

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab, Blida
USDB.

Faculté des sciences.
Département informatique.

**Mémoire pour l'obtention
D'un diplôme d'ingénieur d'état en informatique.**
Option : IA

Sujet :

**Classification connexionniste d'images couleurs :
Application aux images cytologiques**

Présenté par : Djeddi Omar

Promotrices : Mlle N. Benblidia et Mlle F.Z. Reguieg

Organisme d'accueil : LRDSI, université de Blida



-Promotion 2005-2006-

MIG-004-144-1



Remerciements

Je remercie avant tous Dieu tous puissant qui m'a donné la force et le courage pour réalisé ce modeste travail.

Je profite l'occasion de remercier notre chef de département M^r MESSIAD pour son aide particulière aux étudiants de fin de cycle.

Aussi, je remercie, M^{elle} djazouli faculté de biologie, pour son aide

Je remercie a tous ceux que d'une manière ou d'une autre, m'ont apporté de l'aide durant tous notre cursus.

je tien a remercier vivement mes promotrices M^{elle} BENBLIDIA et M^{elle} REGUIEG d'avoir accepté de m'encadré et proposer ce sujet, j'exprime tout ma gratitude et mes sincère remerciement pour tous leurs effort, coopération, aide, patience, encouragement et vaillance pour la réalisation ce travail, prie Dieu de leurs donné toute la force pour continué leurs noble mission.

Aussi, je remercie tous les enseignants du département d'informatique, pour m'avoir formé tout au long de notre cursus, et j' espère pouvoir être à la hauteur de leur espérance dans le futur proche.

Dédicace

Je dédie ce modeste travail à la mémoire de
ma grand mere
A Ma très chère Parents surtout ma mère, et
je les
remercie pour leur soutien pendant $\frac{1}{4}$ de siècle.

A mes sœur fella et Naziha.

A la mémoire de notre amie Djaouan
Nesredine.

A mes chère amies, ramzi poisson, raouf,
mounir, réadh, Ahmed, Med Z, Nouredine,
salim d sans oublié Panucci

A mes chère amies de promotion surtout le
groupe 14d10, commençants par Abdelrezak ,
Mahdi, Mohamed, Morad et younes, sans
oublier the Red, Yahia ,Ibrahim, Hamza,
Omar, Sofian, nouni, Redha,

A tous ceux que je n'est pas cité par oublie,
qu'ils me pardonne

PFE 2006

SOMMAIRE

Introduction générale.....	1
Chapitre I Introduction à la vision humaine couleur et à la cytologie	
I Introduction.....	3
II Les modèles de la vision humaine et les espaces couleurs.....	4
II.1 les espaces couleurs.....	5
III Introduction à la cytologie.....	10
III.1 L'anatomie pathologique.....	10
III.2 Structure anatomique et cytologique.....	11
III.3 Les prélèvements en anatomie et cytologie pathologiques.....	12
III.4 Les colorations.....	13
III.4.1 Types cellulaires en cytologie.....	14
III.5 Le screening.....	17
IV Conclusion.....	20
Chapitre II Segmentation des images couleurs par la morphologie mathématique	
I Intro.....	21
II Principales méthodes de segmentation d'images couleurs.....	21
III Outils principaux de la morphologie mathématique.....	24
IV La ligne de partage des eaux.....	27
IV.1 Définition par immersion.....	27
IV.2 La L.P.E contrainte par marqueurs.....	28
IV.3 La ligne de partage des eaux couleur.....	29
IV.3.1 Définition de la fonction d'agrégation couleur.....	32
IV.3.2 Présentation de la méthode de segmentation couleur.....	32
V Conclusion.....	44
Chapitre III Caractérisation morphologique et classification connexionniste des images cytologiques	
I Caractérisation des zones cellulaires.....	45
I.1 Description des paramètres morphologiques.....	45
I.2 Rapport nucléo-cytoplasmique.....	46
I.3 Hyperchromatisme nucléaire.....	50

II. Classification connexionniste	52
II.1 Le Neurone biologique, structure et fonctionnement.....	52
II.2 Le neurone formel.....	54
II.3 Configuration des réseaux de neurones.....	55
II.4 Apprentissage des réseaux de neurones.....	57
III. Carte Auto organisatrice de Kohonen.....	58
IV. Classification par carte auto organisatrice.....	61
VI.1 Apprentissage.....	63
VI.2 Classification.....	64
V. Conclusion.....	65
Chapitre IV Elaboration et présentation de L'application	
I. Introduction.....	66
II. Présentation de l'application Morphocyt.....	66
II.1 Interface du logiciel.....	67
II.2 Les menus.....	67
II.3 Barre des boutons.....	74
III. Présentation des résultats et discussion.....	74
IV. Conclusion.....	84
Conclusion Générale	
Bibliographie	

FigureI.1 :	Chaîne de la V.A.O	3
FigureI.2 :	Le spectre des couleurs visibles	4
FigureI.3 :	La représentation spatiale du modèle RVB	5
FigureI.4 :	Variation du tri stimulus XYZ en fonction de la longueur d'onde.....	6
FigureI.5 :	La représentation spatiale du modèle CIE L*a*b*	8
FigureI.6 :	La représentation spatiale du modèle HSL	9
FigureI.7 :	Représentation schématique de la cellule.....	11
FigureI.8 :	Schéma des activités séreuses du corps humain.....	13
FigureI.9 :	Quelques types cellulaires en cytologie des séreuses. Les cellules peuvent être isolées ou en amas.....	16
FigureI.10 :	Le screening manuel sur une lame	17
FigureII.1 :	Procédé du traitement par l'utilisation des espaces couleurs.....	22
FigureII.2 :	Erosion de (a) par un élément structurant circulaire.....	24
FigureII.3 :	Erosion de (a) par un élément structurant circulaire.....	25
FigureII.4 :	Ouverture d'un ensemble d'objets X par un élément structurant circulaire...	26
FigureII.5 :	Fermeture d'un ensemble d'objets X par un élément structurant circulaire...	26
FigureII.6 :	Barrages aux endroits où l'eau provenant de deux minima se mélangerait...	28
FigureII.7 :	Image segmentée par la croissance des régions.....	29
FigureII.8 :	Algorithme de croissance de région.....	30
FigureII.9 :	Les différentes configurations spatiales des cellules.....	34
FigureII.10 :	Stratégie de segmentation d'images de cytologie des séreuses.....	35
FigureII.11 :	Division par une image de fond.....	36
FigureII.12 :	Exemple d'extraction des cellules	37
FigureII.13 :	Cellules contenant des globules rouges et le masque correspondant	38
FigureII.14 :	L'image a et sont masque correspondant	39
FigureII.15 :	Les marqueurs (a), le gradient (b), les marqueurs (c) et les régions cytoplasmiques en bleu (d).....	42
FigureII.16 :	Les marqueurs (a), le gradient (b), les marqueurs (c) et les régions nucléaires en jaune (d).....	44
FigureIII.1 :	coupe de tissu cancéreux.....	46
FigureIII.2 :	Exemple d'image à traiter.....	46
FigureIII.3 :	Etapes d'isolation des noyaux.....	48
FigureIII.4 :	Image du cytoplasme.....	50
FigureIII.5 :	Image des noyaux.....	51
FigureIII.6 :	Image indique la représentation. de l'hyperchromatisme nucléaire.....	52
FigureIII.7 :	Le neurone biologique.....	53
FigureIII.8 :	Mise en correspondance : neurone biologique/neurone artificiel.....	54
FigureIII.9 :	Le neurone artificiel.....	55
FigureIII.10 :	Réseau à couches non bouclé.....	56
FigureIII.11 :	Un réseau de neurones bouclé à deux entrées.....	57
FigureIII.12 :	Architecture du réseau de Kohonen montrant la relation entre le vecteur d'entrée, les poids des connexions W_{ij} et le vecteur de sortie y	59
FigureIII.13 :	principe général de la reconnaissance.....	62
FigureIII.14 :	Principe de la classification par le réseau de Kohonen	62
FigureIII.15 :	Algorithme d'apprentissage par le réseau de Kohonen	64
FigureIV.1 :	La fenêtre principale du MorphoCyt.....	67
FigureIV.2 :	Menu fichier.....	67
FigureIV.3 :	Menu segmentation par morphologie mathématique	68
FigureIV.4 :	Sous menu prétraitement	68
FigureIV.5 :	Sous menu filtre numérique.....	69

FigureIV.6 :	Sous menu Binarisation.....	69
FigureIV.7 :	Sous menu gradient	69
FigureIV.8:	Sous menu Morphologie mathématique	70
FigureII.9 :	Sous menu Segmentation	70
FigureIV.10 :	Sous menu marqueur.....	71
FigureIV.11 :	Sous menu espace couleur	71
FigureIV.12 :	Sous menus paramètres	72
FigureIV.13:	Sous menus reconnaissance	72
FigureIV.14 :	Menu outils	73
FigureIV.15 :	La fenêtre à propos.....	73
FigureIV.16 :	La barre des boutons.....	74
FigureIV.17 :	Une image cellulaire.....	75
FigureIV.18 :	Représentation de résultat des différents filtrages sur une image cellulaire..	75
FigureIV.19 :	Détection des cellules d'intérêt	76
FigureIV.20 :	Représentation de l'image cellule l dans différents espaces de couleur.....	77
FigureIV.21 :	Segmentation des cytoplasmes.....	78
FigureIV.22 :	Segmentation des noyaux	79
FigureIV.23 :	Calcul de rapport nucléo-cytoplasmique	80
FigureIV.24 :	L'hyperchromatisme nucléaire	80
FigureIV.25 :	Apprentissage connexioniste des cellules.....	81
FigureIV.26 :	Cellule a classé.....	83
FigureIV.27 :	Classification des cellules.....	83

Liste des tableaux

Tableau II.1 :	Les contrastes suivant les différents fonds des images.....	40
Tableau II.2 :	Les contrastes suivant les différents fonds des images (noyaux).....	43
Tableau III.1 :	Les paramètres calculer des images testes.....	82

Résumé

Dans ce travail, nous avons mis en œuvre un système d'aide au diagnostic cytopathologiste. Son but est d'extraire les cytoplasmes et les noyaux dans le but d'une analyse quantitative pour une classification connexionniste non supervisée des cellules suivant le modèle de Kohonen.

L'extraction de l'information pertinente, est effectuée suivant les opérations de la morphologie mathématique en se basant sur l'information couleur. L'analyse quantitative découle de ces traitements, en évaluant le rapport nuclé-cytoplasmique et l'hyperchromatisme. La classification qui suit, consiste à classer nos cellules suivant leur type.

Mots-clés : Image cytologique; Segmentation; Morphologie mathématique, caractérisation; réseau de Kohonen; classification.

Durant ces dernières années, l'informatique s'est affirmée comme une nouvelle technologie majeure. Les progrès des moyens informatiques et des techniques de traitement ont permis d'ouvrir une voie de développement très prometteuse. L'amélioration et la multiplication des moyens de diffusion des données informatiques ont pris également une grande part à la multiplication de ces techniques. Les outils de communication utilisent de plus en plus les images, ce qui a entraîné un accroissement des recherches dans les domaines du traitement et de l'analyse d'images. Ces techniques ont été développées avec succès dans des domaines divers et très variés (images industrielles, de robotique, biologiques, aériennes ...etc.).

Dans le domaine médical, l'utilisation des images est à présent considérée comme indispensable pour des raisons de facilité de stockage et aussi pour disposer de techniques d'analyse des images automatiques ou semi-automatiques grâce à des systèmes informatiques. Ces systèmes permettent de proposer aux praticiens une nouvelle façon de travailler en obtenant de nouvelles mesures par des techniques quantitatives et objectives, la réalisation de systèmes informatiques aidant le médecin doit permettre de sécuriser le diagnostic, le rendant ainsi plus fiable et sécurisé. Les conséquences peuvent être nombreuses. L'établissement d'un diagnostic fiable peut permettre par exemple, d'améliorer la qualité du dépistage des cancers et donc de favoriser la prévention. Cette recherche de qualité dans l'établissement d'un diagnostic est très présente en anatomie et cytologie pathologiques. En cytologie, le pathologiste doit établir, à l'issue d'examens minutieux de lames de cellules, un diagnostic qui doit être le plus fiable possible. Le pathologiste reconnaît les types cellulaires présents sur une lame et c'est cette étude de la lame qui détermine son diagnostic. Cet examen se fait au microscope, il est donc visuel, manuel et par conséquent très fastidieux. Afin de faciliter le travail de lecture des lames, des recherches ont permis de développer des systèmes semi-automatiques permettant de sécuriser l'établissement du diagnostic.

Le pathologiste est alors assisté dans son travail. De tels systèmes font appel à des techniques de traitement et d'analyse d'images, d'intelligence artificielle et de biologie.

Nous cherchons à réaliser un système informatique semi-automatique permettant de réaliser la détection de cellules anormales sur une lame de prélèvement cytologique.

L'élaboration d'un tel système, est basée sur trois étapes séquentielles:

✓ **La segmentation** : elle consiste à extraire des objets à partir d'images couleurs suivant la morphologie mathématique; il s'agit de la ligne de partage des eaux pour le cas de notre application,

✓ **La caractérisation** : elle fournit des caractéristiques des objets précédemment extraits, suivant une analyse quantitative du noyau et du cytoplasme. Ces données doivent être suffisamment pertinentes et précises pour décrire les objets.

✓ **La classification** : les cellules sont classées dans un type cellulaire grâce à une méthode de classification suivant un modèle de réseaux de neurones.

Le but de notre système est de détecter les cellules anormales ou suspectes qui auraient pu échapper à la vigilance du pathologiste. Notre système traite des images couleurs de la cytologie des séreuses sur lesquelles il doit reconnaître les cellules. Les cellules suspectes ou anormales sont présentées au pathologiste afin de vérifier son diagnostic, le rendant ainsi plus fiable. Nous cherchons à réaliser dans le cadre de ce travail, un système d'aide au diagnostic pour le cytopathologiste, suivant une méthode morphologique de segmentation d'images 2D couleur de cytologie utilisant l'information couleur comme information *a priori*. Cette méthodologie utilise des mesures quantitatives et des informations *a priori* sur les objets à extraire par des pathologistes (rapport du noyau sur le cytoplasme...) dans le but d'une classification connexionniste nous permettant une aide à l'interprétation clinique.

Afin de mieux comprendre notre approche, nous consacrons le premier chapitre à la vision humaine couleurs ainsi qu'aux notions cytologiques,

Le deuxième chapitre, présente la segmentation d'images suivant la morphologie mathématique,

Le troisième chapitre aborde l'analyse quantitative et la classification connexionniste,

Le quatrième chapitre, illustre l'application de la segmentation par morphologie mathématique ainsi que la classification connexionniste sur des images cytologiques.

I. Introduction

La vision artificielle ou à vision assistée par ordinateur (V.A.O) symbolise toute la chaîne visuelle partant de l'image brut et allant jusqu'à l'interprétation de son contenu .il existe donc un aspect décisionnel très important lors de la phase de compréhension , d'où l'interaction avec l'intelligence artificielle. Le traitement d'images s'intègre dans ce processus comme un outil, il n'a aucun pouvoir décisionnel car il se définit comme un ensemble de tâches destinées à extraire de l'image des informations qualitatives et quantitatives qui sont transmises au module de niveau supérieur la figure I.1 résume les composantes d'une chaîne de reconnaissance visuelle [TOU 90].

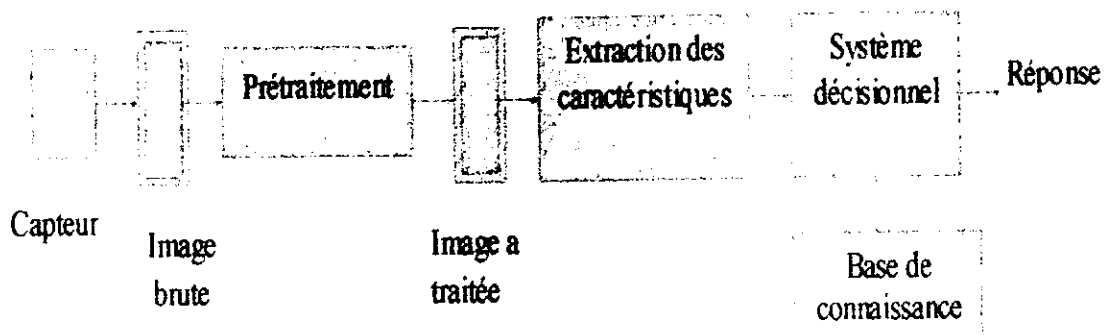


Figure I.1 Chaîne de la V.A.O

Le capteur dépend du domaine d'application mais est essentiellement un scanner, le prétraitement réalise les tâches d'amélioration de l'image. L'extraction des caractéristiques a pour rôle de faire une description de l'image compatible avec les objets contenus dans la base de reconnaissance. Enfin, le système décisionnel réalise la comparaison entre les caractéristiques qui lui sont fournies en amont et le contenu de la base de connaissance proprement dite.

Pour l'ordinateur, l'image est un ensemble de points (pixels) ayant chacun un niveau de gris, ou bien un niveau de couleur correspondant aux composantes rouge, vert ou bleu ou encore aux trois composantes intensité, teinte et saturation [TEM 01].

II. Les modèles de la vision humaine et les espaces couleurs

La première tentative de compréhension du phénomène de la couleur fut entreprise par le philosophe Aristote [LEZ 02] qui étudia, au quatrième siècle avant notre ère, le mélange des couleurs en observant les effets de différents verres colorés sur la lumière. Au dix-septième siècle le physicien britannique Isaac Newton développa la théorie spectrale de la lumière en dirigeant un faisceau de lumière solaire sur la face d'un prisme. En 1801, le médecin anglais Thomas Young établit la théorie trichromatique de la lumière qui fût améliorée et complétée par Hermann Von Helmholtz, cette théorie porte d'ailleurs le nom de loi Young-Helmholtz. Cette loi stipule que la rétine (cf. figure I.2) doit posséder des éléments de trois espèces sensibles respectivement au bleu indigo, au vert jaune et au rouge afin de posséder la vision totale des couleurs.

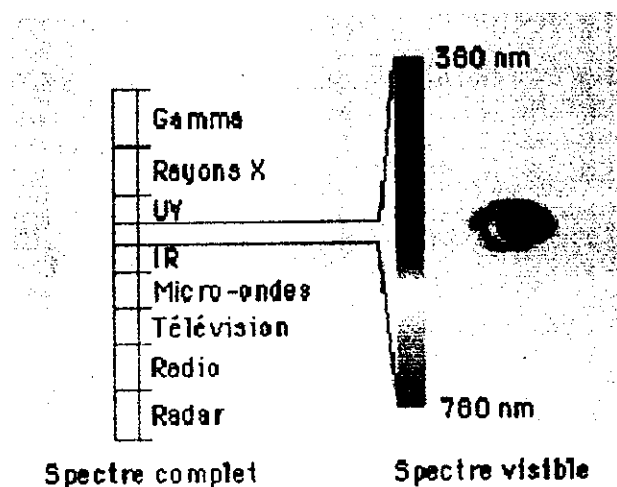


Figure I.2 Le spectre des couleurs visibles

La couleur est le résultat de la perception d'ondes de lumière sur la rétine dans une région visible du spectre, à savoir les longueurs d'onde comprises entre 380nm et 780nm [LEZ 02].

II.1 Les espaces couleurs

Une couleur est généralement représentée par trois composantes. Ces composantes définissent un espace des couleurs. On peut citer l'espace RVB, l'espace CIE XYZ ou Yxy. Ou encore l'espace Lab. Selon l'espace de couleurs choisi pour représenter une image couleur, le nuage des couleurs (c'est-à-dire l'ensemble des couleurs de l'image) n'aura pas la même répartition dans l'espace 3D.

Les espaces de couleurs classiques. Tels que le RVB, CIE, XYZ, etc..., sont issus d'une approche purement physique, sans prise en compte de données psychophysiques [FRU 02]

1- L'espace de couleur RVB (Rouge, Vert, Bleu)

L'espace RVB [FRU 02] est sans doute l'espace de couleur le plus utilisé. Les systèmes de télévision s'y appuient fortement. La représentation des couleurs dans cet espace donne un cube appelé *cube de Maxwell*.

Cet espace est basé sur le modèle du tri stimulus. C'est un système additif qui décompose les couleurs en trois quantités des trois couleurs primaires : le rouge, le vert et le bleu. Le modèle *RVB* utilise le système de coordonnées cartésien. La diagonale du blanc ($1, 1, 1$) au noir ($0, 0, 0$) représente les niveaux de gris (cf Figure I.3).

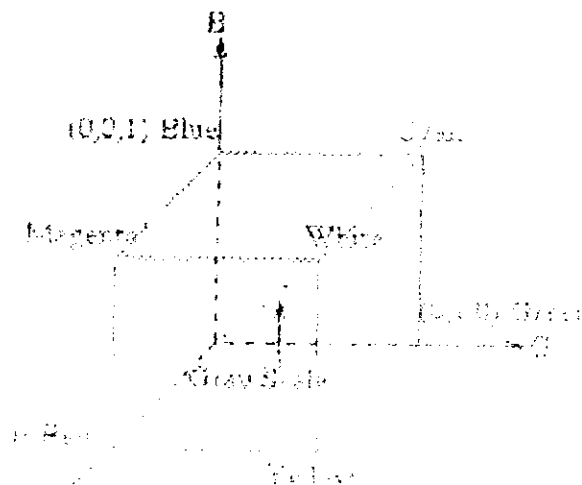


Figure I.3: La représentation spatiale du modèle RVB

2. L'espace de couleur XYZ

L'espace XYZ dépend linéairement du système RVB et a été introduit par la CIE pour remédier aux inconvénients du système RVB pour lequel certaines couleurs sont représentées par une valeur du stimulus négative (cf. Figure I.4).

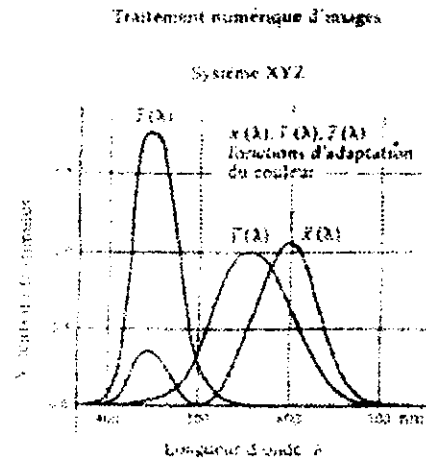


Figure I.4 : Variation du tri stimulus XYZ en fonction de la longueur d'onde.

Le système XYZ pallie cet inconvénient et une couleur est représentée par une combinaison positive de X, Y et Z.

