



République Algérienne Démocratique et Populaire

Université Abou Bakr Belkaid– Tlemcen

Faculté des Sciences

Département d'Informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de Master en Informatique

Option: Mode Intelligent et Décision(M.I.D)

Thème

Fusion des classifieurs supervisés: Application sur la classification pixellaire des images microscopiques

Réalisé par :

- Guermoudi Mohammed el Amine.
- Fekih Mohammed el Amine.

Présenté le 01 Juillet 2013 devant la commission composé de MM.

- | | |
|-----------------------|----------------|
| - Benazzouz .M | (Encadreur) |
| - Baghli .I | (Co-encadreur) |
| - Hajila .F | (Président) |
| - Benmouna .Y | (Examineur) |
| - Chaouche Ramdane .L | (Examineur) |

Année universitaire : 2012 / 2013

Remerciements

Nous tenons à remercier :

Allah le tout puissant ;

Mr Mourtada BENAZZOUIZ, notre encadreur et mlle Baghli Ismahane, pour leurs conseils, disponibilités et son encouragement qui nous ont permis de réaliser ce travail dans les meilleures conditions. Les jurys pour leurs efforts et leur soin apporté à notre travail. Aux enseignants de notre université et département informatique.

Dédicaces

Je dédie ce modeste travail pour mes chers parents par les quelles j'aurais jamais atteintre la place où je suis sans leurs admirables rôles et si encouragent pour accomplir mes études, si ainsi je le dédie pour tous ma familles et mes amis.

A.Guermoudi

Je dédie ce mémoire :

A Ceux qui ont fait de moi l'homme que je suis aujourd'hui : mes très chers parents, que dieu les récompense et les garde, et surtout ma mère qui m'a éclairée mon chemin et qui m'a encouragé et soutenue toute au long de mes études. A Mon frère et mes sœurs. A Tous mes amis et Mes collègues de promotion avec lesquels, j'ai passé mes meilleures années d'études.

A.Fekih

Table des matières

Introduction générale	8
1 Représentation de la couleur et analyse discriminante	10
1.1 Introduction	10
1.2 Représentation de la couleur	10
1.2.1 Les systèmes de primaires	10
1.2.2 Les systèmes luminance-chrominance	11
1.2.3 Les systèmes perceptuels	13
1.2.4 Les systèmes d'axes indépendants	16
1.3 Synthèse	16
1.4 Analyse Linéaire Discriminante (LDA)	17
2 Apprentissage et classifications supervisées	19
2.1 Introduction	19
2.2 Notion d'apprentissage	19
2.2.1 Types d'apprentissage	20
2.3 Notion de la classification	20
2.4 Principales méthodes de la classification supervisée	20
2.4.1 Machine à vecteurs de support (SVM)	20
2.4.2 Réseaux de neurones	25
2.4.3 Modèle de mélanges de gaussiennes	27
2.4.4 Arbres de décision	28
2.4.5 Les K plus proches voisins	31
2.5 Conclusion	31
3 Fusion de décisions	32
3.1 Introduction	32
3.2 Principes et besoins de la combinaison des classifieurs	32

3.3	Définition d'un classifieur dans le cadre de la combinaison	33
3.4	Méthodes sans apprentissage	34
3.5	Méthodes avec apprentissage	35
3.5.1	Approche bayésienne	35
3.5.2	Théorie de l'évidence	36
3.6	Conclusion	39
4	Modélisation de la fusion et expérimentations	40
4.1	Description de la base	40
4.2	Construction de la base d'apprentissage	40
4.2.1	Espaces couleur	40
4.2.2	Construction de la matrice de passage	41
4.2.3	Étiquetage	41
4.2.4	Échantillonnage	42
4.2.5	Base d'apprentissage	43
4.3	Classifications	44
4.4	Fusions	44
4.5	Résultats et évaluations	45
4.5.1	Discussion des résultats	47
4.5.2	Présentation de l'interface graphique	52
	Conclusion générale	53
	Bibliographie	54
	Annexe 1	57
A		57
A.1	Les systèmes de primaires	57
A.1.1	Transformation d'un système (R, G, B) vers le système (X, Y, Z) de la CIE	57
A.2	Les systèmes luminance-chrominance	58
A.2.1	Transformation d'un système (RE, GE, BE) vers le système (Y, U, V)	58
A.2.2	Transformation du système (R, G, B) vers le système (Y, Ch1, Ch2) de Carron	58
A.2.3	Transformation du système (R, G, B) vers le système (I1, I2,I3)	59

Table des figures

1.1	Cube des couleurs	11
1.2	Système de coordonnées polaires	13
1.3	Cube des couleurs(modèle triangulaire)	15
1.4	Les familles de systèmes de représentation de la couleur	17
2.1	Hyperplan de séparation optimal avec marge souple dans un cas non linéairement séparable.	22
2.2	La transformation linéaire des données permet une séparation linéaire dans un nouvel espace	23
2.3	Architecture parallèle de classifieur SVM binaire.	24
2.4	Architecture PMC	26
2.5	fonctions d'activation	26
3.1	Taxonomie des méthodes de combinaison parallèle des classifieurs	34
3.2	Représentation par diagrammes de Venn des notions de crédibilités et de plausibilité	39
4.1	Construction de la matrice de passage	41
4.2	Coloration	42
4.3	Échantillonnage	42
4.4	Construction de la base d'apprentissage	43
4.5	Projection 3D de la base d'apprentissage	43
4.6	Exemple d'écart	47
4.7	Diagramme en bâtons des précisions	48
4.8	Diagramme en bâtons des taux de reconnaissances	48
4.9	image originale(o), résultat arbre de décision(a), résultat RNMC(b),résultat GMM(c) résultat Knn(d), résultat Svm(e),résultat vote maj(f),résultat vote maj pondéré(g), résultat bayésienne(h), résultat T.E H1 R1(i),résultat T.E H1 R2(j), résultat T.E H2 R1(k)	49

4.10	image originale(o), résultat arbre de décision(a), résultat RNMC(b),résultat GMM(c) résultat Knn(d), résultat Svm(e),résultat vote maj(f),résultat vote maj pondéré(g), résultat bayésienne(h), résultat T.E H1 R1(i),résultat T.E H1 R2(j), résultat T.E H2 R1(k)	50
4.11	image originale(o), résultat arbre de décision(a), résultat RNMC(b),résultat GMM(c) résultat Knn(d), résultat Svm(e),résultat vote maj(f),résultat vote maj pondéré(g), résultat bayésienne(h), résultat T.E H1 R1(i),résultat T.E H1 R2(j), résultat T.E H2 R1(k)	51
4.12	Captures de l'interface graphique	52

Liste des tableaux

- 4.1 Taux d'apprentissage 45
- 4.2 Résultats obtenus par les classifieurs 46
- 4.3 Résultats obtenus par les votes 46
- 4.4 Résultats obtenus par l'évidence et bayésienne 47

Liste des abréviations

CIE Commission internationale de l'éclairage

NTSC National Television Standards Committee

FCC Federal Communications Commission

PAL Phase Alternation by Line

EBU European Broadcasting Union

SECAM Séquentiel Couleur À Mémoire

LDA Linear discriminant analysis

MMD Maximum Marginale Diversity

ICA Independent component analysis

GMM Gaussian Mixture Model

KNN K-nearest neighbors

(SVM) Support vector machine

RBF Radial basis function

(R,G,B) Red-Green-Blue

HSV Hue Saturation Value

HSL Hue Saturation Luminance

(I,S,T) intensité saturation teinte

Introduction générale

– Contexte

L'anatomie cytologique est une spécialité médicale qui étudie la composition microscopique des cellules. 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 d'intérêt. Cette étape de lecture de la lame consiste en une évaluation visuelle des cellules présentes sur une lame cytologique. Le but de cette étape est soit la détection de cellules anormales ou suspectes, soit la quantification de cellules. Ceci est donc un intérêt capital pour le pathologiste qui doit établir un diagnostic fiable et valide. Le faible nombre de cellules anormales observées durant l'analyse implique une grande concentration du cytotechnologiste. A cause de ce facteur subjectif, quelques erreurs apparaissent et peuvent causer des faux négatifs.

