

## A.R.C.T.I.C : Un système automatique de Tri Cellulaire par Analyse d'Images

Olivier Lezoray<sup>\*,†</sup>, Abder Elmoataz<sup>†,‡</sup>, Hubert Cardot<sup>†</sup> and Marinette Revenu<sup>‡</sup>

<sup>†</sup> Laboratoire Universitaire des Sciences Appliquées de Cherbourg,  
E.I.C, Site Universitaire, B.P. 78, 50130 Octeville, France;

<sup>\*</sup> Service d'Anatomie et de Cytologie Pathologiques, CHLP,  
46 rue du val de saire, 50100 Cherbourg, France ;

<sup>‡</sup> Groupement de Recherche en Informatique, Image et Instrumentation de Caen, I.S.M.R.A.,  
Bd du Maréchal Juin, 14050 Caen Cedex, France.

Olivier.Lezoray@info.unicaen.fr,

**Résumé**— Nous présentons un système d'aide au diagnostic par la tri cellulaire. Nous présentons la structure du système, ses implications, les méthodes employées et les résultats obtenus. La segmentation d'images couleur est réalisée par des méthodes de morphologie mathématique à travers une extension couleur de la ligne de partage des eaux. La caractérisation des objets utilise la couleur pour l'extraction de critères de texture. La classification des objets est réalisée par une architecture de réseaux de neurones particulière. Les résultats sont probants puisque 96% des cellules anormales et 95% des cellules normales sont reconnues.

**Abstract**— A cellular sorting system is presented through its structure, the methods involved and the results. Segmentation is realized using a color extension of the watershed mathematical morphology process. Feature extraction uses color texture attributes. Classification is achieved with a self-organizing and adaptive architecture of neural networks. Results are accurate since 96% of abnormal cells and 95% of normal cells are recognized.

### 1. Le problème

En Anatomie et Cytologie Pathologiques, on distingue deux types d'examen. L'histologie qui est l'observation de la coupe d'un tissu et la cytologie qui est l'examen d'un étalement de cellules. Nous allons nous intéresser plus particulièrement à l'examen cytologique à travers notre travail de recherche. Les prélèvements sont étalés sur une lame puis fixés et colorés afin de reconnaître les différentes cellules présentes. Les étalements sont ensuite examinés au microscope par un cytotechnicien afin de repérer les cellules intéressantes. Cette étape de lecture (le

*screening*) consiste en un repérage et une évaluation visuelle de toutes les cellules présentes sur une lame. Son but est principalement la détection des cellules anormales ou suspectes afin d'établir un diagnostic. Ceci est donc d'un intérêt capital pour le pathologiste dont le diagnostic dépendra de la bonne reconnaissance des cellules anormales ou suspectes lors du *screening*. Ce *screening* manuel des prélèvements cytologiques est 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 » [7]. Le faible nombre de cellules anormales (comparativement au nombre de cellules normales) observées durant le *screening* implique une grande concentration du cytotechnicien. A cause de ce facteur subjectif, quelques erreurs apparaissent et peuvent causer des faux négatifs. Ces erreurs sont considérées comme étant indissociables du procédé de *screening* manuel. Une approche prometteuse est d'aider le cytopathologiste dans sa recherche des cellules anormales sur une lame. Un système semi-automatique informatique pourrait contribuer à la détection des erreurs de *screening* permettant ainsi une plus grande fiabilité du diagnostic. De tels systèmes sont prévus pour opérer de façon postérieure au *screening* conventionnel mais peuvent opérer éventuellement de façon antérieure. Les lames données négatives par un cytotechnicien sont revues par le système afin de vérifier la non présence de cellules anormales sur la lame. De tels systèmes existent et sont opérationnels. Nous nous proposons de réaliser un tel système d'aide au *screening* : A.R.C.T.I.C (Aide à la Recherche en Cytologie par le Tri Informatique Cellulaire). Nous travaillons sur des liquides d'épanchement des séreuses (plèvre, péritoine,...) contrairement aux autres systèmes qui n'opèrent que sur

des prélèvements gynécologiques. Notre système utilise la segmentation et l'analyse d'images couleur ainsi que la classification par réseaux de neurones et par arbre de décision. Nous allons voir dans cet article sa structure et les trois étapes qui le constituent : Segmentation, Analyse et Classification. Les résultats obtenus seront présentés et discutés.