On a alors :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} x_r & x_v & x_b \\ y_r & y_v & y_b \\ z_r & z_v & z_b \end{pmatrix} \times \begin{pmatrix} R \\ V \\ B \end{pmatrix} \tag{I.1}$$

En normalisant X, Y et Z on obtient le système xyz défini par les relations suivantes :

$$x = \left(\frac{X}{X + Y + Z} \right)$$

$$y = \left(\frac{Y}{X + Y + Z} \right) \tag{I.2}$$

$$z = \left(\frac{Z}{X + Y + Z} \right)$$

$$x + y + z = 1$$

3. L'espace de couleur $L^*u^*v^*$

Le problème avec l'espace XYZ est son manque d'uniformité dans la perception à savoir que deux couleurs à la même distance d'une autre ne sont pas nécessairement perceptuellement équidistantes. Pour corriger cet inconvénient la CIE a proposé en 1976 le modèle $L^*u^*v^*$ [LEZ 02] qui est perceptuellement uniforme, c'est-à-dire que si deux couleurs A et B sont aussi proches entre elles que deux autres couleurs C et D , alors la différence perçue entre A et B sera équivalente à celle perçue entre C et D . Les équations permettant de passer du système XYZ au $L^*u^*v^*$ sont :

$$L^* = 116 \left(\frac{Y}{Y_0} \right)^{\frac{1}{3}} - 16 \quad \text{Si} \left(\frac{Y}{Y_0} \right) > 0.008856$$

$$L^* = 903.3 \left(\frac{Y}{Y_0} \right) \quad \text{Ailleurs}$$

$$u^* = 13 L^*(u - u_0) \quad \text{L3}$$

$$v^* = 13 L^*(v - v_0)$$

$$u = \frac{4X}{X + 15Y + Z} \quad v = \frac{9Y}{X + 15Y + Z}$$

$$u_0 = \frac{4X_0}{X_0 + 15Y_0 + Z_0} \quad v_0 = \frac{9Y_0}{X_0 + 15Y_0 + Z_0}$$

Où X_0, Y_0, Z_0 sont les composantes d'un blanc de référence dans l'espace XYZ . L^* est la luminance et u^* et v^* donnent la chrominance de la couleur.

4. L'espace CIE Lab

Le système Lab est issu du CIE XYZ. Il essaye de prendre en compte la réponse logarithmique de l'oeil. Il possède le grand avantage d'être uniforme (cf. figure I.5). Il est très utile dans le cas de mélanges de pigments, par exemple, pour l'industrie graphique ou du textile [LEZ 02].

très utile dans le cas de mélanges de pigments, par exemple, pour l'industrie graphique ou du textile [LEZ 02].

L'espace $L^*a^*b^*$ est basé sur le modèle des couleurs opposées:

- la variable L^* mesure la luminance (l'opposition noir blanc) par une valeur entre 0 (noir) et 100 (blanc),
- la variable a^* mesure l'opposition rouge vert par une valeur entre -100 et +100, a^* est positif si la couleur contient du rouge, négatif si la couleur contient du vert et nulle si aucun des deux,
- la variable b^* mesure l'opposition jaune bleu par une valeur entre -100 et +100, b^* est positif si la couleur contient du jaune, négatif si la couleur contient du bleu et nulle si aucun des deux,

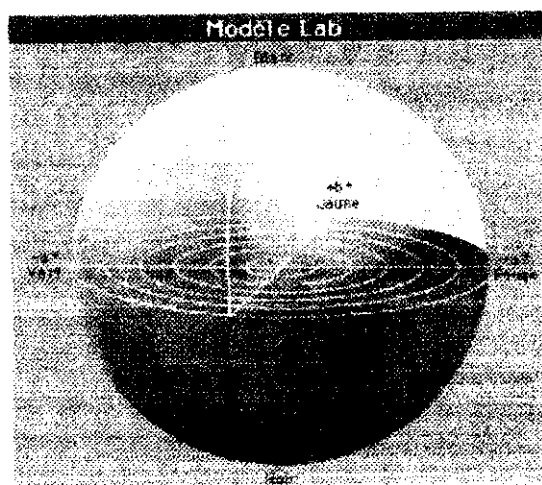


Figure I.5 La représentation spatiale du modèle CIE $L^*a^*b^*$

5. L'espace de couleur HSL (Hue, Saturation, Luminance)

Comme son nom l'indique, il est basé sur les trois grandeurs de Teinte, Saturation et Luminance [FRU 02]. Cet espace est obtenu à partir de l'espace RVB et est décrit par les équations suivantes :

$$H = \arccos \left[\frac{(R - V) + (R - B)}{2\sqrt{(R - V)^2 + (R - V)(V - B)}} \right] \quad \text{Et si } V < B : H = 2\pi - H$$

$$S = 1 - \frac{3 \min(R, V, B)}{L} \quad \text{I.4}$$

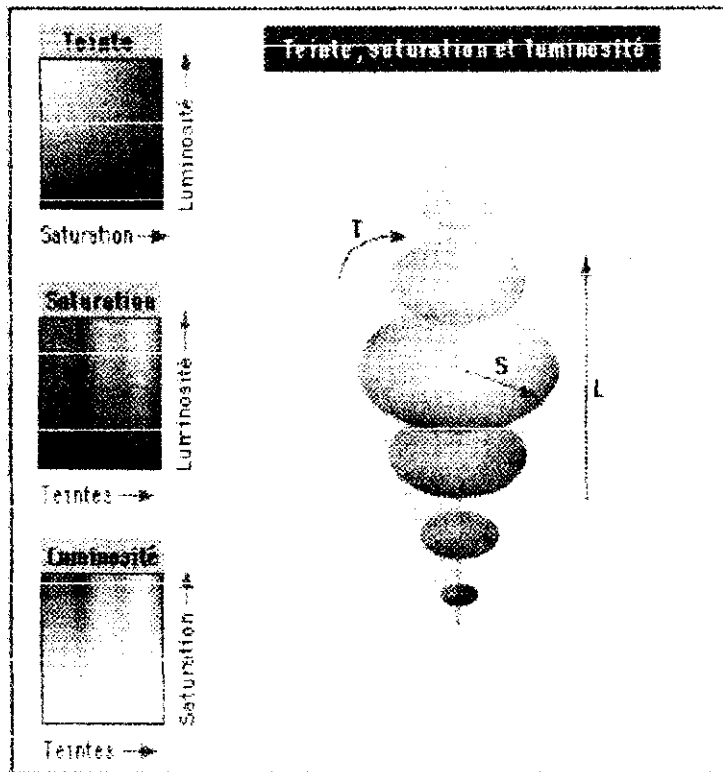


Figure I.6 La représentation spatiale du modèle HSL

La non linéarité des deux équations permettant de calculer H et S entraîne des discontinuités dans leurs valeurs et deux couleurs perceptuellement proches peuvent se transformer en des composantes de teinte et de saturation très différentes. Ceci est dû aux singularités de la teinte pour les points achromatiques ($R=V=B$) et de la saturation pour le noir ($R=V=B=0$) [FRU 02].

Il existe d'autres systèmes d'espaces couleur (YIQ , $CMYK$, etc.). Nous n'avons cité que les espaces principaux utilisés en traitement et analyse d'images. Chacun présente des avantages liés principalement à leur représentation. Il est difficile de prédire a priori

comment une méthode va se comporter dans un espace couleur, si bien que plusieurs espaces couleur ont été conçus afin d'aborder de façon différente les problèmes rencontrés dans les applications [LEZ 02].

III. Introduction à la cytologie

La cytologie [LEZ 95] est une science biologique consacrée à l'étude de la morphologie, de la structure physique et chimique de la cellule et de leurs propriétés ; de manière simplifiée c'est la science de l'observation des cellules.

Les applications de la cytologie sont pratiquement illimitées dans les mondes végétal et animal. Chez l'humain, elles permettent de faire avancer la médecine: par exemple d'élaborer des vaccins et mettre au point de nouveaux traitements. Mais la cytologie a une importance fondamentale pour l'étude des cancers où la cellule est l'unité élémentaire malade. En pratique, elle se consacre surtout à l'examen microscopique des cellules cancéreuses pour reconnaître leur aspect, leur origine et leurs anomalies.

III.1 L'anatomie pathologique

L'anatomie pathologique [LEZ 02] est l'application aux cellules et aux tissus prélevés chez l'homme de diverses méthodes d'analyse basées principalement sur la morphologie, à des fins de diagnostic, de pronostic et de meilleure compréhension des causes et mécanismes des maladies. La pathologie est l'étude et la science des maladies. La tradition littérale de pathologie en grec (pathos, logos) est « les mots de la souffrance ».

Le pathologiste étudie les causes et les conséquences de la maladies, 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 fondamentales et les soins au patient.

L'étude morphologique des lésions qui sont la cause ou la conséquence des maladies peut se faire:

- *à l'œil nu* : c'est l'anatomie pathologique macroscopique,
- *au microscope* : c'est l'histologie pathologique (histo signifie tissu en grec)

où l'on examine des coupes de tissus colorées et la cytologie pathologique (cyto signifie cellule en grec) où l'on examine un étalement de cellules isolées.

III.2 Structure anatomique et cytologique

En 1838, Muller [LEZ 02] fut le premier à décrire et illustrer des cellules malignes de tumeurs observées au microscope. Les premiers atlas de cytologie furent réalisés par des microscopistes français. Des rapports sur les tumeurs en cytologie ont été réalisés ensuite par Beale sur des crachats, par Sanders sur de l'urine, par Dufour sur des fluides cérébraux, par Marrissi sur des spécimens gastrique. En 1928, Papanicolaou présenta ses travaux sur l'étude cytologique du tractus génital féminin ouvrant ainsi la voie vers le dépistage systématique du cancer dans les années 1940.

On peut distinguer quatre niveaux structuraux [LEZ 02]: les cellules, les tissus, les organes et les appareils ou systèmes. A son niveau structural le plus simple, l'organisme est fait de cellules. La cellule (cf. Figure I.7) est l'unité fonctionnelle la plus petite de l'organisme. Les cellules peuvent être différentes entre elles selon leur fonction.

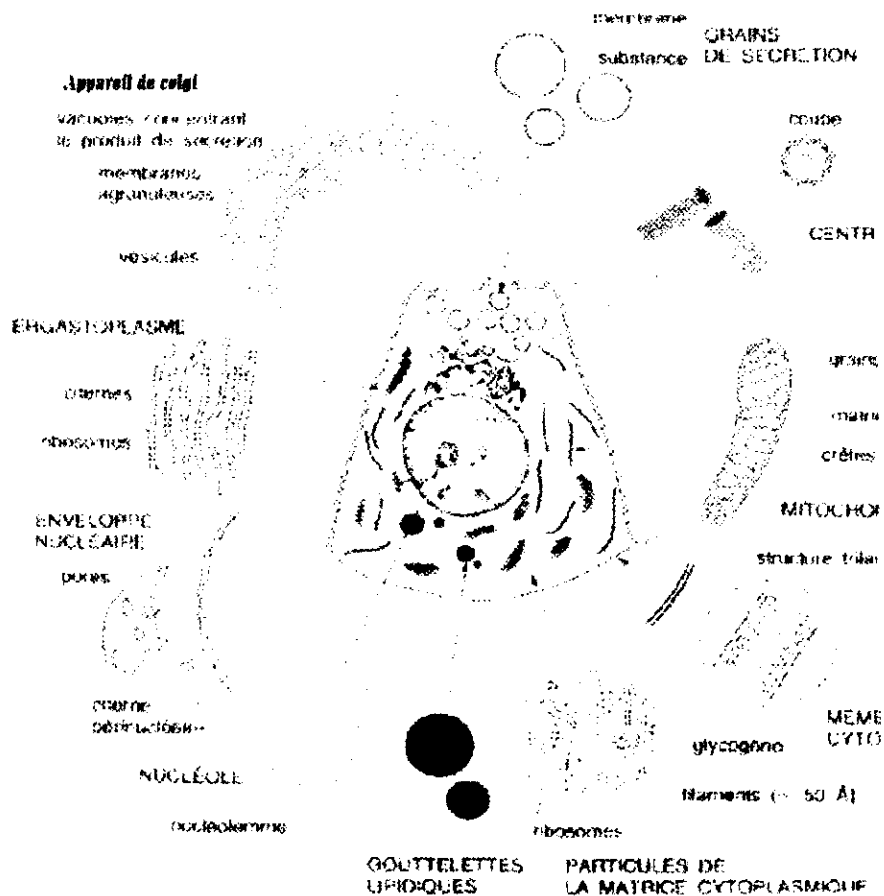


Figure I.7 Représentation schématique de la cellule.

III.3 Les prélèvements en anatomie et cytologie pathologiques

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 :

- 1. 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,*
- 2. les exérèses : ce terme concerne les prélèvements de pièces plus importantes quand on a cherché à retirer la totalité de la lésion,*
- 3. les autopsies : elles servent à déterminer les causes de la mort et les pathologies associées.*

Parmi les prélèvements cytologiques, nous constatons :

- les liquides d'épanchement des séreuses (cf. Figure I.8),
- les urines,
- les liquides céphalo-rachidiens,
- les expectorations, aspirations bronchiques, brossages,
- les frottis génitaux,
- les cytoponctions.

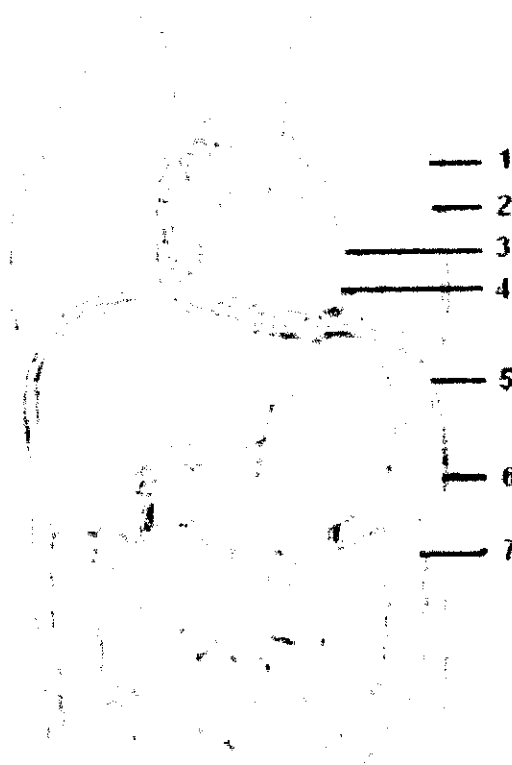


Figure I.8 Schéma des activités séreuses du corps humain. Les cavités pleurales, péricardiques et pétonérales sont comprises entre un feuillet pariétal et un feuillet viscéral. 1 : plèvre pariétale, 2 : plèvre viscérale, 3 : péricarde, 4 : épicarde, 5 : diaphragme, 6 : péritoine pariétal, 7 : péritoine viscéral.

Après le prélèvement on observe une autolyse des cellules (dégénérescence). 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). L'échantillon est « technique » par centrifugation afin d'obtenir une concentration cellulaire. Le culot de centrifugation est étalé sur des lames puis fixées et colorées.

II.4 Les colorations

Il n'existe pas de technique de coloration polyvalente [LEZ97]. 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 sont secs sous réserve d'effectuer des colorations particulières telles que le May- Grunwald-Giemsa, le Bleu Alcian, le Péris et le P.A.S.

Les principales colorations cytologiques sont :

1. la coloration de Papanicolaou,
2. la coloration de Harris-shorr,
3. le may-Grunwald-Giemsa,
4. la coloration de Peris,
5. le P.A.S,
6. le Bleu-Alcian.

Le standard international [LEZ 97] est la coloration de Papanicolaou qui permet d'obtenir les colorations suivantes. Les noyaux cellulaires sont colorés en bleu violet par l'hématoxyline de Harris, les cytoplasmes en vert ou bleu vert par EA50 et OG6.

II.4.1 Types cellulaires en cytologie des séreuses

Les différents types de prélèvements possibles ainsi que les différentes colorations ont été présentés. Nous travaillons sur des cellules provenant de la cytologie des séreuses colorées par le standard international de Papanicolaou. Afin de bien connaître tous les

objets cellulaires pouvant être rencontrés, ceux-ci ont été répertoriés. Ils se répartissent en 6 catégories.

1. Les objets du fond de la préparation

- 1- les hématies ou globules rouges,
- 2- les débris cytoplasmiques,
- 3- les débris nucléaires,
- 4- les noyaux nus,
- 5- les amas de mucus, les spires de Curshmann ou les calcosphérites,
- 6- les cristaux de Charcot-Leyden extra-cellulaires,
- 7- les éléments mycéliens,
- 8- les substances étrangères (grains de colorants, poussières, cellules végétales, grains de talc, dépôts de carbone des mines de crayon, bulles d'air, altemaria).

2. Les cellules épithéliales normales

- 1- Mésothéliales en bague à chaton :
 - Mitoses
 - Binucléées
 - Multinucléées
- 2- Les malphigiennes
 - Intermédiaires
 - Superficielles

3. Les cellules inflammatoires

- 1- Les polynucléaires
 - Neutrophiles
 - Eosinophiles
 - Altérés (Pyocytes)
- 2- Les lymphocytes
 - Activés
 - Non activés
- 3- Les histiocytes
 - Macrophages

4. Les cellules tumorales primaires

Cellules du mésothéliome malin

5. Les cellules tumorales secondaires

- 1- Cellules des lymphomes malins en cellules isolées,
- 2- Sarcomes en cellules isolées,
- 3- Les mélanomes malins

Voici quelques exemples de types cellulaires (isolés ou en amas) en cytologie des séreuses (Cf. Figure I.9). On remarque la diversité des configurations tant au niveau de la couleur (de pale à foncé) que de la répartition spatiale pour des cellules en amas.

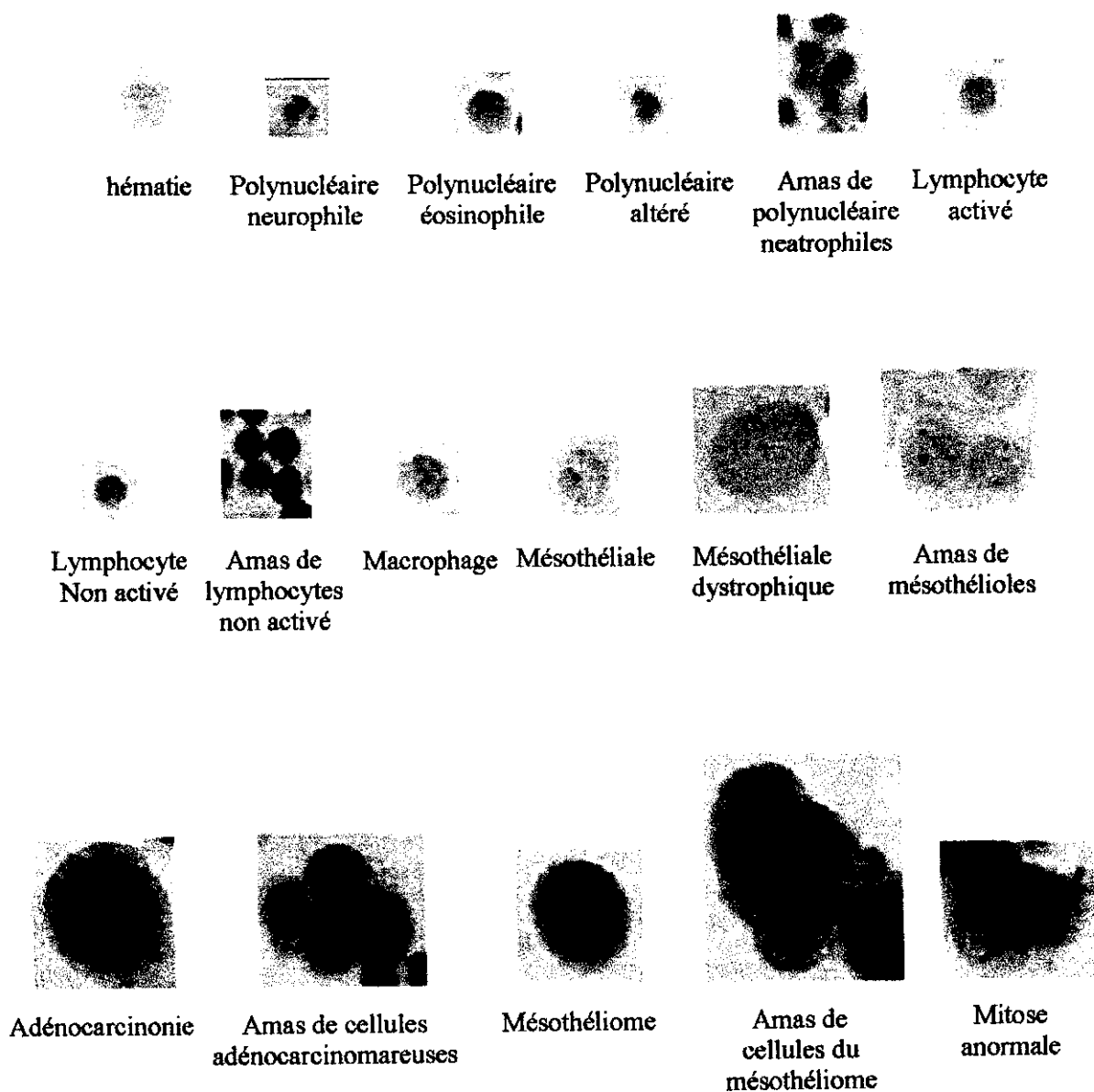


Figure I.9 Quelques types cellulaires en cytologie des séreuses.
Les cellules peuvent être isolées ou en amas.

II.5. Le screening

II.5.1 Le screening Manuel

Une fois le prélèvement concentré, étalé sur une lame, les cellules sont colorées par la coloration de Papanicolaou [LEZ02]. La phase d'expertise de la lame peut commencer. Cette phase s'appelle le *screening*. Cela consiste en une étape de lecture de la lame par un expert en cytologie (un cytotechnicien). Il repère les cellules intéressantes, à savoir les cellules qui sont ou qui pourraient être anormales. Le *screening* est donc un repérage et une évaluation visuelle de toutes les cellules présentes sur une lame afin de détecter et d'identifier d'éventuelles anomalies. Son but est principalement la détection des cellules anormales ou suspectes afin d'établir un diagnostic. Le diagnostic le plus important étant la présence de cellules anormales. Ceci est donc d'un intérêt capital pour le pathologiste dont le diagnostic dépend de la bonne détection et reconnaissance des cellules anormales ou suspectes lors du *screening*.

Un *screening* s'effectue de la manière suivante : le cytotechnicien parcourt la lame de gauche à droite et de haut en bas (technique d'aller-retour). A chaque instant il inspecte une partie de la lame à travers un champ microscopique (cf FigureI.10). Lorsque le cytotechnicien passe au champ suivant, il procède à une superposition entre les champs (précédent et actuel). Cette superposition est horizontale et verticale (les superpositions sont données par les zones hachurées). Ceci permet de sécuriser le *screening* et de rendre le dépistage plus fiable.

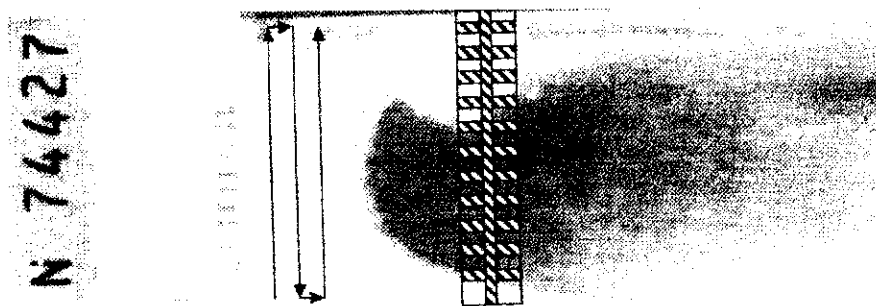


Figure I.10 Le screening manuel sur une lame.

Le *screening* manuel des prélèvements cytologiques est donc difficile, fatigant, consommateur de temps et est décrit comme un travail « intense, particulièrement complexe dont le résultat repose sur le point de vue de l'être humain ». Le faible nombre de cellules anormales (comparativement aux cellules normales) observées durant le *screening* implique une grande concentration du cytotechnicien. Un cytotechnicien voit entre 30 et 200 lames par jour. Sa culture médicale est remise en cause chaque jour. A cause de ce facteur subjectif inévitable, quelques erreurs peuvent apparaître et causer des faux négatifs (une lame donnée comme ne contenant pas de cellules anormales alors que c'est le cas). Or les préoccupations principales en cytologie sont celles qui permettent d'augmenter le potentiel de vérité contenu dans le diagnostic. Un cytopathologiste s'attache donc à réaliser le moins de faux positif possible (une lame donnée comme anormale alors qu'elle ne contient que des cellules normales) mais surtout de faux négatifs.

