

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE DE LA
MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR
MOHAMMED AMINE BENJIRA

CLASSIFICATION AUTOMATIQUE DES DIATOMÉES

NOVEMBRE 2019

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

Remerciements

J'exprime ma profonde gratitude envers Dieu, pour m'avoir donné la santé, la générosité et la persévérence pour réaliser ce mémoire.

Je tiens à remercier mes directeurs de recherche, les professeurs Fathallah Nouboud et Alain Chalifour, qui m'ont accueilli et qui m'ont aidé durant toutes les étapes de ce projet de recherche. Je désire les remercier de m'avoir soutenu, encadré, orienté et conseillé tout au long de mes études. Je les remercie de m'avoir accordé leur confiance en m'offrant une grande autonomie dans mon travail. Leurs conseils, critiques et encouragements m'ont permis de mener à bien ce travail de recherche.

Je voudrais aussi exprimer ma gratitude envers mes chers parents et mes sœurs qui m'ont énormément aidé et encouragé lors de mes études de cycles supérieurs. Je les remercie pour leur soutien moral et financier. Que ce travail soit le reflet de mon immense reconnaissance et de mon amour à leur égard.

J'adresse également mes sincères remerciements à tous les membres du département de Mathématiques et Informatique de l'Université du Québec à Trois-Rivières (UQTR) qui ont toujours été disponibles pour m'aider à réaliser mon cursus universitaire au programme de maîtrise en mathématiques et informatique appliquées.

Résumé

Les diatomées sont des algues microscopiques qui vivent dans différents écosystèmes aquatiques. Puisque les diatomées sont des bio-indicateurs de la qualité de l'eau, leur présence permet une surveillance des principales rivières, ou autres plans d'eau, par exemple le programme d'assainissement des eaux du Québec [1].

Les différentes espèces de diatomées sont identifiées selon leur formes et motifs. Leur classification devient une tâche laborieuse et l'utilisation de méthodes du traitement de l'image peut s'avérer utile lors de l'identification d'échantillons contenant des espèces de diatomées dans le cadre d'études environnementales des milieux aquatiques et dans d'autres applications.

Ce mémoire s'inscrit dans la continuité de plusieurs travaux de recherche entrepris au Laboratoire Interdisciplinaire de Recherche en Imagerie et Combinatoire (LIRIC) de l'UQTR. Un premier travail de classification automatique des diatomées basée sur le traitement d'images a été réalisé par A. Siddour [2]. La classification réalisée dans le cadre de ce travail est basée sur l'utilisation de la courbe des angles normaux du contour 8-connecté des diatomées isolées manuellement dans chaque image et un ensemble restreint de paramètres géométriques. La courbe de contour a été approchée à l'aide de descripteurs de Fourier. Notre projet reprend l'approche de A. Siddour [2] en ajoutant les valeurs propres du Laplacien du problème de Dirichlet [24].

Des tests de classification bayésienne ont été réalisés à partir d'une base de données d'images de diatomées dont les spécimens ont été regroupés (professeur Stéphane Campeau, département des sciences de l'environnement de l'UQTR) dans différentes classes sur la base de leurs formes et de leurs structures internes globales.

Au préalable, nous avons retenu pour l'étape de la classification un nombre restreint de variables issues d'une analyse en composantes principales de regroupements de paramètres. Cette classification a priori n'a pas été réalisée sur la base d'une identification à l'espèce des images de diatomées. Lors de nos tests de classification nous avons analysé trois groupes de paramètres, soit les descripteurs du contour, les paramètres géométriques et les valeurs propres du Laplacien. Nous avons testé isolément le pouvoir de classification de chacun des groupes de paramètres, par paire et pour l'ensemble des trois regroupements. Dans le meilleur des cas nous avons obtenu un taux de classification de 98 %. Les détails sont présentés à la fin de ce mémoire.

Abstract

Diatoms are microscopic algae that live in different aquatic ecosystems. Since diatoms are bioindicators of water quality, their presence allows monitoring of major rivers, or other bodies of water, for example the Quebec water purification program [1].

The different diatom species are identified according to their shapes and patterns. Their classification becomes a laborious task and the use of image processing methods can be useful in identifying samples containing diatom species in environmental studies of aquatic environments and in other applications.

This dissertation is a continuation of several research projects undertaken at the Interdisciplinary Laboratory for Imaging and Combinatory Research (LIRIC) at UQTR. A. Siddour [2] carried out a first work of automatic classification of diatoms based on image processing. The classification performed for this work is based on the use of the normal angles curve of the 8-connected contour of manually isolated diatoms in each image and a restricted set of geometric parameters. The contour curve was approximated using Fourier descriptors. Our project takes up the approach of A. Siddour [2] by adding the Laplacian eigenvalues of the Dirichlet problem [24].

Bayesian classification tests were performed from a database of diatom images whose specimens were grouped together (Professor Stéphane Campeau, Department of Environmental Sciences of UQTR) in different classes based on their global forms and internal structures. Beforehand, we selected for the classification stage a small number of variables resulting from a principal component analysis of parameter groupings. This a priori classification was not carried out on the basis of a species identification of the diatom images. In our classification tests, we analyzed

three groups of parameters, namely the descriptors of the contour, the geometrical parameters and the eigenvalues of the Laplacian. We tested in isolation the power of classification of each of the groups of parameters, in pairs and for all three groups. In the best case, we obtained a classification rate of 98%. Details are presented at the end of this document.

Table des Matières

Chapitre 1	Introduction.....	1
1.1	Contexte et objectifs	1
1.2	Organisation du mémoire.....	2
Chapitre 2	La biologie des diatomées	4
2.1	Description des diatomées	4
2.2	Morphologie des diatomées.....	5
2.3	Classification des diatomées.....	8
2.4	Écologie des diatomées.....	9
2.5	Applications scientifiques des diatomées.....	10
2.6	Utilisation des diatomées au Canada	10
2.7	État de l'art sur la classification automatique des diatomées.....	11
Chapitre 3	Prétraitement des images et extraction du contour et des angles normaux	15
3.1	Prétraitement des images de diatomées.....	15
3.1.1	Lissage d'une image	16
3.1.1.1	Le filtre médian.....	17
3.1.1.2	Le filtre moyen.....	18
3.1.2	Binarisation d'image	19
3.1.3	Rotation d'une diatomée	21
3.2	Extraction du contour et des angles normaux	23
3.2.1	Extraction du contour	24
3.2.2	Approximation polygonale	27

3.3	Extraction des angles normaux.....	29
Chapitre 4	Extraction des caractéristiques.....	32
4.1	Descripteurs de Fourier	32
4.2	Les paramètres de formes	34
4.3	Les valeurs propres du Laplacien	38
4.3.1	Reconnaissance de forme et valeurs propres du Laplacien de Dirichlet.....	38
Chapitre 5	Résultats.....	48
5.1	Analyse en composantes principales	49
5.2	Résultats expérimentaux.....	51
Chapitre 6	Conclusion et perspectives	54
6.1	Les Méthodes appliquées.....	54
6.2	Les perspectives d'amélioration	55
Annexe 1	56
Annexe 2	58
Annexe 3	Implémentation.....	60

Liste des figures

Figure 2.1 Exemples de diatomées	4
Figure 2.2 Principales formes de diatomées : a) centrique, b) pennée avec nervures, c) pennée avec sternum, raphé et strie.....	6
Figure 2.3 Composantes d'une diatomée pennée	7
Figure 2.4 Image microscopique de diatomées en culture (<i>S. robusta</i>).....	7
Figure 2.5 (a) Diatomées centrales, (b) diatomées pénnales.	8
Figure 3.1 Diatomée isolée manuellement.	16
Figure 3.2 Filtre médian à l'aide d'une fenêtre 3*3. À gauche, l'image originale et à droite le résultat de l'application du filtre.....	18
Figure 3.3 Filtre moyen avec une fenêtre 3 × 3. À gauche, l'image originale et à droite le résultat de l'application du filtre.....	19
Figure 3.4 Processus de la binarisation. À gauche, l'image originale, au milieu l'image binaire obtenue avec la méthode OTSU et à droite l'image finale.	21
Figure 3.5 Axe principal d'un contour d'une diatomée.	22
Figure 3.6 Image binaire d'une diatomée avant et après la rotation.	23
Figure 3.7 Indices de Freeman, (1) 4-Connectivité, (2) 8-Connectivité.	24
Figure 3.8 Position du premier pixel de départ.	25
Figure 3.9 Exemple d'une diatomée et code de Freeman avec voisinage V8 [31].....	26
Figure 3.10 Contour initial de la diatomée.....	27
Figure 3.11 Processus de l'approximation polygonale.	28
Figure 3.12 : (a) contour d'une diatomée, (b) contour après l'approximation.....	29
Figure 3.13 : Étapes pour extraire le contour d'une diatomée.....	29
Figure 3.14 : Angle normal avec alpha1 =5 et alpha =6.	30
Figure 3.15 : Exemples de deux diatomées avec leurs fonctions d'angles.....	31
Figure 4.1 Descripteurs de Fourier.	34
Figure 4.2 : Deux diatomées, leurs contours, leurs enveloppes convexes.....	34
Figure 4.3 Exemple de calcul de la convexité	35
Figure 4.4 Exemple de calcul de la circularité	36
Figure 4.5 Exemple de calcul de l'indice de rectangularité	36

Figure 4.6 Exemple de calcul de l'indice d'allongement	37
Figure 4.7 Symétrie d'une diatomée	38
Figure 4.8 Diatomée plongée dans le plan discret : repérage avec 0 pour les pixels hors de la diatomée et 1 à l'intérieur incluant le contour.	40
Figure 4.9 Un pixel (i,j) et les indices voisins dans (4.23).	42
Figure 4.10 Répartition des inconnues.	44

Liste de tableaux

Tableau 3.1 Directions admissibles.	26
Tableau 3.2 : Angles réalisables de α_i	30
Tableau 4.1 Répartition des indices des pixels de la diatomée et de son contour.	44
Tableau 4.2 Positions et numérotation des inconnues à l'intérieur du contour.	45
Tableau 4.3 Répartition des éléments à l'intérieur du contour.	45
Tableau 4.4 Entrées de la matrice A.	46
Tableau 4.5 Valeurs propres de A.....	47
Tableau 5.1 Nombres d'images par classe.	49
Tableau 5.2 Variances et variances cumulées pour l'ensemble des caractéristiques.....	50
Tableau 5.3 Variances et variances cumulées pour les TF+PF.....	50
Tableau 5.4 Résultats de la classification pour chaque cas.....	52

Chapitre 1

Introduction

1.1 Contexte et objectifs Le traitement de l'image est un procédé que tout être humain peut accomplir naturellement. Inconsciemment, ce sont les images que nos yeux envoient à notre cerveau, sous forme de signal, qui nous permettent de nous déplacer dans l'espace, apprécier une œuvre d'art et reconnaître un visage familier. Chez les personnes non-voyantes, la capacité de cerner des formes permet de se mouvoir dans l'environnement. Cependant, le traitement de l'image par la simple vision de l'homme connaît certaines limites, dans la mesure où le point de vue d'un humain n'apporte que des informations sémantiques grâce à l'interprétation qu'en fait le cerveau. D'un point de vu mathématique, une image se définit comme étant un signal 2D déduit de la projection sous le plan 2D d'une scène 3D, qu'on obtient grâce à une matrice de nombres que le cerveau, à lui seul, ne peut facilement décoder, d'où l'utilité du traitement de l'image par ordinateur.

En plus d'être une perception du monde extérieur, une image peut également être représentée de façon numérique, à l'aide de pixels ayant chacun une paire de coordonnées x et y , qui lui sont associées dans le plan discret ainsi qu'une intensité (couleur). Il devient donc possible de manipuler, transformer, cette image à l'aide d'outils numériques, ce que la vision de l'homme ne peut pas accomplir à elle seule pour des problèmes de grande taille. L'imagerie numérique est présente dans plusieurs domaines, dont la biologie, les outils d'analyse d'images nous permettront d'élaborer une stratégie, à l'aide d'algorithmes, visant la classification des diatomées de façon automatique ou semi-automatique.

Les micro-organismes regroupent plusieurs organismes tels que les bactéries, les virus, les champignons unicellulaires, les protistes, les algues, etc. Parmi ceux-ci, on retrouve

les diatomées, des algues unicellulaires comptant des milliers d'espèces, dont les diverses variations morphologiques rendent leur classification laborieuse. Pour ces raisons, le traitement d'images numériques devient un incontournable et notre travail consiste à développer une méthodologie d'aide à la classification des diatomées.

1.2 Organisation du mémoire

Ce mémoire est composé de 6 chapitres. Dans le chapitre 2, nous définissons les diatomées ainsi que leurs caractéristiques morphologiques et environnementales. De plus, nous présentons quelques applications scientifiques justifiant leur étude et leur importance dans plusieurs domaines. Nous conclurons par un court état de l'art des travaux existants sur la classification automatique des diatomées.

Au troisième chapitre, nous présentons les étapes qui précèdent la classification des diatomées proprement dite. Une première étape est celle du prétraitement des images qui vise l'élimination du bruit présent dans l'image. Cette étape de prétraitement vise principalement le lissage des bords de chaque diatomée de notre base de données. Il faut préciser que notre base de données d'images de diatomées origine d'images capturées à l'aide d'une loupe binoculaire couplée à un ordinateur. Les diatomées numérisées sont immergées dans une solution aqueuse provenant d'échantillons de terrain et on observe la présence de déchets végétaux et des fragments d'autres diatomées, elles sont conséquemment très bruitées. Ces échantillons de terrain ont été prélevés, pendant le printemps et l'automne 2002 ainsi qu'à l'automne 2003, par l'équipe du professeur Stéphane Campeau, du département des sciences de l'environnement de l'UQTR, dans une soixantaine de cours d'eau du Québec [22]. À partir de cette base originale d'images, chaque diatomée a été isolée manuellement du son fond bruité par Q. Hu (stagiaire) [23] après des tests d'extraction automatique de chaque diatomée. Plus récemment, E. Guittérèz-Lira [23] a adapté la méthode de seuillage de Rosin [29] pour extraire, de façon automatique, les diatomées du fond des images.

À la suite de cette étape de prétraitement, l'extraction du contour de la diatomée et celle des angles normaux du contour sont réalisées. Nous obtenons une première représentation du contour orienté par une suite ordonnée de pixels 8-connectés et une seconde représentation par les angles normaux (un angle par pixel) calculés en fonction des paires de déplacements successifs basées sur un codage de Freeman le long du contour. Une transformée de Fourier de la courbe périodique des angles normaux est calculée afin d'approximer chaque contour avec un nombre réduit de variables. Ainsi, nous obtenons une troisième représentation approchée du contour. Les premiers descripteurs de Fourier formeront un premier ensemble de paramètres discriminants utilisés dans la classification des images de diatomées.

Dans le quatrième chapitre nous décrivons les différents paramètres de formes utilisés par Siddour [2] auxquels nous ajouterons les valeurs propres du Laplacien discret défini sur le domaine de chaque diatomée. Au chapitre 5 nous présentons les résultats obtenus lors d'une classification bayésienne effectuée sur les variables obtenues à partir d'une analyse en composantes principales effectuée à partir de plusieurs sélections de paramètres retenues.

Au dernier chapitre, nous présentons nos conclusions et les perspectives envisagées pour les futurs travaux.

Chapitre 2

La biologie des diatomées

2.1 Description des diatomées

Les diatomées, ou *Bacillariophycées*, sont des algues microscopiques unicellulaires qui se développent sous forme de colonies (filaments, rubans ou résidus tubulaires) (voir Figure 2.1). Leur nom est issu du mot grec « *dia* » qui veut dire « en divisant » et du mot « *temnein* » qui signifie « couper ». Les parois cellulaires de ces algues sont formées de coquilles siliceuses. Ces microorganismes sont repartis, et de façon omniprésente, principalement dans les écosystèmes aquatiques [3].

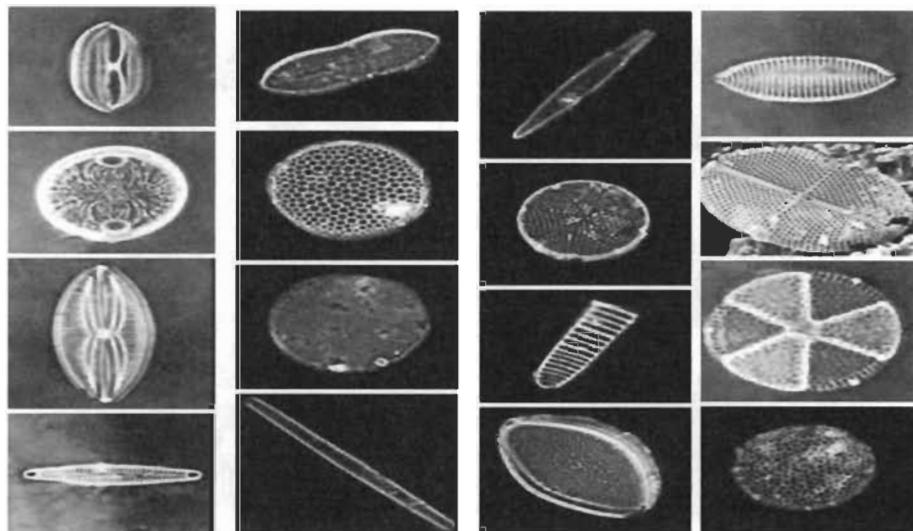


Figure 2.1 Exemples de diatomées.

2.2 Morphologie des diatomées

Les diatomées appartiennent au groupe des algues, un des plus riches groupements d'espèces sur la planète. Ces organismes sont bien conservés sous forme fossile, en raison de leur squelette externe de silice. Cette propriété est fort utile en paléoécologique afin de reconstituer les conditions environnementales des périodes anciennes, et permet actuellement d'identifier plus de 100 000 espèces de diatomées, principalement sur la base de caractères morphologiques [4].

Les cellules individuelles des diatomées sont difficilement détectables à l'œil nu. En effet, la plupart des spécimens mesurent de 10 à 200 µm [5]. Les squelettes ou frustules des diatomées sont composés de deux parties, qu'on appelle des valves, qui s'emboîtent une dans l'autre, comme une boîte et son couvercle au moyen de deux anneaux qui forment les bords et qu'on appelle les zones. Les frustules peuvent prendre des formes variées allant de circulaire à elliptique, en forme d'aiguille, de sigmoïde et polygonale [16] (voir Figure 2.1). Selon la géométrie des frustules, les diatomées sont divisées en deux grandes catégories : les diatomées centrales et les pennées. Les diatomées centrales ont une symétrie radiale avec un frustule ayant plusieurs stries. Les diatomées pennées sont caractérisées par une symétrie bilatérale avec des stries distribuées autour d'un plan longitudinal appelé le sternum [16] (voir Figure 2.2).

