

Rapport de stage de DEA sur

**LA REALISATION D'UNE ETUDE SUR  
LE TRI INFORMATIQUE EN CYTOLOGIE**

en vue de l'obtention du

DEA d'Intelligence Artificielle et Algorithmique

Option Image

de l'université de Caen.

Présenté par

**LEZORAY OLIVIER**



Le Vendredi 22 Septembre 1995

Maître de stage : M. Cardot Hubert  
Examineurs : Mme Marinette Revenu  
M. Jean-Luc Lambert

---

## Remerciements

---

Le travail décrit dans ce mémoire a été réalisé au sein de l'équipe de traitement d'image du GREYC de l'ISMRA de Caen en collaboration avec le Centre Hospitalier Louis Pasteur de Cherbourg (C.H.L.P).

Je voudrais remercier en premier lieu M. Hubert CARDOT, maître de conférences au GREYC, pour m'avoir accepté pour ce stage et pour m'avoir (tâche difficile) trouvé un sujet intéressant. Sa motivation constante et ses conseils m'auront été très précieux.

Je remercie le Docteur Hubert ELIE, les cyto-techniciens M. Michel LECLUSE et Mme Annette ALEXANDRE du C.H.L.P. D'une part pour avoir accepté de participer activement à ce travail de recherche, ce qui représente un travail assez important, et d'autre part pour leur grande patience à m'expliquer les mystères de la cytologie. Leur participation dans ce stage s'est faite à tous les niveaux et leurs remarques ont permis de diriger mon travail. Je remercie également tout le personnel du service anapath pour leur sympathie.

Je remercie Mme Marinette REVENU, professeur à l'ISMRA, pour m'avoir accueilli au sein de son équipe de recherche.

Je remercie M. Jean-Luc LAMBERT d'avoir accepté d'être présent à ma soutenance.

Je remercie également vivement M. Abderrahim ELMOATAZ, maître de conférences au GREYC, et M. Régis CLOUARD, maître de conférences à Alençon (courage), pour m'avoir grandement aidé dans le domaine du traitement d'images.

Je remercie tous les autres membres du GREYC aussi bien à l'ISMRA qu'à l'université.

Enfin le meilleur pour la fin. Je tiens à remercier tout particulièrement Eric CROCHEMORE pour m'avoir aidé dans ma classification ainsi que Stéphane DURAND pour m'avoir fait découvrir les vertus apaisantes de la pause café. Je remercie également tous mes camarades et amis du DEA (DUPART Mickaël, LEPLÉ Franck, FOURNIER Dominique, BENOIST Emmanuel) pour m'avoir motivé dans la recherche du style parfait au baby-foot.

Je pense aussi à Natacha ...

---

## Introduction

---

«

*Tout homme tend à devenir machine, habitude, méthode, maîtrise, enfin cela veut dire machine ...*

*En vérité c'est la machine qui tend à devenir animal... »*

Paul Valéry, Cahier, La Pléiade, p885, 1975

Durant ces dernières années, notre compréhension de l'apparition et de l'évolution des populations dans les systèmes macroscopiques a fait d'importants progrès grâce au développement de nouvelles techniques informatiques et mathématiques.

Dans le domaine qui nous intéresse : la cytologie pathologique, le système (un liquide) qui fait l'objet de l'étude peut se décomposer en parties (des cellules). Les techniques associant des connaissances d'horizons divers ( informatique, mathématique et biologique) permettent d'ouvrir une voie inédite à l'étude de ces cellules.

L'informatique s'affirme à présent comme une nouvelle technologie médicale majeure. Son but n'est pas de remplacer le travail des experts mais de le faciliter. Dans ce rapport nous allons faire une étude sur un système permettant de faciliter le travail d'un expert grâce au tri cellulaire informatisé.

Un expert en cytologie pathologique étudie une lame contenant des cellules. Il reconnaît les différents types de cellules présents sur cette lame et son diagnostic peut ensuite être émis. Mais cet examen visuel et manuel de cellules en suspension est une tâche très fastidieuse et une analyse des lames par un système informatique permettrait d'automatiser une partie du travail. Ce type de système sert donc en vue d'une aide au diagnostic : on ne remplace pas le travail de l'expert en cyto-pathologie mais on le facilite.

Un tel système informatique fait appel à des techniques de traitement d'images, d'intelligence artificielle et évidemment biologiques. Dans cette étude nous disposons d'images de cellules de cytologie qui seront étudiées par des techniques de traitement d'images afin d'extraire des informations qualitatives et quantitatives puis qui seront exploitées par des techniques d'intelligence artificielle chargées de les interpréter.

Nous présenterons le système que nous avons mis en oeuvre : son but est de trier les cellules qui sont présentes sur une image. Nous étudierons les systèmes existants cités dans la littérature et nous dégagerons les spécificités de notre futur système.

Ensuite nous concevrons une maquette de notre système. Nous verrons en détail chaque étape :

- ✓la segmentation qui consiste à extraire des objets (des cellules) de nos images,
- ✓la caractérisation qui consiste à extraire des informations quantitatives sur les objets extraits,
- ✓la classification qui consiste à classer nos cellules suivant leur type grâce aux informations obtenues à la caractérisation.

Nous détaillerons pour chaque étape, les techniques mises en oeuvre et nous examinerons les résultats pour chacune d'entre elles. Pour finir nous conclurons sur l'exactitude des méthodes mises en oeuvre et nous verrons ce que cette étude nous a permis de connaître.

---

## Plan de lecture

---

Ce rapport est composé de 2 parties contenant 6 chapitres.

La partie A « *Situation du problème* » présente dans un premier chapitre le domaine de la cytologie pathologique ainsi que les cellules présentes dans les images dont nous disposerons. Dans un deuxième chapitre, nous faisons un état de l'art des systèmes informatiques existants ainsi que des méthodes mises en oeuvre dans ces systèmes. Un troisième chapitre situe le stage par rapport aux études déjà effectuées et dégage les spécificités de notre système.

La partie B « *Le système mis en place* » présente l'architecture du système que nous avons développé et détaille pour chacune des trois étapes (segmentation, caractérisation, classification) les moyens et méthodes utilisés. Les résultats obtenus à chaque étape sont examinés et des conclusions en sont tirées.

Nous concluons enfin par le bilan de cette étude et les perspectives qu'une telle étude nous permet d'envisager.

---

## **SOMMAIRE**

---

1. Généralités d'anatomie et de cytologie pathologiques .....	10
1.1 Introduction .....	10
1.2 Quelques définitions.....	10
1.2.1 Structure de l'organisme .....	10
1.3 La cytologie et l'histologie.....	11
1.4 Les prélèvements en Anatomie et Cytologie Pathologiques (A.C.P).....	12
1.5 Les différentes colorations .....	13
1.6 Les types de cellules et leurs caractéristiques .....	13
1.7 Le cancer et les traitements .....	15
2. Apport des systèmes informatiques en cytologie clinique .....	16
2.1 Historique de la cytologie clinique.....	16
2.2 Pourquoi un système informatique ?.....	17
2.3 Différents types de systèmes informatiques .....	17
2.4 Historique des systèmes informatiques d'analyse d'images.....	17
2.4.1 Les Systèmes Informatiques d'Analyse d'Images cytologiques .....	20
2.4.1.1 La segmentation.....	20
2.4.1.2 La caractérisation des cellules .....	21
2.4.1.3 La classification des cellules .....	22
3. Position du stage .....	25
3.1 Les objectifs .....	25
3.2 Travaux déjà réalisés sur le sujet et comparaison avec le stage .....	25
3.3 Conclusion.....	25

1. La Segmentation.....	28
1.1 Le but de la segmentation.....	28
1.2 Etude de l'image .....	28
1.3 La stratégie de segmentation .....	29
1.3.1 Séparation cytoplasme, noyau et fond.....	29
1.3.2 Méthodes utilisées dans la littérature .....	29
1.3.3 La méthode choisie.....	30
1.3.3.1 Le seuillage.....	30
1.3.3.2 La ligne de Partage des Eaux ou L.P.E.....	31
1.3.3.2.1 Principe de la L.P.E .....	31
1.3.3.2.2 La Lpe contrainte par des marqueurs .....	33
1.3.3.2.3 Application à notre cas .....	34
1.3.3.3 Séparer les frontières cytoplasmes et noyaux .....	36
1.3.3.3.1 Première méthode : Suivi de contours .....	36
1.3.3.3.2 Deuxième méthode : Deux Lpe .....	40
1.3.3.3.3 Comparaison des deux méthodes.....	41
1.4 La séparation des amas.....	41
1.4.1 Quelques notions de base .....	41
1.4.2 La méthode de séparation des amas .....	41
1.5 Conclusion.....	44

2. La caractérisation .....	45
3. La classification .....	46
3.1 L'étiquetage des images.....	46
3.2 Résultats de l'étiquetage .....	49
3.3 Les réseaux neuronaux .....	49
3.3.1 Brefs rappels sur les réseaux neuronaux .....	49
3.2.2 Application à notre cas .....	50
3.2.3 Les résultats .....	51
3.2.4 Conclusion.....	53

Bilan et perspectives
-----------------------

Références
------------



---

## **Partie A : Situation du problème**

---

---

# Partie A

---

## 1. Généralités d'anatomie et de cytologie pathologiques

### 1.1 Introduction

Grâce aux progrès réalisés ces dernières années, on dispose à présent de techniques (telles que la fibre optique) qui permettent l'exploration d'une grande partie du corps. Les prélèvements (biopsies ou aspirations) obtenus grâce à de telles techniques sont ensuite examinés dans des laboratoires d'anatomie et de cytologie pathologiques (A.C.P) privés ou publics.

Avec le nombre croissant de lames de cellules à analyser (en particulier les frottis), le dépistage s'est considérablement développé ces dernières années mettant en avant le besoin d'automatisation par des méthodes informatiques.

### 1.2 Quelques définitions

Afin de comprendre le système complexe dont fait partie la cellule, nous allons préciser quelques termes médicaux.

La pathologie est l'étude et la science des maladies. La traduction littérale de pathologie en Grec (pathos, logos) est "les mots de la souffrance". Le pathologiste étudie les causes et les conséquences de la maladie, comment la maladie affecte certains tissus ou cellules, la progression de la maladie dans le corps humain, comment la maladie se manifeste et enfin les méthodes pour surveiller la progression de la maladie. La pathologie donne un fondement scientifique pour la médecine clinique et sert de lien entre les sciences de base et les soins au patient.

L'anatomie pathologique est l'étude morphologique des lésions qui sont la cause ou la conséquence des maladies. Cette étude peut se faire :

- à l'oeil nu : c'est l'anatomie pathologique macroscopique,
- au microscope :
  - c'est l'histologie pathologique où l'on examine des coupes de tissus colorées,
  - c'est la cytologie pathologique où l'on examine un étalement de cellules isolées.

Le domaine que nous allons étudier, la cytologie, fait partie intégrante de la pathologie.

#### 1.2.1 Structure de l'organisme

L'organisme possède quatre niveaux structuraux : les cellules, les tissus, les organes et les systèmes.

A son niveau structural le plus simple, l'organisme est fait de cellules. La cellule (Figure I.1) est l'unité fonctionnelle la plus petite de l'organisme. Les cellules peuvent être différentes entre elles selon leur fonction.

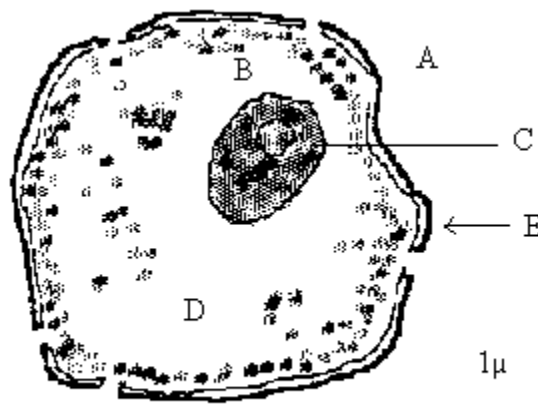


Figure I.1 Représentation schématique du noyau [Menard 93]

- A - Enveloppe nucléaire
- B - Nucléole
- C - Chromatine
- D - Cytoplasme
- E - Enveloppe cytoplasmique

Chez la majorité des animaux, y compris chez l'homme, des groupes de cellules semblables s'associent pour former des tissus. On peut définir un tissu comme groupement de cellules qui possèdent fondamentalement la même structure et qui assurent un même ensemble de fonctions simples. On distingue ainsi [ENSTAC 87] :

- ① Les tissus épithéliaux : ils recouvrent la surface des organes,
- ② Les tissus conjonctifs : ils servent aussi bien de soutien (os, cartilage) que d'attache à d'autres tissus ( tendons, ligaments ...),
- ③ Les tissus musculaires composés de cellules spécialisées capables de se contracter,
- ④ Les tissus nerveux : l'encéphale, la moelle épinière et les nerfs par exemple sont formés de tissus nerveux. Ces cellules sont caractérisées par de longs prolongements,
- ⑤ Le tissu hématopoïétique (ou tissu sanguin ).

Nous nous limiterons à l'étude des tissus épithéliaux. Ils sont caractérisés par une étroite juxtaposition des éléments cellulaires et ils reposent toujours sur une membrane basale : couche de collagène qui sépare le chorion de l'épithélium ( figure I.2).

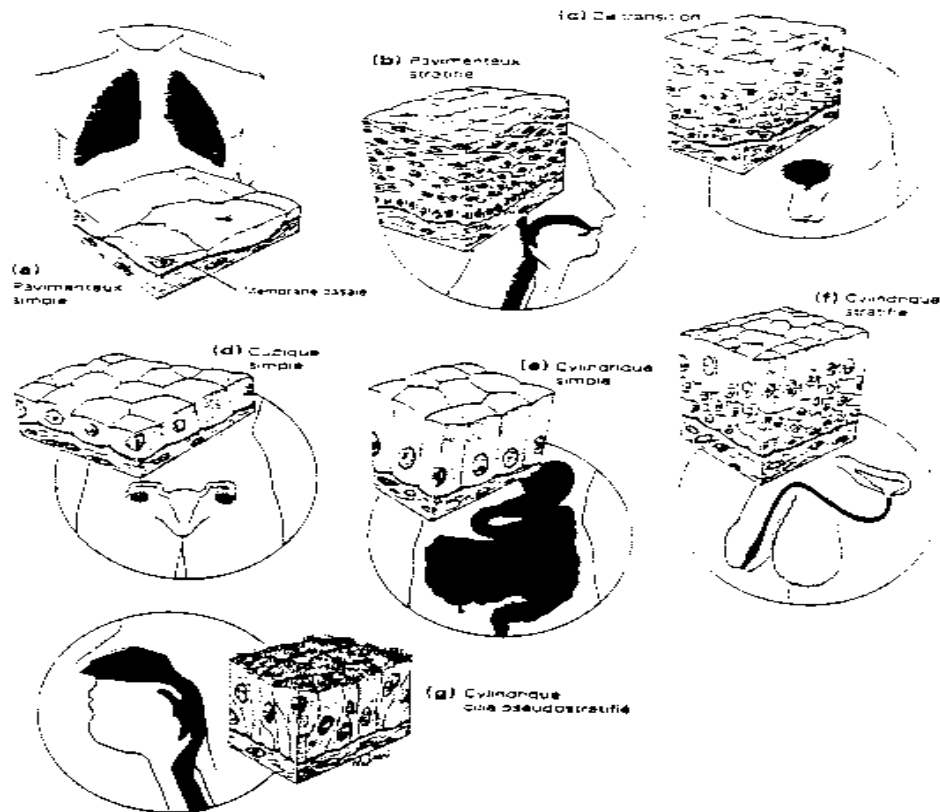


Figure I.3 [Menard 93]  
Différents types d'épithéliums suivant les organes.

Nous pouvons définir les organes comme l'assemblage de plusieurs tissus permettant d'accomplir des fonctions plus complexes. Par exemple, l'estomac est tapissé de tissus épithéliaux, il est entouré de tissus musculaires et innervé par du tissu nerveux. Des tissus conjonctifs les maintiennent ensemble.

On entend enfin par système l'ensemble des organes qui accomplissent une tâche commune.

### 1.3 La cytologie et l'histologie

En anatomie pathologique on distingue deux types d'examen, l'examen histologique et l'examen cytologique. On peut avoir recours aux deux pour donner un diagnostic, mais parfois un seul des deux examens peut être déterminant.

L'histologie est l'observation de la coupe microscopique d'un tissu ( histo signifie tissu en grec ).

En cytologie on examine un étalement de cellules isolées provenant de prélèvements contrairement à l'histologie où l'on examine des coupes de tissus colorés.

La cytologie, science de l'observation des cellules ( cyto signifie cellule en grec ), a connu un développement considérable grâce à des techniques d'observation de plus en plus sophistiquées. Ces techniques permettent d'obtenir une multitude d'images : microscopie photonique, à contraste de phase, électronique, immunocytochimie,...

Les images que nous aurons à analyser sont des images de cytologie.

La cytologie a également évolué grâce à l'apport des domaines complémentaires comme

- La cytochimie qui, en mettant au point des colorants de plus en plus sélectifs, a permis de mettre en évidence certains constituants de la cellule, par exemple le P.A.S ( révélateur des glyco-amino-glycanes ).
- L'immunocytochimie qui permet à partir de marqueurs immunologiques de reconnaître certains sites cellulaires, par exemple l'A.C.E ( Antigène Carcino Embryonnaire ),

## 1.4 Les prélèvements en Anatomie et Cytologie Pathologiques (A.C.P)

Un prélèvement nécessite un personnel qualifié et un matériel adapté. C'est un acte médical et il ne doit pas être traumatisant pour les cellules.

On peut distinguer plusieurs types de prélèvements :

- Les biopsies : c'est un examen pratiqué sur un être vivant. En fait, on réserve ce terme aux petits prélèvements faits dans le but d'obtenir un diagnostic,
- Les exérèses : ce terme concerne les prélèvements de pièces plus importantes quand on a cherché à retirer la totalité de la lésion,
- Les autopsies : elles servent à déterminer les causes de la mort et les pathologies associées.
- Les prélèvements cytologiques :
  - Les liquides d'épanchement des séreuses ( figure I.4),
  - Les urines,
  - Les liquides céphalo-rachidiens,
  - Les expectorations, aspirations bronchiques, brossages,
  - Les frottis génitaux,
  - Les cyto-ponctions,
  - etc.

Les images que nous aurons à analyser proviennent de liquides d'épanchement des séreuses.

Après le prélèvement on observe une autolyse des cellules ( dégénérescence ) et pour les conserver, on procède à une fixation. Les fixateurs sont de deux sortes : les fixateurs chimiques ( Alcool-Ether, Spray,etc...) ou physiques ( séchage à l'air ). La fixation utilisée dans le cadre de notre étude reste dans le domaine d'une fixation standard ( Alcool-Acétone )

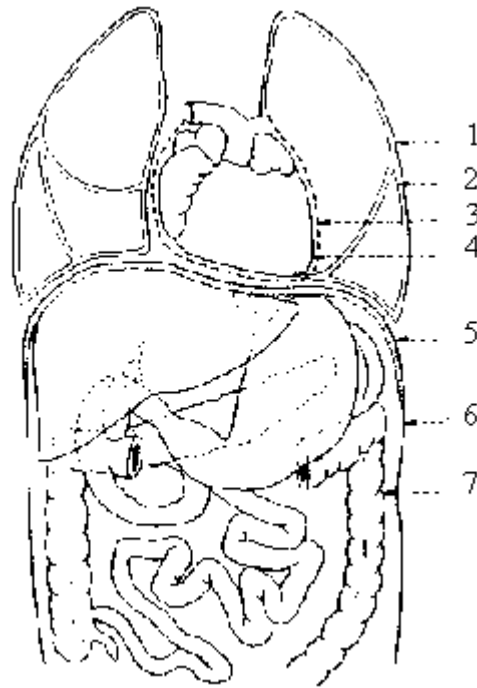


Schéma des cavités séreuses du corps humain Les cavités pleurales, péricardique et péritonéale, sont comprises entre un feuillet pariétal et un feuillet viscéral.

1. plèvre pariétale 2. plèvre viscérale 3. péricarde 4. épicarde  
5. diaphragme 6. péritoine pariétal 7. péritoine viscéral

Figure I.4 [Takahashi 83]

## 1.5 Les différentes colorations

Il n'existe pas de technique de coloration polyvalente. Selon le but à atteindre, en l'occurrence selon la nature chimique des différents constituants cellulaires à mettre en évidence, on fera appel à des techniques différentes.

Dans un grand nombre de cas les méthodes de coloration de topographie générale sont suffisantes. Parfois cependant des techniques histochimiques sont indispensables pour préciser le diagnostic. La coloration ne peut être réussie que si elle est précédée d'une fixation appropriée visant à préserver l'intégrité des structures à colorer.