Des taux de faux négatifs ont été donnés dans la littérature et certains travaux alarmants font état de 50% de faux négatifs pour la cytologie du col. Les principales causes d'erreur en cytologie sont parmi les suivantes:

- Les prélèvements non significatifs (mal faits, mal fixés, acellulaires),
- Les erreurs de mauvaise interprétation,
- Les erreurs d'omission par diminution de la vigilance,
- Le grand nombre de lames à analyser.

Certaines mesures ont été envisagées pour effectuer une vérification de 10% des lames données comme négatives : cela a été prescrit par le gouvernement américain par arrêté fédéral en réponse à un contrôle de qualité dans les services de cytopathologie.

Cependant après plusieurs années, il n'est pas apparu que ce genre de contrôle des lames élimine des erreurs de *screening*. La relecture rapide de toutes les lames donne de meilleurs résultats mais reste consommatrice de temps.

II.5.2 Le screening automatique

Afin d'améliorer le contrôle de la qualité du *screening*, certaines recherches se sont orientées vers une automatisation totale du *screening*. La mise en place d'un système informatique peut apporter au cytopathologiste un plus dans le domaine de l'assurance de

la qualité. En effet l'examen des lames par un cytotechnicien est une tâche subjective et celle-ci n'est pas infaillible. Un système informatique peut permettre de donner une analyse objective de la lame. De tels systèmes basés sur l'analyse d'images sont apparus rapidement.

II.5.3 Structure d'un système informatique de rescreening

Les systèmes informatiques de *rescreening* sont conçus pour opérer de façon postérieure au *screening* manuel. Mais ils peuvent éventuellement être utilisés pour fonctionner de manière antérieure, cependant les pathologistes ont une préférence pour le *rescreening* informatique. L'architecture des systèmes utilisés pour le *rescreening* de lames de cytologie reste identique quel que soit le système employé et est constituée de trois étapes séquentielles:

- La segmentation : elle consiste à extraire des objets (les composantes cellulaires) à partir d'images couleur,
- La caractérisation : elle fournit des caractéristiques des objets précédemment extraits. Ces données doivent être suffisamment pertinentes et précises pour décrire les objets.
- La classification : les cellules sont classées dans un type cellulaire grâce à une méthode de classification (arbre de décision, réseaux de neurones, etc.).

Cela consiste à déterminer le type de la cellule à classer et plus précisément à déterminer son caractère bénin ou malin.

IV. Conclusion

Ce chapitre nous a permis d'étudier les notions de base de la vision couleur ainsi que quelques notions cytologiques en vue d'une aide au screening. Le chapitre suivant, introduit l'extraction de l'information suivant la morphologie mathématique.

I. Introduction

L'utilisation de la couleur en segmentation d'images est un sujet de recherche relativement récent. L'évolution de la puissance et de la capacité de stockage des ordinateurs [LEZ 02] a fourni un regain d'intérêt à l'utilisation d'images couleurs depuis quelques années. Bien que l'on trouve plusieurs sortes d'algorithmes de segmentation d'images couleurs, la littérature n'est pas aussi riche que pour les images en niveaux de gris. Pourtant les chercheurs ont pensé assez tôt à utiliser la couleur pour segmenter des images. En effet, la couleur étant une information que l'on attache directement et naturellement aux régions, une segmentation couleur devrait permettre d'obtenir des régions qui soient davantage significatives. Dans ce chapitre, nous ne traiterons que des méthodes de segmentation dites de vision universelle (de bas niveau) et pour lesquelles l'apport de la couleur consiste en un rapport quantitatif (par un vecteur à trois dimensions).

II. Principales méthodes de segmentation d'images couleurs

Les méthodes de segmentation couleur présentent plusieurs particularités [LEZ 02] que l'on retrouve systématiquement suivant:

A) *La méthode utilisée* : elle caractérise la manière d'extraire les régions d'une image grâce à son information couleur [LEZ 02].

B) *le choix de l'espace de travail* : les images couleurs sont généralement acquises et/ou stockées dans l'espace RVB [COC 95]. Les traitements peuvent être effectués dans cet espace. Cette manière de procéder, présente l'avantage de garantir l'intégrité des données, mais elle entraîne un certain nombre d'inconvénients :

- *Les composantes y sont souvent fortement corrélées,*
- *Il est difficile d'y interpréter la notion de couleur telle qu'elle est habituellement perçue par l'œil humain,*
- *La différence perceptuelle entre deux couleurs ne coïncide pas avec leurs distances*

Dans l'espace RVB, on peut chercher un espace plus approprié. Le traitement est alors effectué dans ce nouvel espace (après une transformation T), et l'application de la transformation inverse (T') permet si nécessaire de revenir dans l'espace RVB (cf. Figure II.1).

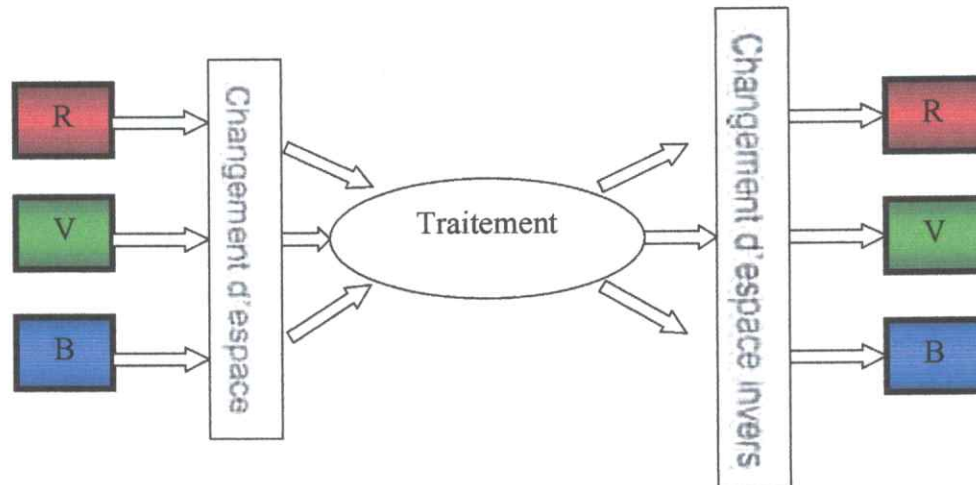


Figure II.1. Procédé du traitement par l'utilisation des espaces couleurs.

Le principal problème réside dans le choix de l'espace couleur pour la résolution d'une application donnée et en fonction de la méthode de segmentation utilisée [LEZ 02].

C) *La métrique utilisée* : La distance euclidienne et la distance de Mahalanobis

[LEZ 98]. sont les plus utilisées. Le problème de la mesure des différences colorimétriques est un des plus importants problèmes qui se pose pour l'analyse de la distribution des couleurs.

Voyons les approches les plus communément utilisées pour segmenter une image couleur.

Elles se répartissent en trois groupes :

- La croissance de régions,
- La classification de pixels,
- La morphologie mathématique, l'objectif de notre étude.

➤ des méthodes qui utilisent des critères de forme et de couleur pour extraire les régions d'une image couleur. Les régions à isoler sont des cellules qui sont caractérisées par leur couleur et leur forme convexe. Le critère couleur utilisé, est une mesure de similarité couleur entre le pixel à ajouter à une région et la région candidate à la croissance [LEZ 02]. Le critère de convexité permet de n'agréger à une région que les points se situant sur l'enveloppe convexe (la croissance de région est initialisée par un seuillage de la composante Verte de l'image couleur *RVB* originale).

➤ Des méthodes qui utilisent une approche coopérative régions frontières afin d'extraire les informations pertinentes contenues dans une image couleur, un opérateur de gradient couleur particulier (opérant sur des fenêtres de taille fixe) est calculé sur l'image couleur. L'image gradient est ensuite seuillée, les pixels de l'image résultante sont fusionnés entre eux dans un voisinage de huit pixels. Les régions finales obtenues sont les régions de couleur homogènes.

II.2 La classification de pixels

Cela consiste à affecter à chaque pixel de l'image une classe qui définit les régions à extraire de l'image. Les réseaux de neurones sont souvent utilisés et les différences entre les méthodes se situent généralement au niveau des réseaux utilisés (hopfield, rétro propagation, minimisation de fonctionnelles) et des paramètres d'entrée [LEZ 02].

II.3 La Morphologie Mathématique

La morphologie mathématique est une science de la forme et de la structure [BEU 90]. Elle est basée sur des concepts topologiques et géométriques de la théorie des ensembles. Au cours des dernières années, la morphologie mathématique s'est établie comme une méthode puissante pour le traitement d'image et elle est devenue une théorie mathématique complète. Elle a été appliquée avec succès dans diverses disciplines, telles que la minéralogie, le diagnostic médical et l'histologie. Elle a trouvé des applications croissantes dans le traitement d'images digitales, la vision par ordinateur, et la reconnaissance de formes.

Le principe de base de la morphologie mathématique est de comparer les objets que l'on veut analyser à un autre objet de référence, de taille et de forme connues, appelé **élément structurant**. En quelque sorte, chaque élément structurant fait apparaître l'objet sous une nouvelle forme. Fondamentalement, la morphologie mathématique a un caractère

ensembliste et ceci correspond à la démarche dont nous disposons pour analyser quantitativement une image [COS 89].

Dans ce qui suit, nous présenterons les transformations de base en morphologie mathématique et leurs propriétés ainsi que son application sur les images.

III. Outils principaux de la morphologie mathématique

Cette théorie est fondée sur les opérations d'érosion et de dilatation. Ces opérations répétées, permettent l'ouverture et la fermeture.

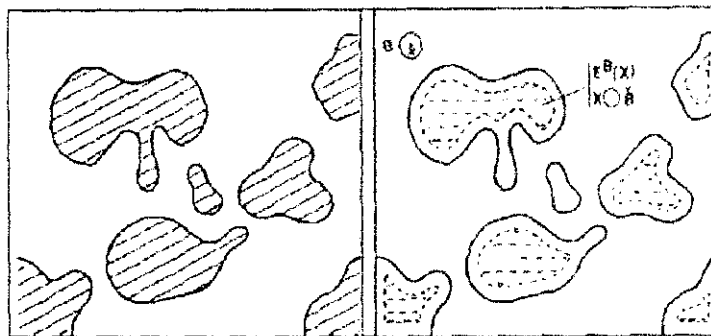
III.1 L'érosion

Soit X l'ensemble à éroder [COS 89], et B un élément structurant de géométrie simple, par exemple un cercle. L'érosion consiste à faire déplacer B de telle sorte que son centre occupe toutes les positions x de l'espace. Pour chaque position, on pose la question : est-ce que B est complètement inclus dans X ?

L'ensemble des positions x correspondant à une réponse positive forme le nouvel ensemble Y , appelé érodé de X par B , noté $X \ominus B$ (cf. Figure II.2).

Cet ensemble satisfait l'équation

$$X \ominus B = \{ x \text{ dans } X : B \subset X \} \quad \text{II.1}$$



a : Ensemble d'objets X

b : Erodé de X

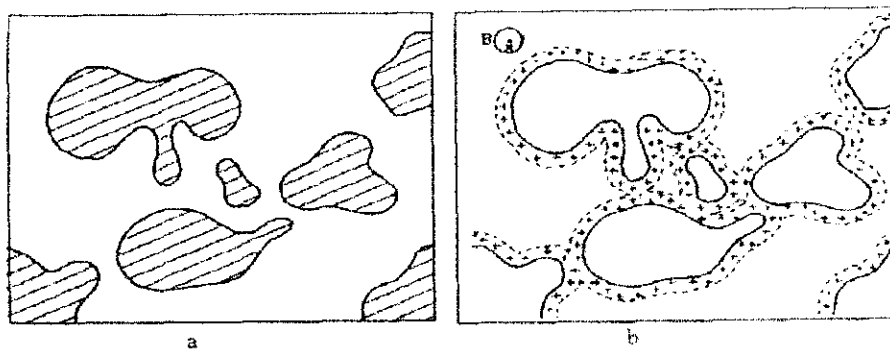
Figure II.2 Erosion de (a) par un élément structurant circulaire

III.2 La Dilatation

L'opération de dilatation [SER 82] se définit de manière analogue à l'érosion. En prenant le même élément structurant B , on pose pour chaque point x la question "Bx touche-t-il l'ensemble X ?", c'est à dire, y a-t-il une intersection non vide entre Bx et X ?

L'ensemble des points de l'image (cf. Figure II.3) correspondant aux réponses positives forme le nouvel ensemble Y des dilatés de X , noté comme suit :

$$y = x \oplus B = \{ x \text{ dans } I : B \cap X \neq \emptyset \} \quad \text{II.2}$$



a : Ensemble d'objets X

b : Dilaté de X

Figure II.3. Dilatation de (a) par un élément structurant circulaire

III.3 Ouverture

Puisque l'érosion et la dilatation sont des transformations itératives, il est possible d'effectuer sur un ensemble X une érosion, puis de dilater l'ensemble érodé par un même élément structurant B . Le résultat est nommé l'ouvert de X par B , noté comme suit :

$$O^B(X) = (X \ominus B) \oplus B \quad \text{II.3}$$

Après une ouverture, on ne retrouve pas l'ensemble de départ; L'ensemble ouvert est plus régulier et moins riche en détails que l'ensemble X initial (cf. Figure II.4). La transformation par ouverture adoucit les contours, coupe les isthmes étroits, supprime les petites îles et les caps étroits [SER 82].

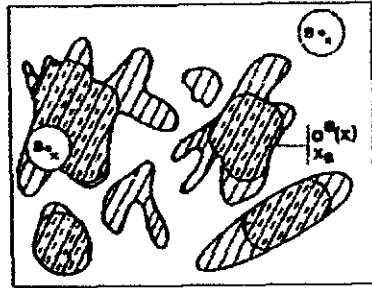


Figure II.4. Ouverture d'un ensemble d'objets X par un élément structurant circulaire

III.4 Fermeture

La fermeture est l'opération "inverse" de l'ouverture, c'est à dire que l'on applique tout d'abord une dilatation puis une érosion (toujours en gardant le même élément structurant) Le résultat est nommé le fermé de X par B, noté comme suit :

$$F^B(X) = (X \oplus B) \ominus B \quad \text{II.4}$$

Un ensemble fermé est également moins riche en détails que l'ensemble initial (cf. Figure II.5). La transformation par fermeture bouche les canaux étroits, supprime les petits lacs et les golfes étroits [COS 89].

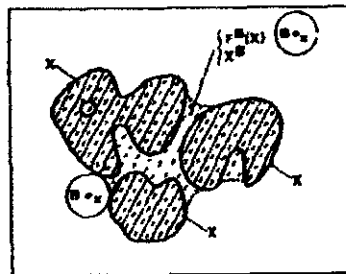


Figure II.5. Fermeture d'un ensemble d'objets X par un élément structurant

IV. La ligne de partage des eaux

La ligne de partage des eaux (L.P.E) est l'outil de base pour la segmentation d'images par morphologie mathématique [COC95], si l'on considère l'image comme une surface dans l'espace de dimension 3, on peut utiliser la terminologie de la géographie qui définit la ligne de partage des eaux, comme la crête qui forme la limite entre deux bassins versants.

Il existe différents algorithmes pour le calcul de la L.P.E qui opèrent généralement sur un gradient de l'image. Pour utiliser la couleur à travers la L.P.E, plusieurs choix sont possibles,

- On peut calculer la L.P.E sur chacune des composantes couleurs séparément et combiner ensuite les résultats,
- On peut calculer la L.P.E directement sur l'image couleur ou encore sur le gradient de l'image couleur [LEZ02].

La principale transformation en morphologie mathématique, a été utilisée pour segmenter des images en niveaux de gris dans un premier temps. Puis son extension aux images couleurs, a permis d'ouvrir une nouvelle voie dans le traitement des images couleurs.

IV.1 Définition par immersion

Pour construire la ligne de partage des eaux [LEZ 02], nous allons procéder en immergeant progressivement dans l'eau l'image par le bas. Pour que l'eau progresse de façon régulière sur toute l'image, supposons que chaque minimum régional soit « percé ». Nous allons ensuite immerger progressivement l'image jusqu'à son immersion complète.

En partant des minima de plus faible altitude, l'eau va progressivement remplir les bassins versants de. En partant des ouvertures que nous avons créées, l'eau va progresser dans chacun des bassins versants correspondant à chaque minimum. Puis lorsque deux eaux en provenance de deux minima différents se rencontrent, on construit un barrage, afin que des eaux d'origines différentes ne se mélangent pas (cf. Figure II.6)

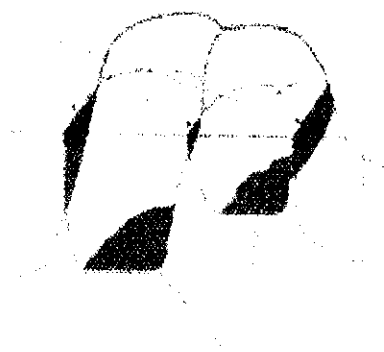


Figure II.6 Barrages aux endroits où l'eau provenant de deux minima se mélangerait

A la fin de la procédure d'immersion, chaque minimum est alors complètement entouré de barrages qui délimitent le bassin versant qui lui est associé (la zone d'influence de ce minimum). L'ensemble des barrages constitue la ligne de partage des eaux de l'image.

IV.2 La L.P.E contrainte par marqueurs

La L.P.E calculée directement sur une image réelle, présente une sur segmentation importante, due au bruit qui crée de nombreux minima parasites. Il est donc nécessaire de mettre en place des procédures permettant de les éliminer [LEZ 02]. Supposons que nous connaissions un ensemble connexe de points par objets à isoler ainsi qu'un ensemble de points de l'extérieur. On appellera ces composantes connexes des marqueurs. Si l'on parvenait à modifier l'image sur laquelle on calcule la L.P.E en lui imposant ces marqueurs comme minima régionaux, on obtiendrait une L.P.E présentant une boucle autour de chacun des objets; chaque bassin versant correspondant soit à un unique objet soit au fond de l'image.

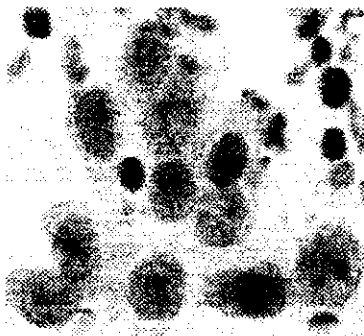
Cette technique de L.P.E contrainte par marqueurs permet de s'affranchir de la recherche exacte des contours et garantit le nombre de contours que l'on trouvera: un autour de chaque composante connexe. Pour résumer, la L.P.E contrainte par marqueurs fonctionne de la manière suivante:

1. *Trouver les marqueurs, une composante connexe par objet et une composante connexe pour le fond de l'image,*
2. *Déterminer l'image sur laquelle on calcule la L.P.E,*
3. *Imposer les minima,*
4. *Calculer la L.P.E.*

IV.3 La ligne de partage des eaux couleur

L'extension de la ligne de partage des eaux en niveaux de gris à la couleur est relativement récente [LEZ 02].

Une méthode de croissance de régions utilise des différences de couleur entre pixels et une région pour réaliser une croissance de régions rapide où les régions se propagent à partir de marqueurs. Ce procédé calcule la ligne de partage des eaux directement sur l'image couleur d'origine. On dispose alors d'un algorithme de croissance de régions en ce sens qu'il est initialisé par des marqueurs qui identifient l'intérieur des régions et ces régions sont étendues jusqu'à occuper tout l'espace disponible (cf. Figure II.7).



a : image cellulaire



b : image segmentée

Figure II.7 Image segmentée par la croissance des régions

Les algorithmes de croissance de régions reposent sur des principes de fonctionnement identiques :

- 1- *Définition d'un prédicat PRED (Rc,Pc) pour mesurer le degré de similarité du pixel examiné Pc avec la région en cours de formation Rc.(Rc : Région courante, Pc : Pixel courant),*
- 2- *Initialisation de la région et de son modèle par un ensemble de pixel, noté G' et appelé noyau ou région-noyau (ou germe : nous l'avons positionné manuellement),*
- 3- *Croissance de la région en respectant la connexité des pixels avec la région en examinant la possibilité d'appartenance à la région en cours de formation,*
- 4- *Le processus s'arrête selon deux conditions :*
 - *Si toutes les régions ont été formées est il n'existe plus de pixel candidat, ou de régions noyau.*
 - *Si le prédicat n'est plus vérifié pour tous les pixels voisins à la région en cours de formation.*

L'algorithme utilisé dans le cadre de notre travail, est indiqué dans ce qui suit:

```

Début
  Rc=0 ;
  G=Le germe initial ;
  Pour tous les points de l'image source Faire
    Rc=0 ;
    G= Le germe choisi ;
    Pour tous les voisins de g Faire
      SI PRED(Rc,Vi) == Vrai Alors
        Rc=Vi ; // affectation du candidat à la région
      Fin Si
    Fin Pour
  Fin Pour
Fin

```

Figure II.8 Algorithme de croissance de région

L'algorithme la croissance de régions (cf. Figure II.8) consiste à extraire un pixel de la file de priorité et à examiner ses régions voisines. Une distance entre le pixel et les régions voisines est alors calculée et le pixel est fusionné avec la région la plus proche (la plus similaire). On note ici une similarité avec les méthodes de croissance de régions à travers cette mesure de distance [BON 93]

C'est la modification de cette fonction distance qui a entraîné des recherches actives; suivant la modification de cette fonction, afin d'améliorer la précision de l'algorithme et calcule la L.P.E directement sur le gradient de l'image couleur. D'autres méthodes [BON 93] ont modifié cette fonction distance en y ajoutant d'autres mesures liées à des critères locaux et globaux..

Une mesure de similarité entre un pixel et une région voisine s'exprime alors de la manière suivante par la fonction d'agrégation :

$$\phi(p, R) = \left(\sum_{i=1}^l \alpha_i \cdot \phi_i(p) + \sum_{i=l+1}^m \alpha_i \cdot \phi_i(p, R) \right) \quad \text{II.5}$$

Où :

Les $\phi_i(p)$ représentent « l critères locaux » directement extraits d'une image (image couleur, gradient,...)

Les $\phi_i(p, R)$ représentent « $m-l$ critères globaux » résultant de comparaisons statistiques entre un point p et une de ses régions voisines R .

Les α_i sont des coefficients de pondération qui respectent la propriété suivante:

$$\sum_{i=1}^l \alpha_i = 1$$

On retrouve la définition de la ligne de partage des $l=m=1$ et $\phi_i(p) = |\text{grad}(i)|$.

IV.3.1 Définition de la fonction d'agrégation couleur

Une ligne de partage des eaux [LEZ 02] utilisant la couleur des objets présents sur l'image et les transitions de couleur entre les objets (que l'on retrouvera toujours quelles que soient les images rencontrées) peut permettre d'extraire de façon très précise et très fiable les objets présents dans l'image. La couleur des objets présents sur l'image se traduit directement par une mesure statistique de la couleur moyenne des régions et les transitions entre les couleurs s'expriment facilement par le gradient couleur. Une ligne de partage des eaux couleur sera précise en utilisant les propriétés suivantes: l'information locale donnée par le gradient couleur traduisant les transitions entre les couleurs et l'information globale donnée par une mesure de la couleur moyenne des régions traduisant l'homogénéité globale des régions. Nous exprimons ces deux critères à travers une fonction d'agrégation. Ceci se traduit par la formule suivante :

$$\phi(p, R) = (1 - \alpha) \left\| \overline{I_{c_1 c_2 c_3}}(R) - I_{c_1 c_2 c_3}(P) \right\| + \alpha \left\| \nabla I_{c_1 c_2 c_3}(P) \right\| \quad \text{II.6}$$

Avec $\overline{I_{c_1 c_2 c_3}}(R)$ le vecteur donnant la couleur moyenne de la région R pour l'image I dans l'espace $C_1 C_2 C_3$, $I_{c_1 c_2 c_3}(p)$ le vecteur donnant la couleur du point p dans l'image I et $\nabla I_{c_1 c_2 c_3}(p)$ le gradient couleur au point p.

