

Extraction d'informations pour la cytologie

Projet Tutoré

Université de Caen Normandie
2015 - 2016

Évariste DALLER

Master 1 - Informatique
Mai 2016

Table des matières

1	Présentation des fonctionnalités	6
1.1	Fonctions générales	6
1.1.1	Lecture et visualisation d'images	7
1.1.2	Outils de traitement d'images	7
1.2	Analyse de la population	8
1.3	Descripteurs de contours	9
1.3.1	Sélection par croissance de région	9
1.3.2	Forme et signature	10
1.3.3	Descripteurs de Fourier	10
1.4	Enregistrement et exploitation des résultats	11
1.4.1	Fichiers de résultat	11
1.4.2	Scripts annexes pour l'exploitation	13
2	Concepts et implémentations	14
2.1	Transformations linéaires	14
2.2	Croissance de région	14
2.3	Signature polaire	15
2.3.1	Trouver le centre de gravité	15
2.3.2	Conversion des coordonnées cartésiennes en polaires	16
2.3.3	Normalisation	16
2.3.4	Descripteurs basés sur la signature polaire	17
2.4	Descripteurs de Fourier	18
2.4.1	Quantifier les variations de la tangente	18
2.4.2	Transformée de Fourier et caractéristiques	19
3	Résultats	20
3.1	Présentation des formes étudiées	20

3.2	Descripteurs basés sur la signature	22
3.2.1	Cellules saines	22
3.2.2	Drépanocytes	22
3.2.3	Dacryocytes	23
3.3	Descripteurs de Fourier	24
3.3.1	Cellules Saines	24
3.3.2	Drépanocytes	25
3.3.3	Dacryocytes	25
4	Améliorations	28
4.1	Interface et fonctionnalités d'extraction	28
4.2	Descripteurs	29

Introduction

Nous présentons une application permettant d'extraire quelques informations de photographies de cellules au microscope optique. Ce projet est pour nous un moyen de découvrir et d'approprier les outils théoriques et mathématiques, mais aussi de développement, liés au traitement d'image et à l'extraction d'information.

Nous nous intéresserons ici à l'étude de deux aspects :

- La population de cellules (et son évolution)
- La caractérisation des contours des cellules

Nous mettrons en place une interface graphique permettant d'utiliser les outils proposés, avec l'API *Qt*, et nous utiliserons les outils de traitements d'image proposés par *OpenCV*.

En outre, nous nous appuierons sur des concepts plus théoriques, comme la signature d'une forme, afin de proposer et implémenter deux descripteurs de contours dont nous présenterons les résultats sur jeux de données réels.

L'application présentée peut être vue comme une boîte à outils pouvant évoluer pour servir à n'importe quel travail d'extraction d'information sur une série d'images.

Nous présenterons dans un premier temps l'interface graphique et ses fonctionnalités, puis nous expliciterons les concepts sous-jacents et les implémentations réalisées, lorsqu'elles ne sont pas proposées par *OpenCV*. Enfin, nous présenterons les résultats obtenus pour l'extraction de descripteurs de contours, ainsi que les formes étudiées.

1 Présentation des fonctionnalités

1.1 Fonctions générales

L'interface utilisateur repose sur une distinction en deux panneaux entre l'affichage des images et les composants d'analyse et d'extraction d'informations.

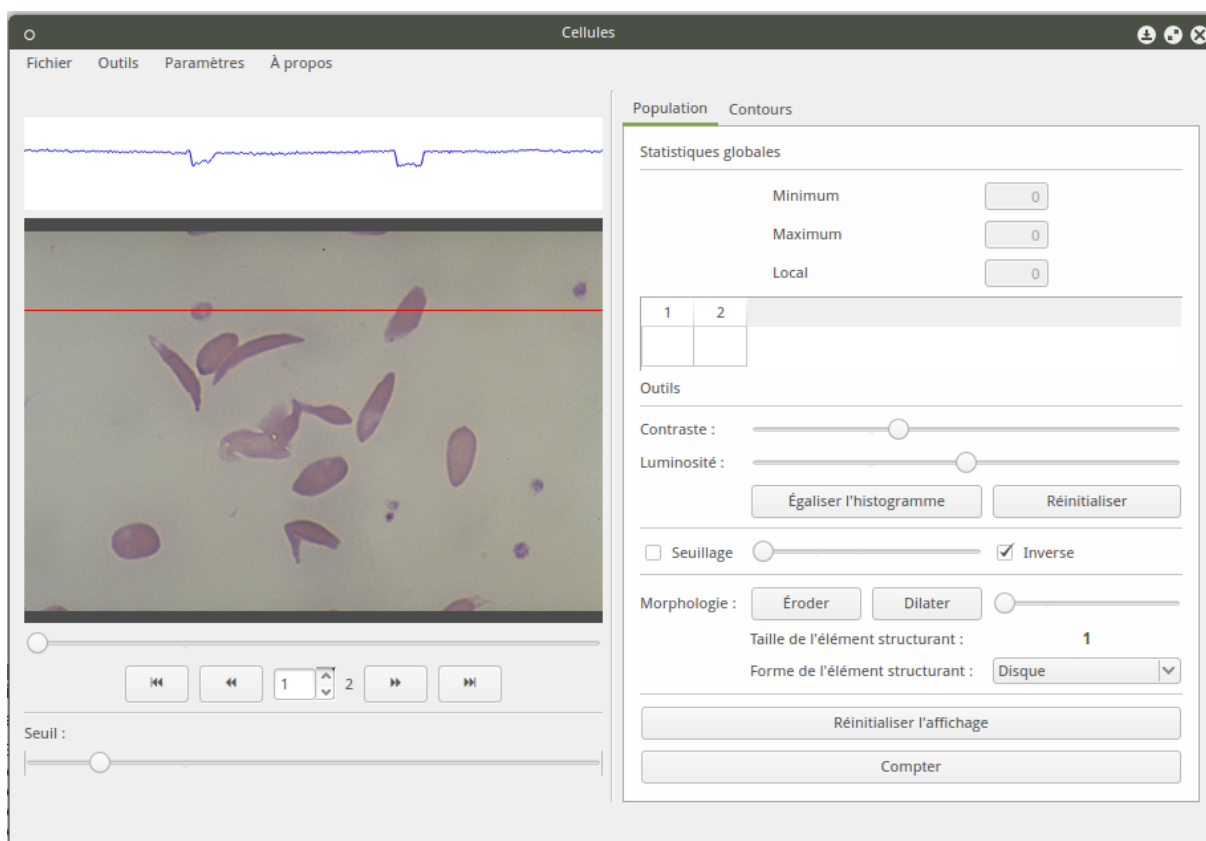


FIGURE 1.1 – Fenêtre principale

L'opérateur utilise le panneau de gauche pour visualiser les images et naviguer dans la liste d'images, ainsi que pour activer des outils simples de visualisation comme des outils de profil ou de seuillage. Il utilise le panneau de droite pour contrôler les outils d'extractions d'informations suivants :

- Population
- Contours

Pour utiliser ces composants, l'utilisateur doit les activer en passant par le menu *Paramètres* → *Composant Actifs*. Un seul composant peut être actif à la fois, car il monopolise l'affichage.

1.1.1 Lecture et visualisation d'images

L'interface permet de visualiser l'image courante (1 sur la figure), de connaître sa position dans la liste d'images et de naviguer dans celle-ci à l'aide de la barre de navigation (2), et de manipuler les outils activés présentés dans la sections suivante (3).