En cytologie les fixateurs essentiellement utilisés sont le mélange alcool-éther ou des laques fixantes, ces deux méthodes permettant ensuite une coloration standard. D'autres étalements seront séchés en réserve d'effectuer des colorations particulières telles que le May-Grunwald-Giemsa, le Bleu Alcian, le Perls et le P.A.S.

Les principales colorations cytologiques sont :

- La coloration de Papanicolaou,
- La coloration de Haris-Shorr,
- Le May-Grunwald-Giemsa,
- La coloration de Perls,
- Le P.A.S,
- Le Bleu-Alcian,
- etc.

Le standard international est la coloration de Papanicolaou.

Les images que nous étudierons par la suite ont été colorées par la méthode de Papanicolaou ( préparation internationale ). Les résultats obtenus sont les suivants :

Noyaux cellulaires : bleu violet par hématoxyline de Harris,  
Cytoplasmes basophiles : vert ou bleu vert par EA50,  
Cytoplasmes acidophiles : rose à orangé.parOG6.

## 1.6 Les types de cellules et leurs caractéristiques

Nous avons vu précédemment les différents types de prélèvements possibles, mais il nous reste à voir quels types de cellules nous pouvons rencontrer dans les images de ces prélèvements ( qui sont des épanchements des séreuses ).

Dans les liquides d'épanchement on trouve les composants cellulaires suivants :

Les cellules mésothéliales : elles composent le mésothélium,

Les macrophages ou hystiocytes		Cellules de défense de l'organisme
Les lymphocytes		
Les polynucléaires		
Les cellules néoplasiques		

Les cellules ne sont jamais observées à l'état normal car à partir du moment où il y a épanchement, il y a anormalité.

Voici décrites brièvement les principales caractéristiques de noyau et de cytoplasme pour ces cellules [Takahashi 83, Jaurand 90, Naylor 91] :

Cellules normales :

	cellules Mésothéliales	Macrophages (hystiocytes)	Lymphocytes	Polynucléaires
Forme du noyau	Rond ou ovale	réniforme	Rond	Polylobés
Situation du noyau	Central ou paracentral	Périphérique	Central	Périphérique
Aspect de la chromatine	Finement granuleuse avec des caryosomes	finement granuleuse et régulièrement répartie	dense non mottée	non mottée
Nucléole	Net	Non visible	Aucun	Aucun
Cytoplasme	Rond	Abondant	Pratiquement pas	Réduit
Bordure nucléaire	Nette	nette	Presque confondue avec la bordure cellulaire	nette
Bordure cellulaire	Assez bien définie	Mal définie	Très nette	Peu visible
Groupement des cellules	En rosettes, en placards ou isolées	Isolées	Isolées mais pouvant se toucher	Isolées
Taille	Grande	Moyenne	Petite	Petite

Cellules anormales ( néoplasiques ou autres )

Noyau irrégulier, boursoufflé,  
Augmentation de la densité optique ( répartition irrégulière de la chromatine),  
Augmentation du nombre de nucléoles,  
Augmentation du rapport Noyau / Cytoplasme,  
Augmentation des mitoses avec anomalie globale.

Le diagnostic d'anormalité d'une cellule repose sur les critères précédent mais une cellule peut avoir certaines des altérations citées ci-dessus sans être malignes.

Une augmentation du taux d'un type de cellule dans un prélèvement peut être significatif :

Augmentation des cellules mésothéliales : hyperplasie mésothéliale (inflammation),  
Augmentation des polynucléaires : inflammation aiguë (pleurésie),  
Augmentation des lymphocytes : inflammation chronique (tuberculose),  
Présence de cellules néoplasiques : cancer dont l'origine peut être primitive ( détecté à l'endroit où il se trouve) ou bien secondaire ( une métastase : les cellules sont à distance de la lésion initiale ).

Dans notre étude, on se propose de faire du tri cellulaire. Il est particulièrement indiqué en A.C.P car lors du screening d'une lame, le cyto-pathologiste procède à un comptage cellulaire lui permettant d'établir la répartition. Les pourcentages sont actuellement faits sur environ 200 cellules. L'analyse d'image permettrait une meilleure précision car elle serait effectuée sur plus de cellules.

Le deuxième avantage du tri cellulaire informatique serait de réduire le taux de faux-négatifs car cela permettrait la détection de cellules néoplasiques que le cyto-pathologiste aurait pu ne pas détecter.

Voici quelques exemples de cellules :

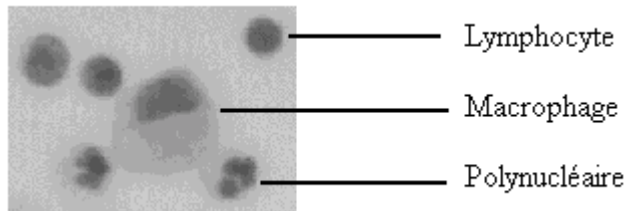


Figure I.5

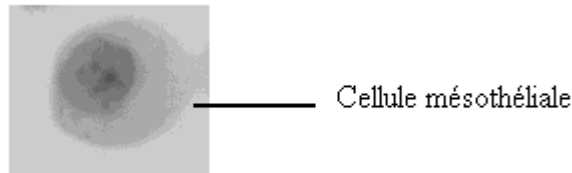
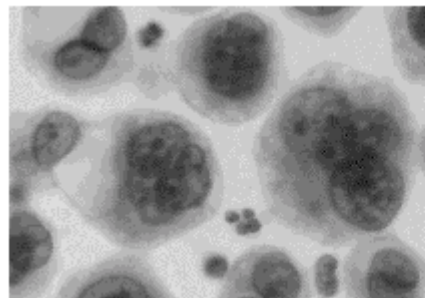
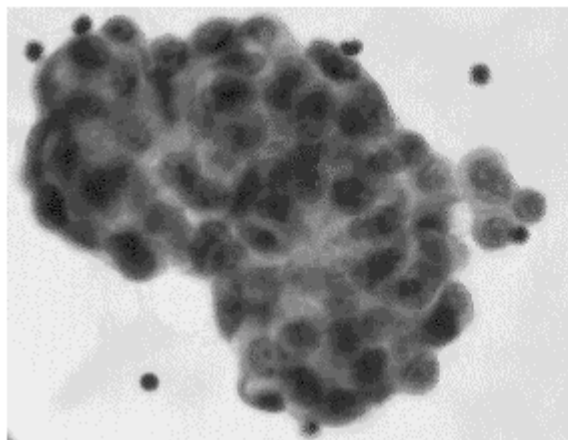


Figure I.6



Plusieurs cellules  
néoplasiques

Figure I.7



Papille de cellules  
néoplasiques

Figure I.8

## 1.7 Le cancer et les traitements

Du point de vue cytologique, une cellule cancéreuse se caractérise par la perte des mécanismes qui collaborent au développement des cellules. Le cancer se présente alors essentiellement comme une division non contrôlée des cellules. On observe alors la naissance de cellules anormales appelées cellules malignes. Parfois si les tumeurs sont découvertes à temps, par un dépistage précoce, elles peuvent être enlevées chirurgicalement. On voit encore une fois l'utilité du tri cellulaire informatique car il permettrait d'améliorer la qualité du dépistage précoce et préventif.

Les cellules cancéreuses peuvent se propager à travers tout l'organisme et se loger dans d'autres organes ou tissus : ce sont les métastases.

Les traitements possibles sont : la chirurgie, la radiothérapie, la chimiothérapie, l'association de plusieurs médicaments ainsi que des traitements combinés, etc.[Ruffé 90]

## 2. Apport des systèmes informatiques en cytologie clinique

### 2.1 Historique de la cytologie clinique

C'est en 1838 que Mueller [Mueller 1838] a été le premier à décrire et illustrer des cellules malignes de tumeurs observées au microscope. Les premiers atlas de cytologie sont apparus réalisés par des microscopistes français, Donné [Donné 1845] et Lebert [Lebert 1845] suivis par un atlas de cytologie sur le cycle menstruel réalisé par le français Pouchet [Pouchet 1847]. Des rapports sur les tumeurs en cytologie ont été réalisés ensuite par Beale [Beale 1860] sur des crachats, par Sanders [Sanders 1864] sur de l'urine, par Dufour [Dufour 04] sur des fluides cérébraux, par Marissi [Marissi 09] sur des spécimens gastriques.

En 1928 sortent simultanément deux publications de Babes [Babes 28] à Bucarest en Roumanie et de Papanicolaou [Papanicolaou 28] à New York sur des travaux de cytologie sur le cancer du col de l'utérus. Mais ces publications bien qu'importantes n'ont pas eu un impact clinique aussi important que les résultats publiés par Papanicolaou [Papanicolaou 47,54] dans les années 1940 sur la détection du cancer utérin. Papanicolaou avait créé une classification sur des lames mais elle n'était pas perçue comme un outil de diagnostic et plusieurs pathologistes refusaient d'accepter ce genre de technique de classification au sein de leur laboratoire.

Aujourd'hui, plusieurs décades après la publication des atlas de Papanicolaou et des méthodes d'estimation sur la valeur d'un examen cytologique, on peut dire sans hésiter que les travaux de Papanicolaou plus connus sous le nom de "Pap smears" ont été le plus grand outil publique qui ait jamais été découvert. Ils ont surtout permis de reconnaître la cytologie comme un outil médical à part entière.

Les préoccupations principales en cytologie sont celles qui permettent d'augmenter le potentiel de vérité contenue dans le diagnostic. Les cytologistes s'attachent à trouver des méthodes pour réaliser le moins possible de faux-positifs ( une cellule classée pré-tumorale ou tumorale alors qu'elle ne l'est pas) mais surtout de faux-négatifs ( une cellule classée bénigne alors qu'elle est tumorale). Des taux de faux-négatifs ont été donnés dans la littérature (certains travaux [Koss 93] font état de 50% pour la cytologie du col de l'utérus), mais une comparaison entre tous les rapports publiés est difficile, étant donnée la variété des différentes méthodes de préparation des lames [Koss 89].

Les causes principales d'erreurs en cytologie sont parmi celles-ci :

- les prélèvements non significatifs ( mal faits, mal fixés, acellulaires),
- les erreurs de mauvaise interprétation,
- les erreurs d'omission par diminution de la vigilance,
- enfin et surtout le grand nombre de lames à analyser dans la journée.

Un cyto-pathologiste en voit entre 30 et 200. Sa culture médicale est remise en cause chaque jour. Mais la monotonie de ce travail et les facteurs de fatigue expliquent les déficiences possibles.

Certaines mesures ont été envisagées pour effectuer une revue de 10% des lames données négatives : cela a été prescrit par le gouvernement américain par arrêté fédéral en réponse à un contrôle de qualité dans les laboratoires cytopathologiques [Revision 67].

Mais il n'est pas apparu que ce genre de contrôle des lames élimine des erreurs dans le screening d'une lame [Koss 93] : d'autres mesures sont nécessaires.



## 2.2 Pourquoi un système informatique ?

Une informatisation des opérations de routine des laboratoires paraît inévitable et un laboratoire peut difficilement travailler sans système informatique lui permettant l'entrée, la sauvegarde et la consultation d'informations sur les lames traitées. Les laboratoires de cyto-pathologie commencent à disposer maintenant de systèmes leur permettant de garder des informations sur chaque lame traitée. Un système informatique peut réellement apporter un plus au cyto-pathologiste dans le domaine de l'assurance de la qualité. En effet, l'examen des lames est une tâche subjective dépendant du savoir-faire et de la compétence du technicien aussi bien que du temps consacré au screening d'une lame.

Ainsi cette tâche n'est pas infaillible et le système informatique peut permettre de donner une analyse objective de la lame.

## 2.3 Différents types de systèmes informatiques

On distingue plusieurs types de systèmes permettant l'aide au diagnostic : les systèmes utilisant l'analyse d'image de cellules, les systèmes fondés sur des techniques de colorations particulières ( marqueurs fluorescents par exemple ) et la cytométrie de flux. Cette dernière n'a pas donné de résultats concluants et les cyto-pathologistes se sont mis d'accord sur le fait que si un système informatique de screening automatique et d'aide au diagnostic efficace doit voir le jour, ce sera un système d'analyse d'images de cellules. A ce titre l'Académie Internationale de Cytologie [IAC 84] a émis les spécifications qu'un tel système doit vérifier.

## 2.4 Historique des systèmes informatiques d'analyse d'images

Nous n'allons nous intéresser qu'à un seul type de système informatique, celui d'analyse d'images. Comme nous l'avons vu précédemment, seul ce type de système est prometteur de résultats appréciables et performants.

Historiquement la première tentative d'automatisation par analyse d'images a été tentée par Tolles avec le cytoanalyser [Tolles 55] mais ce fut un échec. Quelques années plus tard apparut le CYDAC ( Cytometric Data Conversion ) de Prewitt J.M.S et Mendelsohn [Prewitt 66] pour la classification de leucocytes. Suivit ensuite par le TICAS en 1968 de Wied G.L [Wied 68] pour l'analyse des cellules dans des frottis vaginaux.

On remarquera que les travaux de recherche portent principalement sur les frottis car c'est ce type d'examen qui fournit le plus de lames en cytologie ( environ 80%).

Mis à part le fait que le but important dans ces projets de recherche était la conception et la construction du système, ces derniers se sont surtout attachés à l'établissement de standards de caractéristiques cellulaires pour le diagnostic. Ces débuts d'informatisation et d'automatisation par analyse d'image pour les cellules de frottis vaginaux ont été suivis par des essais pour tenter d'automatiser complètement le screening des lames de frottis. L'idée était d'automatiser un travail assez monotone par une machine qui non seulement serait plus fiable que le cyto-pathologiste mais qui donnerait également des résultats valables en un temps rapide et de façon économique.

Un des seuls systèmes ayant pris en compte réellement ce genre de contraintes est le système CYBEST (Japon) [Tanaka 77,87] dont la construction a commencé en 1967 et qui a continué jusqu'en 1986. Ce système regroupait toutes les technologies nécessaires à ce genre d'investigations : micro-ordinateur, scanner vidéo, observation optique, auto changeur de

lames. Le premier prototype apparut en 1972 et était dédié à la détection du cancer du col de l'utérus. Ses auteurs s'étaient fixés les contraintes suivantes :

- Une coloration Papanicolaou pour permettre une évaluation directe des spécimens au microscope,
- Un système d'analyse d'image approprié pour ce type de travail automatisé,
- Un taux de faux-positifs ne dépassant pas 20% et un taux de faux-négatifs inférieur au taux manuel du cytopathologiste,
- Rapidité de l'opération (inférieure à trois minutes ),
- Logique permettant le diagnostic flexible afin de permettre facilement l'addition de paramètres et de corrections,
- Pré traitement automatique des lames pour une dispersion et pour une bonne coloration des cellules.

Bien que prometteur, le fait d'avoir voulu tout intégrer dans la même machine la rendait trop lente à cause de la préparation de la lame. Mais c'est surtout l'absence de coopération entre le système et l'opérateur humain pour la donnée du diagnostic qui est la raison majeure d'échec des systèmes entièrement automatiques. Car dans ces systèmes la séparation entre spécimens normaux et anormaux n'était pas suffisamment efficace et reproductible. La plupart des cellules détectées comme anormales étaient dues à des artefacts. Parmi les systèmes informatiques automatiques qui ont été développés, on peut également citer : BioPEPR (Pays Bas) [Zanisher 82], CERVIFIP (Royaume Uni) [Tucker 82], DIASCANNER (Suède) [Nordin 89], FAZYTAN ( Allemagne) [Erhardt 80] et LEYTAS (Pays Bas) [Ploem 87].

C'est pourquoi un système entièrement automatique n'a encore jamais vu le jour et les cyto-pathologistes se sont plutôt tournés vers le développement d'un système semi-automatique qui laisserait le jugement au spécialiste. Le rôle de ce type de système est alors de détecter les lames suspectes ( à savoir celles présentant des cellules atypiques ou anormales) et de montrer ces cellules à l'observateur pour qu'il puisse juger la lame. Le système peut alors soit bouger le microscope à l'endroit où se trouvent les cellules "anormales" ou bien les afficher sur un écran en précisant les cellules à diagnostiquer. Le rôle de la machine ne serait plus alors de diagnostiquer une lame cytologique mais plutôt à présenter au cyto-pathologiste les zones pour un jugement. On ne cherche pas ainsi à remplacer le travail du cyto-pathologiste mais à le faciliter.

Nous allons maintenant faire une revue exhaustive des différents travaux de recherche sur les systèmes semi-automatiques effectués jusqu'à nos jours :

- meso-TICAS [Dytch 87] qui est la suite du TICAS développé par Wied G.L. Il fut développé en 1986 à l'université de Chicago. Ce système permet les mesures de caractéristiques et présente une originalité car il intègre des paramètres non utilisés dans les autres systèmes, à savoir des paramètres permettant d'avoir des informations de voisinage entre les cellules : la distance à la cellule la plus proche et l'orientation relative de la cellule par rapport aux autres. La coloration utilisée était la coloration Feulgen. Ce système donne les caractéristiques cellulaires brutes à l'utilisateur et un système expert permet de les interpréter [Bartels 87],

- Le SCIMA ( System for Cytologic Image Analysis ) [Hall 87] développé à l'université de Los Angeles de 1977 à 1985. Il intègre un classifieur de type "arbre de décision" pour trouver les cellules atypiques. Il est principalement adapté à la détection de lésions précancéreuses du col de l'utérus,

- Le système IRIS [Kasdan 87] réalisé par International Remote Imaging Systems en 1987. Il permet la mesure des caractéristiques habituelles des cellules de frottis vaginaux et sa principale caractéristique est la sélection des zones d'intérêt par le cyto-pathologiste grâce à un écran sensitif,

Quelques autres systèmes à peu près équivalents ont vu le jour dans les années 80 mais ils n'étaient pas réellement innovants. Par contre la recherche commençait à s'orienter vers les différentes manières d'interpréter les caractéristiques des cellules : systèmes experts, arbres de décisions, réseaux neuronaux, fonctions discriminantes linéaires, etc.

Dans les années 90, on trouve des systèmes intégrant enfin toutes les spécificités d'un système d'analyse d'image semi-automatique, à savoir :

- Screening d'une lame,
- Segmentation de l'image,
- Analyse des caractéristiques des cellules,
- Diagnostic de l'atypicité des cellules,
- Affichage des cellules atypiques.

On trouve donc des méthodes variées à tous les niveaux de ces spécificités qui dépendent beaucoup du but recherché. On retiendra les travaux suivants :

- Les travaux de Kate T.K et Al [Kate 93] pour compter les mitoses dans des sections cancéreuses du sein qui utilise des fonctions discriminantes linéaires pour la classification des cellules.

- Les travaux de Serbouti et Al [Serbouti 93] sur l'analyse d'image et la classification des leucémies. Il utilise des arbres de classification et de régression pour la classification des cellules. Parmi les caractéristiques cellulaires, les plus déterminantes étaient la couleur basophile de base, l'arrangement et la structure de la chromatine,

- Les travaux de Dufer J et Al [Dufer 93] pour la détection du cancer du sein qui définit des paramètres cytomorphométriques du noyau ( densitométriques, géométrique, de texture ). La segmentation est faite par le logiciel d'analyse d'image SAMBA [Brugal 87] La classification utilise des fonctions discriminantes linéaires. Il fonctionne avec la coloration Feulgen.

- Les travaux de Augusto Moragas et Al [Moragas 93] sur les liquides d'épanchement des séreuses où la segmentation est faite par le logiciel VISILOG [Visilog 88] et la classification par des fonctions discriminantes linéaires. C'est un des seuls travaux à s'intéresser aux liquides d'épanchement par rapport aux autres qui travaillent tous sur des frottis,

- Les travaux de Teague T.Kent et Al [Teague 93] sur les lymphocytes. La segmentation est faite par le logiciel du domaine public Image [Rasband 90].

- Les travaux de Antonangelo Leila et Al [Antonangelo 94] qui étudie les liquides d'effusions pleurales. Il y est établi que l'on ne peut se baser uniquement sur des critères cytoplasmiques pour classier les cellules. Malheureusement la conclusion est décevante car aucune caractéristique n'est donnée prédominante,

- Le système HOME [Derder 95] qui introduit un nouveau concept : les systèmes ergonomiques intégrés. Ce type de système permet de fournir des outils de haut niveau nécessaires aux tâches d'observation, de sélection, de marquage etc. Il permet d'assister le cyto-pathologiste et de contrôler ses activités de diagnostic. Des évènements représentent les activités d'interaction du cytopathologiste avec le système, ce qui permet de modéliser l'utilisateur à partir de son activité sous le système. L'objectif de cette modélisation étant d'adapter le comportement du système à l'évolution des caractéristiques de cet utilisateur,

- Enfin PAPNET [Koss 93,94, Boon 93, Mango 93, Sherman 94] qui s'avère être la dernière innovation intéressante en date. Il n'opère que sur des frottis et sa particularité réside dans l'utilisation de réseaux neuronaux pour la classification des cellules, ce qui a donné des résultats 20% meilleurs pour la plupart des frottis examinés.