– Problématique

Une approche prometteuse est d'aider le cytopathologiste dans sa recherche des cellules anormales sur la lame. Un système automatique pourrait contribuer à la détection des erreurs d'analyse pour une meilleure fiabilité.

La première étape du système est la segmentation automatique des images cytologique qui consiste à séparer la scène image en régions homogènes .

L'utilisation de plusieurs espaces couleur permet une description exhaustive des couleurs car les propriétés de chaque espace couleur sont susceptibles d'être exploitées.

Cette description conduit à des données de grande dimensions où la possibilité d'avoir une redondance et une incohérence est élevée l'utilisation d'une méthode de réduction de dimension peut contourner ce problème.

Une analyse discriminante linéaire (LDA) est réalisée pour extraire les caractéristiques discriminantes et pertinentes afin de réduire le temps d'exécution, les redondances et le bruit.

– **Contribution**

Pour cette raison nous proposons une technique de segmentation automatique des images microscopiques dont le but est l'extraction de composantes des cellules sanguines (noyau, cytoplasme, globule rouge et plasma). Cette segmentation est basée sur la fusion des décisions des classifieurs entraînés sur la combinaison par LDA des attributs couleurs issues de 14 espaces.

– **Plan de travail**

Le document est organisé comme suite : Le premier chapitre est divisé en deux parties. La première est consacrée à une représentation de la couleur dont nous présentons les systèmes de primaires, les systèmes luminance-chrominance ainsi que les systèmes d'axes indépendants. Dans la deuxième partie, nous parlons de l'utilisation de l'analyse linéaire discriminante.

Le deuxième chapitre est divisé en trois parties. La première introduit les notions de base du concept apprentissage.

La deuxième partie introduit les notions de classifications. Dans La troisième partie nous présentons quelques principale méthodes de la classification supervisée :les machines à vecteurs de support, les réseaux de neurones, les modèles de mélanges de gaussiennes, les arbres de décision et le K plus proches voisins.

Le troisième chapitre est divisé en quatre parties.La première introduit les principes et besoins de la combinaison des classifieurs.

La deuxième partie introduit la notion de classifieur dans le cadre de la combinaison.

La troisième partie est pour la description des méthodes de fusion sans apprentissage.

La quatrième partie est pour la description des méthodes de fusion avec apprentissage.

Le quatrième chapitre est consacré à la partie modélisation de la fusion et expérimentations, il explique en détaille les différentes modelisations de fusions réalisées ainsi une discussion des résultats et une comparaison des travaux réalisé avec notre travail.

La conclusion générale résume l'ensemble de notre travail, et présente quelques perspectives.

Chapitre 1

Représentation de la couleur et analyse discriminante

1.1 Introduction

Dans ce chapitre, nous présentons quatre familles de systèmes de représentation de la couleur, nous nous arrêtons sur les systèmes les plus utilisés dans la littérature, pour analyser leurs propriétés colorimétriques, et pour pouvoir tirer des informations issue d'une image. Cependant, ces différentes propriétés conduisent à une représentation de l'image d'une grande dimension avec une redondance d'information. Ainsi, une technique de réduction de dimension peut résoudre ce problème, dans notre mémoire, l'Analyse Discriminante Linéaire a été retenue.

1.2 Représentation de la couleur

1.2.1 Les systèmes de primaires

Le système (R,G,B) Dans le domaine de traitement d'images, les composantes couleur d'un pixel dépendent du caméra et du système d'acquisition utilisé. Cela pose différents normes mais le système (Rc, Gc, Bc) de la CIE [1] reste néanmoins le système de référence. Le système RGB est un système à trois dimensions qui peut être représenté sous la forme d'un cube (figure4.9) dont chaque axe correspond à une couleur primaire, d'origine O. La couleur est représenté par un point C qui définit le vecteur couleur \vec{OC}

La CIE définites autres primaires, par exemple, la norme NTSC qui utilise les primaires fixées par la FCC. La norme allemande PAL fixée par EBU ou à la norme SECAM.

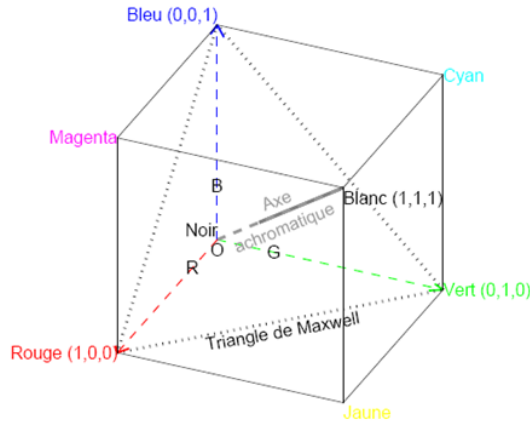


FIGURE 1.1 – Cube des couleurs

Le système (X, Y, Z) Les systèmes (R, G, B) présentent quelques inconvénients majeurs :

- Les composantes trichromatiques peuvent prendre des valeurs négatives.
- Les valeurs des composantes trichromatiques sont liées à la luminance qui est une combinaison linéaire des composantes trichromatiques.
- L'existence de multitude standards de système (R^*, G^*, B^*) .

En 1931, la CIE tente d'établir un système permet de résoudre les inconvénients du système (R_c, G_c, B_c) [2], dont les primaires $[X]$, $[Y]$ et $[Z]$, qui sont imaginaires, une de ces primaires représente l'information de luminance, et toutes les composantes trichromatiques sont positives.

1.2.2 Les systèmes luminance-chrominance

Les composantes d'un système luminance-chrominance sont évaluées à partir des composantes trichromatiques R , G et B par une transformation non linéaire ou une transformation linéaire en utilisant une matrice de passage.

La luminance notée L dans le système luminance-chrominance, qui peut avoir plusieurs sens selon la façon dont elle est considérée. Les systèmes luminance-chrominance sont ainsi notés $(L, Chr1, Chr2)$. Ainsi, il existe différents types de systèmes luminance-chrominance :

Les systèmes perceptuellement uniformes

Judd à proposer de nouvelles composantes de chrominance qui tentent de réduire les disproportions entre les différents ellipses de **MacAdam** [3]. Ces composantes, notées u et v .

Le système (L,u,v) En 1976 la CIE propose le système CIELUV, encore appelé système (L,u,v) [4]. L'information de luminance se calcule par la relation :

$$L = \begin{cases} 116 * \sqrt[3]{Y/Y^W} - 16 & \text{si } Y/Y^W > 0.008856 \\ 903.3 * Y/Y^W & \text{si } Y/Y^W \leq 0.008856 \end{cases} \quad (1.1)$$

Les composantes de chrominance pour le système(L,u,v) sont :

$$u = 13 * L * (u' - u'^w) \quad (1.2)$$

$$v = 13 * L * (v' - v'^w) \quad (1.3)$$

avec :

$$u' = \frac{4X}{X+15Y+3Z} \quad (1.4)$$

$$v' = \frac{9X}{X+15Y+3Z} \quad (1.5)$$

et u' et v' sont les composantes de chrominance respectives de u et v .

Le système (L,a,b) Ou CIELAB, établi par des relations non linéaires à partir du système (X,Y,Z) [5]. Ce système est une réconciliation de l'espace **d'Adams-Nickerson** fréquemment utilisé dans certaines industries. Les composantes de chrominance pour ce système sont :

$$a^* = 500 * (f(X/X^W) - f(Y/Y^W)) \quad (1.6)$$

$$b^* = 200 * (f(Y/Y^W) - f(Z/Z^W)) \quad (1.7)$$

$$f(x) = \begin{cases} \sqrt[3]{x} & \text{si } Y/Y^W > 0.008856 \\ 7.787x * \frac{16}{116} & \text{si } Y/Y^W \leq 0.008856 \end{cases} \quad (1.8)$$

Les systèmes de télévision

Les signaux de télévision séparent l'information de luminance de celle de chrominance. Il existe deux principaux systèmes de télévision :

Le système (Y,I,Q) C'est un système défini par le standard NTSC. On peut réaliser une transformation linéaire des composantes trichromatiques du système par l'utilisation d'une matrice de passage¹.