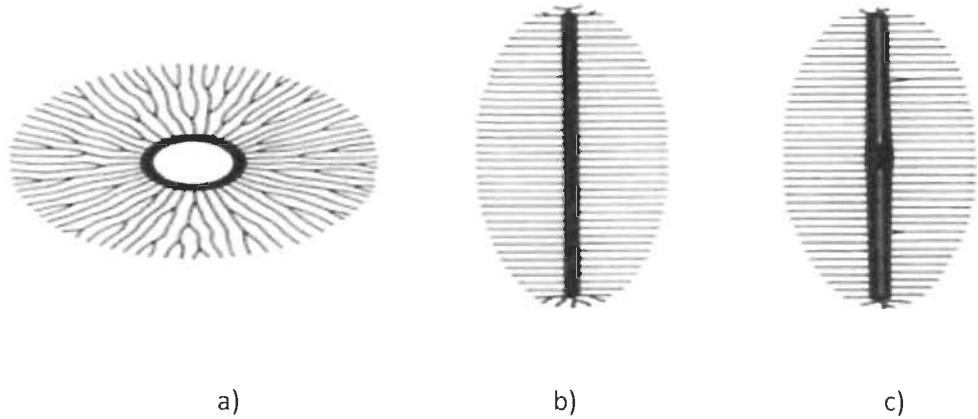


Figure 2.2 Principales formes de diatomées : a) centrique, b) pennée avec nervures, c) pennée avec sternum, raphé et strie.

La présence d'une valve est une caractéristique importante utilisée pour l'identification des espèces des diatomées. La taxonomie des diatomées est principalement basée sur la forme du frustule, la taille et l'ornementation de la valve [6, 17]. Les figures 2.3 et 2.4 illustrent respectivement les différentes parties d'une diatomée pennée et une image microscopique d'une culture de diatomées pennées. Pour la plupart des espèces de diatomées, les valves sont essentiellement composées de deux parties (chacune des faces) en silice, séparées par des rangées de pores [1]. Ces pores facilitent le passage de l'eau à travers la cellule, ainsi que la sécrétion de matière organique. Certaines diatomées sont également caractérisées par une ligne médiane, droite ou sigmoïde, qui traverse la surface des valves dans le sens de sa longueur, appelée raphé et muni généralement d'un nodule central et de deux nodules terminaux [1]. Le nodule central aide à distinguer deux grands ordres de diatomées : les *pennées*, qui contiennent un sternum avec des séries de stries et les *centriques*, qui contiennent une structure de stries radiales [1].

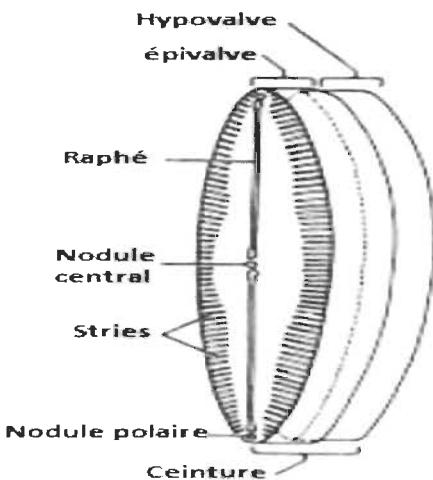


Figure 2.3 Composantes d'une diatomée pennée [17].



Figure 2.4 Image microscopique de diatomées en culture (*S. robusta*).

Le processus d'identification taxonomique est complexe à cause de la variabilité de la forme et de la taille de la valve chez les diatomées. Récemment, des études moléculaires ont été utilisées pour raffiner les descriptions morphologiques des espèces de diatomées et ces études ont révélé que des espèces peuvent être morphologiquement similaires mais génétiquement distinctes [7].

2.3 Classification des diatomées

Selon l'estimation des biologistes, on trouve plus de 200 000 espèces de diatomées. Il existe plusieurs principes de classification des diatomées, soit une classification basée sur la structure et la morphologie de la valve ou encore sur la forme générale du frustule. En règle générale, les diatomées sont divisées en deux grandes classes (voir Figure 2.5); les diatomées centrales et les diatomées pénnales. Les diatomées centrales ou centriques ont des valves circulaires ou elliptiques dont les ornementations sont disposées par rapport à un point central [1]. Les diatomées pénnales ont des valves plus ou moins allongées dont les ornementations sont disposées par rapport à sa médiane. Chez beaucoup d'espèces pénnales, les deux valves ou une seule, sont parcourues en leur milieu, par une fente étroite nommée le raphé. Cette fente est interrompue en son milieu par un épaississement siliceux [2]. Le raphé permet une communication avec le milieu extérieur et l'excrétion de mucilage. Si le raphé est atrophié ou peu marqué, on parle de pseudo-raphé. Les pénnales sans raphé sont aussi appelées diatomées araphidées ou crypto-raphidées [27].

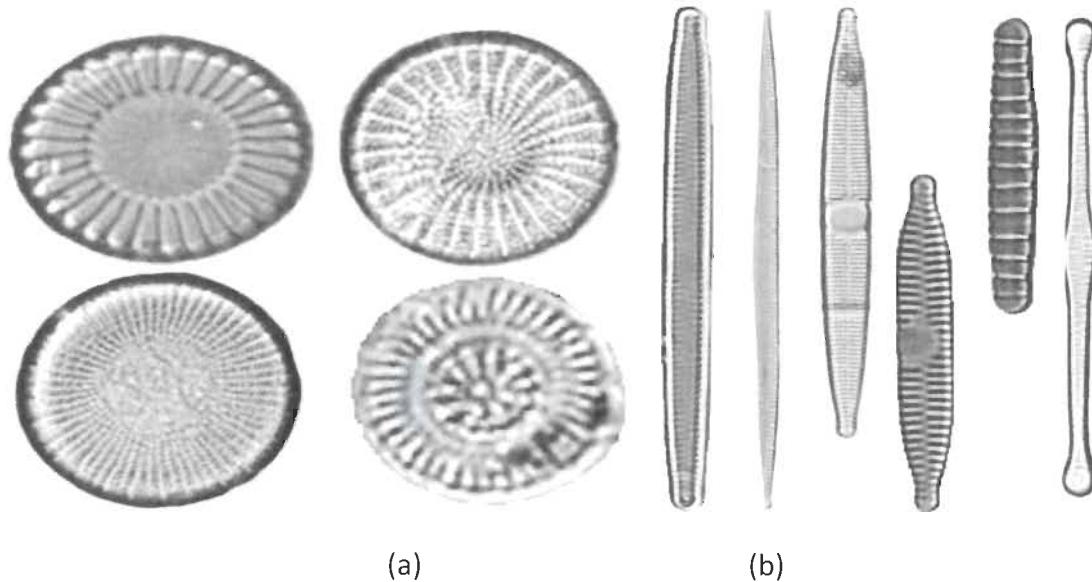


Figure 2.5 (a) Diatomées centrales, (b) diatomées pénnales.

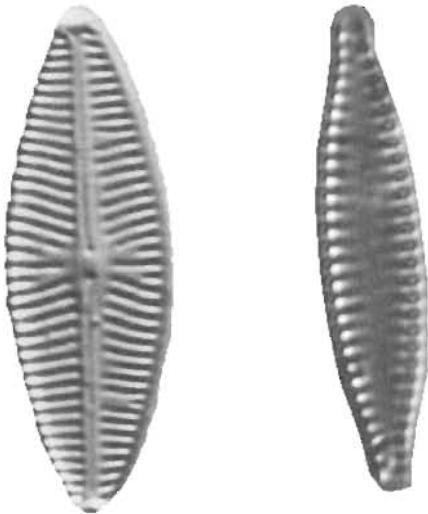


Figure 2.6 Deux diatomées similaires mais d'espèces différentes.

2.4 Écologie des diatomées

Le vent et les animaux participent à la dispersion des diatomées sur de grandes distances. Par conséquent, la distribution des espèces de diatomées est communément cosmopolite [4]. Les diatomées vivent dans de nombreux habitats. Elles sont présentes dans les eaux marines, salées et douces, en zones terrestres, partout sur la planète. Ces organismes constituent également une partie importante de la flore nordique d'algues d'eau douce [11]. En plus de l'eau, les diatomées ont besoin de la présence de lumière et de nutriments adéquats, en particulier pour la formation de la silice du frustule [7]. Plusieurs travaux sur les diatomées font état des préférences écologiques des diatomées et ont révélé que ces communautés sont influencées par les conditions régionales telles que le climat et la géologie ainsi que par des facteurs locaux comme la physico-chimie, la diversité des habitats, la disponibilité de la lumière, l'inclinaison de la pente [14, 15, 16]. La disponibilité des éléments nutritifs affecte fortement la dynamique de la population de diatomées. Les cellules des diatomées sont diploïdes et utilisent la mitose pour se reproduire. La mitose produit deux cellules filles dont une à une taille inférieure à celle de la cellule mère, tandis que la cellule fille a la même taille que la diatomée initiale. En conséquence, lors des divisions cellulaires successives, la taille des diatomées diminue [18, 20]. La limitation des

nutriments à la fin d'une période de floraison des diatomées produit souvent des interruptions dans la phase du cycle de vie de la diatomée. En effet, la cellule passe de la multiplication végétative à la formation de spores ou à la reproduction sexuée [21].

2.5 Applications scientifiques des diatomées

L'étude des diatomées est importante dans plusieurs domaines dont la nanotechnologie, la production de biocarburants, la médecine légale et le suivi des changements climatiques et environnementaux, pour ne citer que ceux-ci. Il est reconnu que l'étude des diatomées est plus contributive si elle se base sur une classification robuste [22]. Les diatomées sont également utiles pour déchiffrer les conditions climatiques et environnementales passées car la composition des communautés de diatomées est une variable sensible aux changements environnementaux, sachant qu'elles se caractérisent par des préférences écologiques spécifiques [5].

2.6 Utilisation des diatomées au Canada

En raison de leur importance, des équipes de biologistes ont établi un « indice diatomée » lequel permet une quantification de l'état de santé écologique des cours d'eau, ainsi que son statut trophique, à partir de l'étude de ses communautés de diatomées [22]. Cet indice permet de mesurer la différence entre l'état des diatomées dans un environnement aquatique sain ou un environnement pollué. Les valeurs de cet indice varient selon une échelle de 0 à 100. Plus la valeur de l'indice est élevée, plus la qualité de l'environnement aquatique est grande et l'intégrité biologique de qualité supérieure. Les informations et les données sur l'environnement obtenues à partir de l'étude des diatomées permettent de révéler des modifications structurales majeures qui touchent une population de diatomées, par exemple la présence dominante des espèces qui prolifèrent dans un environnement pollué. Ceci permet, à l'aide de gradients environnementaux, d'évaluer un écosystème en analysant la structure des populations de diatomées [1].

2.7 État de l'art sur la classification automatique des diatomées

Le sujet de la classification automatique des diatomées a fait l'objet de plusieurs travaux. Dans de récents travaux, le contour, la taille, la forme, l'ornementation et la symétrie de la valve ont été parmi les caractères morphologiques utilisés pour la classification et l'identification automatique des diatomées. Nous présentons dans cette section, des exemples d'approches déjà utilisées pour la classification des diatomées.

En 1998, le programme européen MAST a financé un projet pilote sous le nom d'ADIAC (Automatic Diatom Identification and Classification) dans le but de classifier les diatomées, et aussi de développer des algorithmes de traitement d'images complètement non-supervisés. Ces algorithmes s'appuient sur des caractéristiques visuelles de l'image comme la forme du frustule et les ornementations des valves. La base de données du projet ADIAC contient environ 3400 images de diatomées [36].

Dans le cadre de ce projet, différentes techniques d'identification ont été utilisées, basées sur les caractéristiques géométriques des diatomées telles que les classes de symétrie, la largeur, la longueur, la taille, les descripteurs de Fourier, pour ne citer que celles-ci [39].

Des méthodes de classification ont fait usage de la description du contour des diatomées par des polynômes de Legendre associés à la transformée de Gabor, de la courbure de segments convexes ou concaves du contour, de la densité des motifs des ornementations [39].

Dans Dimitrovski et *al.* [40] les auteurs procèdent de façon standard, en trois étapes, premièrement en segmentant l'image de la diatomée, puis en extrayant les caractéristiques basées sur les contours et la texture, et finalement ils utilisent une classification hiérarchique. La première étape vise à enlever le bruit du fond et à extraire le contour des images de diatomées. Après l'étape d'extraction du contour, des descripteurs de Fourier sont utilisés pour représenter le contour de la diatomée, tandis que les

histogrammes SIFT (Scale Invariant Feature Transform) permettent de détecter et de décrire les caractéristiques locales dans les images, dans ce cas le frustule. Dans l'étape de classification, les chercheurs ont utilisé les PCT (Predictive Clustering Trees) comme classificateurs. Les PCT exploitent la taxonomie hiérarchique et simultanément prévoient toute la hiérarchie taxonomique et les classifient sous différents niveaux : le genre, l'espèce, la variété et la forme. Dans cette étude, un sous-ensemble de 1098 images (saisies sous microscope) sont classées selon la hiérarchie taxonomique. L'approche proposée est validée avec la base de données d'images de diatomées d'ADIAC. Ce système produit un classificateur très efficace offrant des performances prédictives élevées de 97.15%

L'équipe de Jalba et *al.* [41] a proposé une méthode d'identification des diatomées basée sur des espaces d'échelle de courbure morphologique pour l'extraction de caractéristiques basée sur les contours. À nouveau, les images de diatomées sont évaluées sur la base de données ADIAC. Après avoir extrait le contour de la diatomée, celle-ci est lissée de manière adaptative, et représentée à l'aide d'une chaîne de Freeman. Finalement, on calcule la fonction de courbures invariante à la translation et à l'homothétie [41]. Un vecteur de caractéristiques est construit en calculant le nombre de pics (vallées), la courbure moyenne et la variance et l'aire de la diatomée. Les résultats ont montré qu'il est possible d'obtenir une précision d'identification de 84% [41].

Chahboune et *al.* [37] ont classifié les diatomées d'une façon semi-automatique. Dans un premier temps, les auteurs ont construit une base de données des diatomées échantillonnées situées sur le plateau de la Mamora près de la ville de Kénitra (Maroc). Cette base de données est établie tel que les caractéristiques des diatomées sont stockées dans un vecteur indexé en fonction des espèces de diatomées possédant des caractéristiques comme le contour et l'ornementation des valves [37]. L'objectif est de déterminer dans quelle mesure deux diatomées sont similaires à partir des images stockées dans la base de données de départ et de déterminer à quelles classes elles appartiennent. Pour ce faire, une comparaison est faite entre une identification visuelle par la forme de la diatomée et une identification automatique en utilisant ImageJ ; un logiciel de traitement

d'images développé par le National Institute of Health [37]. Les résultats ont montré que le taux de classement est d'environ de 84%.

Une autre approche de classification automatique a été proposée par Claudon [1]. Cette approche consiste à classifier les diatomées d'une base de données construite par l'équipe du professeur Stéphane Campeau [22] et adaptée par le laboratoire LIRIC de l'UQTR, à l'aide d'une analyse des caractéristiques géométriques des structures internes des frustules. Claudon [1] a notamment utilisé l'orientation et la fréquence des stries présentes sur le frustule de chaque spécimen extrait en utilisant une transformée de Fourier bidimensionnelle. Les trois premières composantes du vecteur de représentation d'une diatomée correspondent aux valeurs booléennes (0 pour absence, 1 pour présence) indiquant si le frustule a un raphé, un raphé centré ou nodule central. Deux grandes étapes sont nécessaires pour extraire les trois composantes du vecteur. Pour ce faire, la première étape consiste à extraire un masque binaire de la diatomée, puis à extraire un masque des zones de motifs. L'étape finale est celle de la classification. Cette classification se fait en trois étapes. La première étape est la création d'un vecteur moyen des caractéristiques pour chaque groupe d'espèces à analyser (construites manuellement) afin de constituer une base de référence pour les images testées. La deuxième étape est la présélection des classes possédant le même type de structures que l'image à classer. Enfin, une mesure de distance euclidienne minimale entre les orientations et les fréquences des classes sélectionnées et celles de l'image est testée. Le taux de bonne classification est d'environ 91%.

Pour classifier automatiquement les diatomées, Abdelkader Siddour [2] a, dans un premier temps, extrait les angles normaux du contour 8-connecté antihoraire des images de diatomées binaires, déjà épurées manuellement de leur fond bruité. Pour chaque diatomée, un angle est calculé à chaque pixel du contour en fonction des deux pixels adjacents, est lissée à l'aide d'un nombre restreint de descripteurs de Fourier [2]. Le nombre de descripteurs retenus a été fixé expérimentalement. Ces descripteurs constituent un premier sous-ensemble de paramètres discriminants dans la classification.

Puis, un ensemble de paramètres de formes a été calculé, toujours selon le contour. Les paramètres retenus, après des tests sur un plus grand nombre de paramètres, sont l'indice de circularité, de convexité, d'allongement, de rectangularité et de symétrie.

Une fois les images segmentées et les caractéristiques extraites, le problème de la classification est abordé. Étant donné le nombre élevé de paramètres, une analyse en composantes principales est faite pour diminuer le nombre de ces paramètres. Les valeurs des trois premiers axes sont retenues, lesquels expliquent 91% de la variance dans l'espace initial. Les 449 images de diatomées de la base de données, ont été partagé en 41 classes selon leurs formes par l'équipe du professeur Stéphane Campeau [22]. Dans ce travail, on a conservé une répartition équitable dans chaque regroupement. Sur le premier sous-ensemble de référence, une classification bayésienne est effectuée, soit le calcul des probabilités d'appartenance de chacune des diatomées à sa classe à partir des cotes obtenues sur les axes principaux de l'ACP [2]. Finalement, une classification successive est faite sur 100 échantillons de 41 diatomées tirés au hasard dans le sous-ensemble test. Les résultats montrent un taux de bon classement de 97,81%

Nous avons présenté un état de l'art sur la classification d'images de diatomées. Cela nous a permis d'explorer quelques approches réalisées et les résultats obtenus. À la lumière de ces résultats, nous avons opté pour une approche simple, soit celle initiée par Siddour [2] que nous bonifierons par l'ajout des valeurs propres du problème du Laplacien discrétilisé [33]. Nous testerons quel sous-ensemble de paramètres, parmi ceux retenus, permettent un meilleur classement des diatomées.

Chapitre 3

Prétraitement des images et extraction du contour et des angles normaux

Dans ce chapitre, nous présentons les méthodes du prétraitement des images, tel que le lissage, la binarisation des images en niveaux de gris et enfin la rotation des images. Dans la deuxième partie, nous présenterons les techniques d'extraction des caractéristiques de la diatomée, dont le contour et des paramètres géométriques. À partir du codage de Freeman, nous obtenons une représentation du contour d'une diatomée à l'aide des angles normaux le long de celui-ci. Cette représentation du contour par les angles normaux est une fonction périodique dont nous calculons la transformée de Fourier afin d'obtenir une représentation condensée.

3.1 Prétraitement des images de diatomées

Les prélèvements des diatomées de notre base de données ont été réalisés dans 126 stations du réseau des rivières sous surveillance du ministère du Développement durable, de l'Environnement et des Parcs du Québec, par l'équipe du professeur S. Campeau [22]. Chaque échantillon a été analysé et les diatomées répertoriées. Chaque spécimen d'un inventaire des espèces présentes a été photographié et stocké dans une base de données. Puisque ces images sont saisies à l'aide d'un binoculaire et baignant dans une eau de rivières, il est nécessaire d'isoler, dans l'image, chaque diatomée de son fond bruité, parsemé de débris de végétaux et autres artefacts. Guo et al. [22] ont isolé manuellement (efface électronique) chaque diatomée de la base de données (voir figure 3.1).