Maintenant que nous avons vu les différentes tentatives de semi-automatisation, nous allons voir l'architecture des systèmes informatiques d'analyse d'images cytologiques .

### 2.4.1 Les Systèmes Informatiques d'Analyse d'Images cytologiques

Un système informatique d'analyse d'images cytologiques est constitué de trois parties ou étapes principales :

- La segmentation qui consiste en l'extraction de régions et/ou de primitives adaptées au problème rencontré.

Ici on cherche à isoler les composants cellulaires en l'occurrence,

- La caractérisation qui évalue les caractères extraits à la segmentation. Cela consiste à fournir les données représentatives de chaque cellule : taille, surface, texture etc.

- L'étiquetage qui recherche la meilleure classification pour les régions extraites à la segmentation. Pour cela on se sert des caractéristiques cellulaires trouvées lors de l'étape précédente. Cela consiste en fait à trouver le type de la cellule et à affirmer sa bénignité ou sa malignité.

Tous les systèmes qui ont été développés en cytologie sont de ce type . Nous allons voir maintenant les progrès et les techniques connues pour chacune des trois étapes.

#### 2.4.1.1 La segmentation

D'après Wied G.L [Wied 89] une segmentation robuste et sûre est un pré requis de base pour l'analyse d'images. Elle détermine la consistance et l'exactitude des informations extraites et interprétées. Cette étape reste une des étapes des plus difficiles et des plus imprédictibles. Durant les quinze dernières années un nombre substantiel de procédures et d'enchaînement de procédures ont été développées résultant d'intenses recherches dans ce domaine. On peut dire que le type de segmentation utilisé dépend beaucoup du but recherché. La segmentation est confrontée avec les tâches suivantes : délimiter le cytoplasme du fond et déterminer les limites du noyau. Cela peut généralement être fait par simple seuillage direct, par seuillage adaptatif ou par suivi de contours. Mais la principale difficulté rencontrée est le chevauchement possible des cellules :

- ce qui peut aller du simple contact au chevauchement complet des noyaux interdisant alors toute mesure de caractéristiques cellulaires comme la ploïdie par exemple.

Les techniques de segmentation, sur des images cytologiques, rencontrées dans la littérature sont diverses :

- système CYBEST [Tanaka 77,87] : méthode d'histogramme différentiel du premier et second ordre et seuillage pour le cytoplasme et le noyau,

- système micro-TICAS [Dytch 87] : seuillage et suivi de contours guidés par un expert. Les cellules agglomérées ne sont séparées que si elles sont en simple contact car les amas sont considérés comme inutiles ( ce qui est assez discutable ). Pour séparer deux cellules en simple contact, on parcourt le cytoplasme et l'on cherche les endroits où il y a de grandes variations d'angle dans le parcours,

- système SCIMA [Hall 87] : simple seuillage au choix de l'utilisateur, ce qui est un peu réduit.

- système IRIS [Kasdan 87] : seuillage et suivi de contours guidés par un expert. Certaines cellules sont rejetées sur des critères de circularité,

- par [Kate 93] : accroissement de régions dans la direction du gradient maximal. En premier lieu, un lissage est appliqué à l'image et les points d'exploration pour la croissance de régions sont trouvés en appliquant un seuillage bas à l'image. Les objets trop petits sont éliminés. Les résultats sont assez bons, car 95 à 98% des cellules sont bien délimitées,

- par [Moragas 93] : seuillage interactif, sélection des cellules à analyser, suivi de contours et " ou logique " entre l'image obtenue précédemment et l'image initiale ce qui donne l'image I'. Cette image I' sert de masque binaire pour spécifier les zones d'intérêt. Sur l'image I' on fait des érosions-dilatations par un filtre 4\*4 et finalement on applique un filtre "top-hat" ce qui donne l'image I". Cette image I" sert pour l'extraction de la chromatine.

- par [Teague 93] : rehaussement du contraste puis filtre par chapeau mexicain 13\*13 et enfin détection de contours suivi d'un seuillage dont le seuil a été défini par l'expérience,

- le système PAPNET [Koss 94] : un procédé complexe d'érosions et de dilatations sur l'image initiale qui élimine certaines cellules grâce à des considérations de forme, de taille et de densité optique, ce qui donne l'image I' qui contient les cellules trouvées. Les objets de contraste insuffisant sont éliminés grâce à l'algorithme du top hat renversé,

Au vu de la disparité des techniques utilisées, on ne peut pas dire qu'il existe une segmentation type sur des images de cytologie. Cela s'explique par la diversité des images. Certains logiciels comme VISILOG [Visilog 88], SAMBA [Brugal 87], IMAGE [Rasband 90] permettent de disposer des outils morphologiques de traitement d'images et sont souvent utilisés par les cyto-pathologistes.

On trouve dans la littérature des travaux portant uniquement sur la segmentation d'images cytologiques, les techniques utilisées y sont complexes mais les résultats sont assez concluants [Menard 93, Masson 92].

Des systèmes experts à base de règles de production ont été conçus pour guider la segmentation des images [Bartels 87, Wied 89,92], mais la variété des images a vite rendu leur gestion très lourde compte tenu du nombre de règles assez important.

### **2.4.1.2 La caractérisation des cellules**

Une fois l'image initiale I contenant les cellules segmentée, on obtient une image I' qui contient les différents objets à analyser. Chaque objet de l'image a une étiquette ou référence et pour pouvoir classer chaque cellule (i.e pour connaître son type) , il nous faut des critères mesurables. On conçoit facilement quels genres de critères peuvent être mesurés :

- la surface du cytoplasme,
- la surface du noyau,
- la forme du noyau,
- etc.

Le problème qui se pose est de savoir quelles caractéristiques de la cellule choisir de calculer. Il n'en faut pas trop pour ne pas trop accroître le temps de calcul, mais il en faut assez pour pouvoir effectuer une classification significative.

A ce propos nous allons faire un panorama des caractéristiques les plus généralement retenues. Les caractéristiques que l'on retrouve dans presque la totalité des travaux effectués sur le sujet sont :

- la taille du cytoplasme,
- le rapport Noyau / Cytoplasme,
- la densité optique,
- la forme du noyau et/ou du cytoplasme,
- la texture du noyau et/ou du cytoplasme.

Par exemple ce sont les caractéristiques retenues par le système CYBEST [Tanaka 77,87]. Mais pour d'autres systèmes elles peuvent être très différentes :

- système TICAS [Dytch 87] : caractéristiques générales ainsi que l'orientation relative de la cellule par rapport aux autres et la distance à la plus proche cellule,
- système IRIS [Kasdan 87]: caractéristiques générales,
- par [Kate 93] : caractéristiques générales plus le plus petit niveau de gris de la cellule et la convexité,
- par [Serbouti 93] : 62 caractéristiques sont extraites par cellule, 31 pour le noyau et 31 pour le cytoplasme. Ces caractéristiques ont été retenues après des tests discriminants sur 100 caractéristiques différentes,
- par [Dufur 93] : des tests ont été effectués sur 18 caractéristiques différentes dont 6 de densité optique, 9 de texture et 3 de géométrie. Parmi celles-ci les plus discriminantes sont la densité optique, la déviation standard de l'histogramme de la densité optique, la taille du noyau, la distribution des niveaux de gris et la distribution " run-length ",
- par [Moragas 93] : critères uniquement de texture sont retenus,
- par [Giménez 93] : étude morphologique de la texture de la chromatine,
- par [Jackway 95] : extraction d'éléments de texture par la ligne de partage des eaux,
- par [Antonangelo 94] : caractéristiques générales plus le plus grand et le plus petit diamètre des noyaux et des cytoplasmes ainsi qu'un facteur de forme sur le cytoplasme et le noyau.

On se rend alors compte de la variété des traits cellulaires possibles, mais lesquels choisir et surtout pourquoi ? A cet égard, nombre de cyto-pathologistes se sont intéressés à l'étude des traits cytométriques des cellules. Des travaux de comparaison entre les cellules sur des traits cellulaires particuliers et des standards cytométriques ont pu voir le jour. On constate cependant que le besoin de standards mesurables est sur le déclin avec l'apparition de méthodes empruntées à l'intelligence artificielle en particulier les réseaux de neurones. Mais les standards cytométriques peuvent servir de connaissance issues du domaine pour guider la segmentation.

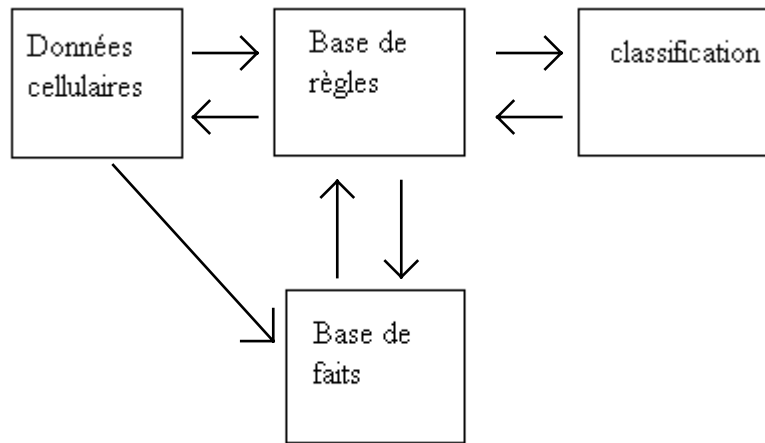
### **2.4.1.3 La classification des cellules**

Une fois chaque cellule caractérisée par différentes valeurs mesurables, le système informatique de cytologie se sert de ces valeurs pour classer la cellule. C'est à ce stade qu'intervient donc une décision. Le système de décision doit intégrer les résultats de mesures quantitatives avec éventuellement des données cliniques.

L'introduction de systèmes de décision a été un effort significatif pour le développement de la cyto-pathologie. Ces systèmes doivent avoir des capacités d'apprentissage et doivent être capables de réactualiser leurs connaissances compte tenu des données qui leur sont proposées. A ces fins, un système de décision doit être construit le plus déclarativement possible. Nous allons voir les différents types de systèmes de décision qui ont existé jusqu'à nos jours.

Au départ les systèmes conçus donnaient les données cellulaires en vrac [Tanaka 87, Dytch 87, Kasdan 87] et le cyto-pathologiste devait se débrouiller avec celles-ci. Réalisant rapidement la somme de travail astronomique que cela représentait, les premiers systèmes de décision sont apparus. Ceux-ci étaient les fonctions discriminantes que l'on retrouve dans différents travaux [Kate 93, Dufur 93, Moragas 93, Antonangelo 93, Taylor 78]. Ce type de classifieur statistique a été très pratique mais il a des désavantages certains : ils sont restreints par le nombre de classes qu'ils maintiennent et ils ne fonctionnent pas par identification positive mais par similarité.

Dans le but de se rapprocher de ce concept d'identification positive, des systèmes experts ont été conçus. Ils peuvent être définis pour être extrêmement spécifiques et sélectifs mais leur structure leur permet toujours d'incrémenter les informations de classification. Pour mémoire on rappelle la structure d'un système expert :



La base de faits contient des informations vérifiées et la base de règles contient des règles établies par un expert en cyto-pathologie. Ce genre de système de décision peut-être extrêmement efficace mais des limitations apparaissent rapidement. Elles sont dues aux tests de redondance de l'information, aux contradictions et aux boucles possibles. La maintenance d'une base de faits et de règles correctes est donc très difficile si l'on veut avoir un système très général. On pourra consulter différents travaux sur ces systèmes experts [Wied 89, Bartels 87].

D'autres types de systèmes de décision comme les arbres de décision ont été conçus [Hall 87, Serbouti 93]; ce sont des arbres où chaque noeud représente une étape et où chaque arc signifie une transition dans le raisonnement. Cependant ils présentent les mêmes limitations que les systèmes experts.

Pour pallier ce type de désavantage, une approche de type connexionniste par les réseaux de neurones a été imaginée. On ne rappellera pas ici le fonctionnement d'un réseau de neurones, voir le chapitre § B.3.2.1 pour plus de détails. Ce type de technique, inspirée par les réseaux biologiques, utilise des éléments hautement interconnectés ( les neurones ) qui encapsulent leur connaissance dans le poids de leurs interconnexions. Ils possèdent surtout la possibilité d'apprendre et de généraliser à partir d'exemples connus. La variété des types de réseaux neuronaux permet d'avoir un champ d'application très étendu, en effet la combinaison des différents paramètres des réseaux neuronaux liée à son architecture constitue un paradigme de réseau qui permet de donner une solution à des types de problème bien précis. Ceci intéresse donc grandement les cyto-pathologistes qui veulent faire le moins de faux-négatifs possible car cela a des graves conséquences pour le patient et le laboratoire. L'avantage majeur pour les cyto-pathologiste est donc que les réseaux neuronaux peuvent classer des cas où il est difficile de formuler des règles simples pour distinguer plusieurs catégories de cellules.

Ce genre de système est d'après [Wied 89] 20% meilleur que les systèmes fondés sur des fonctions statistiques. L'avantage des réseaux neuronaux vient surtout de leur capacité à tolérer des données bruitées, ambiguës ou erronées et à généraliser à partir d'exemples connus pour gérer les cas inconnus. Plusieurs essais de systèmes de décision basés sur les réseaux neuronaux ont été faits [Wied 89, Mango 93, Dytch 90, Yahiaoui 92] et les résultats paraissent assez concluants : PAPNET est un système utilisant un réseau neuronal et sa sensibilité moyenne est de 97% avec une efficacité moyenne de 78% [Boon 93].

D'autres travaux sur les réseaux neuronaux [Truong 95, Dytch 90] ont été menés afin de tester ses différents paramètres afin d'optimiser la classification de cellules : taille, fonction de transfert, taux de tolérance de test et d'entraînement, taux d'entraînement et réaction au bruitage.

Beaucoup de travaux de tests ont été réalisés sur Papnet afin de vérifier ses performances [Koss 93,94, Boon 93, Mango 93, Sherman 94], mais ce type de système ne satisfait pas encore tous les cyto-pathologistes. Les performances d'un système dans la donnée d'une classification sont testées en terme de bons ou mauvais cas de classification en considérant que la classification d'un expert humain est la vérité. Il apparaît alors un gros problème de calibrage dû aux différences de jugement variant d'un expert cyto-pathologiste à un autre. La seule solution serait de réaliser une base de données d'images cytologiques au niveau mondial avec les différents avis des pathologistes pour permettre aux futurs logiciels des développements identiques au sens de l'exactitude du diagnostic humain.



### 3. Position du stage

#### 3.1 Les objectifs

Notre objectif est de développer une étude sur le tri cellulaire sur des images de cytologie en collaboration avec le service d'anatomie et de cytologie pathologiques du Centre Hospitalier Louis Pasteur de Cherbourg ( C.H.L.P ). Ce service effectue régulièrement des mesures sur les lames cytologiques qui leur sont proposées dont surtout des pourcentages sur les différents types de cellules trouvées sur une lame.

Le travail consiste à créer un système qui permette une automatisation du tri cellulaire sur ces images de cytologie. Cette automatisation du tri cellulaire servirait dans un laboratoire médical en vue d'une aide au screening ( parcours de la lame ). Ce type de système de tri nécessite comme nous l'avons vu précédemment ( § 2.5 ) l'architecture suivante :

Segmentation de l'image,  
Analyse des caractéristiques cellulaires,  
Classification des cellules ( le tri ).

Ce genre de système nécessite donc de regrouper et de modéliser des connaissances d'horizons divers : pathologie, traitement d'images, intelligence artificielle.

L'étude s'applique à des images de liquides d'épanchements et la préparation appliquée aux lames conserve le cytoplasme et le noyau. De plus les images dont nous disposons sont en couleur ce qui sert aux cyto-pathologistes pour identifier les cellules.

Il faudra également mettre en place un logiciel permettant de constituer une base de données qui est nécessaire pour l'apprentissage et les tests des méthodes de classification. Ceci consistera en un étiquetage des images par des experts en cyto-pathologie.

Le but de ce stage est donc de :

- dégager les spécificités que nécessite la mise en oeuvre d'un système informatique de tri cellulaire,
- mettre en oeuvre un début de système de tri cellulaire utilisant des techniques de traitement d'images et d'intelligence artificielle,
- créer une base de données sur des images de cytologie.

#### 3.2 Travaux déjà réalisés sur le sujet et comparaison avec le stage

Les images cytologiques du CHLP proviennent de liquides d'épanchement et les travaux qui ont été réalisés jusqu'à présent portent majoritairement sur des frottis car ils constituent la majorité des lames de cytologie et également pour des raisons financières. En ce sens l'outil de tri cellulaire que nous nous proposons de développer se distingue de par l'origine des images de ses lames.

La coloration utilisée pour les images au CHLP est celle Papanicolaou et est également majoritairement utilisée dans les autres systèmes de tri.

Tous les travaux de recherche qui ont été réalisés portent sur des images qui sont examinées en niveaux de gris, nous envisageons d'exploiter les ressources d'une image couleur afin d'améliorer le tri .

Les systèmes de tri cellulaire actuels convergent vers une classification par réseaux neuronaux, nous nous inscrirons dans cet axe car nous utiliserons ce type d'architecture pour le tri.

#### 3.3 Conclusion

Le stage s'inscrit donc dans une logique de développement de systèmes de tri cellulaires en vue d'une aide au screening. Nous resterons dans la lignée des derniers travaux de recherche réalisés sur le sujet au niveau de la classification mais certaines différences apparaissent au niveau de l'interprétation et l'utilisation de la couleur et la provenance des images. On ne fait pas de comparaison avec les techniques de segmentation et de caractérisation de la littérature car elles sont tellement diverses qu'une homogénéisation est impossible. On retiendra par contre que dans ces deux derniers éléments des systèmes informatiques d'aide au screening, le traitement des amas de cellules reste un travail difficile non encore résolu et les principales difficultés de segmentation et de caractérisation y sont rattachées. Un traitement approprié des amas de cellules sera donc un des objectifs majeurs.

---

## **Partie B : le système mis en place**

---

---

# Introduction

---

Comme nous l'avons vu dans la partie précédente, dans ce stage nous nous proposons de faire une étude sur le tri cellulaire sur des images de cytologie du C.H.L.P. Les images dont nous disposons proviennent de lames de liquides d'épanchement. La numérisation a été faite au centre François Baclesse à Caen et les images sont des images couleurs 8 bits de 750 x 572 pixels.

Le tri cellulaire consiste à trier les cellules présentes dans une image selon les classes de cellules qui nous ont été assignées. L'architecture d'un tel système de tri cellulaire repose sur trois parties s'enchaînant :

- La segmentation de l'image
- Analyse des caractéristiques cellulaires
- La Classification des cellules

Nous allons voir ces trois parties tout au long de ce deuxième chapitre.

# 1. La Segmentation

## 1.1 Le but de la segmentation

La segmentation consiste en la construction d'une représentation structurée et condensée synthétisant les informations présentes dans la scène, sous forme de régions ou de contours. Les contours s'obtiennent par l'analyse des discontinuités sur certains types d'informations (niveaux de gris, texture), tandis que les régions sont obtenues par l'analyse de leur homogénéité. La segmentation a donc pour objectif l'extraction d'indices visuels dans une image et leur regroupement en structures "perceptives" (régions, contours). La segmentation n'est pas considérée comme un but en soi, mais plutôt comme un moyen efficace et indispensable de réduction de données permettant leur analyse par des processus de plus haut niveau.

Dans notre cas la segmentation doit permettre d'extraire pour chaque cellule son cytoplasme et son ou ses noyaux. La segmentation se devra d'être robuste et sûre car de ses résultats dépendent la consistance et l'exactitude des informations qui en seront extraites et interprétées. Cette étape est des plus difficiles et des plus imprédictibles.

## 1.2 Etude de l'image

L'étude de l'image passe par l'étude de son contexte. Un contexte est constitué de trois niveaux de description [Elmoataz 90, Clouard 94] :

❶ Les informations de niveau physique décrivent la physique de formation de l'image et les particularités "techniques" induites sur l'image. Ce sont notamment le type de scène représentée, liée au type de capteur utilisé (microscope, satellite, ...), les conditions d'acquisition (brouillard, sur éclairage, ...), la nature du bruit (fort, faible, ...) la qualité (floue, nette, ...), ...

❷ Les informations de niveau perceptif décrivent les indices visuels de l'image, et renseignent sur les caractéristiques "syntaxiques" de l'image. Il y est fait état du type d'information représentée (nuage de points, image de densité, image de contours...), des caractéristiques des contours et frontières entre les objets (contrasté, type, longueur, ...), des caractéristiques des régions (texturée, homogène, surface, ...), s'il existe un fond sur lequel reposent tous les objets...