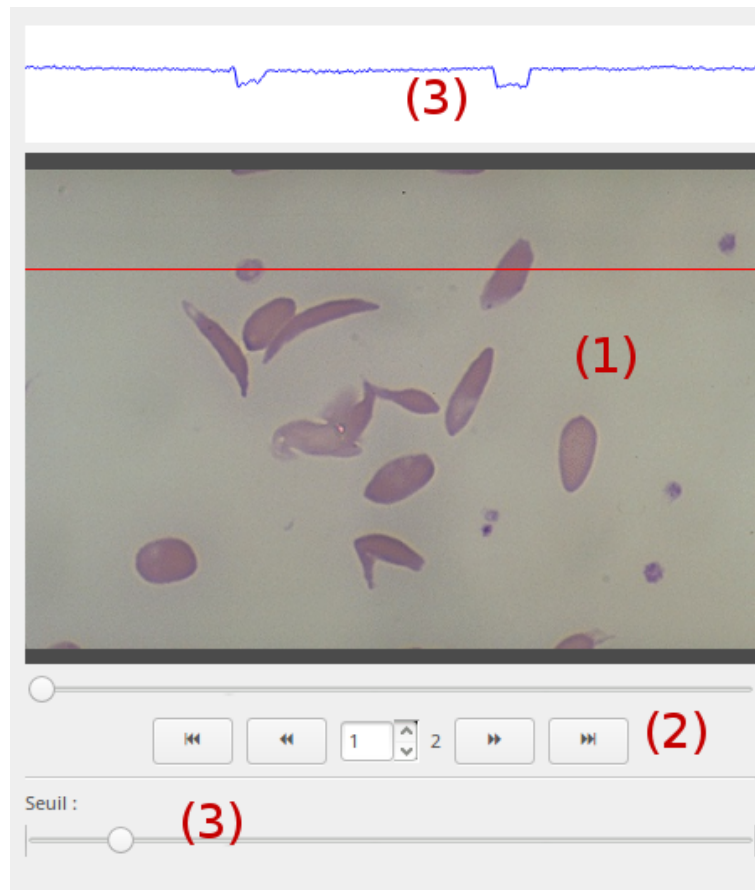


FIGURE 1.2 – Panneau de visualisation

On importe une liste de fichiers images avec l'action *Fichier* → *Ouvrir un répertoire*, ou *Fichier* → *Ouvrir une liste d'images*.

1.1.2 Outils de traitement d'images

On active les outils de traitement d'image par le menu *Outils*. Un outil n'est affiché dans l'interface que s'il est activé.

Outils de coupe Les outils de coupe (ou de profil) permettent de voir le signal d'une image en niveau de gris le long d'une ligne ou d'une colonne. Pour sélectionner une ligne (resp. une colonne),

on clique sur la vue de l'outil approprié (une fois activé), puis on clique dans l'image en un point correspondant à la ligne (resp. colonne) voulue. L'outil se met à jour en affichant le profil de l'image.

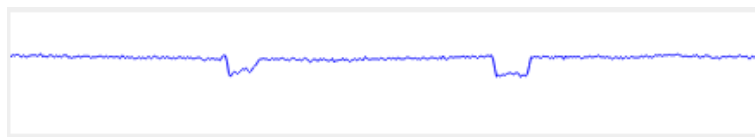


FIGURE 1.3 – Outils de coupe horizontal

Outil de seuillage L'outil de seuillage permet de définir un niveau de déclenchement sur les différents canaux de l'image en dessous duquel la valeur des pixels est mise à zéro. La valeur des pixels au dessus du seuil est inchangée.

1.2 Analyse de la population

Une fois que le composant *Population* est activé, on peut l'utiliser (dans l'onglet de droite du même nom) pour compter les cellules présentes sur la photographie en utilisant des outils de morphologie mathématique. Le principe est de binariser l'image de façon à voir des formes correspondant aux cellules se dessiner. Une fois qu'on a obtenu une forme séparée des autres pour chaque cellule, on demande à l'outil de les compter.

L'interface présente les fonctionnalités décrites ci-après.

Statistiques : Tient à jour les statistiques de base sur la population de la liste d'image depuis le début de l'exécution du programme. On y trouve le minimum de cellule dans une image, le maximum, ainsi que la population de l'image courante dans la rubrique *Local*.

Tant qu'aucun comptage n'a été effectué, ces informations restent à zéro.

Tableau de résultats : Le cadre suivant est un tableau contenant, pour chaque image, le nombre de cellules détectées. Ce tableau est mis à jour à chaque comptage effectué.

Luminosité et contraste : Afin de réaliser un bon seuillage (nécessaire pour la détection), on peut utiliser cet outil pour ajuster la luminosité et le contraste de l'image. Les transformations appliquées sont des transformations linéaires décrites en section 2.1

Seuillage (binarisation) : On active le seuillage en cochant la case correspondante. Le niveau de déclenchement est contrôlé par le curseur adjacent. Si la case *Inverse* est cochée, le seuillage est inversé, ceci permet de mieux appréhender le fait que les éléments sombres sont en général ceux qui nous intéressent dans les photographies. En effet, sur l'image en noir et blanc obtenue, les formes correspondant aux cellules doivent être blanches sur fond noir.

Morphologie : Il n'est pas rare d'obtenir à ce stade une image présentant des imperfections qui pourraient altérer le comptage :

- Du bruit ou des parasites peuvent apparaître et générer de petites formes à ne pas prendre en compte comme des cellules.
- Certaines cellules peuvent être collées, si bien que le programme ne les distinguera pas.

On a alors recours à l'érosion et la dilatation pour corriger ces problèmes.

On peut définir la forme de l'élément structurant à utiliser, ainsi que sa taille. Cette taille est définie comme le rayon maximal en pixels. Cela signifie que si la taille sélectionnée est n , alors l'élément utilisé aura une taille réelle de $2n + 1$ pixels.

Les trois formes d'éléments disponibles sont :

- Un disque : ●
- Un carré : □
- Une croix : ×

On peut réaliser des opérations successives, et ainsi produire des ouvertures ou fermetures. On peut aussi changer la forme de l'élément structurant entre deux opérations pour faciliter la séparation des formes.

Si les opérations ne sont pas concluantes et que l'on veut recommencer, on peut utiliser la fonction *Réinitialiser l'affichage*, qui efface les opérations de morphologies effectuées et permet de reprendre notre travail avec l'image binaire préalablement obtenue.

Compter : Une fois que l'image (binaire) présente des formes séparées pour chaque cellule, on utilise la fonctionnalité *Compter* pour exécuter un comptage sur cette image. Le programme extrait alors les contours extérieurs des formes (les trous ne sont pas pris en compte), les affiche sur l'image, et met à jour le tableau de résultats et les statistiques.

1.3 Descripteurs de contours

Le deuxième composant, placé sous l'onglet *Contours*, permet d'extraire deux descripteurs du contour d'une cellule. Il s'agit d'un descripteur basé sur la signature polaire de la forme extraite (voir sections 2.3 et 2.3.4) et de descripteurs de Fourier (section 2.4).

1.3.1 Sélection par croissance de région

Pour commencer l'analyse, on sélectionne une cellule en procédant à une segmentation par croissance de région (détails au paragraphe 2.2). Afin de réaliser cette opération, on place un germe au sein d'une cellule sur l'image en cliquant préalablement sur *Placer le germe*. Suivant le seuil de tolérance et le critère d'homogénéité choisi, la région sélectionnée sera plus ou moins grande et fidèle à la forme de la cellule.

On peut voir la région exacte en image binaire si la case *Afficher uniquement le contour* est décochée, sinon le contour externe est affiché sur l'image d'origine. Si la région présente des trous, ils ne seront pas pris en compte pour la suite (on ne s'intéresse qu'au contour externe des formes).

Si la case *Afficher uniquement le contour* est cochée, le contour est affiché en rouge sur

l'image et le centre de gravité est pointé par un petit disque bleu. On s'aidera de cet affichage ainsi que des autres outils décrits ci-après pour s'assurer que la forme est correctement sélectionnée (centre de gravité et signature cohérents *etc*)

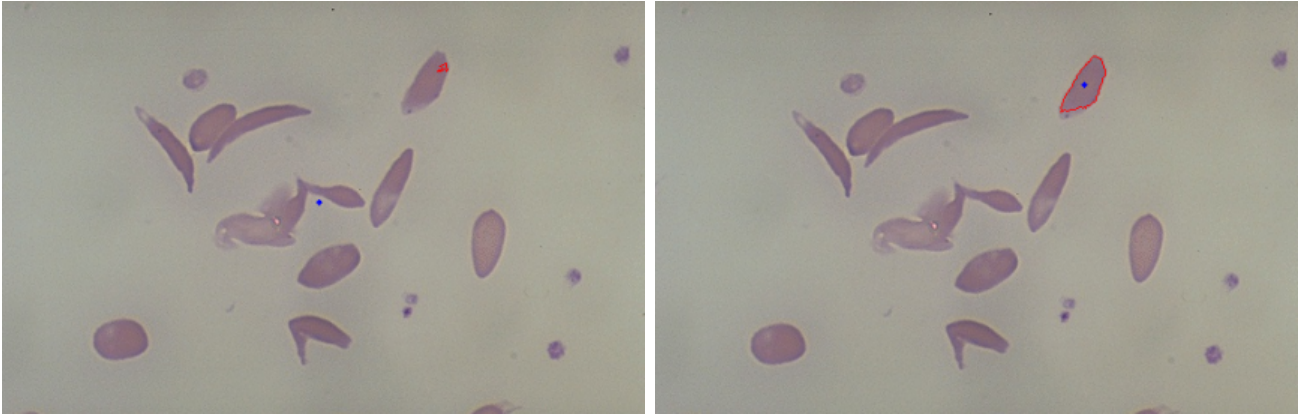


FIGURE 1.4 – À gauche : mauvaise sélection, à droite : bonne sélection

1.3.2 Forme et signature

La section *Forme sélectionnée* présente la forme précédemment segmentée sous deux aspects : la forme centrée sur son centre de gravité (calculé suivant la méthode décrite en section 2.3.1), et sa signature polaire normalisée (voir section 2.3).