α est un coefficient de pondération qui permet de modifier la relation et l'influence des critères global et local durant le processus de croissance (avec $\alpha = 1$, on retrouve l'algorithme classique de la L.P.E mais dont le gradient est calculé sur une image couleur. Avec $\alpha = 0$ on retrouve l'algorithme qui n'utilise que la distance couleur entre un pixel et une région).

IV.3.2 Présentation de la méthode de segmentation d'images couleurs

Nous avons vue précédemment une méthode générale de segmentation d'images couleurs. Nous allons voir des stratégies complètes d'extraction d'objets qui utilisent la ligne de partage des eaux d'images couleurs. Pour chaque application, nous exposerons la stratégie d'analyse d'images puis nous détaillerons chacun de ses points.

IV.3.2 Présentation de la méthode de segmentation d'images couleurs

Nous avons vu précédemment une méthode générale de segmentation d'images couleurs. Nous allons voir des stratégies complètes d'extraction d'objets qui utilisent la ligne de partage des eaux d'images couleurs. Pour chaque application, nous exposerons la stratégie d'analyse d'images puis nous détaillerons chacun de ses points.

Pour chacune des applications, la résolution est guidée par l'information couleur [LEZ 95] C'est cette information qui caractérise principalement nos images et c'est cette information qui guidera les différentes stratégies de segmentation (à travers la ligne de partage des eaux ainsi que d'autres opérations de traitement d'images).

1. Notations utilisées

Afin de faciliter la lecture de ce document nous allons d'abord commencer par présenter les notations que nous allons utiliser:

- I désigne une image en niveaux de gris,
- $I_{c_1c_2c_3}$ désigne une image couleurs dans l'espace couleur $C_1C_2C_3$,
- $I_{c_1c_2c_3}^{c_1}$ désigne la composante C_1 de l'image couleur $I_{c_1c_2c_3}$,
- $S(I)$ désigne le seuillage d'une image I ,
- $E^B(I)$ désigne l'érosion morphologique d'une image I avec l'élément structurant B ,
- $D^B(I)$ désigne la dilatation morphologique d'une image I avec l'élément structurant B ,
- $O^B(I)$ désigne l'ouverture morphologique d'une image I avec l'élément structurant B ,
- $I_1 \cap I_2$ désigne l'intersection morphologique entre deux images binaires I_1 et I_2 ,
- $Exp.(I)$ désigne le lissage exponentiel d'une image I ,
- $\nabla(I_{c_1c_2c_3})$ désigne le gradient couleur de l'image couleurs $I_{c_1c_2c_3}$,
- $LAB(I)$ désigne l'étiquetage d'une image binaire I .

2. Description des images

Les images que nous devons segmenter pour cette application sont des images de la cytologie des séreuses. Sur ces images, on trouve des cellules qui doivent être isolées dans le but de l'extraction de l'information pertinente. Pour cela, il faut extraire le cytoplasme et le noyau de chacune des cellules présentes

Avant de segmenter nos images, il faut décrire précisément les types de fond et les types d'objets que nous pouvons rencontrer. Dans ce cadre, on distingue trois types de fond: les fonds clairs homogènes, les fonds hémorragiques (présence de globules rouges en plus des cellules) et les fonds muco-protéiniques (présence de mucus et de protéines colorés en vert). Une cellule est soit un globule rouge coloré en rouge, soit une cellule dont le noyau est coloré en bleu et le cytoplasme en vert. La stratégie de segmentation que nous avons adopté, doit extraire les cellules présentes quelles que soient leurs configurations spatiales (cf. Figure II.9) à l'exception des globules rouges qui ne présentent aucun intérêt d'interprétation dans le cas étudié.

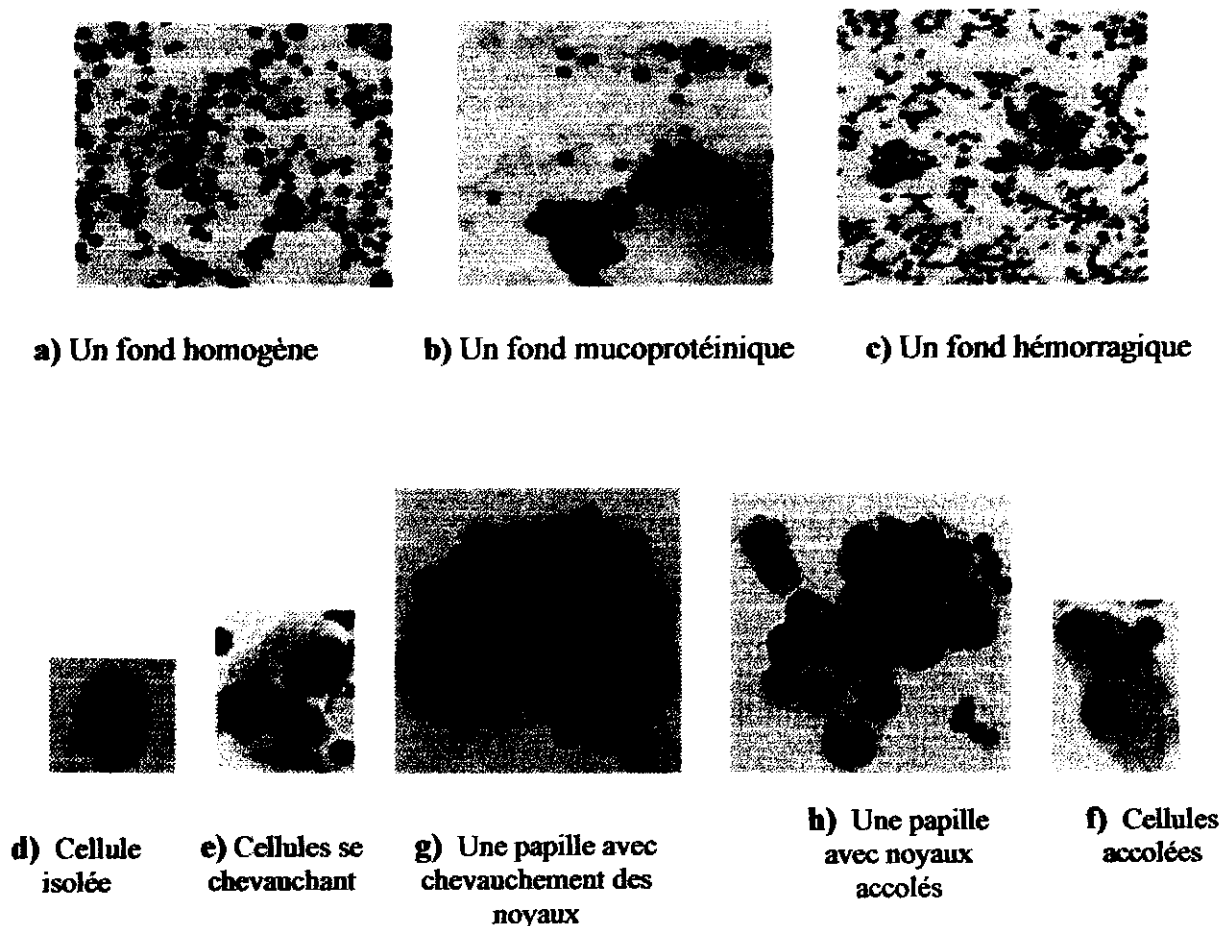


Figure II.9 Les différentes configurations spatiales des cellules

3. Processus de la segmentation des images cellulaires couleurs

La stratégie de segmentation doit isoler le noyau et le cytoplasme de toutes les cellules présentes sur une image à l'exception des globules rouges [LEZ 02]. Nous procédons en quatre étapes successives:

- ✓ La première étape consiste à déterminer si il y a présence ou non de noyaux sur l'image à analyser.
- ✓ Si des cellules sont présentes, on procède à l'extraction des globules rouges afin de les éliminer.
- ✓ Les étapes suivantes concernent l'extraction des cytoplasmes et des noyaux et sont réalisées par la méthode de la ligne de partage des eaux.

La stratégie d'extraction, est exprimée sur l'organigramme de la figure II.10

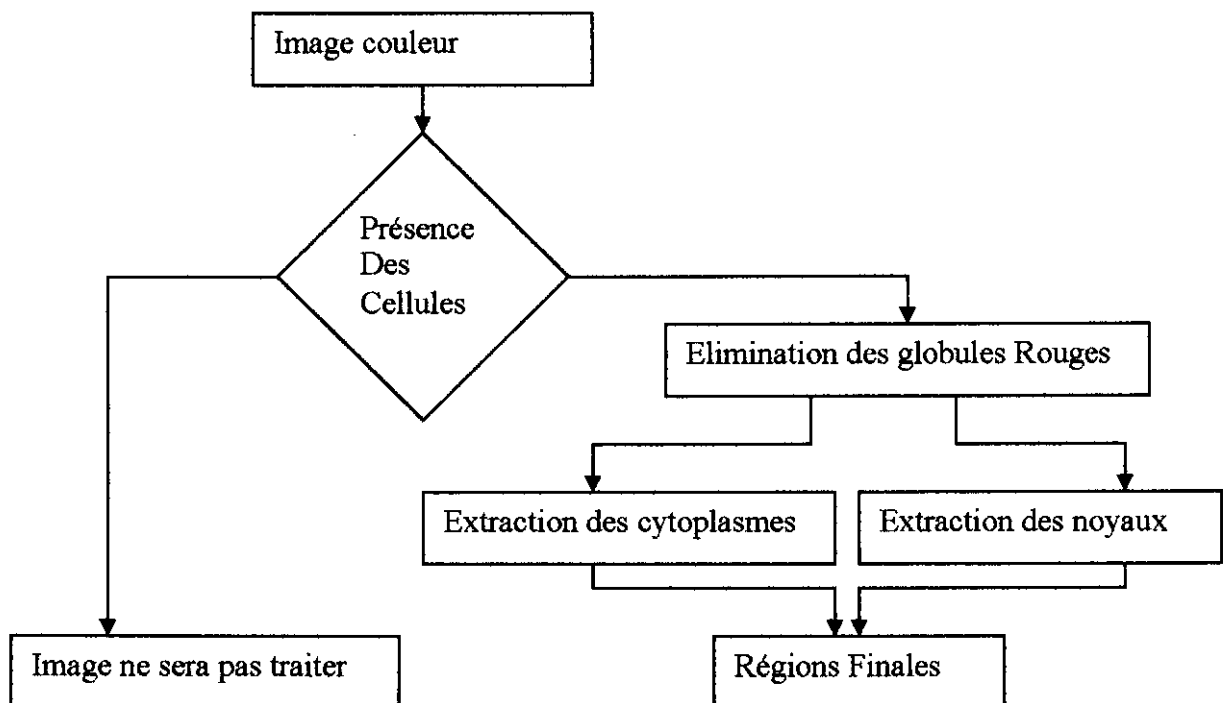


Figure II.10 Stratégie de segmentation d'images de cytologie des séreuses.

➤ **Élimination du bruit**

Lors de la segmentation, nous effectuons des lissages afin d'améliorer la qualité de l'image. Afin d'obtenir des images ne présentant pas d'aberrations, nous procédons à une division entre chaque image acquise et une image de fond (une image d'un champ microscopique vide). Les flèches montrent les aberrations présentes sur l'image originale (cf. Figure II.11.a). On constate que ces dernières n'apparaissent pas sur l'image corrigée (cf. Figure II.11.b). Ceci a plusieurs avantages: le bruit est éliminé, le contraste de l'image s'en trouve augmenté et cela permet d'homogénéiser les images entre elles (au niveau de la couleur).

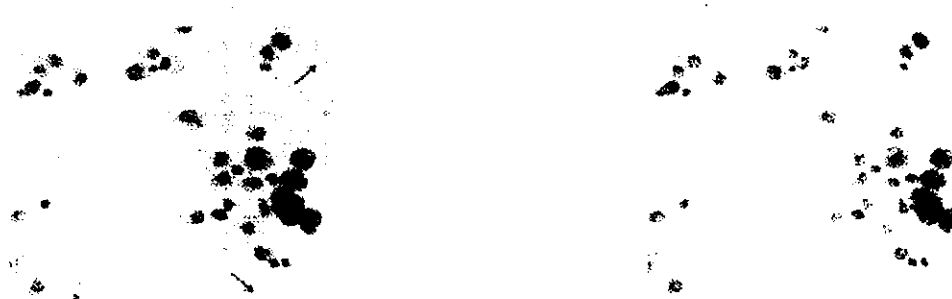


Figure II.11 Une image des séreuses avant (a) et après (b) division par une image de fond

➤ **Détection des cellules pathologiques**

Avant de segmenter une image, il faut déterminer si des cellules intéressantes sont présentes sur celle-ci. Les cellules présentant un intérêt sont les cellules possédant un noyau bleu. L'information a priori que représente la détection d'objets bleus peut se traduire directement à travers l'expression de la teinte [LEZ 02]. Un simple seuillage de la teinte dans l'espace *HSL* permet de déterminer si l'image contient du bleu ou non. En prenant deux seuils $S1$ et $S2$ définissant les bornes supérieure et inférieure d'un intervalle de teinte,

L'image résultante I_{bleue} est exprimée par :

$$I_{bleue} = E(E(S(I_{HSL}^H))) \quad \text{II.7}$$

Où

$$S(I_{HSL}^H(p)) = \begin{cases} 255 & \text{si } I_{HSL}^H(p) \in \left[\pi + \frac{\pi}{3}, \frac{3\pi}{2} \right] \\ 0 & \text{sinon} \end{cases}$$

Si L'image binaire finale I_{bleue} est égale à l'ensemble vide (i.e. si $\exists p(x, y) \in I_{bleue}$ tel que $I_{bleue}(p) > 0$), alors aucun noyau bleu n'est pas présent sur l'image et l'image ne sera pas traitée (cf. Figure II.12).

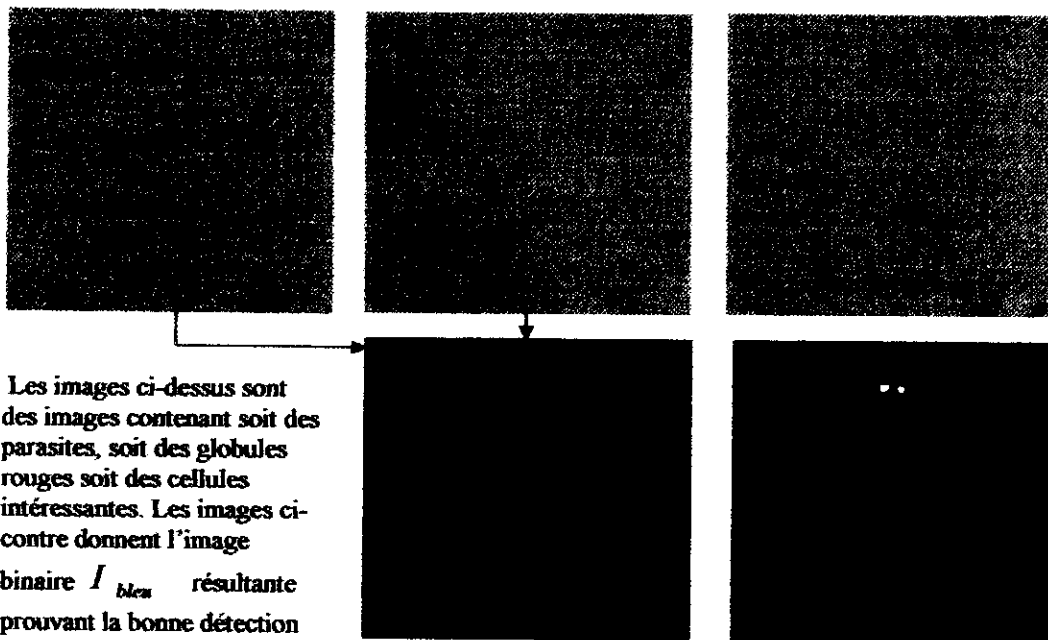


Figure II.12 Exemple d'extraction des cellules

➤ **Élimination des globules rouges**

Nous avons pu déterminer la présence de cellules intéressantes caractérisées par leur noyau bleu (cf. Figure II.12). Cependant sur certaines images présentant un fond hémorragique, certaines particules (les globules rouges) peuvent gêner des traitements ultérieurs. Parfois, seules quelques cellules intéressantes sont présentes et par contre beaucoup de globules rouges apparaissent sur l'image.

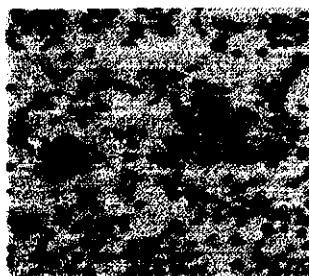
Pour éliminer les objets non intéressants, nous avons mis en oeuvre une méthode identique. Le résultat de ces opérations est un masque I_{Rouge} divisant l'image en deux régions: les globules rouges et le reste de l'image (les cellules et le fond).

Par ce masque, nous réduisons considérablement le nombre de cellules à segmenter sans pénaliser la future reconnaissance des cellules intéressantes [LEZ 02].

I_{Rouge} s'exprime par :

$$I_{Rouge} = O(S(I_{HSL}^H)) \text{ Avec } S(I_{HSL}^H(p)) = \begin{cases} 255 & \text{si } I_{HSL}^H(p) \in \left[0, \frac{\pi}{3}\right] \cup \left[\frac{3\pi}{2}, 2\pi\right] \\ 0 & \text{sinon} \end{cases} \quad \text{II.8}$$

Ce masque sera utilisé lors des étapes suivantes pour déterminer les marqueurs dans le processus de la ligne de partage des eaux couleur (cf. Figure II.13). On remarquera que le masque a été conçu pour éliminer également certains débris cellulaires (cf. Figure II.14).

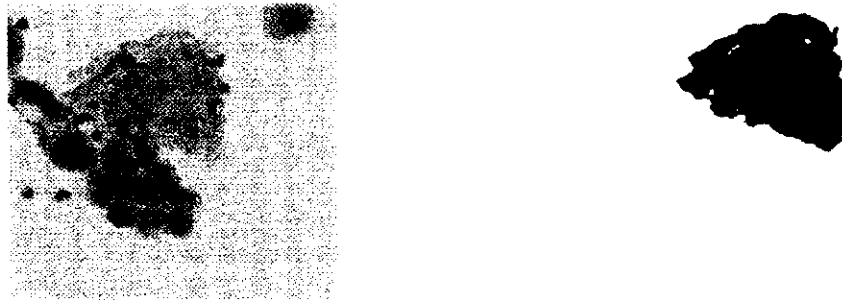


a) Une image à fond hémorragique



b) Le masque I_{Rouge} correspondant

Figure II.13 Une image de cellules contenant des globules rouges et le masque correspondant



a) Une image contenant un débris cellulaire

b) le masque correspondant

Figure II.14 L'image a et son masque correspondant

➤ Segmentation des régions cytoplasmiques

Pour segmenter les régions cytoplasmiques des images, nous avons utilisé la ligne de Partage des eaux couleur. Nous allons voir chacune des étapes de la stratégie d'optimisation de la fonction d'agrégation pour les images traitées.

1- **Le choix de l'espace couleur** a été réalisé sur des images représentatives des différentes cellules et des différents fonds pouvant être rencontrés. On trouve également sur ces images les diverses configurations spatiales possibles des cellules ou amas de cellules. Sur ces images nous avons appliqué la méthode de contraste qui sert à déterminer quel espace est le mieux adapté à un processus de segmentation.

La première étape consiste à segmenter manuellement l'image afin d'isoler les zones ou objets à reconnaître. Ensuite nous calculons le contraste couleur local de l'image dans chacun des espaces couleur C_1, C_2, C_3 parmi $RVB, XYZ, L*u*v, L*a*b, HSL$ puis nous calculons le contraste global (somme des contrastes locaux) pour les points se trouvant aux frontières des régions extraites à l'étape précédente [LEZ 02]. Les contrastes couleur locale et globale sont obtenus de la manière suivante :

Soit $p(x, y) \in I$ avec I une image couleur (\forall son espace couleur).

$$c(P) = \left\{ \max (\text{dist. } (P, M)) / M \text{ soit un 8 voisin de } P \right\}.$$

Le contraste couleur global est alors : $C = \sum (c(P))$ tel que P est un point de frontière entre deux régions.

Les résultats sont donnés dans le Tableau II.1, L'espace couleur présentant le meilleur contraste pour toutes les images est l'espace RVB. Cet espace sera utilisé pour calculer la L.P.E couleur

FOND	RVB	XYZ	L*U*V*	L*A*B*	HSL
Homogène	67.20	64.25	45.19	27.83	64.49
Hémorragique	87.59	84.45	60.61	34.53	85.23
Muco-rotéinique	60.59	58.05	53.35	25.95	59.39
Moyenne	71.79	68.91	53.05	29.43	69.70

Tableau II.1 Les contrastes suivant les différents fonds des images

2- la métrique est une métrique euclidienne : la notion de distance entre les couleurs se traduit généralement par une distance euclidienne. Voici les distances pour les espaces couleurs que nous avons vus précédemment :

$$\text{RVB} : \Delta d = \sqrt{\Delta R^2 + \Delta V^2 + \Delta B^2}$$

$$\text{XYZ} : \Delta d = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2}$$

$$\text{L}^* \text{u}^* \text{v}^* : \Delta d = \sqrt{\Delta L^{*2} + \Delta u^{*2} + \Delta v^{*2}} \quad \text{II.9}$$

$$\text{L}^* \text{a}^* \text{b}^* : \Delta d = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}}$$

$$\text{HSL} : \Delta d = \sqrt{\Delta H^2 + \Delta S^2 + \Delta L^2}$$

Cette distance est évidemment liée aux performances de l'espace couleur. Elle est meilleure si l'espace est perceptuellement uniforme.

3- Le calcul du gradient couleur : le calcul du gradient [LEZ 02] dans un espace couleur peut se faire sur une seule, sur deux ou trois composantes. Cependant l'utilisation d'un gradient calculé sur les trois composantes à la fois tiendra mieux compte de la dispersion des couleurs et sera mieux approprié à un processus de type croissance de régions. Une image couleur peut être vue comme une fonction $F : R_2 \rightarrow R_m$ où m est le nombre de composantes.

Les filtres utilisés pour calculer le gradient suivant les direction x et y sont respectivement FX et FY définis pour un voisinage huit (i.e. on a une trame carrée) :

$$FX = \frac{1}{12} \begin{pmatrix} 1 & 4 & 1 \\ 0 & 0 & 0 \\ -1 & -4 & -1 \end{pmatrix} \quad FY = \frac{1}{12} \begin{pmatrix} 1 & 0 & -1 \\ 4 & 0 & -4 \\ 1 & 0 & -1 \end{pmatrix} \quad \text{II.11}$$

4- Pour déterminer les marqueurs, les informations a priori que nous pouvons utiliser sont la couleur verte du cytoplasme ainsi que la transition franche entre le fond et le cytoplasme quel que soit le type du fond. Cette dernière propriété se traduit directement par l'information de la Teinte (H) [LEZ 95]. Pour que l'extraction des marqueurs se fasse de façon précise et reproductible, nous avons choisi la teinte de l'espace couleur *HSL* car elle satisfaisait à ces exigences. Après un lissage par un filtre exponentiel, l'image présente des transitions plus franches et moins bruitées entre les cytoplasmes et le fond. Pour éliminer les globules rouges, il suffit de calculer une intersection morphologique entre l'image I_{bleue} et le masque des globules rouges.