## 2. Place et structure du système

ARCTIC est un système permettant une forme d'aide au diagnostic. Celui-ci procède à la segmentation des cellules ainsi qu'à leur classification. Ceci afin de vérifier qu'aucune cellule anormale ou suspecte n'est présente sur la lame. Si c'est le cas, les cellules anormales détectées sont affichées à l'écran pour une relecture par un pathologiste afin d'établir un diagnostic. Le système aidera donc le pathologiste en vue d'un diagnostic en recherchant les éléments rares ou suspects qui auraient pu échapper à la vigilance du cyto-technicien lors de son *screening*. Il permettra de réduire le nombre de faux-négatifs par la mise en place d'un contrôle qualité. L'architecture d'un système d'aide au *screening* est commune à tous les systèmes existants [6,13,14,15]. Dans une première étape on procède à l'acquisition d'une image couleur, celle-ci est ensuite segmentée (extraction des objets se trouvant sur l'image à savoir les cellules). Les objets extraits sont ensuite caractérisés (extraction de données représentatives pour chaque cellule) et classés dans une ultime étape. Les cellules identifiées comme anormales sont alors présentées à l'expert qui peut établir son diagnostic.

Des travaux utilisant une telle architecture ont déjà été conçus par le GREYC [12] avec lequel nous collaborons. Nous avons utilisé ces acquis pour mettre en place notre système.

## 3. Matériels et méthodes

Les lames sont examinées par un microscope sur lequel est fixé une caméra couleur. L'image est numérisée via une carte d'acquisition. Les images obtenues sont des images couleur 24 bits de 512x512 pixels. Notre logiciel a été développé en Visual C++ en utilisant la librairie de traitement d'images PANDORE [3] à laquelle nous avons rajouté la gestion et le traitement des images couleur.

Par la suite les images que nous utilisons sont des images multispectrales. La caméra est calibrée par une image représentant un champ microscopique vide afin d'assurer une dynamique correcte des niveaux de gris de chacune des composantes couleur RVB. La procédure d'acquisition des images est figée : grossissement x20, diaphragme ouvert à 0.3, la source lumineuse est stabilisée à une tension constante de 9V aux bornes du microscope.

Cette dernière recommandation est indispensable pour assurer une illumination constante et linéaire du champ microscopique. La tension a été fixée à 9V en accord avec les spécifications des constructeurs mais également pour disposer d'une température de couleur de 5500K (température qui correspond aux spécifications des illuminants standard spécifiés par la CIE [18]). Le bruit présent sur les images correspond à une variable gaussienne aléatoire. Ce bruit sera corrigé ultérieurement par la segmentation d'images grâce à un lissage. Cependant outre le bruit thermique, quelques non uniformités peuvent apparaître sur chaque image. Pour corriger ceci, après chaque acquisition une division entre l'image acquise et une image de fond est réalisée. L'image obtenue permet de corriger des variations de fond qui sont dépendantes de la lame analysée.

## 4. Description des images

Les images que nous avons à traiter sont donc des images couleur. Sur ces images sont présentes des cellules qui doivent être extraites. Il nous faut isoler à la fois leur cytoplasme et leur noyau. Le cytoplasme afin d'obtenir une information de contexte (pour caractériser les cellules isolées ou en amas) et le noyau pour caractériser la cellule et évaluer sa malignité.

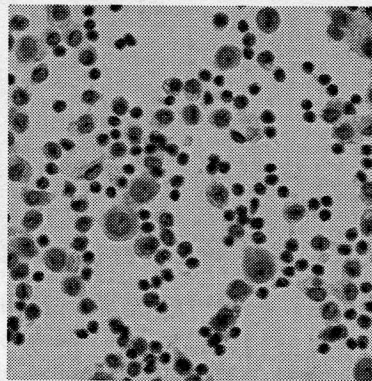


Figure 1. Fond homogène

Les deux informations doivent amener à une reconnaissance des différents types cellulaires. Avant d'entamer la première étape de notre système à savoir la segmentation, nous devons connaître précisément la nature et le contexte des images. Nos images sont des images couleur présentant des cellules provenant de la cytologie des séreuses et colorées par le standard international de coloration de Papanicolaou [10]. Ces images peuvent être réparties en trois groupes selon leurs fonds : homogène (figure 1), hémorragique (figure 2) ou muco-protéique (figure 3). Les cellules ont un noyau bleu et un cytoplasme vert à l'exception des globules rouges qui sont totalement colorés en rouge. La configuration spatiale

des cellules et leur couleur sont extrêmement variables. On trouve des cellules isolées, accolées mais également des amas dont les cellules peuvent se chevaucher (noyau ou cytoplasme). La couleur du noyau peut varier de bleu très pâle à bleu très foncé. Cette grande variété dans la configuration spatiale et la couleur des cellules pose des problèmes de segmentation et requiert une méthode de segmentation fine et robuste à la fois.

