notre connaissance, seule une étude (Wang et al., 2011) propose une extraction des fibres d'ADN par analyse de la hessienne (Frangi et al., 1998) sans proposer une approche robuste pour prolonger les larges discontinuités.

Nous rappelons que l'objectif principal de notre étude est de proposer un logiciel en accès-libre permettant la segmentation automatique des brins d'ADN et la mesure de leur longueur respective.

Notre projet de recherche comporte les objectifs spécifiques suivant :

1. Développer une méthode automatique d'extraction des brins d'ADN
2. Développer une méthode de prolongement des discontinuités de la segmentation précédemment obtenue
3. Mesurer la longueur de chaque section d'analogues de nucléotides (CLdU et IdU) par analyse des canaux rouge et vert de l'image.
4. Évaluer et valider notre algorithme en le comparant à la segmentation de plusieurs utilisateurs humains ainsi qu'en le testant dans divers contextes de stress réplicatif.

## CHAPITRE 3 ARTICLE 1 : AN OPEN-SOURCE IMPLEMENTATION FOR FULLY AUTOMATED QUANTIFICATION OF DNA FIBERS

Pierre Ghesquière<sup>1</sup>, Abdelhamid Elsherbiny<sup>1,2</sup>, Emile Fortier<sup>1,2</sup>, Mary McQuaid<sup>1,2</sup>, François Bélanger<sup>1</sup>, Farida Cheriet<sup>1,3</sup>, Javier Mazzaferri<sup>1</sup>, Elliot Drobetsky<sup>1,2,4</sup>, Hugo Wurtele<sup>1,2,4</sup>, Santiago Costantino<sup>1,5</sup>

1- Centre de Recherche de l'Hôpital Maisonneuve-Rosemont, Montréal, Québec, Canada

2- Programmes de Biologie Moléculaire, Université de Montréal, Montréal, Québec, Canada

3- Département de Génie Informatique et Génie Logiciel, Polytechnique Montréal, Québec, Canada.

4- Département de Médecine, Université de Montréal, Montréal, Québec, Canada

5- Département d'Ophtalmologie, Université de Montréal, Montréal, Québec, Canada.

Dans cette section, nous reproduisons notre article accepté dans la revue *DNA Repair* le 7 janvier 2019.

### Abstract

DNA fiber fluorography is widely employed to study the kinetics of DNA replication, but the usefulness of this approach has been limited by the lack of freely-available automated analysis tools. Quantification of DNA fibers usually relies on manual examination of immunofluorescence microscopy images, which is laborious and prone to inter- and intra-operator variability. To address this, we developed Automated DNA Fiber Analysis (ADFA), an unbiased fully-automated algorithm which quantifies fluorescent DNA fibers length from microscopy images. ADFA is an open-source image processing tool based on edge detection and a novel segment splicing approach. Here, we describe the algorithm in detail, validate our results experimentally, and benchmark the analysis against manual assessments. Our open-source implementation is offered under the General Public License to the scientific community (<https://github.com/pierreghesqui/FiberQ>).

### 3.1 Introduction

DNA replication is tightly regulated by a myriad of molecular interactions that ensure accurate transmission of genetic information to daughter cells. The fidelity of this process can be compromised by DNA replicative stress, i.e., the abnormal slowing down or stalling of DNA Replication Forks (RF) (Zeman and Cimprich, 2014). Indeed, stalled RF must be resolved

in a timely manner to prevent their “collapse” into highly-genotoxic double strand breaks (DSB) which engender chromosomal rearrangements and genomic instability (Alexander and Orr-Weaver, 2016). Replicative stress may arise from various impediments to DNA synthesis, such as DNA secondary structures (eg. G-quadruplexes, palindromes) (Ribeyre et al., 2009), RNA:DNA hybrids (R-loops) (Gan et al., 2011), collisions between replication and transcription machineries (Liu et al., 1993), dNTP pool imbalances (Poli et al., 2012), or DNA adducts induced by any among a plethora of genotoxins (including endogenous agents, environmental mutagens, and chemotherapeutic drugs). Mutation or defective regulation of essential DNA replication factors, as well as activation of certain oncogenes including Ras Myc, and Bcl-2, have also been shown to generate abnormal DNA replication dynamics (Xie et al., 2014; Cottini et al., 2013). The resultant replicative stress-induced genomic instability constitutes a critical determinant in both cancer development and treatment. Replicative stress is also implicated in the molecular pathogenesis of aging and neurodegenerative disease, as well as developmental syndromes such as primordial dwarfism (Negrini et al., 2010).

DNA fiber fluorography is commonly used to evaluate RF progression at the level of individual DNA molecules (11). This method is based on the incorporation of halogenated nucleotide analogs, such as chloro- (CldU), iodo- (IdU), or bromo-deoxyuridine (BrdU) in nascent DNA at RF in living cells. In a typical experiment, sequential incorporation of two nucleotide analogs, e.g., IdU and CldU, is performed. Cells are exposed to DNA replication stress-inducing treatments during or after the second labeling period (Jackson and Pombo, 1998). Following cell lysis and spreading of DNA on microscopy slides, DNA molecules are labeled using anti-IdU and anti-CldU antibodies coupled to different fluorophores. Two-color images generated by fluorescence microscopy then reveal contiguous labeled regions in elongated DNA fibers. Measurement of the respective lengths of these labeled stretches of DNA permits quantification of RF progression. Variations of this general experimental strategy have been used extensively to quantify DNA replication dynamics in the context of replicative stress induced by a plethora of experimental conditions, including exposure to chemotherapy drugs and expression of oncogenes (Chaudhuri et al., 2016; Belanger et al., 2018).

Evaluation of DNA fiber length is generally performed manually using simple image manipulation tools. This procedure is laborious and subject to inter- and intra-user variability stemming in part from unintended bias in the choice of fibers to be measured. These problems highlight the need for a reliable computational method for unbiased analysis of DNA fiber immunofluorescence images. To the best of our knowledge, the only available tool is not free (Wang et al., 2011), and the source-code not open, rendering it unavailable for wide distribution and public validation. Here we present Automated DNA Fiber Analysis (ADFA), a novel, fully-automated algorithm to segment (ie. delineate) and quantify labeled DNA fiber

length from fluorescence microscopy images. ADFA is based on edge detection filters and splicing techniques, and provides rapid, reliable, and unbiased analysis of DNA fibers. We describe our algorithm in full detail, and use images obtained under different experimental conditions to compare its performance with manual segmentation. Our open-source software is offered free of charge to the scientific community.

## 3.2 Results

### 3.2.1 Inter-user variability upon manual quantification of DNA fibers on immunofluorescence images

DNA fiber immunofluorescence images display variations in fiber density, straightness, branching, and staining intensity. Inter-user variability due to biased identification and inaccurate measurement of isolated fibers is expected. To evaluate this variability, three experienced users were asked to manually segment the same batch of 6 images obtained from two different experiments (3 images per experiment). Cells were pulsed sequentially with IdU and CldU such that bicolor contiguous regions of DNA fibers represent progressing RF which first incorporated IdU, and then CldU, into nascent DNA. Figure 3.1-A shows an example of such images extracted from the second sample. Using a widely-available open-source image manipulation software, GIMP (GNU Image Manipulation Program), users colored CldU and IdU fiber sections in red and green, respectively.

We found that the number of segmented fibers was very different between users (Table 3.1). For example, User 2 measured 60 fibers more than User 1 and 37 fibers more than User 3 in sample 1, which represents respectively 21% and 13% of all the 285 segmented fibers (Table 3.1). Moreover, only 16% and 23% of the total number of segmented fibers were quantified by every user for experiment 1 and 2 respectively. As an illustration, Figure 3.1-B shows how many users chose each segmented fiber in an image extracted from experiment 2. Finally, statistically significant differences were observed between User 3 and the other two users (see Table 3.1 column 9, 10 and 11) when quantifying the distribution of ratios  $r$  ( $r = (\text{CldU length}) / (\text{IdU length})$ ), a widely-used metric to quantify the dynamics of RFs (Nieminuszczy et al., 2016; Chaudhuri et al., 2016; Belanger et al., 2018).

Such disagreement among users illustrates the challenge surrounding manual selection of fibers. This can be explained by several factors. First, overlap between IdU and CldU signals complicates the precise localization of label changes, leading to variations in  $r$  ratio quantification between users for a given fiber. Second, the frequent occurrence of small gaps that split fibers in smaller segments may in some cases cause disparities between users.

Table 3.1 Inter-user variability: manual quantification of CldU- and IdU-labeled DNA fibers by three experienced users. The results of two independent experiments are presented. The last 3 columns show the P-value when comparing distribution of ratios  $r$  ( $r = \frac{\text{CldU length}}{\text{IdU length}}$ ) between users.

	Total nb of segmented fibers	Fibers segmented by User 1	Fibers segmented by User 2	Fibers segmented by User 3	Fibers chosen by only one user	Fibers chosen by only two users	Fibers chosen by the three users	P value (User 1 vs User 2)	P value (User 1 vs User 3)	P value (User 2 vs User 3)
Experiment 1	285	87 (31%)	148 (52%)	111 (39%)	166 (58%)	72 (25%)	47 (16%)	0.37	<b>0.0027</b>	<b>1.6e-5</b>
Experiment 2	273	97 (36%)	143 (52%)	98 (36%)	100 (37%)	111 (41%)	62 (23%)	0.057	0.084	<b>4.0e-5</b>

Indeed, merging such segments during the quantification is a subjective choice: for example, some users may tolerate high radii of curvature in single fibers whereas others prefer straight fibers. Finally, entangled fibers, debris and non-specific antibody staining can interfere with accurate measurements.

These factors are not only a source of variability for human segmentation, they also pose a challenge for the development of image processing algorithms toward this task. Thus, our initial objectives were: i) the identification of DNA fibers within a noisy image ii) the design of a robust fiber splicing criteria for bridging the gaps along elongated straight fibers, and iii) the removal of unexploitable tangled fibers.

To tackle those challenges, we devised the following approach (detailed in the Materials and Methods section and summarized in Figure 3.2). After a preprocessing step in which fibers are enhanced with respect to the background, DNA fibers are detected using an ad hoc edge detection method inspired from Marr Hildreth’s (Marr and Hildreth, 1980) and Canny’s (Canny, 1986). The splicing of nearby segments that may belong to a single fiber is based on their curvature and distance. Unexploitable clusters of fibers are removed by establishing the maximum local fiber density that can be processed. Finally, color transitions are determined by analyzing the difference in fluorescence between IdU and CldU channels.

### 3.2.2 Comparing ADFA vs manual quantifications

Manual quantifications were compared with our algorithm by computing the correlations between both methods as well as inter-operator differences for humans and ADFA. We used a database of 98 images and measured the median CldU/IdU length ratios for each image. We observed a very good correlation between the algorithm and users (Pearson’s coefficient of 0.79), demonstrating that ADFA is consistent with trained users’ observations (Figure 3.3-A). To evaluate inter-operator differences for individual fibers, 12 images were segmented manually by three experienced users or ADFA. For each fiber segmented by two operators, we measure the difference between the two segmentations using three metrics:  $C_{opi,opj}^{green}$ ,  $C_{opi,opj}^{red}$  and  $\Delta ratio_{opi,opj}$ .



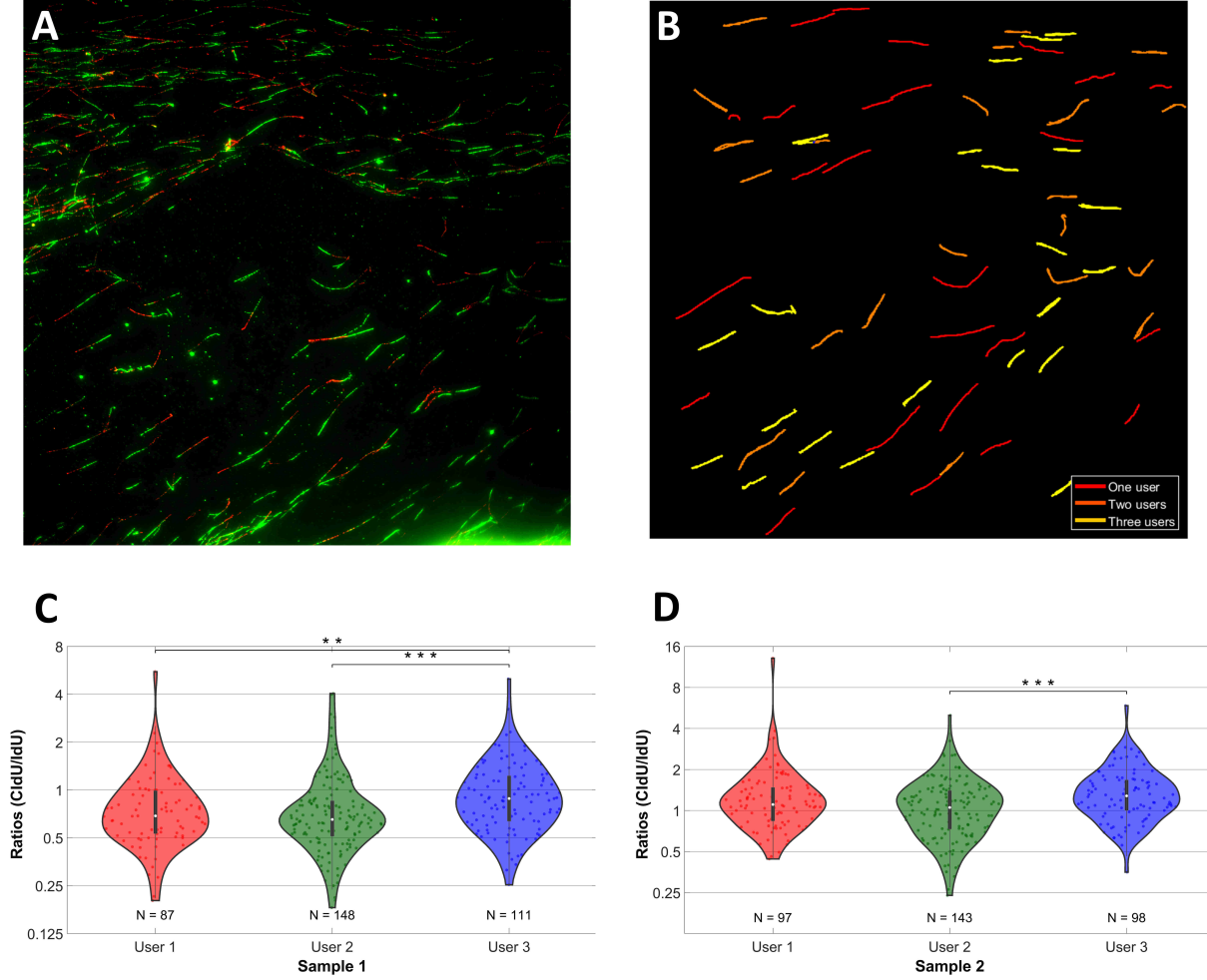


Figure 3.1 Inter-user variability. Three experienced users segmented the same set of 6 images obtained from two samples. A- Example of one fluorescent image from the second sample. B- Illustration of the different segmentations performed by three experienced users. Yellow fibers were segmented by the 3 users, orange fibers by only 2 users and red fibers by only 1 user. C- Ratio (CldU/IdU) distribution for each user (sample 1). The distribution of the 3rd user shows a significant difference ( $p < 10^{-2}$  between User 1 and User 3,  $p < 10^{-4}$  between User 2 and User 2, Mann Whitney test). D-Ratio (CldU/IdU) distribution for each user (sample 2). The distribution of User 3 shows a significant difference with User 2 ( $p < 10^{-4}$ )

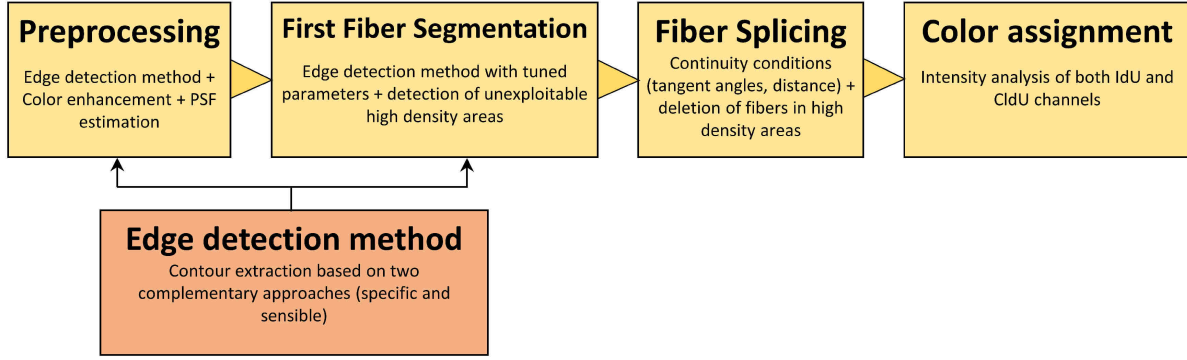


Figure 3.2 General Framework of ADFA. After a preprocessing step in which fiber -roughly segmented with the Edge detection method- are enhanced with respect to the background, the Point Spread Function (PSF) of the imaging system is calculated. The PSF aims at tuning spatial parameters (convolution filter size, maximum splicing distance, etc.) Then, a better segmentation is obtained on the enhanced image with tuned parameters. Unexploitable clusters of fibers are detected by measuring local fiber density. A Fiber Splicing algorithm connects nearby segments that belong to the same fiber and deletes fibers passing through unusable zones. Finally, a color label (e.g. red or green) is assigned to differentiate CldU vs IdU signals. This color assignment is based on the analysis of the intensity of both IdU and CldU channels.