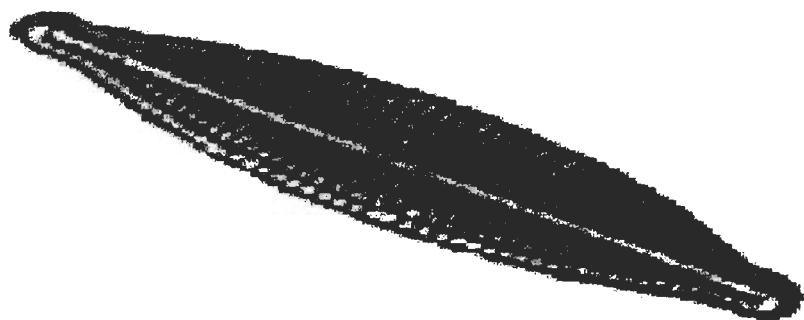
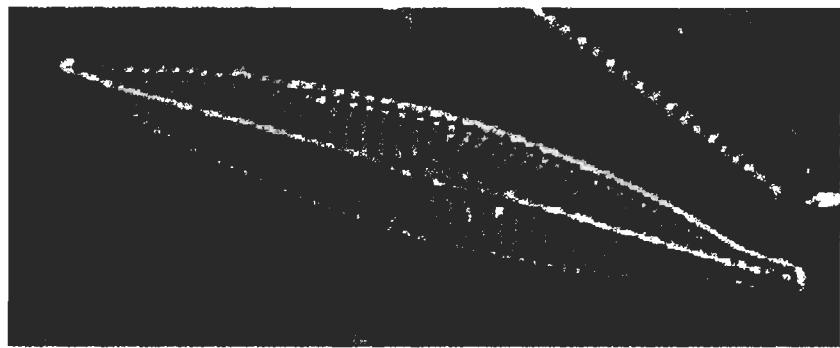


Figure 3.1 Diatomée isolée manuellement.

Le prétraitement d'une image correspond à toutes les transformations appliquées aux images de notre base de données afin de supprimer le bruit dans celles-ci et de corriger le contraste. Les opérateurs de prétraitement appliqués sont, dans l'ordre :

- Le lissage de l'image;
- La binarisation d'image;
- La rotation d'une image de diatomée autour de son axe principal.

3.1.1 Lissage d'une image

Le prétraitement d'images vise la réduction du bruit parasite d'une image acquise par une caméra ou tout autre système d'acquisition et aussi l'amélioration du contraste. L'image peut être corrompue par des variations dues à l'éclairage ou à la présence

d'artéfacts indésirables. Le lissage est une opération de traitement d'image importante car elle permettra idéalement de préciser les contours des structures étudiées [42].

Filtrer une image signifie convoluer la fonction impulsionnelle du filtre $f(x, y)$ avec la fonction d'intensité $I(x, y)$ [42]. Cette convolution consiste à remplacer chaque niveau de gris $I(x, y)$ par une combinaison linéaire des niveaux de gris des points voisins dont les coefficients de la combinaison sont définis par la réponse impulsionnelle du filtre $f(x, y)$

$$I'(x, y) = f(x, y) * I(x, y). \quad (3.1)$$

Dans la prochaine section, nous présentons les filtres utilisés dans notre approche.

3.1.1.1 Le filtre médian

Le filtrage médian réalise un lissage plus performant surtout dans différents modèles de bruits.

Chaque pixel est traité en considérant ses voisins dans une fenêtre de taille donnée. Le pixel traité est remplacé par la valeur médiane des pixels de la fenêtre [43]. La fenêtre peut être carrée $n \times n$, une croix $n \times n$, une ligne ou une colonne. Les effets de ce filtre changent en fonction de la taille de la fenêtre.

Le filtre médian offre deux avantages :

- Le filtrage médian ne change pas les bords des contours.
- Étant donné que la valeur médiane doit être la valeur de l'un des pixels du voisinage, le filtre médian ne crée pas de nouvelles valeurs de pixel aberrantes comme le filtre moyen, qui calcule la moyenne des pixels aux voisins pour nous donner une nouvelle valeur.

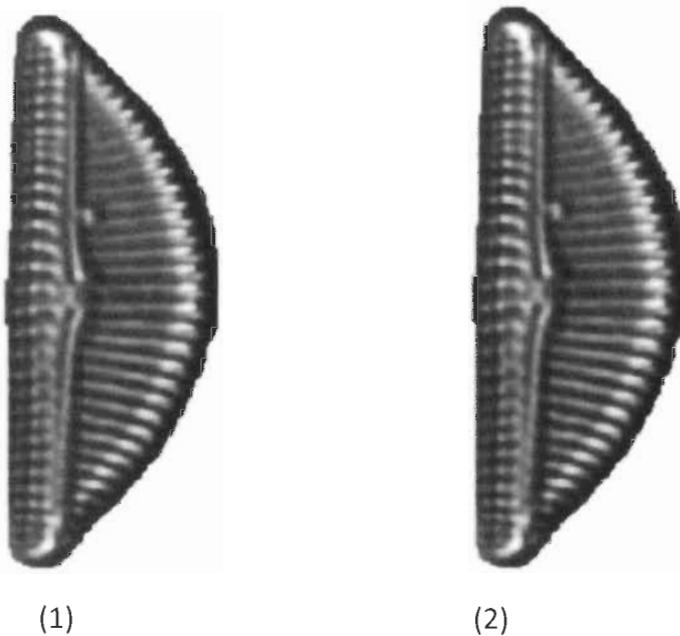


Figure 3.2 Filtre médian à l'aide d'une fenêtre 3×3 . À gauche, l'image originale et à droite le résultat de l'application du filtre.

3.1.1.2 Le filtre moyen

Le filtre moyen est la méthode la plus utilisée pour réduire le bruit dans une image. Chaque pixel est traité en considérant ses voisins sur un voisinage donné. Le pixel traité et ses voisins forment un ensemble dont on calcule la valeur moyenne, le pixel sera donc remplacé par celle-ci [43]. Le filtre moyen se base autour d'un noyau qui représente la forme et la taille du voisinage à échantillonner lors du calcul de la moyenne. Généralement, un noyau carré de 3×3 est utilisé, cependant l'utilisateur peut choisir une autre taille.

Les deux principaux problèmes du filtrage moyen sont :

- Un seul pixel avec une valeur aberrante peut affecter la valeur moyenne de tous les pixels de son voisinage.

- Lorsque le voisinage du filtre chevauche un bord du contour, le filtre interpole de nouvelles valeurs pour les pixels sur le bord et va donc flouter ce bord.

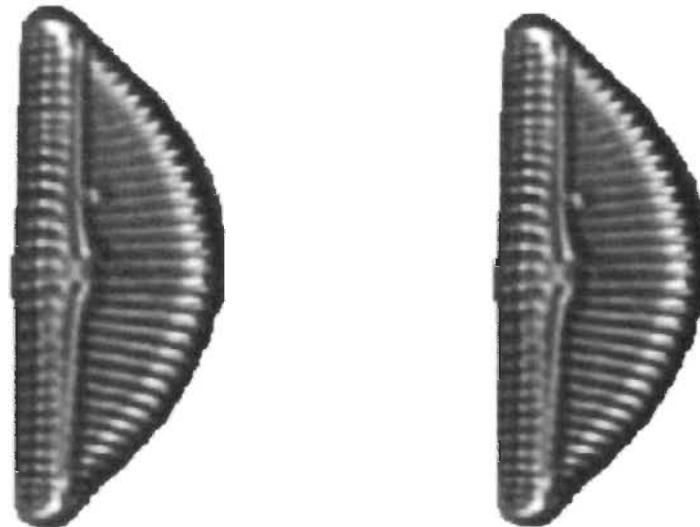


Figure 3.3 Filtre moyen avec une fenêtre 3×3 . À gauche, l'image originale et à droite le résultat de l'application du filtre.

Comme on peut le constater sur les deux images ci-dessus les deux filtres n'ont qu'une faible répercussion sur la qualité de l'image, ce qui rend ces filtres inadaptés dans le cas du lissage de l'image d'une diatomée. La diminution du bruit n'étant presque pas visible à l'œil nu. On constate donc que ces deux filtres ne sont pas appropriés dans le cadre de notre recherche.

3.1.2 Binarisation d'image

Dans l'analyse d'images, la binarisation est souvent une des premières étapes utilisées avant l'étape de la reconnaissance. Elle a donc une grande influence sur la performance des étapes suivantes et sur le résultat final [21]. C'est une technique importante dans les applications de traitement d'images. La binarisation d'une image est un processus de transformation d'une image en niveaux de gris en une image en noir et blanc. Dans une image en niveau de gris, un pixel peut prendre 256 valeurs d'intensité différentes tandis

que dans une image binaire chaque pixel est soit noir ou blanc [21]. Cette conversion est effectuée en établissant, idéalement de façon automatique, la valeur d'un seuil pour départager les pixels d'une image en deux classes. Si le seuil est établi trop bas, l'image binaire résultante sera principalement composée de pixels blancs et inversement, si le seuil est trop haut, l'image résultante mettra en évidence un grand nombre de pixels noirs, éventuellement indésirables [21]. Plusieurs algorithmes ont été étudiés et testés dans le cadre de notre travail, nous avons retenu l'algorithme d'OTSU qui offrait le meilleur résultat.

La méthode d'Otsu permet un seuillage automatique à partir de la forme de l'histogramme des valeurs des niveaux de gris dans l'image et montre de meilleures performances que les autres méthodes de binarisation [44].

Dans la méthode d'Otsu, nous prenons comme valeur initiale du seuil t , la valeur moyenne de l'image d'origine, ce qui produit deux classes de pixels de valeurs d'intensité réparties de part et d'autre du seuil. Ensuite, nous trouvons les valeurs moyennes de chacun des deux classes. Puis, nous calculons le nouveau seuil. Le seuil optimal est celui pour lequel la variance intra-classe est minimale.

L algorithme Otsu se défini comme la somme pondérée des variances des deux classes :

$$\delta^2_w(t) = \omega_0(t) \delta_0^2(t) + \omega_1(t) \delta_1^2(t). \quad (3.2)$$

Où ω_0 et ω_1 sont les probabilités des deux classes séparés par le seuil t et δ_0, δ_1 sont les variances de chaque classe.

$$\omega_0 = \sum_{k=1}^t P(k). \quad (3.3)$$

$$\omega_1 = \sum_{k=t+1}^{256} P(k). \quad (3.4)$$

$$\delta_0^2 = \frac{\sum_{i=1}^t N_1(i) - \text{Moy}_1(t) * P(i)}{\omega_0}. \quad (3.5)$$

$$\delta_1^2 = \frac{\sum_{i=t+1}^{256} N_2(i) - \text{Moy}_1(t) * P(i)}{\omega_1}. \quad (3.6)$$

Où N_1 est un vecteur de de 0 à $t - 1$ et N_2 est un vecteur de t à 255. La probabilité de chaque niveau de gris $P(k)$ est calculé en divisant le nombre de pixels présent pour chaque niveau de gris par le nombre total de pixel dans l'image.

La moyenne de chaque classe est calculée en sommant le vecteur N qui est multiplié par la probabilité de chaque niveau de gris. Le tout est ensuite divisé par la probabilité de la classe.

$$moy1(t) = \frac{\sum_{i=1}^t N_1(i)*P(i)}{\omega_0(t)} \quad (3.7)$$

$$moy2(t) = \frac{\sum_{i=1}^{256} N_2(i)*P(i)}{\omega_1(t)}. \quad (3.8)$$

Finalement, pour remplir les trous noirs à l'intérieur de la diatomée, nous appliquons une fonction appellé *Imfill* de matlab (voir figure 3.4). Ce remplissage est basé sur les points noirs à l'intérieur de premier contour trouvé.



Figure 3.4 Processus de la binarisation. À gauche, l'image originale, au milieu l'image binaire obtenue avec la méthode OTSU et à droite l'image finale.

3.1.3 Rotation d'une diatomée

Notre objectif est d'extraire des paramètres de classification d'images de diatomées, pour cela nous devons standardiser nos images, soit l'orientation de celles-ci. Nous avons des images déjà isolées de leur fond et le contour est facile à extraire. Avec ce contour, on calcule

l'axe principal et on fait une rotation horizontale. Une rotation en mathématiques consiste à déplacer le long d'un arc de cercle chacun des pixels, en préservant la forme de l'objet (ici une diatomée) et en appliquant un même angle de rotation sur toute l'image. Une rotation dans le plan s'effectue à l'aide d'une matrice de rotation. Un point de coordonnée $P = (x, y)$ subira un déplacement dans le sens anti-horaire d'un angle θ à l'aide d'une multiplication matricielle [2] dont la matrice qui multiplie le vecteur position est donnée par;

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \quad (3.9)$$

L'angle de la rotation de la diatomée est défini à partir de l'axe principal de la diatomée. L'axe principal de la diatomée comme étant le plus long segment liant deux points O et O' du contour de la diatomée. Ici, nous supposons que le contour a été isolé (voir plus loin). Le segment de longueur $|OO'|$ vérifie;

$$|OO'| = \max_{1 \leq i, j \leq L} d(C_i, C_j) \quad (3.10)$$

où C_i et C_j sont deux points distincts pris dans le domaine de la diatomée (pixels blanc) et L est le nombre de pixels contenus dans la diatomée.

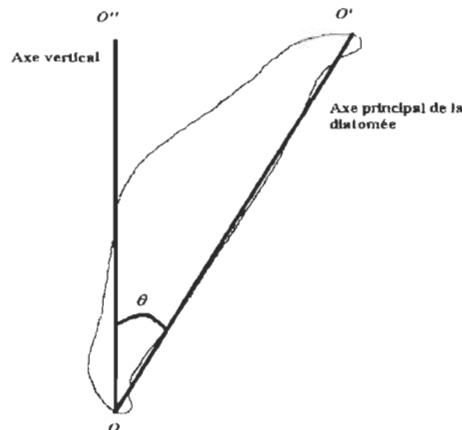


Figure 3.5 Axe principal d'un contour d'une diatomée.

L'angle de rotation d'une diatomée correspond à l'angle que forment l'axe de la diatomée et l'axe vertical passant par le point O . Un point (x, y) aura comme image, après rotation, le point (x', y') suivant;

$$\begin{cases} x' = (x - x_0)\cos\theta - (y - y_0)\sin\theta + x_0 \\ y' = (x - x_0)\sin\theta + (y - y_0)\cos\theta + y_0 \end{cases} \quad (3.11)$$

où (x_0, y_0) sont les coordonnées du centre $(0,0)$ par rapport auquel est effectuée la rotation d'angle θ . Nous avons privilégié une position à l'horizontale comme orientation du plus grand axe [25], cela pour chaque image (voir figure 3.7).



Figure 3.6 Image binaire d'une diatomée avant et après la rotation.

3.2 Extraction du contour et des angles normaux

L'analyse d'image est une science récente nécessitant l'obtention d'images à travers une chaîne d'acquisition et des outils mathématiques de traitement adaptés. Il existe un nombre considérable d'algorithmes qui permettent d'extraire les principaux constituants d'une image [26]. Par exemple, le contour d'une forme présente sur l'image est un descripteur de forme puissant et facile à obtenir. L'extraction du contour est une étape préliminaire à de nombreuses applications de l'analyse d'images [26]. En effet, les contours constituent des indices riches qui peuvent nous aider dans notre travail. Nous calculons les contours 8-connectés des images des diatomées. Une fois le contour est structuré, grâce aux propagations progressives des pixels de contour, nous déterminons une suite d'indices

de Freeman. Ces indices nous permettent de lier chaque pixel de contour à un angle normal [2].

3.2.1 Extraction du contour

Dans le domaine de l'analyse d'image, il existe plusieurs algorithmes qui aident à extraire les objets constituant une image [30]. Nous considérerons le cas où un objet est caractérisé par son contour, il existe un algorithme simple et efficace qui permet de décrire le contour d'une forme basé sur la définition du code de Freeman [30].

On peut examiner les propriétés d'une image en exploitant le voisinage à quatre points noté V4(i, j) ou le voisinage à huit points noté V8 (i, j) d'un pixel (i, j) où i et j sont respectivement l'indice de ligne et l'indice de colonne du pixel dans l'image. Ces voisinages sont illustrés par la figure 3.7 [31]. Les valeurs possibles sont {0, 2, 4, 6} pour V4 et {0, 1, 2, 3, 4, 5, 6, 7} pour V8. Freeman sauvegarde le point P_1 (le point en haut à gauche qui est notre point de départ).

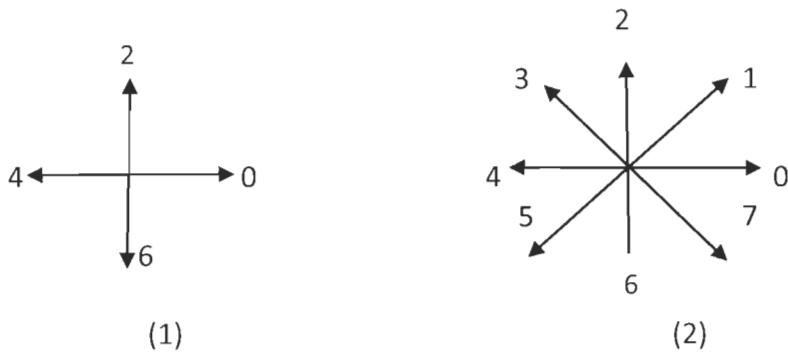


Figure 3.7 Indices de Freeman, (1) 4-Connectivité, (2) 8-Connectivité.

Tous les points du voisinage sont à une distance égale du point central. Cette distance égale à 1 pixel. On parle alors de 4-connectivité. (Un objet pour lequel le passage d'un pixel à l'un de ses voisins ne se fait qu'à la verticale ou l'horizontale est 4-connexe.), Tandis que pour le voisinage V8, cette distance égale à 1 pixel pour les pixels situés horizontalement ou

verticalement par rapport au pixel central et $\sqrt{2}$ pixels pour les pixels situés en diagonale du point central [31].

Le contour de la diatomée est un chemin fermé, antihoraire formé de pixels noirs et composant une chaîne de longueur au sens de la connectivité V8.

La procédure d'extraction du contour des diatomées est la suivante. Tout d'abord, nous sélectionnons un pixel P_1 du contour. Ce pixel a alors la position la plus haute à gauche. Deuxièmement, à partir de P_1 , nous cherchons tous les pixels voisins disponibles au sens de la 8 connectivité. Parmi les pixels voisins de P_1 , tous les pixels disponibles sont uniquement des pixels avec au moins un pixel vert adjacent parmi les 8 voisins connectés qui leur sont propres. Notons $P_1 = (x_1, y_1)$ les coordonnées du notre pixel de départ du contour C, le deuxième pixel du contour orienté est le pixel noir avec coordonnées $P_2 = (x_2, y_2) = (x_1+1, y_1-1)$. On associera au premier contour $[P_1, P_2]$, l'indice de Freeman $\alpha = 5$ [45] (voir figure 3.8). Notre but est de construire un contour 8-connecté formé de pixels noirs en s'inspirant des travaux de Chalifour et al. [45].