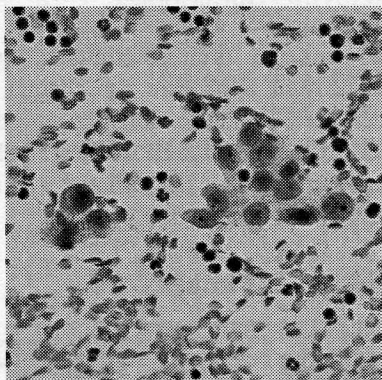


Figure 3. Fond hémorragique

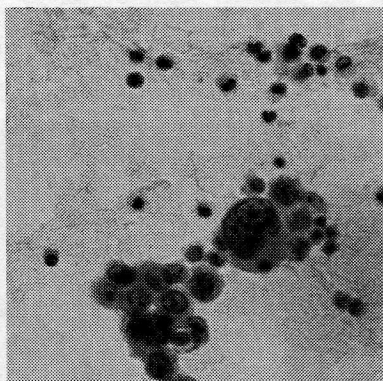


Figure 4. Fond muco-protéique

## 5. Stratégie de segmentation couleur

La stratégie de segmentation que nous avons mise en place est une stratégie ascendante. Dans un premier temps il nous faut déterminer si des cellules nucléées (caractérisées par un noyau bleu) sont présentes sur l'image à segmenter, ceci afin de ne traiter que les images où celles-ci sont présentes. Si des cellules sont présentes sur l'image nous procédons à l'élimination des globules rouges car ils ne présentent aucun intérêt diagnostique et pourraient être gênants pour extraire les autres cellules intéressantes. Nous extrayons ensuite les cytoplasmes et les noyaux des cellules de façon séparée. Les noyaux accolés sont séparés et les cytoplasmes sont divisés de façon à ne retrouver qu'un seul noyau ou amas de noyaux par cytoplasme. La stratégie complète est résumée sur la

figure 4 par un graphe de niveau opérationnel selon la méthodologie précisée dans ATHENA[4].

La stratégie de segmentation d'images couleur a déjà été publiée, on se référera à [8] pour plus de détails. Nous allons néanmoins rappeler dans ses grands axes en quoi notre façon d'aborder la segmentation des images couleur est à la fois originale et efficace.

### Principe de la Segmentation d'images couleur

Nous avons choisi de nous placer dans le contexte de la morphologie mathématique et d'utiliser une méthode de segmentation couleur dérivée de celle proposée par [1]. Aux vues d'une image couleur, quelles informations nous permettent d'interpréter la sémantique contenue dans cette image ? En fait ceci peut se résumer à deux informations : la couleur des objets présents sur l'image et les transitions de couleur entre les objets. Une ligne de partage des eaux [2,16] utilisant ces deux informations (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

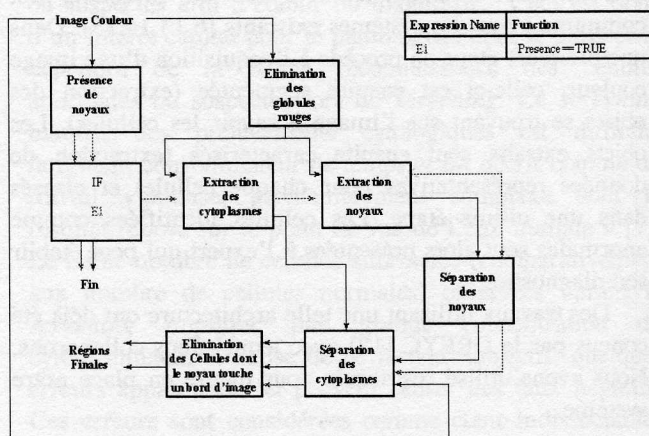


Figure 2. La stratégie de segmentation d'images couleur de cytologie des séreuses.

présents dans 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. L'utilisation de la couleur de tenir compte des changements plus significatifs dans les images 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. Ceci se traduit par la fonction distance suivante. Soient  $I_{C_1 C_2 C_3}(R)$  le vecteur donnant la couleur moyenne de la

région  $R$  pour l'image  $I$  dans l'espace couleur  $C_1C_2C_3$ ,  $I_{C_1C_2C_3}(p)$  le vecteur donnant la couleur du point  $p$  dans l'image  $I$  et  $\nabla I_{C_1C_2C_3}(p)$  le gradient couleur au point  $p$ .

$$\phi(p, R) = (1 - \alpha) \left\| \overline{I_{C_1C_2C_3}(R)} - I_{C_1C_2C_3}(p) \right\| + \alpha \left\| \nabla I_{C_1C_2C_3}(p) \right\| \quad (1)$$