L'image résultante doit être suivie par un calcul de gradient afin d'effectuer la binarisation. Les marqueurs sont érodés afin d'augmenter leur zone d'influence pour obtenir une meilleure précision dans la définition des régions cytoplasmiques. L'image des marqueurs cytoplasmiques (cf. Figures II.15. a. et c) s'exprime par :

$$I_{CM} = E(B(\nabla(I_{Rouge} \cap S(Exp(I_{(HSL)}^H)))))) \quad \text{II.12}$$

5- détermination de l'image cytoplasmique : Comme nous l'avons vu, la limite entre le fond et les cytoplasmes est bien marquée et l'information gradient est suffisante pour extraire les cytoplasmes des cellules. Nous avons fixé α à 0.5 afin d'utiliser les deux informations (locale et globale). L'image résultante obtenue (cf. Figure II.15. d) par la ligne de partage des eaux couleur est :

$$I_{cytoplasme} = I_C = LPEC(\nabla(I_{RVB}), LAB(I_{CM}), I_{RVB}). \quad \text{II.13}$$

Par cette méthode les cytoplasmes des cellules sont correctement segmentés (cf. Figure II.15) quelles que soient les cellules à segmenter mais cela ne sépare pas les amas de cytoplasmes qui présentent peu ou pas de différences d'intensité entre tous les cytoplasmes de l'amas.

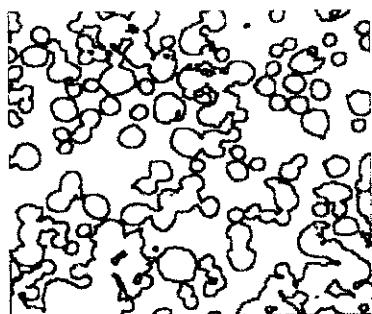
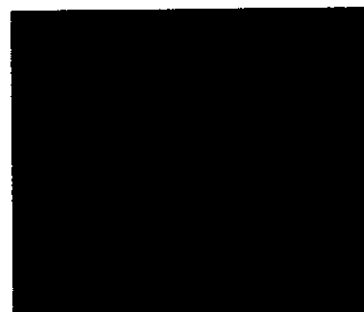
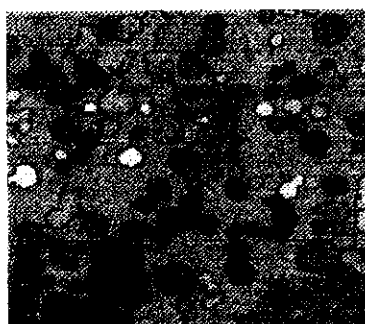
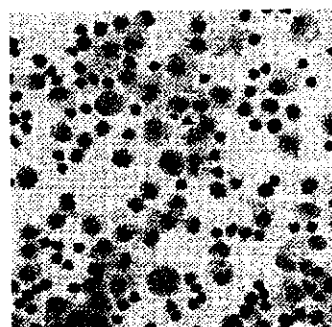
a) Les marqueurs I_{CM} b) Le gradient $\Delta(I_{RVB})$ c) Les marqueurs étiquetés LAB (I_{CM})d) Les régions cytoplasmiques I_C

Figure II.15 Les marqueurs (a), le gradient (b), les marqueurs (c) et les régions cytoplasmiques en jaune (d).

➤ Segmentation des régions nucléaires

La segmentation des régions nucléaires est également basée sur la ligne de partage des eaux couleur :

1. Le choix de l'espace couleur se fait sur des images représentatives des noyaux des cellules pouvant être rencontrés. Comme nous l'avons vu précédemment les noyaux ont une extrême variabilité dans leur configuration spatiale et de leur couleur. Nos images

contiennent suffisamment de cellules dont les noyaux sont représentatifs pour avoir une mesure du contraste valide et fiable. Le *Tableau 4* donne les mesures de contraste pour chacun des contrastes valide et fiable ainsi que les mesures de contraste pour chacun des espaces couleur. Nous avons pris trois types de cellules. Le type 1 correspond à une image contenant des cellules aux noyaux isolées ou accolés, le type 2 à des cellules aux noyaux isolés ou en amas et le type 3 à des cellules dont le noyau est isolé ou accolé dont le noyau est bleu très clair.

L'espace *RVB* qui présente la meilleure moyenne du contraste est choisi.

2. La métrique est une métrique euclidienne.

3. Le calcul du gradient couleur est effectué dans l'espace *RVB*; puisque c'est cet espace qui a été sélectionné (cf. Figure 16. b).

4. Pour trouver les marqueurs des noyaux il faut déterminer une composante connexe et une seule par noyau ou amas de noyaux. La difficulté posée par l'extraction des marqueurs réside, dans l'extraction des noyaux clairs et des noyaux foncés ainsi que dans les faibles transitions couleurs que l'on peut rencontrer entre les noyaux et les cytoplasmes.

Pour que l'extraction des marqueurs se fasse de façon précise et reproductible, nous avons choisi la luminosité de l'espace couleur *HSL* car elle satisfaisait à ces exigences. Après un lissage par un filtre exponentiel ; les étapes d'extraction des noyaux se font de la même manière que pour l'extraction des cytoplasmes.

L'image finale représente les marqueurs nucléaires (*Figure 15.a et c*) et s'exprime par

$$I_{NM} = E(B(\nabla(I_{Rouge} \cap S(Exp(I_{HSL}^L)))))) \quad \text{II.14}$$

5. Les noyaux sont obtenus par la ligne de partage des eaux couleur et l'image finale s'exprime par : $I_{Noyaux} = I_N = LPEC(\nabla(I_{RVB}), LAB(I_{MN}), (I_{RVB}))$. Le coefficient de pondération α est fixé à .5 ; ceci permet de prendre en compte à la fois les informations locales et globales. L'utilisation de ces deux informations donne des régions nucléaires plus précises car l'une d'entre elles ne suffirait pas. Ainsi l'utilisation du critère d'homogénéité couleur permet d'obtenir des régions nucléaires uniformes et dont les contours sont bien déterminés (cf. Figure 15.d).

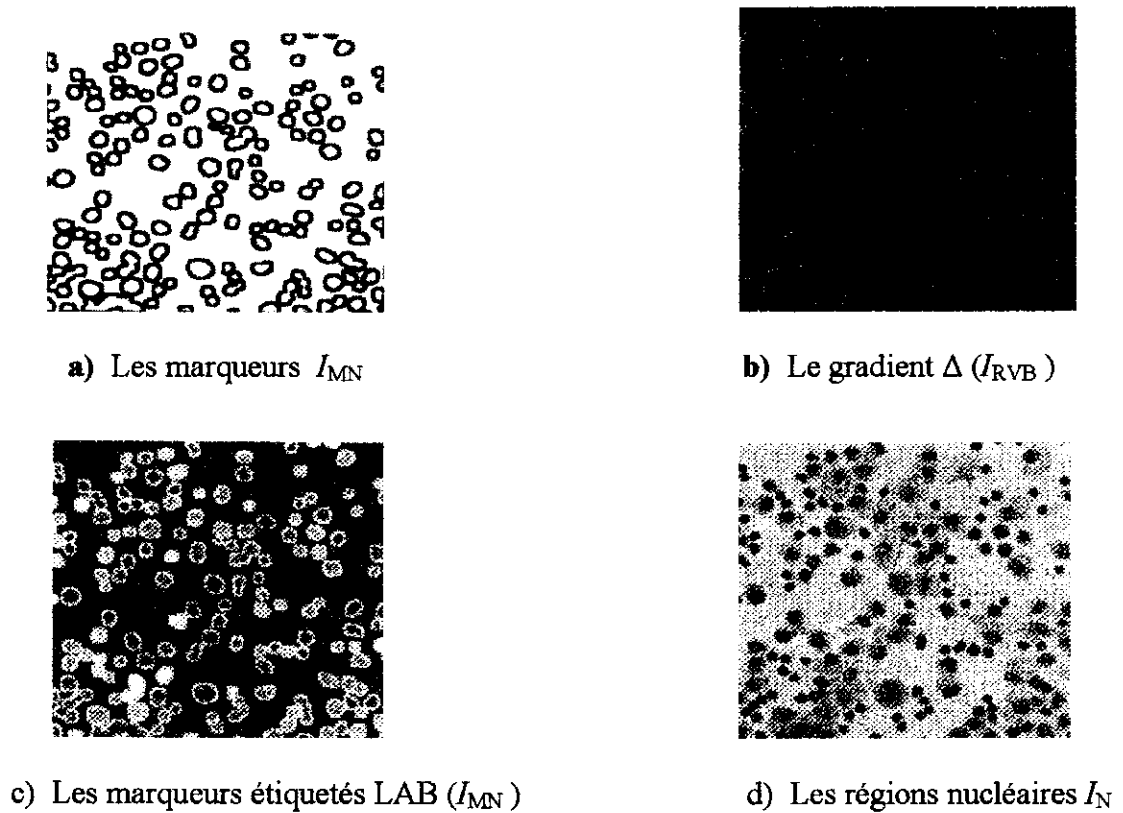


Figure II.16 Les marqueurs (a), le gradient (b), les marqueurs (c) et les régions nucléaires en jaune (d).

V. Conclusion

Dans cette deuxième partie, nous nous sommes intéressés à la segmentation d'images cytologiques couleurs. La ligne de partage des eaux étant particulièrement adaptée pour segmenter des images complexes telles que les images couleurs, la méthode que nous avons adoptée permet d'extraire les cytoplasmes et les noyaux en éliminant d'autres particules. Cette méthode utilise l'information couleur comme une information a priori pour l'extraction des régions intéressantes; examine aussi les différents espaces de couleur afin de choisir l'espace le mieux approprié au traitement étudié.

Le chapitre suivant, aborde une analyse quantitative de régions extraites suivant des paramètres texturaux et morphologiques, pour une classification suivant les réseaux de neurones.

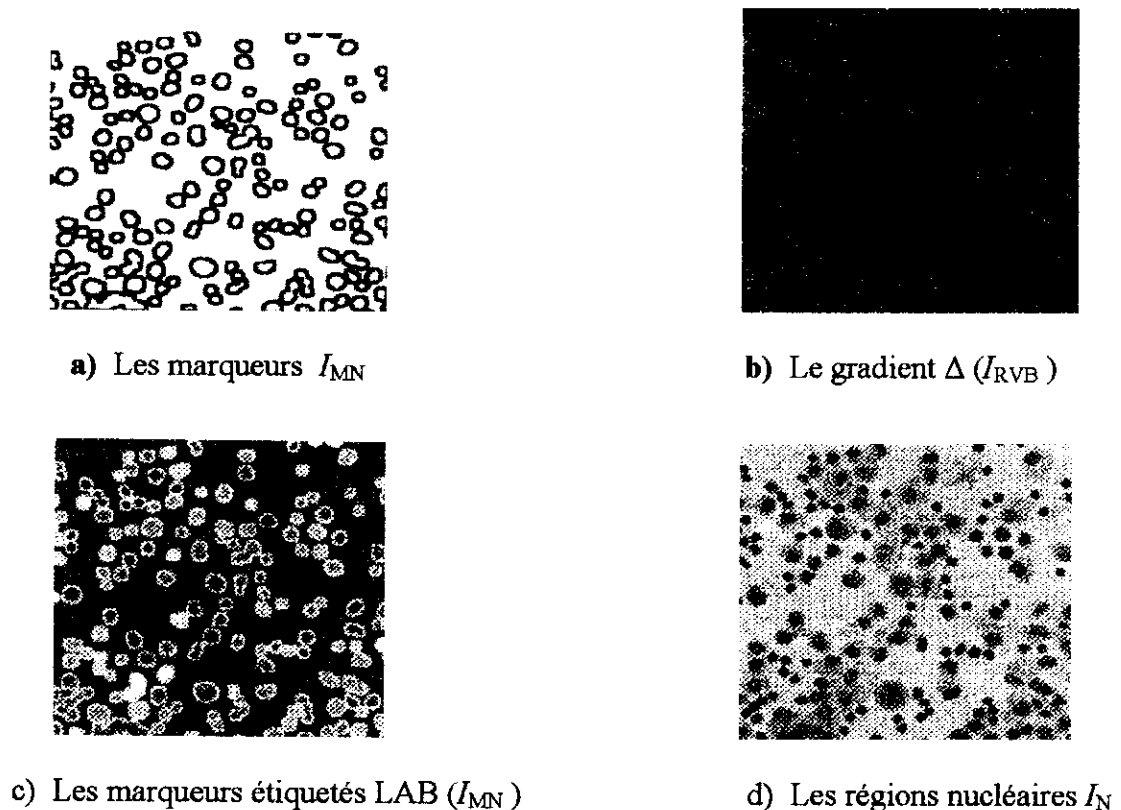


Figure II.16 Les marqueurs (a), le gradient (b), les marqueurs (c) et les régions nucléaires en jaune (d).

V. Conclusion

Dans cette deuxième partie, nous nous sommes intéressés à la segmentation d'images cytologiques couleurs. La ligne de partage des eaux étant particulièrement adaptée pour segmenter des images complexes telles que les images couleurs, la méthode que nous avons adoptée permet d'extraire les cytoplasmes et les noyaux en éliminant d'autres particules. Cette méthode utilise l'information couleur comme une information a priori pour l'extraction des régions intéressantes; examine aussi les différents espaces de couleur afin de choisir l'espace le mieux approprié au traitement étudié.

Le chapitre suivant, aborde une analyse quantitative de régions extraites suivant des paramètres texturaux et morphologiques, pour une classification suivant les réseaux de neurones.

L'objectif de cette étude, et de procéder à une analyse des caractéristiques des cellules que nous avons obtenues par application de la ligne de partage des eaux suivant la théorie de la morphologie mathématique.

I. Caractérisation des zones cellulaires

Les caractéristiques que nous avons choisies de calculer, ce sont celles qui nous paraissent les plus pertinentes pour modéliser numériquement nos cellules de façon correcte. Elles découlent de la morphologie mathématique; elles permettent d'identifier les cellules cancéreuses en se basant sur le cytoplasme et le noyau.

I.1 Description des paramètres morphologiques

Un certain nombre de critères cytologiques, permettent d'identifier la plupart des cellules cancéreuses [BEU 90]. Nous en avons retenu deux, qui figurent parmi les symptômes de cancer les plus fréquents.

- Tout d'abord, une augmentation du **rapport nucléo-cytoplasmique**, c'est-à-dire du rapport entre la surface prise par les noyaux et celle que constitue le cytoplasme, liée à une augmentation du noyau plus importante que celle du cytoplasme. En effet, dans la plupart des cellules cancéreuses, les noyaux croissent de manière anarchique et envahissent presque toute la cellule. Cette évolution anormale des noyaux permet d'identifier la plupart des cellules cancéreuses.

- Un **hyperchromatisme nucléaire**, dû à l'abondance du matériel chromatinien et à la répartition de la chromatine en mottes [BEU 90]. Cet hyperchromatisme a pour effet, à l'image, l'apparition d'une texture importante, d'un caractère "granuleux" dans les noyaux. Une granulosité trop élevée est donc également un caractère de malignité.

Nous pouvons distinguer sur l'image de la figure III.1, qui représente une coupe d'un tissu cancéreux, un noyau de grande taille à côté de noyaux plus petits, ainsi que certains noyaux allongés, et d'autres présentant une texture granuleuse.

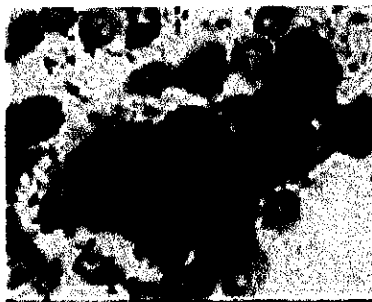


Figure III.1 coupe d'un tissu cancéreux

Ces deux paramètres ont été choisis non seulement à cause de leur importance pour le diagnostic, mais aussi de par leur caractère quantifiable. Nous allons développer une méthode qui permet d'assigner à chacun de ces deux paramètres, une valeur numérique.

Nous comptons extraire ainsi, pour chaque coupe, un vecteur de dimension deux, où chaque composante représente la valeur d'un des deux paramètres extraits. La classification d'images que nous allons effectuer se fera par un classificateur connexionniste construit grâce à une séquence d'apprentissage de vecteurs extraits, par cette méthode, d'un ensemble d'images représentatives des cellules étudiées.

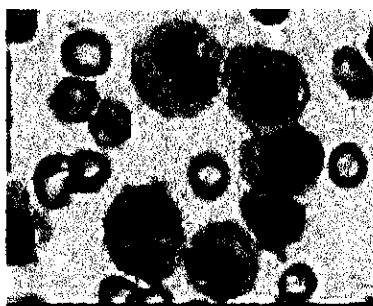


Figure III.2 Exemple d'image à traiter

I.2 Rapport nucléo-cytoplasmique

Comme nous l'avons dit précédemment, le rapport entre la surface occupée par les noyaux et celle occupée par le cytoplasme est un critère déterminant, dans l'analyse des cellules. A cet effet, nous avons créé, à partir de l'image prétraitée, deux images, l'une

isolant les noyaux, l'autre le cytoplasme. Un simple comptage des pixels, de chacune des images suffit pour extraire le rapport nucléo-cytoplasmique.

I.2.1 Création de l'image des noyaux

La création de cette image, consiste en une suite d'opérations sur l'image, permettant un seuillage aisé (où le seuil peut être choisi dans une plage assez large de valeurs de luminance).

Contrairement aux autres éléments présents dans l'image, les noyaux ont une taille importante, il s'agit là d'un caractère qui les distingue du reste plus que la luminance elle-même. Nous faisons alors usage de cette caractéristique pour les isoler grâce à une suite d'opérations de la morphologie mathématique.

Nous présentons dans ce qui suit, étapes permettant d'extraire l'image des noyaux (cf. Figure III.3) en considérant les valeurs typiques que nous avons utilisées:

- ✓ **Ouverture par un grand élément structurant :**
 - filtrage morphologique,
 - élimination de taches isolées,
 - altération de la forme.
- ✓ **Reconstruction de l'image ouverte:**
 - récupération de la forme des éléments restants.
- ✓ **Seuillage, extraction des éléments fonces (luminance < 140).**
- ✓ **Fermeture par un petit élément structurant (5-6 pixels):**
 - remplissage des trous éventuels apparus au sein des noyaux.
- ✓ **Élimination des filaments et globules ($\alpha = 15$ pixels)**
- ✓ **Ouverture par un petit élément structurant :**
 - élimination d'éventuelles taches de petite taille

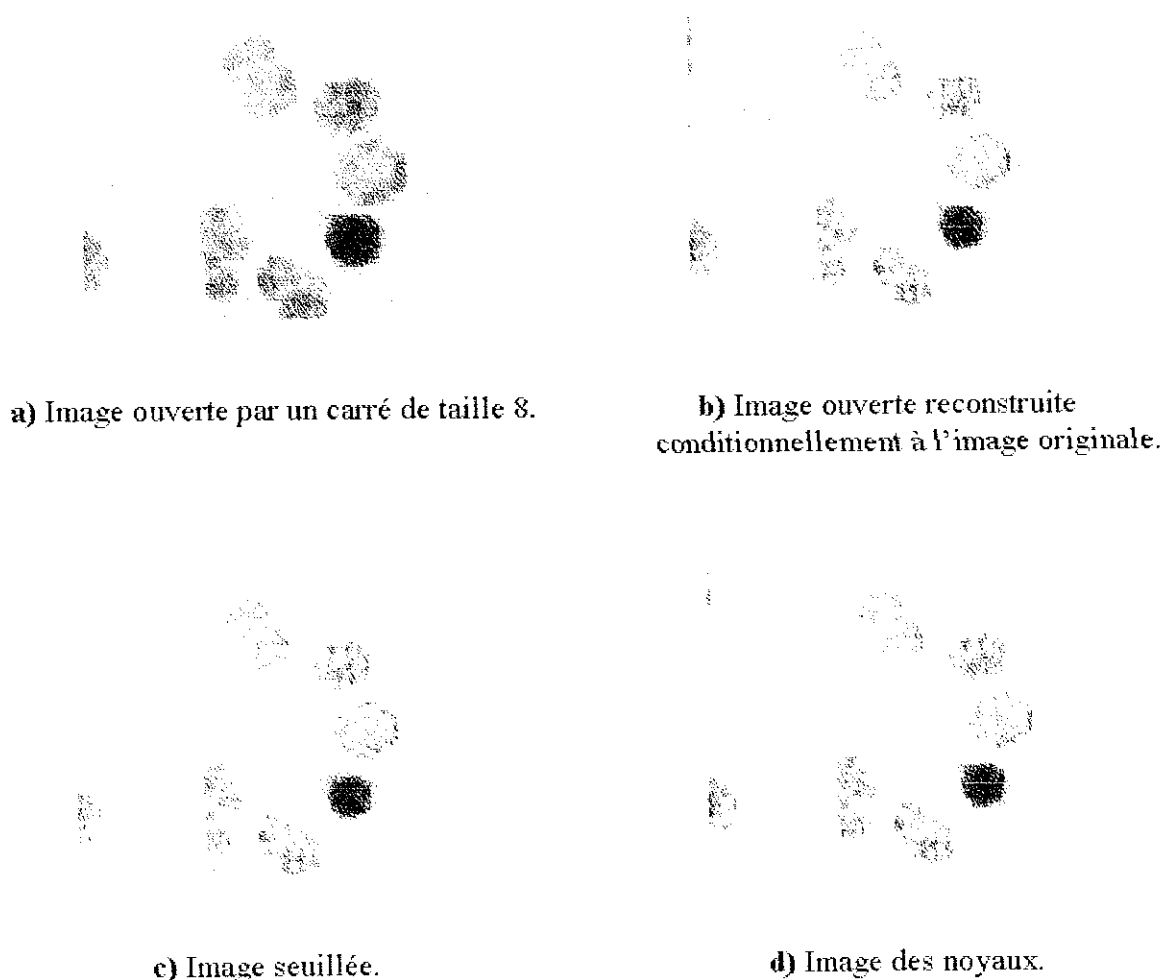


Figure III.3 Etapes d'isolations des noyaux

Nous opérons ainsi, par une ouverture sur l'image, à l'aide d'un élément structurant (carré dans notre cas), dont les dimensions ne dépassent toutefois pas la taille du plus petit noyau de l'image. L'image III.3.a, représente l'image de la figure III.2 qui a subi une ouverture par un élément structurant carré de 8 pixels. On remarque que cette ouverture, permet aux détails isolés de taille inférieure à celle de l'élément structurant de disparaître. De plus, certaines taches foncées dans des zones plus claires, ont subi le filtrage morphologique qu'effectue l'ouverture, à savoir que ces pics de luminance ont été aplanis. Néanmoins, les formes des éléments restants ont été altérées. En effet, les noyaux ont une forme plus carrée, ce qui est dû à l'élément structurant utilisé. La reconstruction permet alors, de récupérer les détails des formes sur l'image III.2.b. L'image III.2.c, permet de ne garder que les éléments foncés grâce à un seuillage, effectué sur l'image reconstruite, suivi

d'une fermeture, permettant de remplir ainsi les trous éventuels dans les noyaux. La dernière étape consiste en l'élimination des filaments, qui ne sont pas considérés comme des noyaux. L'opération, permet de distinguer les éléments filiformes de l'image en analysant le voisinage de chaque pixel non blanc, dans les huit directions cardinales (on calcule dans chacune de ces directions, la distance du pixel considéré au bord de l'objet), les distances caractéristiques devant être supérieures à une certaine longueur minimale α , de l'ordre de la largeur d'un filament. L'image III.3.d, est ainsi obtenue après élimination des filaments et à une ouverture par un élément structurant de 10 pixel.