Sur le graphique présentant la forme, le centre de gravité est l'origine, et les axes sont contractés ou étendus pour faire tenir la forme dans un carré de dimension 1 : 1.

La signature étant quand à elle normalisée sur ses deux dimensions, le graphique l'affichant présente toujours les mêmes intervalles sur ses axes : $x \in [0; 2\pi]$ et $y \in [0; 1]$.

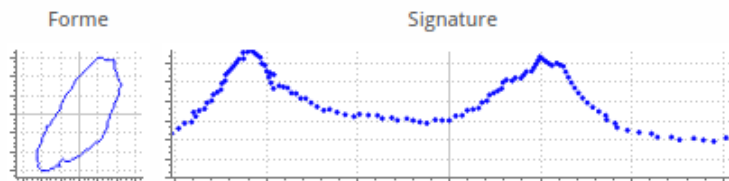


FIGURE 1.5 – Affichage de la forme et de sa signature

1.3.3 Descripteurs de Fourier

Les descripteurs de Fourier (voir section 2.4) ont pour but de quantifier les variations du contour dans le domaine fréquentiel. Ils permettent de distinguer les composantes principales du signal (basses fréquences) des variations rapides (bords déchiquetés, bruit, *etc* : hautes fréquences).

Variations de la tangente

Comme expliqué au paragraphe 2.4, on se base ici sur les variations de la tangente au contour en fonction de la distance parcourue sur celui-ci (relativement à la première mesure). Ici, l'intervalle en abscisse correspondant au nombre de points du contour est normalisé dans $[0; 2\pi]$ et l'angle de la tangente varie entre $-\pi$ et π .

La droite grise déjà présente sur le graphique représente un cercle parfait (variation de l'angle linéaire de $-\pi$ à π entre 0 et 2π).

Ainsi, plus les points affichés s'éloignent de la droite pour rejoindre $(0; -\pi)$ à $(2\pi; \pi)$, plus la forme est différente d'un cercle.

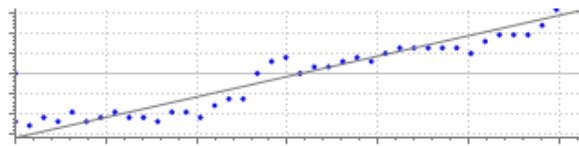


FIGURE 1.6 – Angle de la tangente pour chaque point du contour

Affichage de l'histogramme

La transformée de Fourier des variations de la tangente (dérivée de l'angle) est automatiquement calculée, et les n premières harmoniques sont affichées sous forme d'histogramme (n étant le nombre choisi dans la boîte *Harmoniques*).

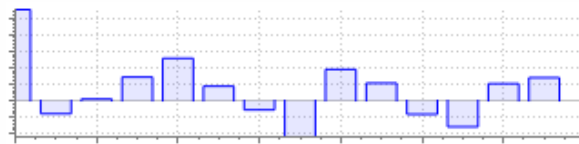


FIGURE 1.7 – Histogramme (14 harmoniques)

1.4 Enregistrement et exploitation des résultats

Une fois qu'on a obtenu des résultats, concernant des populations ou un contour, on peut les enregistrer au format XML afin de les exploiter avec des programmes externes.

On enregistre un résultat à l'aide de l'action *Fichier* \rightarrow *Enregistrer les résultats*.

1.4.1 Fichiers de résultat

Le résultat alors enregistré correspondra au composant actuellement actif. Il s'agit toujours d'un fichier XML, mais le contenu varie d'un composant à l'autre.

Population

Les fichiers concernant une population reproduisent le contenu du tableau listant le nombre de cellules en fonction du numéro de l'image. Pour chaque image, une balise **frame** est créée, portant un attribut **id** correspondant au numéro de l'image concernée. Le contenu textuel de chacune de ces balises est le nombre de cellules compté correspondant.

```
1 <!DOCTYPE Population>
2 <population>
3   <frame id="0">85</frame>
4   <frame id="1">87</frame>
5   <frame id="2">80</frame>
6   <frame id="3">53</frame>
7   <frame id="4">84</frame>
8   <frame id="5">73</frame>
9 </population>
```

Listing 1.1 – Exemple de résultat de population

Contours

Les fichiers de résultats associés aux contours listent les valeurs des deux descripteurs concernés :

- Les deux dimensions du descripteur de la signature polaire
- Les n harmoniques correspondant au descripteur de Fourier

Le descripteur pour la signature est défini à la section 2.3.4.

On voit donc qu'un fichier XML correspond à un résultat donné (une cellule).

```
1 <!DOCTYPE Contours>
2 <descripteurs>
3   <signature>
4     <variance>0.0195555</variance>
5     <pics>2</pics>
6   </signature>
7   <fourier>
8     <harm n="0">4.71239</harm>
9     <harm n="1">-1.93341</harm>
10    <harm n="2">0.0788933</harm>
11    <harm n="3">0.786645</harm>
12    <harm n="4">0.263966</harm>
13    <harm n="5">1.60437</harm>
14    <harm n="6">0.277774</harm>
15    <harm n="7">1.87989</harm>
16    <harm n="8">-0.155759</harm>
17    <harm n="9">2.54605</harm>
18   </fourier>
19 </descripteurs>
```

Listing 1.2 – Exemple de résultat pour les contours

1.4.2 Scripts annexes pour l'exploitation

Afin de plus facilement exploiter les fichiers XML fournis, quelques scripts python annexes sont présentés ici.

compactXmlDescriptors : Permet de compacter tous les fichiers XML d'un répertoire en un seul fichier. L'arbre XML de chaque fichier sera présent sous un nœud **shape** portant un attribut **id** correspondant au nom du fichier concerné.

plotPolar : Affiche un graphique en deux dimensions (correspondant aux deux dimensions du descripteur) où sont construits les nuages de points correspondant aux descripteurs de la signature en provenance de fichiers compactés avec le script précédent. Chaque fichier fourni en entrée (jusqu'à trois) correspondra à une classe de points sur le graphique.

plotFourier et plotFourier3d Ces outils permettent d'afficher des points sur un graphique de la même manière que pour le précédent script, par rapport aux harmoniques sélectionnées pour les descripteurs de Fourier. **plotFourier** représente les points suivant deux dimensions (deux harmoniques), **plotFourier3d** permet de représenter les points suivant trois harmoniques.

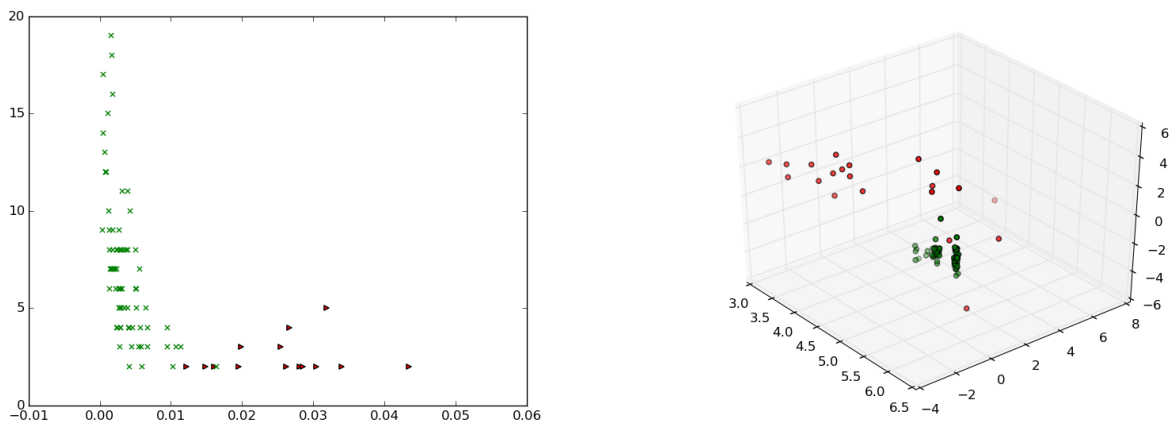


FIGURE 1.8 – Exemples d'images obtenues avec ces outils

2 Concepts et implémentations

Nous décrivons dans cette partie les outils mathématiques et de traitement d'image que nous utilisons au sein de ce projet. Nous présentons et commentons aussi les implémentations de ces concepts lorsqu'ils ne font pas partie de modules externes (OpenCV par exemple).