Cette distance combine l'information locale (le module du gradient couleur) et l'information globale (résultant d'une comparaison statistique entre la couleur d'un point  $p$  et une région voisine  $R$ ).  $\alpha$  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  $\alpha$  égal à 1 on retrouve l'algorithme classique de la ligne de partage des eaux mais dont le

configuration spatiale. Aucun globule rouge n'a été segmenté : tous les objets extraits par notre stratégie sont des cellules et quelques débris. Grâce au procédé d'inspection visuelle toutes les cellules mal segmentées sont identifiées. Le taux de réussite de la segmentation est présenté dans le tableau 1 pour chaque expert. On note une variation sensible de 10% entre les experts, mais ceci est considéré comme normal étant donné la difficulté d'interprétation pour des erreurs minimales et pour les cas extrêmes. Ceci prouve que bien qu'une inspection visuelle de la segmentation par un expert permet d'obtenir une mesure objective du taux de réussite de la segmentation, la difficulté d'interprétation de certaines configurations cellulaires rajoute un caractère subjectif à ce type d'évaluation (ce qui la rapproche des limitations du *screening* manuel).

Les résultats de la segmentation sont très satisfaisants.

Expert N°	Noyaux corrects	Cytoplasmes corrects	Erreurs sur les noyaux	Erreurs sur les cytoplasmes
1	98.3%	99.0%	1.7%	1.0%
2	96.1%	91.5%	3.9%	8.5%
3	89.2%	88.7%	10.8%	11.3%
Moyenne	94.5%	93.0%	5.5%	6.9%

Tableau 1. Taux de réussite de la segmentation pour les noyaux et les cytoplasmes.

gradient est calculé sur une image couleur. Avec  $\alpha$  égal à 0 on retrouve l'algorithme de [9] qui n'utilise que la distance couleur entre un pixel et une région. Notre stratégie de segmentation couleur utilise cette méthode pour extraire les noyaux et les cytoplasmes. L'originalité de la méthode réside dans l'extraction des marqueurs de la ligne de partage des eaux qui utilise plusieurs composantes couleurs de différents espaces couleurs mais également dans le choix automatique de l'espace couleur  $C_1C_2C_3$  par une méthode d'extraction du contraste.

Nous avons également défini une méthodologie d'utilisation de notre procédé de segmentation couleur. Cette méthodologie permet de formaliser les différentes étapes de la mise au point des différents paramètres pris en compte lors de la segmentation couleur (choix de l'espace couleur, choix d'une métrique pour comparer  $\overline{I_{C_1C_2C_3}(R)}$  et  $I_{C_1C_2C_3}(p)$ , choix d'une méthode de calcul du gradient, choix des marqueurs, choix de  $\alpha$ ). Ceci permet d'appliquer la méthode facilement à d'autres types d'images couleur.

#### Résultats de la segmentation

Les résultats de la segmentation ont été inspectés visuellement par trois experts sur 50 images soit environ 2000 cellules. Toutes les cellules possédant un noyau ont été correctement segmentées quelle que soit leur

94.5% des noyaux et 93% des cytoplasmes sont correctement segmentés. Le taux de réussite de la segmentation dépend en grande partie de la population cellulaire présente sur une image. Si beaucoup de cellules isolées sont présentes, notre stratégie sera très efficace. Mais pour des configurations plus complexes (comme la superposition de noyaux), le taux de réussite n'est pas représentatif car l'expert est également incapable de segmenter manuellement les cellules de façon précise. De ce point de vue, notre stratégie de segmentation est différente de celles employées par les autres systèmes : nous avons choisi de segmenter non seulement les cellules isolées (figure 5) mais également les amas les plus complexes (figure 6) et les résultats sont très satisfaisants.

## 6. Analyse d'images couleur

Une fois les cellules segmentées, il reste à les caractériser par des paramètres suffisamment discriminants pour permettre une bonne classification des cellules. Le choix des paramètres est très important, celui-ci est guidé par les informations a priori dont on dispose sur les cellules. Dans notre cas nous avons répertorié les différents types d'objets qui peuvent être rencontrés en cytologie des séreuses. On distingue les cellules isolées (composées de 18 classes), les amas (constitués de

superpositions de cellules isolées) et les débris. Ceci représente un nombre de classes de cellules à reconnaître assez important. Les critères qui permettent aux experts de les différencier sont des critères de taille, de couleur, de forme et de texture. Ce sont donc ceux-ci que nous avons mis en place. Nous utilisons actuellement 46 paramètres qui sont mesurés sur chacune des cellules qui sont extraites de la segmentation. Les paramètres de taille concernent noyau et cytoplasme. Les paramètres de forme concernent essentiellement le noyau.