- $C_{opi,opj}^{green}$  and  $C_{opi,opj}^{red}$  measure the difference of length between operator i and operator j normalized by the mean of the 2 lengths:

$$C_{opi,opj}^{green} = \frac{l_{opi}^{green} - l_{opj}^{green}}{\frac{l_{opi}^{green} + l_{opj}^{green}}{2}}$$

where  $l_{opi}^{green}$  (resp.  $l_{opj}^{green}$ ) are the length of the green part of the fiber measured by operator i or j.

- $\Delta ratio_{opi,opj}$  is the difference of CldU/IdU ratios for a fiber segmented by opi and opj. For the above metrics, only bicolor fibers were considered. Metrics distributions were computed for each possible pair of operators (Figure 3.3-B, C, D). We observed similar inter-operator variability when comparing either ADFA vs users or user i vs user j.

### 3.2.3 Performance of ADFA in biologically-relevant experimental conditions

We next sought to validate ADFA experimentally by comparing the progression of RF in samples treated with hydroxyurea (HU) vs untreated controls. Hydroxyurea inhibits the activity of the ribonucleotide reductase enzymatic complex (Yarbro, 1992), thereby depleting

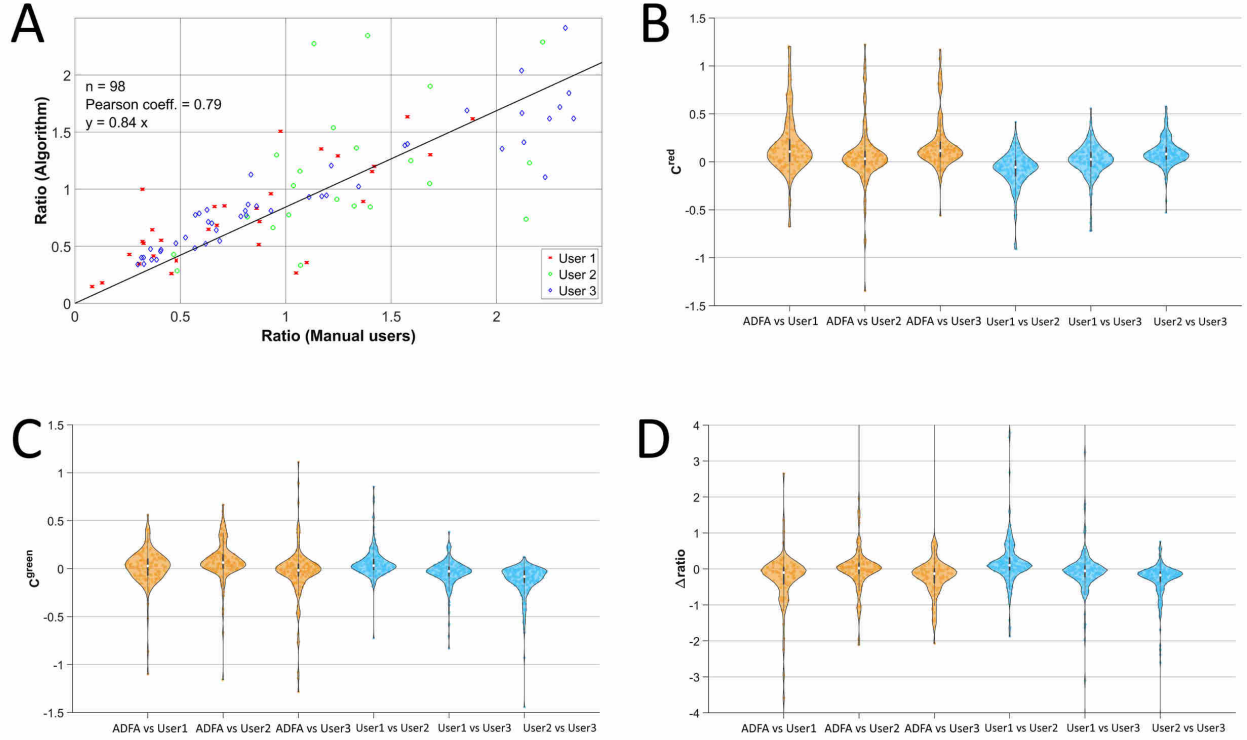


Figure 3.3 Comparison of ADFA vs human users. A- Correlation plot for 98 images quantified by ADFA and manual users. B- Distribution of  $C^{green}_{opi,opj}$  (normalized difference length for green portions) for each pair of operators. C- Distribution of  $C^{red}_{opi,opj}$  (normalized difference length of red portions) for each pair of operators. D- Distribution of  $\Delta ratio_{opi,opj}$  (difference of ratios) for each pair of operators

deoxyribonucleotide pools and strongly slowing down RF progression. We first quantified images acquired from a typical experiment in which HU was included in the culture media of HeLa cells during the second pulse with halogenated nucleotides (Figure 3.4 A-B). IdU/CldU pulses periods were 30 min/60 min for the first experiment, and 30 min/90 min for the second experiment. As expected, both ADFA and manual quantifications reveal that CldU-labeled tracks are shorter in HU-treated samples than in control untreated samples, leading to reduced CldU/IdU length ratio (Figure 3.4 A-B).

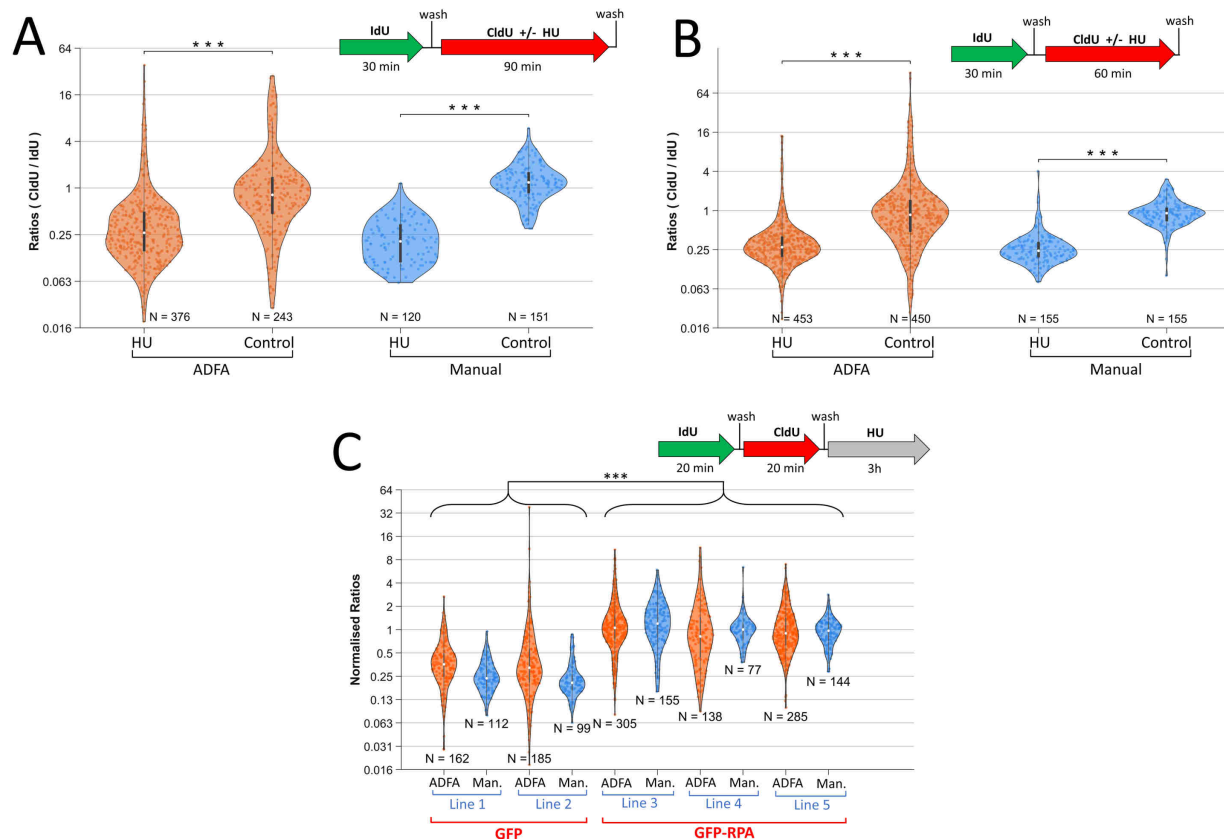


Figure 3.4 Comparison between control and HU-treated samples. A, B- HU is incorporated during the second pulse. Orange and Blue violins are respectively the measures of our algorithm (ADFA) and a manual user. Experiments in A and B only differ in the incubation time of CldU (90 min vs 60 min). \*\*\*:  $p < 10^{-31}$  (Mann Whitney test). N: number of quantified fibers. C- Comparison between OV1946 cells overexpressing RPA vs GFP. Cells are incubated in HU-containing medium for 3 hours after the CldU pulse. Distributions of ratios of the HU-treated sample normalized by the median of the control sample are displayed. ADFA quantification is in orange, manual quantification in blue. \*\*\*:  $p < 10^{-15}$  (Mann-Whitney test). N: number of quantified fibers.

Our group and others have previously shown that nascent DNA is unstable in certain cell lines

due to aberrant nuclease activity at stalled DNA RF (Chaudhuri et al., 2016; Belanger et al., 2018). We also recently showed that overexpression of all three subunits of the Replication Protein A complex suppresses such nascent DNA instability in the ovarian cancer cell line OV1946 (Belanger et al., 2018). To evaluate the stability of nascent DNA at stalled RFs, OV1946 cells were exposed to HU for 3 hours after the second pulse. Reduction in the length of the second label (CldU) upon incubation with HU reflects nuclease-mediated degradation of nascent DNA, which is rescued by RPA overexpression. Both manual quantification and our algorithm confirmed this result, indicating that GFP-RPA expression leads to higher CldU/IdU ratios vs GFP alone, as expected (Figure 3.4 C). We note that measurements performed with ADFA generally display higher variance, partly due to a much larger number of measured fibers.

We also validated ADFA by varying the duration of the nucleotide analog pulse and evaluating the effect on labeled track length. Incubation time for the second nucleotide analogue (CldU) was incrementally increased from 10 to 90 minutes, whereas the IdU pulse period remained constant at 20 min. Manual and ADFA quantification of the images indicate, as expected, an increase in CldU/IdU ratio for CldU labeling periods of up to 60 minutes (Figure 3.5 A-B). Intriguingly, we observed that the length of the CldU tracks reaches a plateau between 60 and 90 minutes of labeling. Examination of the images reveal that this is likely due to the presence of extremely long fibers, which are almost invariably entangled in clusters or cut out of the image field. Only very short isolated DNA molecules can be detected under these conditions, which introduces biases in both manual and automatic segmentations.

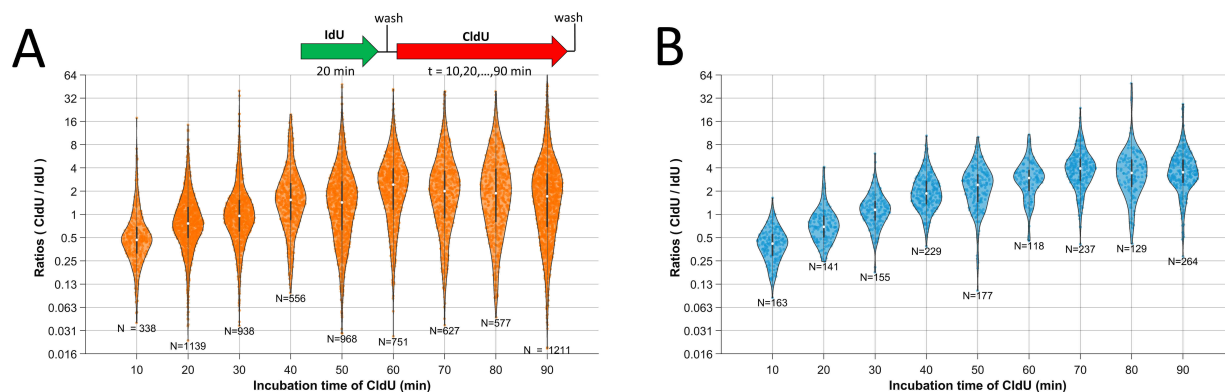


Figure 3.5 Variation of the incubation time of CldU. A- Algorithm quantification for CldU incubation time varying from 10 to 90 minutes. N: number of quantified fibers. B- Manual quantification for CldU incubation time varying from 10 to 90 minutes. N: number of quantified fibers.

### 3.3 Materials and Methods

#### 3.3.1 DNA Fiber- Assay

Exponentially growing Hela cells were labeled with 10  $\mu$  M IdU for a duration T1. Cells were washed twice with 3 ml PBS, then were labeled with 250  $\mu$  M CldU for a duration T2. Cells were washed, harvested, and resuspended in PBS at a final concentration of 500 cells/ $\mu$ L. Two  $\mu$  L were transferred to a slide, overlaid with 7.5  $\mu$  L lysis buffer (0.5% SDS, 200 mM Tris-HCl (pH 7.4), and 50 mM EDTA), and incubated at room temperature for 3 min. Slides were tilted to allow DNA to spread due to gravity, air-dried for 7 min, fixed for 10 min with freshly prepared 3:1 methanol/acetic acid, and air-dried for 7 min. DNA was denatured by incubating the slide in 2.5 M HCl for 80 minutes, followed by three washes with PBS. Blocking was performed with 200  $\mu$  L 5% BSA for 20 minutes. For immunostaining, slides were incubated for 2 hours with primary antibodies; ab6326 anti-BrdU (cross-reacts with CldU) antibody (rat) (1:400) and BD Biosciences 347580 anti-BrdU (cross-reacts with IdU) antibody (mouse) (1:25) in 5% BSA in PBS. Slides were washed three times with PBS-T (PBS + 0.05% tween), then once with PBS. Next, slides were incubated for one hour with the secondary antibodies; anti-rat Alexa-594 (1:100) and goat anti-mouse Alexa-488 (1:100) in 5% BSA in PBS. Slides were washed three times with PBS-T (PBS + 0.05% tween), then one time with PBS. Slides were allowed to dry in the air for few minutes then mounting medium was added and images were acquired using two different microscopes: either GE Healthcare Deltavision or ZEISS Axio Imager 2.

#### 3.3.2 ADFA algorithm: method

From a raw image made up of two color channels (eg. IdU and CldU), DNA fibers are segmented and the length of each fluorescent marker is quantified to evaluate the dynamic of the replication process. Our framework (summarized in Figure 3.2) is divided in four steps. Fibers are first enhanced with respect to the background and then extracted with an ad-hoc edge detection method. As this first segmentation is flawed by gaps within strands, a splicing step reconnects sections belonging to single DNA fibers. Finally, we quantify the length of each fluorophore by analysing channel intensity differences.

In our analysis pipeline, some parameters have been experimentally optimized using a large image database obtained with two imaging systems (GE Healthcare Deltavision and ZEISS Axio Imager 2). All these Experimentally Optimized Parameters (noted  $EOP_i$  hereafter) are expressed related to the diameter of the Point Spread Function of the imaging system to derive spatial metrics. If necessary, they can easily be tuned by the user. Table 3.3 displays

the values that we have established in our implementation of ADFA.