Le système (Y,U,V) C'est un système défini par le standard PAL (norme allemande). Il est aussi possible d'évaluer les composantes trichromatiques à partir du système (Rc,Gc,Bc) de la CIE en utilisant une matrice de passage.

1.2.3 Les systèmes perceptuels

Les systèmes perceptuels ressemblent à des systèmes luminance-chrominance puisqu'ils sont composés d'une information quantifiant la luminance ainsi que de deux informations quantifiant la chrominance. Ces notions mises en évidence en 1909 par **A.Henry Munsell**

Il existe de nombreux systèmes de ce type dans la littérature telles que : ISH, HSL, HSV, TLS, LCH, LSH, LST, ITS, ... Nous distinguons deux familles de systèmes perceptuels :

Les systèmes de coordonnées polaires

Cette famille de systèmes décomposant l'information couleur en un axe de luminance et un plan de chrominance, par transposition des coordonnées cartésiennes en coordonnées polaires (figure1.2).

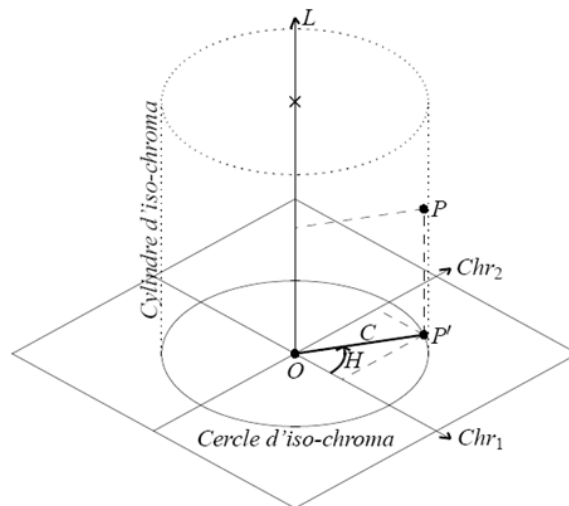


FIGURE 1.2 – Système de coordonnées polaires

1. L'annexe A propose plusieurs transformations standards entre les systèmes à l'aide des matrices des passage.

La première composante d'un système de coordonnées polaires représente l'information de luminance L, identique à la première composante du système luminance-chrominance correspondant, qui est défini par l'équation :

$$L = \left\| \vec{OP'} \right\| = \sqrt{Chr_1^2 + Chr_2^2} \quad (1.9)$$

Géométriquement on peut exprimer la valeur de H de la façon suivante :

$$H = \widehat{(\vec{Chr}_1, \vec{OP'})} \begin{cases} 0 \leq H < \frac{\pi}{2} & \text{si } Chr_1 > 0 \text{ et } Chr_2 \geq 0 \\ \frac{\pi}{2} \leq H < \pi & \text{si } Chr_1 \leq 0 \text{ et } Chr_2 > 0 \\ \pi \leq H < \frac{3\pi}{2} & \text{si } Chr_1 < 0 \text{ et } Chr_2 \leq 0 \\ \frac{3\pi}{2} \leq H < 2\pi & \text{si } Chr_1 \geq 0 \text{ et } Chr_2 < 0 \end{cases} \quad (1.10)$$

L'utilisation des relations de trigonométrie permet d'évaluer C, qui peut être exprimé de la façon suivante :

$$C = \arctan\left(\frac{Chr_1}{Chr_2}\right) \quad (1.11)$$

Les systèmes humains de perception de la couleur

Dans ce type d'espace, la couleur est décrite comme l'homme la qualifie, exprimé par trois composantes forment un système (I,S,T), La première composante I représente l'intensité correspond à l'information de luminance, ensuite la saturation S représente le niveau de coloration, et T la teinte.

Le système (I,S,T) Ce modèle correspond à une expression des composantes I, S et T dans le cube (figure1.3) des couleurs d'un système (R,G,B). Dans cette représentation du cube des couleurs, Nous retrouvons l'axe achromatique qui correspond à l'axe d'intensité dans le système (I,S,T) [1].

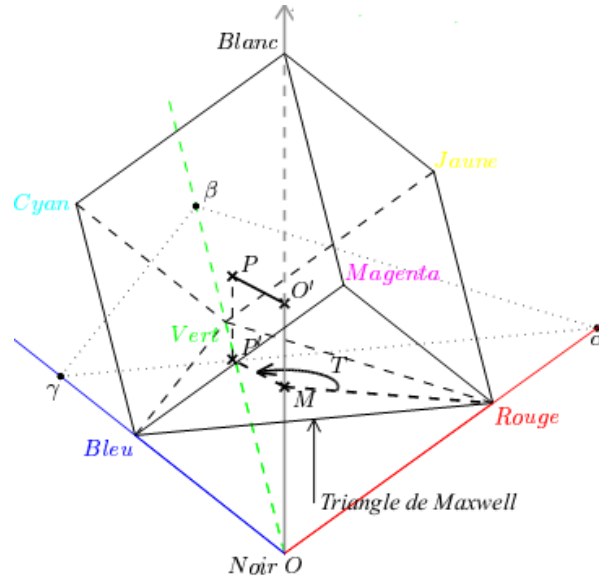


FIGURE 1.3 – Cube des couleurs(modèle triangulaire)

L'intensité I correspond à la norme du vecteur $\vec{OO'}$ Elle est donc évaluée par la formule :

$$I = \left\| \vec{OO'} \right\| = \frac{R+G+B}{\sqrt{3}} \quad (1.12)$$

La saturation S correspond au rapport entre la norme du vecteur \vec{OP} et la norme du vecteur $\vec{OO'}$ qui représente l'intensité, alors la saturation évaluée par la formule :

$$S = \frac{\left\| \vec{OP} \right\|}{\left\| \vec{OO'} \right\|} = \sqrt{2} * \frac{\sqrt{R^2+G^2+B^2-RG-GB-RB}}{R+G+B} \quad (1.13)$$

ou encore :

$$S = \frac{\sqrt{(R-G)^2+(G-B)^2+(B-R)^2}}{R+G+B} \quad (1.14)$$

Liang[6], présente la formulation de la teinte sans utiliser les coordonnées trichromatiques.

$$T = \begin{cases} \frac{1}{3} * \frac{G-B}{S} & \text{si } \min(R, G, B) = B \\ \frac{1}{3} * \frac{B-R}{S} & \text{si } \min(R, G, B) = R \\ \frac{1}{3} * \frac{R-G}{S} & \text{si } \min(R, G, B) = G \end{cases} \quad (1.15)$$

Le Système (L,S,H) Ce système a été défini par la CIE en même temps que les systèmes (L,C,H). On référence les travaux de **Pujas** [7] , **Carron** [8] , **Shih** [9] ou **Galichet** [10] qui présentent d'autres systèmes de perception humaine de la couleur.

1.2.4 Les systèmes d'axes indépendants

Le système (I1,I2,I3) Une expérience menée par **Ohta** en 1980 [11] montre qu'il existe une transformation unique appelée (I1,I2,I3) modélisant la transformation de **Karhunen-Loeve**[12], et se montrant aussi l'efficacité de cette dernière.

Cette transformation est définie par les équations suivantes :

$$\begin{cases} I_1 = \frac{R+G+B}{3} \\ I_2 = \frac{R-B}{2} \\ I_3 = \frac{2G-R-B}{4} \end{cases} \quad (1.16)$$

La première composante représente la luminance, qui est la plus discriminante, et les deux autres composantes apportent moins d'informations représentent les oppositions bleu-rouge et magenta-vert.