La texture n'est mesurée que sur le noyau car c'est celui-ci qui permet d'évaluer la malignité d'une cellule. Quatre méthodes sont utilisées : une méthode statistique (moyenne, variance et écart-type pour chacune des composantes couleur ainsi qu'une mesure globale sur la couleur), une méthode d'extraction de textons [17] (division du noyau en trois zones, montagne désigne les zones claires, vallée désigne les zones foncées et inter désigne les zones intermédiaires. On mesure pour chacune de ces zones la moyenne et l'écart-type de la taille, de la couleur et de la position des textons dans le noyau), une méthode basée sur le chapeau haut de forme [2] (division du noyau en zones de forte intensité chromatinienne et mesure de leur nombre, taille, couleur et position) et enfin une méthode basée sur la division du noyau par une ligne de partage des eaux [5] où des mesures identiques sont

autre innovation de notre système. Nous avons donc utilisé un nombre de paramètres de texture assez élevé de façon à disposer de différentes représentations de la texture (les paramètres de texture sont de plus en plus fins entre les méthodes, la précision va croissante entre les méthodes Top-Hat, Watershed et Textons). Ceci permet d'avoir un point de vue multi-échelle de la texture et entraîne une meilleure classification des cellules).

## 7. Classification des cellules

Pour classer les cellules nous avons mis en place une architecture basée sur les réseaux de neurones et les arbres de décision. Pour classer un objet extrait à la segmentation il nous faut d'abord déterminer s'il s'agit d'un débris, d'un amas de cellules ou bien d'une cellule. Ceci est réalisé par un arbre de décision généré par C4.5 [11]. Cet arbre de décision a un taux d'erreurs de 6.3%. L'arbre de décision est représenté par une série de 30 règles qui permettent de classer un objet en tant que débris, cellule ou amas. Les principales erreurs proviennent de cellules classées comme débris, ou de cellules classées comme amas. Afin de remédier à ce problème nous avons rajouté une règle sur la taille des cellules afin de n'avoir que des petites objets dans la classe débris et des très grands objets dans la classe amas. Ceci permet de réduire le taux d'erreurs à 1.9%. Les erreurs restantes sont dues à de très

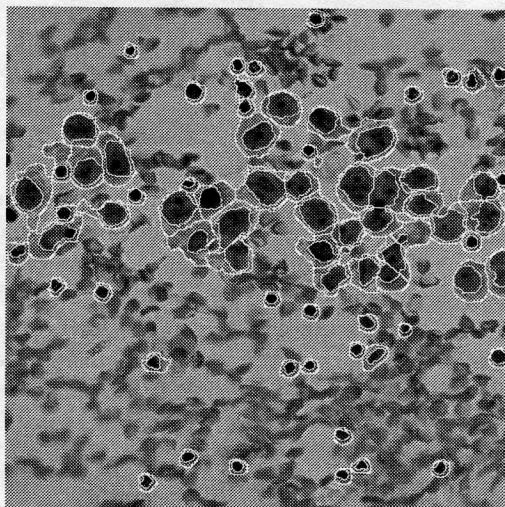


Figure 5. Cellules isolées et accolées après segmentation.

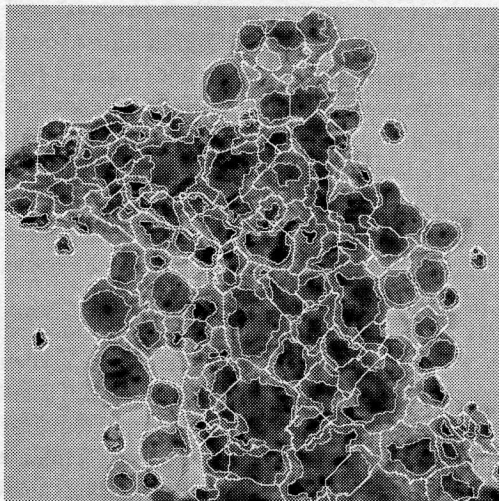


Figure 6. Cellules accolées ou en amas après segmentation.

effectuées.

On remarque que beaucoup de critères de texture ont été utilisés car la texture est l'élément le plus important pour différencier les cellules entre elles. Mais la couleur est un paramètre également très important, c'est pourquoi tous nos paramètres de texture ont été calculés vis à vis de la couleur et non pas en niveaux de gris. Ceci présente une

petites cellules qui sont classées comme débris.