### Preprocessing: Color enhancement and Point Spread Function estimation

To smooth the raw image without altering the edge, a 4x4x1 median filter is applied to both channels. For simplicity, we call  $I_1$ , the channel of the first nucleotide analog, and  $I_2$  the channel of the second one. The two channels are combined into a grayscale image,  $I_{gray1} = \frac{I_1 + I_2}{2}$ . An ad-hoc edge detection algorithm (see details below) is applied on  $I_{gray1}$  to obtain a first rough segmentation of DNA fibers. This rough segmentation ( $BW_{DNA1}$ ) is a binary image in which true pixels represent fiber pixels and false pixels are considered background.

Intensity normalisation is performed separately on both  $I_1$  and  $I_2$ . On each of those two channels, we calculate the 5th and the 95th intensity percentiles of fiber pixels (ie. pixels belonging to the foreground mask of  $BW_{DNA1}$ ). We linearly map the intensity values of both channels by saturating the bottom and top intensities to those two percentiles. The new normalised channels  $I_{1N}$  and  $I_{2N}$  which are matrices of doubles in the interval  $[0,1]$ , have enhanced fiber fluorescence with respect to the background (Figure 3.1B). A new grayscale image,  $I_{gray2}$ , is obtained by combining  $I_{1N}$  and  $I_{2N}$ .

The rough DNA mask,  $BW_{DNA1}$ , is also used to estimate the diameter of the Point Spread Function (PSF) of the imaging system in pixel units. Throughout the analysis pipeline, this PSF is used as a characteristic metric to adjust all spatial parameters (morphological operators, structuring elements, convolutional filters kernels, etc.) to the individual images. The diameter of the PSF is estimated by measuring the fiber width: for each fiber, the intensity distribution on a cross section  $s$  is fitted by a Gaussian function  $P(\mu, \sigma) = A \cdot \exp(-\frac{(s-\mu)^2}{2\sigma^2})$ . The PSF diameter is set to the median of all  $\sigma$  measured.

### First fiber segmentation

To obtain a first fiber segmentation adapted to the structure of the input image, the ad-hoc edge detection method is applied again but this time on the enhanced grayscale image ( $I_{gray2}$ ) with tuned spatial parameters based on the PSF (see details below). At the output of the edge detection method, we get a rough fiber segmentation  $BW_{DNA2}$ .

Large clusters of overlapping DNA molecules cannot be adequately analysed and therefore need to be removed of  $BW_{DNA2}$ . For each pixel  $(x,y)$  of  $BW_{DNA2}$ , we first calculate its foreground local pixel density  $d(x,y)$  (i.e. density of foreground pixels in a kernel of  $EOP_1 \cdot PSF \times EOP_1 \cdot PSF$  around  $(x,y)$ ). To get rid of the influence of the Point Spread Function on

$d(x,y)$ , this metric is normalised by the PSF value:  $d_N(x,y) = \frac{d(x,y)}{PSF}$ . Therefore, a threshold independent from the imaging system can be experimentally set: zones where  $d_N < EOP_2$  are discarded for fiber analysis (Figure 3.6-C).

A second filter removes objects of width larger than  $EOP_3$  times the PSF. More precisely, we discard objects for which the minor axis of the ellipse that has the same normalized second central moments are larger than such threshold.

### Edge detection method

The edge detection method aims at roughly segmenting DNA fibers. It is applied twice in our whole algorithm: the first time to compute  $I_{gray_1}$  during the pre-processing step with a default PSF ( $PSF = 2$  pixels), and the second time to obtain  $I_{gray_2}$  during the first fiber segmentation step using the measured value of the PSF.

This edge detection method is made up of two parts. Briefly, we first use an edge detection that produces many false positives and then use a very selective edge detection that yields numerous false negatives. Images with false positives and false negatives are then combined.

First, the input image ( $I_{gray_i}$  with  $i=1$  or  $2$ ) is convolved with a Laplacian of Gaussian (LoG) filter, whose standard deviation is set to  $EOP_4 \cdot PSF$ . Edge pixels, defined as zero-crossing pixels, are closed 8-connected contours delimiting fibers that we fill. The resulting binary image (Figure 3.7-B) called  $I_{LoG}$  contains a high number of false positive contours originated from noise and debris.

Next, a more selective contour detection is performed. We compute two smooth gradients of  $I_{gray_i}$  in x and y direction:  $\nabla_x I_{gray_i}, \nabla_y I_{gray_i}$ . Those smooth gradients are obtained by convolving  $I_{gray_i}$  with a 1D-derivative of gaussian ( $\delta = PSF$ ). Then, the gradient modulus M is calculated:  $M = \sqrt{\nabla_x I_{gray_i}^2 + \nabla_y I_{gray_i}^2}$ . Finally, a pixel P of  $I_{gray_i}$  is a contour if it fulfills two conditions: (i) its gradient magnitude  $M(P)$  is bigger than Otsu's threshold  $t_{otsu}$  applied on M, and (ii) P is a maximum in the gradient direction  $\theta = \arctan(\frac{\nabla_y I_{gray_i}}{\nabla_x I_{gray_i}})$ . Note that this selective contour detection is equivalent to Canny edge detection with both thresholds equal to  $t_{otsu}$ . The contours obtained with this selective method are morphologically dilated with a disk of diameter  $\frac{PSF}{EOP_5}$ . The resulting binary image is called  $I_{Canny}$ .

We combine these two binary images ( $I_{Canny}$  and  $I_{LoG}$ ) by deleting all objects from  $I_{LoG}$  that have no intersection with  $I_{Canny}$ . The resulting image ( $BW_{DNA_i}$ ) is the output of the edge detection method.



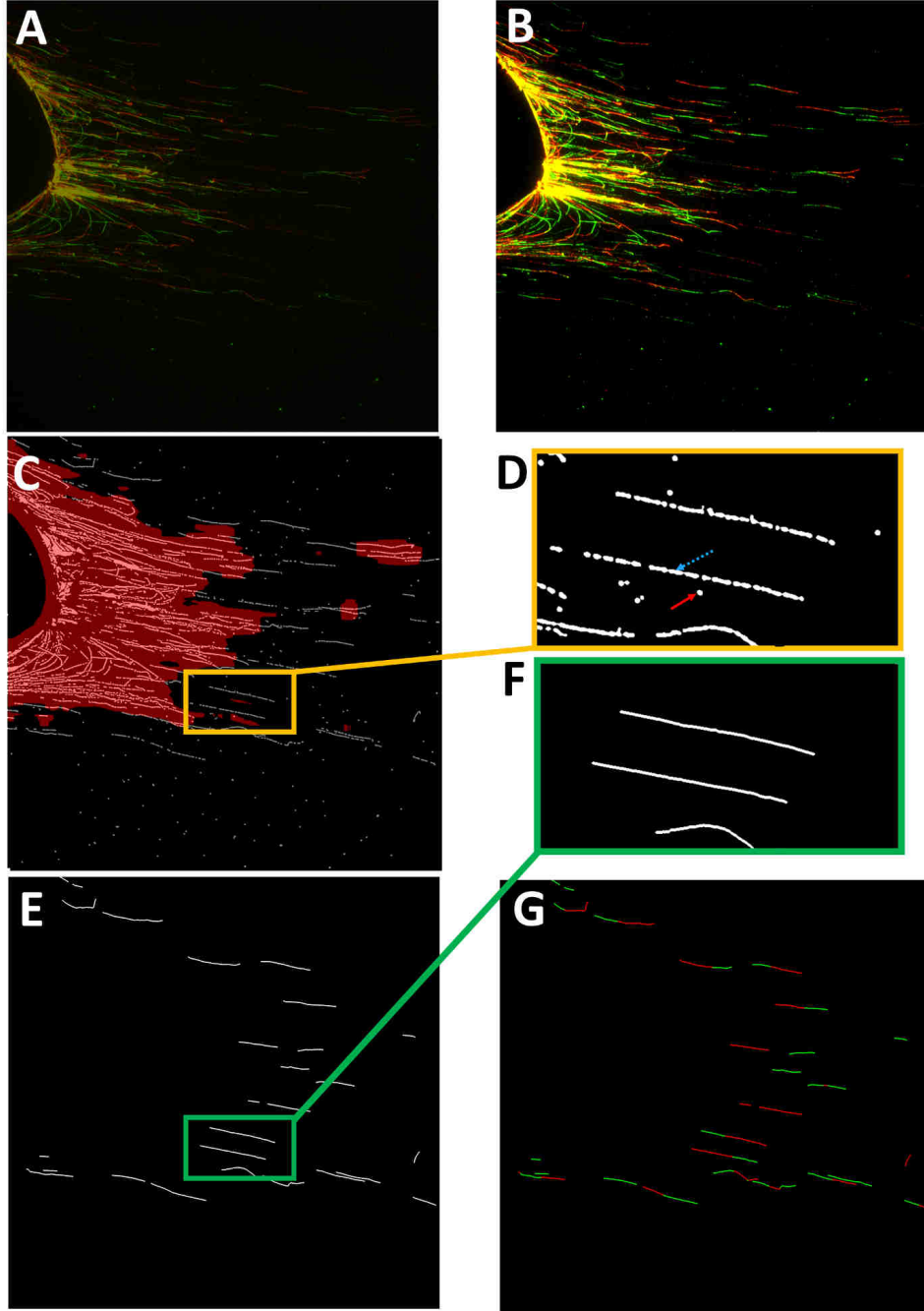


Figure 3.6 From the raw image to the final segmentation. A- Raw image [  $I_1$  : red channel (CldU),  $I_2$  : green channel (IdU) ]. B- Color enhancement after the preprocessing step [  $I_{1N}$ ,  $I_{2N}$  ]. C- Superposition of  $BW_{DNA2}$  in white (segmentation of DNA fibers with the tuned edge detection method) and unexploitable high density areas in red obtained by thresholding the local white pixel normalized density  $dN$ . D- Zoom of the yellow rectangle in C. DNA Fibers are fragmented. An example of blob and an example of strand are respectively flagged by a red full line arrow and a blue dotted arrow. E-  $SKEL_{DNA}$ : Result of the splicing method. F- Zoom of the green rectangle in E. Fragmented fibers have been spliced. G- Color assignment: analysis of the red and green channels of the color enhanced image (B)

## Fiber splicing

The main drawback of this first segmentation ( $BW_{DNA_2}$ ) is the frequent fragmentation of DNA fibers (Figure 3.6-C, D). A splicing method is thereby necessary to reconnect portions of the same DNA fiber (Figure 3.6-E, F). Briefly, large objects of  $BW_{DNA_2}$  are successively spliced with nearby objects if several continuity criteria (based on distance and fiber orientation) are fulfilled.

All objects of  $BW_{DNA_2}$  are classed in two different groups: Blobs and Strands (Figure 3.6-D). A Blob is a small and compact object which can be modeled by an ellipse with a relatively low eccentricity. In  $BW_{DNA_2}$ , even though many blobs are the consequence of noise in the original image, some of them are portions of a longer fiber. On the other hand, a strand is a longer curvilinear object that is a fraction or the totality of a DNA fiber. Practically, a strand fulfills two criteria:

- (i) High eccentricity  $e$ :  $e > \frac{\sqrt{15}}{4} \approx 0.968$  (ie. the ratio of the large axis of the ellipse over the small axis has to be higher than 4)
- (ii) Low solidity  $s$ :  $s = \frac{\text{Object Area}}{\text{Area Of Its Convex Hull}} < 0.7$ .
- (iii) Minimum length  $l$ :  $l > EOP_6 \cdot PSF$

All other objects are blobs.

Strands are skeletonized by applying successive morphological erosions. For each strand, we store their pixel coordinates (x, y), the position of their two endpoints (EP) and the orientation of the tangents at each one of the two endpoints (Figure 3.8-A). Such two tangents are computed after smoothing the coordinates (x,y) using a local least square regression with a 1<sup>st</sup> degree polynomial model spanning a length of  $EOP_7 \cdot PSF$ .

For each blob, only the coordinates  $(x_c, y_c)$  of its centroid are stored.

With all this data, a graph G is built (Figure 3.9) where each object (strand or blob) is a node. The edges of G link objects whose EP are separated by a distance D inferior to  $D_{Max}$  ( $D_{Max} = EOP_8 \cdot PSF$ ). Actually, an edge symbolises a potential connection between the EPs of two different objects. Each edge is characterised by a doublet  $edge = (s, D)$  where s is a score based on the angular continuity of the potential connection and d is the distance between the two objects. Those two parameters (s and D) will be used to rank all potential connections from a given strand.

Calculation of the score s

- **Connection between two strands:**

Figure 3.8-A shows the different parameters used in the calculation of the score.  $\theta_{EP1}$  and  $\theta_{EP2}$  are the orientations of the tangents for both endpoints EP.  $\theta_1$  and  $\theta_2$  are the angles between the tangents and the segment connecting EP1 to EP2. We define  $\Delta\theta_{tan} = \theta_{tan2} - \theta_{tan1} - \pi$ . If the connection is continuous,  $\Delta\theta_{tan}, \theta_1$  and  $\theta_2$  should be minimal. We also define  $\theta_{Max} = \max(\Delta\theta_{tan}, \theta_1, \theta_2)$ . The value of  $s$  from 1 to 5 is given according to the value of  $\theta_{Max}$  (see 3.2). The advantage of splitting the connections into five discrete classes rather than assigning a continuous score value is that potential connections with similar  $\theta_{Max}$  have equal scores. The distance parameter  $D$  will be used to rank the connections with the same score  $s$ .

- **Connection between a strand and a blob:**

Figure 3.8-B illustrates this configuration. Here,  $\theta_{Max}$  is the angle between the tangent of the strand and the segment that links the endpoint and the centroid of the blob.

- **Connection between two blobs:**

As the splicing process begins from a strand and iteratively merges nearby objects (blobs or strands), we never consider blob-blob edges.

Table 3.2 Score  $s$  with respect to  $\theta_{Max}$ . The lower  $\theta_{Max}$  is, the higher the angular score  $s$  is.

Value of $\theta_{Max}$ (in degree)	Angular Score $s$
$\theta_{Max} \leq 2$	5
$2 < \theta_{Max} \leq 5$	4
$5 < \theta_{Max} \leq 10$	3
$10 < \theta_{Max} \leq 15$	2
$15 < \theta_{Max}$	1

The following splicing procedure is applied iteratively from the longest to the smallest strand of the graph. Let  $str_i$  be the  $i$ th strand processed. First, all edges connected to  $str_i$  are sorted in decreasing order of score  $s$ . Strands with equal  $s$  are ranked in decreasing order of distance  $D$ . Let's note  $edge_1 = (s_1, D_1)$  the first edge of the ranking. This edge links  $str_i$  to another object that we call  $obj_1$ . We splice the strand  $str_i$  to  $obj_1$  if two conditions are fulfilled: (i)  $s_1 \leq 4$ , (ii)  $obj_1$  is not connected to a better edge (i.e. an edge whose score  $s$  is strictly higher than  $s_1$ , or  $s = s_1$  and  $D < D_1$ ). If those two conditions are not fulfilled for  $edge_1$ , we try with the following edges in the ranking.