Troisièmement, de cette manière, de nombreux sous-ensembles sont extraits. Nous sélectionnons le sous-ensemble de longueur minimale comme contour de la diatomée.

$[P_1, P_2, \dots, P_1, P_{i+1}, \dots, P_L]$. Où L est la longueur de la chaîne du contour.

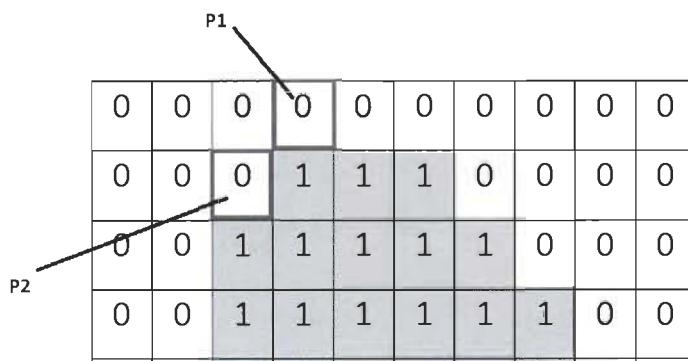


Figure 3.8 Position du premier pixel de départ.

Certains cas sont inadmissibles. Il y a 8 directions adjacentes possibles. Le tableau 3.1 explique les arcs $[P_i, P_j]$ possibles.

P_j/P_i	0	1	2	3	4	5	6	7
0	X	X			X	X	X	X
1	X	X	X	X		X	X	X
2	X	X	X	X			X	X
3	X	X	X	X		X	X	X
4	X	X	X	X	X	X		
5		X	X	X	X	X	X	X
6		X	X	X	X	X	X	X
7	X	X		X	X	X	X	X

Tableau 3.1 Directions admissibles.

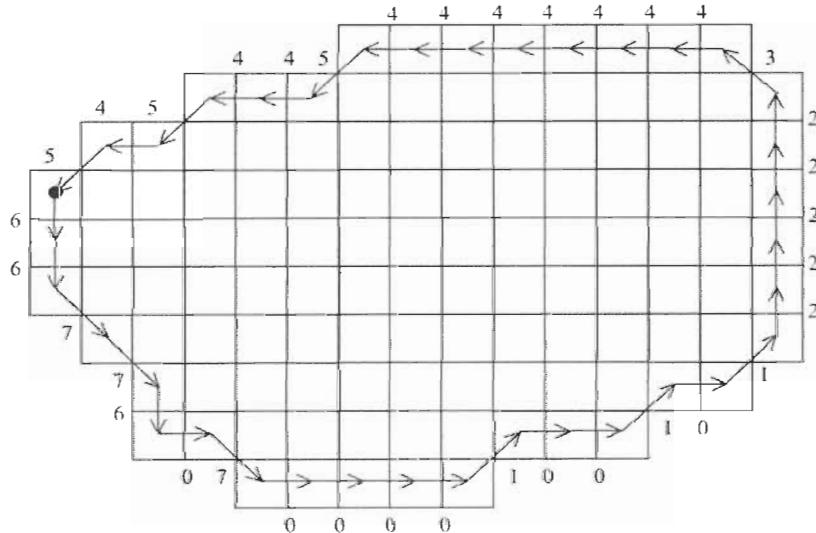


Figure 3.9 Exemple d'une diatomée et code de Freeman avec voisinage V8 [31].

On lie au contour C une suite de directions de Freeman $[\theta_1, \theta_2, \dots, \theta_k, \dots, \theta_l]$ où θ_k est la direction de l'arc $[P_k, P_{k+1}]$. Ces directions avec la suite de pixels nous aiderons à calculer les angles normaux au contour.

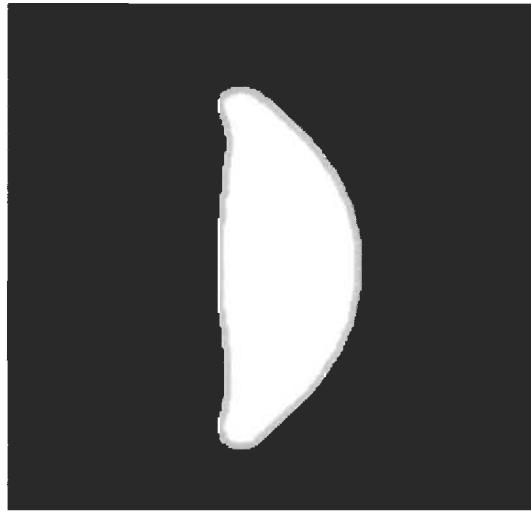


Figure 3.10 Contour initial de la diatomée.

3.2.2 Approximation polygonale

L'approximation polygonale consiste à supprimer et minimiser les points correspondant au bruit, tout en préservant des informations importantes sur la courbe [20]. D'une autre façon, elle consiste à transformer une chaîne de points de contours (le contour ici est la frontière des points blancs et noirs) en une suite de segments de droite. Au cours des dernières décennies, plusieurs méthodes d'approximation ont été proposées. Malgré le nombre important de méthodes d'approximation, il existe toujours des problèmes de robustesse, de stabilité et de complexité lors de l'application de transformations géométriques [28]. Notre décision se base sur la méthode Split-and-merge. Cette méthode est simple et facile à programmer. De plus, les résultats sont intéressants.

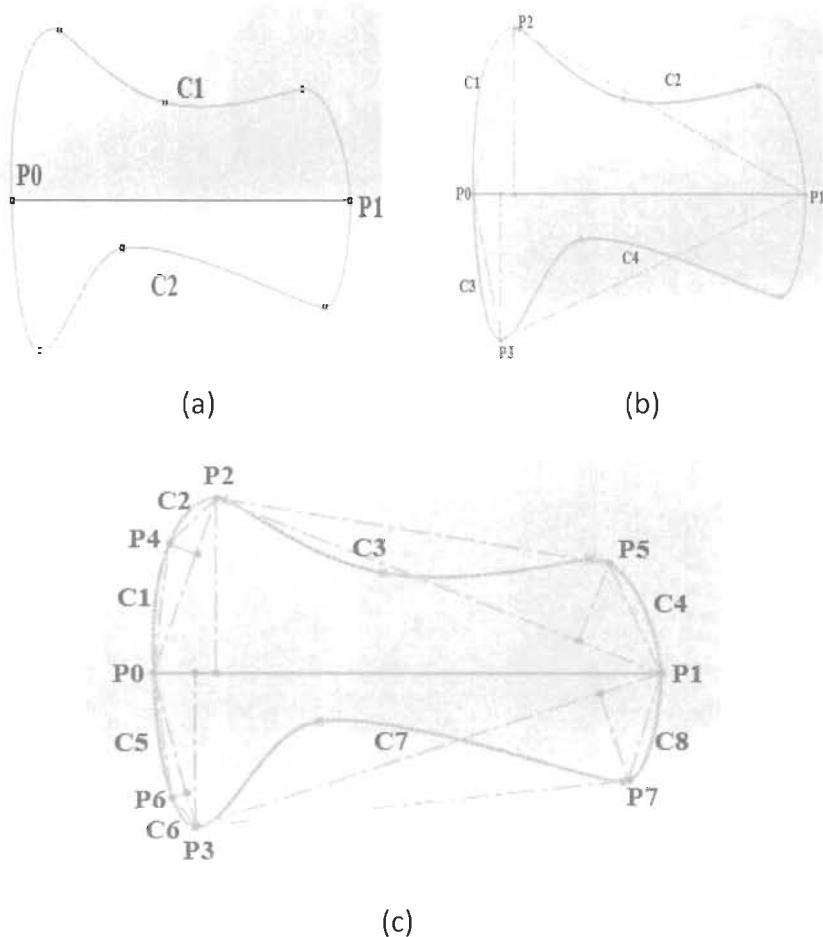


Figure 3.11 Processus de l'approximation polygonale.

Définissons une courbe polygonale P à N sommets fermée dans un espace à 2 dimensions comme l'ensemble ordonné de sommets $P = \{P_1, P_2, \dots, P_N\}$. L'algorithme d'approximation du seuil est le suivant. Tout d'abord, nous fixons $\epsilon > 0$, et nous sélectionnons les deux pixels avec la plus grande distance entre eux dans le contour. Ces deux pixels P_0 et P_1 forment un segment de droite et le contour est divisé en courbes C_1 et C_2 (voir figure 3.11 (a)). Après nous sélectionnons le point P_2 de C_1 le plus éloigné du segment $[P_0-P_1]$, ensuite nous recommençons de l'autre côté, nous sélectionnons le point P_3 de C_2 le plus éloigné de $[P_0-P_1]$ (voir figure 3.11 (b)). Cette procédure est répétée jusqu'à ce que la valeur maximale de la distance entre les segments et les courbes soit inférieure au seuil ϵ .

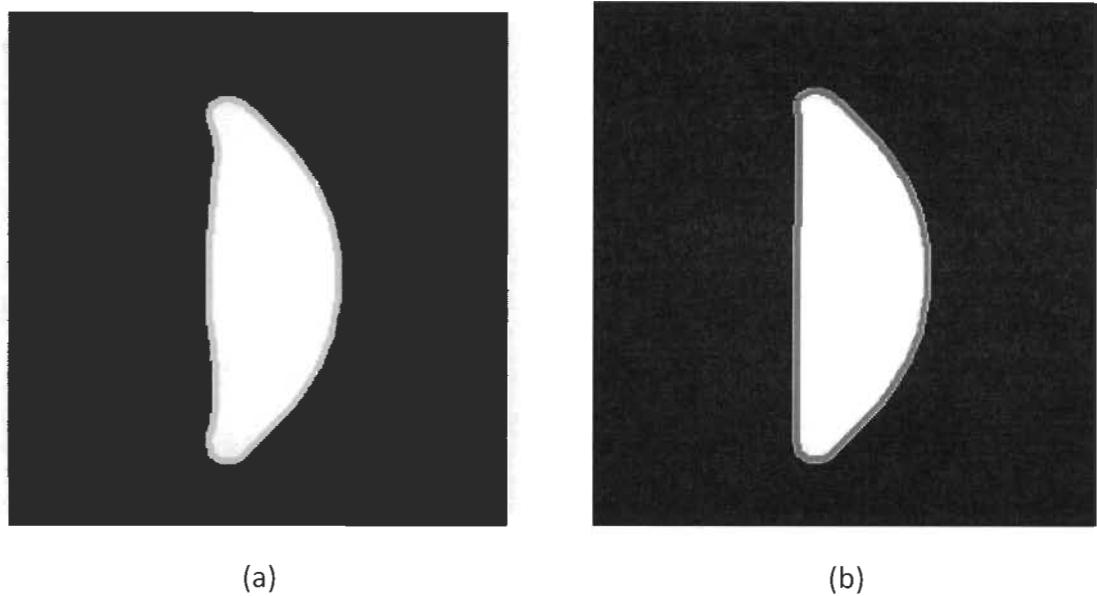


Figure 3.12 : (a) contour d'une diatomée, (b) contour après l'approximation.

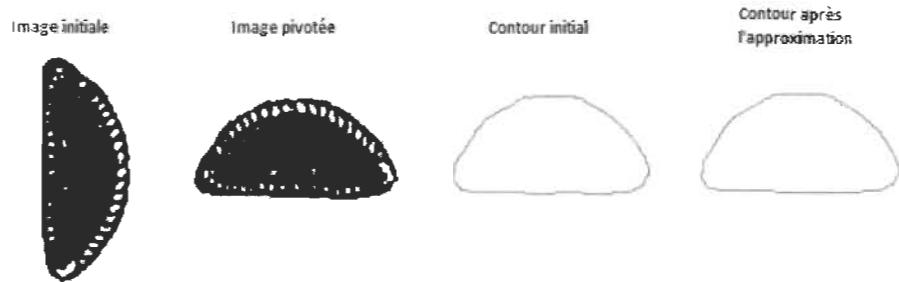


Figure 3.13 : Étapes pour extraire le contour d'une diatomée.

3.3 Extraction des angles normaux

Soit $[P_1, p_2, \dots, P_i, P_{i+1}, \dots, P_L, P_{L+1}]$ le contour orienté positivement d'une diatomée, nous associons en radians la valeur moyenne θ_i des angles normaux à chaque pixel p_i d'une paire de déplacements $[p_{i-1}, p_i, p_{i+1}]$ [3].

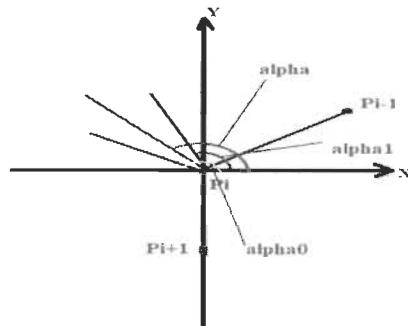


Figure 3.14 : Angle normal avec $\alpha_1 = 5$ et $\alpha = 6$.

α_i	0	1	2	3	4	5	6	7
0	$\pi/2$	β	X	X	X	0	$\pi/4$	B
1	B	$\pi/4$	α	0	X	X	$\pi/2$	$\pi/2$
2	$\pi/4$	α	0	α	X	X	$\pi/2$	$\pi/2$
3	0	0	α	$\pi/4$	β	$\pi/2$	X	X
4	X	0	$\pi/4$	β	$\pi/2$	B	X	X
5	X	X	$\pi/2$	$\pi/2$	β	$\pi/4$	α	0
6	X	X	X	$\pi/2$	$\pi/4$	A	0	A
7	B	$\pi/2$	X	X	0	0	α	$\pi/4$

Tableau 3.2 : Angles réalisables de α_i .

Les angles réalisables sont : $[0, \arctg(1/2), \arctg(2), \pi/4]$ par des rotations de $\theta + k\pi/2$ avec $k=0,1,2,3$ et $\alpha = \pi/2 - \arctg(2)$ et $\beta = \pi/2 - \arctg(1/2)$. Nous obtenons une fonction des angles normaux périodique, en fonction de la position des pixels dans la chaîne de contour [2]. Pour l'ensemble de nos images, une représentation standardisée des courbes des fonctions d'angles normaux indépendante de la longueur du contour a été décrite. Nous avons ramené par contraction de l'intervalle de définition de ces fonctions dans l'intervalle $[0,1]$ (voir figure 3.15).

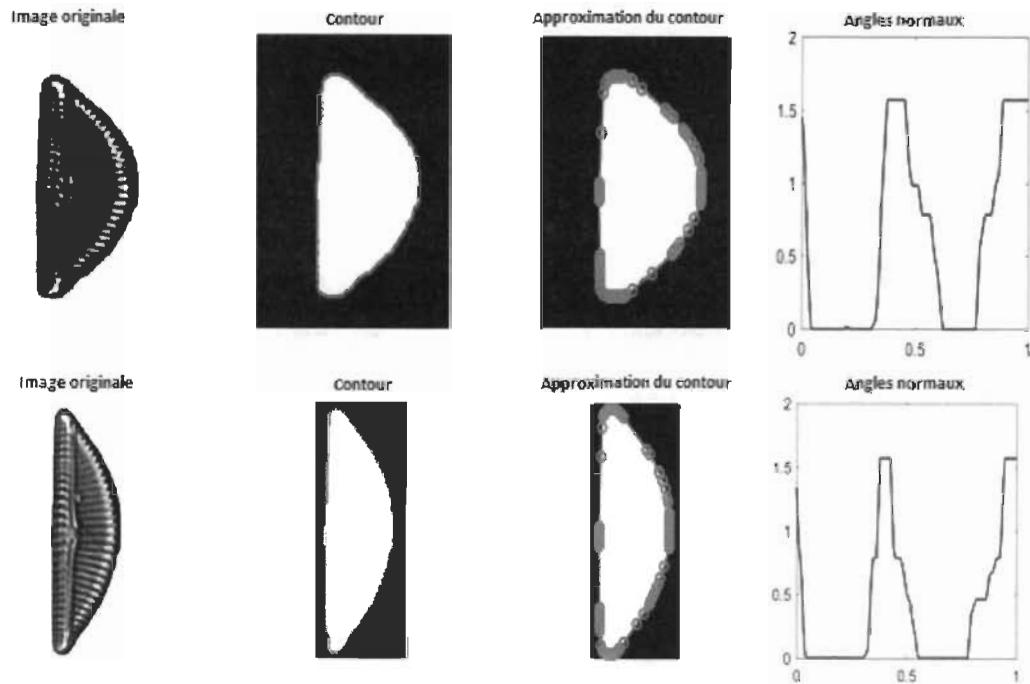


Figure 3.15 : Exemples de deux diatomées avec leurs fonctions d'angles.

Chapitre 4

Extraction des caractéristiques

L'extraction des caractéristiques est une composante du traitement de l'image qui va souvent de pair avec la classification. En effet l'extraction des caractéristiques est une étape principale pour la reconnaissance automatique des formes. Nous retenons des caractéristiques typiquement adaptées à la nature des objets étudiées de manière à ce qu'ensuite les méthodes de classification puissent être robustes et plus simples [3].

Ce chapitre est divisé en trois parties, dans la première partie nous présentons les descripteurs de Fourier grâce à la fonction périodique des angles normaux. Dans la deuxième partie nous présentons quelques paramètres de forme. Ces paramètres sont : la convexité, la circularité, l'indice de rectangularité et la symétrie. Dans la dernière partie nous présentons les valeurs propres du Laplacien de Dirichlet.

4.1 Descripteurs de Fourier

Les descripteurs de Fourier sont à la base d'une méthode classique de reconnaissance de forme et sont devenus une méthode générale pour coder diverses signatures de forme. L'idée générale est que si une courbe fermée (dans notre situation c'est la fonction des angles normaux), nous pouvons la présenter par une fonction périodique par les coefficients de Fourier.

Soit f une fonction continue, périodique de période T , Joseph Fourier, assure qu'il est possible dans certaines conditions de décomposer une fonction périodique f sous la forme d'une somme infinie de signaux sinusoïdaux.

$$f(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t). \quad (4.1)$$

On peut considérer alors f comme la somme d'un terme constant a_0 et d'un nombre infini de termes sinusoïdaux appelées harmoniques.