A l'issue de cette première classification on ne s'intéresse alors qu'aux cellules (qui peuvent être des cellules anormales ou suspectes). Elles doivent être réparties dans les différents classes exposées dans le tableau 2. Pour cela, nous avons mis en place un système

de réseaux de neurones multi-couches à rétropropagation du gradient de l'erreur. Les réseaux que nous utilisons ne se présentent pas de la même façon que les réseaux de neurones habituels. Typiquement un réseau de neurones est constitué d'une couche d'entrée (les paramètres caractérisant un objet), d'une ou de deux couches cachées et d'une couche de sortie fournissant la classe de l'objet. Pour des problèmes très complexes où les objets sont difficiles à différencier, ce type de réseau peut présenter certaines limites, c'est pourquoi nous avons choisi une architecture de réseaux de neurones sensiblement différente.

#### Architecture du système de réseaux de neurones

Le système de classification basé sur les réseaux de neurones se compose d'un ensemble de réseaux, chacun étant destiné à séparer les éléments de deux classes distinctes. Si l'on dispose de n classes cela nous conduit à

avoir  $\frac{n*(n-1)}{2}$  réseaux de neurones servant à la

classification. La difficulté à séparer n classes est simplifiée par la spécialisation de chaque réseau car un réseau ne s'intéresse qu'à la séparation de deux classes. Chacun des réseaux procède à son propre processus d'apprentissage et possède la structure usuelle : une couche d'entrée, une couche cachée et une couche de sortie constituée d'un unique neurone décidant de la classe de l'objet dont la probabilité d'appartenance est la plus forte parmi les deux classes que le réseau sépare. Les entrées de chacun des réseaux sont constituées des paramètres extraits lors de l'analyse des objets et sont normalisés entre 0 et 1. La phase d'apprentissage s'opère sur 80% de l'échantillon d'apprentissage et les 20% restants servent de base de test. L'apprentissage que nous utilisons est un apprentissage automatique qui détermine l'architecture optimale du réseau à savoir son nombre de neurones de la couche cachée. Cela consiste à effectuer une suite d'apprentissages sur chacun des réseaux du système en augmentant à chaque itération le nombre de neurones de la couche cachée. Le processus continue jusqu'à ce que la qualité obtenue atteigne un certain seuil de qualité. Ainsi chaque réseau détermine sa structure lors de l'apprentissage et sa qualité de classification est la meilleure possible.

Pour classer un objet, chaque réseau du système est caractérisé par les deux classes qu'il sépare. La valeur du neurone de sortie appartient à  $[-1,1]$ , ce qui permet, suivant le signe du résultat, de classer l'objet dans l'une ou l'autre classe. De cette manière la classe d'appartenance de l'objet est déterminée par élimination grâce aux résultats obtenus par les différents réseaux. En effet, si un réseau séparant deux classes  $C_0$  et  $C_1$  désigne un objet comme appartenant à la classe  $C_0$ , la classe  $C_1$  est éliminée. Au fur et à mesure de l'examen des résultats de

différents réseaux, on arrive finalement à obtenir la classe d'appartenance de cet objet. L'ordonnement des réseaux d'après leur qualité de classification conduit au processus d'élimination de classes le plus juste possible et permet d'obtenir la classe d'appartenance la plus probable (figure 7). Pour classer les cellules nous disposons donc d'un système à base de réseaux de neurones qui prennent la décision liée à la classification.

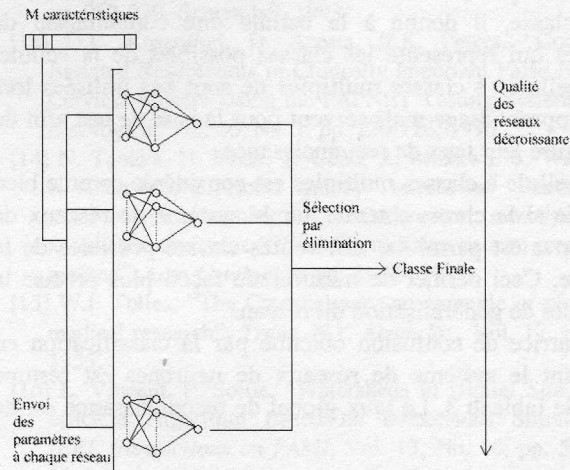


Figure 7. Schéma d'organisation du système à base de réseaux de neurones.