2.1 Transformations linéaires

Pour régler le contraste et la luminosité d'une image, on utilise une transformation linéaire : soit α le contraste et β la luminosité, pour chaque pixel (x, y) de valeur $p(x, y)$, on opère la transformation $p'(x, y) \leftarrow \alpha.p(x, y) + \beta$

```
Data :  $I$  une image,  $\alpha \in \mathbb{R}^+$ ,  $\beta \in \mathbb{R}$   
Result :  $I'$  une image  
foreach pixel  $I_{(x,y)}$  do  
|  $I'_{(x,y)} \leftarrow \alpha.I_{(x,y)} + \beta$   
end
```

Algorithme 1 : Contraste et Luminosité

2.2 Croissance de région

Afin de sélectionner une cellule sur l'image lors de l'utilisation du composant *Contours*, on met en place une segmentation par croissance de région à partir d'un germe.

On définit les termes suivants :

Germe : Un germe (*seed* en anglais) utilisé pour une croissance de région est un pixel de l'image concernée. On notera $G = \langle (x, y) \rightarrow val \rangle$ le fait que le germe G est de coordonnées (x, y) et de valeur val (nuance de gris).

Critère d'homogénéité : Un critère d'homogénéité est ici une fonction booléenne h pouvant prendre comme paramètre l'image d'origine, la portion de région déjà découverte (dans laquelle se trouve le germe) et le germe lui-même. On dit que le critère h est satisfait lorsqu'il renvoie \top (vrai).

En partant du germe, l'algorithme de croissance de région va tester pour chaque voisin le critère d'homogénéité. Si celui-ci est satisfait, le pixel voisin concerné est stocké dans une pile. On

ne visite pas deux fois le même pixel (on sauvegarde les passages dans un masque). Une fois cette itération terminée, on recommence avec le premier pixel de la pile. On réitère jusqu'à ce que la pile soit vide.

Data : I une image, $G = \langle (x, y) \rightarrow val \rangle$ un germe et h un critère d'homogénéité

Result : M une image masque de la région

Initialisation :

Soit M une image binaire de même taille que I

Soit V une image binaire carte des visites (même taille que I)

Soit P une pile de pixels, vide

$\forall p$ pixel, $M_p \leftarrow V_p \leftarrow \perp$

empiler le germe (x, y)

while P n'est pas vide **do**

$pix \leftarrow \text{dépiler}(P)$

$M_{pix} \leftarrow \top$

foreach v voisin de pix **do**

if $\neg V_v \wedge h(I, M, G)$ **then**

 empiler v

$V_v \leftarrow \top$

end

end

end

Algorithme 2 : Croissance de région

2.3 Signature polaire

Un des deux descripteurs utilisés ici pour discriminer les formes suivant leur contour est basé sur la signature polaire de la forme. Cette signature correspond à un "déroulage" du contour de la forme relativement à son centre de gravité. On exprime donc la distance entre chaque point P du contour et le centre de gravité G en fonction de l'orientation de la droite $[GP]$ (entre 0 et 2π), comme expliqué dans [1].

2.3.1 Trouver le centre de gravité

En reprenant la définition du centre de gravité par les moments de [1] (§2.4.2), on a le centre de gravité $G = (x, y)$ avec :

$$x = \frac{M_{1,0}}{M_{0,0}} \text{ et } y = \frac{M_{0,1}}{M_{0,0}} \quad (2.1)$$

En considérant l'équation générale des moments :

$$M_{i,j} = \sum_{k=1}^K \sum_{l=1}^L k^i \cdot l^j \cdot g(k, l) \quad (2.2)$$

Ce qui revient à la moyenne des lignes et des colonnes (séparément) pondérée par les niveaux de gris g .

$$x = \frac{\sum_{\forall(k,l)} k \cdot g(k,l)}{\sum_{\forall(k,l)} g(k,l)} \quad \text{et} \quad y = \frac{\sum_{\forall(k,l)} l \cdot g(k,l)}{\sum_{\forall(k,l)} g(k,l)} \quad (2.3)$$

Ici, le niveau de gris g d'un pixel est 0 si ce pixel n'appartient pas à la forme. Dans notre implémentation, on utilise un masque pour savoir si un pixel appartient à la forme. On inverse aussi la pondération des niveaux de gris en remplaçant $k \cdot g(k,l)$ par $k \cdot (MAX - g(k,l))$ car, sur les photographies, les formes foncées (de faible niveau de gris) sont généralement plus importantes que les formes claires.

Centrage du repérage

Dans la suite, on centre le repère en G par translation. On translate chaque point (x, y) du contour en $(x, y) - G$. Le centre de gravité devient donc $G - G = O$.

2.3.2 Conversion des coordonnées cartésiennes en polaires

Le principe de la signature polaire est d'exprimer les points du contour par leur distance au centre de gravité en fonction de l'angle de la droite passant par G et les interceptant (comme décrit différemment dans [1]). Ceci revient ici à exprimer chaque point du contour en coordonnées polaires, car G est maintenant considéré comme l'origine.

Soit un point du plan cartésien $A = (x, y)$. On l'exprime en coordonnées polaires par $A = (\theta, r)$ avec $\theta \in \mathbb{R}$ et $r \in \mathbb{R}^+$. On exprimera θ dans $] -\pi, \pi]$ avec la fonction $atan2$.

$$\theta = atan2(y, x) \quad \text{et} \quad r = \sqrt{x^2 + y^2} \quad (2.4)$$

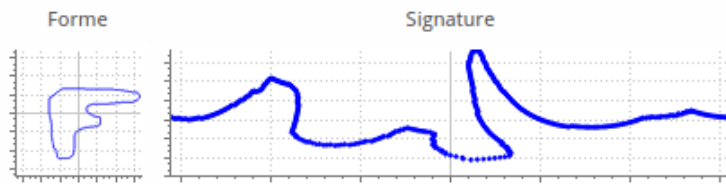


FIGURE 2.1 – Signature d'un F

2.3.3 Normalisation

Afin de ne traiter que l'apparence des formes, mais pas leur échelle, on normalise la signature polaire telle qu'on l'a obtenue. Remarquons tout d'abord que l'angle n'a pas besoin de traitement particulier, car il est toujours dans l'intervalle $] -\pi; \pi]$. Cependant, la distance (magnitude) doit être comparable d'une forme à l'autre. On normalise donc cette donnée entre 0 et 1.

$$\forall r, \quad r \leftarrow r \cdot \frac{1}{r_{max}} \quad (2.5)$$

2.3.4 Descripteurs basés sur la signature polaire

La signature polaire peut être vue comme une mesure de dissimilarité avec le cercle. En effet un cercle aura toujours une signature polaire constante égale à son rayon. On peut donc caractériser visuellement assez facilement la forme à partir de sa signature.

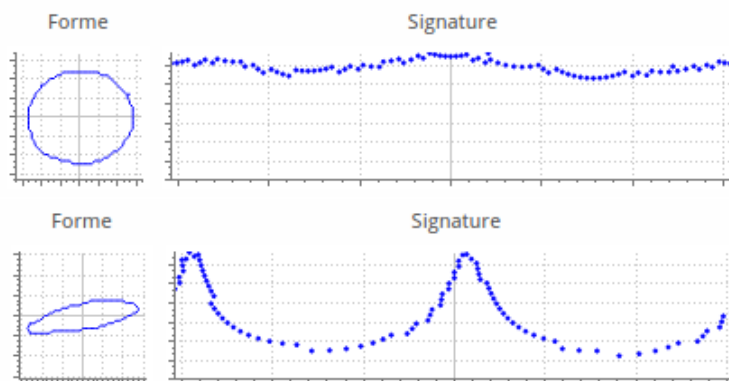


FIGURE 2.2 – En haut : hématie saine, en bas : hématie drépanocytaire

Cependant, la signature polaire n'est pas une fonction (plutôt un nuage de points). Nous allons donc la quantifier par des méthodes statistiques.

Nous proposons comme descripteur un vecteur à deux dimension :

- La variance de la magnitude du nuage de points
- Le nombre de "pics" présents

Pics Les formes que nous voulons distinguer (voir section 3.1) sont différenciables par leur étirement global ou local, c'est-à-dire par les groupes de points de la signature polaire qui ont une magnitude importante. Nous appelons ces groupes des pics (voir Figure 2.2). Remarquons que ces pics, ainsi que leur nombre, quantifient l'allongement et la quantité d'appendices ou d'angles qu'a la forme.