❸ Les informations de niveau sémantique doivent traduire les notions attachées au concept d'objet pour le domaine d'application, ainsi que les relations qui unissent ces objets. Ce sont les descriptions des différents types d'objets possibles (taille, forme, couleur...), et leur organisation dans la scène (répartis, amassés, juxtaposés).

Le contexte des images dont nous disposons est le suivant :

### DESCRIPTION PHYSIQUE

type de la scène   microscopique,  
qualité de l'image   nette,  
bruit                                    moyen

### DESCRIPTION PERCEPTIVE

présence d'un fond            oui  
aspect du fond                    homogène  
aspect des objets                homogène  
couleur du fond                claire  
couleur des objets               foncée

### DESCRIPTION SEMANTIQUE

taille des objets   moyenne ou grosse  
répartition                        dissociée  
densité                               moyenne  
forme des objets   convexe

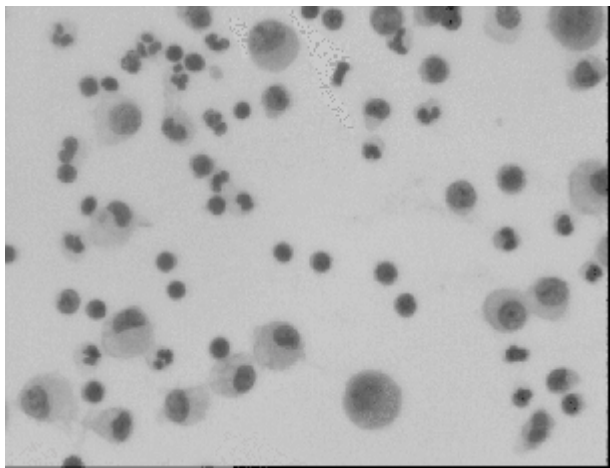


figure II.1. un exemple d'image

Les images contiennent des cellules et sont constituées de trois parties : le fond, les cytoplasmes des cellules et les noyaux des cellules.

Les images dont nous disposons sont en couleurs données dans le système RVB : un pixel couleur est un vecteur à trois composantes qui traduisent l'intensité lumineuse pour les trois couleurs de base ( RVB - rouge, vert, bleu ). Ces trois grandeurs des couleurs primaires définissent un espace tridimensionnel, appelé modèle RVB, qui constitue la représentation numérique de base de la couleur.

Pour étudier les images il nous faudra donc étudier chacune des trois composantes rouge, verte et bleue. Nous avons étudié en particulier la répartition des niveaux de gris dans l'image ( ce qui nous a permis de savoir si le fond et les objets étaient homogènes). On cherche à partir de quel seuil haut le fond est séparé des cytoplasmes des cellules, et à partir de quel seuil bas les cytoplasmes sont séparés des noyaux. Les résultats que nous avons obtenus sont donnés dans le tableau suivant ( figure II.2 )

Une Image	Composante Rouge	Composante Verte	Composante Bleue
Image 1	Seuil Bas : 162 Seuil Haut : 178	Seuil Bas : 136 Seuil Haut : 180	Seuil Bas : 162 Seuil Haut : 178
Image 2	Seuil Bas : 154 Seuil Haut : 180	Seuil Bas : 132 Seuil Haut : 180	Seuil Bas : 154 Seuil Haut : 180
Image 3	Seuil Bas : 172 Seuil Haut : 206	Seuil Bas : 136 Seuil Haut : 204	Seuil Bas : 172 Seuil Haut : 206
Image 4	Seuil Bas : 160 Seuil Haut : 188	Seuil Bas : 136 Seuil Haut : 184	Seuil Bas : 160 Seuil Haut : 188

Seule la composante verte donne des seuils vraiment homogènes et l'écart entre le Seuil Bas et le Seuil haut est important. Cette composante permettra d'arriver à des résultats concluants. Nous avons donc décidé de procéder à une segmentation seulement sur la composante verte des images, les autres composantes pouvant être utilisées pour des traitements complémentaires.

Nous avons à notre disposition 8 images obtenues à deux grossissements différents, mais nous n'avons retenu que les quatre plus simples car les quatre autres paraissaient nécessiter un traitement beaucoup trop approfondi pour cette étude (trop d'amas sont présents). Notre étude sera donc effectuée sur ces quatre images obtenues à deux grossissements différents.

## 1.3 La stratégie de segmentation

### 1.3.1 Séparation cytoplasme, noyau et fond

La stratégie que nous allons appliquer aux images de cytologie doit permettre d'extraire des régions qui représenteront des cellules, il va donc nous falloir séparer le cytoplasme du fond ainsi que les noyaux des cytoplasmes, ce qui paraît possible étant donné le contexte de l'image (voir § 1.2 ).

### 1.3.2 Méthodes utilisées dans la littérature

Les méthodes que l'on trouve dans la littérature sont très diverses et comme nous l'avons vu dans la partie A, elles dépendent beaucoup du but recherché. Peu de travaux ont été réalisés sur les liquides d'épanchement des séreuses auxquels

nous nous intéressons dans cette étude. Si bien que les résultats que nous obtiendrons sont imprédictibles et devront être testés sur une plus grande variété d'images.

### 1.3.3 La méthode choisie

Nous allons maintenant voir en détail le principe de la méthode de segmentation que nous avons choisi d'implanter et les différentes étapes qui nous y ont conduit.

#### 1.3.3.1 Le seuillage

On rappelle ici qu'on ne travaille que sur la composante Verte des images couleurs dont on dispose, c'est donc sur une image en niveaux de gris que la segmentation sera effectuée.

Nous avons vu au § 1.2 qu'un double seuillage permettait déjà d'isoler assez grossièrement les cytoplasmes et les noyaux ( voir figure II.3), ceci n'est pas suffisant mais peut servir de base à une méthode d'extraction de frontières.

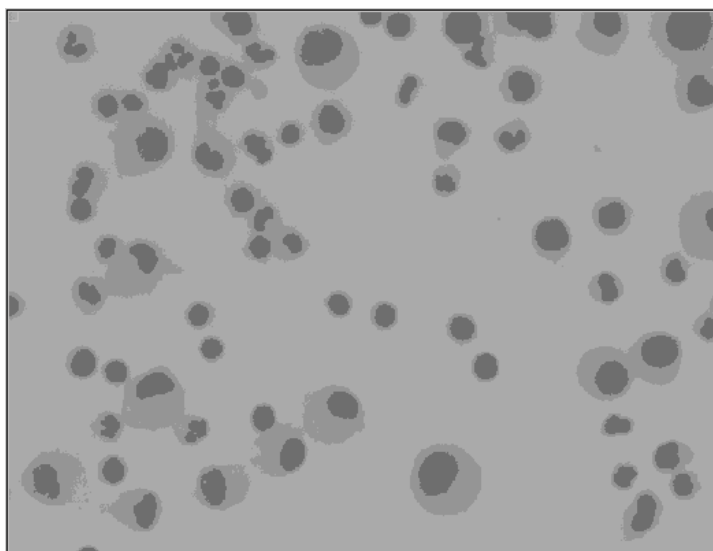


figure II.3 : image obtenue par double seuillage

Etant donné que l'on connaît les seuils bas et hauts on peut trouver les frontières ( figure II.4)

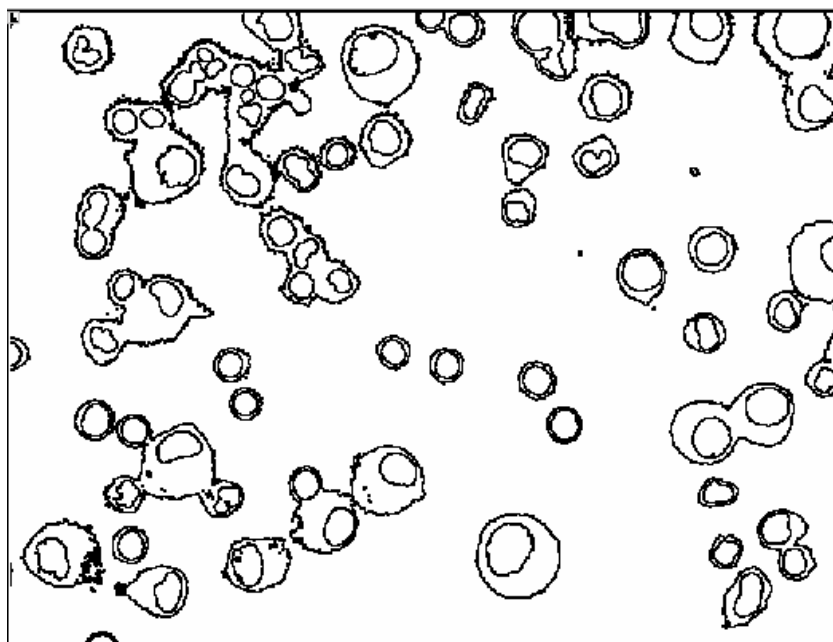


Figure II.4 : image seuillée où l'on a détecté les transitions de niveaux de gris

Nous allons nous aider de cette image pour trouver exactement les frontières des noyaux et des cytoplasmes à l'aide de la méthode de la Ligne de Partage de Eaux ou L.P.E.

### 1.3.3.2 La ligne de Partage des Eaux ou L.P.E

Dans ce paragraphe nous allons définir ce qu'est la L.P.E et nous allons voir son application sur nos images.

#### 1.3.3.2.1 Principe de la L.P.E

Nous allons considérer qu'une image en niveaux de gris est définie par une fonction  $f(x)$ , où  $f(x)$  représente l'amplitude du signal ( donc les niveaux de gris ) au point  $x$  de l'image (figure II.5).

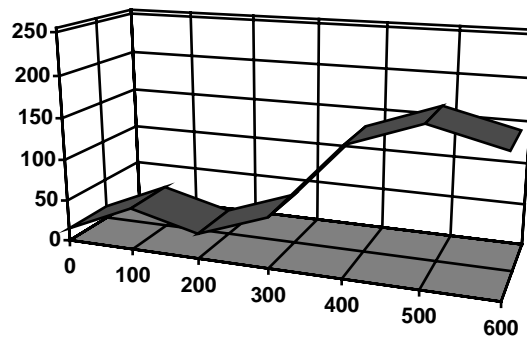


figure II.5: une image de niveaux de gris  
A un point X de l'image correspond un niveau de gris.

Pour mieux comprendre le principe de la L.P.E nous allons comparer une image en niveaux de gris à un relief terrestre, l'altitude jouant le rôle des niveaux de gris ( on retrouve la définition de la fonction donnée ci-dessus ). Dans cette comparaison, une image peut se diviser en plusieurs zones. Chaque zone sera définie comme un bassin versant. Un bassin versant correspond à l'espace géographique dans lequel toutes les eaux convergent vers le même minimum. Les différents bassins versants sont délimités par une ligne de partage des eaux appelée aussi ligne de crête des point les plus élevés joignant les bassins versants.

L'idée pour segmenter l'image serait de déterminer cette ligne de partage des eaux de l'image c'est à dire les lignes le long desquelles les niveaux de gris varient rapidement.

Pour cela nous allons voir une approche par immersion dans le cas d'une fonction  $f$  monodimensionnelle. Nous appellerons plateaux de  $f$  l'ensemble des points au voisinage desquels la fonction  $f$  est constante.

Afin de construire la ligne de partage des eaux, nous allons procéder en immergeant progressivement dans l'eau la fonction  $f$  par le bas. Pour que l'eau monte partout de façon régulière, on supposera que les minima régionaux sont "percés", comme le montre la figure II.6

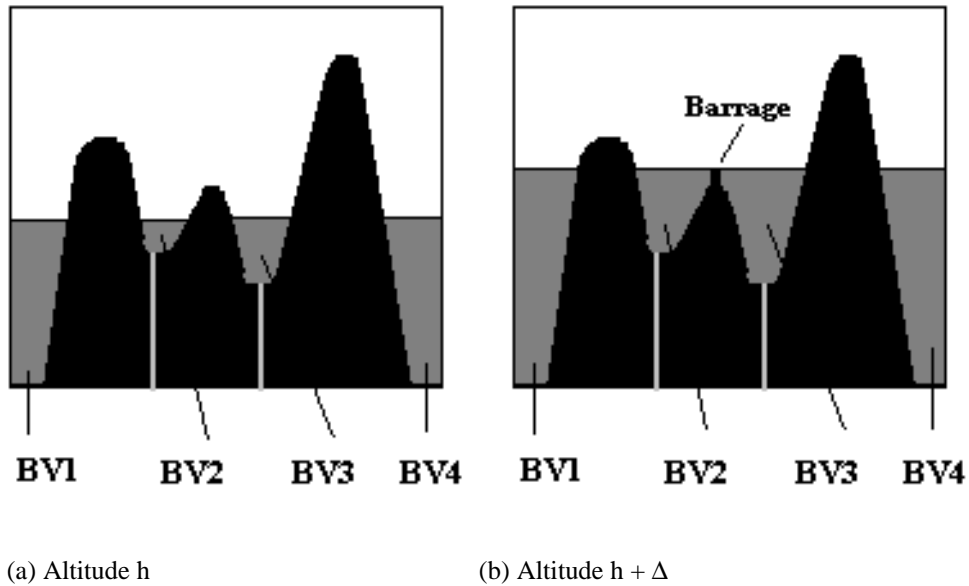


Figure II.6  
 Construction de la ligne de partage des eaux pour une fonction monodimensionnelle.  
 A l'altitude h, les quatre bassins versants sont encore séparés. Puis à l'altitude h + Δ, on construit un barrage pour séparer les eaux de BV2 et BV3

Passant par les ouvertures que nous avons créés, l'eau progresse alors dans les bassins correspondants à chaque minimum ( figure II.6.a). Lorsque les eaux en provenance de minima différents se rencontrent, on place un barrage, de telle sorte que deux eaux différentes ne se mélangent pas (figure II.6.b). La ligne de partage des eaux est constituée par l'ensemble de ces barrages. Elle partage donc l'image en différentes zones, que l'on peut considérer comme les zones d'influence ou d'attraction des minima régionaux : ce sont ces zones qui sont les bassins versants.

En dimension un, le lieu des barrages se calcule naturellement. Ils sont localisés sur les maxima locaux de l'image. En dimension deux ( le cas de nos images ), il est beaucoup plus délicat de caractériser ces barrages directement. Comme ces barrages sont les frontières des bassins versants, ce sont des lignes. Ces lignes contiennent bien plus que les maxima locaux : le contact entre deux bassins versants s'effectue sur un plateau. On suppose alors que cette rencontre a lieu au milieu de ce plateau.

Une définition formelle serait [Schmitt 94] :

Soit  $f : \mathbb{R}^2 \rightarrow \mathbb{N}$  une fonction bornée. Notons :

1.  $h_{\min} = \inf f(x)$  et  $h_{\max} = \sup f(x)$ ,
2.  $f^h$  la section inférieure de f au niveau h :  $f^h = \{x, f(x) \leq h\}$ ,
3.  $\text{Reg\_Min}_h f(x)$  l'ensemble des minima régionaux de f d'altitude h.

L'ensemble des bassins versants de f est l'ensemble  $X_{h_{\max}}$  obtenu après la récurrence suivante :

- (i)  $X_{h_{\min}} = f_{h_{\min}}$ ,
- (ii)  $X_{h+1} = \text{Reg\_Min}_{h+1}(f) \cup \text{IZ}_{f^{h+1}}(X_h), \forall h \in [h_{\min}, h_{\max}-1]$

où  $\text{IZ}_A(B)$  est l'union des zones d'influence géodésique de B dans A. La ligne des partage des eaux de f est alors le complémentaire de  $X_{h_{\max}}$  ( Figure II.7)



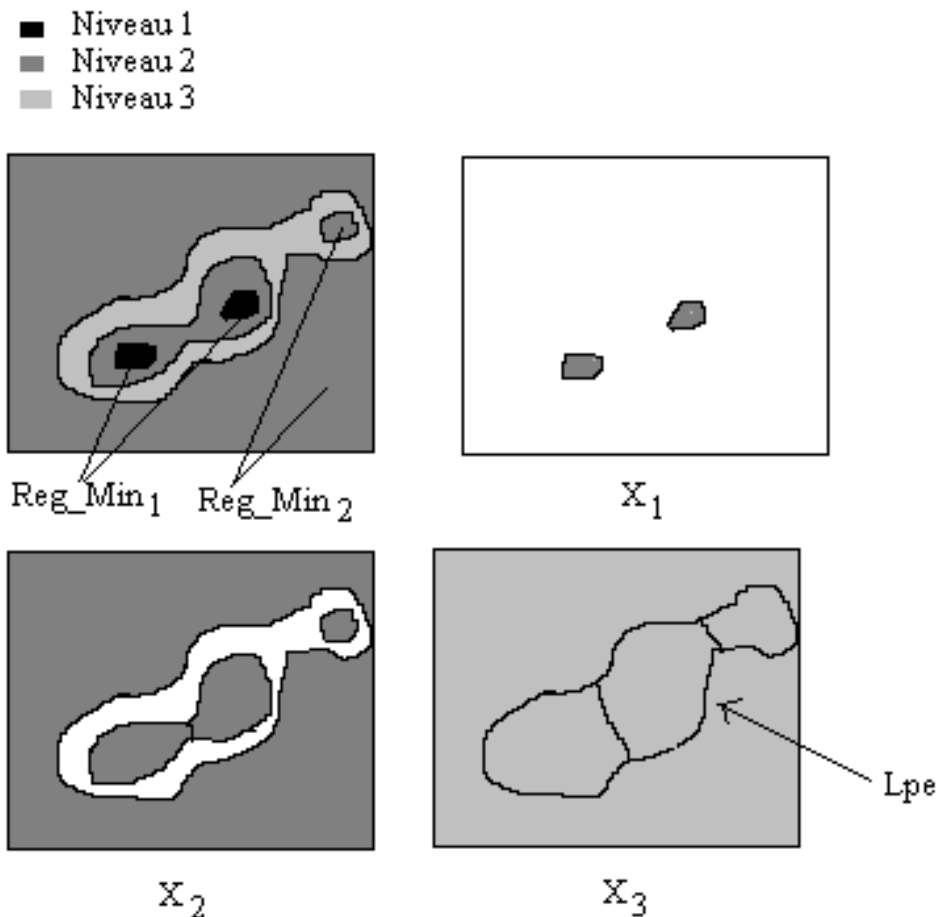


Figure II.7 - Exemple du processus d'immersion

Pour plus de détails on consultera [Schmitt 94, Coster 89].

### 1.3.3.2 La Lpe contrainte par des marqueurs

Nous allons voir maintenant une utilisation pratique de la Lpe pour des tâches de segmentation d'images et constater que la Lpe est un outil morphologique puissant de la segmentation. Cette utilisation de la Lpe va nous permettre de trouver des objets dans une image en niveaux de gris.

La ligne de partage des eaux calculée directement sur une image réelle représente une sur-segmentation importante, due au bruit qui crée de nombreux minima parasites. Il est donc nécessaire de mettre en place des procédures permettant d'éliminer certains minima régionaux. Pour cela nous allons nous servir de marqueurs.

Supposons que l'on connaisse un ensemble connexe de points faisant partie des objets. On appellera ces composantes connexes des marqueurs. Si l'on parvenait à modifier l'image sur laquelle on calcule la Lpe en lui imposant des ensembles comme minima régionaux, on obtiendrait une Lpe présentant une boucle autour de chacun des objets, chaque bassin versant correspondant à un unique objet.

Cette technique de Lpe contrainte par des marqueurs permet de s'affranchir de la recherche exacte des contours et garantit le nombre de contours que l'on trouvera : un autour de chaque objet s'il a été marqué par une seule composante connexe. La difficulté réside donc dans la détermination des marqueurs qui permettent de localiser grossièrement les objets.

Cette méthode consiste donc en :

- ❶ Trouver les marqueurs i.e une composante connexe par objet
- ❷ Déterminer l'image sur laquelle on calcule la Lpe
- ❸ Imposer les minima

#### ④ Calculer la Lpe

Pour plus de détails on consultera [Schmitt 94, Coster 89].

### 1.3.3.2.3 Application à notre cas

Comme nous l'avons vu au 1.3.3.1 on peut disposer, grâce à un double seuillage, d'une image donnant assez grossièrement les contours des objets que l'on veut caractériser ( à savoir le cytoplasme et le noyau des cellules ). Ces contours sont trop grossiers pour être exploités directement, l'idée serait alors de se servir de cette image obtenue par double seuillage comme d'une image de marqueurs et procéder à une Lpe contrainte par ces mêmes marqueurs. On calculerait la ligne de partage des eaux sur l'image du module du gradient de l'image en niveaux de gris avec les minima imposés par l'image obtenue par double seuillage.