If a candidate  $obj_k$  meeting those requirements is found,  $str_i$  merges with  $obj_k$  ( $str_i \leftarrow str_i \cup obj_k$ ): (i)  $str_i$  is linked to  $obj_k$  by a straight line, (ii) the node of  $str_i$  in the graph

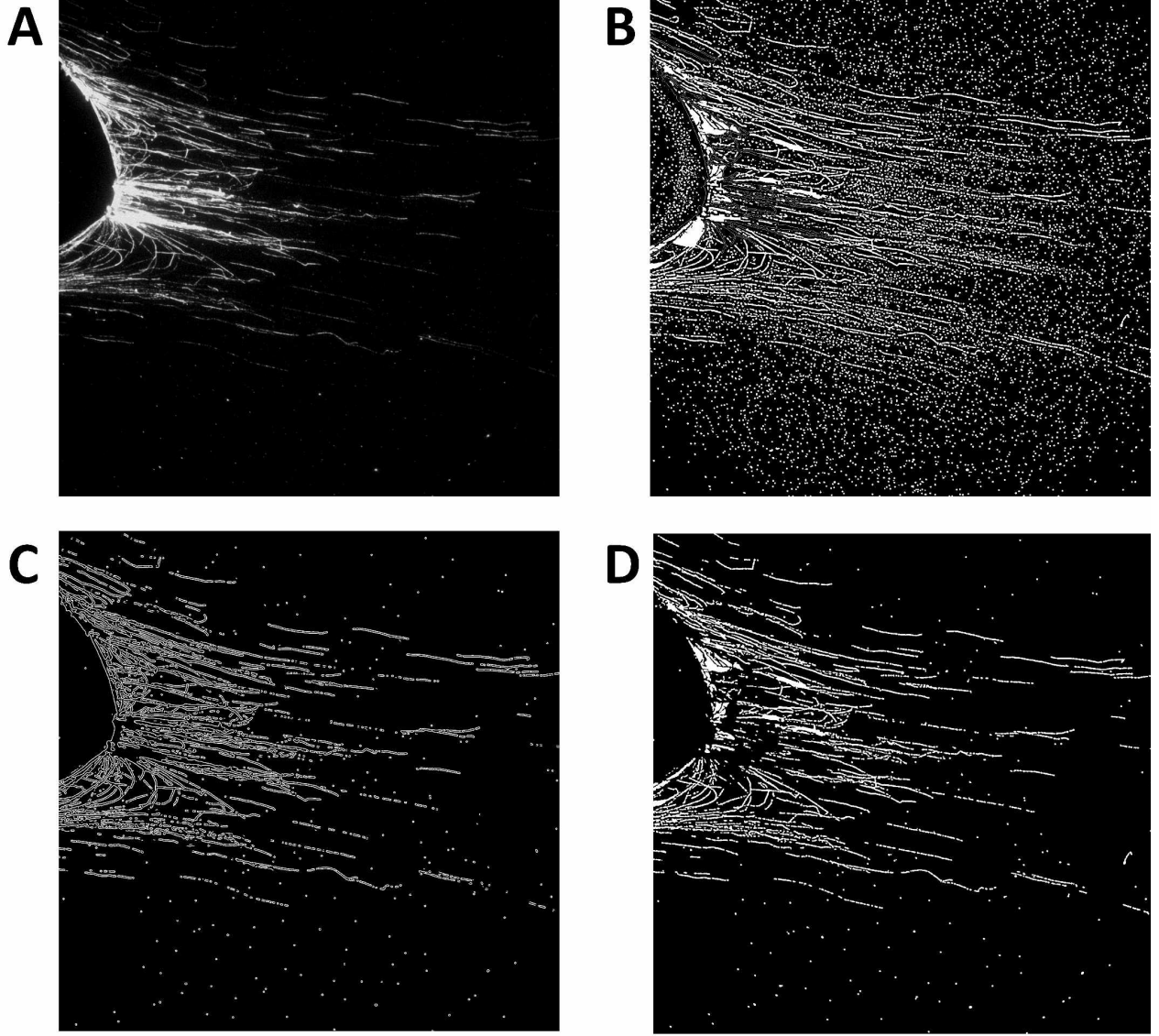


Figure 3.7 Edge detection method. A–  $I_{gray_2}$  : combination of the normalized intensity channel :  $I_{gray_2} = \frac{I_{1N} + I_{2N}}{2}$ . B–  $I_{LoG}$  : First edge detection [Laplacian of Gaussian (LoG)]. C–  $I_{Canny}$  : Second edge detection. D–  $BW_{DNA2}$  : Combination of  $I_{LoG}$  and  $I_{Canny}$  : Objects of  $I_{LoG}$  that intersect a white pixel of  $I_{Canny}$  are kept.

merges with the node of  $obj_k$ . The new node preserves the links of  $str_i$  and  $obj_k$  (except the link between both objects) but with updated value  $(s, D)$ .

At the end of this splicing step, we obtain a binary image containing the skeletons of DNA fibers:  $SKEL_{DNA}$  (Figure 3.6-E). All skeletons with pixels in high-density areas (as defined before in the first fiber segmentation) are deleted. We also remove skeletons whose length is inferior to  $l_{min}$  ( $l_{min} = EOP_9 \cdot PSF$ ) because they often are artefacts due to debris or non-specific staining.

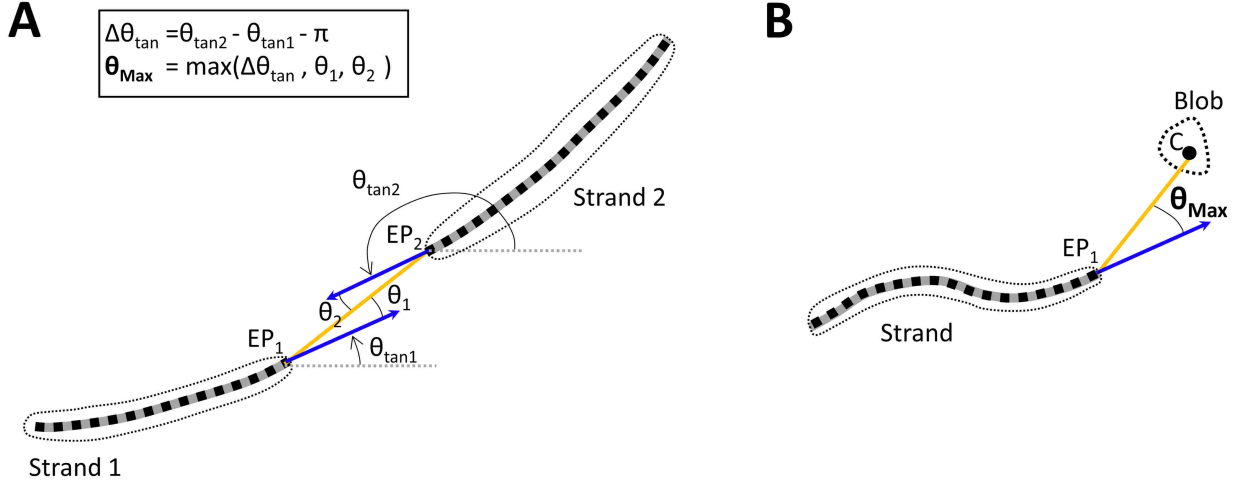


Figure 3.8 Splicing parameters. A- Connection between 2 strands.  $\theta_{tan1}$  and  $\theta_{tan2}$  are the orientation of the tangents at each endpoint (EP1, EP2) for each strand.  $\theta_1$  and  $\theta_2$  are the angles between the tangents and the connection segment [EP1, EP2] in yellow.  $\theta_{Max}$  is the maximum of  $\Delta\theta_{tan}$ ,  $\theta_1$ ,  $\theta_2$ . B- Connection between a strand and a blob.  $\theta_{Max}$  is the angle between the tangent and the connection segment [EP1,C] in yellow.

### Color assignment

Once DNA fibers are segmented and skeletonized, we estimate the color at each pixel of the skeletonized fibers by comparing intensities of each channel:  $I_{1N}$  and  $I_{2N}$ . The objective is to convert color intensities (doubles between 0 and 1) to a color label (e.g. IdU or CldU).

Each foreground pixel of  $SKEL_{DNA}$  is assigned a pair of intensities referring to the normalised fluorescence of the two nucleotide analogs in the vicinity of the pixel. To do so, each one of the two normalized intensity images  $I_{1N}$  and  $I_{2N}$  is convolved with a Gaussian kernel of standard deviation PSF, and multiplied element-wise by  $SKEL_{DNA}$ . Following the pixels

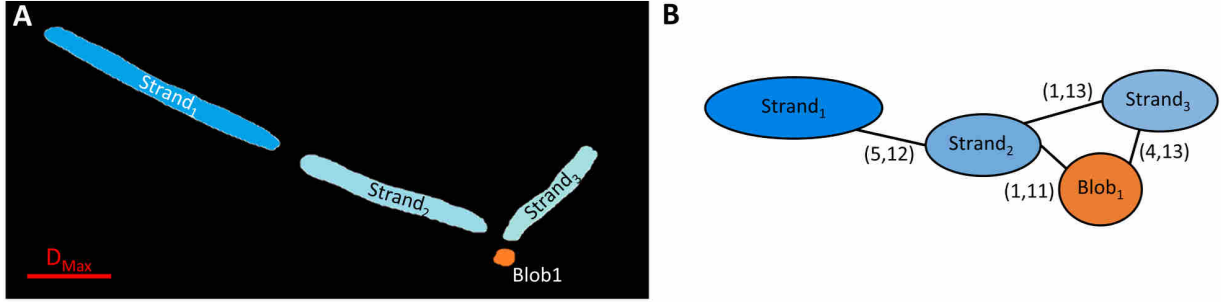


Figure 3.9 Graph of Strands and Blobs. A - Image of 3 strands (in blue) and one blob (in orange) B— Objects whose endpoints are separated by less than  $d_{Max}$  are linked by an edge  $(s,D)$ . The angular score  $s$  is an integer between 1 and 5. The parameter  $D$  is the distance between the endpoints.

of each skeleton from one endpoint to the other, it is possible to define two color intensity profiles:  $S_1$  (intensity profile of the first nucleotide analog) and  $S_2$  (intensity profile of the second one).

We compute the color difference function as  $\Delta S = S_1 - S_2$ , which can be interpreted as the difference between the normalized fluorescence of both channels. The zero-crossing of the function  $\Delta S$  are computed to partition fibers into segments of a predominant nucleotide analog. Colors are assigned in two steps:

1. All segments where  $|mean(\Delta S)|$  is larger than a threshold (empirically set to 7%), are assigned to the predominant nucleotide analog color.
2. Each remaining section is assigned as follows:
  - When the segment is surrounded by neighbours of the same color, the same color is assigned.
  - When the segment is surrounded by neighbours of different colors, the segment is split in halves and colors are assigned to match the color of neighbours.
  - When the segment is located at the end of a fiber, the neighbour segment color is assigned.

Thus, each skeleton is partitioned in sections of different colors. If some sections are too small ( $EOP_{10} \cdot PSF$ ) they are considered as mistakes: the color label of those sections is inverted so that they are merged with their two neighbours.

Table 3.3 List of Experimentally Optimized Parameter. All these parameters are set as multiplication factors of the PSF of the imaging system

Parameter name	Value	Description
$EOP_1$	54	Sets the size of the area used to calculate fiber density
$EOP_2$	0.0185	Normalized density threshold
$EOP_3$	3.6	Sets the maximum fiber width
$EOP_4$	1	Sets the kernel size of the Laplacian of Gaussian filter
$EOP_5$	2	Sets the length of the structural element used for dilating Canny's edges
$EOP_6$	9	Set the minimum length for a strand
$EOP_7$	6	Sets the spanning length of the 1st degree polynomial for smoothing fiber skeletons
$EOP_8$	13.5	Sets the maximum splicing distance
$EOP_9$	15	Sets the minimum length for fibers after splicing
$EOP_{10}$	6	Sets the minimum length of a pulse section within a fiber

### 3.4 Discussion

Fluorescent imaging of DNA fibers is widely used to study the dynamics of RF progression, as a proxy for several aspects of genomic stability and replicative stress. The vast majority of studies using this technology are based on manual segmentation of fibers, using simple image processing software to facilitate record keeping and annotation. Our results demonstrate a significant degree of variability when manually measuring DNA fibers, which compromises data reproducibility and renders analyses prone to bias. The automated segmentation method that we present here (ADFA) is devoid of subjectivity and enables rapid analysis of large image databases, thereby increasing statistical power.

The proposed framework is robust and can be adapted to various experimental conditions. Indeed, the use of the Point Spread Function, which is computed for each image using fiber width auto-calibrates free parameters of the algorithm. In addition, this characteristic metric enables removal from the analysis of anomalous objects such as small nonspecific stainings, unusually large fibers and areas characterized by excessively high fiber density.

ADFA was tested on a database of different images of varying quality, signal to noise ratio, or fiber length which were acquired from two different microscopes. Segmenting fibers in such images represents a complex task because (i) fiber fluorescence is not homogeneous, (ii) fibers are split in several segments, tangled in clusters, and/or mixed with nonspecific stained objects. We note that our algorithm was implemented specifically for the analysis of DNA

fibers and was not optimized for DNA combing images. In our experience, the latter generally present a much larger number of split DNA segments, which renders adequate joining and segmentation of labeled tracks challenging. Further optimization of the ADFA algorithm will therefore be required to make it reliable for DNA combing analysis.

We observed that in the case of very long fibers split in a high number of segments or blobs, our algorithm can yield aberrant results. This problem was highlighted in the experiment where the incubation time is varied from 10 to 90 minutes, where both manual and automatic segmentation fail to detect the extension of fibers for long incubation time (i.e. when fibers are very long; Figure 3.5). Examination of the images revealed that in the case of ADFA, the above difference is the consequence of mistakes during the splicing stage of the algorithm which fails to connect all strands and blobs belonging to very long fibers. As second pulse tracks are expected to be much longer than contiguous first pulse ones, their ratio is often underestimated. While our results suggest that such images also cannot be reliably quantified manually, artifacts and mistakes appear to be worse with ADFA. We also note that we observed larger standard deviation in ADFA measurements as compared to manual quantification. A small fraction of the outliers measured only by ADFA are probably segmentation errors, but a significant portion are not. Indeed, careful inspection of manual segmentations showed that users tend to preferentially select fibers of similar lengths within a given image, and avoid very long or short fibers. To improve accuracy, we propose that two main experimental parameters should be optimized for automated analysis: (i) incubation duration should be kept relatively short to reduce fiber length, and ii) dilution of fibers on the slide might reduce clustering of fluorescent DNA molecules.

The open-source implementation we provide was programmed using Matlab. We also made available a free a compiled version of the code which can be used by investigators without programming experience or access to this commercial language. The output of ADFA consists of three images outlining the results and a spreadsheet with single fiber details. Images show 1) which of the fibers were chosen for analysis, 2) skeleton versions of such fibers with a tag that allows identification in the spreadsheet, and 3) the ratio (second/first pulse) for each one of them. The spreadsheet contains all necessary information for statistical analysis: the file name of origin, the fiber tag, the combination of colors found, and the length of each nucleotide label. ADFA is an algorithm that greatly facilitates investigations on the dynamics of DNA replication by automatically measuring the length of fluorescently labelled DNA fibers. Contrary to manual techniques, ADFA is devoid of inter/intra user variability. Our algorithm should therefore be useful to reduce the tediousness, bias, and poor reproducibility associated with manual quantification of DNA fiber length.



## CHAPITRE 4 DISCUSSION GÉNÉRALE

La segmentation des brins d'ADN est une tâche complexe. En effet, les images sont caractérisées par des enchevêtrements de fibres non exploitables, des brins fragmentés en petites structures qu'il faut fusionner, la présence d'un bruit de fond causé par un appariement non spécifique des anticorps.

Jusqu'à présent, la quantification des fibres d'ADN se faisait manuellement. La comparaison de la segmentation manuelle de trois opérateurs expérimentés a montré des variations statistiquement significatives. Effectivement, le choix des fibres et leur segmentation sont des critères très subjectifs. L'avantage majeur de notre algorithme, au delà du gain de temps considérable qu'il procure, est de s'affranchir de ces biais inter-opérateurs.

Pour segmenter automatiquement les brins d'ADN, nous avons privilégié une approche par analyse de gradient (Marr and Hildreth, 1980; Canny, 1986) pour plusieurs raisons : Premièrement, comme les brins d'ADN ont tendance à être fragmentés en pointillés peu filiformes, les métriques dérivées de la hessienne (eg. métriques de Frangi) nous paraissaient peu appropriées. Deuxièmement, les contours des brins d'ADN sont la plupart du temps clairement définis bien qu'ils peuvent être noyés au sein de petites tâches de coloration non spécifique (non specific staining). Une analyse de la distribution du gradient dans l'image permet d'établir des seuils adaptatifs afin de supprimer les structures bruitées en arrière plan tout en conservant les structures d'intérêt. L'utilisation de la segmentation de Marr-Hildreth avec un seuil de 0 nous permet d'extraire des contours fermés et donc des objets. Enfin, bien que les autres techniques d'extraction d'objets curvilignes (tensor voting, orientation score, RORPO ...) sont très attrayantes du point de vue théorique, nous avons préféré nous concentrer sur des outils plus simples. Les performances réelles des autres algorithmes sont difficiles à évaluer objectivement (cf. section 2.2.9). La plupart de ces méthodes ne sont pas disponibles en libre accès ce qui rend leur validation d'autant plus difficile.