La Figure 2.3 montre la différence entre la signature polaire d'une ellipse, d'un triangle et d'un carré. On peut caractériser ces formes par le nombre et l'intensité de leurs pics (des dégradés ont été mis en place pour décentrer le centre de gravité).

On définit numériquement un pic comme un groupe de points successifs au-dessus de la médiane. De cette manière, les formes allongées présenteront deux pics correspondants aux deux extrémités éloignées du centre de gravité. En revanche, les formes proches du cercle en présenteront beaucoup plus.

Extraction du descripteur Le nombre de pics est calculé simplement en parcourant les points du contour dans l'ordre, et en incrémentant un compteur lorsque l'on passe dans un groupe supérieur à la médiane. Celle-ci est calculée selon la méthode proposée par N. Wirth dans [2] et implémentée

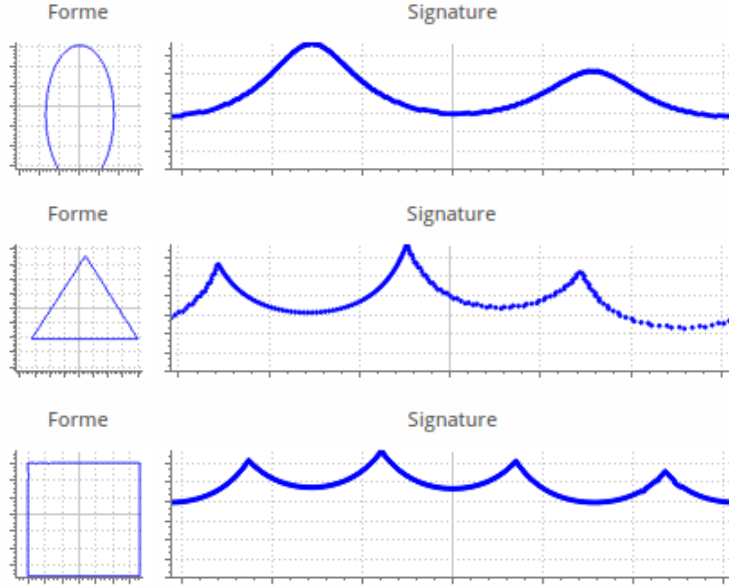


FIGURE 2.3 – Signature de trois formes géométriques

en C dans [3]. On tient compte de la périodicité du signal en vérifiant, lorsque le dernier point est dans un de ces groupes, si le premier y est aussi. Dans ce cas on retrace 1 au compteur.

2.4 Descripteurs de Fourier

On peut décomposer n'importe quel signal périodique en une somme de sinusoides (potentiellement une infinité) : une série de Fourier. On généralise ce procédé à n'importe quel signal par une transformée de Fourier. OpenCV propose une implémentation de transformée de Fourier discrète que l'on va utiliser pour extraire des quantificateurs liés aux variations du contour d'une forme.

La transformée de Fourier discrète nous donnera ainsi une représentation spectrale du contour, que l'on utilisera pour discriminer la forme, en se basant sur [1] et [4].

2.4.1 Quantifier les variations de la tangente

Tout d'abord, on va extraire un signal en bijection avec le contour. Ce signal sera l'inclinaison de la tangente à la courbe en fonction de l'azimut du point tangent (coordonnées polaires).

On définit la tangente au contour en un point P_n comme la droite passant par les points P_n et $P_{n+\epsilon}$ où ϵ est un petit entier supérieur à 1. Si on prenait deux points consécutifs, on risquerait d'avoir des sauts d'angle correspondants à des escaliers de pixels, ce qui ne décrirait pas du tout correctement le contour.

L'angle de cette droite est $\alpha = \text{atan2}(P_{n+k}^y - P_n^y, P_{n+k}^x - P_n^x)$ et l'azimut de chaque point a déjà été calculé pour la signature polaire.

On obtient ainsi une fonction périodique décrivant le contour. Pour un cercle, en admettant que le premier point du contour donne une tangente d'un angle de $-\pi$, cette fonction sera périodique

en dents de scies : c'est-à-dire linéaire sur sa période.

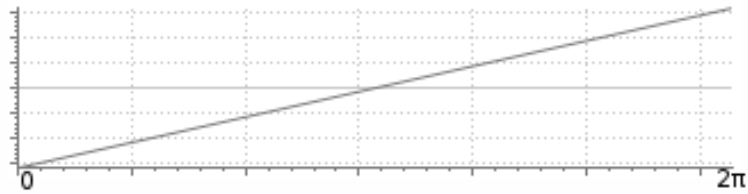
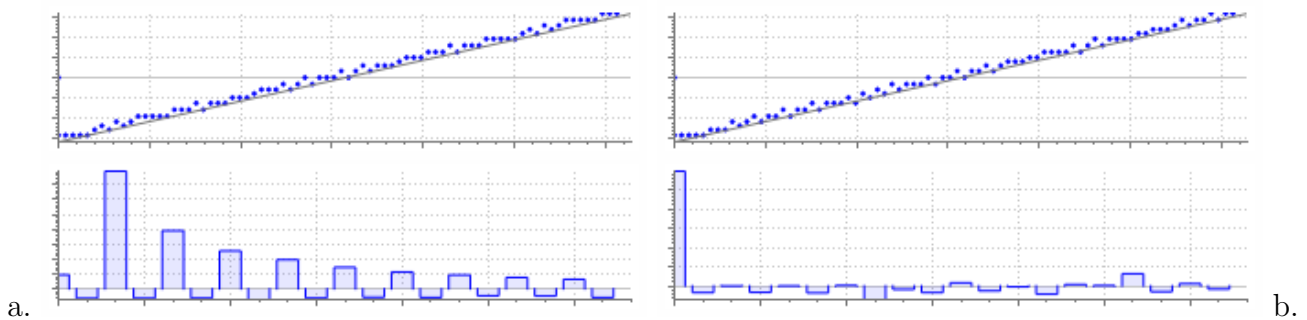


FIGURE 2.4 – Inclinaison de la tangente pour un cercle

2.4.2 Transformée de Fourier et caractéristiques

On pourrait effectuer une transformée de Fourier sur ce signal pour l'analyser ensuite. Cependant on caractériserait des signaux plus ou moins en dents de scie (donc plus ou moins linéaires sur la fenêtre) modulés par quelques variations. Il sera alors plus judicieux de procéder comme dans [1], et de se baser sur la dérivée du signal précédent.

De cette manière, on s'affranchit de la composante linéaire et le contour d'un cercle devient un signal constant (cf Figure 2.5).



Transformée de fourier sur les signaux extraits du contour d'un cercle. a. Orientation de la tangente. b. Dérivée

FIGURE 2.5 – Comparaison entre les transformées du signal et de sa dérivée

Caractéristiques spectrales

La transformée discrète donne une représentation spectrale (ordonnée) du signal correspondant aux variations du contour. Les harmoniques graves (ou de basse fréquence) correspondent aux composantes de grande longueur d'onde sur le signal tandis que les harmoniques de haute fréquence sont liées aux variations rapides du signal. Ainsi, une image bruitée donnera une représentation spectrale dont les harmoniques hautes sont plus intenses. À contrario, un cercle quasi-parfait ne présentera qu'une fondamentale (harmonique 0) forte.

3 Résultats

Nous présentons dans cette partie les résultats obtenus sur des jeux de données réels (photographies au microscope optique). Nous tenterons de distinguer trois types de cellules de formes différentes à l'aide de nos deux descripteurs et de l'interface présentée dans les chapitres précédents.

3.1 Présentation des formes étudiées

Les trois formes à discriminer sont caractéristiques de trois type d'hématies (globules rouges ou érythrocytes) :

- Hématies saines
- Drépanocytes (atteintes de drépanocytose)
- Dacryocytes (hématies en forme de larme)

On voit dans [5] que ces cellules sont généralement caractérisées par les biologistes d'après leur forme particulière.

Cellules saines Selon [5] :

Les globules rouges normaux (...), de forme discoïde aplatie, ils apparaissent régulièrement arrondis et ont une coloration rosée homogène à renforcement périphérique.

Ces cellules ont une forme proche du disque, parfois légèrement aplatie, avec une coloration homogène (ce qui devrait donner des régions dont le contour est lisse –et non déchiqueté).