Rappelons brièvement ce qu'est un gradient :

L'opérateur gradient est :

$$Df(x,y) = \frac{\partial f}{\partial x} u + \frac{\partial f}{\partial y} v \quad \text{où } u \text{ et } v \text{ sont les vecteurs unitaires des axes } x \text{ et } y.$$

son module est :

$$|Df(x,y)| = \sqrt{\frac{\partial f^2}{\partial x} + \frac{\partial f^2}{\partial y}}$$

La principale propriété du module du gradient réside dans le fait qu'il est élevé pour des points au voisinage desquels la fonction  $f$  représentant l'image est fortement discontinue, et il est faible pour les points au voisinage desquels la fonction image  $f$  varie peu. Le module du gradient est donc un maximum régional pour un point de contour, ce qui en fait un outil morphologique très bien adapté à la Lpe.

Voyons maintenant comment nous procédons pour obtenir l'image contenant la ligne de partage des eaux :

On commence par calculer une image de marqueurs obtenue par un double seuillage, les seuils ont été déterminés manuellement pour chaque image que nous traitons ( voir § 1.2 et § 1.3.3.1). Une image de marqueurs est celle de la Figure II.4

Ensuite on calcule une image du module du gradient (cette image est calculée sur l'image originale en niveaux de gris sur laquelle on a fait un lissage exponentiel décroissant ). Une image de gradient est donnée par la figure II.8

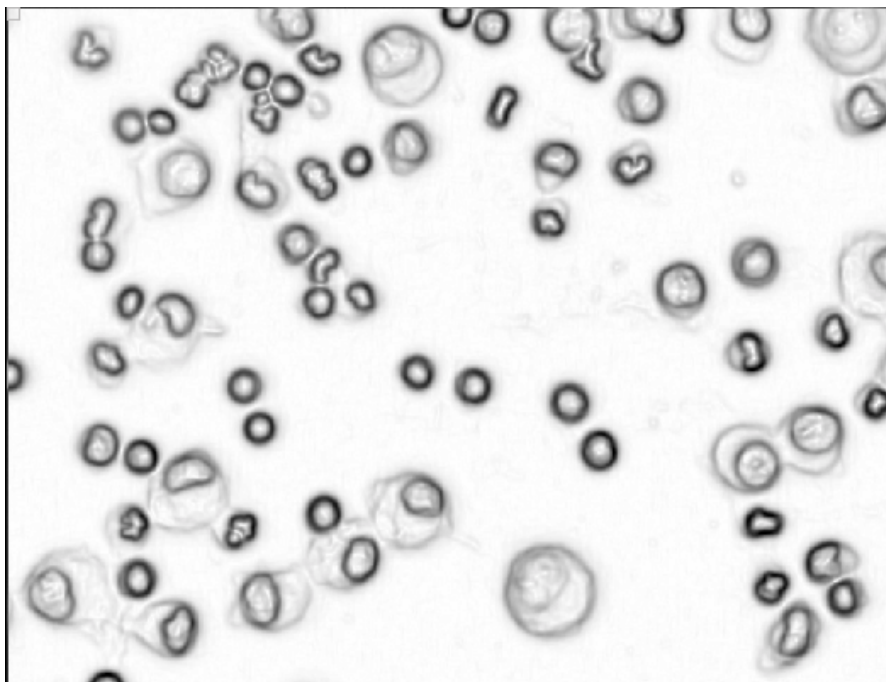


Figure II.8 : une image de gradient

Puis on calcule la ligne de partage des eaux sur l'image du module du gradient et on obtient l'image finale ( Figure II.9).

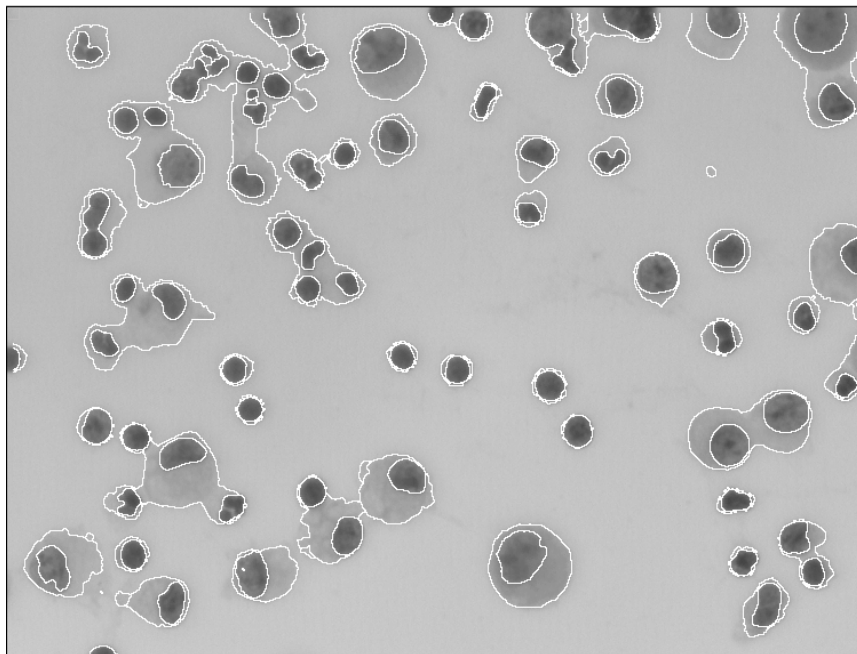


Figure II.9 : L'image obtenue par la Ligne de Partage des Eaux

On remarque que l'avantage de la Lpe est de fournir directement des contours fermés, ce qui en fait sa force. C'est donc cette méthode que nous avons retenue pour caractériser les contours des cytoplasmes et des noyaux des cellules. Les résultats obtenus sont très satisfaisants, les cellules simples aussi bien que les amas plus complexes sont bien délimités ( voir figures II.10).

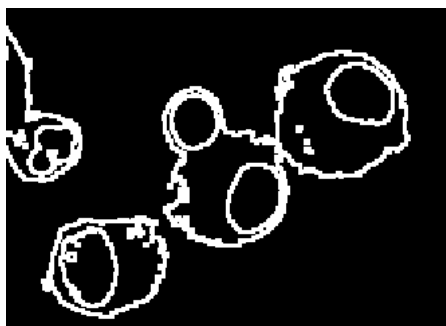


Figure II.10.a : les marqueurs

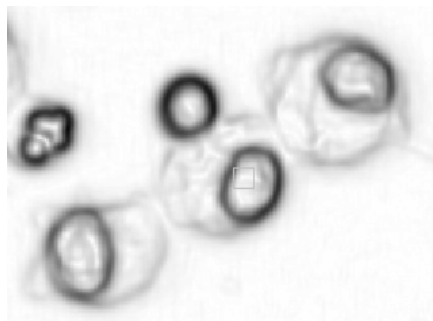


Figure II.10.b : le gradient

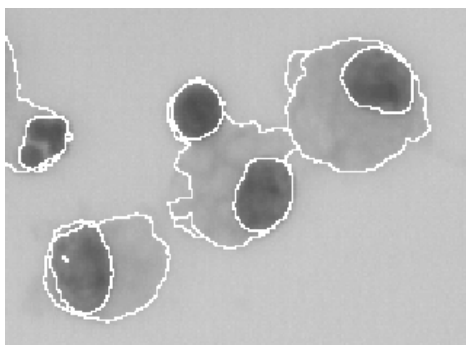


Figure II.10.c : la Ligne de Partage des Eaux

Figure II.10 : on constate que les frontières des cytoplasmes ont bien été trouvées.

### 1.3.3.3 Séparer les frontières cytoplasmes et noyaux

On dispose maintenant d'une image nous donnant les contours des objets contenus dans l'image, si l'on ne garde que les contours ( les points ayant leur valeur à 255), on obtient l'image binaire suivante ( Figure II.11)

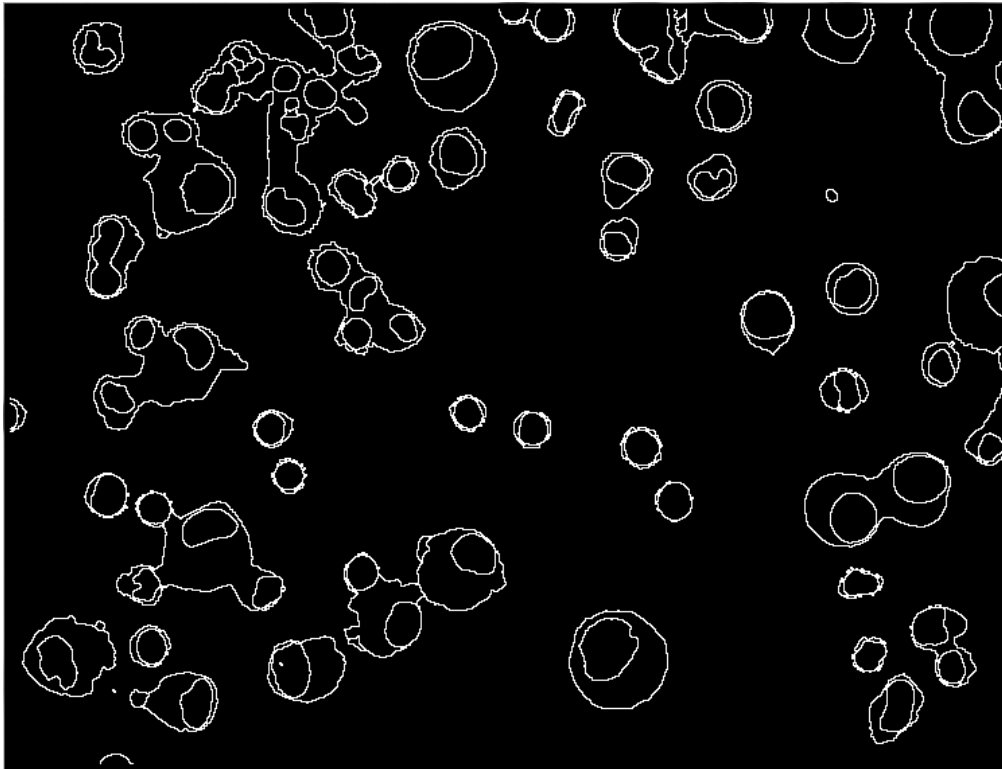


Figure II.11 : Image de Lpe binarisée

Mais ce qui nous intéresse est de séparer les cytoplasmes des noyaux. On voudrait donc avoir une image contenant les contours des cytoplasmes et une image contenant les contours des noyaux. Cela nécessite donc de savoir si un point de contour appartient à la frontière du cytoplasme, à la frontière du noyau ou bien encore si il est commun aux deux frontières. C'est ce problème de séparation des frontières des cytoplasmes et des noyaux que nous allons voir en détail maintenant. Deux méthodes ont été testées, nous allons les expliquer et les comparer.

#### 1.3.3.3.1 Première méthode : Suivi de contours

Tout d'abord quelques notations :  $NG_1(P)$  signifiera niveau de gris du point P dans une image I,  $E_8(P)$  signifiera l'ensemble des voisins en 8-voisinage du point P dans une image.

On rappelle que pour chaque image initiale en niveau de gris on connaît deux seuils haut et bas ( voir § 1.2 et 1.3.3.1) séparant grossièrement pour l'un le fond du cytoplasme et pour l'autre le cytoplasme du noyau. On dispose de deux images : l'image initiale en niveau de gris que nous appellerons I1 et l'image Lpe binarisée ( voir § 1.3.3.3 ) que nous appellerons I2. Dans I2 les points sont soit à 255 soit à 0. Un point à 255 représente un point de contour. Notre but est de séparer I2 en deux images I3 et I4. I3 est une image contenant les contours des cytoplasmes et I4 une image contenant les contours des noyaux. On procède ainsi :

❶ Dans cette première étape on va analyser tout point de contour ( à 255) de l'image I2. On va déterminer pour chacun de ces points si il appartient à une frontière d'un cytoplasme ou d'un noyau et l'ajouter à l'image I3 ou l'image I4. Pour cela on se sert des seuils bas et hauts : soit P le point de contour analysé, si  $\exists J / J \in E_8(P)$  si  $NG_1(J) < \text{Seuil bas}$  alors ajouter P à I3 ( l'image des noyaux ) sinon si  $NG_1(J) > \text{Seuil haut}$  alors ajouter P à I4 (l'image des cytoplasmes ). Voici un exemple grâce aux figures II.12

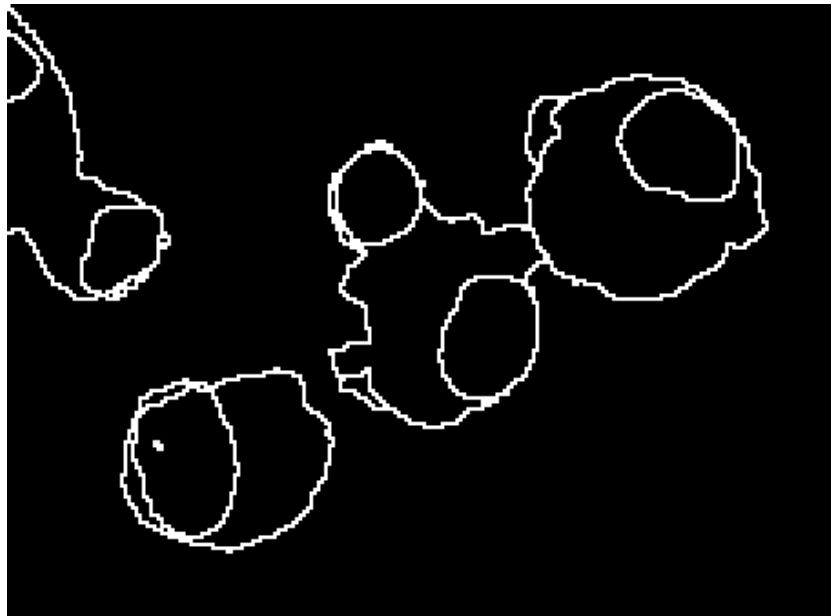


Figure II.12.a : les contours obtenus par Lpe



Figure II.12.b : les cytoplasmes



Figure II.12.c : les noyaux

On a donc juste réparti dans deux images I3 et I4 les points de l'image I2. Mais les contours que l'on obtient ne sont pas fermés, on va donc appliquer un suivi de contours.

② Dans cette deuxième étape nous allons analyser les images I3 et I4. Pour fermer les contours des images nous allons "marquer" les points de contour qui sont des points terminaux. En 8-voisinage on distingue plusieurs sortes de points de contour donnés par les masques suivants :

- Les points isolés

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

- Les points terminaux

$$\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

ou bien

$$\begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

ainsi que les points donnés par ces masques ayant subi une rotation.

- Les points intérieurs à une chaîne de contours

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{matrix} \quad (\text{plus rotations})$$

- Les points de jonction qui sont les points d'intersection de plusieurs chaînes de contour.

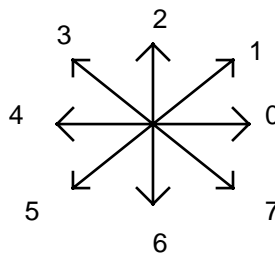
Cette étape consiste à marquer tous les points terminaux. Cela nous donne une image (voir figure II.13 ) que nous appellerons I5.



Figure II.13 : Les points terminaux ont été marqués.

➊ Ensuite on procède à un suivi de contours.

Rappelons le codage de Freeman qui nous indique les directions en 8-connexité :



Ce que l'on veut faire c'est raccorder des points marqués appartenant à un même contour de cytoplasme ou de noyau. Pour cela on va partir des points terminaux marqués et on va suivre la direction du contour jusqu'à retomber sur un point terminal marqué, on aura alors refermé un contour.

On procède de la façon suivante :

Pour chaque point P de l'image qui est un point marqué :

(i) Déterminer la direction du contour à ce point, soit  $Dir(P)$  celle-ci donnée suivant le codage de Freeman.

(i i) Chercher dans l'image I2 un point voisin dans la direction du contour ( voir Figure II.14 ), si on en trouve plusieurs possibles prendre celui dont le niveau de gris dans l'image I1 est le plus élevé. Si le point trouvé est un point terminal marqué, on arrête car on a fermé un contour et on passe à un autre point marqué non exploré de l'image. Sinon le point trouvé devient le point courant et on retourne en (ii).

Configuration suivant les directions du contour

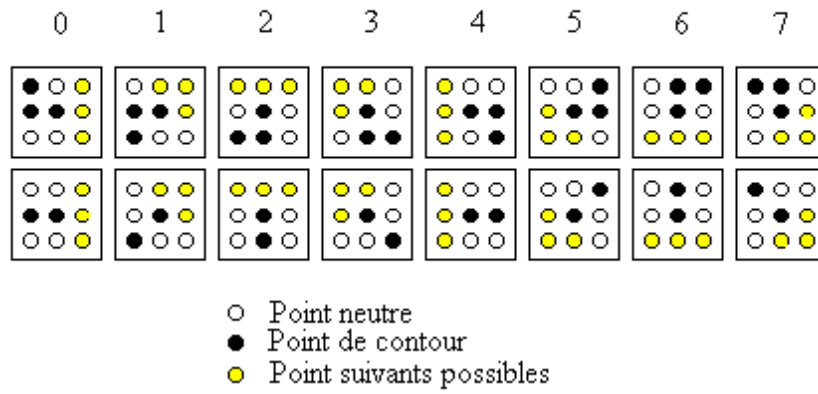


Figure II.14 :

Le point courant est celui du milieu de chaque masque. Suivant la direction du contour que l'on étudie on a plusieurs points possibles.



Figure II.15 : le contour du cytoplasme après suivi de contours.  
 Tout les points terminaux ont été connectés.

### 1.3.3.2 Deuxième méthode : Deux Lpe

Une deuxième approche du problème serait de se servir de la Lpe pour non pas caractériser sur une même image les contours des cytoplasmes et noyaux mais sur deux images différentes. La façon de le faire est très simple : au lieu d'avoir une image de marqueurs contenant des composantes connexes des cytoplasmes et des noyaux, on crée deux images de marqueurs : une contenant des composantes connexes des cytoplasmes et une autre des composantes connexes des noyaux. Ensuite on applique la Lpe contrainte par marqueurs avec ces deux images de marqueurs. On obtient alors deux images : une nous donnant les contours des cytoplasmes et une autre nous donnant les contours des noyaux.

Voici un exemple pour mieux comprendre.

On utilise les images de marqueurs suivantes ( figures II.16 )

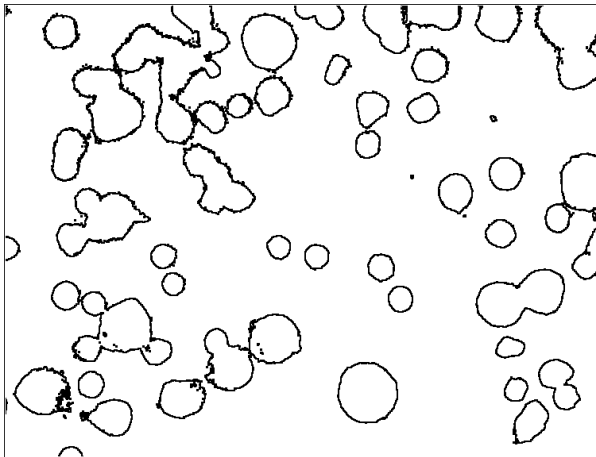


Figure II.16.a : les marqueurs du cytoplasme

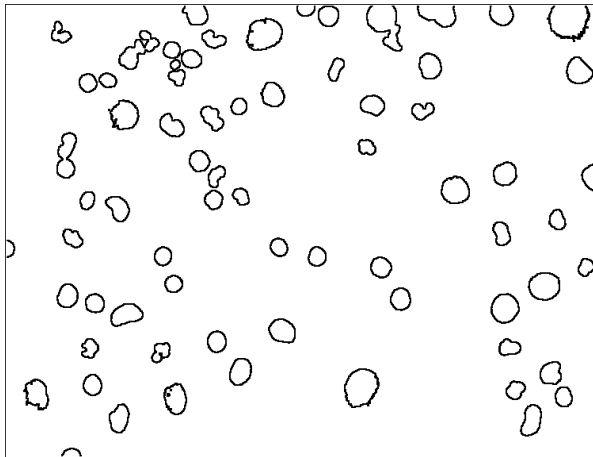


Figure II.16.b : les marqueurs des noyaux

Après avoir calculé la Lpe sur l'image gradient avec les marqueurs précédents, on obtient les deux images suivantes ( figures II.17 )

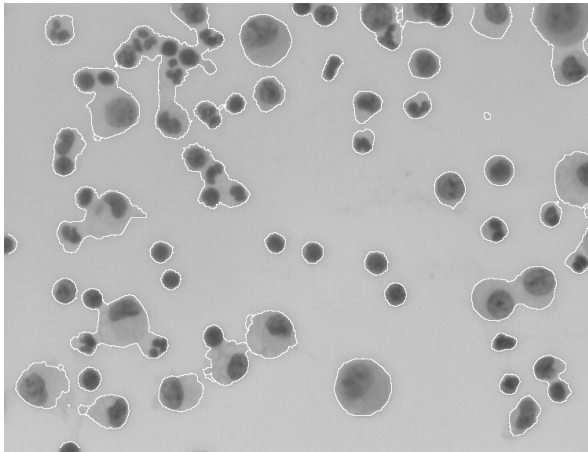


Figure II.17.a : les cytoplasmes par Lpe

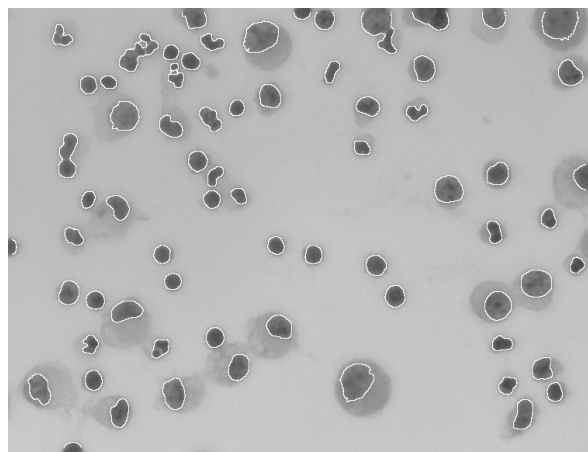


Figure II.18.b : les noyaux par Lpe



### 1.3.3.3 Comparaison des deux méthodes

La première méthode que nous avons utilisée a une grande faiblesse : les contours que l'on obtient ne seront pas obligatoirement fermés car aux points de jonction, on peut se tromper de direction dans le suivi de contour et ne pas obtenir le contour désiré (Figure II.18.a).