1.3 Synthèse

Sur la figure (figure1.4) , **Vandenbroucke** a proposé des regroupements en 4 familles [13] telles que nous les avons présentées dans ce premier chapitre :

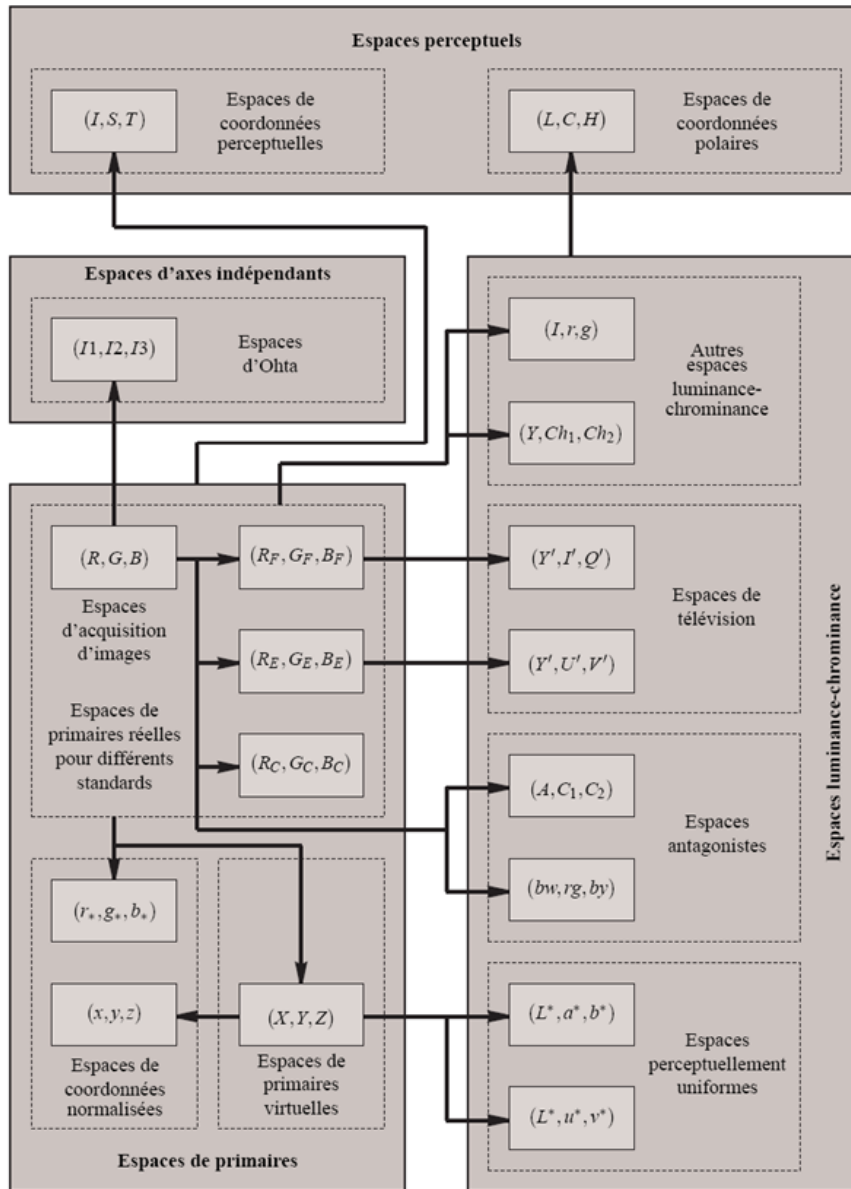


FIGURE 1.4 – Les familles de systèmes de représentation de la couleur

Les rectangles grisés correspondent à ces quatre familles et les flèches indiquent les transformations possibles entre les espaces.

1.4 Analyse Linéaire Discriminante (LDA)

L'objectif de l'Analyse Linéaire Discriminante (LDA), est de réduire le nombre de dimensions en présentant le maximum des données. Elle cherche les axes tels que la projection des données dans l'espace engendré par ces axes permette une plus grande discrimination

entre les classes [14].

Algorithme

La matrice de dispersion totale (total scatter matrix) est définie par :

$$S_T = \sum_{\vec{x}_i \in A} (\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T \quad (1.17)$$

où $\vec{\mu}$ est le vecteur moyen des données. La matrice de dispersion intraclasse (within-class scatter matrix) est définie comme la dispersion des moyennes de chaque classe :

$$S_W = \sum_{k=1}^{nc} \sum_{\vec{x}_i \in C_k} (\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T = \sum_{k=1}^{nc} S_k \quad (1.18)$$

où S_k est la matrice de dispersion des éléments de la classe . La matrice de dispersion interclasse (between-class scatter matrix) est définie comme la moyenne des dispersions de chaque classe :

$$S_B = \sum_{k=1}^{nc} nC_k (\vec{\mu}_k - \vec{\mu})(\vec{\mu}_k - \vec{\mu})^T \quad (1.19)$$

où $\vec{\mu}_k$ est le vecteur moyen des éléments de la classe . Nous cherchons une transformation linéaire de matrice qui maximise la dispersion des classes (mesurée par la matrice de dispersion interclasse) par rapport à la dispersion des éléments au sein des classes (mesurée par la matrice de dispersion intraclasse). Une mesure simple de dispersion est la valeur du déterminant des matrices de dispersion. Le critère de Fisher propose de choisir la transformation linéaire qui maximise le ratio :

$$J_{LDA}(\Omega) = \frac{|S'_B|}{|S'_W|} = \frac{|\Omega^T S_B \Omega|}{|\Omega^T S_W \Omega|} \quad (1.20)$$

où S'_B et S'_W sont les matrices de dispersion interclasse et intraclasse des projetés . On montre que les vecteurs colonnes sont les vecteurs propres qui correspondent aux plus grandes valeurs propres solutions de l'équation :

$$S_B \omega_k = \lambda_k S_W \omega_k \quad (1.21)$$

Chapitre 2

Apprentissage et classifications supervisées

2.1 Introduction

La présence d'images numériques dans des applications devient importante de jour en jour. Plusieurs formalismes existent pour modéliser l'information issue d'une image ou d'une mesure [15]. Ces modélisations posent problème au niveau de la taille des données, ce problème peut être réglé par l'utilisation d'une méthode de réduction de dimension. Nous avons choisi la méthode LDA qui nous permet de réduire la taille des données d'image en combinant les composantes couleurs et donc elle nous facilite la classification.

2.2 Notion d'apprentissage

L'apprentissage automatique (machine-Learning en anglais) est une discipline scientifique, qui est aussi l'un des champs d'étude de l'intelligence artificielle. L'apprentissage automatique fait référence au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine (au sens large) d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques. Cette notion englobe toute méthode permettant de construire un modèle de la réalité à partir de données, soit en améliorant un modèle partiel ou moins général, soit en créant complètement le modèle ou bien capacité à améliorer les performances au fur et à mesure de l'exercice d'une activité

2.2.1 Types d'apprentissage

Les algorithmes d'apprentissage peuvent se catégoriser selon le mode d'apprentissage qu'ils emploient :

1. **L'apprentissage supervisé**

Si les classes sont prédéterminées et les exemples connus, le système apprend à classer selon un modèle de classement ; on parle alors d'apprentissage supervisé (ou d'analyse discriminante).

2. **L'apprentissage non-supervisé**

Quand le système ou l'opérateur ne disposent que d'exemples, mais non d'étiquettes, et que le nombre de classe et leur nature n'ont pas été prédéterminés, on parle d'apprentissage non supervisé ou clustering.

3. **L'apprentissage par renforcement**

L'algorithme apprend un comportement étant donné une observation. L'action de l'algorithme sur l'environnement produit une valeur de retour qui guide l'algorithme d'apprentissage.