Pour prolonger les discontinuités de la segmentation obtenue par analyse du gradient, nous avons développé une approche de suivi reposant sur l'analyse de l'orientation et de la continuité des objets adjacents. Cette approche reconstruit l'intégralité d'un brin global en partant d'un sous-fragment puis en l'étendant de manière itérative. La principale limite de cette méthode d'accroissement successif est son confinement à une analyse très locale des discontinuités sans comparer les configurations globales des objets comme c'est le cas pour les méthodes de chemin minimaux, de tensor voting ou de score d'orientation. Néanmoins, cette méthode permet d'établir des critères de liaison simples, facilement modifiables par

l'utilisateur.

Notre algorithme permet aussi l'identification des sections IdU et CldU par analyse des canaux vert et rouge de l'image. Cette tâche est parfois difficile car malgré les étapes de rinçage, les deux types d'analogues de nucléotides peuvent coexister ce qui engendre des brins aux couleurs hybrides (jaune orangé). Pour pallier ces problèmes nous proposons une normalisation des canaux rouge et vert. Cette normalisation est effectuée par rapport au 95<sup>e</sup> percentile de l'intensité des fibres (en excluant l'arrière plan). Ensuite, l'analyse du profil d'intensité le long du brin permet d'extraire des régions où un des deux canaux est largement prédominant et des sections où la couleur est hybride. Pour ces dernières régions, on attribue le label (CldU ou IdU) par analyse des sections adjacentes.

Notre algorithme s'adapte à chaque image durant l'étape de pré-traitement. En effet, l'ensemble de nos paramètres morphologiques (écart-type des filtres de convolution, distance maximale entre 2 brins, longueur minimale d'un segment) n'est pas codé "en dur" dans l'algorithme. Ces paramètres dépendent de la largeur de la réponse impulsionnelle du système d'acquisition (PSF). Cette PSF est estimée pendant l'étape de pré-traitement, à partir du calcul de la largeur moyenne des fibres modélisées par un profil gaussien.

Plusieurs expériences ont permis d'évaluer les performances de notre algorithme. Tout d'abord, il est comparé à la segmentation manuelle de plusieurs opérateurs à l'échelle d'une image (comparaison de la médiane des rapports CldU/IdU pour 98 images) et à l'échelle de chaque fibre (comparaison de la différence des longueurs des fibres segmentées en commun). Ces deux expériences montrent une haute corrélation entre la segmentation manuelle et automatique et des différences similaires au sein des opérateurs humains et entre les humains et la machine. Ensuite, nous avons réalisé trois expériences où l'on compare la dynamique de répllication de l'ADN entre un échantillon soumis à un stress réplicatif et un échantillon témoin. Dans chacun des cas, les mesures de l'algorithme sont conformes aux résultats attendus et sont confirmées par une seconde quantification manuelle. Enfin, nous avons fait une expérience où nous faisons varier le temps d'incubation du second pulse d'analogues de nucléotide entre 10 et 90 minutes. Notre algorithme mesure un accroissement global des rapports CldU/IdU entre 10 et 60 minutes, puis une saturation. Ces observations seront aussi confirmées par la quantification manuelle.

Malgré ces bons résultats, notre approche comporte des limites. Plus les fibres d'ADN sont longues et fragmentées, plus la méthode de prolongement des discontinuités peine à relier l'ensemble des parties de la structure globale. La grande variabilité dans l'agencement des discontinuités (de l'accumulation rectiligne de points aux croisements de longues fibres) ne nous a pas permis d'établir un modèle de connexion efficace tout le temps. D'autres erreurs

sont dûes à de fortes disparités d'intensité et de contexte au sein d'une fibre : alors qu'une section est largement visible, une autre a une intensité à peine détectable ou est masquée dans un enchevêtrement d'autres fibres. Contrairement à l'opérateur humain, notre algorithme n'est pas capable de rejeter ces fibres dont seulement une partie est camouflée.

Certains critères expérimentaux permettent d'optimiser les performances de notre algorithme :

- Temps d'incubation des analogues de nucléotides : Favoriser les temps courts permet de réduire le nombre de discontinuités au sein d'une fibre et de réduire les enchevêtrements. Cela entraîne une diminution de la probabilité de commettre des erreurs lors de la prolongation des discontinuités
- Concentration des fibres sur la lame : Des fibres espacées réduisent les difficultés dûes aux croisements et enchevêtrements.

## CHAPITRE 5 CONCLUSION ET RECOMMANDATIONS

### 5.1 Synthèse des travaux

A partir d'une image de fibres d'ADN, notre algorithme segmente automatiquement les brins fluorescents et quantifie pour chacun d'entre eux la longueur des segments pulsés d'IdU et de CldU. Cette méthode se base sur (i) une combinaison de l'approche de Marr and Hildreth (1980) et de Canny (1986) pour segmenter les fragments fluorescents de l'image, (ii) une méthode de prolongement des discontinuités de la segmentation précédente par analyse de l'orientation des structures adjacentes, (iii) une analyse des canaux rouge et vert de l'image pour détecter les deux pulses d'analogues de nucléotide (IdU et CldU).

Nous avons testé notre algorithme sur de nombreuses images de qualité variable dans différentes conditions de stress réplicatif et en le comparant à plusieurs opérateurs humains. Les résultats montrent une haute corrélation entre la segmentation manuelle et automatique ainsi que des différences significatives entre échantillons témoins et échantillons soumis à un stress.

Notre algorithme disponible en libre accès pourra être largement distribué afin de servir de référence non sujette aux variabilités inter et intra opérateurs.

### 5.2 Améliorations futures

Dans certaines expériences, des anticorps anti-ssDNA sont utilisés pour rendre fluorescent l'intégralité des brins d'ADN. Cette coloration pourrait être ajoutée comme troisième canal à l'image et ainsi servir de support supplémentaire à la détection des brins d'ADN et plus particulièrement à la prolongation des discontinuités.

Notre approche a utilisé des outils simples pour prolonger les discontinuités (analyse locale de l'orientation des objets adjacents). Il serait intéressant d'intégrer à notre approche des outils plus évolués comme les méthodes par score d'orientation ou par vote de tenseur.

Enfin, notre algorithme a été optimisée pour traiter les images de fibres d'ADN (DNA spreading) et non pas les images obtenues par peignage de l'ADN (DNA combing). Cette dernière technique, moins utilisée a l'avantage de disposer les brins d'ADN de manière rectiligne et parallèle mais présente un niveau de bruit plus important. Il serait intéressant d'adapter notre algorithme à ce type d'imagerie.

## RÉFÉRENCES

- S. Abbasi-Sureshjani, M. Favali, G. Citti, A. Sarti, et B. M. t. H. Romeny, “Curvature Integration in a 5D Kernel for Extracting Vessel Connections in Retinal Images”, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 27, no. 2, pp. 606–621, FEB 2018. DOI: 10.1109/TIP.2017.2761543
- J. L. Alexander et T. L. Orr-Weaver, “Replication fork instability and the consequences of fork collisions from rereplication”, *GENES & DEVELOPMENT*, vol. 30, no. 20, pp. 2241–2252, OCT 15 2016. DOI: 10.1101/gad.288142.116
- K. Alsaih, G. Lemaitre, J. M. Vall, M. Rastgoo, D. Sidibe, T. Y. Wong, E. Lamoureux, D. Milea, C. Y. Cheung, et F. Meriaudeau, “Classification of SD-OCT volumes with multi pyramids, LBP and HOG descriptors: application to DME detections”, dans *2016 38TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC)*, série IEEE Engineering in Medicine and Biology Society Conference Proceedings, Patton, J and Barbieri, R and Ji, J and Jabbari, E and Dokos, S and Mukkamala, R and Guiraud, D and Jovanov, E and Dhaher, Y and Panescu, D and Vangils, M and Wheeler, B and Dhawan, AP, éd. IEEE Engn Med & Biol Soc, 2016, pp. 1344–1347, 38th Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC), Orlando, FL, AUG 16-20, 2016.
- R. Annunziata et E. Trucco, “Accelerating Convolutional Sparse Coding for Curvilinear Structures Segmentation by Refining SCIRD-TS Filter Banks”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 35, no. 11, pp. 2381–2392, NOV 2016. DOI: 10.1109/TMI.2016.2570123
- R. Annunziata, A. Kheirkhah, P. Hamrah, et E. Trucco, “Boosting Hand-Crafted Features for Curvilinear Structure Segmentation by Learning Context Filters”, *MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION, PT III*, vol. 9351, pp. 596–603, 2015, 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, GERMANY, OCT 05-09, 2015. DOI: 10.1007/978-3-319-24574-4\_71
- L. Ballerini, R. Lovreglio, M. d. C. V. Hernandez, J. Ramirez, B. J. MacIntosh, S. E. Black, et J. M. Wardlaw, “Perivascular Spaces Segmentation in Brain MRI Using Optimal 3D Fil-

tering”, *SCIENTIFIC REPORTS*, vol. 8, FEB 1 2018. DOI: 10.1038/s41598-018-19781-5

K. A. Barella, V. P. Costa, V. G. Vidotti, F. R. Silva, M. Dias, et E. S. Gomi, “Glaucoma Diagnostic Accuracy of Machine Learning Classifiers Using Retinal Nerve Fiber Layer and Optic Nerve Data from SD-OCT”, *JOURNAL OF OPHTHALMOLOGY*, 2013. DOI: 10.1155/2013/789129

C. Becker, R. Rigamonti, V. Lepetit, et P. Fua, “Supervised Feature Learning for Curvilinear Structure Segmentation”, dans *MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION (MICCAI 2013), PT I*, série Lecture Notes in Computer Science, Mori, K and Sakuma, I and Sato, Y and Barillot, C and Navab, N, éd., vol. 8149. Nagoya Convent & Visitors Bur; Murata Sci Fdn; Daiko Fdn; Japan Soc Comp Aided Surg; Sci Council Japan; Nagoya Univ, Informat & Commun Headquarters; Nagoya Univ, Grad Sch Informat Sci, 2013, pp. 526–533, 16th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Nagoya Univ, Nagoya, JAPAN, SEP 22-26, 2013.

E. Bekkers, R. Duits, T. Berendschot, et B. t. H. Romeny, “A Multi-Orientation Analysis Approach to Retinal Vessel Tracking”, *JOURNAL OF MATHEMATICAL IMAGING AND VISION*, vol. 49, no. 3, pp. 583–610, JUL 2014. DOI: 10.1007/s10851-013-0488-6

F. Belanger, E. Fortier, M. Dube, J.-F. Lemay, R. Buisson, J.-Y. Masson, A. Elsherbiny, S. Costantino, E. Carmona, A.-M. Mes-Masson, H. Wurtele, et E. Drobetsky, “Replication Protein A Availability during DNA Replication Stress Is a Major Determinant of Cisplatin Resistance in Ovarian Cancer Cells”, *CANCER RESEARCH*, vol. 78, no. 19, pp. 5561–5573, OCT 1 2018. DOI: 10.1158/0008-5472.CAN-18-0618

F. Benmansour et L. D. Cohen, “Tubular Structure Segmentation Based on Minimal Path Method and Anisotropic Enhancement”, *INTERNATIONAL JOURNAL OF COMPUTER VISION*, vol. 92, no. 2, pp. 192–210, APR 2011. DOI: 10.1007/s11263-010-0331-0

A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot, et D. Bensimon, “ALIGNMENT AND SENSITIVE DETECTION OF DNA BY A MOVING INTERFACE”, *SCIENCE*, vol. 265, no. 5181, pp. 2096–2098, SEP 30 1994. DOI: 10.1126/science.7522347

D. Bensimon, A. Simon N, V. Croquette, et A. Bensimon, “STRETCHING DNA WITH A RECEDING MENISCUS - EXPERIMENTS AND MODELS”, *PHYSICAL REVIEW LETTERS*, vol. 74, no. 23, pp. 4754–4757, JUN 5 1995. DOI: 10.1103/PhysRevLett.74.4754

P. Bibiloni, M. Gonzalez-Hidalgo, et S. Massanet, “A survey on curvilinear object segmentation in multiple applications”, *PATTERN RECOGNITION*, vol. 60, pp. 949–970, DEC 2016. DOI: 10.1016/j.patcog.2016.07.023

D. Bizios, A. Heijl, J. L. Hougaard, et B. Bengtsson, “Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT”, *ACTA OPHTHALMOLOGICA*, vol. 88, no. 1, pp. 44–52, FEB 2010. DOI: 10.1111/j.1755-3768.2009.01784.x

M. L. Bochman, K. Paeschke, et V. A. Zakian, “DNA secondary structures: stability and function of G-quadruplex structures”, *NATURE REVIEWS GENETICS*, vol. 13, no. 11, pp. 770–780, NOV 2012. DOI: 10.1038/nrg3296

B. Bouraoui, C. Ronse, J. Baruthio, N. Passat, et P. Germain, “3D segmentation of coronary arteries based on advanced mathematical morphology techniques”, *COMPUTERIZED MEDICAL IMAGING AND GRAPHICS*, vol. 34, no. 5, pp. 377–387, JUL 2010. DOI: 10.1016/j.compmedimag.2010.01.001

T. Byun, M. Pacek, M. Yee, J. Walter, et K. Cimprich, “Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint”, *GENES & DEVELOPMENT*, vol. 19, no. 9, pp. 1040–1052, MAY 1 2005. DOI: 10.1101/gad.1301205

J. Canny, “A computational approach to edge detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov 1986. DOI: 10.1109/TPAMI.1986.4767851

A. R. Chaudhuri, E. Callen, X. Ding, E. Gogola, A. A. Duarte, J.-E. Lee, N. Wong, V. Lafarga, J. A. Calvo, N. J. Panzarino, S. John, A. Day, A. V. Crespo, B. Shen, L. M. Starnes, J. R. de Ruiter, J. A. Daniel, P. A. Konstantinopoulos, D. Cortez, S. B. Cantor, O. Fernandez-Capetillo, K. Ge, J. Jonkers, S. Rottenberg, S. K. Sharan, et A. Nussenzweig, “Replication fork stability confers chemoresistance in BRCA-deficient cells”, *NATURE*, vol. 535, no. 7612, pp. 382+, JUL 21 2016. DOI: 10.1038/nature18325

Y. Chen, Y. Zhang, J. Yang, Q. Cao, G. Yang, J. Chen, H. Shu, L. Luo, J.-L. Coatrieux, et Q. Feng, “Curve-Like Structure Extraction Using Minimal Path Propagation With Backtracking”, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 25, no. 2, pp. 988–1003, FEB 2016. DOI: 10.1109/TIP.2015.2496279

A. Christodoulidis, T. Hurtut, H. Ben Tahar, et F. Cheriet, “A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images”, *COMPUTERIZED MEDICAL IMAGING AND GRAPHICS*, vol. 52, pp. 28–43, SEP 2016. DOI: 10.1016/j.compmedimag.2016.06.001

A. Ciccia et S. J. Elledge, “The DNA Damage Response: Making It Safe to Play with Knives”, *MOLECULAR CELL*, vol. 40, no. 2, SI, pp. 179–204, OCT 22 2010. DOI: 10.1016/j.molcel.2010.09.019

F. Cottini, T. Hideshima, G. Tonon, et K. C. Anderson, “The Oncogene MYC Triggers Replicative Stress and DNA Damage In Multiple Myeloma”, *BLOOD*, vol. 122, no. 21, NOV 15 2013, 55th Annual Meeting of the American-Society-of-Hematology, New Orleans, LA, DEC 07-10, 2013.