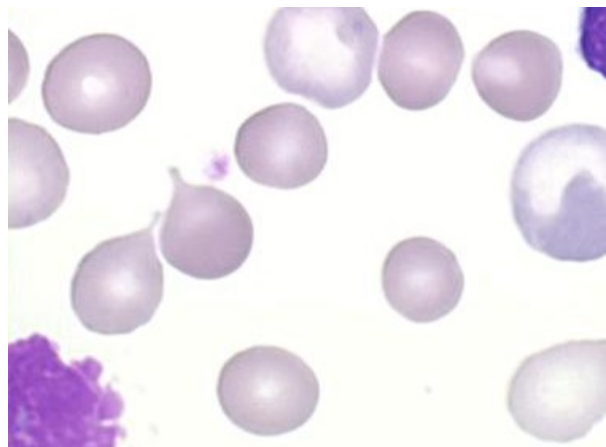


FIGURE 3.1 – Hématies normales

Drépanocytes La drépanocytose est la première maladie génétique en France¹, et est due à une malformation de l'hémoglobine (la protéine permettant le transport du dioxygène au sein des hématies), qui déforme les globules rouges. Cette déformation se traduit par une forme d'hématie très allongée, voir en forme de faucille (du grec *drepanon*).

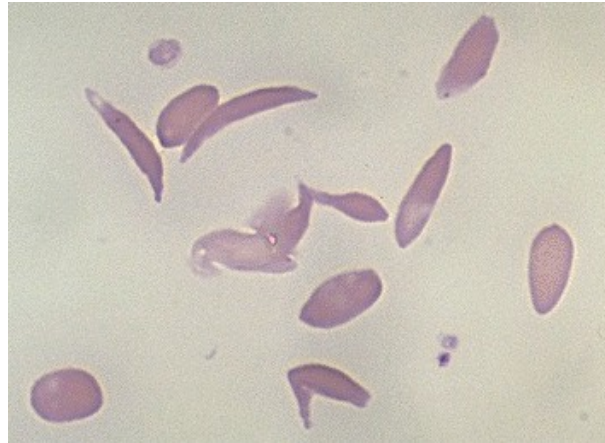


FIGURE 3.2 – Hématies drépanocytaires

Dacryocytes Les dacryocytes sont des hématies en forme de larme, caractéristiques de la myélofibrose avec métaplasie myéloïde (MMM). Cette anémie est décrite par *vulgaris medical*² comme :

caractérisant une perte de fonctionnement normal de la moelle osseuse entraînant une perte d'élasticité de celle-ci (...) ne l'autorisant plus à fabriquer les éléments essentiels du sang

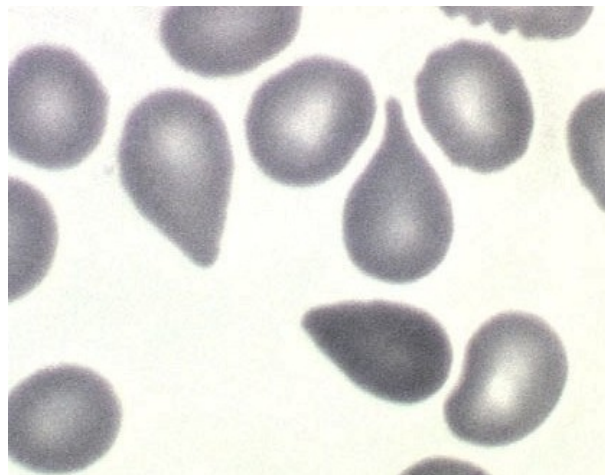


FIGURE 3.3 – Hématies dacryocytes

1. et probablement dans le monde, selon l'association Drepavie (<http://www.drepavie.org/>)
2. <http://www.vulgaris-medical.com/encyclopedie-medicale/metaplasie-myeloide>

3.2 Descripteurs basés sur la signature

Nous présentons ici les résultats obtenus pour le descripteur basé sur la signature polaire.

3.2.1 Cellules saines

Les hématies normales sont presque des disques, leur contour est proche du cercle. La signature polaire de ces cellules sera donc presque constante, avec très peu de variations.

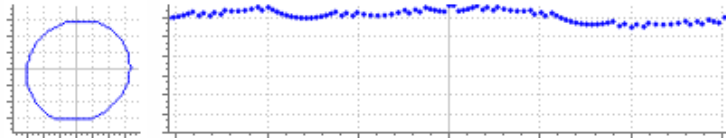


FIGURE 3.4 – Forme et signature polaire d'une cellule saine

On a généralement une variance très faible et potentiellement beaucoup de groupes de points au-dessus de la médiane (pics).

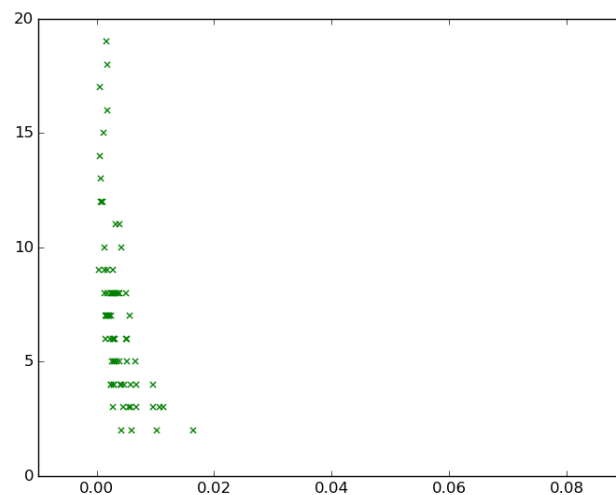


FIGURE 3.5 – Descripteurs SP pour 74 cellules saines

3.2.2 Drépanocytes

Les cellules drépanocytaires sont très allongées et potentiellement recourbées. Leur signature polaire va donc présenter systématiquement deux pics et une variance prononcée.

On peut comparer en Figure 3.7 les nuages de points des cellules saines et drépanocytaires pour 21 drépanocytes. Ces nuages sont clairement séparés et permettent de réaliser sans problème une classification.

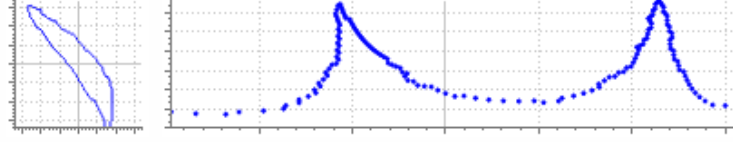
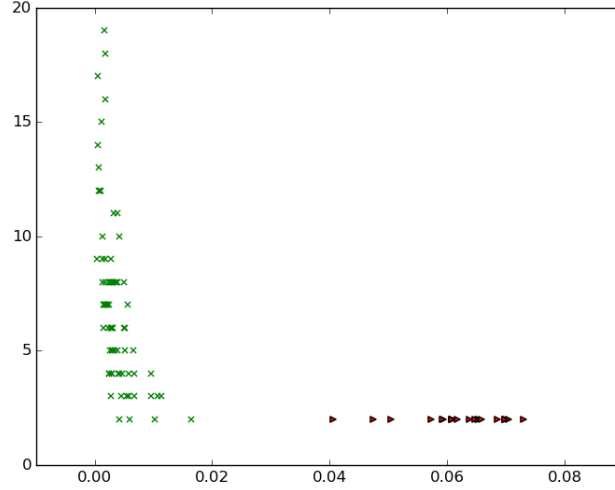


FIGURE 3.6 – Forme et signature polaire d'une drépanocyte



Drépanocytes en rouge

FIGURE 3.7 – Descripteurs SP pour 74 saines et 21 drépanocytes

3.2.3 Dacryocytes

Les dacryocytes ont une forme intermédiaire entre les deux autres types de cellules étudiés. Une partie de leur contour est semblable à un cercle, tandis que, diamétralement opposé, un pic s'éloigne du reste de la forme. Cette excroissance déplace légèrement le centre de gravité, ce qui a pour effet de gonfler la signature polaire pour la partie qui devrait être quasi-continue.