2.3 Notion de la classification

La classification est une opération de structuration qui vise à organiser un ensemble d'observation en groupes homogènes et contrastés afin de faciliter l'analyse des informations et d'effectuer des prédictions [16].

2.4 Principales méthodes de la classification supervisée

2.4.1 Machine à vecteurs de support (SVM)

Les machines à vecteurs support ont été introduites en 1995 par **Cortes** et **Vapnik** [17]. Elles sont utilisées dans de nombreux problèmes d'apprentissage : reconnaissance de forme, catégorisation de texte ou encore diagnostic médical.

Les SVM reposent sur deux notions : celle de marge maximale et celle de fonction noyau. Elles permettent de résoudre des problèmes de discrimination non linéaire. La marge est la distance entre la frontière de séparation et les échantillons les plus proches appelés vecteurs support. Dans un problème linéairement séparable les SVM trouvent une séparatrice qui maximise cette marge.

Cas à deux classes

1. Principe

Les machines à vecteurs de support forment une classe d'algorithmes d'apprentissage supervisés. Nous nous intéressons à une fonction notée f qui à toute entrée x fait correspondre une sortie $y = f(x)$. Le but est d'essayer d'apprendre f à partir d'un ensemble de couple $(x_i; y_i)$. Dans ce problème les machines à vecteurs de support vont être utilisées pour classifier une nouvelle observation x en se limitant à deux classes $y \in \{-1, 1\}$. Nous allons donc construire une fonction f qui à chaque valeur d'entrée dans un ensemble \mathbb{R}^d va faire correspondre une valeur de sortie $y \in \{-1, 1\}$:

$$f : \mathbb{R}^d \rightarrow \{-1, 1\} f(x) = y. \quad (2.1)$$

Dans le cas linéaire, une fonction discriminante h est obtenue par combinaison linéaire d'un vecteur d'entrée $x = (x_1; \dots; x_d)$ et s'écrit :

$$h(x) = w * x + b. \quad (2.2)$$

La classe est donnée par le signe de $h(x)$: $f(x) = \text{sign}(h(x))$. Si $h(x) \geq 0$ alors x est de classe 1 sinon x est de classe -1. Le séparateur est alors un hyperplan affiné d'équation :

$$w * x + b = 0. \quad (2.3)$$

Si (x_i, y_i) est un des p éléments de la base d'apprentissage noté A_p , on veut trouver le classifieur h tel que $y_i(w * x_i + b) \geq 0, i \in [1, p]$ Dans le cas simple linéairement séparable il existe de nombreux hyperplans séparateurs. Selon la théorie de **Vapnik** [17] l'hyperplan optimal (optimum de la distance interclasse) est celui qui maximise la marge. Cette dernière étant définie comme la distance entre un hyperplan et les points échantillons les plus proches. Ces points particuliers sont appelés vecteurs support. La distance entre un point x quelconque et l'hyperplan est donnée par l'équation :

$$d(x) = \frac{|w * x + b|}{\|w\|} \quad (2.4)$$

2. Forme Primale

Les paramètres w et b étant définis à un coefficient multiplicatif près, on choisit de les normaliser pour que les échantillons les plus proches (x_s) vérifient l'égalité suivante : $y_s(w * x_s + b) = 1$. Donc quelque soit l'échantillon x_i on obtient : $y_i(w * x_i + b) \geq 1$.

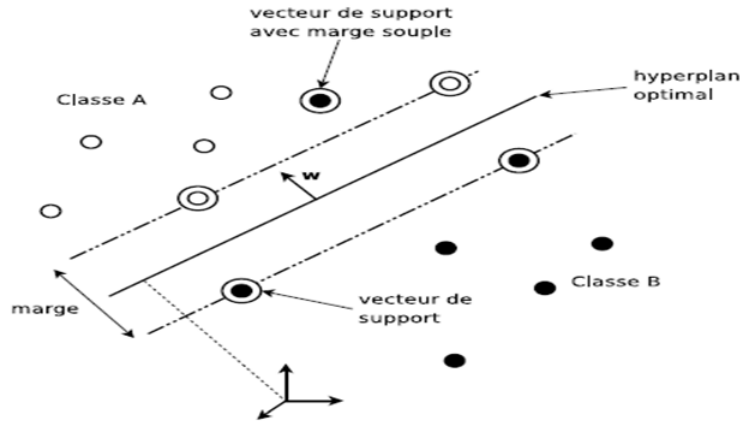


FIGURE 2.1 – Hyperplan de séparation optimal avec marge souple dans un cas non linéairement séparable.

La distance entre l'hyperplan et un point support est donc définie par $\frac{1}{\|w\|}$.

La marge géométrique entre deux classes est égale à $\frac{2}{\|w\|}$.

La forme primale (qui dépend seulement de w et b) des SVM est donc un problème de minimisation sous contrainte qui s'écrit :

$$Y = \begin{cases} \min(1/2 \|w\|^2) \\ \forall (x_i, y_i) \in A_p, y_i(w * x_i + b) \geq 1 \end{cases} \quad (2.5)$$

3. Forme duale

La formulation primale peut être transformée en formulation duale en utilisant les multiplicateurs de Lagrange. L'équation 2.5 s'écrit alors sous la forme suivante :

$$L(w, b, \alpha) = \frac{1}{2} * \|w\|^2 - \sum_{i=1}^p \alpha_i * (y_i(w * x + b) - 1) \quad (2.6)$$

La formulation de Lagrange permet de trouver les extremums en annulant les dérivées partielles de la fonction $L(w, b, \alpha)$. Le lagrangien L doit être minimisé par rapport à w et b et maximisé par rapport à α .

Nous donnons la valeur optimale pour w noté w^* : $w^* = \sum_{i=1}^p \alpha_i^* * y_i * x_i$, avec α_i^* les coefficients de Lagrange optimaux. En utilisant l'équation de l'hyperplan nous obtenons l'hyperplan de marge maximale.

$$h(x) = \sum_{i=1}^p \alpha_i^* y_i * x * x_i + b. \quad (2.7)$$

4. Astuce du noyau

Le cas linéairement séparable est peu intéressant, car les problèmes de classification sont souvent non linéaires. Pour résoudre ce point la méthode classique est de projeter les données dans un espace de dimension supérieur appelé espace de redescription. L'idée étant qu'en augmentant la dimensionnalité du problème on se retrouve dans le cas linéaire vu précédemment. Nous allons donc appliquer une transformation non linéaire $\Phi(\bullet)$ aux vecteurs d'entrée x_i tel que $x_i \in \mathfrak{R}$ et $\Phi(x_i) \in \mathfrak{R}(e > d)$. Ce changement va conduire à passer d'un produit scalaire dans l'espace d'origine $x_i \cdot x_j$ à un produit scalaire $\Phi(x_i) \cdot \Phi(x_j)$ dans l'espace de redescription l'astuce est d'utiliser une fonction noyau notée K qui évite le calcul explicite du produit scalaire dans l'espace de redescription. Les fonctions noyaux doivent satisfaire le théorème de Mercer. Nous avons alors l'égalité suivante :

$$k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (2.8)$$

Il existe de nombreuses fonctions noyau prédéfinies, les deux les plus usitées sont le noyau gaussienne (2.9) et le noyau polynomial (2.10)

$$k_\gamma(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (2.9)$$

$$k_{\gamma,d,r}(x_i, x_j) = (\gamma \cdot x_i \cdot x_j + r)^d. \quad (2.10)$$

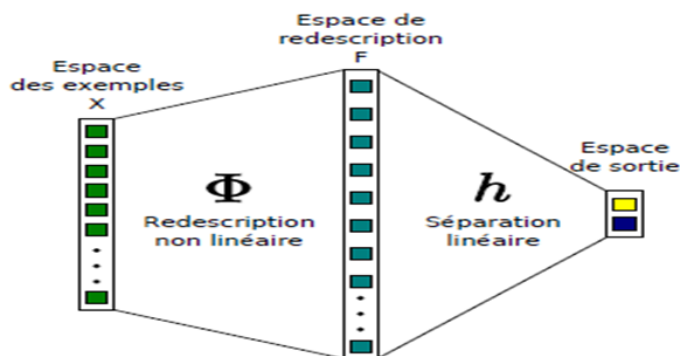


FIGURE 2.2 – La transformation linéaire des données permet une séparation linéaire dans un nouvel espace

Classification multi-classes

Les machines à vecteurs support vu précédemment effectuent une classification binaire. Dans la plupart des contextes nous cherchons à résoudre des problèmes multi-classes. Plu-

sieurs méthodes ont été développées, elles se regroupent en deux approches : les approches parallèles et hiérarchiques.