Número de la classe	Nom de la Classe	Nombre de cellules
0	polynucléaires neutrophiles	300
1	polynucléaires éosinophiles	298
2	polynucléaires altérés	375
3	lymphocytes non activés	514
4	lymphocytes activés	336
5	macrophages	635
6	sidérophages	0
7	mésothéliales	64
8	mésothéliales dystrophiques	22
9	mésothéliome	86
10	adénocarcinome	320
11	carcinome épidermoïde	0
12	lymphome	0
13	sarcome	0
14	mélanome malin	0
15	mitoses anormales	28
16	mitoses normales	1
17	lobe de polynucléaire	266
Combinaison	Multiple	74

Tableau 2. La distribution des cellules dans la base d'apprentissage.

## 8. Résultats et discussion

Nous présentons les résultats que nous avons obtenus avec le système à base de réseaux de neurones sur une base de 3630 cellules. Le tableau 2 donne la distribution des cellules dans les différentes classes. Certaines cellules peuvent correspondre à une combinaison de classes : ce sont des cellules à classes multiples. Lorsqu'un expert ne peut de façon certaine affirmer qu'une cellule appartient à une classe, il donne à la cellule une combinaison de classes qui représente les classes possibles de la cellule. Ces cellules à classes multiples ne sont pas utilisées lors de l'apprentissage mais servent pour la base de test afin de connaître leur taux de reconnaissance.

Une cellule à classes multiples est considérée comme bien classée si la classe obtenue par le système de réseaux de neurones est parmi les différentes classes possibles de la cellule. Ceci permet de mesurer de façon plus précise la capacité de généralisation du réseau.

La matrice de confusion obtenue par la classification en utilisant le système de réseaux de neurones est résumé dans le tableau 3. Le taux global de reconnaissance est de

Si l'on s'attarde sur la matrice de confusion du tableau 3 (les classes ne contenant pas d'objets n'y figurent pas), on remarque que les erreurs les plus flagrantes se situent au niveau de la différenciation entre polynucléaires neutrophiles et altérés, entre macrophages et cellules mésothéliales ainsi qu'entre macrophages et lymphocytes activés. Les mitoses ne sont pas reconnues en tant que telles même si elles sont classées parmi les cellules anormales. Nous travaillons actuellement à une extension de la base afin d'équilibrer et d'étoffer le nombre de cellules dans chaque classe. En effet certaines classes sont peu représentées et donc peu apprises par le système, ce qui ne permet de donner des résultats assez significatifs pour ces classes. D'autre part la classe macrophage se situe très à la limite entre les lymphocytes activés et les mésothéliales et souvent l'expert peine à prendre une décision de classification (lors de l'étiquetage manuel). Cette classe est donc une classe très hétérogène et sa reconnaissance est des plus difficile.

D'autre part nous avons vu précédemment que certains objets ont été classés en tant qu'amas, ceux-ci n'ont pas

Classes	Polynucléaires classes 0,1,2,17	Lymphocytes classes 3,4	Macrophages ou Mésothéliales classes 5,7	Anormales classes 8 à 15
Polynucléaires	224	9	11	5
Lymphocytes	30	123	15	3
Macrophages ou Mésothéliales	6	25	95	14
Anormales	0	0	5	89

Tableau 4. La matrice de confusion du Tableau 3 après regroupement.

60.24% pour les cellules à classes simples et de 71.62% pour les cellules à classes multiples (qui ne figurent pas dans la matrice de confusion).

	0	1	2	3	4	5	7	8	9	10	15	16	17
0	13	12	25	1	2	5	0	0	0	0	2	0	0
1	5	48	0	1	0	2	0	0	0	1	2	0	1
2	13	1	41	3	2	4	0	0	0	0	0	0	11
3	3	5	6	65	11	5	0	0	0	0	0	0	8
4	2	1	5	13	34	10	0	0	1	1	1	0	0
5	1	4	1	6	19	77	7	4	0	3	5	0	0
7	0	0	0	0	0	3	8	1	0	1	0	0	0
8	0	0	0	0	0	1	1	3	0	0	0	0	0
9	0	0	0	0	0	0	1	2	7	8	0	0	0
10	0	0	0	0	0	1	1	0	3	57	2	0	0
15	0	0	0	0	0	0	0	0	0	4	2	0	0
16	0	0	0	0	0	0	0	0	1	0	0	0	0
17	3	9	3	0	0	0	0	0	0	0	0	0	39

Tableau 3. La matrice de confusion du système de réseaux de neurones.

encore fait l'objet d'une classification. Nous mettons en place une classification pour ces derniers afin de les répartir selon qu'il sont composés de cellules normales ou anormales. Le problème majeur qui se pose lors de la reconnaissance des cellules provient de la présence de flou sur l'image. Certaines cellules peuvent avoir été acquises floues et leur reconnaissance peut poser certains problèmes car on ne dispose plus d'informations pertinentes sur la cellule. Même si une classification de ces cellules amène à des résultats équivalents à la classification de cellules non floues, ceci sera corrigé lors de la phase d'acquisition des images par une méthode de mise au point automatique du microscope.