J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O’Donoghue, D. Visentin, G. van den Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, C. O. Hughes, R. Raine, J. Hughes, D. A. Sim, C. Egan, A. Tufail, H. Montgomery, D. Hassabis, G. Rees, T. Back, P. T. Khaw, M. Suleyman, J. Cornebise, P. A. Keane, et O. Ronneberger, “Clinically applicable deep learning for diagnosis and referral in retinal disease”, *NATURE MEDICINE*, vol. 24, no. 9, pp. 1342+, SEP 2018. DOI: 10.1038/s41591-018-0107-6

P. A. Dufour, L. Ceklic, H. Abdillahi, S. Schroeder, S. De Dzanet, U. Wolf-Schnurrbusch, et J. Kowal, “Graph-Based Multi-Surface Segmentation of OCT Data Using Trained Hard and Soft Constraints”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 32, no. 3, pp. 531–543, MAR 2013. DOI: 10.1109/TMI.2012.2225152

R. Duits, M. Duits, M. van Almsick, et B. ter Haar Romeny, “Invertible orientation scores as an application of generalized wavelet theory”, *Pattern Recognition and Image Analysis*, vol. 17, no. 3, pp. 438–438, Sep 2007. DOI: 10.1134/S105466180703011X. En ligne: <https://doi.org/10.1134/S105466180703011X>

J. Fan, J. Yang, Y. Wang, S. Yang, D. Ai, Y. Huang, H. Song, A. Hao, et Y. Wang, “Multichannel Fully Convolutional Network for Coronary Artery Segmentation in X-Ray Angiograms”, *IEEE ACCESS*, vol. 6, pp. 44 635–44 643, 2018. DOI: 10.1109/ACCESS.2018.2864592



L. Fang, C. Wang, D. Cunefer, R. H. Guymer, et S. Farsiu, “Automatic segmentation of nine layer boundaries in OCT images using convolutional neural networks and graph search”, *INVESTIGATIVE OPHTHALMOLOGY & VISUAL SCIENCE*, vol. 58, no. 8, JUN 2017, Annual Meeting of the Association-for-Research-in-Vision-and-Ophthalmology (ARVO), Baltimore, MD, MAY 07-11, 2017.

N. D. Forkert, A. Schmidt-Richberg, J. Fiehler, T. Illies, D. Moeller, H. Handels, et D. Saering, “Automatic Correction of Gaps in Cerebrovascular Segmentations Extracted from 3D Time-of-Flight MRA Datasets”, *METHODS OF INFORMATION IN MEDICINE*, vol. 51, no. 5, pp. 415–422, 2012. DOI: 10.3414/ME11-02-0037

A. Frangi, W. Niessen, K. Vincken, et M. Viergever, “Multiscale vessel enhancement filtering”, *1st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 98)*, CAMBRIDGE, MA, OCT 11-13, 1998, vol. 1496, pp. 130–137, 1998.

H. Fu, Y. Xu, D. W. K. Wong, et J. Liu, “RETINAL VESSEL SEGMENTATION VIA DEEP LEARNING NETWORK AND FULLY-CONNECTED CONDITIONAL RANDOM FIELDS”, *2016 IEEE 13TH INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI)*, pp. 698–701, 2016, IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, CZECH REPUBLIC, APR 13-16, 2016.

W. Gan, Z. Guan, J. Liu, T. Gui, K. Shen, J. L. Manley, et X. Li, “R-loop-mediated genomic instability is caused by impairment of replication fork progression”, *GENES & DEVELOPMENT*, vol. 25, no. 19, pp. 2041–2056, OCT 2011. DOI: 10.1101/gad.17010011

K. A. Goatman, A. D. Fleming, S. Philip, G. J. Williams, J. A. Olson, et P. F. Sharp, “Detection of New Vessels on the Optic Disc Using Retinal Photographs”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 30, no. 4, pp. 972–979, APR 2011. DOI: 10.1109/TMI.2010.2099236

C. Hennemersperger, M. Baust, P. Waelkens, A. Karamalis, S.-A. Ahmadi, et N. Navab, “Multi-Scale Tubular Structure Detection in Ultrasound Imaging”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 34, no. 1, pp. 13–26, JAN 2015. DOI: 10.1109/TMI.2014.2340912

Q. Huang, J. Sun, H. Ding, X. Wang, et G. Wang, “Robust liver vessel extraction using 3D U Net with variant dice loss function”, *COMPUTERS IN BIOLOGY AND MEDICINE*,

vol. 101, pp. 153–162, OCT 1 2018. DOI: 10.1016/j.compbiomed.2018.08.018

M. A. Hussain, A. Bhuiyan, C. D. Luu, R. T. Smith, R. H. Guymer, H. Ishikawa, J. S. Schuman, et K. Ramamohanarao, “Classification of healthy and diseased retina using SD-OCT imaging and Random Forest algorithm”, *PLOS ONE*, vol. 13, no. 6, JUN 4 2018. DOI: 10.1371/journal.pone.0198281

B. Ibragimov, D. Toesca, D. Chang, A. Koong, et L. Xing, “Combining deep learning with anatomical analysis for segmentation of the portal vein for liver SBRT planning”, *PHYSICS IN MEDICINE AND BIOLOGY*, vol. 62, no. 23, pp. 8943–8958, DEC 7 2017. DOI: 10.1088/1361-6560/aa9262

D. Jackson et A. Pombo, “Replicon clusters are stable units of chromosome structure: Evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells”, *JOURNAL OF CELL BIOLOGY*, vol. 140, no. 6, pp. 1285–1295, MAR 23 1998. DOI: 10.1083/jcb.140.6.1285

V. S. Joshi, M. K. Garvin, J. M. Reinhardt, et M. D. Abramoff, “IDENTIFICATION AND RECONNECTION OF INTERRUPTED VESSELS IN RETINAL VESSEL SEGMENTATION”, dans *2011 8TH IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING: FROM NANO TO MACRO*, série IEEE International Symposium on Biomedical Imaging. Inst Elect & Electron Engn; Engn Med & Biol Soc (EMBS); IEEE Signal Proc Soc (SPS), 2011, pp. 1416–1420, 8th IEEE International Symposium on Biomedical Imaging (ISBI) - From Nano to Macro, Chicago, IL, MAR 30-APR 02, 2011.

V. Kaul, A. Yezzi, et Y. J. Tsai, “Detecting Curves with Unknown Endpoints and Arbitrary Topology Using Minimal Paths”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 34, no. 10, pp. 1952–1965, OCT 2012. DOI: 10.1109/TPAMI.2011.267

A. Lahiri, K. Ayush, P. K. Biswas, et P. Mitra, “Generative Adversarial Learning for Reducing Manual Annotation in Semantic Segmentation on Large Scale Microscopy Images: Automated Vessel Segmentation in Retinal Fundus Image as Test Case”, *2017 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW)*, pp. 794–800, 2017, 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, JUL 21-26, 2017. DOI: 10.1109/CVPRW.2017.110

M. W. K. Law et A. C. S. Chung, “Three Dimensional Curvilinear Structure Detection Using Optimally Oriented Flux”, dans *COMPUTER VISION - ECCV 2008, PT IV, PROCEEDINGS*, série LECTURE NOTES IN COMPUTER SCIENCE, Forsyth, D and Torr, P and Zisserman, A, éd., vol. 5305, no. 4. INRIA; Ville Marseille; Reg Province Alpes-Cote Azur; Deutsch Telekom Lab; Microsoft Res; Orange; INRIA, Microsoft Res; EADS; TOSHIBA; Springer, 2008, pp. 368–382, 10th European Conference on Computer Vision (ECCV 2008), Marseille, FRANCE, OCT 12-18, 2008.

S. Lee, E. Lebed, M. V. Sarunic, et M. F. Beg, “Exact Surface Registration of Retinal Surfaces From 3-D Optical Coherence Tomography Images”, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 62, no. 2, pp. 609–617, FEB 2015. DOI: 10.1109/TBME.2014.2361778

S. Lee, N. Charon, B. Charlier, K. Popuri, E. Lebed, M. V. Sarunic, A. Trouve, et M. F. Beg, “Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (fshape) framework”, *MEDICAL IMAGE ANALYSIS*, vol. 35, pp. 570–581, JAN 2017. DOI: 10.1016/j.media.2016.08.012

H. Li et A. Yezzi, “Vessels as 4d curves: Global minimal 4d paths to extract 3d tubular surfaces”, dans *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW’06)*, June 2006, pp. 82–82. DOI: 10.1109/CVPRW.2006.210

K. Li, X. Wu, D. Chen, et M. Sonka, “Optimal surface segmentation in volumetric images - A graph-theoretic approach”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 28, no. 1, pp. 119–134, JAN 2006. DOI: 10.1109/TPAMI.2006.19

Z. Li, Y. Zhang, H. Gong, G. Liu, W. Li, et X. Tang, “An automatic and efficient coronary arteries extraction method in CT angiographies”, *BIOMEDICAL SIGNAL PROCESSING AND CONTROL*, vol. 36, pp. 221–233, JUL 2017. DOI: 10.1016/j.bspc.2017.04.002

W. Liao, S. Woerz, C.-K. Kang, Z.-H. Cho, et K. Rohr, “Progressive Minimal Path Method for Segmentation of 2D and 3D Line Structures”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 40, no. 3, pp. 696–709, MAR 2018. DOI: 10.1109/TPAMI.2017.2691709

B. Liu, M. Wong, R. TINKER, E. GEIDUSCHEK, et B. ALBERTS, “THE DNA-REPLICATION FORK CAN PASS RNA-POLYMERASE WITHOUT DISPLACING THE

NASCENT TRANSCRIPT”, *NATURE*, vol. 366, no. 6450, pp. 33–39, NOV 4 1993. DOI: 10.1038/366033a0

Y.-Y. Liu, H. Ishikawa, M. Chen, G. Wollstein, J. S. Duker, J. G. Fujimoto, J. S. Schuman, et J. M. Rehg, “Computerized Macular Pathology Diagnosis in Spectral Domain Optical Coherence Tomography Scans Based on Multiscale Texture and Shape Features”, *INVESTIGATIVE OPHTHALMOLOGY & VISUAL SCIENCE*, vol. 52, no. 11, pp. 8316–8322, OCT 2011. DOI: 10.1167/iovs.10-7012

C. Lorenz, I. Carlsen, T. Buzug, C. Fassnacht, et J. Weese, “Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images”, *Lecture Notes in Computer Science*, vol. 1205, pp. 233–242, 1997, 1st Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine / Medical Robotics and Computer-Assisted Surgery (CVRMed-MRCAS 97), GRENOBLE, FRANCE, MAR 19-22, 1997.

L. A. Loss, G. Bebis, et B. Parvin, “Iterative Tensor Voting for Perceptual Grouping of Ill-Defined Curvilinear Structures”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 30, no. 8, pp. 1503–1513, AUG 2011. DOI: 10.1109/TMI.2011.2129526

K. Maninis, J. Pont-Tuset, P. A. Arbeláez, et L. J. V. Gool, “Deep retinal image understanding”, *CoRR*, vol. abs/1609.01103, 2016. En ligne: <http://arxiv.org/abs/1609.01103>

D. Marr et E. Hildreth, “Theory of edge detection”, *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980. DOI: 10.1098/rspb.1980.0020. En ligne: <http://rspb.royalsocietypublishing.org/content/207/1167/187>

G. Medioni, M.-S. Lee, et C.-K. Tang, “computational framework for segmentation and grouping”, *Elsevier*, 2000.

O. Merveille, H. Talbot, L. Najman, et N. Passat, “Curvilinear Structure Analysis by Ranking the Orientation Responses of Path Operators”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 40, no. 2, pp. 304–317, FEB 2018. DOI: 10.1109/TPAMI.2017.2672972

R. Moreno et O. Smedby, “Gradient-based enhancement of tubular structures in medical images”, *MEDICAL IMAGE ANALYSIS*, vol. 26, no. 1, pp. 19–29, DEC 2015. DOI: 10.1016/j.media.2015.07.001

E. Nasr-Esfahani, N. Karimi, M. H. Jafari, S. M. R. Soroushmehr, S. Samavi, B. K. Nalamothu, et K. Najarian, “Segmentation of vessels in angiograms using convolutional neural networks”, *BIOMEDICAL SIGNAL PROCESSING AND CONTROL*, vol. 40, pp. 240–251, FEB 2018. DOI: 10.1016/j.bspc.2017.09.012

S. Negrini, V. G. Gorgoulis, et T. D. Halazonetis, “Genomic instability - an evolving hallmark of cancer”, *NATURE REVIEWS MOLECULAR CELL BIOLOGY*, vol. 11, no. 3, pp. 220–228, MAR 2010. DOI: 10.1038/nrm2858

J. Nieminuszczy, R. A. Schwab, et W. Niedzwiedz, “The DNA fibre technique - tracking helicases at work”, *METHODS*, vol. 108, pp. 92–98, OCT 1 2016. DOI: 10.1016/j.ymeth.2016.04.019

K. Paeschke, M. L. Bochman, P. D. Garcia, P. Cejka, K. L. Friedman, S. C. Kowalczykowski, et V. A. Zakian, “Pif1 family helicases suppress genome instability at G-quadruplex motifs”, *NATURE*, vol. 497, no. 7450, pp. 458+, MAY 23 2013. DOI: 10.1038/nature12149

J. Park, N. T. Kien, et G. Lee, “Optic disc detection in retinal images using tensor voting and adaptive mean-shift”, dans *ICCP 2007: IEEE 3RD INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTER COMMUNICATION AND PROCESSING, PROCEEDINGS*, série IEEE International Conference on Intelligent Computer Communication and Processing ICCP, Letia, IA, éd. IEEE, 2007, pp. 237+, IEEE 3rd International Conference on Intelligent Computer Communication and Processing, Cluj Napoca, ROMANIA, SEP 06-08, 2007.

R. Phellan et N. D. Forkert, “Comparison of vessel enhancement algorithms applied to time-of-flight MRA images for cerebrovascular segmentation”, *Medical Physics*, vol. 44, pp. 5901–5915, Nov. 2017. DOI: 10.1002/mp.12560

J. Poli, O. Tsaponina, L. Crabbe, A. Keszthelyi, V. Pantesco, A. Chabes, A. Lengronne, et P. Pasero, “dNTP pools determine fork progression and origin usage under replication stress”, *EMBO JOURNAL*, vol. 31, no. 4, pp. 883–894, FEB 15 2012. DOI: 10.1038/emboj.2011.470

G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, et M. Sonka, “Three-Dimensional Analysis of Retinal Layer Texture: Identification of Fluid-Filled Regions in SD-OCT of the Macula”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 29, no. 6, pp. 1321–1330, JUN 2010. DOI: 10.1109/TMI.2010.2047023

C. Ribeyre, J. Lopes, J.-B. Boule, A. Piazza, A. Guedin, V. A. Zakian, J.-L. Mergny, et A. Nicolas, “The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo”, *PLOS GENETICS*, vol. 5, no. 5, MAY 2009. DOI: 10.1371/journal.pgen.1000475

R. Rigamonti et V. Lepetit, “Accurate and Efficient Linear Structure Segmentation by Leveraging Ad Hoc Features with Learned Filters”, dans *MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION - MICCAI 2012, PT I*, série Lecture Notes in Computer Science, Ayache, N and Delingette, H and Golland, P and Mori, K, éd., vol. 7510. GE HealthCare; Philips; Siemens; Canon Median; ERC MedYMA; Medtronic; Aviesan; Dosisoft; IHU Strasbourg; IRCAD France; Kitware; Microsoft Res, 2012, pp. 189–197, 15th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Nice, FRANCE, OCT 01-05, 2012.

L. Risser, F. Plouraboue, et X. Descombes, “Gap filling of 3-D microvascular networks by tensor voting”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 27, no. 5, pp. 674–687, MAY 2008. DOI: 10.1109/TMI.2007.913248

D. Rivest-Henault et M. Cheriet, “3-D Curvilinear Structure Detection Filter via Structure-Ball Analysis”, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 22, no. 7, pp. 2849–2863, JUL 2013. DOI: 10.1109/TIP.2013.2240005

Y. Ruzankina, C. Pinzon-Guzman, A. Asare, T. Ong, L. Pontano, G. Cotsarelis, V. P. Zediak, M. Velez, A. Bhandoola, et E. J. Brown, “Deletion of the developmentally essential gene ATR in adult mice leads to age-related phenotypes and stem cell loss”, *CELL STEM CELL*, vol. 1, no. 1, pp. 113–126, JUL 2007. DOI: 10.1016/j.stem.2007.03.002

A. Salic et T. J. Mitchison, “A chemical method for fast and sensitive detection of DNA synthesis in vivo”, *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 105, no. 7, pp. 2415–2420, FEB 19 2008. DOI: 10.1073/pnas.0712168105

Y. Sato, S. Nakajima, H. Atsumi, T. Koller, G. Gerig, S. Yoshida, et R. Kikinis, “3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images”, *Springer Berlin Heidelberg*, pp. 213–222, 1997.