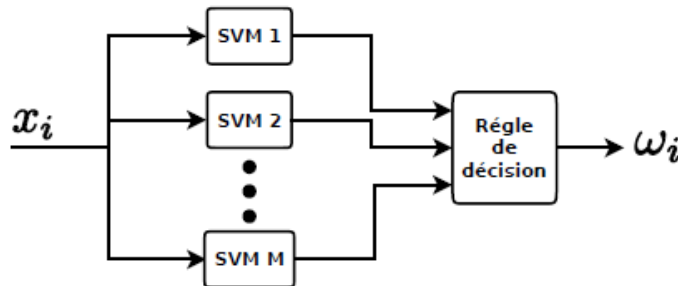


FIGURE 2.3 – Architecture parallèle de classifieur SVM binaire.

1. Un contre un

Cette approche utilise $k*(k-1)/2$ classifieurs où chaque SVM est entraîné pour départager deux classes w_i et w_j . On construit ainsi autant de classifieurs que de couples de classe $(w_i; w_j)$; $i \neq j$ possibles. Nous avons donc de nouveau l'architecture présentée à la figure 2.3 mais avec $M = k*(k-1)/2$. La règle de décision s'obtient en utilisant la méthode du vote majoritaire. Chaque classifieur va incrémenter le score S_i (respectivement S_j) associé à la classe w_i (respectivement w_j). Ensuite le score $S_i(x)$ le plus élevé va permettre d'attribuer l'étiquette i à l'entrée x . Certaines implémentations ajoutent une pondération au vote de chaque classifieur.

2. Un contre tous

Cette approche utilise une architecture parallèle de k machines à vecteur de support, une pour chaque classe. Chaque machine à vecteur de support résout un problème à deux classes : une classe ω_i ($\omega_i \in \Omega$), et toutes les autres $\Omega - \omega_i$. Sur la figure 2.3 nous avons M classifieurs binaires, dans cette première approche $M = k$. La règle de décision finale est l'application du principe « winner take all ». Pour chaque classifieur un score va être établi et l'étiquette attribuée à l'entrée x_i est celle associée au classifieur qui obtient le meilleur score. Son principal inconvénient est d'effectuer un apprentissage qui peut être grandement déséquilibré. Ainsi pour un classifieur ω_i nous pouvons avoir un très petit nombre d'exemples de la classe i et un grand nombre de contre exemples [18].

2.4.2 Réseaux de neurones

Historique et définition

En 1943, Warren McCulloch et Walter Pitts ont proposé un des premiers modèles de neurone artificiel, qui deviendra la base des réseaux de neurones artificiels. Ce modèle de neurone (également appelé neurone de McCulloch-Pitts) utilise la sommation des intrants pesés (weighted inputs) comme intrant « net » X (intrant direct de la fonction du neurone) et comme fonction d'activation, la fonction de signe (comparé avec le seuil θ)

$$X = \sum_{i=1}^n x_i * w_i \quad (2.11)$$

$$Y = \begin{cases} +1 & \text{si } X > \theta \\ -1 & \text{si } X < \theta \end{cases}$$

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire (neurone artificiel) calcule une sortie unique sur la base des informations qu'il reçoit.

Perceptron multi couche

Il y a principalement deux facteurs qui influent sur l'apprentissage. Ce sont la qualité de l'échantillonnage d'apprentissage (les exemples qui constituent la base d'apprentissage) et la diversité des valeurs. En effet, le réseau de neurones généralisera mieux (aura plus de chances de répondre correctement en lui donnant en entrée des informations non présentes dans les exemples d'apprentissage) si la qualité de l'échantillonnage est meilleure et si les données des exemples d'apprentissage sont variées. Intuitivement, on est conscient que s'il sait répondre correctement pour un nombre fini de situations les plus diverses, il sera alors plus proche de ce que l'on veut dans une situation nouvelle.

RÉTRO-PROPAGATION DU GRADIENT (back-propagation) La méthode de modification des poids est très simple avec l'algorithme de **Rosenblatt**, mais il implique quelques limitations d'apprentissage. Dans le cas de perceptrons multicouches, comme on ne sait pas les extrants (outputs) désirés des couches cachées, mais seulement de la dernière couche (figure 2.4), il faut propager la responsabilité des erreurs de la dernière couche à la première, dans le sens contraire de l'exécution de réseau, d'où le nom rétro-propagation. De plus, les perceptrons multicouches utilisent des neurones munis de la fonction d'activation sigmoïde (figure 2.5), celle-ci permettant les nuances nécessaires à une bonne utilisation de la

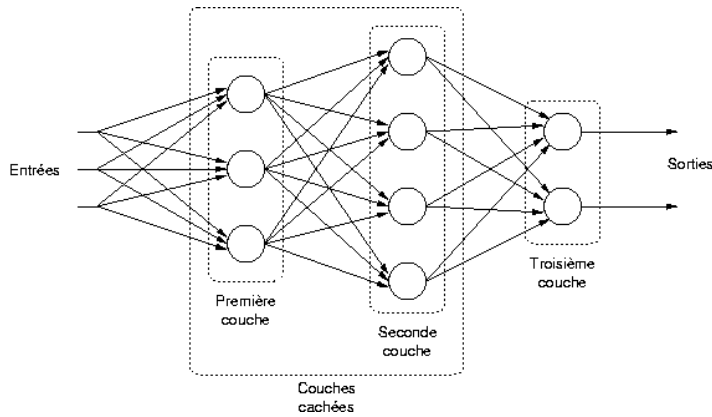


FIGURE 2.4 – Architecture PMC

Avec	Avec
$f'(x) = f(x)(1 - f(x))$ $f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = (1 + f(x))(1 - f(x))$ $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
fonction sigmoïde	fonction tangente hyperbolique

FIGURE 2.5 – fonctions d'activation

rétro-propagation. Voici les principales lignes de la méthode de rétro-propagation :

Calcul de l'erreur :

$$e_k(p) = y_{d,k} - y_k \quad (2.12)$$

L'erreur du neurone k est la différence entre la valeur de l'extrant désiré et la valeur actuelle de l'extrant du neurone k) Correction de poids (par la loi delta) :

$$\Delta w_{jk} = \alpha * y_j(p) \delta_k(p) \quad (2.13)$$

La correction du poids du lien du neurone j au neurone k, où $\delta_k(p)$ est le gradient d'erreur du neurone k à l'itération p. Gradient d'erreur (pour les neurones de la couche de sortie) :

$$\delta_k(p) = F' X_k(p) e_k(p) = y_k(p) [1 - y_k(p)] e_k(p) \quad (2.14)$$

La deuxième partie est obtenue par la dérivée de la fonction sigmoïde appliquée à l'intrant net du neurone k à l'itération p Gradient d'erreur (pour les neurones d'une couche cachée) :

$$\delta_j(p) = y_j(p)[1 - y_j(p)] \sum_{k=1}^l \delta_k(p) w_{j,k}(p) \quad (2.15)$$

où l est le nombre de neurones sur la couche suivante [19].