La deuxième méthode est par contre beaucoup plus sûre car elle permet d'avoir directement des contours fermés (Figure II.18.b).

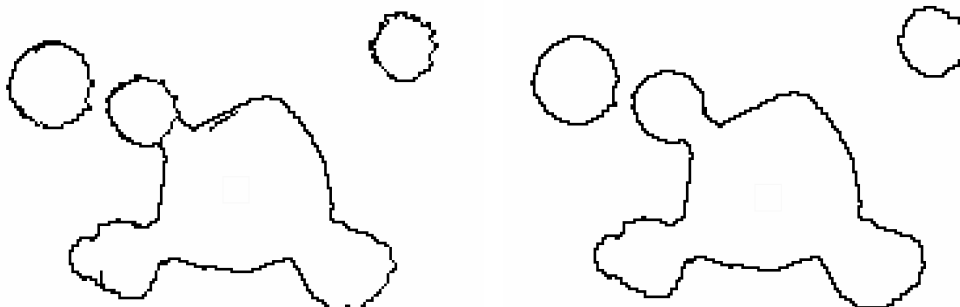


Figure II.18.a : le contour n'est pas parfaitement déterminé.

Figure II.18.b : le contour est exactement déterminé.

La méthode que nous avons retenue est donc la deuxième.

## 1.4 La séparation des amas

On dispose maintenant d'images nous donnant précisément les contours des cytoplasmes et des noyaux. Mais il subsiste un problème : certains cytoplasmes contiennent plusieurs noyaux, ce sont ce que l'on appelle des amas de cellules. Ce que nous nous proposons de faire ici est de séparer ces amas en plusieurs cellules. Par exemple pour l'amas de la figure II.18.b, on veut partitionner l'amas en quatre cellules.

### 1.4.1 Quelques notions de base

Pour comprendre ce qui va suivre, il nous faut introduire quelques notions.

- Erodés ultimes

Ce sont les composantes connexes obtenues par une érosion de taille  $n$  et qui disparaissent par une érosion de taille  $n+1$

- Image de distance

Etant donné une image contenant des points dit de référence, on appelle image de distance l'image telle que la valeur attribuée en tout point  $P$  est égale à la distance entre  $P$  et le point de référence le plus proche. Dans notre étude nous n'utiliserons que la distance euclidienne.

### 1.4.2 La méthode de séparation des amas

Dans une première approche nous avons tenté une méthode qui s'applique directement sur l'image de contours des noyaux des cellules. Cette méthode n'ayant pas donné de résultats appréciables, nous ne nous attarderons pas sur son explication, cependant nous pouvons en tirer quelques conclusions. En observant l'image que nous avons obtenue (Figure II.19), on s'aperçoit que les cellules qui intuitivement paraissent les plus faciles à séparer ne l'ont pas été alors que les amas plus complexes l'ont été. C'est pourquoi nous avons mis en place une méthode de séparation des amas prenant en compte le type des cellules à séparer.

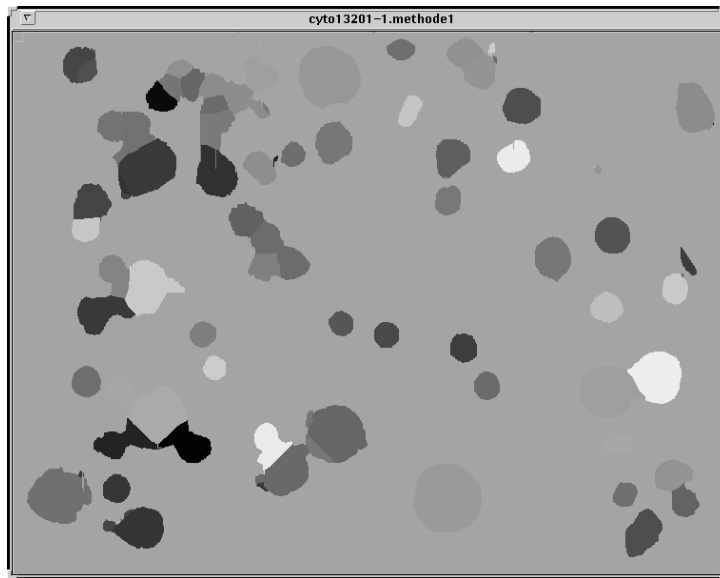


Figure II.19

Voici maintenant la méthode que nous avons mis en place, elle se compose de plusieurs étapes:

❶ On procède à un partitionnement de l'image. L'idée est que les cellules qui n'ont qu'un seul noyau n'ont pas besoin d'être traitées par une méthode de séparation des amas, seules les cellules ayant plus d'un noyau le doivent. On divise l'image des cytoplasmes ( Figure II.17.a) en trois sous images : les cellules ne possédant qu'un noyau ( Figure II.20.a), les cellules possédant plus d'un noyau ( Figure II.20.b) et les cellules qui sont coupées ( Figure II.20.c).

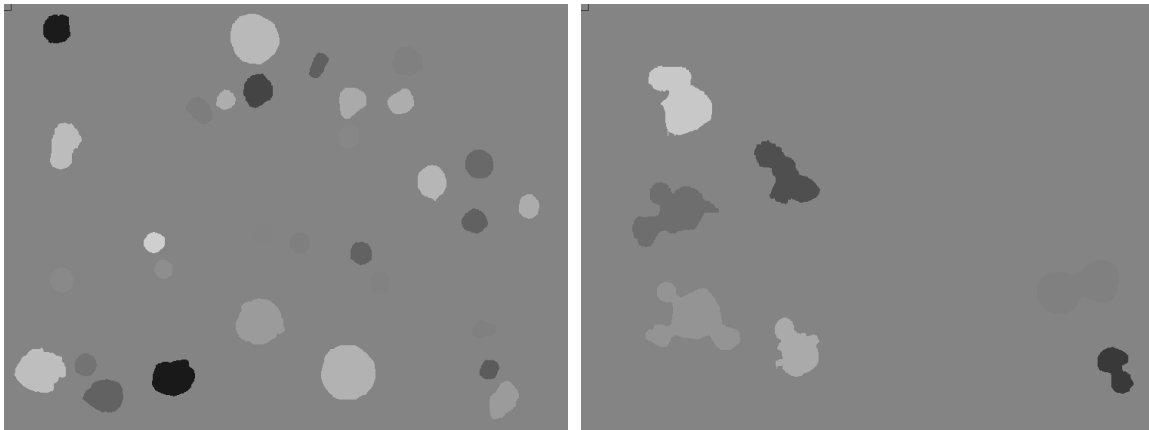


Figure II.20.a : les cellules à un noyau      Figure II.20.b : les cellules à plus d'un noyau



Figure II.20.c : les cellules coupées

Les cellules à un noyau n'ont pas besoin d'être traitées, on ne va traiter que les cellules à plus d'un noyau et les cellules coupées. Pour cela on va se servir de la Lpe contrainte par marqueurs appliquée à l'inverse de la fonction distance. Si l'on représente schématiquement le cas extrême d'une cellule que l'on veut séparer par la figure II.21.a, la fonction distance correspondante est alors donnée par la figure II.21.b. En prenant son inverse donné par la figure II.21.c, on voit qu'en procédant à une immersion par le bas (voir le § 1.3.3.2.1), on peut séparer la cellule en deux parties qui correspondent aux deux cellules contenues dans l'amas.

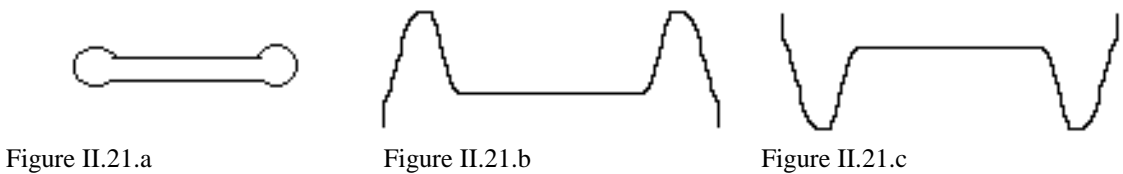
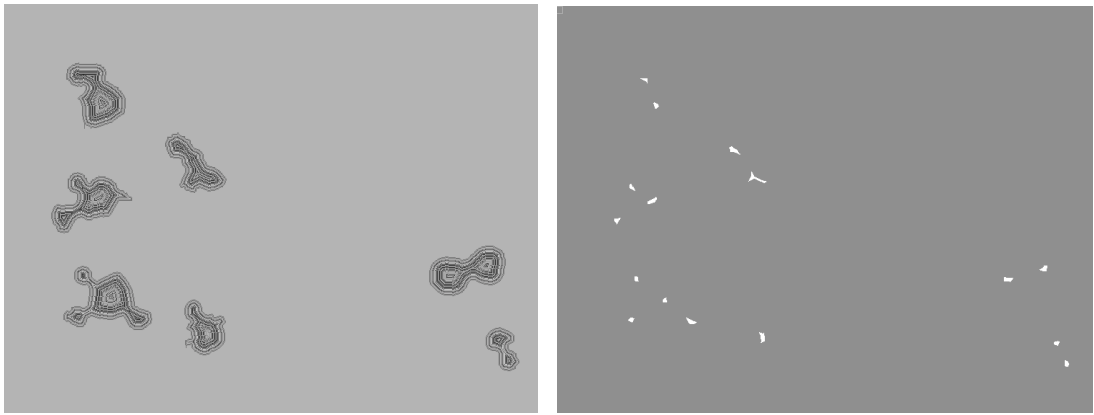


Figure II.21 : Illustration du principe de séparation

En plus de se servir de cette propriété de la fonction distance, nous allons procéder à une Lpe contrainte par marqueurs où les marqueurs seront les érodés ultimes de la fonction distance (ses minimas régionaux en quelque sorte). Voici le principe résumé :

- ❶ Calculer une image de distance de l'image des contours des cytoplasmes (Figure II.22),
- ❷ Calculer les érodés ultimes de l'image de distance ( Figure II.23),
- ❸ Calculer l'inverse de la fonction distance,
- ❹ Calculer la Lpe de l'inverse de la fonction distance avec les érodés ultimes comme marqueurs ( Figure II.24).



Sur les cellules qui ont plus d'un noyau

Figure II.22

Figure II.23

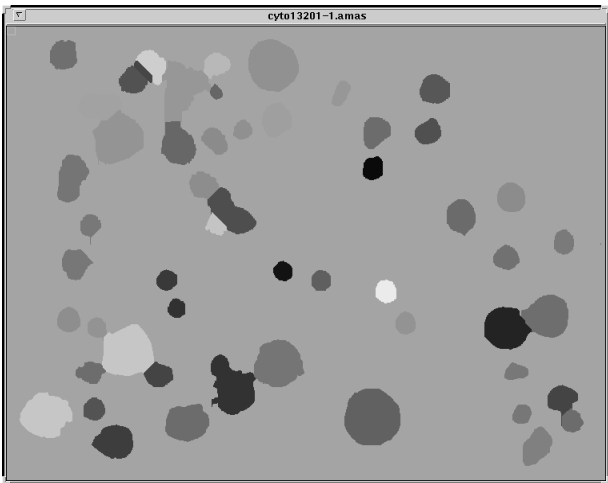


Figure II.24 : le résultat final avec toutes les cellules

Les résultats obtenus sont très concluants. Les amas sont bien séparés au niveau des étranglements. L'image finale est une image de régions.

## **1.5 Conclusion**

Nous avons réussi à segmenter nos images de façon très satisfaisante : les contours des cytoplasmes et des noyaux sont bien déterminés et les amas sont bien séparés. Les images finales dont nous disposons sont des images de régions : chaque cellule représente une région. Maintenant que nos images sont segmentées, nous pouvons passer à la partie caractérisation.

## 2. La caractérisation

Maintenant que nous disposons d'images correctement segmentées, on peut procéder à une analyse des caractéristiques des cellules que nous avons obtenues. Nous avons déjà vu au §A.2.4.1.2 quelles caractéristiques sont généralement employées dans les travaux de recherche.

Celles que nous avons choisies de calculer font partie des caractéristiques "classiques" mais ce sont celles qui nous paraissent les plus pertinentes pour modéliser numériquement nos cellules de façon correcte.

Les caractéristiques retenues pour une cellule sont les suivantes :

- ☞ La surface du cytoplasme ( SC ),
- ☞ La surface du ou des noyaux ( SN ),
- ☞ Le rapport SC / SN,
- ☞ Le périmètre du cytoplasme,
- ☞ L'élongation du cytoplasme,
- ☞ La moyenne des niveaux de gris du ou des noyaux,
- ☞ La variance des niveaux de gris du ou des noyaux.

Soit en tout 7 caractéristiques pour chaque cellule. Nous allons voir comment à partir des images dont nous disposons, ces caractéristiques ont pu être mesurées. On rappelle que l'on dispose d'images de régions et que les cellules et noyaux sont tous étiquetés avec pour chacun une étiquette différente.

❶ Les surfaces sont données simplement par le nombre de pixels appartenant au cytoplasme pour SC ou au(x) noyau(x) pour SN.

❷ Le périmètre est donné par le nombre de points se situant sur la frontière, il suffit de chercher dans nos images les points qui ont un voisin qui a une étiquette différente de la leur.

❸ La moyenne des niveaux de gris et la variance sont calculées grâce au niveau de gris des pixels des noyaux. Pour chaque cellule on calcule  $\sum X$  et  $\sum X^2$  où X représente le niveau de gris d'un pixel du noyau. La moyenne nous est donnée par :

$$\text{Moyenne} = \frac{\sum X}{\text{SN}} \quad \text{car SN donne bien le nombre de points du noyau,}$$

$$\text{Variance} = \frac{\sum X^2 - \frac{(\sum X)^2}{\text{SN}}}{\text{SN}}$$

❹ L'élongation calculée est l'élongation par rapport à l'axe des abscisses. Elle est donnée par la formule suivante :

$$\text{elongation} = \frac{\sqrt{M_{20} - M_{02} + 4 \times M_{11}}}{\text{SC}} \quad \text{avec } M_{pq} = \sum_i \left( \sum_j ((i-y_G)^p + (j-x_G)^q) \right)$$

$$\text{et } (x_G, y_G) = (M_{10}/M_{00}; M_{01}/M_{00})$$

$M_{pq}$  est le moment d'inertie d'ordre (p+q) d'une cellule par rapport à l'axe des abscisses.

En faisant varier (p,q) dans  $N^2$ , les moments  $M_{ij}$  décrivent une forme de façon unique.

Cette étape dans la construction d'un système de Tri Cellulaire Informatisé en Cytologie n'est pas des plus difficiles mais elle est des plus déterminantes pour la suite des traitements. Nous avons choisi des caractéristiques géométriques sur le noyau et le cytoplasme et des caractéristiques de texture sur le noyau. Elles nous paraissaient discriminer nos classes de cellules, mais si les valeurs qui sont extraites pour chaque cellule ne sont pas significatives et ne permettent pas de séparer correctement les classes de cellules dont nous disposons, il faudra choisir de nouvelles caractéristiques cellulaires à extraire. On pourra par exemple ajouter des mesures de convexité sur le noyau et/ou le cytoplasme, rajouter des critères de texture, des facteurs de forme sur le cytoplasme et/ou le noyau, et tout autre caractéristique permettant de séparer deux classes de cellules.

### 3. La classification

On dispose maintenant de données numériques sur chaque cellule, on peut donc passer à la troisième phase de notre étude : la classification. Mais avant toute chose, il va nous falloir étiqueter manuellement nos images.

#### 3.1 L'étiquetage des images

En effet pour que le réseau neuronal qui nous servira à la classification puisse fonctionner, il faut passer par une phase d'apprentissage supervisé. Il nous faut donc connaître le type de chaque cellule pour les images dont nous disposons. Pour cela il nous faut réaliser un logiciel permettant de visualiser les images et de les étiqueter en même temps.

Ce logiciel a été réalisé sous Unix en C++ avec la librairie Xview [Barkakati 92, Heller 90]. Nous allons voir les différentes possibilités qu'offre ce logiciel.

Une fois le programme lancé, on peut charger une image au format "rasterfile", la fenêtre suivante apparaît (figure II.25)



Figure II.25

Puis l'image apparaît dans une fenêtre (Figure II.26) et différents menus nous permettent de réaliser plusieurs actions : charger une autre image, étiqueter une image, segmenter une image et enfin faire des mesures de caractéristiques sur les images obtenues par segmentation.

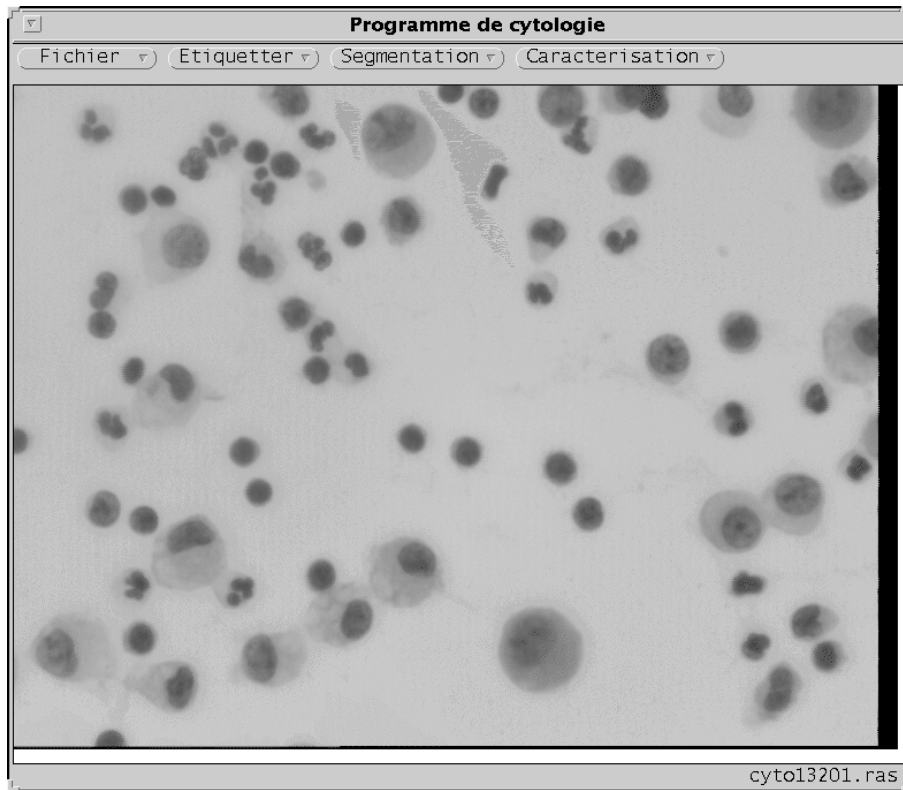


Figure II.26

Si l'on procède à un étiquetage, la fenêtre suivante (Figure II.27) apparaît nous demandant de choisir quel expert va étiqueter l'image. Les experts sont au nombre de trois :  
M. Michel Lecluse et Mme Annette Alexandre, cyto-techniciens, et le Docteur Hubert Elie tous trois du C.H.L.P.