L'harmonique de rang n est :

$$u_n(t) = a_n \cos n\omega t + b_n \sin n\omega t. \quad (4.2)$$

On peut écrire la première équation avec d'autres variables :

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_k \left(\cos \frac{2\pi kx}{T} \right) + b_k \left(\sin \frac{2\pi kx}{L} \right) \right). \quad (4.3)$$

Avec : $a_k = \frac{2}{T} \int f(x) \cos \left(\frac{2\pi kx}{T} \right) dx \quad t = 1, 2, \dots, L. \quad (4.4)$

$$b_k = \frac{2}{L} \int f(x) \sin \left(\frac{2\pi kx}{L} \right) dx \quad t = 1, 2, \dots, L. \quad (4.5)$$

Soit C le graphe des angles normaux d'une diatomée, et θ_i l'angle normal de ce graphe est présenté par :

$$C(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_k \left(\cos \frac{2\pi kt}{T} \right) + b_k \sin \left(\frac{2\pi kt}{L} \right) \right). \quad (4.6)$$

Avec : $a_k = \frac{2}{T} \int f(x) \cos \left(\frac{2\pi kx}{T} \right) dx. \quad (4.7)$

$$b_k = \frac{2}{L} \int f(x) \sin \left(\frac{2\pi kx}{L} \right) dx. \quad (4.8)$$

Soit $N=449$ images, le nombre de diatomées de notre base de données. Pour $k \in [1, 2, \dots, 15]$ nous avons calculé la série de Fourier approximative du graphe C_i pour chaque valeur de k :

$$f_{i,k}(x) = \frac{a_0}{2} + \sum_{k=1}^{15} \left(a_k \left(\cos \frac{2\pi kx}{T} \right) + b_k \sin \left(\frac{2\pi kx}{L} \right) \right). \quad (4.9)$$

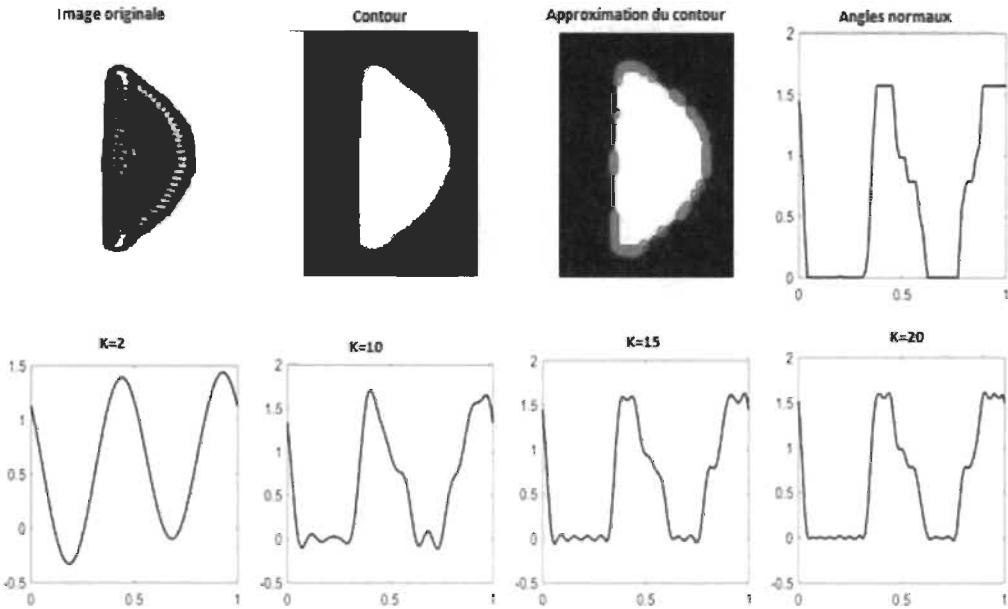


Figure 4.1 Descripteurs de Fourier.

4.2 Les paramètres de formes

- **Convexité**

Une mesure de convexité peut être obtenue en formant le rapport du périmètre de l'objet lui-même $P(D)$ au périmètre de la surface d'un objet coque convexe $P(C)$. Nous pouvons la calculer aussi en formant le rapport de l'aire de l'objet à l'aire d'un objet coque convexe [32]. Cette mesure de convexité est maximale et vaut 1 lorsque l'objet est convexe.

$$C_p(D) = \frac{P(D)}{P(C)}. \quad C_a(D) = \frac{A(D)}{A(C)}. \quad (4.11)$$

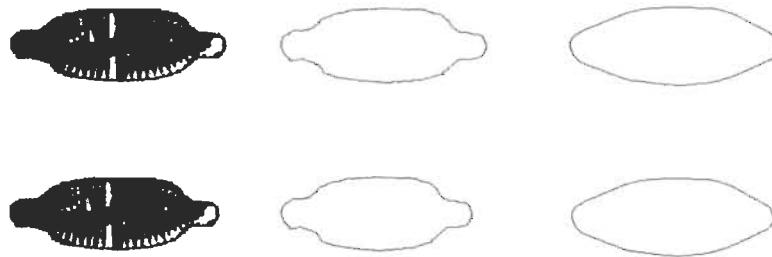


Figure 4.2 : Deux diatomées, leurs contours, leurs enveloppes convexes.

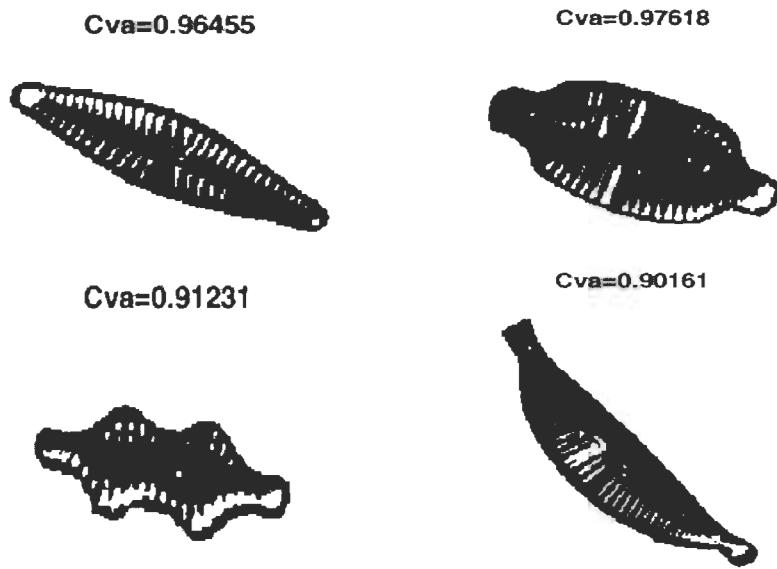


Figure 4.3 Exemple de calcul de la convexité.

• Circularité

Une mesure de circularité peut être obtenue en formant le rapport du périmètre de l'objet sur son aire [32]. Cette statistique est égale à 1 pour un objet circulaire.

$$C(D) = \frac{P^2(D)}{4\pi A(C)}. \quad (4.12)$$





Figure 4.4 Exemple de calcul de la circularité.

- **L'indice de rectangularité et d'allongement**

L'indice de rectangularité peut être obtenu en formant le rapport de l'aire de l'objet $A(D)$ sur la surface du rectangle circonscrit, d'aire minimum, d'un objet après une rotation selon son axe principal $R(D)$.

$$R_c(D) = \frac{A(D)}{R(D)}. \quad (4.13)$$

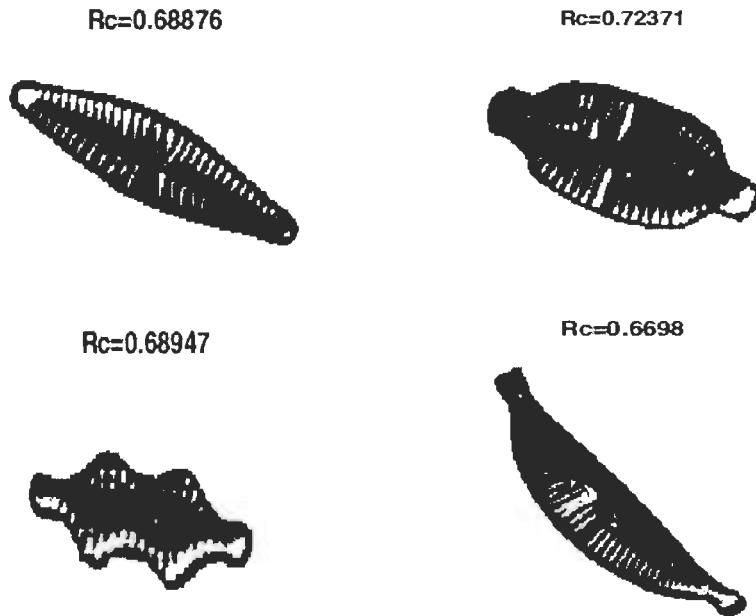


Figure 4.5 Exemple de calcul de l'indice de rectangularité.

On définit l'indice d'allongement d'un objet comme étant;

$$A_l = \frac{L}{l}. \quad (4.14)$$

Où L et l sont respectivement les longueurs du côté le plus long et le plus court du rectangle circonscrit.

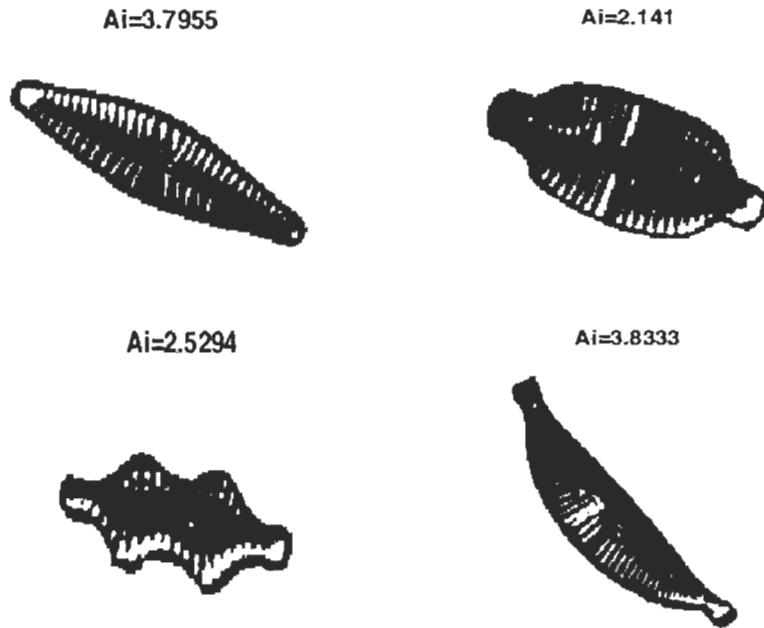


Figure 4.6 Exemple de calcul de l'indice d'allongement.

• Symétrie

Il existe deux paramètres de symétrie d'une diatomée. Un premier axe vertical, passant par le centre de gravité, divise la diatomée en deux parties, gauche et droite, A_l and A_r [3]. Un deuxième axe horizontal, passant par le centre de gravité, partage la diatomée en deux parties (sections supérieure et inférieure), A_s et A_i . Les paramètres S_h et S_V sont donnés par les expressions suivantes :

$$S_h = \frac{|A_s - A_i|}{A_s + A_i}, \quad S_V = \frac{|A_d - A_g|}{A_d + A_g}. \quad (4.15)$$

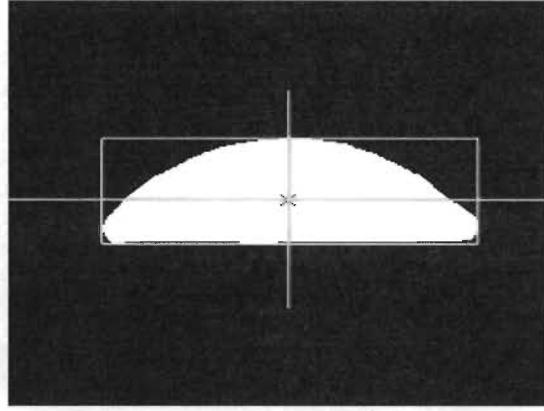


Figure 4.7 Symétrie d'une diatomée.

4.3 Les valeurs propres du Laplacien

En plus des descripteurs de Fourier et des paramètres de formes, nous avons ajouté dans notre étude des caractéristiques basées sur les valeurs propres du Laplacien. Le calcul des valeurs propres du Laplacien en fonction du domaine a été étudié dans de nombreuses recherches en traitement d'images afin de classer des images. Notre approche est celle présentée par Rhouma et al. [33]. Dans les sections suivantes nous décrivons cette approche.

4.3.1 Reconnaissance de forme et valeurs propres du Laplacien de Dirichlet

Rhouma et al. [33] ont utilisé trois ensembles de caractéristiques pour la classification des formes dans des images binaires. Ces caractéristiques sont basées sur les valeurs propres du Laplacien du problème de Dirichlet. Les valeurs propres sont invariantes en translation, en rotation et en taille [33]. Dans leur travail, 250 images dessinées à la main ont été numérisées. Ces images artificielles sont des disques, des triangles, des rectangles et des ellipses, de tailles et d'orientations différentes. Après avoir calculé les valeurs propres du Laplacien du problème de Dirichlet discret défini sur chaque image, trois ensembles de paramètres, notés $[F_1, F_2, F_3]$, sont calculés avec ces valeurs propres. Les valeurs propres sont en lien avec la forme de l'image, en l'occurrence celle d'une diatomée. Des tests ont été réalisés afin de démontrer le pouvoir discriminant de ces ensembles de

caractéristiques et aussi la tolérance en fonction du bruit sur le contour, cela en supprimant des pixels dans les images construites. Les caractéristiques des images bruitées sont comparées avec les caractéristiques des images originales non bruitées. Les caractéristiques des images bruitées se sont avérées proches de celles des images originales. La classification a été réalisée à l'aide d'un réseau de neurones. Les réseaux de neurones ont été entraînés à l'aide de la moitié des images (250), l'autre moitié sert aux tests de classification. Un taux de bonne classification de 97,5% est obtenu avec les paramètres de la classe F_1 , de 88% avec F_2 et de 94.5% avec F_3 . Ces trois ensembles sont définis comme suit :

$$F_1 = \left(\frac{\lambda_1}{\lambda_2}, \frac{\lambda_1}{\lambda_3}, \frac{\lambda_1}{\lambda_4}, \dots, \frac{\lambda_1}{\lambda_n} \right), \quad (4.16)$$

$$F_2 = \left(\frac{\lambda_1}{\lambda_2}, \frac{\lambda_2}{\lambda_3}, \frac{\lambda_3}{\lambda_4}, \dots, \frac{\lambda_{n-1}}{\lambda_n} \right), \quad (4.17)$$

$$F_3 = \left(\frac{\lambda_1}{\lambda_2} - \frac{d_1}{d_2}, \frac{\lambda_1}{\lambda_3} - \frac{d_1}{d_3}, \frac{\lambda_1}{\lambda_4} - \frac{d_1}{d_4}, \dots, \frac{\lambda_1}{\lambda_n} - \frac{d_1}{d_n} \right). \quad (4.18)$$

Ici, n est la taille du problème discret de Dirichlet et $\lambda_1, \dots, \lambda_n$ sont les valeurs propres associées. Ce nombre n correspond au nombre d'inconnues discrètes dans le système d'approximation. Dans notre cas spécifique, n sera le nombre total de pixels intérieurs à la diatomée sans son contour car la solution est égale à zéro sur les pixels du contour. Les valeurs $0 < d_1 < d_2 \leq d_3 \dots \leq d_k \leq \dots \rightarrow \infty$ sont les valeurs propres calculées sur un disque discret. Les valeurs de F_1 et de F_2 sont entre 0 et 1 et celles de F_3 varient entre -1 et 1 [33].

4.3.2 Calcul des valeurs propres du problème de Dirichlet discret

L'opérateur Laplacien intervient dans de nombreux domaines comme l'électromagnétisme, l'électricité mais aussi en vision par ordinateur [45]. Étant donné un domaine borné Ω du plan, on cherche le problème de Dirichlet s'énonce comme suit :

$$\begin{cases} \Delta u + \lambda u = 0, \text{ dans } \Omega \\ u = 0, \quad \text{sur } \partial\Omega \end{cases} \quad (4.19)$$

où $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. Dans notre cas, Ω est le domaine (plan réel en théorie et dans le plan discret lors de la discrétisation) de la diatomée, $\partial\Omega$ est le contour de la diatomée.

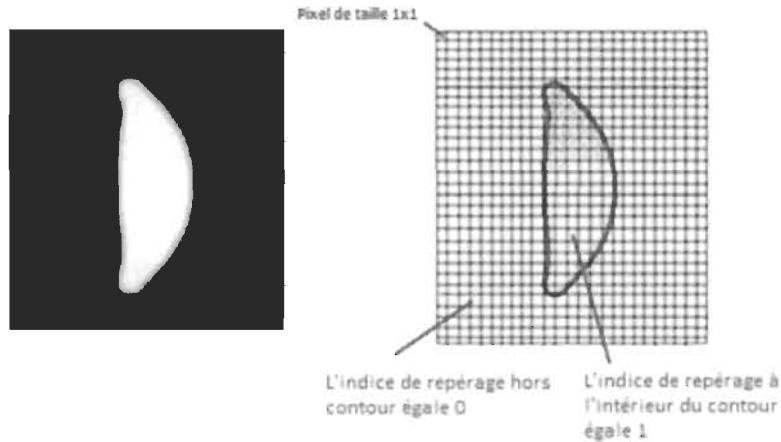


Figure 4.8 Diatomée plongée dans le plan discret : repérage avec 0 pour les pixels hors de la diatomée et 1 à l'intérieur incluant le contour.

Les valeurs propres recherchées sont celles du problème discret du Laplacien (équation 4.19). Le problème approché est défini dans le plan discret avec une maille naturelle définie par des pixels unitaires. De façon générale, on définit un pas h de discrétisation selon l'axe des x et un pas k selon l'axe des y . Ici, on posera $h=k=1$ (voir figure 4.8).

Dans le processus d'approximation de la solution au problème (4.19) nous discrétisons l'équation aux dérivées partielles à l'aide d'un schéma aux différences finis. Cet opérateur différentiel sera appliqué à chaque nœud intérieur (excluant le contour) de la diatomée dans le plan discret (Figure 4.8 à gauche) et en tenant compte du fait que la solution approchée u_h (l'approximation de $u(x, y)$) sera égale à zéro pour les pixels du contour. Pour cela, nous considérons les nœuds de la partition (maille) centrés sur les pixels (nœud $N_{ij} = (i, j)$) et nous notons u_{ij} la solution approchée de $u(x, y)$, solution du problème continu, en chacun de ces nœuds.

Au préalable, nous identifions les nœuds intérieurs et ceux de la frontière (voir Figure 4.8). Les indices (i, j) des valeurs de la solution approchée aux nœuds sont ceux correspondant aux coordonnées discrètes des pixels de la diatomée.

Dans l'équation 4.19 chacune des dérivées secondes du Laplacien (selon x avec un pas h et selon y avec un pas k) est approchée par les schémas aux différences suivants :

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x+h, y) - 2f(x, y) + f(x-h, y)}{h^2}, \quad (4.20)$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{f(x, y+k) - 2f(x, y) + f(x, y-k)}{k^2}. \quad (4.21)$$

Le Laplacien en un point (x, y) est approché, et si $h=k$, par l'expression suivante

$$\Delta f \simeq \frac{f(x+h, y) + f(x-h, y) + f(x, y+h) + f(x, y-h) - 4f(x, y)}{h^2}. \quad (4.22)$$

Si nous remplaçons dans l'équation (4.19) l'expression du Laplacien par l'équation (4.22) en chacun des nœuds intérieurs de la diatomée, avec des pas unitaires ($h=k=1$) et cela pour des ensembles de valeurs de i et j, que nous noterons I et J, correspondant aux indices de parcours de la diatomée, le

problème approché consiste à trouver les valeurs $u_{i,j}$ telles que: pour $i \in I, j \in J$,

$$4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} = \lambda u_{i,j} . \quad (4.23)$$

Certaines des valeurs $u_{l,k}$ dans l'équation (4.23) sont nulles si les indices (positions dans l'image) correspondent à des pixels du contour, en concordance avec les conditions limites du problème aux dérivées partielles (4.22).