I.2.2 Création de l'image des cytoplasmes

Nous analysons dans ce cas (cf. Figure III.4), les étapes permettant d'obtenir une image ne gardant que le cytoplasme de l'image originale.

- **Élimination des noyaux** (dans l'image originale par la différence entre l'image originale et l'image des noyaux).
- **Seuillage, élimination d'éléments foncés** (luminance < 200):
 - élimination des taches foncées,
 - élimination des filaments et globules rouges,
 - élimination d'éventuels noyaux non détectés précédemment.
- **Seuillage, élimination d'éléments clairs** (luminance > 250):
 - élimination des zones non cytoplasmiques, trop claires
- **Ouverture par un petit élément structurant** (5 pixels) :
 - remplissage de trous au sein du cytoplasme

Nous effectuons, en premier lieu, une différence d'images, à savoir que nous effaçons de l'image originale, l'image de noyaux. Toutefois, ce qui reste n'est pas forcément le cytoplasme: on y retrouve certains globules rouges isolés, d'éventuels noyaux, ainsi que des filaments. Nous procédons alors à un seuillage ne gardant que les éléments clairs. Nous effectuons en dernier, une ouverture pour combler les vides, afin de bien calculer la surface du cytoplasme.

Figure III.4 Image du cytoplasme.

I.2.3 Calcul du rapport nucléo-cytoplasmique

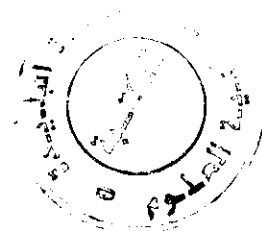
Ce calcul consiste en un simple comptage de pixels non blancs dans les deux images créées (noyaux et cytoplasme), suivi du rapport des deux nombres. Ce paramètre sera ramené à une valeur normalisée, dans le vecteur paramètre.

I.3 Hyperchromatisme nucléaire

Rappelons que le cancer provoque un accroissement anormal de l'activité nucléaire, ce qui se traduit à l'image par des noyaux granuleux, présentant une texture particulière. Il s'agit donc ici de travailler à nouveau à partir de l'image "noyaux" (cf. Figure III.3.d). Maintenant, les valeurs originales de la luminance sont à prendre en compte, puisque le but est de quantifier les sauts de luminance au sein des noyaux.

Or, l'image "noyaux" obtenue précédemment était le résultat d'une succession d'étapes dont l'une était la reconstruction d'une image ouverte. Cela avait pour effet un filtrage morphologique, avantageux pour le seuillage, mais indésirable dans le cas qui nous occupe à présent, puisque la luminance a été altérée (lissage).

Nous devons donc commencer par obtenir l'image des noyaux avec la luminance qu'ils présentaient dans l'image originale. Cela se fait aisément, en remplaçant la luminance de chaque pixel non blanc de l'image "noyaux" filtrée, par sa valeur dans l'image originale. La figure III.5 représente cette image "noyaux" originale, qui s'obtient à partir de l'image "noyaux" filtrée (figure III.3.d).



Les étapes suivantes du processus sont des opérations bien connues, utilisant la Morphologie Mathématique, dont le but est de ne garder que le caractère irrégulier de la luminance des noyaux. Comme les noyaux sont des éléments foncés, nous désirons détecter l'existence de zones claires à l'intérieur des noyaux, c'est-à-dire des creux de luminance (qui traduisent la granulosité nucléaire).

Pour ce faire, nous effectuons d'abord une fermeture de l'image "noyaux" originale à l'aide d'un élément structurant de petite taille, ce qui a pour effet un filtrage morphologique, en ce sens que des taches claires dans les noyaux, deviennent plus foncées (les petits creux de luminance sont comblés). La forme des noyaux n'est quasi pas altérée par la fermeture, de par la faible taille de l'élément structurant.

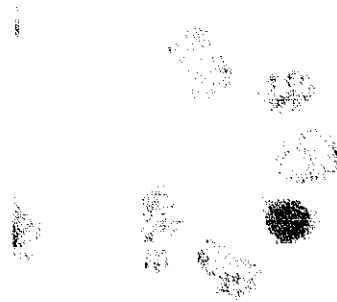


Figure III.5 Image des noyaux

L'opération suivante consiste à effectuer la différence de luminance entre l'image "noyaux" originale et son image fermée: nous obtenons une image où ne sont présentes que les irrégularités de luminance, qui avaient été filtrées par la fermeture. La figure III.6 représente une telle image, obtenue en soustrayant la figure III.6 de son image fermée. Les étapes suivantes, permettent de ne garder que le caractère irrégulier de la luminance des noyaux.

Figure III.6 Image indiquant la représentation de l'hyperchromatisme nucléaire

Les étapes permettant de quantifier l'hyperchromatisme nucléaire sont donc :

- ✓ obtention de l'image "noyaux" originale,
- ✓ fermeture par un petit élément structurant (5 pixels),
- ✓ différence entre les 2 images précédentes (obtention des creux de luminance),
- ✓ Calcul du paramètre "hyperchromatisme nucléaire".

Nous pouvons donc caractériser l'hyperchromatisme nucléaire, en totalisant les luminances de chaque pixel de l'image, puis en divisant cette somme par le nombre de pixels de l'image des noyaux (c'est-à-dire la surface nucléaire). Ce que nous obtenons est donc une moyenne des creux de luminance au sein des noyaux, qui traduisent bien l'importance de la granulosité dans les noyaux. Plus il y a une texture apparente au sein des noyaux, plus les creux de luminance seront nombreux et/ou profonds.

Les deux paramètres extraits nous permettront de classifier les tissus suivant des classes saines, pathologiques et suspectes.

II. Classification connexionniste

La classification connexionniste adoptée, répond à une classification non supervisée, suivant le réseau de Kohonen [TOU 92].

II.1 Le Neurone biologique, structure et fonctionnement

Le cerveau est composé d'environ 10^{12} neurones (mille milliards) avec 100 à 100000 synapses (connexions) par neurone qui est l'élément de base du système nerveux central

[Tou 92]. Comme illustré sur la figure III.7, le neurone est une cellule nerveuse constituée principalement de trois parties qui vis-à-vis du transfert d'informations, ont un rôle bien défini. Ces trois parties sont: les dendrites, le soma et l'axone [HER 99]. Les neurones sont interconnectés à travers des synapses.

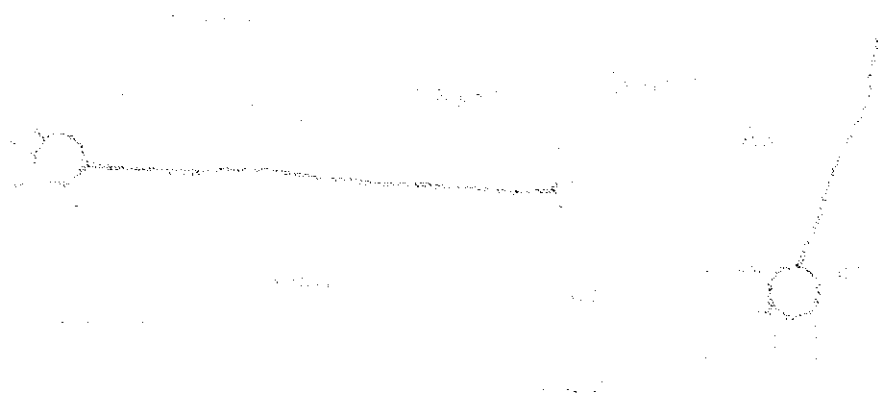


Figure III.7 Le neurone biologique

- **Les dendrites** : reçoivent l'information au niveau des points de contact avec les autres neurones, ces points de contact sont appelés synapses. L'information est ensuite acheminée vers le soma.
- **Le soma** : recueille et concentre l'ensemble des informations reçues par les dendrites où on effectue une sommation spatio-temporelle:
 - ✓ **Spatio** : parce qu'en raison de son étendue, l'arbre dendritique fait converger sur le soma des signaux venant d'un grand domaine spatial autour du neurone.
 - ✓ **Temporelle** : parce que la transmission des signaux le long d'une dendrite est caractérisée par un retard, une atténuation et un effet de filtrage passe bas.
- **L'axone** : son rôle est de transmettre sans atténuation l'information traitée par le neurone, puis la répartir sur les synapses des neurones cibles grâce à l'arborisation terminale.

Les principales fonctions des neurones sont :

- recevoir des signaux en provenance des neurones voisins,
- intégrer ces signaux,
- engendrer un flux nerveux,
- le conduire,
- le transmettre à un autre neurone capable de le recevoir.

II.2 Le neurone formel

Comme illustré sur la figure III.8, la représentation du neurone formel est issue d'une tentative de modélisation du mode de fonctionnement du neurone biologique [TOU 92].

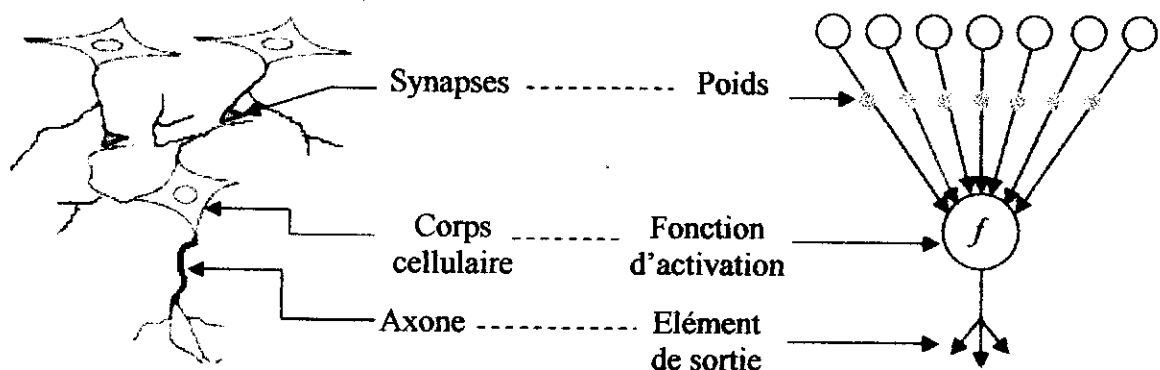


Figure III.8 Mise en correspondance : neurone biologique/neurone artificiel

Nous pouvons traduire cette mise en correspondance, par le fait que le neurone biologique (NB) et le neurone artificiel (NA) possèdent tous les deux :

- **des entrées pondérées**
 - ✓ les entrées du NA correspondent aux dendrites du NB,
 - ✓ les poids du NA correspondent aux synapses excitatrices ou inhibitrices du NB.
- **Un seuil de déclenchement,**
- **Une sortie :** la sortie (transmission de l'état) du NA correspondent à l'axone du NB.

D'après la figure III.9 qui représente la structure du neurone artificiel, la sortie du neurone est une fonction non linéaire d'une combinaison des entrées x_i pondérées par les paramètres w_i . Ces derniers sont appelés «poids» ou «poids synaptiques». A cette combinaison linéaire est ajouté un terme constant w_0 nommé «biais» [DRE 02].

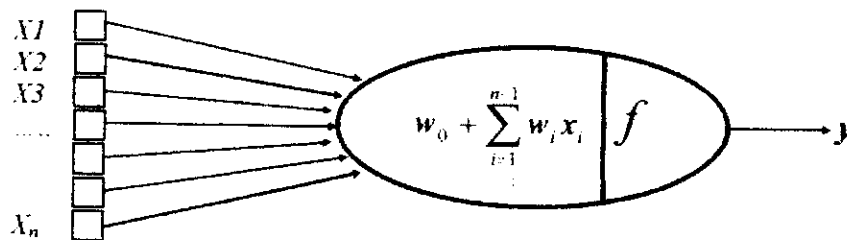


Figure III.9 Le neurone artificiel

La somme de ces deux termes permet d'obtenir l'équation suivante:

$$v = w_0 + \sum_{i=1}^{n-1} w_i x_i \quad \text{III.1}$$

La fonction f est appelée fonction d'activation. L'expression de la sortie du neurone est alors; donnée par :

$$y = f \left[w_0 + \sum_{i=1}^{n-1} w_i x_i \right] \quad \text{III.2}$$

Il existe plusieurs types de fonction d'activation. Elles peuvent être discrètes ou continues, selon l'utilisation du réseau [DRE 02].

II.3 Configuration des réseaux de neurones

Comme nous venons de le voir, un neurone réalise simplement une fonction non linéaire, paramétrée, de ses variables d'entrée. L'intérêt des neurones [HER 99] réside dans

les propriétés qui résultent de leur association en réseaux, c'est-à-dire de la combinaison des fonctions non linéaires réalisées par chacun des neurones.

On distingue deux types de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés.

II.3.1 Les réseaux de neurones non bouclés

Comme il est montré sur la figure III.10, un réseau de neurones non bouclé est représenté par un ensemble de neurones «connectés» entre eux, de façon à ce que l'information circule des entrées vers les sorties sans retour en arrière.

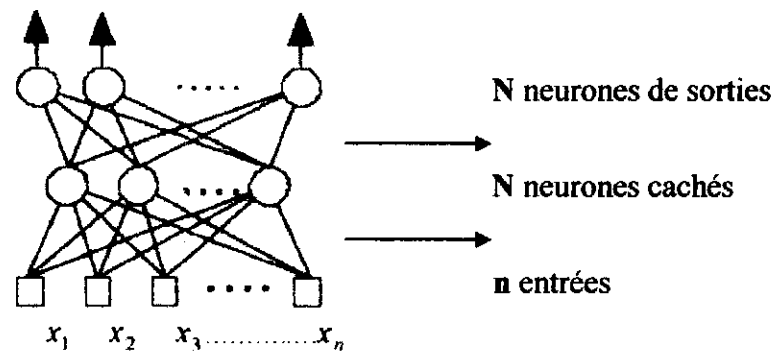


Figure III.10 Réseau à couches non bouclé

Remarques

✓ Dans un réseau de neurones non bouclé, le temps ne joue aucun rôle fonctionnel: si les entrées sont constantes, les sorties le sont également.

✓ Le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone est négligeable et, fonctionnellement, on peut considérer ce calcul comme instantané. Pour cette raison, les réseaux non bouclés sont souvent appelés «réseaux statiques», par opposition aux réseaux bouclés ou «dynamiques».

II.3. 2 Les réseaux de neurones bouclés

Lorsqu'on se déplace dans un réseau de neurones bouclé (récurrents ou dynamiques) en suivant le sens des connexions [DRE 02], il est possible de trouver au moins un chemin qui revient à son point de départ (un tel chemin est désigné sous le terme de « cycle »). La sortie d'un neurone du réseau peut donc être fonction d'elle même; cela n'est évidemment concevable que si la notion de *temps* est explicitement prise en considération.

La figure III.11 représente un exemple de réseau de neurones bouclé. Les chiffres dans les carrés indiquent le retard attaché à chaque connexion, exprimé en multiple de l'unité de temps (ou période d'échantillonnage) T . Ce réseau contient un cycle, qui part du neurone 3 et revient à celui-ci en passant par le neurone 4; la connexion de 4 vers 3 ayant un retard de 1 unité de temps.



Figure III.11 Un réseau de neurones bouclé à deux entrées

II.4 Apprentissage des réseaux de neurones

On appelle « apprentissage » des réseaux de neurones la procédure qui consiste à estimer les paramètres des neurones du réseau (les poids synaptiques), afin que celui-ci remplisse au mieux la tâche qui lui est affectée.

L'apprentissage [TOU 92] est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

On peut distinguer deux types d'apprentissages:

➤ **apprentissage supervisé** : dans ce type d'apprentissage, on connaît en tous les points ou seulement en certains points, les valeurs que doit avoir la sortie du réseau en fonction des entrées correspondantes: c'est en ce sens que l'apprentissage est « supervisé », métaphore qui signifie qu'un « professeur » peut fournir au réseau des « exemples » de ce que celui-ci doit faire, donc les exemples d'apprentissage sont des couples (entrée, sortie associée) [TOU 92].

➤ **Apprentissage non supervisé** : dans ce type d'apprentissage, il n'y a pas de « professeur », car c'est au réseau de découvrir les ressemblances entre les éléments de la base de données, et de les traduire par un entourage dans la « carte » qu'il doit produire, ainsi les exemples d'apprentissage sont seulement des valeurs d'entrées [TOU 92].

Les réseaux à apprentissage non supervisé les plus étudiés et utilisés sont les « cartes auto organisatrices » ou « cartes de Kohonen » qui sont à la base de notre étude et qui font l'objet de la suite de ce chapitre.

III. Carte Auto organisatrice de Kohonen

On entend par « auto-organisation » un processus par lequel un système trouve sans apprentissage supervisé, une solution optimale à un problème c'est-à-dire qu'aucune intervention humaine n'est requise et que peu d'informations sont nécessaires relativement aux caractéristiques des données d'entrées [BOU 04].

Globalement, il s'agit de modèles de quantification vectorielle qui ne font appel à aucun retour d'information de l'environnement. Un tel réseau se chargera donc lui-même de découvrir les relations qui existent entre les données d'entrée. De façon générale, un réseau de neurones peut être décrit par trois caractéristiques [HUG 98] soient:

- **une architecture**: l'architecture des cartes de Kohonen comme le montre la figure III.12, est constituée:
 - d'une couche d'entrée qui comporte p unités d'entrées,
 - d'une couche de sortie qui contient N neurones.

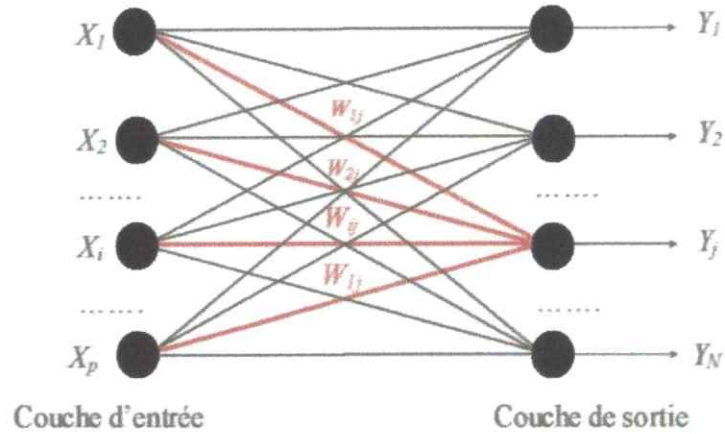


Figure III.12 Architecture du réseau de Kohonen montrant la relation entre le vecteur d'entrée, les poids des connections w_{ij} et le vecteur de sortie y

➤ **Une règle de transmission**

Soient :

$x = [x_1, x_2, \dots, x_p]$: Le vecteur d'entrée appliqué au réseau où p est le nombre d'unités d'entrée.

$w_j = [w_{j1}, w_{j2}, \dots, w_{jp}]$: Le vecteur des poids synaptiques relatif au neurone de sortie j .

$y = [y_1, y_2, \dots, y_N]$: Le vecteur de sortie du réseau où N est le nombre d'unités de sortie du réseau.

La sortie du réseau peut être exprimée sous la forme générale suivante :

$$y_j = \phi \left[\sum_{i=1}^p w_{ji} \cdot x_i \right] \quad \text{Avec } j=1, \dots, N \quad \text{III.3}$$

Où :

ϕ : est une fonction non linéaire, elle limite la valeur de y_j et assure que $y_j \geq 0$.

$$\sum_{i=1}^p w_{ij} \cdot x_i : \text{Réponse pondérée des signaux d'entrée sur un neurone } j.$$

➤ Règle d'apprentissage

L'apprentissage du réseau consiste à trouver le neurone gagnant i à l'aide d'une mesure de similarité et de mettre à jour les poids des neurones concernés, c'est-à-dire ceux qui se trouvent à l'intérieur du voisinage topologique [BOU 04].

La première étape de l'apprentissage s'effectue en choisissant aléatoirement un vecteur de l'ensemble d'apprentissage x . Ce vecteur est ensuite comparé à chacun des vecteurs poids. Le neurone gagnant $i(x)$ est ensuite identifié à l'aide d'une mesure de similarité calculée par la distance euclidienne entre les deux vecteurs :

$$d_i = \|x_i - w_{wi}\| = \sqrt{\sum_{i=1}^n (x_i - w_{ji})^2} \tag{III.4}$$

$$i(x) = \arg \min \|x - w_j\| \tag{III.5}$$

Où :

$$\| \| : \text{est la norme euclidienne telle que définie par } \|E\| = \sqrt{\sum_{i=1}^m E_i^2}$$

Une fois le vecteur poids gagnant trouvé, le réseau s'ajuste de manière à mieux représenter ce vecteur d'apprentissage. La règle d'apprentissage [BOU 04] pour un neurone j appartenant au voisinage topologique variable du neurone gagnant $i(x)$, décrit une relation des vecteurs poids vers le vecteur d'entrée et s'énonce comme suit :

$$w_j(n+1) = \begin{cases} w_j(n) + \eta(n)[x - w_j(n)] & \text{ssi } j \in V_{i(x)}(n) \\ w_j(n) & \text{sin on} \end{cases} \tag{III.5}$$

Où :

$V_{i(x)}(n)$: dénote le voisinage topologique autour du neurone gagnant au temps discret n .

$\eta(n)$: Fonction de voisinage qui contient une fonction de taux d'apprentissage décroissante en fonction du temps; elle détermine aussi le taux d'adaptation des neurones situés à l'intérieur de $V_{i(x)}(n)$.

Remarques

- ✓ Plusieurs fonctions peuvent être utilisées pour le voisinage tant qu'elles représentent la condition de non croissance.
- ✓ La convergence du réseau est établie quand la carte résultante ne change plus de façon significative.

IV. Classification par carte auto organisatrice

Comme nous l'avons précisé précédemment, nous allons pour ce faire utiliser un réseau de Kohonen. Nous l'avons choisi à deux entrées, puisque l'espace des vecteurs d'entrée a cette dimension, et possédant trois neurones de sortie, correspondant à trois régions possibles de l'espace d'entrée: la région regroupant les cellules saines, celle correspondant à des cellules cancéreuses, et une région intermédiaire, que nous qualifierons de cellules suspectes. La propriété d'ordonnement des noeuds de sortie, propre à Kohonen, joue un rôle prépondérant, car elle nous permettra de placer notre diagnostic sur une échelle linéaire. En effet, comme le montre la figure III.13, si l'on identifie par exemple le neurone 1 comme étant celui qui correspond à la région des cellules saines, alors nous avons la certitude que le neurone 3 correspondra à une région de l'espace d'entrée regroupant des

cellules plus éloignées des cellules saines que celles de la région 2. Nous pouvons alors, affirmer que la région 3 sera celle des cellules cancéreuses, et la région médiane celle des cellules suspectes.

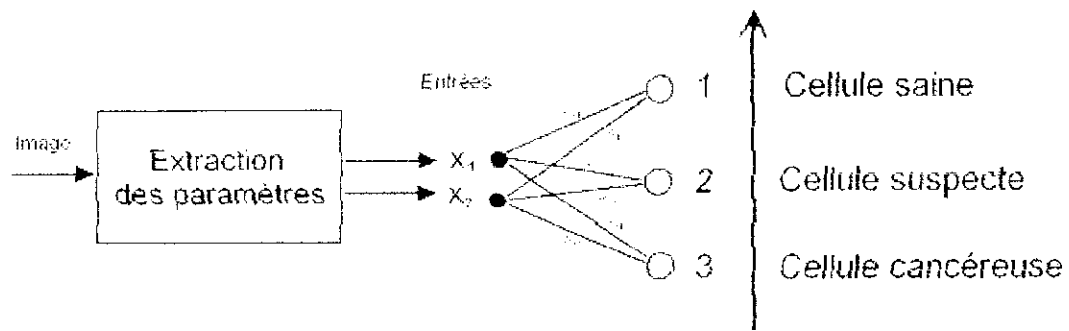


Figure III.13 principe général de la classification

Le principe général de la classification étant posé, il nous faut maintenant procéder à l'apprentissage du réseau de Kohonen, afin d'en déterminer les trois vecteurs poids.