Validation croisée La validation croisée est une technique qui permet de tester un modèle d'apprentissage. La validation croisée se décline en plusieurs sous méthodes. La plus répandue est la méthode « k-Fold » avec typiquement $k \in [4; 10]$. Si l'on a une base d'apprentissage A_p contenant p éléments : $A_p = \{x_1, \dots, x_p\}$ la validation croisée consiste à appliquer les cinq étapes suivantes :

1. Découper l'ensemble des exemples en k sous-ensembles disjoints de taille p/k
2. Apprendre sur les $k-1$ sous-ensembles.
3. Calculer l'erreur sur la k -ième partie.
4. Répéter le processus k fois.
5. Obtenir l'erreur finale en calculant la moyenne des k erreurs précédentes.

La validation croisée est simple à mettre en œuvre et utilise toutes les données.

Elle permet d'obtenir une estimation de l'erreur de généralisation. Cela permet d'éviter le sur-apprentissage .

2.4.3 Modèle de mélanges de gaussiennes

Le terme "Fonction Gaussien" est un autre terme pour la loi normale. Si les événements sont issus d'une composition de " N " phénomènes, la densité $p(x)$ prendra la forme d'une composition de lois normales. Dans un tel cas, on peut approximer par une somme pondérée de densités normales.

$$p(x) = \sum_{n=1}^N \alpha_n N(x; \mu_n, \sigma_n) \quad (2.16)$$

Une telle somme est connue par le terme "mélange de Gaussiennes" pour chaque Gaussien il faut estimer trois paramètres : $v(\alpha_n, \mu_n, \sigma_n)$. En total il y a $3 * N$ paramètres à estimer.

$v(\alpha_1, \mu_1, \sigma_1, \alpha_2, \mu_2, \sigma_2, \dots, \alpha_N, \mu_N, \sigma_N)$.

Si tous les μ et σ sont fixes (et égales), on pourrait calculer les α_n directement. Mais parce qu'ils sont variables, il faut les estimer par un processus itératif. Un tel processus est composé de deux étapes. Ces étapes fournissent une estimation des variables cachées. Pour un mélange de gaussiennes, les variables cachées sont les sources des événements. $hm = N$ la source de

l'événement M . On suppose chaque événement est issue d'un des N sources. Nous allons construire une table de probabilités. $h(m, n) = \Pr$ l'événement E_m est issu de la source N les probabilités, $h(m, n)$, nous donnera les facteurs de Mélange, α_n , ainsi que μ_N et σ_N .

L'algorithme d'estimation s'appelle "Expectation-Maximization" ou "EM".

Ébauche de l'Algorithme EM

Soit une ensemble ("training set") de M observations $T = \{X_m\}$.

On fait une première estimation des paramètres $v(0)$ et puis on alterne "E" et "M".

E : Faire une estimation des valeurs manquantes, $h(m, n)$, pour les événements.

$h(m, n)(i) = p(hm | X_1, X_2, \dots, X_M, v(i))$ pour chaque terme "n" [20].

$$h(m, n)(i) = \frac{\alpha_n(i)N(X_m; \mu_n(i), \sigma_n(i))}{\sum_{j=1}^N \alpha_j(i)N(X_m; \mu_j(i), \sigma_j(i))} \quad (2.17)$$

M : Recalculer $v(i+1)$ avec $p(hm | X_1, X_2, \dots, X_M, v(i))$.

$$S_n(i+1) = \sum_{m=1}^M h(m, n)(i) \quad (2.18)$$

$$\alpha_n(i+1) = \frac{1}{M}(i+1) \quad (2.19)$$

$$\mu_n(i+1) = \frac{1}{S_n(i+1)} \sum_{m=1}^M h(m, n)(i) X_m \quad (2.20)$$

$$\sigma_n^2(i+1) = \frac{1}{S_n(i+1)} * \sum_{m=1}^M h(m, n)(i) (X_m - \mu_n(i+1))^2 \quad (2.21)$$

2.4.4 Arbres de décision

Les arbres de décision sont l'une des structures de données majeures de l'apprentissage statistique. Leur fonctionnement repose sur des heuristiques qui, tout en satisfaisant l'intuition, donnent des résultats remarquables en pratique (notamment lorsqu'ils sont utilisés en « forêts aléatoires »). Leur structure arborescente les rend également lisibles par un être humain, contrairement à d'autres approches où le prédicteur construit est une « boîte noire ». Un arbre de décision modélise une hiérarchie de tests sur les valeurs d'un ensemble de variables appelées attributs. À l'issue de ces tests, le prédicteur produit une valeur numérique ou choisit un élément dans un ensemble discret de conclusions.

Construction d'un arbre de décision

Idée centrale : Diviser récursivement et le plus efficacement possible les exemples de l'ensemble d'apprentissage par des tests définis à l'aide des attributs jusqu'à ce que l'on obtienne des sous-ensembles d'exemples ne contenant (presque) que des exemples appartenant tous à une même classe. À partir d'un ensemble d'observations $T = \{(x, y)\}$, on souhaite construire un arbre de décision prédisant l'attribut y en fonction des nouvelles instances x . Pour ce faire, il existe essentiellement deux familles d'algorithmes à ce jour : les arbres de QUINLAN et les arbres CART. Les deux approches suivent le paradigme diviser-pour-régner, que l'on peut schématiser (dans le cas d'attributs à valeurs discrètes) par le pseudo-code suivant :

Arbre Décision(T)

```

if "condition d'arrêt" then
  feuille( $T$ )
else
  return le meilleur attribut  $i$  entre 1 et  $m$ 
  for  $v$  de l'attribut  $i$  do
     $T[v] = \{(x, y) \text{ de } T \text{ tels que } xi = v\}.t[v] = \text{ArbreDcision}(T[v]).$ 
  end for
  nœud ( $i, v \rightarrow t[v]$ )
end if

```

Condition d'arrêt :

Elle influe sur la profondeur et la précision du prédicteur produit. Par exemple, la condition $|T| = 1$ produira des arbres très précis sur l'ensemble d'entraînement (i.e. pour une instance $(x, y) \in T$, la prédiction pour x sera exactement y) mais également très profonds, donc longs à calculer, et qui risquent de sur apprendre les données d'entraînement.

Meilleur attribut :

Il s'agit d'évaluer localement quel attribut apporte « le plus d'information » (ou encore « est le plus corrélé ») au résultat à prédire. On verra plusieurs de ces critères par la suite. Lorsque l'attribut x_i est à valeurs réelles, on adapte l'algorithme ci-dessus en choisissant une valeur de partage (split value) v et en effectuant le test $x_i \leq v$. On notera « nœud ($i, v, t \leq t$) » le constructeur associé. En particulier, si tous les attributs sont réels, l'arbre de décision obtenu est binaire.

Classification

```

Classifier(x, t)
  if "t = feuille(Tf)" then
    return la classe majoritaire de Tf
  else if t = noeud(i, v, t left, t right) then
    if  $x[i] \leq v$  then
      return Classifier(x, t left)
    else
      return Classifier(x, t right)
    else
      t = noeud(i, {v → t[v]})
      return Classifier(x, t[x[i]])
  end if
end if

```

Mesurer l'homogénéité des feuilles

Les observations précédentes soulignent qu'un arbre de classification est d'autant meilleur que (les instances de) ses feuilles sont de classes homogènes. On souhaite alors définir une mesure de l'hétérogénéité d'une feuille qui jouera le même rôle que la variance dans le cas de la régression. Soient p_1, \dots, p_C les fréquences relatives des classes 1, ..., C dans Tf, et c^* la classe la plus fréquente. Voici trois mesures fréquemment rencontrées dans la littérature :

Taux d'erreur :

$$e(T_f) = 1 - p_{C^*}. \quad (2.22)$$

Il s'agit du taux d'erreurs de classification sur l'ensemble d'entraînement.

Critère de Gini :

$$e(T_f) = \sum p_c(1 - p_c). \quad (2.23)$$

Il s'agit du taux d'erreur sur l'ensemble d'entraînement d'un algorithme randomisé qui retournerait la classe c avec probabilité p_c (au lieu de toujours retourner la classe c^*). C'est la mesure d'erreur utilisée dans CART.)