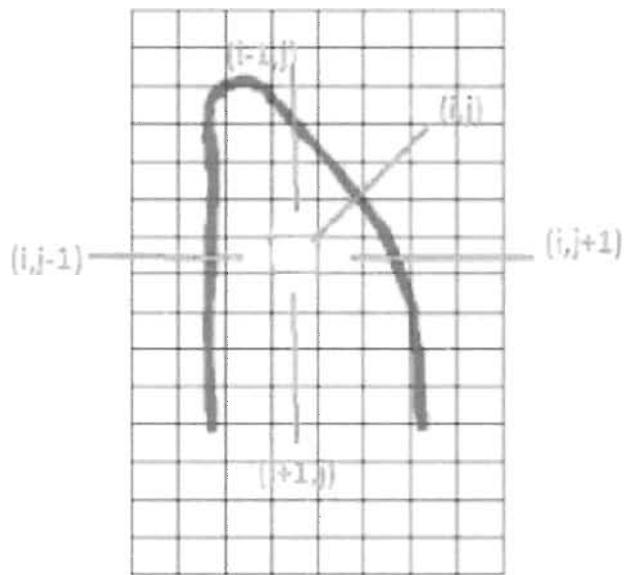


Figure 4.9 Un pixel (i,j) et les indices voisins dans (4.23).

Afin de trouver la solution approchée, le vecteur des valeurs u_{ij} (notons $u = (u_{ij})$) approchant la solution $u(x,y)$ aux nœuds (pixels) intérieurs de la maille, ici la diatomée, on construit le système linéaire dont le vecteur u est solution. Pour cela, on assemble toutes les équations (4.23) en parcourant tous les couples d'indices (i,j) correspondant aux nœuds intérieurs de la diatomée. En résumé, une équation par nœud intérieur. Le système linéaire conséquent aura comme taille, celui du nombre de pixels de la diatomée, sauf ceux sur les bords car nous savons que les valeurs prises sont égales à zéro. Après assemblage des équations discrètes (4.23), le vecteur $u = (u_{ij})$ de la solution approchée aux nœuds intérieurs, vérifie l'équation matricielle suivante,

$$A u = \lambda u . \quad (4.24)$$

Une solution triviale à ce système linéaire est $u=0$ partout. Or, nous cherchons une solution non-triviale à ce problème et ce sera le cas quand λ est une valeur propre du système et donc u , son vecteur propre associé. Nous devons donc calculer les n valeurs propres de ce système. Idéalement, si le domaine de définition du problème aux dérivées partielles est un carré avec une maille unitaire ayant n pixels intérieurs et un contour extérieur, la matrice A a la forme suivante;

$$A = \begin{bmatrix} 4 & -1 & 0 & \cdots & 0 \\ -1 & 4 & -1 & \cdots & 0 \\ 0 & -1 & 4 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 4 \end{bmatrix}_{n \times n} .$$

La matrice A d'ordre n est définie positive. Dans notre cas, la matrice A est carrée d'ordre n mais sa structure interne (entrées de la matrice) dépendra de forme de la diatomée.

Illustrons l'assemblage de la matrice A , à partir des équations (4.23), à l'aide d'un exemple. Considérons la diatomée (voir figure 4.10), où les inconnues coïncident avec les indices des pixels. Cette diatomée possède 10 pixels intérieurs, donc le système aura $n=10$ inconnues. Les pixels du contour, avec $u_{ij}=0$, ne sont pas montrés.

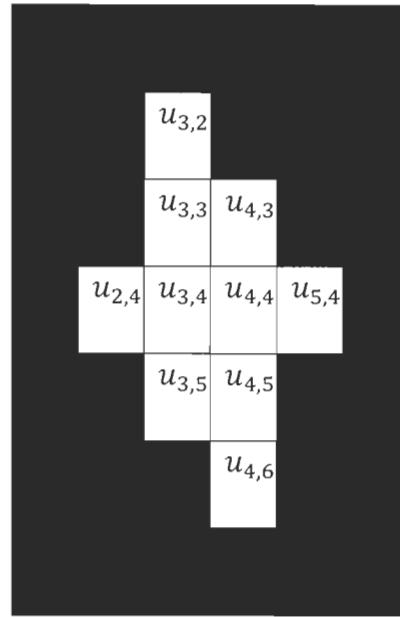


Figure 4.10 Répartition des inconnues.

Le tableau 4.1 illustre, par des 1, la position des inconnues dans l'image. Si un indice est égal à 0 alors $u = 0$ a un pixel du contour.

$\rightarrow i$	1	2	3	4	5	6
1			0			
2		0	1	0		
3		0	1	1	0	
4	0	1	1	1	1	0
5		0	1	1	0	
6			0	1	0	
7				0		

Tableau 4.1 Répartition des indices des pixels de la diatomée et de son contour.

La fonction *find* de Matlab renvoie les indices de chaque pixel de valeur 1. Ainsi, cette fonction renvoie les indices de rangée et de colonne de pixels à l'intérieur du contour sous

forme d'un tableau (voir tableau 4.2). L'entier M est un compteur qui parcourt les positions d'indices 1, $M = \{1, 2, \dots, 10\}$, et ainsi repère le nombre d'équations et d'inconnues ($n=10$) du système. La fonction *axis_data* de Matlab nous permet de graduer nos axes (voir tableau 4.3).

M	Ligne	Colonne
1	2	3
2	3	3
3	3	4
4	4	2
5	4	3
6	4	4
7	4	5
8	5	3
9	5	4
10	6	4

Tableau 4.2 Positions et numérotation des inconnues à l'intérieur du contour.

	1	2	3	4	5	6
1			0			
2		0	1	0		
3		0	2	3	0	
4	0	4	5	6	7	0
5		0	8	9	0	
6			0	10	0	
7				0		

Tableau 4.3 Répartition des éléments à l'intérieur du contour.

L'assemblage de la matrice A se fait en parcourant les pixels de chaque ligne en commençant par la ligne du haut et en insérant les coefficients des équations 4.23 dans A,

et dans les colonnes correspondantes aux numéros des inconnues des pixels voisins (voir tableau 4.3). Dans cette approche de recherche des valeurs propres afin de constituer les quotients des ensembles F_i , il n'est pas nécessaire de calculer les solutions (vecteurs propres) du système 4.5. Dans cet exemple, les entrées de la matrice A sont illustrées dans le tableau 4.4.

	1	2	3	4	5	6	7	8	9	10
1	4	-1	0	0	0	0	0	0	0	0
2	-1	4	-1	0	-1	0	0	0	0	0
3	0	-1	4	0	0	-1	0	0	0	0
4	0	0	0	4	-1	0	0	0	0	0
5	0	-1	0	-1	4	-1	0	-1	0	0
6	0	0	-1	0	-1	4	-1	0	-1	0
7	0	0	0	0	0	-1	4	0	0	0
8	0	0	0	0	-1	0	0	4	-1	0
9	0	0	0	0	0	-1	0	-1	4	-1
10	0	0	0	0	0	0	0	0	-1	4

Tableau 4.4 Entrées de la matrice A.

Pour trouver les valeurs propres, nous devons résoudre l'équation polynomiale de degré M,

$$\det(A - \lambda I) = 0 . \quad (4.25)$$

Pour toutes nos diatomées, nous avons utilisé la fonction *eig* de Matlab. Dans notre exemple ci-dessus, nous obtenons les valeurs propres suivantes (voir tableau 4.5) :

No	λ
1	6.6996
2	5.4605
3	5.0000
4	4.7609
5	4.0000
6	4.0000
7	3.2391
8	3.0000
9	2.5395
10	1.3004

Tableau 4.5 Valeurs propres de A.

Le nombre de valeurs propres retenues dans [33] est n=13 et le taux de classification dépasse 85%. Nous avons utilisé cette valeur (13) dans nos tests de classification et nous avons obtenu le même résultat. Dans nos tests, nous commençons par garder 10 valeurs propres, et nous faisons une classification bayésienne, ensuite nous gardons 11 et nous refaisons la classification bayésienne juste pour les valeurs de Laplacien et ainsi de suite. Lorsqu' on prend 13 valeurs pour la plupart des diatomées le taux de la classification ne change plus. C'est pour cette raison que nous retenons 13 valeurs propres.

Chapitre 5

Résultats

Après l'extraction des caractéristiques des images de diatomées nous procédons à l'étape de la classification. Afin de réduire le nombre de paramètres traités lors de la classification (approche bayésienne), nous effectuons une analyse en composantes principales (ACP) sur chacun des regroupements de variables initiales (voir plus loin). Notre objectif est de tester la performance de chacun des regroupements de variables et de comparer le pouvoir discriminant de ceux-ci. Nous avons partagé en deux sous-ensembles les diatomées de notre base de données (449 images), réparties dans plusieurs classes a priori. Les spécimens de diatomées ont été regroupés par le professeur Stéphane Campeau de l'Université du Québec à Trois-Rivières [22]. Un premier sous-ensemble, noté E_1 , sert au calcul des probabilités d'appartenance de chacune des diatomées à sa classe. Le complémentaire E_2 servira à la classification proprement dite. L'ensemble E_1 est choisi arbitrairement mais à la condition de conserver dans E_1 et E_2 un nombre suffisant d'images de chaque classe.

Nous avons procédé à la classification de 100 échantillons de 41 diatomées tirées au hasard dans l'ensemble E_2 . Le tableau 5.1 montre le nombre de diatomées dans chaque classe.

A	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
11	6	6	4	6	5	3	5	3	7	4
AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
8	7	6	5	8	15	5	6	6	4	5
AV	AW	AX	AY	AZ	B	BA	BB	BC	C	Di
6	7	6	8	6	16	13	7	5	10	13
E	F	G	H	I	J	K	L	M	N	O
13	17	33	17	16	12	13	14	5	5	5
P	Q	R	S	T	U	V	W	X	Y	Z
20	5	5	5	5	4	5	5	5	5	5

Tableau 5.1 Nombres d'images par classe.

5.1 Analyse en composantes principales

L'analyse en composantes principales (ACP) est un outil statistique qui permet une synthèse de l'information en présence d'un nombre important de variables quantitatives à traiter et à interpréter.

Cette analyse statistique dans un espace multidimensionnel permet la représentation d'échantillons selon un nombre d'axes principaux, combinaisons linéaires des variables initiales, selon la variance totale [3] des axes retenus, de la plus grande à la plus petite, et indépendants les uns des autres [35].

Nous avons effectué un ACP sur les caractéristiques des diatomées de l'ensemble E₁ en ne retenant que les caractéristiques des regroupements testés. Le plus grand vecteur testé est $\vec{x}_i = (x_{1i}, x_{2i}, \dots, x_{35i})$; les 15 premières sont les descripteurs de Fourier du contour, les 7 suivantes sont les paramètres de forme calculés antérieurement à savoir, la circularité Cr, les deux indices de convexité C_{va}, C_{vp}, l'indice d'allongement A_l, l'indice de rectangularité R et les deux indices de Symétrie S_h,S_v, les 13 dernières composantes sont les paramètres retenus (voir chapitre 4) et issus du calcul des valeurs propres du problème

discret du Laplacien. Nous appellerons regroupement singleton, pris séparément, les descripteurs de Fourier, les paramètres de forme et finalement les quotients de valeurs propres. Nous avons testé ces singltons séparément pour la classification, puis les regroupements de deux singltons et l'ensemble des trois.

Pour chacun des singltons et des regroupements de caractéristiques, nous avons conservé les trois premiers axes principaux de l'ACP, qui expliquent dans tous les cas plus de 90% de la variance. Le tableau 5.2 illustre les variances pour l'ensemble des caractéristiques.

Ensemble des caractéristiques	Axe ₁	Axe ₂	Axe ₃	Axe ₄
Variance	0.447	0.356	0.11	0.054
Variance cumulée	0.447	0.801	0.912	0.966

Tableau 5.2 Variances et variances cumulées pour l'ensemble des caractéristiques.

Le tableau 5.3 montre les variances pour le regroupement des descripteurs de Fourier (DF) et les paramètres de forme (PF).

DF+PF	Axe ₁	Axe ₂	Axe ₃	Axe ₄
Variance	0.432	0.368	0.122	0.067
Variance cumulée	0.432	0.80	0.922	0.989

Tableau 5.3 Variances et variances cumulées pour les TF+PF.

L'étape suivante, avant la classification, est d'établir les probabilités d'appartenance à une classe pour les diatomées de l'ensemble E₁. Ces probabilités sont calculées à l'aide des

variables sur les axes principaux résultants de l'ACP dans chacun des cas (singletons, paires et le triplet de variables).

La classification bayésienne est basée sur le théorème de Bayes. Les classificateurs bayésiens sont des classificateurs statistiques. Les classificateurs bayésiens peuvent prédire les probabilités d'appartenance à une classe, dans notre cas, la probabilité qu'une diatomée appartienne à une classe particulière et d'indiquer la classe la plus probable [34]. En utilisant la formule de Bayes, la probabilité qu'une diatomée représentée par un vecteur $\vec{x}_i = (x_{1i}, x_{2i}, \dots, x_{35i})$ appartienne à une classe $C_i \in \{C_1, C_2, \dots, C_n\}$ est donnée par l'expression suivante ;

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{\sum_{j=1}^n P(X|C_j)P(C_j)}. \quad (5.1)$$

où $P(C_i|X)$ est la probabilité a posteriori que la diatomée représentée par le vecteur X appartienne à la classe C_i . $P(C_i)$ est la probabilité a priori que la diatomée appartienne à C_i . La décision finale de la classe à laquelle une diatomée appartient se fait en comparant les probabilités a posteriori d'appartenance à toutes les classes et bien sûr nous sélectionnons la plus grande.

5.2 Résultats expérimentaux

Nous avons procédé à la classification de 100 échantillons de 41 diatomées tirées au hasard dans l'ensemble de E_2 avec remise. Le calcul des probabilités et la classification bayésienne dans chaque essai sont effectués avec les valeurs sur les axes principaux. La moyenne et l'écart-type de la classification pour chaque combinaison sont représentée dans le tableau 5.4.

	Moyenne	Écart-type
Descripteurs de Fourier	3.93	45%

	Moyenne	Écart-type
Paramètres de forme	3.27	73.03%

	La moyenne	Écart-type
Paramètres du Laplacien	2.65	84.30%

	Moyenne	Écart-type
Descripteurs de Fourier + paramètres de forme	2.06	91.54%

	Moyenne	Écart-type
Descripteurs de Fourier+ paramètres du Laplacien	1.99	92.11%

	Moyenne	Écart-type
Paramètres de forme + paramètres du Laplacien	1.30	96.45%

	Moyenne	Écart-type
Tous les paramètres	1.11	97.84%

Tableau 5.4 Résultats de la classification pour chaque cas.

Suite à l'analyse des résultats pour l'ensemble des 449 images de diatomées, nous observons que les résultats des 7 tests donnent un bon taux de classification d'approximativement 98% pour les trois singltons, et un taux de bonne classification de

96% pour les deux singletons (Paramètres de forme + paramètres de Laplacien). Le taux de bonne classification pour les paramètres de Fourier seuls ne dépasse pas 50%.

Chapitre 6

Conclusion et perspectives

Pour solutionner le problème de la classification des images de diatomées, nous avons présenté une solution rassemblant des techniques connues et robustes de traitement d'images basé sur le contour et la géométrie. Ces techniques forment un système de classification automatique intéressant pour l'échantillon des images de diatomées utilisées dans cette recherche. Nous proposons dans notre dernier chapitre un rappel des méthodes appliquées pour notre classification et les perspectives d'évolution de notre recherche.

6.1 Les Méthodes appliquées

Dans notre travail, nous avons proposé une méthode de classification automatique des diatomées basée sur le contour et la géométrie. Différentes étapes ont été essentielles, allant du prétraitement des images jusqu'à l'extraction et l'analyse des motifs.

Pour l'aspect programmation de notre approche, nous avons utilisé un logiciel de calcul scientifique permettant de développer des solutions à des problèmes techniques appelé Matlab. Matlab est associé à des boîtes à outils appelées *toolbox* permettant d'accéder à des fonctions spécifiques à un domaine d'application comme le traitement d'image.

Le prétraitement des images est notre première étape, le but de cette étape est de réduire le bruit présent sur les images de diatomées. À la fin de prétraitement, nous obtenons des images non bruitées, ces images sont transmises à l'étape de détection du contour 8-connecté et les angles normaux au contour.

Ensuite, nous avons procédé à l'étape d'extraction des caractéristiques. Dans un premier temps, nous avons proposé trois familles de descripteurs : les descripteurs de Fourier puis les paramètres de formes et les paramètres du Laplacien. Nous avons pu constater que les trois descripteurs sont bien adaptés pour identifier les diatomées.

L'étape finale est la classification des images de diatomées. Pour classer les images, nous avons utilisé la classification bayésienne.

Après plusieurs tests sur les images de différentes classes, nous avons conclu que l'utilisation des trois descripteurs fournit de très bons résultats de l'ordre de 98%.

6.2 Les perspectives d'amélioration

Nous avons seulement 449 images de diatomées, ce nombre est peu élevé. Cette base de données devrait être élargie.

Dans l'étape d'extraction, il serait avantageux de trouver une méthode automatique pour isoler les diatomées, au lieu de le faire manuellement.

D'autres méthodes de classification pourraient être étudiées. Notre méthode pour la classification est simple, d'autres méthodes comme les réseaux de neurones pourraient être plus efficaces.