M. Schneider, S. Hirsch, B. Weber, G. Szekely, et B. H. Menze, “Joint 3-D vessel segmentation and centerline extraction using oblique Hough forests with steerable filters”, *MEDICAL IMAGE ANALYSIS*, vol. 19, no. 1, pp. 220–249, JAN 2015. DOI: 10.1016/j.media.2014.09.007

F. R. Silva, V. G. Vidotti, F. Cremasco, M. Dias, E. S. Gomi, et V. P. Costa, “Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry”, *ARQUIVOS BRASILEIROS DE OFTALMOLOGIA*, vol. 76, no. 3, pp. 170–174, MAR-JUN 2013. DOI: 10.1590/S0004-27492013000300008

P. Soille, E. Breen, et R. Jones, “Recursive implementation of erosions and dilations along discrete lines at arbitrary angles”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 18, no. 5, pp. 562–567, MAY 1996. DOI: 10.1109/34.494646

P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, et S. Farsiu, “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images”, *BIOMEDICAL OPTICS EXPRESS*, vol. 5, no. 10, pp. 3568–3577, OCT 1 2014. DOI: 10.1364/B0E.5.003568

L. Sukkaew, B. Uyyanonvara, et S. Barman, “Comparison of edge detection techniques on vessel detection of infant’s retinal image”, *Proceeding of ICIM*, pp. 6.1–6.5, 2005.

O. Tankyevych, H. Talbot, et P. Dokladal, “Curvilinear morpho-Hessian filter”, *2008 IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING: FROM NANO TO MACRO, VOLS 1-4*, pp. 1011+, 2008, 5th IEEE International Symposium on Biomedical Imaging, Paris, FRANCE, MAY 14-17, 2008. DOI: 10.1109/ISBI.2008.4541170

B. Thackray et A. Nelson, “SEMIAUTOMATIC SEGMENTATION OF VASCULAR NETWORK IMAGES USING A ROTATING STRUCTURING ELEMENT (ROSE) WITH MATHEMATICAL MORPHOLOGY AND DUAL FEATURE THRESHOLDING”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 12, no. 3, pp. 385–392, SEP 1993. DOI: 10.1109/42.241865

- E. Turetken, C. Becker, P. Glowacki, F. Benmansour, et P. Fua, “Detecting Irregular Curvilinear Structures in Gray Scale and Color Imagery using Multi-Directional Oriented Flux”, *2013 IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV)*, pp. 1553–1560, 2013, IEEE International Conference on Computer Vision (ICCV), Sydney, AUSTRALIA, DEC 01-08, 2013. DOI: 10.1109/ICCV.2013.196
- E. Turetken, F. Benmansour, B. Andres, P. Glowacki, H. Pfister, et P. Fua, “Reconstructing Curvilinear Networks Using Path Classifiers and Integer Programming”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 38, no. 12, pp. 2515–2530, DEC 2016. DOI: 10.1109/TPAMI.2016.2519025
- F. G. Venhuizen, B. van Ginneken, F. van Asten, M. J. J. P. van Grinsven, S. Fauser, C. B. Hoyng, T. Theelen, et C. I. Sanchez, “Automated Staging of Age-Related Macular Degeneration Using Optical Coherence Tomography”, *INVESTIGATIVE OPHTHALMOLOGY & VISUAL SCIENCE*, vol. 58, no. 4, pp. 2318–2328, APR 2017. DOI: 10.1167/iovs.16-20541
- Y. Wang, P. Chastain, P. Yap, J. Cheng, D. Kaufman, L. Guo, et D. Shen, “Automated dna fiber tracking and measurement”, *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1349–1352, March 2011. DOI: 10.1109/ISBI.2011.5872650
- M. Xie, Y. Yen, T. K. Owonikoko, S. S. Ramalingam, F. R. Khuri, W. J. Curran, P. W. Doetsch, et X. Deng, “Bcl2 Induces DNA Replication Stress by Inhibiting Ribonucleotide Reductase”, *CANCER RESEARCH*, vol. 74, no. 1, pp. 212–223, JAN 1 2014. DOI: 10.1158/0008-5472.CAN-13-1536-T
- J. Yarbrow, “MECHANISM OF ACTION OF HYDROXYUREA”, *SEMINARS IN ONCOLOGY*, vol. 19, no. 3, 9, pp. 1–10, JUN 1992.
- F. Zana et J. Klein, “Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation”, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 10, no. 7, pp. 1010–1019, JUL 2001. DOI: 10.1109/83.931095
- M. K. Zeman et K. A. Cimprich, “Causes and consequences of replication stress”, *NATURE CELL BIOLOGY*, vol. 16, no. 1, pp. 2–9, JAN 2014. DOI: 10.1038/ncb2897



D. Zhang et G. Lu, “Review of shape representation and description techniques”, *PATTERN RECOGNITION*, vol. 37, no. 1, pp. 1–19, JAN 2004. DOI: 10.1016/j.patcog.2003.07.008

H. Zhang, Y. Yang, et H. Shen, “Detection of Curvilinear Structure in Images by a Multi-Centered Hough Forest Method”, *IEEE ACCESS*, vol. 6, pp. 22 684–22 694, 2018. DOI: 10.1109/ACCESS.2018.2823726

J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. W. Pluim, R. Duits, et B. M. t. H. Romeny, “Robust Retinal Vessel Segmentation via Locally Adaptive Derivative Frames in Orientation Scores”, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 35, no. 12, pp. 2631–2644, DEC 2016. DOI: 10.1109/TMI.2016.2587062

J. Zhang, E. Bekkers, D. Chen, T. T. J. M. Berendschot, J. Schouten, J. P. W. Pluim, Y. Shi, B. Dashtbozorg, et B. M. t. H. Romeny, “Reconnection of Interrupted Curvilinear Structures via Cortically Inspired Completion for Ophthalmologic Images”, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 65, no. 5, pp. 1151–1165, MAY 2018. DOI: 10.1109/TBME.2017.2787025

H. Zhao, J. Kumagai, M. Nakagawa, et R. Shibasaki, “Semi-automatic road extraction from high-resolution satellite image”, *Remote Sens. Spatial Inf. Sci.*, no. 34, 2002.

## ANNEXE A CLASSIFICATION AUTOMATIQUE DE PATHOLOGIES OCULAIRES A PARTIR DE L'ÉTUDE DES CARTES D'ÉPAISSEUR DE LA CHOROÏDE ET DE LA RÉTINE SUR DES IMAGES OCT DU FOND DE L'OEIL

Le travail présenté dans cette annexe est indépendant du reste de ce mémoire. Il s'agit du second projet réalisé pendant la maîtrise.

### A.1 Introduction

La Tomographie par Cohérence Optique (OCT) est une technique d'imagerie non invasive permettant de reconstituer la structure d'un tissu à partir de l'étude de l'interférence d'une onde lumineuse. Cette modalité, utilisée en ophtalmologie a l'avantage de pouvoir imager les couches profondes de la rétine et de la choroïde. L'étude des images OCT par le médecin ophtalmologue permet d'extraire des informations utiles au diagnostic et à la compréhension des différentes pathologies oculaires.

L'automatisation du diagnostic des pathologies oculaires grâce à des outils d'apprentissage automatique a suscité un intérêt au cours de la dernière décennie. En avril 2018, la *Food and Drug Administration* (FDA) a autorisé la mise sur le marché du premier système de diagnostic automatique de la rétinopathie diabétique à partir de l'étude de la photo du fond de l'oeil. Ce type d'algorithme pourrait être utile pour faire du dépistage dans les zones géographiques marquées par la pénurie de médecins ophtalmologues. Des algorithmes de détection de pathologies oculaires pourraient aussi être très utiles dans les hôpitaux afin de réduire le temps d'attente des patients. En effet, un diagnostic automatique en amont de la rencontre avec le médecin permettrait d'orienter le patient vers un ophtalmologue ou bien vers une infirmière praticienne selon la pathologie détectée. D'autre part, l'utilisation d'algorithmes pour créer un modèle de classification est aussi utile afin de mettre en relation un très grand nombre de descripteurs et découvrir de nouvelles propriétés des pathologies.

L'objectif de ce travail est de classer automatiquement la Dégénérescence Maculaire Liée à l'Âge, le glaucome, l'hyperTension oculaire et l'uvéite de type birdshot à partir de l'étude des cartes d'épaisseur de la rétine et de la choroïde.

Ce chapitre se décomposera de la manière suivante : Dans une première partie, nous définirons brièvement les pathologies étudiées et ferons un tour d'horizon des algorithmes de classification existants. Dans une seconde partie, nous détaillerons la méthode de notre approche.

Dans une troisième partie, nous exposerons nos résultats que nous discuterons dans la dernière partie de ce chapitre

## A.2 Revue de Littérature

### A.2.1 Anatomie rétinienne et choroïdienne sur les images OCT

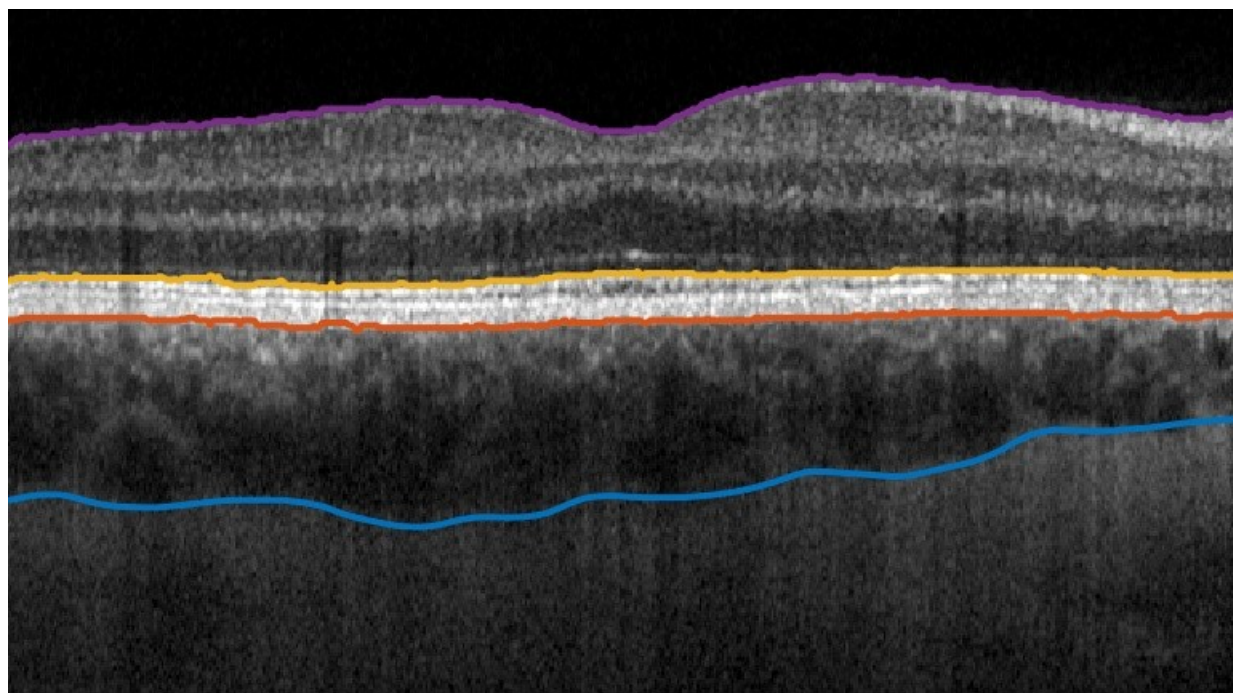


Figure A.1 BScan d'un sujet sain. Les interfaces des différentes couches rétiniennes et choroïdiennes sont représentées : RVI (violet), RPEa (jaune), RPEp (orange) et CSI (bleu)

Nous décrivons dans ce paragraphe la structure des différentes couches rétiniennes et choroïdiennes utilisées dans notre étude. L'Interface entre le Vitrée et la Rétine (RVI) est représentée en violet sur la Figure A.3. Elle sépare la partie vitrée (substance gélatineuse transparente) de la rétine. La couche la plus postérieure de la rétine est l'Épithélium Rétinien Pigmentaire (RPE). Cette couche a pour but de nourrir les cellules réceptrices de la rétine. Elle est séparée du reste de la rétine par son interface antérieure (RPEa) et de la choroïde par sa membrane postérieure (RPEp).

Sous la RPE, se trouve la choroïde contenant les vaisseaux sanguins irrigant la rétine. L'Interface entre la Choroïde et la Sclère est notée CSI.

### **A.2.2 Brèves définitions des pathologies oculaires étudiées**

Notre travail consiste en l'étude des images OCT pour des patients atteints de différentes maladies oculaires : La Dégénérescence Maculaire Liée à l'Âge (DMLA), le glaucome, l'uvéite et l'hypertension oculaire. Nous décrivons brièvement ces différentes pathologies ci-après.

La Dégénérescence Maculaire Liée à l'Âge (DMLA) est la première cause de cécité dans les pays développés. Les deux principales formes sont la DMLA sèche et humide. La DMLA sèche se caractérise par le dépôt de débris cellulaires appelés drusens localisés entre la rétine et la choroïde. La DMLA humide, plus grave, est marquée par l'apparition de vaisseaux sanguins entre la rétine et la choroïde.

Le glaucome est la troisième cause de cécité dans les pays développés. Cette maladie se caractérise par une dégradation du nerf optique entraînant une réduction du champ de vision. Bien que d'origine neuronale, le glaucome agit aussi sur la rétine en détruisant les cellules ganglionnaires et leurs axons.

L'Hypertension Oculaire (HTO) se caractérise par une pression de l'œil anormalement élevée. Cette maladie se traduit par la difficulté de l'œil à évacuer l'humeur aqueuse. L'hypertension oculaire peut dégénérer en glaucome.

L'Uvéite de type Birdshot est une forme rare d'inflammation de l'uvée postérieure (composée de l'iris, du corps ciliaire et de la choroïde). Cette maladie cause une altération de la vision chez le patient ; vision trouble et distinction des couleurs plus difficile. Les principaux marqueurs physiologiques de cette maladie pour les stades avancés sont des lésions choroïdiennes formant des tâches blanches ou jaunâtres sur le fond de l'œil (d'où le nom «birdshot»). D'autres marqueurs tels que des œdèmes maculaires et une forte vascularisation peuvent apparaître. La présence du gène HLA-A29 peut présumer un risque de maladie.

### **A.2.3 Les algorithmes de diagnostic existant se basent sur des descripteurs texturaux et morphologiques**

De nombreuses études s'intéressent à la classification automatique de pathologies oculaires à partir de l'étude d'images OCT. Le diagnostic automatique de la Dégénérescence Maculaire Liée à l'Âge (DMLA) (Liu et al., 2011; Srinivasan et al., 2014; Venhuizen et al., 2017; Hussain et al., 2018), de la Rétinopathie Diabétique (RD), de l'Oedème Maculaire Diabétique (DME) (Liu et al., 2011; Hussain et al., 2018; Alsaih et al., 2016; Quéllec et al., 2010) et du glaucome (Bizios et al., 2010; Barella et al., 2013; Silva et al., 2013; Lee et al., 2017) ont déjà fait l'objet de nombreux travaux. Ces études utilisent 2 types de descripteurs pour caractériser ces maladies :

- **Descripteurs texturaux** : Ces descripteurs ont pour but de caractériser les textures entre les tissus sains et pathologiques. Les descripteurs les plus populaires sont les Patterns Binaires Locaux (PBL) (Liu et al., 2011; Alsaïh et al., 2016) ou les Histogrammes de Gradients Orientés (HGO) (Srinivasan et al., 2014). L’approche par PBL [16] repose sur l’analyse des différences d’intensité de chaque pixel avec ses 8 pixels voisins. Liu et al. (2011) proposent d’extraire les PBL à différentes échelles de l’image à partir de la décomposition pyramidale par ondelette. Les HGO quant à eux étudient la distribution de l’orientation du gradient de l’image.
- **Descripteurs morphologiques** : Le principal défaut des descripteurs texturaux est qu’ils dépendent beaucoup du bruit généré par l’appareil de mesure. En revanche, les descripteurs morphologiques comme l’épaisseur des différentes couches de la rétine (Hussain et al., 2018; Bizios et al., 2010; Barella et al., 2013), ou encore la taille des oedèmes rétiens et des drusens (Quellec et al., 2010) ne sont pas sujets à de tels biais. La détection automatique de ces descripteurs morphologiques est l’objet de la partie suivante.