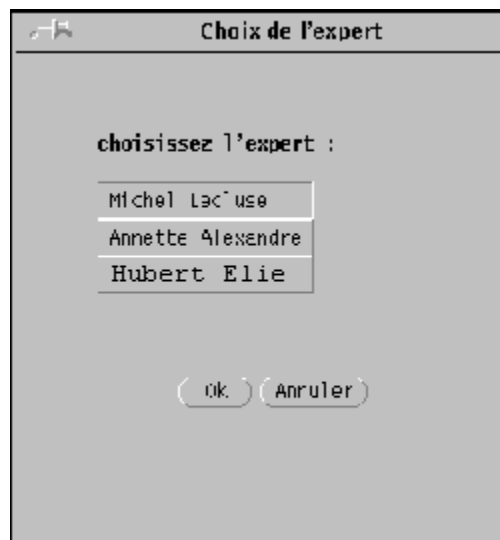


Figure II.27

Ensuite une fenêtre permettant de commencer la classification apparaît ( Figure II.28).

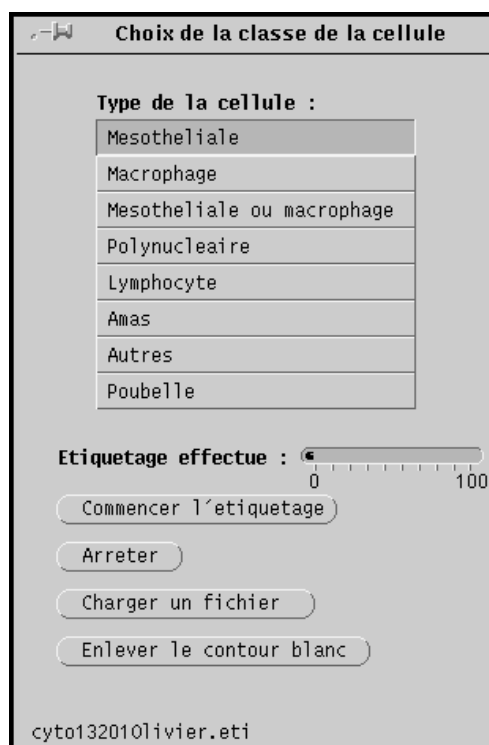


Figure II.28

On peut alors soit charger un fichier d'une image qui n'a pas été entièrement étiquetée, soit commencer l'étiquetage, soit arrêter. Si l'on commence l'étiquetage, la cellule à étiqueter est entourée en blanc sur l'image ( ce contour peut être enlevé ou rajouté pour faciliter la classification), il faut alors choisir la classe de la cellule, elles sont parmi Mésothéliale, Macrophage, Mésothéliale ou Macrophage, Polynucléaire, Lymphocyte, et trois autres classes Amas, Autres et Poubelle qui demandent à l'expert plus de précision sur la cellule : une fenêtre apparaît Figure II.29). Nous verrons plus tard pourquoi ces classes ont été choisies. Une jauge permet d'apprécier l'avancement de l'étiquetage.

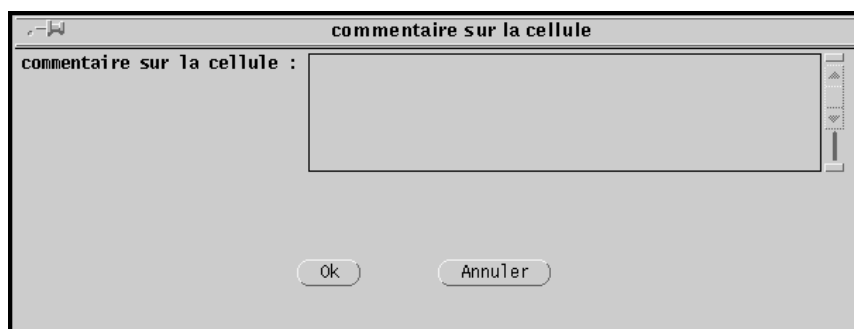


Figure II.29

L'étiquetage est sauvegardé au fur et à mesure dans deux fichiers, un pour les classes des cellules et un pour les commentaires éventuels sur les cellules.



### 3.2 Résultats de l'étiquetage

Nous avons étiqueté les quatre images sur lesquelles nous avons travaillé dans cette étude. Les résultats sont les suivants ( on donne le nombre de cellules par image et par classe pour chaque expert) :

Image	Expert	Meso	Macro	Meso ou Macro	Poly	Lymp	Amas	Autres	Poub	Nbre Total
Image1	A.A	17	3	2	12	12	5	1	4	56
	M.L	8	7	5	12	13	5	1	5	
	H.E	9	9	2	14	12	5	2	3	
Image2	A.A	2	5	0	20	19	5	0	1	52
	M.L	1	0	5	19	21	5	0	1	
	H.E	1	5	0	19	21	5	0	1	
Image3	A.A	11	19	7	46	62	34	11	1	191
Image4	H.E	19	17	11	51	69	38	20	16	241

Figure II.30 : les résultats de l'étiquetage par les experts  
 A.A : Annette Alexandre, M.L : Michel Lecluse, H.E : Hubert Elie

On voit que les polynucléaires et les lymphocytes sont majoritaires dans les images. Les images à fort grossissement (Image 1 et 2) contiennent peu d'amas contrairement aux images 3 et 4 obtenues à un grossissement plus faible de moitié environ. Les variations de jugement entre les experts se situent surtout au niveau de la distinction entre les cellules macrophages et mésothéliales. L'étiquetage des images 1 et 2 a été fait facilement par les experts car l'image n'est pas très difficile à juger à ce grossissement, mais les images 3 et 4 leur ont posé beaucoup plus de problèmes car le grossissement faible ne leur permettait pas un jugement très fiable.

On retiendra de cet étiquetage trois choses : la première est qu'il est meilleur de travailler à fort grossissement, la seconde est que nos images contiennent principalement des lymphocytes et des polynucléaires, la troisième est que certains amas ne sont pas encore très bien segmentés et il restera à étudier ces cas pour améliorer la segmentation.

### 3.3 Les réseaux neuronaux

Une fois que l'étiquetage des images a été effectué, on dispose pour chaque image de :  
 La classe de chaque cellule  
 Caractéristiques sur chaque cellule

Ceci peut constituer une base d'apprentissage à notre réseau neuronal. Avant de voir en détail l'application d'un réseau à notre cas, nous allons faire une brève présentation des réseaux neuronaux.

#### 3.3.1 Brefs rappels sur les réseaux neuronaux

Les réseaux neuronaux sont une modélisation de ce que la nature a produit de plus élaboré : notre cerveau. Chaque neurone est interconnecté avec d'autres neurones et l'information circule de neurones en neurones. Les réseaux de neurones sont organisés en "couches". On distingue trois couches : la couche d'entrée, la ou les couche(s) intermédiaires(s) et la couche de sortie (figure II.31). Chaque couche est reliée à la suivante par des synapses.

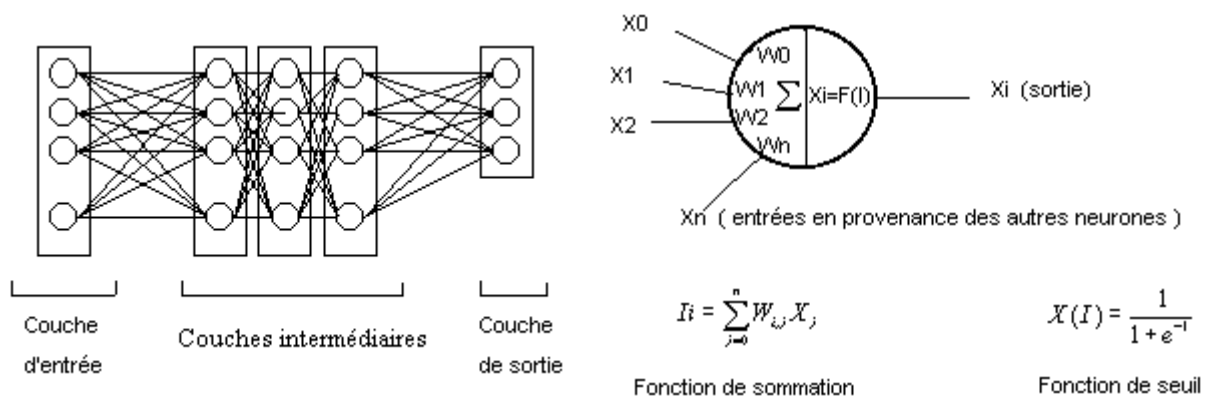


Figure II.31 d'après [Dreue 92]

Les neurones de la couche d'entrée représentent les données de la question et les neurones de la couche de sortie le résultat.

Chaque neurone possède une valeur ou activité (appelée  $X_i$ ) et chaque synapse possède un poids (appelé  $W_{ij}$  : la synapse part du neurone  $j$  et arrive au neurone  $i$ ). L'information circule dans le réseau des couches d'entrées aux couches de sortie. On distingue deux phases pour un réseau de neurones : la phase d'apprentissage supervisé et la phase de fonctionnement. Au début les poids de nos synapses et des activités des neurones sont fixées aléatoirement entre -1 et 1. Ensuite nous passons dans la phase d'apprentissage supervisé. On va apprendre au réseau à reconnaître les cas que l'on veut lui faire apprendre.

Pour cela on transmet nos valeurs aux neurones de la couche d'entrée qui agit à la manière de buffers et n'a d'autre rôle que de transmettre à la couche supérieure les données qui lui sont appliquées. Ces neurones d'entrée constituent ce que l'on appelle la "rétine". Le processus d'activation pour les neurones intermédiaire ou de sortie est calculé à partir des sommations pondérées des entrées ( fonction de sommation). Cette activation est ensuite transformée en sortie au travers d'une fonction sigmoïde. (fonction de seuil). Chaque connexion transmet sa valeur de sortie aux différents neurones de la couche supérieure ( Figure II.31). Cette phase constitue la phase de propagation des valeurs des entrées. Les valeurs que nous avons fixées aux neurones de la couche d'entrée ont donc permis à l'information de se propager le long du réseau.

On a donc théoriquement dans nos neurones de sortie le résultat de la question posée au réseau. Dans cette phase d'apprentissage supervisé on connaît à l'avance la question posée au réseau ainsi que la réponse, c'est ce qui permettra au réseau d'apprendre. En effet on compare ensuite la réponse obtenue avec la réponse souhaitée et l'on calcule l'erreur commise par le réseau. Ensuite on modifie les poids des synapses en fonction de l'erreur, cette étape est la rétro-propagation du gradient de l'erreur qui part des neurones de la couche de sortie aux neurones d'entrée. Chaque présentation d'une nouvelle question au réseau permet le changement progressif des poids des synapses. On continue l'apprentissage supervisé du réseau jusqu'à ce que le réseau converge vers une réduction minimale de l'erreur. Lorsque celle-ci a atteint un niveau satisfaisant, on peut considérer l'apprentissage comme achevé.

Ensuite peut commencer la phase de fonctionnement. Notre réseau a appris les cas que nous lui avons proposé et maintenant il peut fonctionner seul. On lui fournit les valeurs des neurones d'entrée et il les propage dans le réseau jusqu'aux neurones de sortie ce qui nous donne la réponse à la question posée.

Pour plus de détails on consultera [Kosko 92, Freeman 91, Pao 89]

### **3.2.2 Application à notre cas**

L'utilité d'un réseau de neurones ayant déjà été vue dans la Partie A de ce rapport, nous ne reviendrons pas sur ce point. Le réseau que nous devons mettre en place fonctionnera de la façon suivante :

Son but est de donner la classe des cellules qu'on lui propose. Il aura donc 7 neurones d'entrées, chacun de ces neurones correspondant aux paramètres qui ont été extraits lors de la phase de caractérisation. Ensuite un certain nombre de neurones et de couches intermédiaires transmettront la solution aux neurones de sortie qui sont au nombre de 8. Chaque neurone de sortie correspond à une classe de cellule, nous avons retenu 8 classes : Mésothéliale, Macrophage, Mésothéliale ou Macrophage, Polynucléaire, Lymphocyte, et Amas, Autres et Poubelle.

On reconnaît les quatre principales classes de cellules que l'on retrouve habituellement à savoir Mésothéliale, Macrophage, Polynucléaire et Lymphocyte. Nous avons rajouté la classe Mésothéliale ou Macrophage car dans certains cas il est très difficile de faire la différence entre ces deux classes de cellules. Lorsque le cyto-technicien procède à l'analyse d'une lame et a un doute entre ces deux types de cellules, il ne fait pas forcément de choix et la cellule est alors d'un type ou de l'autre. Il nous a paru important de modéliser cette démarche du cyto-technicien par une classe.

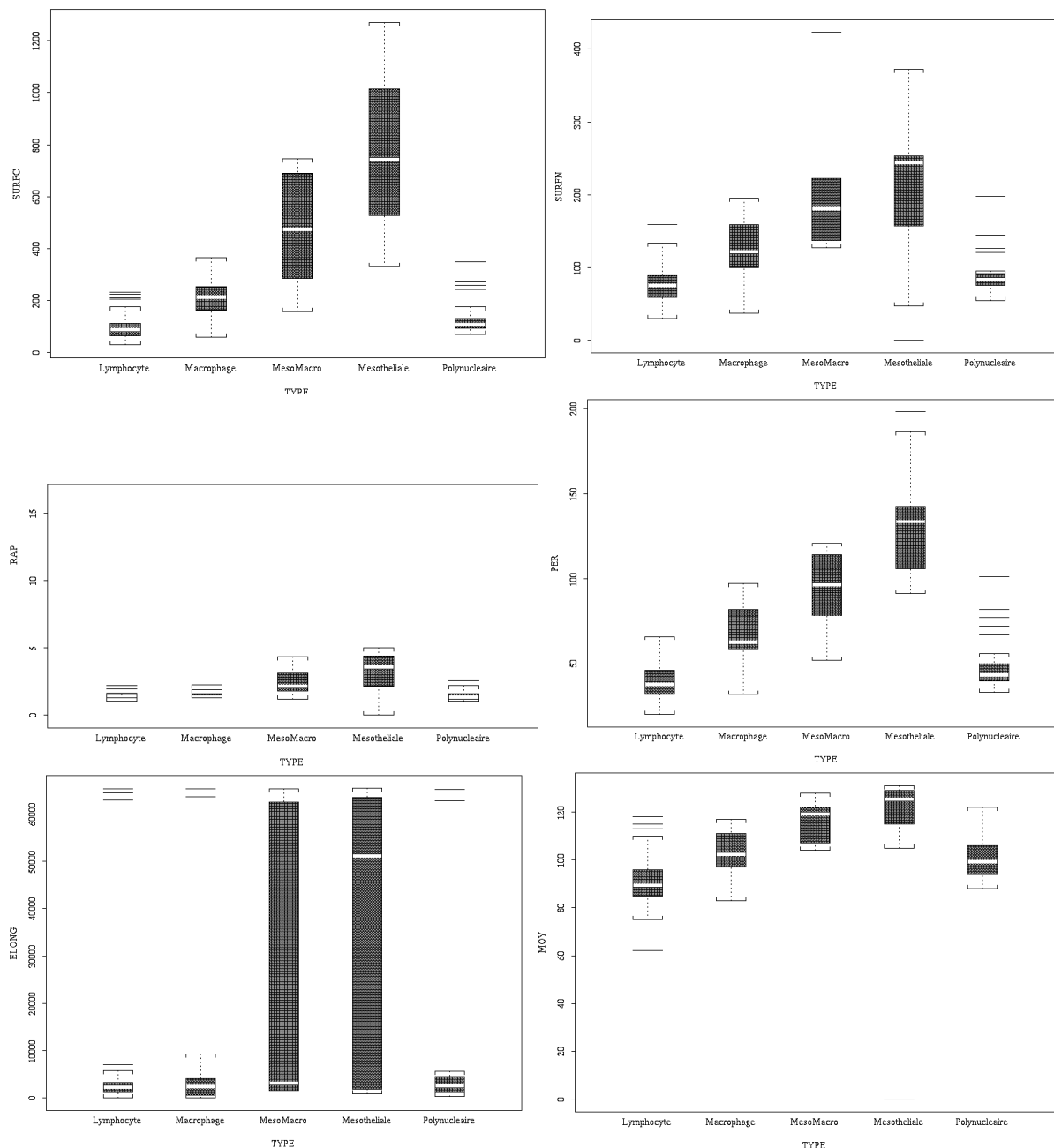
Ensuite trois autres classes apparaissent. La classe des "amas" englobe toutes les cellules polynuclées ou les attroupement de cellules, la classe des "autres" qui englobe les cellules anormales ou bien cancéreuses, les cellules inflammatoires et toutes autres cellules présentant une anomalie, la classe "poubelle" englobe tous les déchets qui peuvent venir de la préparation de la lame ou bien de la segmentation de l'image. Lors de l'étiquetage des images on demande pour ces trois dernières classes de commenter la cellule car ces cellules ont souvent des propriétés que l'on ne peut pas modéliser seulement par l'appartenance à une classe.

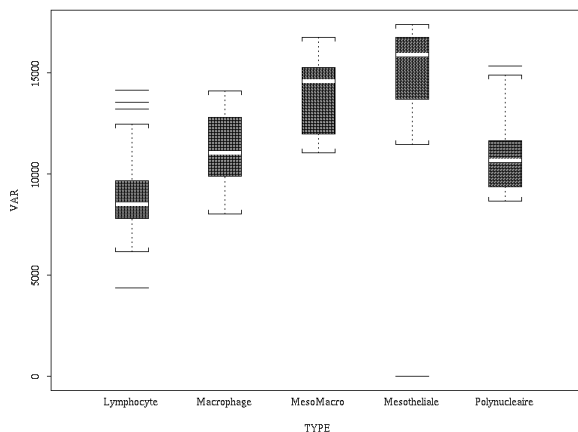
Le fait de posséder en plus d'une base d'étiquettes sur nos images, d'une base de commentaires sur les cellules les plus délicates pourra ultérieurement nous aider de façon significative pour une classification plus poussée.

### 3.2.3 Les résultats

Les essais de réseaux de neurones ont été faits avec le logiciel statistique S-PLUS [Baumgartner 94]. S-PLUS est un logiciel statistique permettant l'analyse de données, pour plus de détails, on consultera [Crochemore 95].

Avant de chercher à classifier nos cellules, nous avons étudié la répartition des différentes caractéristiques extraites pour chaque cellule. Nous n'avons fait les essais de classification que sur les classes suivantes : Mésothéliale, Macrophage, Mésothéliale ou Macrophage, Polynucléaire et Lymphocyte car les autres classes sont trop complexes et nécessiterons ultérieurement de modifier la segmentation des amas et d'éliminer les cellules qui ont une anomalie évidente risquant de fausser les résultats (cellules sans noyaux, résidus de la préparation de la lame etc.). Voici la répartition des caractéristiques des cellules de l'image 1 étiquetée par l'expert Mme Annette Alexandre (Figure II.32).





Figures II.32 : la répartition des caractéristiques

Au vu de ces répartitions on peut conclure que la surface du noyau, le périmètre du cytoplasme et la variance de niveaux de gris du noyau permettent déjà de séparer assez facilement les cellules mésothéliales ou macrophages des polynucléaires ou lymphocytes. Ces deux derniers types cellulaires ne se séparant pas vraiment pour toutes les caractéristiques dont nous disposons. On constate aussi que l'élongation par rapport à l'axe des abscisses n'est pas une caractéristique discriminante et qu'il nous faudra l'améliorer.

Les résultats que nous obtenons avec un réseau neuronal sont ceux attendus étant donné les répartitions des caractéristiques. Le réseau apprend avec une image étiquetée par un expert au plus gros grossissement et il teste sur une autre image étiquetée par le même expert au même grossissement. Les résultats sont les suivants ( on n'a gardé que trois classes : Lymphocyte, Mésothéliale ou Macrophage et Polynucléaire) :

Cellules Trouvées / Cellules Réelles	Lymphocytes	Mésothéliale ou Macrophage	Polynucléaire
Lymphocytes	11	0	1
Mésothéliale ou Macrophage	1	21	1
Polynucléaire	1	0	11

Cellules Trouvées / Cellules Réelles	Lymphocytes	Mésothéliale ou Macrophage	Polynucléaire
Lymphocytes	20	0	0
Mésothéliale ou Macrophage	1	3	3
Polynucléaire	9	1	11

Figure II.33.a : les résultats de l'apprentissage

Figure II.33.b: les résultats des tests

Soit 8,51% d'erreur de classification sur l'apprentissage et 29,78% d'erreur sur les tests.