Comme le montre la Figure III.14, notre méthode classification connexionniste passe par deux étapes.

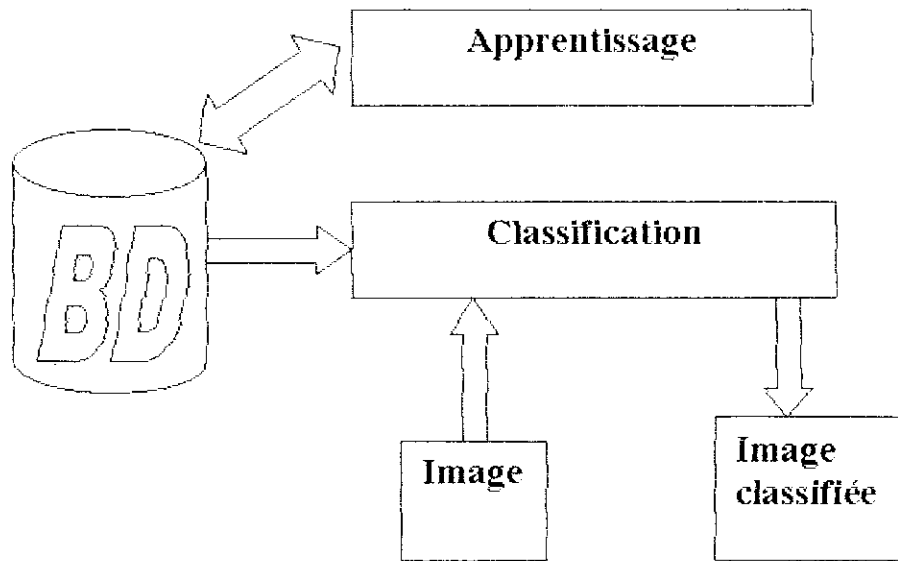


Figure III.14 Principe de la classification par le réseau de Kohonen

VI.1 Apprentissage

- Nous avons donc besoin d'une séquence d'entraînement, composée de vecteurs (de dimension 2) représentatifs de l'ensemble des vecteurs susceptibles d'être traités. Pour cela, nous utilisons des images que nous utiliserons comme test lors de la phase de reconnaissance.
- Une fois obtenus, ces paramètres morphologiques sont introduits dans la carte auto organisatrice de Kohonen qui va se charger de les grouper en classes distinctes.

L'algorithme d'apprentissage donné par les étapes ci-dessous :

Algorithme du réseau de Kohonen

- **Initialisation** : choix aléatoire des vecteurs poids $w_j(0)$, $j=1, 2, \dots, N$ où N est le nombre de neurones dans la structure.
- **Présentation** : tirage aléatoire d'une entrée x dans la base d'apprentissage afin de la présenter au réseau.
- **Similarité** : recherche du neurone gagnant $i(x)$ en utilisant le critère de la distance euclidienne minimale.
- **Modification des poids** : ajuster les poids des neurones appartenant au voisinage topologique $V_{i(x)}(n)$ suivant le pas d'adaptation $\eta(n)$.
- **Continuation** : retour à l'étape **présentation** si la base d'apprentissage contient d'autres entrées à apprendre ou bien si le but d'apprentissage n'est pas atteint.

L'algorithme utilisé dans le cadre de notre travail, est indiqué dans ce qui suit:

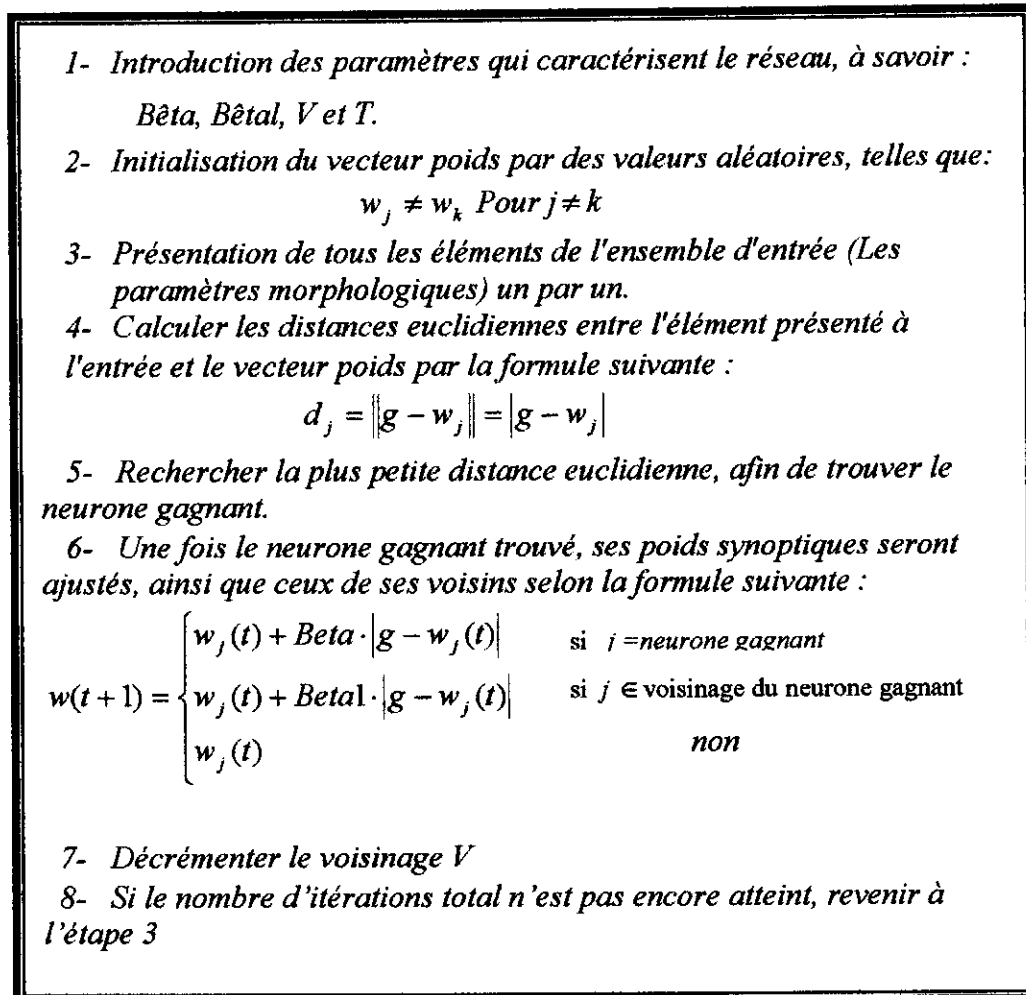


Figure III.15 Algorithme d'apprentissage de réseau

de Kohonen

VI.2 Classification

Nous disposons à présent des trois centroïdes correspondant aux trois régions envisagées. Nous pouvons donc procéder à la classification d'une image inconnue, afin de décider si elle représente des cellules cancéreuses ou non, selon les critères retenus.

La première étape de cette démarche consiste à extraire de l'image à classifier le vecteur caractéristique, correspondant aux deux valeurs des critères malignité étudiés. La classification s'effectuera donc en deux étapes.

I. Introduction

Dans ce chapitre, nous présentons chacune des étapes de la méthodologie développée, illustrée par des résultats sur des images cytologiques, dans le but d'une classification connexionniste, étudiée dans le chapitre précédent.

➤ Matériel utilisé

Notre application a été développée sur un PC qui présente les caractéristiques suivantes:

- ✓ un microprocesseur Pentium IV,
- ✓ une mémoire vive (RAM) de capacité 128 Mbits,
- ✓ un disque dur de 40 GHz,
- ✓ une horloge de 1.70 GHz.

➤ Langage utilisé

Nous avons réalisé notre application *morphocyt*, sous environnement windows avec « C++ Builder 5.0 » comme langage de programmation car il présente des avantages dont:

- ✓ la gestion automatique de la mémoire,
- ✓ la possibilité d'accéder au (Windows), et d'utiliser ses propres fonctions,
- ✓ langage très riche (bibliothèque graphique, mathématique...),
- ✓ la possession d'un mode hybride (texte, graphique),
- ✓ la spécification de manière interactive les propriétés des divers composants (fenêtres, boutons, zones d'édition, etc.),
- ✓ la conception de l'interface de l'utilisateur de manière particulièrement visuelle et interactive.

➤ Description des images

Les images manipulées dans notre logiciel, sont des images cytologiques couleurs de format bitmap.

II. Présentation de l'application Morphocyt

Morphocyt est un logiciel d'analyse d'images qui a pour objectif de faire la classification des images cytologiques couleurs en se basant sur la morphologie mathématique, en vue d'une aide à l'interprétation cytopathologiste.

II.1 Interface du logiciel

La fenêtre principale de Morphocyt est présentée par la figure suivante:

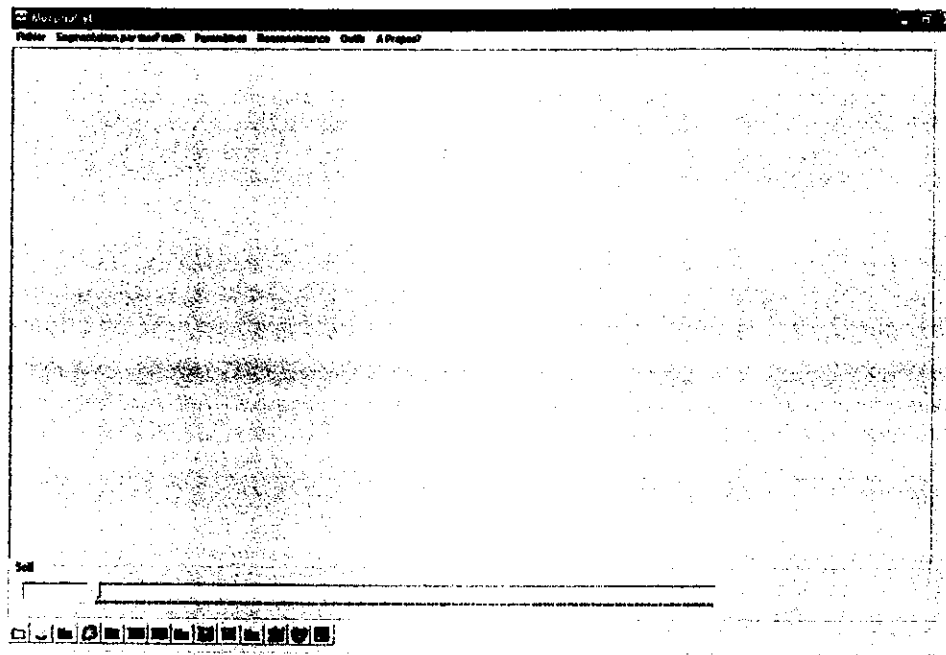


Figure IV.1 Fenêtre principale de morphocyt

Sur cette fenêtre, sont indiquées les différentes barres ci-dessous;

- ✓ barre du menu principal,
- ✓ barre des boutons,
- ✓ barre d'état.

II.2 Les menus

Dans la fenêtre principale, se trouve la barre des menus qui contient les différents menus nécessaires pour notre application.

II.2.1 Menu "fichier"

Il contient trois commandes:

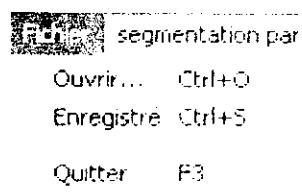


Figure IV.2 Menu "fichier"

- ✓ **Ouvrir** : L'exécution de cette commande permet d'afficher une image de type Bitmap.
- ✓ **Enregistrer** : permet de sauvegarder une image sous un nom et dans un répertoire choisi par l'utilisateur.
- ✓ **Quitter** : cette commande permet de fermer l'application après confirmation.

II.2.2 Menu "Segmentation par morphologie mathématique"

Il contient quatre sous menus:

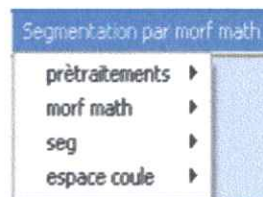


Figure IV.3 Menu "segmentation par morphologie mathématique"

- **Prétraitement** : contient trois sous menus. Cette étape permet de filtrer l'image pour l'amélioration de sa qualité.

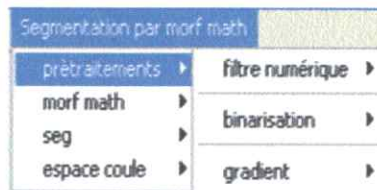


Figure IV.4 Sous menu "prétraitement"

- **Filtrage numérique**: il contient les différents filtres permettant l'amélioration des images couleurs.
 - Filtre exponentiel,
 - filtre gaussien,
 - filtre médian,
 - filtre moyen.

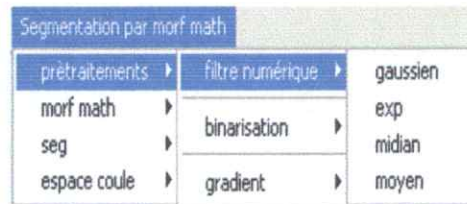


Figure IV.5 Sous menu ‘filtre numérique’

- Binarisation: c’est un sous menu composé de 3 commandes, permet de réaliser une binarisation manuelle suivant un seuil choisi par l’utilisateur:

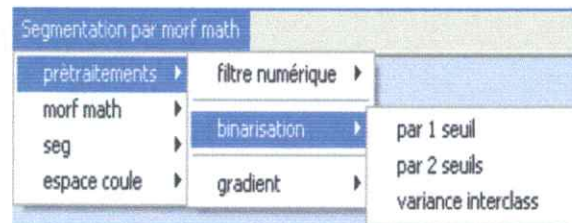


Figure IV.6 Sous menu ‘Binarisation’

- Binarisation par un seuil,
 - binarisation par deux seuils,
 - binarisation variance interclasses.
- Gradient: c’est un sous menu composé de 4 commandes qui permet l’obtention du gradient de l’image suivant;
 - Robert,
 - Sobel,
 - Prewitt,
 - gradient couleur.

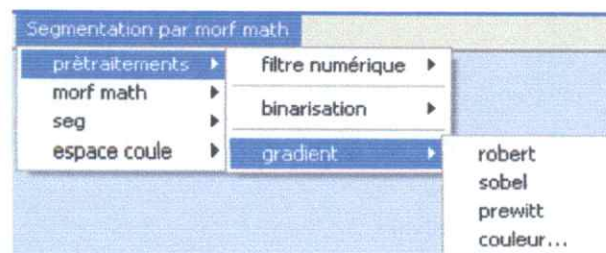


Figure IV.7 Sous menu ‘gradient’

- **Morphologie mathématique** : c'est un sous menu qui nous donne la possibilité d'exécuter les opérateurs de la morphologie mathématique sur des images cytologiques. Ce sous-menu se subdivise en cinq opérateurs:

- dilatation,
- érosion,
- ouverture,
- fermeture,
- amincissement,

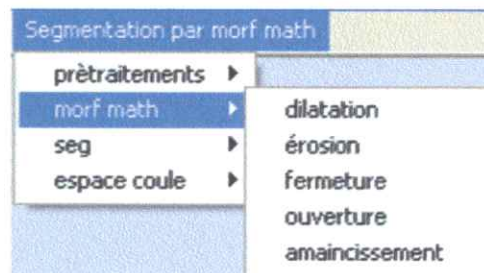


Figure IV.8 Sous menu Morphologie mathématique

- **Segmentation**: c'est un sous menu composé de cinq commandes et un sous menu.

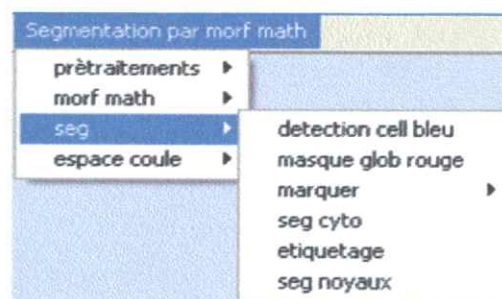


Figure IV.9 Sous menu Segmentation

- Détection des cellules bleues : cette commande permet de détecter les cellules possédant des noyaux.
- Masque des globules Rouges: cette opération permet d'afficher l'image masque des globules rouges en vue de les éliminer.
- Marqueurs : c'est un sous menu composé de 2 commandes:

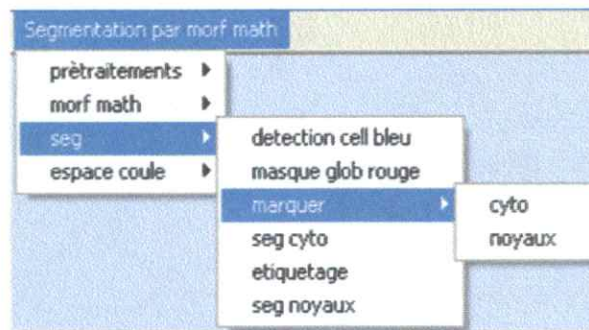


Figure IV.10 Sous menu "marqueurs"

- Marqueurs cytoplasmiques: opération qui permet d'extraire uniquement les régions cytoplasmiques.
 - Marqueurs nucléaires: opération qui permet d'extraire uniquement les régions nucléaires.
- Segmentation des cytoplastes: menu qui permet d'afficher l'image des régions cytoplasmiques.
 - Etiquetage: permet d'évaluer le nombre de contours des marqueurs.
 - Segmentation des noyaux: menu qui permet d'afficher l'image des régions nucléaires.

➤ **Espaces couleurs** : c'est un sous menu composé de quatre commandes:

- HSL1 : permet d'afficher l'image dans l'espace de couleurs HSL1.
- HSL2 : permet d'afficher l'image dans l'espace de couleurs HSL2.
- XYZ : permet d'afficher l'image dans l'espace de couleurs XYZ.
- LUV : permet d'afficher l'image dans l'espace de couleurs LUV.

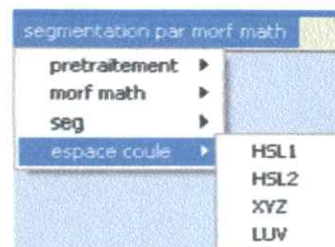


Figure IV.11 Sous menu "espaces couleurs"

II.2.3 Menu Paramètres

C'est un sous menu composé de trois commandes:

- Rapport nucléo-cytoplasmique: commande qui permet le calcul du rapport entre le cytoplasme et le noyau, l'affichage de l'image cytoplasme ainsi que celle du noyau.
- Hyperchromatisme nucléaire : l'exécution de cette commande permet de donner la valeur de l'hyperchromatisme.



Figure IV.12 Sous menus "paramètres"

II.2.4 Menu "Reconnaissance"

C'est un sous menu composé de deux commandes:

- Apprentissage : l'exécution de cette commande permet de déterminer les trois vecteurs-poids, après une séquence d'entraînement.
- Classification : permet d'afficher un message qui contient le neurone actif.



Figure IV.13 Sous menu "reconnaissance"

II.2.5 Menu Outils

C'est un sous menu composé de quatre commandes:

- Inverseur des niveaux de gris: opération permettant d'inverser les niveaux de gris dans une image.
- Histogramme: opération permettant d'afficher l'histogramme d'une image.
- Niveau de gris: commande qui permet d'afficher l'image en niveau de gris.
- Caractérisation: cette commande permet d'évaluer les différents paramètres (de taille, de forme, et texturaux).

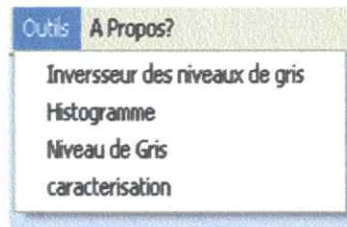


Figure IV.14 Sous menus outils

II.2.6 Menu A propos

Il affiche une fenêtre décrivant les caractéristiques de notre logiciel.

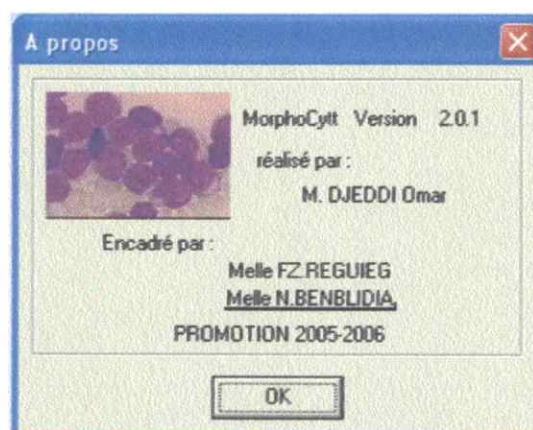


Figure IV.15 Menu à propos

II.3 Barre des boutons

Elle contient des boutons de raccourci des fonctions du menu principal (ouvrir, filtre exponentiel, détection des cellules bleues,...).

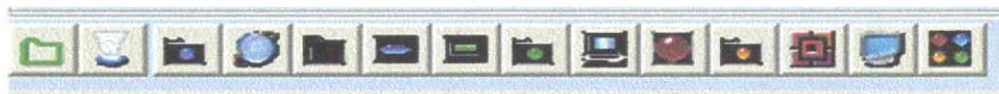


Figure IV.16 Barre des boutons

- 1- Nouvel image ;
- 2- Enregistrer ;
- 3- Filtre exponentiel ;
- 4- Détection des cellules bleues ;
- 5- Marqueurs cytoplasmiques ;
- 6- Marqueurs nucléaires ;
- 7- Le Gradient couleur ;
- 8- Marqueurs étiquetées ;
- 9- Segmentation du cytoplasme ;
- 10- Segmentation du noyau ;
- 11- Rapport nucléo-cytoplasmique;
- 12- L'hyperchromatisme nucléaire;
- 13- Apprentissage;
- 14- Classification.

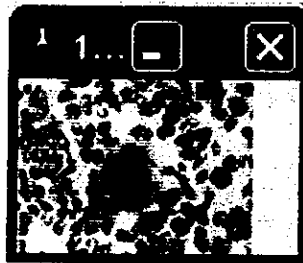
III. Présentation des résultats et discussion

Nous avons appliqué notre méthodologie sur différentes images couleurs, dont les cytologiques, objet de notre étude. L'application *morphocyt* a parcouru les étapes suivantes en vue d'une aide au diagnostic cytopathologiste.

➤ Première étape : Elimination du bruit.

L'image 'cellule 1' de la figure IV.17.a, a subi les différents filtres implantés pour améliorer la qualité de l'image, pour des raisons aussi diverses dues à l'éclairage, au type d'appareillage, ainsi qu'aux perturbations extérieures.

Il existe plusieurs opérateurs de prétraitement générés, mais le choix d'un opérateur par rapport à un autre, est étroitement lié à la nature du bruit dans l'image et aussi à ses caractéristiques. Nous avons choisi les opérateurs de réduction du bruit, tels, le filtre moyen, gaussien et exponentiel (cf. Figures IV.17)



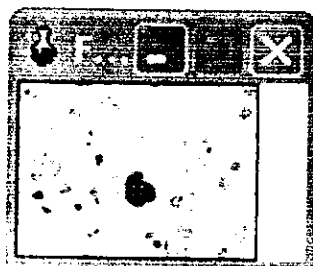
(a) Image cellulaire.