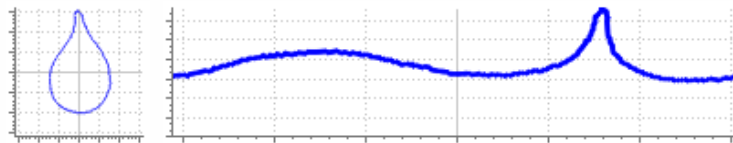
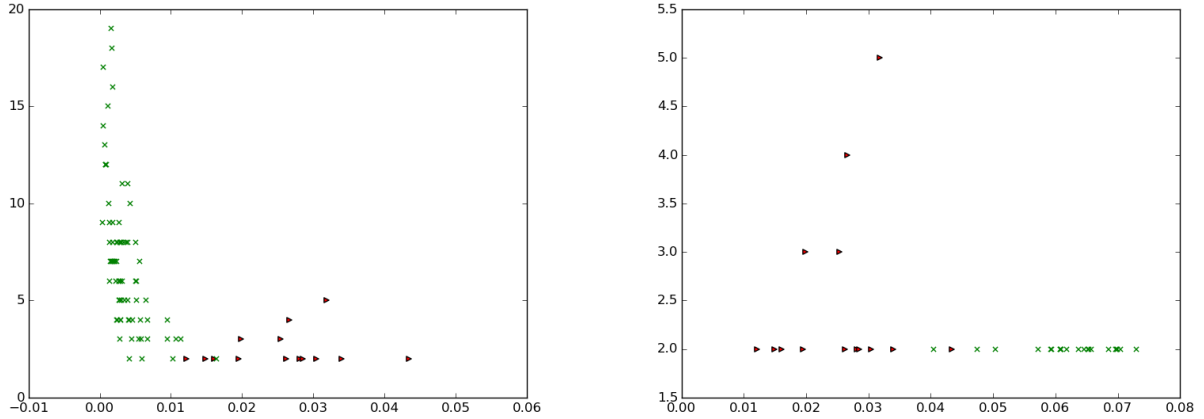


FIGURE 3.8 – Forme et signature polaire d'une dacryocyte

Ceci a pour conséquence que la signature présente généralement deux "pics" (groupes consécutifs de points au dessus de la médiane). La variance est quant à elle modérée et se situe entre les deux autres situations. Il arrive cependant, comme les variations ne sont pas très prononcées, que plus de deux pics soient détectés.



Gauche : descripteurs SP pour 74 saines et 15 dacryocytes
Droite : comparaison entre dacryocytes et drépanocytes

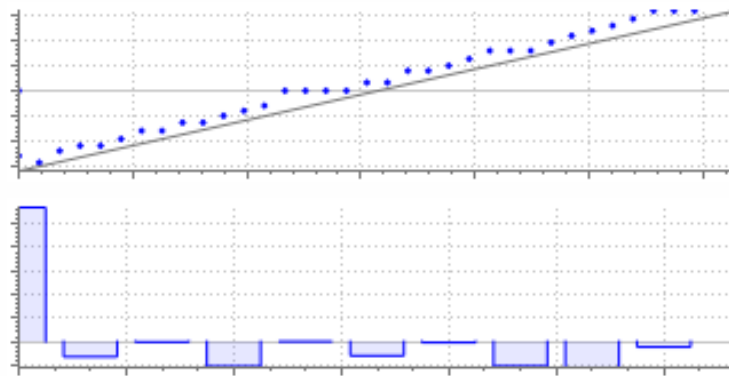
FIGURE 3.9 – Comparaison entre les différents nuages de points

3.3 Descripteurs de Fourier

Les descripteurs de Fourier ont un nombre de dimension fixé par l'utilisateur, via l'interface graphique. Suivant la forme à caractériser, on aura besoin d'harmoniques de plus ou moins haute fréquence.

3.3.1 Cellules Saines

Comme on l'a vu précédemment, les hématies saines ont un contour proche d'un cercle. La représentation spectrale va donc être concentrée sur la fondamentale, les autres harmoniques caractérisant les faibles variations autour du cercle.



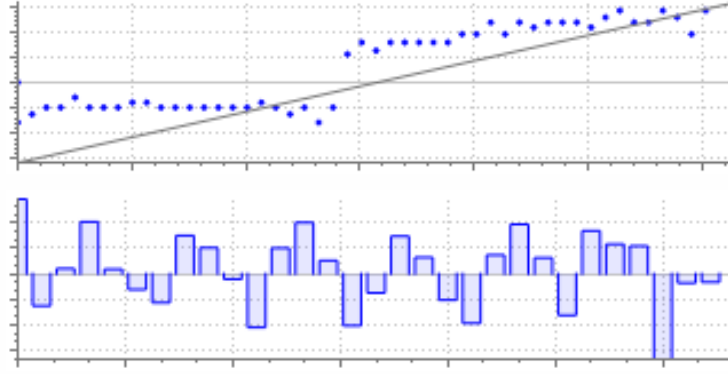
En haut : orientation de la tangente en fonction de l'azimut
En bas : transformée de Fourier discrète (10 harm.)

FIGURE 3.10 – Descripteurs de Fourier d'une hématie saine

Globalement, pour les cellules saines, la fondamentale est élevée (autour de 5) et les autres harmoniques, même jusqu'à la 30^e, sont faibles (entre -1 et 1).

3.3.2 Drépanocytes

Les cellules drépanocytaires ont une forme assez éloignée du cercle. Elles sont allongées et présentent souvent des extrémités pointues, ce qui grossit les hautes fréquences. L'étirement global de leur forme se traduit en terme de basse fréquence dans la variation de la tangente, comme une combinaison des premières harmoniques.



On a généralement des valeurs importantes pour h_0 , h_1 et h_3

FIGURE 3.11 – Descripteurs de Fourier d'une hématie drépanocytaire

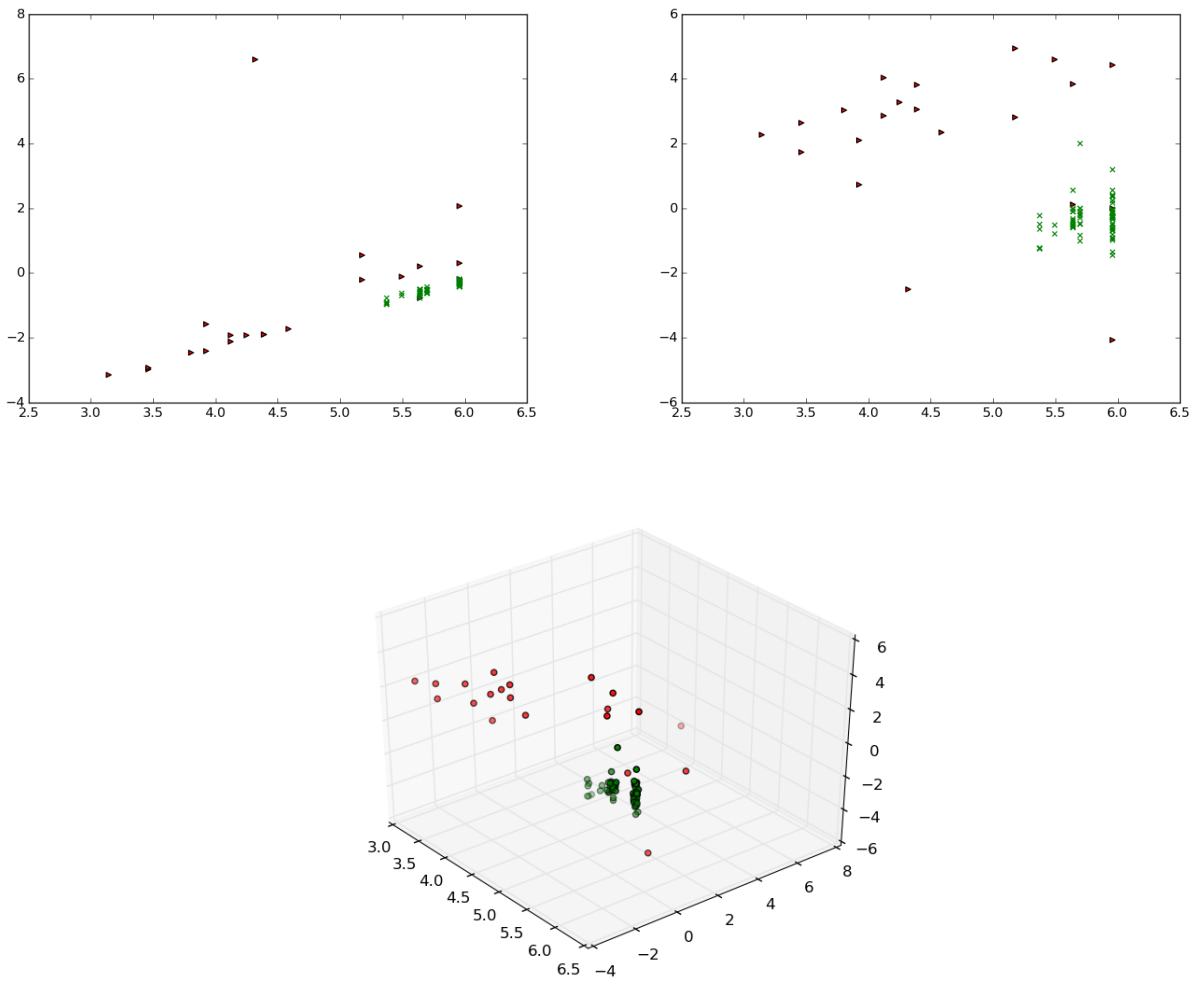
Si l'on souhaite utiliser peu de dimensions pour différencier des hématies saines de drépanocytes, on peut se concentrer sur les basses fréquences (moins sensibles aux variations) notamment les harmoniques h_0 , h_1 et h_3 .