Ces erreurs proviennent surtout de confusions entre les lymphocytes et les polynucléaires ainsi qu'entre les petites cellules mésothéliales ou macrophages et les polynucléaires.

Les résultats sont équivalents pour les images de plus faible grossissement. En apprenant sur un fichier étiqueté par un expert et en testant sur un fichier étiqueté par un autre, les résultats du réseau sont les suivants ( Figure II.34) :

Cellules Trouvées Cellules Réelles	Lymphocytes	Mésothéliale ou Macrophage	Polynucléaire
Lymphocytes	48	3	11
Mésothéliale ou Macrophage	2	29	6
Polynucléaire	11	3	33

Cellules Trouvées Cellules Réelles	Lymphocytes	Mésothéliale ou Macrophage	Polynucléaire
Lymphocytes	63	3	4
Mésothéliale ou Macrophage	7	40	0
Polynucléaire	26	1	24

Figure II.34.a : résultats de l'apprentissage Figure II.35.b : résultats des tests

Soit 24 % d'erreur sur les échantillons d'apprentissage et de test. Les erreurs étant les mêmes que pour des images à un grossissement plus fort, on peut donc conclure que cela ne vient pas du grossissement des images bien que le plus fort grossissement soit plus précis pour l'étiquetage.

### 3.2.4 Conclusion

Le réseau neuronal que nous avons mis en place n'est pas optimal et le taux d'erreur que nous avons est encore trop élevé pour espérer bien classer nos cellules. Mais cela peut être assez facilement corrigé car nous pouvons diagnostiquer l'origine des erreurs de classification du réseau. On voit que les cellules que le réseau a du mal à distinguer sont les lymphocytes et les polynucléaires ainsi que les petites cellules mésothéliales ou macrophages. Ceci vient du fait que le réseau ne dispose d'aucune caractéristique réellement discriminante entre ces deux classes de cellules. Une caractéristique pourrait facilement améliorer les résultats : la convexité du noyau des cellules. En effet le noyau des polynucléaires est polylobé (constitué de plusieurs lobes) alors que celui des lymphocytes est à peu près rond et que celui des macrophages est en forme de croissant de lune. Cette caractéristique est donc déterminante pour réussir à séparer correctement nos classes de cellules. Nous revenons donc au problème que nous avons évoqué dans la partie §B.2 : l'étape de classification ne peut être efficace que si l'étape de caractérisation nous donne des résultats permettant de bien répartir nos classes entre elles. Pour l'instant toutes les caractéristiques que nous avons extraites ne vérifient pas ce critère, l'élongation par exemple ne discrimine en rien, alors que la convexité du noyau permettrait une meilleure classification.

On se rend alors bien compte de la difficulté que représente l'étape de caractérisation et son impact sur l'étape de classification. Une suite à cette étude consistera donc à chercher des critères cellulaires propres à chaque type de cellule afin d'améliorer la classification.

---

## Bilan et perspectives

---

Nous nous sommes intéressés à la réalisation d'une maquette d'un système de Tri Cellulaire Informatique en cytologie. Les motivations de cette étude n'étaient pas de réaliser un système parfait qui arrive à trier correctement toutes nos cellules. Le but était de réaliser une maquette, c'est à dire de dégager les spécificités et de tenter une première conception du système afin d'en avoir la trame générale. Nous avons pu voir qu'aucun autre système ne s'est réellement intéressé aux images de liquides d'épanchement que nous avons étudiées. Par rapport aux motivations initiales, il nous a fallu renoncer aux images trop complexes car contenant trop d'amas de cellules et nous nous sommes restreints aux images en niveaux de gris.

☞ Les résultats de la segmentation sur les images sont bons : les cytoplasmes et les noyaux des cellules sont bien délimités grâce à la méthode la ligne de partage de eaux par marqueurs. Le fait d'avoir une méthode sûre et robuste permettant de bien délimiter nos contours cellulaires est déjà une assurance pour le système. La séparation des amas a donné des résultats assez satisfaisants même sur des cellules assez complexes. Les problèmes restant à résoudre se situent à ce niveau. Notre méthode de séparation des amas sépare les cellules qui ont plusieurs noyaux grâce à la forme du cytoplasme. Un amas de forme à peu près ronde ne sera donc pas séparé en plusieurs cellules. Il nous faudra donc améliorer ce point en réalisant une segmentation des amas plus robuste et plus adaptative,

☞ La classification n'a pas donné de résultats vraiment appréciables mais cela est dû au fait que les d'informations extraites à la phase de caractérisation ne sont pas toutes pertinentes. La faiblesse de notre système vient donc du manque de caractéristiques pertinentes. Nous avons pu en dégager quelques unes importantes telles que la surface et le périmètre du cytoplasme, mais il nous reste à trouver une ou plusieurs caractéristiques permettant de séparer les lymphocytes et les polynucléaires.

Enfin nous pouvons conclure que nos images peuvent être traitées par un système informatique du type de celui que nous avons conçu. Le bilan de notre étude est positif : nous avons pu trouver des méthodes donnant des résultats et nous savons exactement où se situent les problèmes dans l'élaboration d'un système de tri informatique en cytologie.

Les points qui restent à explorer sont les suivants :

☞ Pour trouver nos images de marqueurs servant à la ligne de partage de eaux, nous nous servons de seuils qui ont été déterminés manuellement. Il sera impératif que ces seuils soient calculés de façon automatique par le système,

☞ Améliorer la séparation des amas de façon "adaptative". Les cellules avec un noyau ne nécessitent pas de retraitement, par contre il faudra améliorer la séparation des cellules à plusieurs noyaux. Dans notre étude nous séparons les amas grâce à un critère de forme du cytoplasme. Dans certains cas cela ne sépare pas l'amas de façon satisfaisante et il nous faudra essayer de se servir des distances entre les noyaux de l'amas ainsi que de la taille des noyaux. Les difficultés seront alors de savoir quels critères nous permettent de savoir si un amas est suffisamment segmenté et comment répartir le cytoplasme entre plusieurs noyaux. Pour cela une collaboration avec les experts en cytologie sera absolument nécessaire,

☞ Trouver des caractéristiques cellulaires pertinentes, c'est-à-dire permettant de bien séparer nos cellules par un réseau de neurones. Encore une fois l'aide des experts en cytologie sera nécessaire, car ils pourront nous guider dans la recherche de ces caractéristiques étant données leur connaissances sur les cellules,

☞ Une fois que nous disposerons d'une segmentation et d'une caractérisation satisfaisantes, nous pourrons apprendre un maximum de cellules à notre réseau. Cette étape sera déterminante car plus les exemples seront bons, plus le réseau reconnaîtra de cellules. Un nombre de cellules important sera alors exigé, ce qui nécessitera d'avoir constitué au préalable une base de données d'image suffisamment représentative.

---

## REFERENCES

---

- [Antonangelo 94] Anatonangelo Leïla. Hilario Nascimento Saldiva P. Amaro E. Capelozzi V.L.  
"Utility of computerized morphometry combined with AgNor staining in distinguishing benign from malignant pleural effusions", *Analytical and Quantitative Cytology and Histology*, Vol 16, n° 4, 1994
- [Babes 28] Babes A.  
"Diagnostic du cancer du col utérin par les frottis", *Presse Med*, Vol 36, p 451, 1928,
- [Barkakati 92] Barkakati Naba  
"X Window programmation", *Editions Sybex*, 1992
- [Bartels 87] Bartels P.H. Graham A. Kuhn W. Wied G.L  
"Knowledge engineering in quantitative histopathology", *Applied Optics*, Vol 26, p3330, 1987
- [Baumgartner 94] Baumgartner M.  
"Une introduction à S-PLUS", *Ecole Polytechnique Fédérale de Lausanne*, 1994
- [Beale 1860] Beale L.S.  
"Examination of sputum from a case of cancer of the pharynx and adjacent parts", *Arch Med*, Vol22, p 44, 1860
- [Boon 93] Mathilde E. Boon. Kok L.P.  
"Neural network processing can provide means to catch errors that slip through human screening of pap smears", *Diagnostic Cytopathology*, Vol 9, n° 4, 1993
- [Brugal 87] Brugal B.  
"Required facilities for image analysis at the microscope in biological and medical applications : the SAMBA image processor, clinical cytometry and histometry", *Academic Press*, 1987
- [Clouard 94] Clouard Régis  
"Raisonnement incrémental et opportuniste appliqué à la construction dynamique de plans de traitement d'images", *Thèse de l'université de Caen*, 1994
- [Clouard 95] Clouard Régis  
"Horus 1.0. Guide d'utilisation, Manuel de référence, Manuel de programmation", GREYC, 1995
- [Coster 89] Coster Michel. Chermant Jean-Louis  
" Précis d'analyse d'images ", *Editions du CNRS*, Paris 1989



- [Crochemore 95] Crochemore Eric  
 "Etudes d'indices morphologiques et texturels. Comparaison et étude de classifieurs. Application aux images aériennes", *Rapport de stage de DEA*, Université de Caen, 95
- [Derder 95] Derder Adel,  
 "Système de support et de supervision des activités en cytologie", *Thèse de l'Institut National de Polytechnique de Grenoble*, These IMAG, 95
- [Donné 1845] Donné A.  
 "Cours de microscopie élémentaire des études médicales d'anatomie microscopique et physiologie des fluides", *J.B Baillère*, Paris, 1845
- [Dufer 93] Dufer J. Liautaud-Roger F., Barbarin D. Coninx P.  
 "Nucleus image analysis as a possible prognostic tool in grading breast cancer", *Biomed& Pharmacother*, Vol 47, p 131-135, 1993
- [Dufour 04] Dufour H.M.  
 "Meningite sarcomateuse diffuse avec envahissement de la moelle et des racines : cytologie positive et spéciale du liquide céphalo-raphidien," *Rev Neurol Soc Neurol*, Vol 12, p 204, 1904
- [Dytch 87] Dytch H.E. Bibbo M. Bartels P.H. Puls J.H. Wied G.L.  
 "A PC-based system for the objective analysis of histologic specimens through quantitative contextual karyometry", *Applied Optics*, Vol 26, n° 16, p 3270, 1987
- [Dytch 90] Dytch H.E. Wied G.L.  
 "3 Artificial neural networks and their use in quantitative pathology ", *Analytical and Quantitative Cytology and Histology*, Vol 12, n°6, p 379-393, 1990
- [Elmoataz 90] Elmoataz-Billah Abderrahim  
 " Mécanismes d'opérateurs d'un segmenteur d'images non dédié : définition d'une base d'opérateurs", *Thèse de l'université de Caen*, 1990
- [ENSTAC 87]  
 "Cours de cyto-pathologie" , Ecole nationale supérieure de techniques anatomo et cyto-pathologiques, *Université de Paris VII*
- [Erhardt 80] Erhardt R. Reinhart E.R Shipf W. Bloss W.H  
 "FAZYTAN : a system for fast automated cell segmentation, cell image analysis and feature extraction, based on T-V image pickup and parallel processing", *Analytical and Quantitative Cytology and Histology*, Vol 2, p 25-40, 1980
- [Freeman 91] Freeman J.A. Skapura D.M  
 "Neural networks : algorithms, applications and programming techniques", *Addison- Wesley*, 1991

- [Giménez 95] Giménez-Mas J.A. Sanz-Moncasi M.P. Remon L. Gambo P.  
 "Automated textural analysis of nuclear chromatin : a mathematical morphology approach", *Analytical and Quantitative Cytology and Histology*, Vol 17, n° 1, Février 1995
- [Hall 87] Hall Thomas L. Keyani D.H.L. Rosenthal D.L.  
 "Microcomputer-based image processing workstations for cytology", *Applied Optics*, Vol 26, n° 16, p 3266, 1987
- [Heller 90] Heller Dan  
 "Xview programming manuel", *Editions O'Reilly & Associates*, Juillet 90
- [IAC 84] International Academy of Cytology  
 "Specifications for automated cytodagnostic systems proposed by the I.A.C", *ActaCytologica*, Vol 28, p 352, 1984
- [Jackway 95] Paul Jackway  
 "Morphological multiscale gradient watershed image analysis", *The 9th Scandinavian Conference on Image Analysis*, Uppsala, Sweden, Juin 1995
- [Jaurand 90] Jaurand M-C. Bernaudin J-F. Bignon J.  
 "La cellule mésothéliale", *La revue du praticien*, Vol 40, n° 20, p 1823-1828, 1990
- [Kasdan 87] Kasdan Harvey L.  
 "High performance pathology workstation using an automated multispectral microscope", *Applied Optics*, Vol 26, n° 16, p 3295, 1987
- [Kate 93] Kate T.K.T. Beliën J.A.M. Smeulders A.W.M. Baak J.P.A.  
 "Method for counting mitoses by image processing in feulgen stained breast cancer sections", *Cytometry*, Vol 14, p 241-250, 1993
- [Kosko 92] Kosko Bart  
 "Neural networks and fuzzy systems : a dynamical systems approach to machine intelligence", *Prentice-Hall*, 1992
- [Koss 89] Koss L.G.  
 "The Papanicolaou test for cervical cancer detection, a triumph and a tragedy", *J. Am. Med. Assoc.*, Vol 261, p737-743, 1989
- [Koss 93] Koss L.G.  
 "Cervical Pap smear", *Cancer*, 1993
- [Koss 94] Koss L.G.  
 "Evaluation of PAPNET cytologic screening system for quality control of cervical smears", *American Journal of clinical pathology*, Vol 101, n° 2, Fev 1994
- [Lebert 1845] Lebert H.  
 "Physiologie et pathologie ou recherches cliniques, expérimentales et microscopiques", *J.B Baillière*, Paris, 1845

- [Mango 93] Mango L.J.  
"Aide au dépistage du cancer du col par ordinateur à réseaux neuraux", *Cancer Letters*,1993
- [Marissi 09] Marissi G.  
"Ueber die diagnose des magencarcinomas auf Grund der Cytologie des Spuelwasser", *Arch Verdauungskrankh*, vol 15, p 251, 1909
- [Masson 92] E. Masson.  
"Cytométrie automatisée par traitement numérique d'images : application à la caractérisation et à la classification des cellules épithéliales normales et tumorales",*Thèse de l'université de Caen*, 1992
- [Menard 93] Menard Michel.  
"Développement d'algorithmes et d'un système de traitement associé, pour l'aide au diagnostic d'images de cellules épithéliales en vue d'examens en routine clinique",  
*Thèse de l'université de Poitiers*, 1993
- [Moragas 93] Moragas A Garcia-Bonafé M. De Torres I. Sans M.  
"Textural analysis of lymphoid cells in serous effusions : a mathematical morphologic approach",  
*Analtical and Quantitative Cytology and histology*, Vol 15, n° 3, 1993
- [Mueller 1838] Mueller J.  
"Ueber den feineren Bau und die Formen der krankenhaften geshwuelste", *G Reimer*, Berlin 1838
- [Nordin 89] Nordin B.  
"The development of an automated prescreener for the early detection of cervical cancer: Algorithms and Implementation", *Ph D Thesis*, Upsalla University, Sweden,1989
- [Pao 89] Pao Yoh-Han  
"Adaptative pattern-recognition and neural networks", *Addison-Wesley*, 1989
- [Papanicolaou 28] Papanicolaou G.N.  
"New cancer diagnosis", *Proc third race betterment conference*,p 528, 1928
- [Papanicolaou 47] Papanicolaou G.N. Traut HF.  
"Diagnosis of Uterine Cancer by the Vaginal Smear" , *Commonwealth Fund*, NewYork, 1947
- [Papanicolaou 54] Papanicolaou G.N.  
"Atlas of Exfoliative Cytology", *Harvard University Press*, Cambridge, Massachusetts,1954

- [Ploem 87] Ploem J.S. Tanke H.J.  
 "Introduction to fluorescence microscopy", *Roy. Micr. Soc*, Microscopy handbooks  
 10,Oxford, University Press, 1987
- [Pouchet 1847] Pouchet F.A.  
 "Théorie positive de l'ovulation spontanée et de la fécondation des mamifères et de l'espèce  
 humaine", *J.B Baillière*, Paris, 1847
- [Prewitt 66] Prewitt J.M.S. Mendelsohn M.L  
 "The analysis of cell images", *Ann. N.Y. Acad. Sc.*, Vol 128, p 2035, 1966
- [Rasband 90]  
 "Wayne Rasband Image Software", *National Technical Information Software*, Springfield
- [Revision 67]  
 "Revision of laboratory requisitions for the medicara, medicaid and clinical  
 laboratoriesimprovement act of 1967 ( CLIA'67) programs", *Federal Register*, Vol 5, p 9600,  
 1967
- [Ruffé 90] Ruffé P.  
 "Traitement des tumeurs pleurales", *La revue du praticien*, Vol 40, n° 20, p 1851-  
 1853,1990
- [Sanders 1864] Sanders W.R.  
 "Cancer of the bladder : fragments forming urethral plugs discharged in the urine",  
*Med J*, Edinburgh, Vol 1, p 273, 1864
- [Serbouti 93] Serbouti S. Harms H. Gunzer U. Aus H.M. Beuscart R. Mary J-Y.  
 "Image Analysis and classification methods for leukemias", *Innov. Tech. Biol. Med.*,  
 Vol 14, n° 4, 1993
- [Sherman 94] Mark E.Sherman.  
 "Papnet analysis of reported negative smears preceding the diagnosis of a high grad squamous  
 intraepithelial lesion or carcinoma", *Modern Pathology*, Vol 7, n° 5, p 578, 1994
- [Schmitt 94] Schmitt M. Mattioli J.  
 "Morphologie mathématique", *Masson*, 1994
- [Tanaka77] Tanaka Noburu. Hideo Ideka. Tetsuo Ueno. Akio Mukawa. Sadakazu Watanabe  
 Kozo Okamoto. Sugao Hosoi. Sho Tsunekawa.  
 "Fundamental study of cyto-screening for uterine cancer : new system for  
 automatedapparatus ( CYBEST) utilizing pattern recognition method",  
*Acta Cytologica*,Vol 21, p85, 1977

- [Tanaka 87] Tanaka Noburu.Hideo Ideka. Tetsuo Ueno. Akio Mukawa. Sadakazu Watanabe Kozo Okamoto. Sugao Hosoi. Sho Tsunekawa.  
"Automated cytologic screening system ( CYBEST model 4) an integrated image cytometry system", *Applied Optics*, Vol 26, n° 16, p 3301-3307, 1987
- [Takahashi 83] Takahashi M.  
"Atlas en couleur de cytologie du cancer", deuxième édition, *VIGOT*, 83
- [Taylor 78] Taylor J. Bartels P.H. Bibbo M. Wied G.L.  
"Automated hierarchic decision structures for multiple category classification byTICAS", *Acta Cytologica*, Vol 22, n° 4, 1978
- [Teague 93] Teague T.K. Munn L. Zygourakis K. McIntyre B.W.  
"Analysis of lymphocyte activation and proliferation by video microscopy and digitalimaging", *Cytometry*, Vol 14, p 772-782, 1993
- [Tolles 55] Tolles W.E.  
"The cytoanalyser : an example of physics in medical research", *Trans. N.Y Acad. Sc.*, Vol 17, p 250, 1955,
- [Truong 95] Hanh Truong. Morimoto R. Walts A.E. Marchevsky A.  
"Neural network as an aid in the diagnosis of lymphocyte-rich effusions", *Analytical and Quantitative Cytology and Histology*, Vol 17, n°1, Février 1995
- [Tucker 82] Tucker J.M. Shippey G.A. Farrow S. Bayley R. Stark M.  
"CERVIFIP : an automated cytologic prescreening system using a linear diode array", *Analytical Quantitative Cytology*, Vol 4, p 160, 1982
- [Visilog 88] Visilog User's guide  
"A introduction to Visilog computer Vision Software", *Noesis*, Versailles, France, 88
- [Wied 92] Wied G.L.  
"History of clinical cytology and outlook for its future", *Compendium on Diagnostic Cytology*, AIC 92
- [Wied 68] Wied G.L. Bartels P.H. Bart G.F. Oldfield D.G.  
"Taxinomic intra-cellular Analytic System (TICAS) for cell identification", *ActaCytologica*, Vol 12, p 180-204, 1968
- [Wied 89] Wied G.L. Bartels P.H. Bibbo M. Dytch H.E  
"Image analysis in quantitative cytopathology and histopathology", *Human Pathology*, Vol 20, n° 6, 1989
- [Yahiaoui 92] Yahiaoui G. Azmy N. Vibert J-F.  
" Réseaux de neurones biomimétiques : applications potentielles en traitement d'images", *Science Technique Technologie*, n°20, 1992
- [Zanisher 82] Zanisher D.J. Oud P.S. Raaijmakers M.C.T. Vooijs G.P. Van de Walle R.T  
"the BioPEPR system for cevical cytology automation", *Analytical Quantitativecytology*, Vol 4, p 164, 1982