(b) image filtrée par le filtre Gaussien



(c) image filtrée par le filtre Exponentiel



(d) image filtrée par le filtre Médiane



(e) image filtrée par le filtre Moyen

Figure IV.17 Résultats des différents filtrages sur l'image cellulaire (a)

L'application des différents types des filtres sur les images couleurs, plus particulièrement les images cellulaires montre que pour un meilleur résultat de segmentation doit se faire sur une image obtenue par un filtre exponentiel.

- **Deuxième étape :** La présence des cellules sur l'image a analysé les cellules présentant un intérêt pour l'information pathologique, sont les cellules possédant un noyau bleu (cf. figure IV.18).

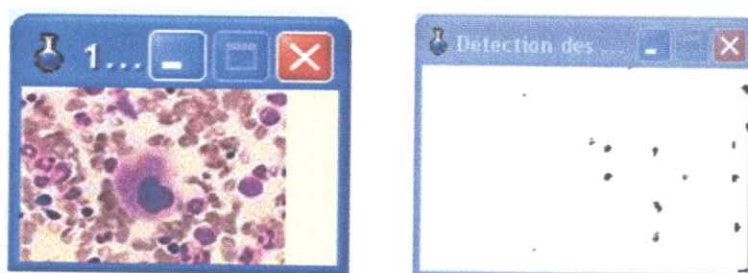


Figure IV.18 Détection des cellules d'intérêt

- **Troisième étape**

C'est la Segmentation par la morphologie mathématique. Celle-ci est basée sur la ligne de partage des eaux (L.P.E) par marqueur. Nous avons choisi comme espace couleurs, l'espace RVB car il présente un meilleur contraste par rapport à l'espace de couleurs HSL, XYZ, et $L^*u^*v^*$, dans le cas de notre étude (cf. Figures IV.19).

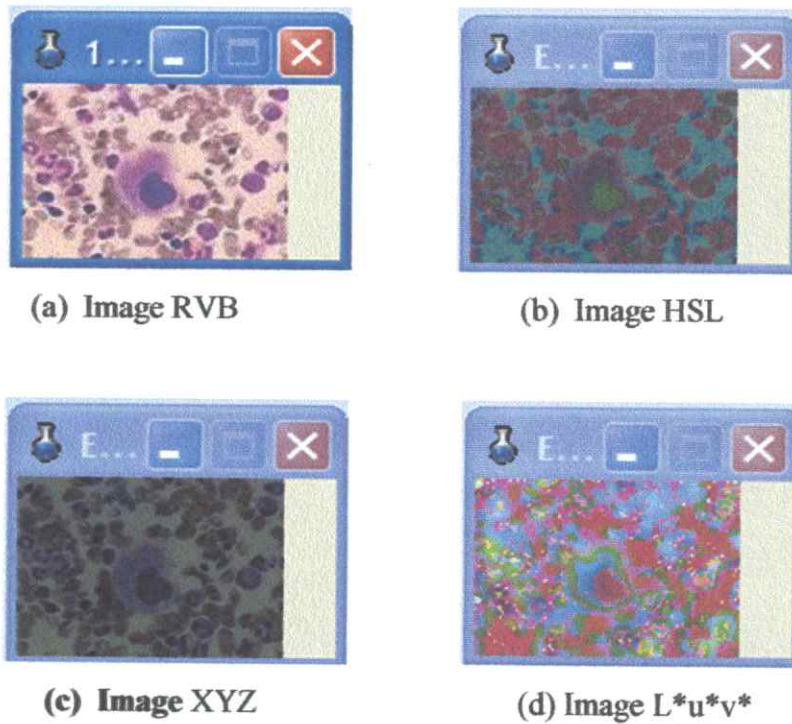


Figure IV.19 Représentation de l'image cellule (IV.14.a) dans différents espaces couleurs

Pour déterminer les marqueurs du cytoplasme et du noyau; nous avons utilisé le gradient couleur. Ces Marqueurs permettent d'extraire uniquement les régions cytoplasmiques et nucléaires. La figure suivante donne les différentes phases de segmentation pour les régions cytoplasmiques (cf. Figure IV.20), ainsi que pour les régions nucléaires (cf. Figure IV.21).



Image cellulaire

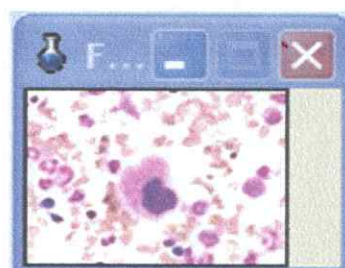
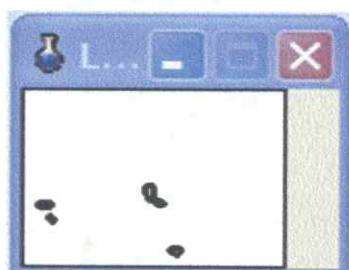
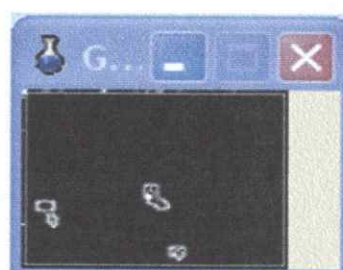


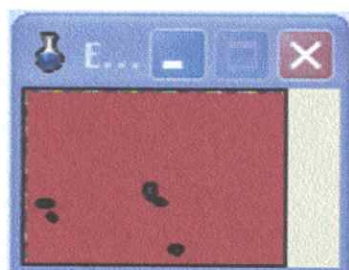
Image filtrée



Marqueurs cytoplasmiques



Gradient couleur



Marqueurs étiquetés



Régions cytoplasmiques

Figure IV.20 Segmentation des cytoplasmes

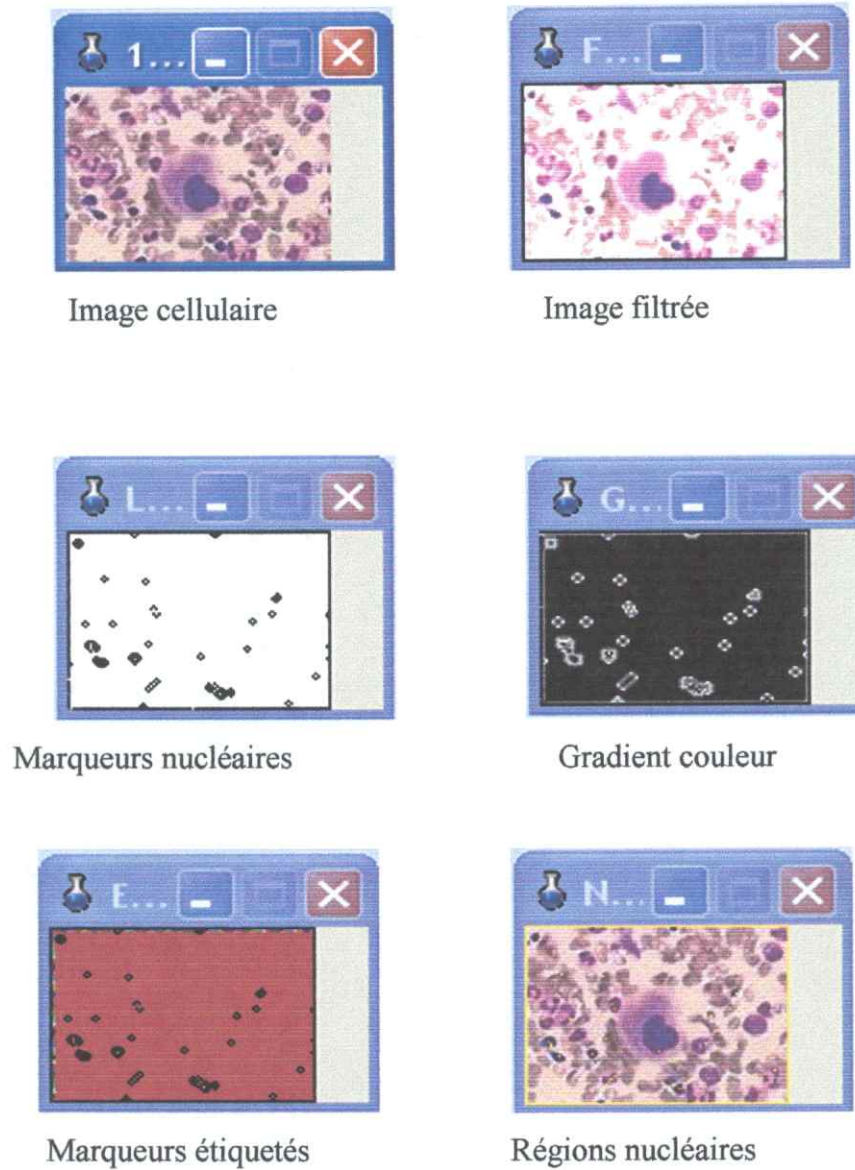


Figure IV.21 Segmentation des noyaux

- **Quatrième étape:** cette phase illustre la caractérisation. Celle-ci est déterminée par les critères morphologiques (le rapport nucléo-cytoplasmique, et l'hyperchromatisme nucléaire), que nous allons utiliser pour la classification (cf. Figure IV.22).
Le calcul du rapport nucléo-cytoplasmique nous donne les images des noyaux et des cytoplasmes pour quantifier la valeur du rapport des surfaces.

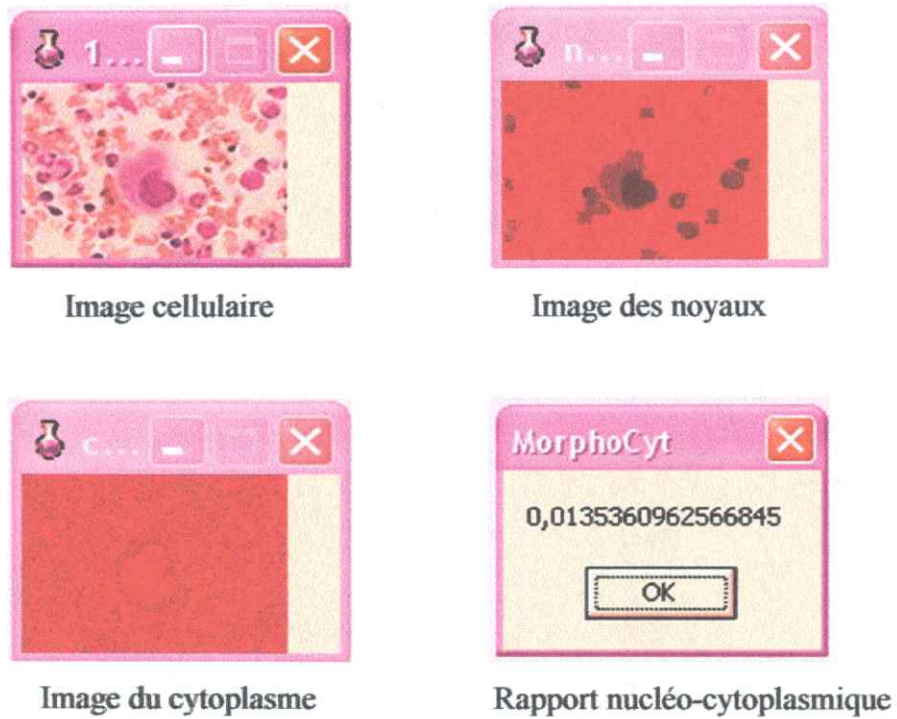


Figure IV.22 Calcul du rapport nucléo-cytoplasmique

Le calcul d'hyperchromatisme nous donne l'image de l'accroissement anormal de l'activité nucléaire et la valeur de la texture apparente au sein des noyaux (cf. Figure IV.23).



Figure IV.23 Hyperchromatisme nucléaire

Ces valeurs nous permettent de classifier nos cellules suivant le modèle de Kohonen.

- **Cinquième étape:** c'est la dernière phase de notre méthodologie, l'étape de la classification. Nous avons utilisé les paramètres précédents pour l'apprentissage, en indiquant le nombre de séquences de 50 et le nombre de neurones (3 classes). Nous obtenons trois centroïdes relatifs aux trois régions à identifier. L'application *morphocyt* effectue une comparaison entre les critères morphologiques (valeurs de la base d'apprentissage) et les centroïdes en indiquant ainsi le neurone actif, c'est à dire la classe de la cellule présentée.

Dans ce cadre, nous utilisons 12 images dans notre apprentissage, et nous gardons une à l'écart, que nous utiliserons comme test lors de la phase de reconnaissance. Nous avons donc construit une séquence de vecteurs d'apprentissage en appliquant sur ces 11 images la méthode d'extraction de paramètres décrite ci-dessus (cf. Figure IV.24).

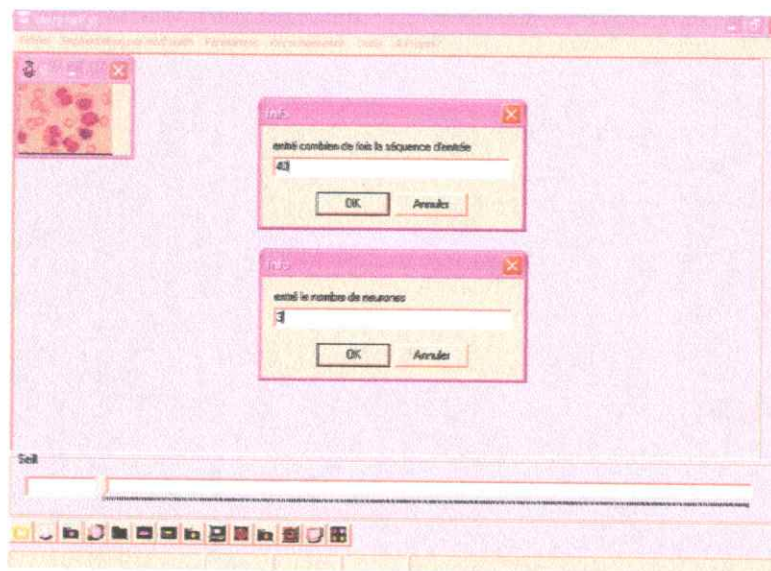


Figure IV.24 Apprentissage connexionniste des cellules

Nous avons donc obtenu une suite de 11 vecteurs de dimension deux. Le tableau III.1 ci-dessous indique, les valeurs des deux paramètres.

Image N°	Rapport Nucléo cytoplasmique	Hyperchromatisme nucléaire	Classe de l'image
1	0.82	2.74	1
2	0.22	28	3
3	8.16	24	3
4	0.85	3.91	1
5	0.85	4.89	2
6	2.11	25	3
7	1.53	29	3
8	0.19	37	3
9	1.54	2.47	1
10	1.81	1.80	1
11	0.85	4.89	2

Le tableau III.1 paramètres morphologiques d'images testée

Disposant de cette séquence, nous pouvons donc procéder à l'apprentissage de notre réseau de Kohonen. Pour notre exemple, nous avons choisi les paramètres suivants:

- fonction d'adaptation η proportionnelle au temps,
- Un voisinage initial de 3, rétrécissant avec le temps,
- La séquence d'apprentissage a été présentée 50 fois.

Nous obtenons donc trois centroïdes, correspondant aux trois régions que nous voulons identifier trois centroïdes correspondant aux trois régions envisagées. Nous pouvons donc procéder à la classification d'une image inconnue (cf. Figure IV.25), afin de décider si elle représente une cellule cancéreuse ou non, selon les critères retenus.

La carte de Kohonen qui effectue la classification, est une carte ayant deux entrées et une la sortie contenant trois neurones (le nombre de classes).

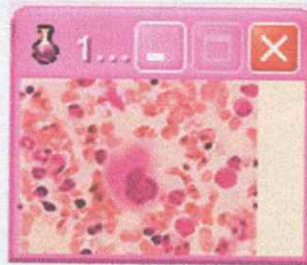


Figure IV.25 Cellule à classer

Après avoir procédé à l'apprentissage de notre réseau de Kohonen sur nos 11 images, la reconnaissance effectuée sur la dernière image (l'image 12), indique les valeurs des critères de malignité (cf. Figure IV.26) obtenues suivantes:

- Rapport nucléo-cytoplasmique : 0,014,
- Hyperchromatisme : 55.

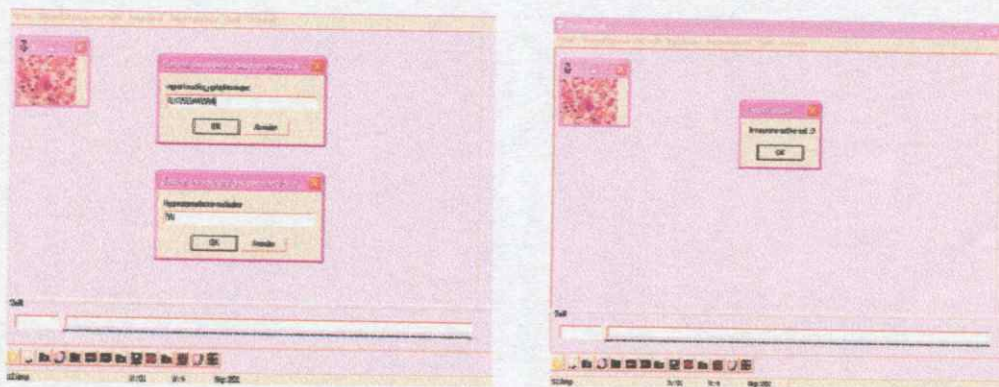


Figure IV.26 Classification des cellules

Cette cellule présente donc un hyperchromatisme supérieur à celui de la moyenne des cellules cancéreuses, sera donc classée parmi les *cellules cancéreuses*.

Le projet dans lequel, nous nous sommes investi, s'intègre dans le cadre d'un travail de recherche appréhendé aux laboratoires "LRDSI" (laboratoire de recherche pour le développement de systèmes informatiques) et "LATSI" (laboratoire de traitement de signal et de l'image); dont l'un des intérêts, est la détection de cellules cancéreuses à partir d'images cytologiques couleurs pour l'aide au diagnostic.

Une des applications principales de la cytologie au sein d'un laboratoire d'hématologie est d'aider le biologiste à caractériser au mieux les cellules qui se multiplient de manière anarchique dans les leucémies aiguës. Nous avons donc pu mettre en œuvre une méthode d'analyse d'images basée sur une segmentation par la ligne de partage des eaux, où nous nous sommes plus particulièrement intéressés au choix de l'espace couleur et à l'obtention de marqueurs. Notre méthode de segmentation permet ainsi, l'extraction de l'image des cytoplasmes et celle des noyaux dans le but d'une caractérisation des cellules pathologiques. L'intérêt de cette méthode, est qu'elle n'extrait que l'information utile; dans notre cas, il s'agit des cellules cancéreuses. Les globules rouges se trouvent ainsi, éliminés, puisqu'ils ne présentent aucune information nécessaire pour la classification des cellules cancéreuses.

Le processus développé peut être appliqué à divers cas de cellules cancéreuses. Le traitement d'images différentes, ne nécessitera qu'un simple ajustage des paramètres du système (taille et forme de l'élément structurant).

Quant, à la construction du classificateur, elle peut être totalement adaptée à l'utilisation que l'on veut en faire. Par exemple, un médecin intéressé uniquement par des coupes provenant d'un organe donné, utilisera des images d'apprentissage caractéristiques des cellules qu'il désire traiter. Le classificateur obtenu pourra donc être très pointu et spécialisé: il suffira de lui présenter des images spécifiques en grand nombre lors de la phase d'apprentissage. Si, de plus, les images sont très homogènes (conditions d'acquisition identiques, ...), les performances du classificateur seront encore accrues. Ajoutons que, dans ce cas, on pourra prévoir un nombre plus important de noeuds de sortie du réseau de Kohonen, et donc obtenir une gradation plus fine dans le diagnostic proposé. On pourrait donc imaginer un praticien

Conclusion générale et perspectives

possédant plusieurs classificateurs, chacun étant adapté à un organe ou à un tissu particulier. Par contre, si l'on se choisit des images d'apprentissage très générales, provenant d'un ensemble d'organes différents, mais répondant bien sûr aux mêmes critères de malignité, on a alors la possibilité de construire un classificateur de type plus "généraliste".

L'élaboration de notre système est donc réussie, cependant certains points restent à approfondir. La base doit être complétée par davantage de cellules pour améliorer l'apprentissage et équilibrer les proportions des différentes classes de cellules existantes. Signalons enfin qu'en ce qui concerne l'apprentissage, chaque nouvelle image analysée lors de la phase de classification, pourra venir enrichir la séquence d'entraînement et permettra d'améliorer de façon continue la qualité de la classification.

Le traitement d'images, n'a pas pour objet de remplacer le cytopathologiste, pour le diagnostic. Toutefois, la cytogénétique et la biologie moléculaire, permettent d'obtenir des informations complémentaires permettant de mieux définir les entités tumorales, de prédire la réponse à une thérapie particulière ou encore de suivre l'évolution de la maladie. La méthode étudiée, peut être élargie à tout problème d'analyse d'images, où la forme et la taille des objets présents, fournissent l'information nécessaire à la prise de décision. Elle constitue donc la porte ouverte à de nombreuses applications, dans tous les domaines où l'analyse d'images joue un rôle important.

Le traitement d'images n'a pas pour objet de remplacer les techniques traditionnelles utilisées pour le diagnostic des pathologies. La cytologie et l'histologie restent les techniques de référence pour la classification de ces entités. Toutefois, la cytogénétique et la biologie moléculaire permettent d'obtenir des informations complémentaires permettant de mieux définir ces entités et de fournir des éléments importants pour définir l'agressivité d'un processus néoplasique, de prédire la réponse à une thérapie particulière ou encore pour suivre l'évolution de la maladie au cours du traitement. La méthodologie étudiée constitue une porte ouverte aux systèmes futurs pour l'aide à la décision.

Bibliographie

- [BEU 90] Serge Beucher, « *Segmentation d'image et morphologie mathématique* », thèse de doctorat, école Nationale Supérieure des Mines de Paris, 1990.
- [BOU 04] : S. Bouzid, S. Fekhari « *Détection de contour et classification de textures par la carte de Kohonen* », mémoire d'ingénieur, Université de Blida 2004.
- [COS 89] Machel Coster & J.L. Cherman, « *Traitement d'image par morphologie mathématique application aux domaines médical et industriel* », thèse de doctorat, université Blaise Pascal, 1987.
- [DRE 02] : G. Dreyfus, J. Martinez, M. Samuelides, M. Gordon, F. Bardan, « *Réseaux neurones : Méthodologie et applications* », édition Eyrolles, 2002
- [FRU 02] Jean Fruit, « *Outils et méthodes pour le traitement des images par ordinateur* », Université de Marne la Vallée, France, 2002.
- [HER 99] : J. Héroult, C. Jutten, « *Réseaux neurones et traitement de signal* », édition Hermes, 1999.
- [LEZ 02] Olivier Lezoray, « *segmentation d'images couleurs par morphologie mathématique et classification des données par réseaux de neurones : application à la classification de cellules en cytologie des séreuses* », Thèse de doctorat, université de Caen France, 2002.
- [LEZ 95] Olivier Lezoray, « *La réalisation d'une étude sur le tri informatique en cytologie* », Rapport de stage DEA, université de Caen France, 1995.
- [SER 82] J. Serra, « *Image analysis and mathematical morphology* », academic Press, 1982.
- [TEM 01] W. Temam, A. Slimani, « *Reconnaissance des caractères numériques manuscrits Par la transformée de fourier-mellin et la carte auto organisatrice* », mémoire d'ingénieur, université de Blida, 2001.
- [TOU 90] Jean Jacques Tournazet, « *traitement de l'image par l'exemple* », éditions Sybex, juillet 1990.