Si l'on se place dans l'optique de notre système à savoir la détection de cellules anormales ou suspectes, notre système est très satisfaisant (tableau 4). En effet 96% des cellules anormales et 95% des cellules normales sont reconnues ce qui est largement supérieur au taux de réussite d'un expert. Etant donné que notre système opère en post *screening*, on peut penser, aux vues des résultats obtenus, qu'une cellule anormale omise par le

cytotechnicien sera détectée par notre système. Evidemment celui-ci reste perfectible et nous travaillons à son amélioration.

## 9. Références

- [1] P. Belhomme, A. Elmoataz, P. Herlin. and D. Bloyet, "generalized region growing operator with optimal scanning : application to segmentation of breast cancer images". *Journal of Microscopy*, Vol. 186, pp.41-50, 1997.
- [2] S. Beucher, *Segmentation d'images et morphologie mathématique*, Thèse de l'Ecole Nationale des Mines de Paris, 1990.
- [3] R. Clouard, A. Elmoataz, F. Angot, O. Lezoray, *Pandore : une bibliothèque et un environnement de programmation d'opérateurs de traitement d'images*, Rapport interne du GREYC, Caen, France, Mars 1997.
- [4] R. Clouard, A. Elmoataz, M. Revenu, "Une modélisation explicite et opérationnelle de la connaissance en traitement d'images", *Actes de RFIA'98*, Clermont-Ferrand, Vol 2, pp. 65-74, 1998.
- [5] P. Jackway, "Morphological Multiscale Gradient Watershed Image Analysis", *9<sup>th</sup> Scandinavian Conference on Image Analysis*, Uppsala, Sweden, pp. 87-93, June 1995.
- [6] E.A. Knesel Jr., "Roche Image Analysis System", *Acta Cytologica*, Vol. 40, No. 1, pp. ; 60-66, 1996.
- [7] L.G. Koss, "The Papanicolaou test for cervical cancer detection : a triumph and a tragedy", *J. Am. Med. Assoc.*, Vol. 261, pp. 737-743, 1989.
- [8] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, H. Elie and M. Revenu, "Color Images Segmentation : Application to serous effusions cytology for Computer Cellular Sorting", *Actes de RFIA'98*, Clermont-Ferrand, Vol. 2, pp. 287-295, 1998.
- [9] F. Meyer , "Color Image Segmentation", *Proceedings of the 4<sup>th</sup> Conf Image Processing and its Applications*, Maastricht, pp. 303-306, 1992.
- [10] G.N. Papanicolaou, "A new procedure for staining vaginal smears", *Science* , Vol. 95, pp. 432, 1942.
- [11] J.R. Quinlan, *C4.5 : programs for machine learning*, Morgann Kauffman, San Mateo, CA, 1993.
- [12] M. Revenu, A. Elmoataz, C. Porquet, H. Cardot, "An automated system for the classification of cellular categories in cytological images", *Intelligent Robot and Computer Vision XII, Algorithms and Techniques*, SPIE, Boston, USA, pp. 967-970, September 1993.
- [13] D.L. Rosenthal, D. Acosta, R.K. Peters, "Computed-Assisted Rescreening of Clinically Important False Negative Cervical Smears Using the PAPNET Testing System", *Acta Cytologica*, Vol. 40, No. 1, pp. 120-126, 1996.
- [14] N. Tanaka, H. Ikeda, T. Ueno, A. Mukawa, S. Watanabe, K. Okamoto, S. Hosoi and S. Tsunekawa, "Fundamental study of cyto-screening for uterine cancer : new system for automated apparatus (CYBEST) utilizing pattern recognition method", *Acta Cytologica*, pp. 21-85, 1977.
- [15] W.E Tolles, "The Cytoanalyser : an example of physics in medical research", *Trans. N.Y. Acad. Sc.*, Vol. 17, pp. 250, 1955.
- [16] L. Vincent, P. Soille, "Watersheds in Digital Spaces : An efficient Algorithm Based on Immersions Simulations", *IEEE transactions on PAMI*, Vol. 13, No. 16, pp. 583-598, 1991.
- [17] G. Wolf, M. Beil, H. Guski, "Chromatin Structure Analysis Based on a Hierarchic Texture Model", *Analytical and Quantitative Cytology and Histology*, Vol. 17, No.1, 1995.
- [18] G. Wyszeccki G., W.S Styles, *Color Science : Concepts and Methods, Quantitative Data and Formulae*, John Wiley & Sons, Second Edition, New-York, 1982.