Annexe 1



Classe A



Classe AA



Classe AB



Classe AC



Classe AD



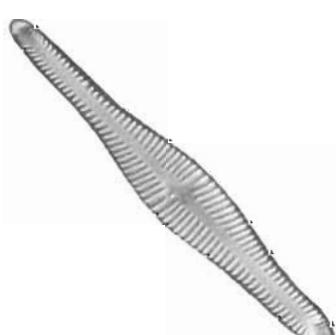
Classe AE



Classe AF



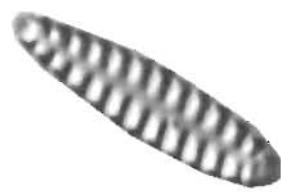
Classe AG



Classe AH



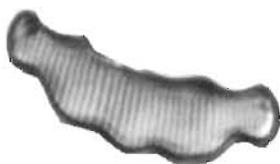
Classe AI



Classe AJ



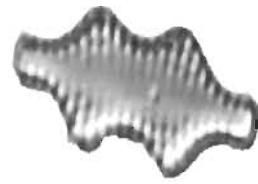
Classe Ak



Classe AL



Classe AM



Classe AN

Annexe 2

No.	Fourier descriptor	Paramètres de formes	Laplacian descriptor	Fourier Et formes	Fourier Et Laplacian	Formes Et Laplacian	Fourier, Formes Et Laplacian
Test 1	48.95495856	76.48842359	84.09586775	88.64727747	89.82567995	97.22007159	95.8482669
Test 2	45.1236677	70.96037879	79.67647814	89.97162689	90.3962418	95.63106086	95.94641037
Test 3	39.51795944	74.85278668	84.27314996	93.93967606	92.69484354	97.49902045	97.57496421
Test 4	47.32762203	70.51448095	82.75537514	90.17814876	91.15736567	96.76344204	98.05494303
Test 5	42.26486629	75.8931244	84.99837681	90.972347	90.89497348	98.60767494	97.83662097
Test 6	52.86602295	66.26907948	80.08460142	88.32043193	91.11167855	93.73750945	96.17550347
Test 7	40.9777302	76.50571906	87.93982139	91.49029991	90.34977114	96.56977507	99.28892203
Test 8	51.31457621	74.12505171	86.07017728	92.05152452	88.7544285	95.52751042	96.84033999
Test 9	43.62501905	73.63480853	86.91125675	88.74693011	93.53566612	96.28381158	97.71380969
Test 10	42.80557243	70.04223521	82.98154212	85.02691635	93.94786914	94.15804726	96.15932683
Test 11	45.42748197	68.14746545	79.35417953	89.86354373	91.42798581	94.46333986	97.81657409
Test 12	49.97803497	73.6759657	85.70308073	89.37479278	91.78107104	96.42006834	96.49079384
Test 13	48.21185155	69.05985601	82.90128424	88.90816104	91.4069127	96.66226869	97.77064388
Test 14	42.44705085	69.46283447	80.23948774	89.5486348	92.94665101	94.64214547	97.19709423
Test 15	46.32532335	73.86612547	84.1415292	90.8958948	92.2182136	97.92914954	96.07174584
Test 16	45.86087212	70.03588238	82.06084732	89.4251302	90.73334382	94.62854479	97.12496839
Test 17	47.76440273	74.04014911	82.90978993	90.42955503	94.05706126	95.98258978	99.85087976
Test 18	42.75179342	72.12659578	87.75350804	92.16989879	93.26539518	96.38974261	98.39853844
Test 19	45.53197722	74.90246028	86.32077347	92.53911736	92.33048153	95.67445698	96.95089797
Test 20	44.43630483	79.95799993	84.0436438	93.21517011	90.45054521	95.29499352	97.3026316
Test 21	39.410842	68.47960583	84.39909502	91.13578325	92.29745695	96.67517348	98.83246885
Test 22	45.52536277	73.32141557	83.20523269	87.33567077	94.96494829	94.37540898	98.98289931
Test 23	38.48134676	75.53492929	86.22270869	94.21587961	92.00892934	97.22814805	98.30155109
Test 24	43.34964411	72.06704823	84.47819822	89.00910976	89.5508905	95.28231564	97.86573327
Test 25	47.16745888	70.35513735	88.7638032	92.05017881	96.31551451	95.78545343	98.24821072
Test 26	45.73618569	79.00411083	86.49661126	95.23872602	95.3552757	97.41691456	99.8862524
Test 27	46.28342796	73.07554572	85.9879591	90.82322027	94.11240744	94.9902221	98.83958087
Test 28	43.40291896	74.12736247	85.04517539	87.3853315	94.45857037	96.80836776	97.29364037
Test 29	46.26063078	76.8782663	87.31252755	92.89747747	93.22709735	98.99183197	99.25363749
Test 30	34.20768222	74.28914015	80.30138237	92.00453758	90.08401329	95.5699672	98.39235612
Test 31	42.36985289	74.53442003	82.29628741	93.36518602	93.4483195	97.30670379	98.97571407
Test 32	50.63512641	73.14820541	83.42933914	93.19596507	89.98404821	96.86244648	99.2467409
Test 33	43.52648272	74.93329329	81.99649633	93.85809257	89.12669508	97.30284182	98.87239158
Test 34	45.21456813	76.48002328	91.86667539	93.1929242	92.17126518	98.84828017	98.04853509
Test 35	44.6193571	69.24922869	85.06235599	90.5303232	92.757905	96.78734584	97.55043684
Test 36	46.29360449	67.87784211	86.5327678	91.80802418	94.90560467	96.35853749	99.10533298
Test 37	37.07938303	73.99095761	81.64615216	94.40799104	89.14026528	95.51476081	99.12340057
Test 38	50.21662891	71.13339718	84.34621438	90.89827063	93.62870044	97.13568091	98.39860285
Test 39	39.32901749	69.36493631	81.90163324	92.05375882	90.31904649	95.01484807	99.13490087
Test 40	39.64875318	73.50988768	83.24573043	93.26173251	89.92747388	95.56026499	98.49962395
Test 41	47.34164769	73.89158545	86.67895089	90.87581463	93.29063471	96.68194734	96.93846973
Test 42	47.422666807	71.09401612	84.13501183	90.48090197	90.72929295	95.81556229	96.19952482
Test 43	41.10371043	73.3043798	83.70553252	93.14866732	93.8020664	97.55397279	97.36088596
Test 44	46.51809536	65.67435525	79.843195	89.20845275	92.86113551	97.55523415	97.73429572
Test 45	47.68344294	73.58437994	86.04449257	91.11336965	92.78064618	96.55593478	97.86705145
Test 46	42.63263515	71.60460873	89.49271637	90.59795834	87.6851513	96.93311717	96.34745539
Test 47	51.89377738	70.25719574	80.03499415	91.08769852	94.55923033	97.55459182	97.40140779
Test 48	41.40904757	73.35585562	79.96336666	91.62453394	91.05409943	94.62405928	98.28650624
Test 49	49.18176573	74.5003479	86.0886187	92.78566225	93.2089631	96.14302606	97.84480155
Test 50	49.03893785	66.39576852	81.40229896	88.40229913	89.31912869	96.76922167	97.89174479
Test 51	40.18301097	65.2360993	82.11090332	90.98951156	91.69334762	94.164938	96.45379843
Test 52	39.16460768	74.9242026	84.50349686	92.68486149	90.16475906	96.20233913	97.24691544
Test 53	43.07013765	69.15299981	82.05298735	89.85358046	91.66878019	96.4276796	98.15214618
Test 54	41.43425977	79.88102945	85.70481604	94.50819489	95.26931321	98.98615229	98.28847309
Test 55	43.21692421	65.90187768	85.37416637	89.92924627	91.68829441	95.48692659	98.97922001
Test 56	41.36367263	74.23789037	86.83603129	93.38897443	94.12657343	97.5452708	98.64796189
Test 57	38.27382148	72.07279935	83.75611654	88.7725884	92.06814266	94.7453801	97.77850652

Test 58	43.41126566	73.97348479	88.33666736	90.10473514	90.8915175	96.13690998	97.86913554
Test 59	35.82011709	78.06101512	87.33358847	92.35182801	93.77344277	97.70710819	98.81562453
Test 60	48.96363349	71.58724489	87.05051027	93.35520799	92.07176001	97.51957428	98.10063612
Test 61	39.66614362	70.36902038	85.85418297	91.6809188	92.43168376	96.46676702	97.10819107
Test 62	45.5893343	77.8675708	86.00870736	91.71605086	93.50085938	97.27783918	97.70531332
Test 63	44.94962387	79.93735063	85.42989283	94.54569331	95.39640444	98.50727447	97.8996099
Test 64	44.90214336	73.26422913	84.34012349	90.87429731	92.45145554	97.21960502	96.64545665
Test 65	41.83623621	75.51876536	85.11733255	93.03176325	94.62238169	98.45876142	97.47461575
Test 66	40.82255258	74.08641224	83.9215166	93.38992743	89.7109726	96.43853654	97.25159043
Test 67	35.94557625	80.13744176	88.01217382	94.41093722	94.84013287	96.81815427	98.12771919
Test 68	44.21731599	74.86471796	85.2478825	95.72851468	91.61844285	96.75459041	99.22019094
Test 69	45.60589089	80.88597693	85.67204964	92.03667313	93.74021631	98.80063936	99.89318338
Test 70	45.53610979	72.36172441	84.26257086	94.10519025	88.45302794	95.77977428	98.59415513
Test 71	48.58236037	72.2244281	87.654043	91.62530713	95.15538183	97.3399801	99.34011437
Test 72	47.89573087	69.02684558	82.37690461	90.36196619	94.5821423	94.27991001	98.84948521
Test 73	42.00613402	72.79549836	79.99100595	94.13726387	91.42166145	95.37352535	97.3443549
Test 74	44.54538558	75.89943087	83.72539089	92.88481769	90.890105	97.73536553	97.83477218
Test 75	47.46100591	71.13774116	84.71494023	89.02018326	89.03214892	95.75180242	94.33703778
Test 76	40.55723003	74.26675601	85.81689528	93.6490081	93.47559819	95.5293833	97.74339687
Test 77	47.06708006	73.28400267	83.20796313	94.7138936	89.09825007	96.45966832	99.28040242
Test 78	49.3278198	69.96825521	83.08196923	88.51517244	93.04469399	94.70876963	98.81892664
Test 79	39.99469505	70.71137053	77.31997806	91.07495228	92.00860531	95.03146723	96.95967712
Test 80	41.75250499	73.42536542	87.62976427	90.68400245	91.41859289	97.7608283	97.51670041
Test 81	46.11607454	72.41352384	80.43437262	89.67389002	91.51901216	96.84453669	96.49840582
Test 82	39.89312184	73.88643932	87.59101744	92.78395067	94.94919374	97.018547	97.80067357
Test 83	40.51065544	73.70646762	84.54908298	92.86969663	92.68503775	97.93078656	97.05501399
Test 84	42.00231815	74.41240347	84.23959916	89.34590001	89.28945883	97.09402271	95.62496318
Test 85	40.68564175	74.73117513	86.77139785	92.9386053	95.77541603	97.99572512	98.87846547
Test 86	42.15285986	72.2837107	81.66829314	89.03265385	86.75989738	95.66945473	97.44270698
Test 87	41.77514384	73.026218	83.75413938	93.86997616	90.24538625	97.75614324	97.31257865
Test 88	39.25867458	77.4093438	88.85004903	94.00111584	92.39920142	99.61225071	99.93808608
Test 89	44.34613022	70.89867322	83.43846631	89.95555339	92.30492065	95.05715941	97.52737461
Test 90	38.78768855	73.70786358	87.32451019	90.86775814	93.3399675	97.93618876	97.76830767
Test 91	53.20713713	78.03888675	86.42926958	93.3963927	95.14436175	96.7383656	96.59269449
Test 92	49.34270165	68.7353752	81.73542668	91.32920009	93.92405227	94.34675212	98.96021508
Test 93	47.74873352	69.63775797	78.60586884	89.95867395	90.73888539	94.95379186	96.74231545
Test 94	38.76782426	70.89161522	83.32136326	89.59623675	90.55896288	95.55755917	94.63823903
Test 95	40.55335451	76.25443383	85.60397033	95.24520795	93.4428521	98.98429577	99.19555464
Test 96	46.35754162	74.86734816	84.47227102	93.66151258	92.59099167	94.13933421	97.63808797
Test 97	48.24117426	73.70647484	84.2030439	92.01911066	92.10227433	95.10762648	98.55905477
Test 98	46.17490023	74.57544557	83.78997752	93.53014468	91.04783724	96.69351283	99.43629312
Test 99	45.43115546	72.00455236	79.82518455	93.33691624	93.45513885	95.45690333	98.48648613
Test 100	42.05492306	71.30685373	84.57013872	89.12153218	88.98419711	97.12577313	96.93494352
Écart-Type	3.933183625	3.278561526	2.652095723	2.060049505	1.990767544	1.305030679	1.114330214
Moyenne	44.13071169	73.03134779	84.30762308	91.54368138	92.11923201	96.45521055	97.84445801

Annexe 3 Implémentation

```
%% Loading of images
fname = mat2str(cell2mat(imgSet.ImageLocation(i)));
[path, FileName, ext] = fileparts(fname(2:end-1));
set(handles.FileName, 'string', [FileName ext]);
img = imread(fname(2:end-1));
%img = read(handles.imgSet, i);
set(gcf,'CurrentAxes',handles.axes1);
[r,c,z] = size(img); % get size of image i.e. r=height, c=width, z=3 for RGB and z=1 for Gray
if z == 3
    img = rgb2gray(img); % if image is rgb then convert it to grayscale
end
imshow(img,[]);
title('Original Image');

%% Binarize image
for k=1:r
    for j=1:c
        if img(k,j)<255
            img(k,j) = 0;
        end
    end
end
set(gcf,'CurrentAxes',handles.axes2);
bimg = im2bw(img,0); % make image black and white using otsu thresholding and
% save in handles.bimg
bimg = (~bimg); % as diatom is black, so invert the image i.e. make diatom white
% and background black
bimg = imfill(bimg, 'holes'); % if there is any holes in the diatom, fill them, because we
% need only outside contour
imshow(bimg,[]);
title('Black & White Image');
hold on

%% Detection of contour
mask = zeros(size(bimg));
mask(10:end-10, 10:end-10) = 1;

InsideOfCntrs = activecontour(bimg, mask, 400);
```

```

visboundaries(InsideOfCntrs , 'color', 'g');
title('Initial Contour');
hold off;
cntrs = bwboundaries(InsideOfCntrs ,8); % find the contours in the black and white
image
if ~isempty(cntrs) % if any contour is found
L = length(cntrs); % find number of contours in black and white image
PointsOfContour = cntrs(1); % choose first contour
for ii = 1 : L % loop over all contours and find largest contour
cntr = cell2mat(cntrs(ii)); % convert contour to matrix from cell array
if (size(cntr,1) > size(PointsOfContour,1)) % check if this contour has more points then
select this one
PointsOfContour = cntr;
end
end
L = size(PointsOfContour,1); % number of points in contour
cntr = PointsOfContour;
ind = 1;
for ii = 1 : L % loop over the number of points in largest contour and find the top left
point in contour
if cntr(ii,1) < cntr(ind,1) && cntr(ii,2) < cntr(ind,2)
ind = ii;
elseif cntr(ii,1) < cntr(ind,1) && cntr(ii,2) == cntr(ind,2)
ind = ii;
elseif cntr(ii,1) == cntr(ind,1) && cntr(ii,2) < cntr(ind,2)
ind = ii;
end
end
for ii = 1 : L % circular shift the contour so that top left point should be first point of
contour
cntr(ii,:) = PointsOfContour(ind,:);
ind = ind + 1;
if ind > L
ind = 1;
end
end
PointsOfContour = cntr; % save the largest contour
% plot(PointsOfContour(:,2),PointsOfContour(:,1),'r','LineWidth',1.5); % show the
contour in axes2
% hold off
% title('Original Contours');
end

%% Approximation of Contour

```

```

[leng wide] = size(PointsOfContour);
% Compute distances between each combination of 2 rows
T = nchoosek(1:leng,2); % pairs of indexes for all combinations of 2 rows
d = zeros(1, length(T));
for mm=1:length(T)
d(mm) = norm(PointsOfContour(T(mm,1),:)-PointsOfContour(T(mm,2),:));
end
% Find maximum distance
[~, maxIndex] = max(d);
PointsOfContour = circshift(PointsOfContour, [-(T(maxIndex,1) - 1), 0]);
threshold = str2double(get(handles.Threshold, 'string'));
IndexBoundary = boundary(PointsOfContour,threshold);
FinalPointsOfContour = PointsOfContour(IndexBoundary,:);
[row col] = size(bimg);
for nn=1:row
for gg=1:col
if bimg(nn,gg) == 1
if inpolygon(nn,gg,FinalPointsOfContour(:, 1), FinalPointsOfContour(:, 2))==0
bimg(nn,gg) = 0;
end
end
if bimg(nn,gg) == 0
if inpolygon(nn,gg,FinalPointsOfContour(:, 1), FinalPointsOfContour(:, 2))==1
bimg(nn,gg) = 1;
end
end
end
end
% Solve the eigenvalues of image using Dirihlet-Laplacian
[size_row, size_col] = size(bimg);
max_length = max(size_row, size_col);
scale = 0;
if max_length>183
scale = 183/max_length;
else
scale = 1;
end
bimg_reduced = imresize(bimg, scale);
[row1, col1] = find(bimg_reduced);
LengthOfObject = length(row1);
IndexOfInsideList = zeros(LengthOfObject, 6);
for xx=1:LengthOfObject
IndexOfInsideList(xx, 1) = row1(xx);
IndexOfInsideList(xx, 2) = col1(xx);

```

```

if bimg_reduced(row1(xx)+1, col1(xx))== 0
IndexOfInsideList(xx, 3) = 0;
else
IndexOfInsideList(xx, 3) = row1(xx)+1;
end
if bimg_reduced(row1(xx)-1, col1(xx))== 0
IndexOfInsideList(xx, 4) = 0;
else
IndexOfInsideList(xx, 4) = row1(xx)-1;
end
if bimg_reduced(row1(xx), col1(xx)-1)== 0
IndexOfInsideList(xx, 5) = 0;
else
IndexOfInsideList(xx, 5) = col1(xx)-1;
end
if bimg_reduced(row1(xx), col1(xx)+1)== 0
IndexOfInsideList(xx, 6) = 0;
else
IndexOfInsideList(xx, 6) = col1(xx)+1;
end
end
CofMatOfLaplac = zeros(LengthOfObject, LengthOfObject);
for kk=1:LengthOfObject
CofMatOfLaplac(kk, kk) = 4;
if IndexOfInsideList(kk, 3) ~= 0
[~, up] = ismember([IndexOfInsideList(kk, 3), ...
IndexOfInsideList(kk, 2)], IndexOfInsideList(:, [1, 2]), ...
'rows');
CofMatOfLaplac(kk, up) = -1;
end
if IndexOfInsideList(kk, 4) ~= 0
[~, down] = ismember([IndexOfInsideList(kk, 4), ...
IndexOfInsideList(kk, 2)], IndexOfInsideList(:, [1, 2]), ...
'rows');
CofMatOfLaplac(kk, down) = -1;
end
if IndexOfInsideList(kk, 5) ~= 0
[~, left] = ismember([IndexOfInsideList(kk, 1), ...
IndexOfInsideList(kk, 5)], IndexOfInsideList(:, [1, 2]), ...
'rows');
CofMatOfLaplac(kk, left) = -1;
end
if IndexOfInsideList(kk, 6) ~= 0
[~, right] = ismember([IndexOfInsideList(kk, 1), ...