#### A.2.4 L’extraction automatique des descripteurs morphologiques repose sur un algorithme de segmentation de la rétine et de la choroïde

L’extraction automatique des descripteurs morphologiques repose sur un algorithme de segmentation de la rétine et de la choroïde. L’extraction automatique de descripteurs morphologiques comme l’épaisseur de la rétine ou de la choroïde nécessite la segmentation automatique de ces couches. L’algorithme développé par Li and Yezzi (2006) permet de segmenter des surfaces dans des images en 3D. L’approche consiste à rechercher des surfaces lisses passant par le plus de contours (“edge”) possibles. Un algorithme de recherche par graphe 3D permet de trouver ces surfaces optimales. La quantification des contours dans les images OCT est la plupart du temps effectuée en calculant les gradients axiaux pour chaque pixel (Lee et al., 2015; Dufour et al., 2013). Plus récemment, Fang et al. (2017) utilisent un réseau de neurone convolutionnel pour calculer une carte de probabilité des contours avant d’utiliser la méthode de recherche par graphe susmentionnée.

Plus récemment, De Fauw et al. (2018) ont proposé une approche de segmentation des interfaces totalement basée sur les réseaux de neurones convolutionnels. Un réseau de type U-net segmente les différentes structures du volume OCT et un second réseau fait un diagnostic à partir de l’étude des cartes d’épaisseur précédemment obtenues. Le type d’architecture utilisé nécessite des ressources computationnelles très importantes. Notons aussi que leur réseau de neurones ne permet pas la segmentation de l’interface entre la choroïde et la sclère.

### A.2.5 Méthode

#### a- Base de Données

Nos images ont été acquises en utilisant un système FD-OCT (Spectralis OCT Plus, Heidelberg Engineering, Germany). Chaque volume est constitué d'un nombre ajustable de Bscans (49 à 193). Chaque Bscan a une résolution axiale de  $3.9 \mu m$  et latérale de  $11 \mu m$ . Le contenu de la base de données est synthétisé dans le tableau A.1 :

#### b- Segmentation des différentes interfaces de la rétine et de la choroïde

**Etape 1 : Création du Volume OCT : alignement des Bscans** Pour pouvoir aligner les Bscans, nous décidons de les aplatir au niveau de la membrane de Bruch. Pour chacun des Bscans, nous segmentons la RVI ainsi que la membrane antérieure et postérieure de la RPE. La segmentation repose sur un algorithme de la théorie des graphes : l'algorithme du plus court chemin. Les nœuds du graphe sont les pics du gradient de chaque A-scan. Les poids de chaque connexion sont définis à partir de la distance séparant chaque noeud. Une fois la RPE segmentée, on obtient la membrane de Bruch en traçant l'enveloppe convexe de l'interface postérieure de la RPE. Chaque A-scan est décalé de manière à ce que la membrane de Bruch soit aplatie. Enfin, pour réduire le 'speckle noise', nous moyennons chaque Bscan avec les Bscans adjacents.

#### Etape 2 : Segmentation de la RPE et de la RVI

Dans l'étape précédente, la segmentation de la RPE et de la RVI est faite indépendamment pour chaque Bscan grâce à un graphe en 2D. Maintenant que les Bscans sont alignés, chaque couche peut être segmentée, grâce à un graphe en 3D. Cette technique décrite dans Li et al. (2006) permet de trouver la surface S qui maximise une fonction objectif Fo. Cette fonction objective est définie par la somme des poids P des pixels constituant la surface S :

Table A.1 Base de données

	Nombre d'yeux imagés	Nombre de patients
Contrôle (sain)	121	106
Glaucome	115	98
DMLA	60	43
Uvéite Birdshot	36	20
HTO	29	27
TOTAL	361	294

$$Fo = \sum_{x \in S} P(x) \quad (\text{A.1})$$

avec  $P(x) = I(x) - I(x - 1)$  si  $x$  est un point d'inflexion sur son A scan ou

$$P(x) = -\gamma \text{ sinon } (\gamma > 0)$$

Ainsi, notre surface passera par les points d'inflexion ayant le plus fort gradient. La technique du graphe 3D (Li et al., 2006) permet de segmenter plusieurs surfaces en même temps si l'on spécifie la distance maximale et minimale séparant ces surfaces. Pour notre problème, un graphe segmentera l'interface antérieure et postérieure de la RPE, puis un autre graphe segmentera la RVI. Pour chacun des graphes, On ne considère uniquement les pixels situés dans un voisinage de la segmentation préliminaire obtenue dans l'étape 1.

**Etape 3 : Segmentation de la CSI** Dans un premier temps, considérons chaque Bscan indépendamment. La segmentation de la CSI repose sur un algorithme de classification (Random Forrest). Chaque pixel postérieur à la RPE est classé comme appartenant à la sclère ou bien comme appartenant à la choroïde. Les caractéristiques de chaque pixel sont basées sur le gradient, le laplacien de chaque A-scan et le filtre de gabor du Bscan. Un fois la classification faite, on obtient une image binaire. Les zones d'interface entre les pixels de classe 'sclère' et les pixels de classe 'choroïde' constituent les nœuds d'un graphe 2D. L'algorithme du plus court chemin permet de relier les nœuds et d'obtenir ainsi une première approximation de la CSI. Pour affiner cette segmentation et éliminer les nœuds dont la position est aberrante par rapport aux Bscans adjacents, nous appliquons de nouveau l'algorithme du plus court chemin mais cette fois-ci dans chaque plan orthogonal aux Bscans. En interpolant les nœuds restants, on obtient la CSI.

Au terme de ces trois étapes nous obtenons 3 couches, ce qui nous permettra d'obtenir 3 cartes d'épaisseur : une carte de l'épaisseur de la rétine, de l'épithélium pigmentaire rétinien, et de la choroïde.

### c- Extraction des descripteurs

Dans cette étude, nous nous limitons aux descripteurs extraits des cartes de l'épaisseur des différentes couches de la rétine et de la choroïde. Après avoir décrit la construction des cartes, nous extrairons deux types de descripteurs : les descripteurs spatiaux, les descripteurs de Fourier.

#### 1. Descripteurs spatiaux : épaisseur locale et gradients

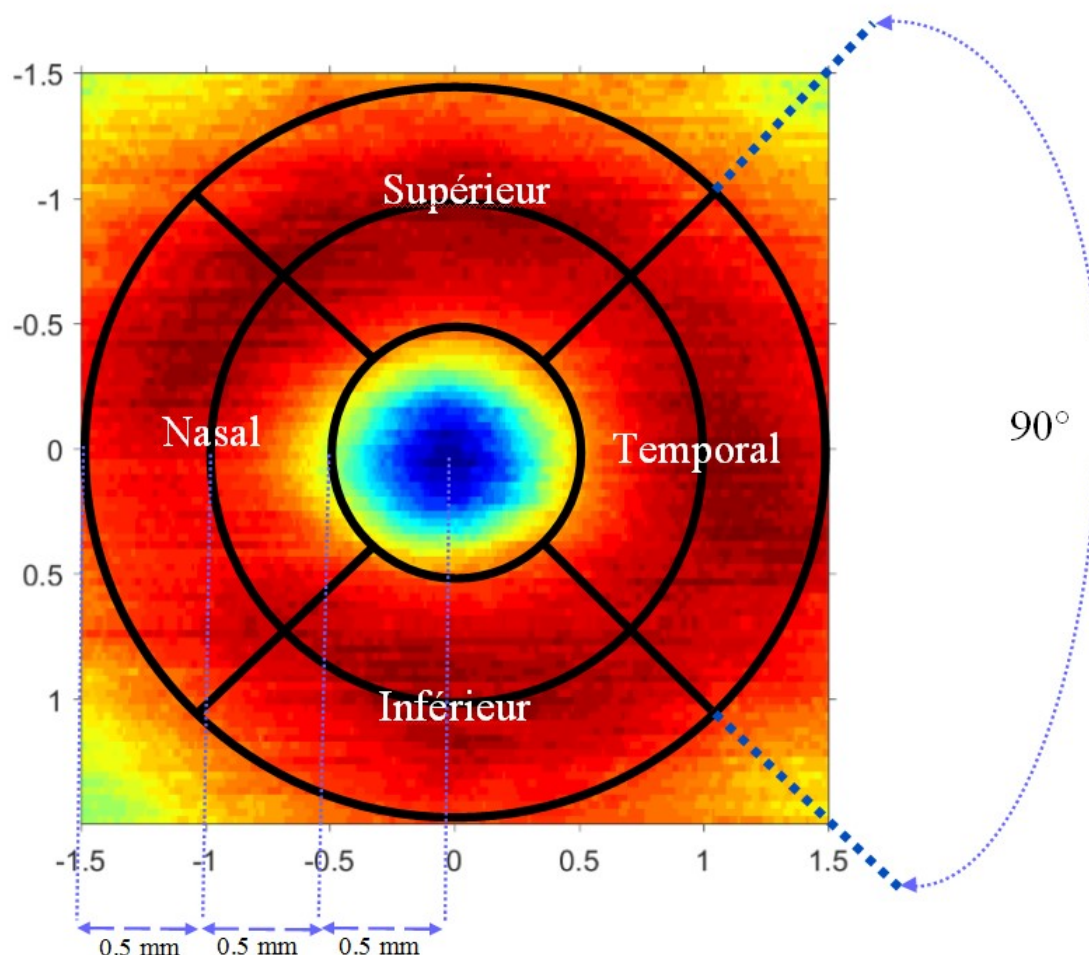


Figure A.2 Découpage d'une carte d'épaisseur en 9 sous-régions

Chaque carte d'épaisseur est divisée en 9 sous régions (Figure A.2) : Une région centrale délimitée par un cercle de rayon 0.5mm centré sur la macula, et deux sous régions par quadrant. Nous extrayons des descripteurs d'épaisseur et de gradient :

- Descripteurs d'épaisseur : Pour chaque sous région, nous calculons moyenne, écart-type, 10e percentile, 90e percentile, minimum, maximum (6 caractéristiques).
- Descripteurs de gradient : histogramme du gradient vertical et horizontal (14 bins de  $3.9 \mu m$  chacun = 28 descripteurs) On obtient donc  $(6+28)*9=306$  descripteurs par carte. Comme chaque œil dispose de 3 cartes d'épaisseur, on obtient donc 918 descripteurs dits "classiques" par œil.



2. **Descripteurs de Fourier** Les descripteurs de Fourier sont particulièrement utilisés dans le domaine du « shape matching » (Zhang and Lu, 2004). Pour chaque carte nous calculons la transformée de Fourier 2D. Ensuite, nous réalisons un filtrage passe-bas idéal. La fréquence de coupure pour l’axe des X et l’axe des Y sont respectivement  $F_{xc} = 15\text{kHz}$  et  $F_{yc} = 14\text{kHz}$  (Figure 4). Les descripteurs de Fourier sont les magnitudes de la fft pour les fréquences filtrées. On se retrouve à la fin du filtrage avec un vecteur de 465 descripteurs par carte. Comme on dispose de 3 cartes par œil, on se retrouve avec  $465 \cdot 3 = 1395$  descripteurs par œil.

#### d- Sélection des descripteurs

Pour chaque descripteur, nous calculons son critère de Fisher (Duda, Hart, & Stork, Non-parametric techniques, 2000). Ce critère évalue la divergence entre deux classe  $i$  et  $j$  en se basant sur la moyenne  $\mu$  et l’écart-type  $\sigma$  estimés du descripteur  $F$  pour ces deux classes :

$$D(F) = \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2} \quad (\text{A.2})$$

avec  $\mu_i$  et  $\sigma_i$  respectivement la moyenne et l’écart-type du descripteur pour la classe  $i$ . Le critère  $D(F)$  doit être maximisé car on veut que la séparation entre les classes soit la plus grande possible. On ne retiendra uniquement les descripteurs ayant un critère de Fisher supérieur au 75e percentile de l’ensemble des critères de Fisher. Enfin, on effectue une analyse par composante principale sur les descripteurs filtrés. On ne garde les projections dont l’inertie regroupe 95% de l’inertie totale.

#### e- Classification

Nous utilisons le classifieur Random Forest (RF) pour discriminer les différentes classes.

Entraîner un RF consiste à construire une forêt d’arbre de décision en sélectionnant aléatoirement l’ensemble d’entraînement et les descripteurs. La classification des échantillons test consiste à agréger les votes des différents arbres (Breiman, Random forests, 2001). Deux paramètres doivent être choisis par l’utilisateur : le nombre d’arbres ( $N_a$ ) et le nombre de descripteurs ( $N_f$ ) utilisés par chaque arbre. Nous choisissons  $N_a = 300 \cdot nb_{classes}$  et  $N_f = \sqrt{N_{ft}}$  avec  $N_{ft}$  le nombre total de descripteurs après sélection.

Nous employons la méthode de validation croisée pour tester les performances de notre approche.

### A.2.6 Résultats

Nous avons d'une part entraîné notre algorithme à discriminer les 5 classes (sain, DMLA, Glaucome, HTO, Uvéite Bidshsot). Les résultats de cette expérience sont représentés dans la figure A.3

Ensuite, nous avons entraîné le classifieur pour discriminer une maladie vs les sujets sains ( Figure A.4).

<i>prédiction</i> <i>réalité</i>	Sain	DMLA	Glauc.	HTO	Bird.	total
Sain	29	1	20	54	17	121
DMLA	0	55	0	0	5	60
Glauc.	30	2	35	25	23	115
HTO	8	0	6	14	1	29
Bird.	2	3	4	4	23	36
Total	69	61	65	97	69	361

Figure A.3 Classification globale. DMLA : Dégénérescence Maculaire Liée à l'Âge; Glauc.: Glaucome; HTO: HyperTension Oculaire; Bird.:Uvéite de type Birdshot

	Spécificité (%)	Sensibilité (%)	Précision (%)
DMLA vs. Sain	96	96	96
Glauc. vs. Sain	74	71	73
HTO vs. Sain	67	72	69
Birdshot vs. Sain	83	80	83

Figure A.4 Classification binaire

### A.2.7 Discussion

#### De bons résultats pour la DMLA

On remarque que notre algorithme est particulièrement performant pour diagnostiquer les patients atteints de DMLA. 92% des patients atteints de DMLA sont bien diagnostiqués dans

la classification globale. Dans la classification binaire (DMLA vs sain) 96 % des sujets ont été correctement classés. Les descripteurs extraits discriminent très efficacement la DMLA par rapport à toutes les autres maladies. Il est intéressant de noter que contrairement à (Farsiu, Chiu et al. 2014), nous n'avons pas exclu de notre étude les patients dont les drusens sont trop petits ( $<125 \mu m$ ).

Parmi toutes les maladies étudiées dans ce travail, la DMLA est celle qui est la plus visible au niveau de l'OCT. Les lésions dans l'épithélium pigmentaire entraînent une forte variation de l'épaisseur de ces tissus.

### **Des résultats prometteurs pour le diagnostics automatique de l'Uvéite**

La sensibilité et la spécificité de la classification binaire sont respectivement 80% et 83%. Il est intéressant de remarquer que peu de patient birdshot sont classifiés comme étant des patients atteints de DMLA alors que ces deux maladies affectent l'Epithélium Rétinien Pigmentaire. A notre connaissance, il n'existe pas d'autres études s'intéressant au diagnostic automatique de l'uvéite de type birdshot.

### **Les performances de notre algorithme pour discriminer les patients atteints de glaucome et d'hypertension oculaire sont modestes**

La précision de la classification binaire pour discriminer le glaucome et l'hypertension oculaire est plus modeste. La précision atteint respectivement 73% et 69%. Cela est peu surprenant car nos volumes OCT sont centrés sur la macula alors que le glaucome est une maladie qui touche le nerf optique invisible sur nos images. L'hypertension oculaire quant à elle est quasiment invisible sur l'OCT. C'est pourquoi notre algorithme classe une majorité de patients sains dans cette catégorie (Figure A.3). Aujourd'hui, cette maladie est diagnostiquée non pas par imagerie mais par analyse de la pression intra-oculaire.

## **A.3 Conclusion**

Nous avons développé un algorithme de classification de quatre pathologies oculaires (la Dégénérescence Maculaire Liée à l'Age (DMLA), le glaucome, l'uvéite de type Birdshot, et l'HTO). Nous obtenons des résultats très bons pour la DMLA, et prometteurs pour l'uvéite.

Les travaux futurs consisteraient en l'interprétation des résultats afin de mieux comprendre les modèles fabriqués par l'algorithme, l'extension de la base de données des patients uvéites pour tenter d'améliorer les performances et l'inclusion de descripteurs texturaux.