Comme on le voit sur la Figure 3.12, les hématies saines sont très regroupées car très semblables. En revanche, le nuage de points correspondants aux cellules drépanocytaires est assez étalé, mais reste différenciable des cellules saines.

3.3.3 Dacryocytes

Les dacryocytes présentent généralement une partie du contour stable et proche du cercle ainsi qu'une excroissance qui perturbe les variations de la tangente. Cela se traduit dans le domaine fréquentiel par une forte fondamentale et quelques harmoniques médiums, mais peu de hautes fréquences car le contour est relativement lisse (voir Figure 3.13).

On a utilisé (voir Figure 3.14) les harmoniques h_2 , h_8 , et h_{18} pour différencier des dacryocytes de cellules saines. De cette façon, les hématies normales sont concentrées autour de l'origine, et les dacryocytes sont réparties dans l'espace à une distance minimum de cet origine.



Dans l'ordre : projections pour (h_0, h_1) , pour (h_0, h_3) , et (h_0, h_1, h_3)
 En vert : hématies saines

FIGURE 3.12 – Comparaisons entre hématies saines et drépanocytaires

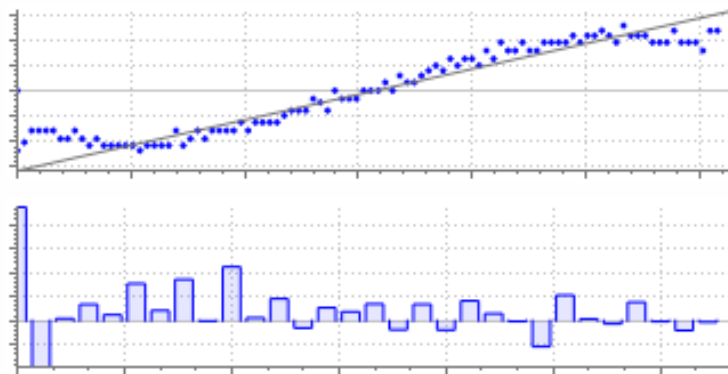
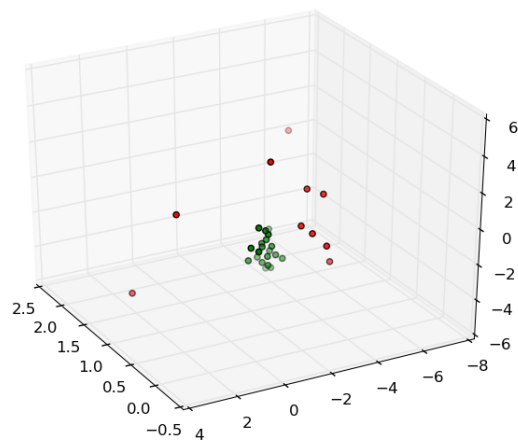


FIGURE 3.13 – Descripteurs de Fourier d'une dacryocyte



Cellules saines en vert (regroupées)

FIGURE 3.14 – Répartition des cellules saines et dacryocytes pour h_2 h_8 et h_{18}

4 Améliorations

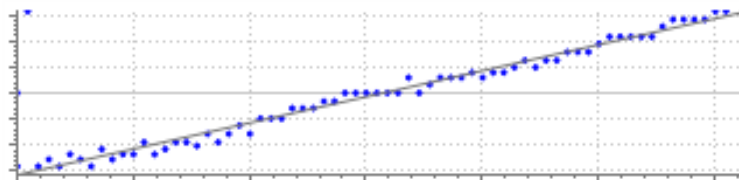
Nous regroupons ici les idées d'améliorations au vue de l'utilisation de l'interface et des résultats obtenus. Nous distinguerons les points relevant des fonctionnalités et de l'interface utilisateur de ceux portant sur les descripteurs et les résultats obtenus.

4.1 Interface et fonctionnalités d'extraction

Gestion des résultats Il est assez laborieux d'enregistrer les résultats obtenus sur les contours cellule par cellule, puis de les compacter à l'aide d'un script externe. Outre la multiplication des manipulations, la portabilité du programme est aussi réduite du fait d'utiliser Python pour une tâche potentiellement automatisable. Il serait judicieux d'intégrer à l'interface un système de gestion de classes, qui permettrait de classer à la volée les cellules que l'on sélectionne. Le fichier XML enregistré contiendrait alors toutes les données utiles.

Comptage des cellules On voit que les deux outils morphologiques proposés ne permettent pas toujours d'avoir un compte exact des cellules présentes, même lorsque celles-ci ne sont pas entassées. On pourrait ajouter au système existant la possibilité de sélectionner des zones de l'image à traiter (ce qui laisserait les autres intactes), ce qui permettrait de réaliser des érosions à certains endroits, sans que d'autres formes ne disparaissent.

Problème de gestion des angles (tangente) Pour calculer les descripteurs de Fourier, on a besoin de quantifier l'orientation de la tangente au contour, en fonction de l'azimut du point tangent. Cependant, comme l'angle est exprimé entre $-\pi$ et π , il y a un point de rupture qui peut être gênant.



Le deuxième point est placé à $y = \pi$, ce qui crée un saut

FIGURE 4.1 – Mauvaise gestion des angles

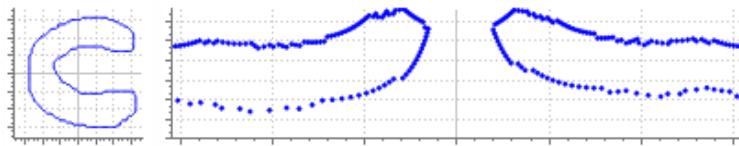
Ce problème influe évidemment sur la dérivée, donc sur les descripteurs de Fourier extraits. Il faut donc corriger cette mauvaise gestion des angles.

4.2 Descripteurs

Signature polaire

Le descripteur proposé permet de différencier très correctement les cellules drépanocytaires des cellules saines. On voit cependant que la différence est moins nette lorsque les magnitudes des points sont moins dispersées. La mesure des pics en fonction du nombre de groupes au-dessus de la médiane n'est pas très fiable.

On pourrait aussi ajouter une dimension pour quantifier la convexité (ou non-convexité) d'une forme. Ceci pourrait être utile pour la reconnaissance de caractères (Voir Figure 4.2).



Le C est une forme non-convexe

FIGURE 4.2 – Signature d'un C

Cette dimension pourrait par exemple être le nombre moyen de points partageant le même azimut (ou en dessus d'une distance seuil), mais une magnitude éloignée.

Descripteurs de Fourier

Les descripteurs de Fourier sont de grande dimension, et on peut difficilement prendre arbitrairement telle ou telle harmonique pour discriminer un échantillon. Nous n'avons pas réalisé de classification en utilisant des descripteurs de grande dimension, ce qui ne nous permet pas d'émettre un jugement.

En revanche, pour des classifications simple, il peut être intéressant de proposer des descripteurs de petite dimension résumant la répartition spectrale d'un descripteur de Fourier. Un exemple simple est de regrouper les fréquences en bandes, et de caractériser chaque bande (basses, moyennes et hautes par exemple) par :

- Une moyenne
- La variance
- L'étendue

Dans notre cas, ce sont ces caractéristiques qui distinguent bien les formes. En effet, lorsque les hautes fréquences sont sollicitées par exemple, les harmoniques élevées sont souvent isolées . Le numéro de ces harmonique peut varier d'une cellule à l'autre, au sein de la même classe. Il est alors plus intéressant de les regrouper en bandes de fréquences.

Bibliographie

- [1] M. Kunt. *Reconnaissance des formes et analyse de scènes*, volume 3. 2000.
- [2] N. Wirth. *Algorithms + data structures = programs*. Prentice-Hall Series in Automatic Computation. 1976.
- [3] N. Devillard. Fast median search : an ANSI C implementation [consulté : mai 2016]. <http://n-devilla.free.fr/median/median.pdf>.
- [4] R. Roskies C. Zahn. Fourier descriptors for plane curves. *IEEE*, C-21(3) :269–281, Mars 1972.
- [5] A. Bernadou. Cytologie sanguine - sang normal : normes [consulté : mai 2016]. <http://www.cytologie-sanguine.com/html/sangnorme01.php>.