```

```

IndexOfInsideList(kk, 6)], IndexOfInsideList(:, [1, 2]), ...
'rows');
CofMatOfLaplac(kk, right) = -1;
end
end
EigvalueOfLaplac = eig(CofMatOfLaplac);
LP1 = zeros(1,13);
LP2 = zeros(1,13);
LP3 = zeros(1,13);
for kkk = 1:13
LP1(kkk) = EigvalueOfLaplac(1)/EigvalueOfLaplac(kkk+1);
LP2(kkk) = EigvalueOfLaplac(kkk)/EigvalueOfLaplac(kkk+1);
LP3(kkk) = LP1(kkk) - LamdaOfDisk(1)/LamdaOfDisk(kkk+1);
end

set(gcf,'CurrentAxes',handles.axes3);
imshow(bimg,[]); % show black and white image
hold on;
visboundaries(bimg, 'color', 'r');
title('Approximated Contour');
hold off;
%% Calculation of Normal angles
set(gcf,'CurrentAxes',handles.axes4);
X = FinalPointsOfContour(:,2); Y = FinalPointsOfContour(:,1);
angs = zeros(1, length(X));
slop1 = [-(Y(1)-Y(end)),(X(1)-X(end))];
slop2 = [-(Y(2)-Y(1)),(X(2)-X(1))];
xAxis1 = [sign(slop1(1)), 0];
xAxis2 = [sign(slop2(1)), 0];
nn1 = norm(slop1)*norm(xAxis1);
nn2 = norm(slop2)*norm(xAxis2);
if nn1 == 0
ang1 = pi/2;
else
ang1 = acos(dot(slop1,xAxis1)/(norm(slop1)*norm(xAxis1)));
end
if nn2 == 0
ang2 = pi/2;
else
ang2 = acos(dot(slop2,xAxis2)/(norm(slop2)*norm(xAxis2)));
end
angs(1) = (ang1 + ang2)/2;
for ff=length(X)-1:-1:2
slop1 = [-(Y(ff)-Y(ff-1)),(X(ff)-X(ff-1))];

```

```

slop2 = [-(Y(ff+1)-Y(ff)),(X(ff+1)-X(ff))];
xAxis1 = [sign(slop1(1)), 0];
xAxis2 = [sign(slop2(1)), 0];
nn1 = norm(slop1)*norm(xAxis1);
nn2 = norm(slop2)*norm(xAxis2);
if nn1 == 0
ang1 = pi/2;
else
ang1 = acos(dot(slop1,xAxis1)/(norm(slop1)*norm(xAxis1)));
end
if nn2 == 0
ang2 = pi/2;
else
ang2 = acos(dot(slop2,xAxis2)/(norm(slop2)*norm(xAxis2)));
end
angs(ff) = (ang1 + ang2)/2;
end

slop1 = [-(Y(end)-Y(end-1)),(X(end)-X(end-1))];
slop2 = [-(Y(1)-Y(end)),(X(1)-X(end))];
xAxis1 = [sign(slop1(1)), 0];
xAxis2 = [sign(slop2(1)), 0];
nn1 = norm(slop1)*norm(xAxis1);
nn2 = norm(slop2)*norm(xAxis2);
if nn1 == 0
ang1 = pi/2;
else
ang1 = acos(dot(slop1,xAxis1)/(norm(slop1)*norm(xAxis1)));
end
if nn2 == 0
ang2 = pi/2;
else
ang2 = acos(dot(slop2,xAxis2)/(norm(slop2)*norm(xAxis2)));
end
angs(end) = (ang1 + ang2)/2;
angle = angs(end:-1:1);
XOfOrigin = 0:1/(length(angle)-1):1;
plot(XOfOrigin, angle,'b','LineWidth',1.5);
title('Original Normal Angle');
grid minor

%% Fourier Transformation of Normal Angles
Period = 1;
K = 20;
t = 0:Period/(length(angle)-1):Period;

```

```

a0 = trapz(t, angle)/Period;
[ak bk] = FourierCoeff(angle, t, Period, K);
fk2 = FourierTrans(a0, ak, bk, t, Period, 2);
set(gcf,'CurrentAxes',handles.axes5);
plot(XOfOrigin, fk2,'b','LineWidth',1.5);
title('Fourier Transform(K=2)');
grid minor
fk10 = FourierTrans(a0, ak, bk, t, Period, 10);
set(gcf,'CurrentAxes',handles.axes6);
plot(XOfOrigin, fk10,'b','LineWidth',1.5);
title('Fourier Transform(K=10)');
grid minor
fk15 = FourierTrans(a0, ak, bk, t, Period, 15);
set(gcf,'CurrentAxes',handles.axes7);
plot(XOfOrigin, fk15,'b','LineWidth',1.5);
title('Fourier Transform(when K=15)');
grid minor
fk20 = FourierTrans(a0, ak, bk, t, Period, 20);
set(gcf,'CurrentAxes',handles.axes8);
plot(XOfOrigin, fk20,'b','LineWidth',1.5);
title('Fourier Transform(K=20)');
grid minor
%% Calculation of Fourior descriptor
FD = zeros(1, 15);
for ss=1:15
fk = FourierTrans(a0, ak, bk, t, Period, ss);
FD(ss) = mean(fk);
end
%% Calculation of Perimeter
perimeter = regionprops(bimg, 'Perimeter');
Perimeter = perimeter.Perimeter;
%% Calculation of Area
area = regionprops(bimg, 'Area');
Area = area.Area;
%% Calculation of Area of Convex Hull
Area_ConvHull = regionprops(bimg, 'ConvexArea');
%% Calculation of Cva
Cva = Area/Area_ConvHull.ConvexArea;
%% Calculation of Perimeter of Convex Hull
bimgOfConvHull = regionprops(bimg, 'ConvexImage');
perimeter = regionprops(bimgOfConvHull.ConvexImage, 'Perimeter');
PerimeterOfConvHull = perimeter.Perimeter;
%% Calculation of Cvp
Cvp = Perimeter/PerimeterOfConvHull;

```

```

%% Calculation of Cr
Cr = Perimeter^2/(4*pi*Area);
%% Rotation of Image in terms of major axis
set(gcf,'CurrentAxes',handles.axes9);
Org = regionprops(bimg,'Orientation');
Orient = Org.Orientation;
bimg = imrotate(bimg, -Orient);
imshow(bimg,[]);
hold on;
title('Shape Parameters');
%% Calculation of Bounding Box
Dimension = regionprops(bimg, 'BoundingBox');
corner = [Dimension.BoundingBox(1) Dimension.BoundingBox(2)];
XLength = Dimension.BoundingBox(3);
YLength = Dimension.BoundingBox(4);
R = XLength*YLength;
CornerPointsOfBoundBox = [corner(1) corner(2); corner(1)+XLength corner(2); ...
corner(1)+XLength corner(2)+YLength; corner(1) corner(2)+YLength; ...
corner(1) corner(2)];
plot(CornerPointsOfBoundBox(:,1), CornerPointsOfBoundBox(:, 2), 'y');
%% Calculation of Rc
Rc = Area/R;
%% Calculation of Ai
Ai =max([XLength YLength])/min([XLength YLength]);
%% Calculation of Center of Gravity
center = regionprops(bimg,'Centroid');
COG = center.Centroid;
plot(COG(1), COG(2), 'b*');
%% Calculation of Main Axis
majorLength = regionprops(bimg,'MajorAxisLength');
LengthOfMajorAxis = majorLength.MajorAxisLength;
minorLength = regionprops(bimg,'MinorAxisLength');
LengthOfMinorAxis = minorLength.MinorAxisLength;
EndPonitsOfMajorAxis = [COG(1)+LengthOfMajorAxis COG(2); ...
COG(1)-LengthOfMajorAxis COG(2)];
EndPonitsOfMinorAxis = [COG(1) COG(2)+LengthOfMinorAxis; ...
COG(1) COG(2)-LengthOfMinorAxis];
plot(EndPonitsOfMajorAxis(:,1), EndPonitsOfMajorAxis(:,2), 'g');
plot(EndPonitsOfMinorAxis(:,1), EndPonitsOfMinorAxis(:,2), 'g');
hold off;
%% Calculation of Sv and Sh
[Xi, Yi, Index] = polyxpoly(CornerPointsOfBoundBox(:,1), CornerPointsOfBoundBox(:,2), ...
EndPonitsOfMajorAxis(:,1),EndPonitsOfMajorAxis(:,2));
PolygonOfMajor = [Xi(1) Yi(1)];

```

```

index = 1;
for kk = Index(1, 1)+1:Index(2, 1)
index = index+1;
PolygonOfMajor(index, :) = [CornerPointsOfBoundBox(kk,1)
CornerPointsOfBoundBox(kk,2)];
end
PolygonOfMajor(index+1,:) = [Xi(2) Yi(2)];
PolygonOfMajor(index+2,:) = [Xi(1) Yi(1)];
bimg_As = bimg;
[row col] = size(bimg);
for nn=1:row
for gg=1:col
if bimg(nn,gg) == 1
if inpolygon(nn,gg,PolygonOfMajor(:, 2), PolygonOfMajor(:, 1))==1
bimg_As(nn,gg) = 0;
end
end
end
end
area3 = regionprops(bimg_As, 'Area');
Area_As = area3.Area;
Area_Ai = Area - Area_As;
Sh = abs(Area_As-Area_Ai)/(Area_As+Area_Ai);

[Xi, Yi, Index] = polyxpoly(CornerPointsOfBoundBox(:,1), CornerPointsOfBoundBox(:,2), ...
EndPonitsOfMinorAxis(:,1),EndPonitsOfMinorAxis(:,2));
PolygonOfMinor = [Xi(1) Yi(1)];
index = 1;
for kk = Index(1, 1)+1:Index(2,1)
index = index+1;
PolygonOfMinor(index, :) = [CornerPointsOfBoundBox(kk,1)
CornerPointsOfBoundBox(kk,2)];
end
PolygonOfMinor(index+1,:) = [Xi(2) Yi(2)];
PolygonOfMinor(index+2,:) = [Xi(1) Yi(1)];
bimg_Ad = bimg;
for nn=1:row
for gg=1:col
if bimg(nn,gg) == 1
if inpolygon(nn,gg,PolygonOfMinor(:, 2), PolygonOfMinor(:, 1))==1
bimg_Ad(nn,gg) = 0;
end
end
end
end

```

```

end
area1 = regionprops(bimg_Ad, 'Area');
Area_Ad = area1.Area;
Area_Ag = Area - Area_Ad;
Sv = abs(Area_Ad-Area_Ag)/(Area_Ad+Area_Ag);
TestParameter(i+1,:) = {[FileName ext], FD(1), FD(2), FD(3), ...
FD(4), FD(5), FD(6), FD(7), FD(8), FD(9), FD(10), FD(11), ...
FD(12), FD(13), FD(14), FD(15), Cva, Cvp, Cr, Rc, Ai, Sh, Sv, ...
LP1(1), LP1(2), LP1(3), LP1(4), LP1(5), LP1(6), LP1(7), ...
LP1(8), LP1(9), LP1(10), LP1(11), LP1(12), LP1(13), ...
LP2(1), LP2(2), LP2(3), LP2(4), LP2(5), LP2(6), LP2(7), ...
LP2(8), LP2(9), LP2(10), LP2(11), LP2(12), LP2(13), ...
LP3(1), LP3(2), LP3(3), LP3(4), LP3(5), LP3(6), LP3(7), ...
LP3(8), LP3(9), LP3(10), LP3(11), LP3(12), LP3(13)};
set(handles.StateOfSampling, 'string', ['Completed image: ' FileName ext]);
disp(['Completed image: ' FileName ext]);
pause(2);
end
xlswrite('TestParameter.xlsx', TestParameter);
set(handles.StateOfSampling, 'string', 'Completed calculation of Parameters');
disp('Calculation of Parameters was completed!');
guidata(hObject, handles);

```

Bibliographie

- [1] Claudon, Nicolas. *Classification automatique des diatomées: une approche par les motifs des structures internes*. Diss. Université du Québec à Trois-Rivières, 2007.
- [2] Siddour, Abdelkader, et al. "Classification des diatomées par les descripteurs de Fourier du contour Diatom Classification by Fourier Descriptors of Outlines." (2007).
- [3] Seckbach, Joseph, and Patrick Kociolek, eds. *The diatom world*. Vol. 19. Springer Science & Business Media, 2011.
- [4] Finlay, Bland J. "Global dispersal of free-living microbial eukaryote species." *Science* 296.5570 (2002): 1061-1063.
- [5] Spanbauer, Trisha L. "Sizes, shifts, & structures: A statistical exploration of changes in diatom morphologies and communities." (2015).
- [6] Mann, David G., and Pieter Vanormelingen. "An inordinate fondness? The number, distributions, and origins of diatom species." *Journal of eukaryotic microbiology* 60.4 (2013): 414-420.
- [7] Seckbach, Joseph, and Patrick Kociolek, eds. *The diatom world*. Vol. 19. Springer Science & Business Media, 2011.
- [8] Loke, R. E., et al. "Diatom recognition by convex and concave contour curvature." *OCEANS'02 MTS/IEEE*. Vol. 4. IEEE, 2002.
- [9] Lundholm, Nina, et al. "Cryptic and pseudo-cryptic diversity in diatoms—with descriptions of *pseudo-nitzschia hasleana* sp. nov. and *p. fryxelliana* sp. nov. 1." *Journal of Phycology* 48.2 (2012): 436-454.

- [10] Verdy, Ariane, Mick Follows, and Glenn Flierl. "Optimal phytoplankton cell size in an allometric model." *Marine Ecology Progress Series* 379 (2009): 1-12.
- [11] Litchman, E., C. A. Klausmeier, and K. Yoshiyama. "Contrasting size evolution in marine and freshwater diatoms." *Proceedings of the National Academy of Sciences* 106.8 (2009): 2665-2670.
- [12] Mann, David G. "The species concept in diatoms." *Phycologia* 38.6 (1999): 437-495.
- [13] Jones, J. "The diversity, distribution and ecology of diatoms from Antarctic inland waters." *Biodiversity & Conservation* 5.11 (1996): 1433-1449.
- [14] Michelutti, Neal, et al. "PERIPHERYIC DIATOM ASSEMBLAGES FROM ULTRA-OLIGOTROPHIC AND UV TRANSPARENT LAKES AND PONDS ON VICTORIA ISLAND AND COMPARISONS WITH OTHER DIATOM SURVEYS IN THE CANADIAN ARCTIC 1." *Journal of Phycology* 39.3 (2003): 465-480.
- [15] Bouchard, Giselle, K. Gajewski, and P. B. Hamilton. "Freshwater diatom biogeography in the Canadian Arctic Archipelago." *Journal of Biogeography* 31.12 (2004): 1955-1973.
- [16] Douglas, Marianne SV, John P. Smol, and Weston Blake. "Marked post-18th century environmental change in high-arctic ecosystems." *Science* 266.5184 (1994): 416-419.
- [17] Theriot, Edward, and Diatom Herbarium. "FE Round, RM Crawford, and DG Mann, The Diatoms. Biology and Morphology of the Genera." (1992): 125-126.
- [18] Davies, Andrew, Alan ES Kemp, and Jennifer Pike. "Late Cretaceous seasonal ocean variability from the Arctic." *Nature* 460.7252 (2009): 254.
- [19] Armstrong, Howard, and Martin Brasier. *Microfossils*. John Wiley & Sons, 2013.
- [20] Huysman, Marie JJ, Wim Vyverman, and Lieven De Veylder. "Molecular regulation of the diatom cell cycle." *Journal of Experimental Botany* 65.10 (2013): 2573-2584.

- [21] Smetacek, Victor. "Making sense of ocean biota: How evolution and biodiversity of land organisms differ from that of the plankton." *Journal of biosciences* 37.4 (2012): 589-607.
- [22] Guo, Q., et al. "Diatom classification by image analysis." International Conférence: Science of Electronic, Technologie of Information and Télécommunications. 2004.
- [23] Lira, Emanuel Gutiérrez, et al. "Image Segmentation and Object Extraction for Automatic Diatoms Classification." International Conference on Image and Signal Processing. Springer, Cham, 2018.
- [24] Khabou, Mohamed A., Lotfi Hermi, and Mohamed Ben Hadj Rhouma. "Shape recognition using eigenvalues of the Dirichlet Laplacian." *Pattern recognition* 40.1 (2007): 141-153.
- [25] Halvorsen, K., M. Lesser, and A. Lundberg. "A new method for estimating the axis of rotation and the center of rotation." *Journal of biomechanics* 32.11 (1999): 1221-1227.
- [26] Trouillot, X. A. V. I. E. R., M. I. C. H. E. L. Jourlin, and JEAN-CHARLES Pinoli. "Geometric parameters computation with Freeman code." 12th International Congress for Stereology, St Etienne, France. 2007.
- [27] Bignot, Gérard. "Introduction à la micropaléontologie." (2001).
- [28] Debled-Rennesson, Isabelle, Salvatore Tabbone, and Laurent Wendling. "Multiorder polygonal approximation of digital curves." *ELCVIA: electronic letters on computer vision and image analysis* 5.2 (2005): 98-110.
- [29] Rosin, Paul L. "Measuring sigmoidality." *Pattern Recognition* 37.8 (2004): 1735-1744.
- [30] Cheriet, Mohamed, et al. *Character recognition systems: a guide for students and practitioners*. John Wiley & Sons, 2007.

- [31] Trouillot, Xavier. Étude de paramètres géométriques à partir du code de Freeman. Diss. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008.
- [32] Wirth, Michael A. "Shape analysis and measurement." University of Guelph. CIS 6320 (2001).
- [33] Rhouma, Mohamed Ben Haj, Mohamed Ali Khabou, and Lotfi Hermi. "Shape recognition based on eigenvalues of the Laplacian." Advances in Imaging and Electron Physics. Vol. 167. Elsevier, 2011. 185-254.
- [34] Cheeseman, Peter, et al. "Autoclass: A Bayesian classification system." Machine learning proceedings 1988. Morgan Kaufmann, 1988. 54-64.
- [35] Guerrien, Marc. "L'intérêt de l'analyse en composantes principales (acp) pour la recherche en sciences sociales. Présentation à partir d'une étude sur le mexique." Cahiers des Amériques latines 43 (2003): 181-192.
- [36] Du Buf, Hans, et al. "Diatom identification : à double challenge called ADIAC." Proceedings 10th International Conference on Image Analysis and Processing. IEEE, 1999.
- [37] Chahboune, Nouzha, Mohamed Mehdi, and Allal Douira. "Automatic classification of diatoms of Merja fouarate." bioRxiv(2017): 099135.
- [38] KELLY, MARTYN G., et al. "Human error and quality assurance in diatom analysis." Automatic diatom identification. 2002. 75-91.
- [39] Du Buf, Hans. Automatic diatom identification. Vol. 51. World Scientific, 2002.
- [40] Dimitrovski, Ivica, et al. "Hierarchical classification of diatom images using ensembles of predictive clustering trees." Ecological Informatics 7.1 (2012): 19-29.

[41] Jalba, Andrei C., et al. "Automatic diatom identification using contour analysis by morphological curvature scale spaces." *Machine Vision and Applications* 16.4 (2005): 217-228.

[42] Belguesmia, Said. Modèle basé sur le filtre de Deriche et sur la transformée de Hough et reconnaissance des formes, Application à la vision robotique. Diss. Université Mouloud Mammeri, 2010.

[43] Kulkarni, Rutuja Nandkumar, and Pradip C. Bhaskar. "Decision based median filter algorithm using resource optimized FPGA to extract impulse noise." *International Journal of Reconfigurable and Embedded Systems* 3.1 (2014): 1.

[44] Liu, Dongju, and Jian Yu. "Otsu method and K-means." *2009 Ninth International Conference on Hybrid Intelligent Systems*. Vol. 1. IEEE, 2009.

[45] S. Beauchemin, P. Bose, A. Chalifour, F. Nouboud and L. vought. Analyse Morphologique des Filets de Tricoptères: Un Problème d'Ecotoxicologie et de Géométrie Discrète. *Visual Interface (VI 98)*, pp.479-486, 1998