Entropie :

$$e(T_f) = \sum p_c \log(p_c) \quad (2.24)$$

Il s'agit d'un estimateur de l'entropie de la classe d'une instance de Tf tirée uniformément au hasard. C'est la mesure d'erreur utilisée dans les arbres ID3 et C4.5

Choix de l'attribut de partage

La mesure d'erreur est utilisée de même pour choisir le meilleur attribut de partage : si un attribut i partitionne T en $T = \bigcup_{v_i} T_{v_i}$, chaque ensemble T_{v_i} a sa propre erreur $e(T_{v_i})$ et

l'erreur attendue après un branchement sur cet attribut (pour un élément tiré uniformément au hasard de T) est :

$$e_i = \sum_{vi} \frac{|T_{vi}|}{T} e(T_{vi}) \quad (2.25)$$

L'attribut qui minimise cette erreur est (heuristiquement) considéré comme le meilleur choix possible [21].

2.4.5 Les K plus proches voisins

Principe

La forme à classer est comparée aux autres formes ayant déjà été classées, et on lui affecte la classe la plus représentée parmi les k plus proches d'elle. Dans le cas particulier $k=1$, c'est la classe de la forme la plus proche de la forme à classer qui est affectée à cette dernière. La notion de proximité utilisée est quantifiée par une mesure de similarité. La mesure de similarité la plus utilisée est la distance euclidienne. Pour que l'algorithme puisse démarrer, il faut utiliser un certain nombre d'exemples dont la classe est connue, auxquels vont être comparés les premiers vecteurs à classer.

Paramètres

Les différents paramètres de l'algorithme, réglables par l'utilisateur, sont :

- le nombre d'exemples dont la classe est connue, utilisés au démarrage de l'algorithme.
- la valeur de k .
- la mesure de similarité (ex. : distance euclidienne).

Algorithme

Choisir des exemples initiaux, dont la classe est connue

Pour chaque vecteur à classer

- Mesurer la similarité entre le vecteur à classer et tous les vecteurs déjà classés .
- Déterminer les k vecteurs pour lesquels la similarité est la plus grande (kPPV).
- Déterminer la classe la plus représentée parmi ces k vecteurs et affecter cette classe au vecteur à classer [22].

2.5 Conclusion

Dans ce chapitre nous avons défini brièvement le fonctionnement de chacun des classifieurs, ces classifieurs sont supervisés, et ils sont connus par leur efficacité et robustesse.

Chapitre 3

Fusion de décisions

3.1 Introduction

La fusion de données est souvent un passage nécessaire pour l'analyse de systèmes ou de situations complexes. Il en est de même en analyse d'images ou elle permet de tirer profit de la redondance et de la complémentarité des données. L'idée principale de la fusion de données est de combiner différentes informations relatives à un problème. Pour cette raison, on trouve souvent dans la littérature le terme fusion d'informations au lieu de fusion de données. Les informations fusionnées pouvant être de natures très variées. Elles peuvent être des données provenant de sources différentes ou bien présentant des caractéristiques différentes, extraites des mêmes données initiales ou encore des décisions provenant des sorties des classifieurs [23].

3.2 Principes et besoins de la combinaison des classifieurs

L'idée principale derrière la combinaison de classifieurs est l'augmentation de la qualité des résultats. En effet, de plus en plus, les chercheurs se rendent compte qu'il n'existe aucune méthode de classification pouvant manifester une supériorité sur les autres méthodes dans tous les problèmes et toutes les situations [24]. Si aucune d'entre elles ne peut satisfaire entièrement les exigences d'une application envisagée, l'utilisation simultanée de plusieurs méthodes en même temps peut, éventuellement, permettre d'en cumuler les avantages sans en cumuler les inconvénients. Différents schémas de combinaison de classifieurs ont été proposés dans la littérature selon les critères de combinaison [25]. Ainsi, on trouve le critère de l'ordre d'exécution des classifieurs où on observe trois types de combinaison : séquentielle,

parallèle et série parallèle. Un autre critère de combinaison est la combinaison avec bouclage où le résultat de combinaison peut à son tour être utilisé par les classifieurs et modifier ainsi leurs sorties. Il y a aussi le critère de combinaison avec interaction entre les classifieurs où les classifieurs interagissent entre eux et modifient leurs résultats avant que la combinaison ne soit réalisée. Enfin, le critère de combinaison utilise directement la forme x d'entrée en même temps que les réponses des classifieurs[26]. Les méthodes de fusion de données ont été élaborées pour permettre de gérer des informations multiples, en tenant compte éventuellement de leur conflit éventuel ou de leur désaccord [27]. Ces informations redondantes et complémentaires seront alors combinées numériquement par un formalisme adapté afin d'obtenir la meilleure connaissance possible de la situation étudiée en en réduisant l'incertitude et l'imprécision. [28].

3.3 Définition d'un classifieur dans le cadre de la combinaison

Nous appelons classifieur tout outil de reconnaissance qui reçoit une forme x en entrée et donne des informations à propos de la classe de cette forme. Cet outil est vu comme une fonction qui, à l'aide des descripteurs de la forme x à reconnaître, décide d'attribuer à x la classe C_i parmi un nombre fini de classes possibles, $i=1, \dots, M$.

$$x \rightarrow e(x) = C_i \quad (3.1)$$

Les réponses fournies par un classifieur peuvent être divisées en trois catégories suivant le niveau d'information apporté par le classifieur.

Type classe : $e_j(x) = C_i (i \in \{1, \dots, M\})$,

indique que le classifieur j a attribué la classe C_i à x .

Type rang : $e_j(x) = [r_1^j, r_2^j, \dots, r_M^j]$ ou r_i^j est le rang attribué à la classe i par le classifieur j .

Type mesure : $e_j(x) = [m_1^j, m_2^j, \dots, m_M^j]$. ou m_i^j est la mesure attribuée à la classe i par le classifieur j .

La figure (figure3.1) montre la taxonomie des différents formalismes de fusion :

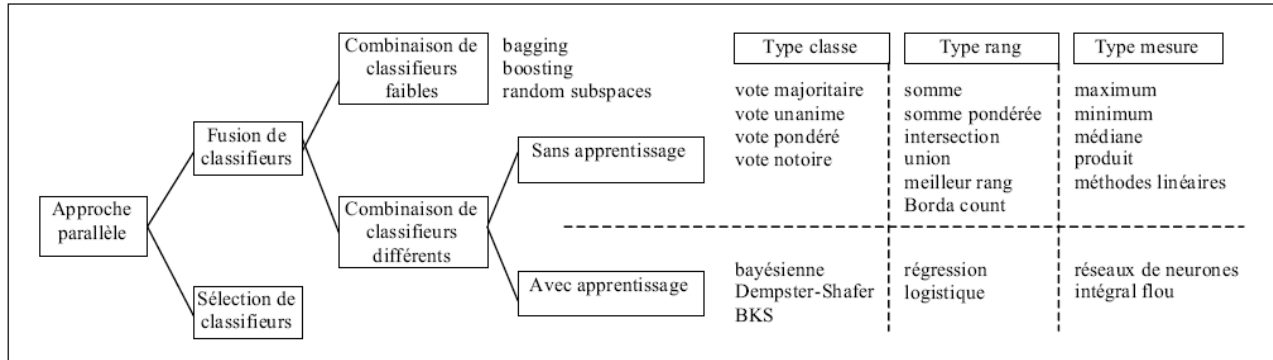


FIGURE 3.1 – Taxonomie des méthodes de combinaison parallèle des classifieurs

3.4 Méthodes sans apprentissage

La méthode à appliquer pour combiner les résultats sans apprentissage sont basés sur le principe de vote. Toutes les méthodes de votes peuvent être dérivés de la règle de majorité avec seuil exprimé par :

$$E(x) = \begin{cases} C_i \text{ si } \sum_i e(i) = \max_{C_i \in \{1, \dots, M\}} \sum_j e(j) \geq \alpha k \\ \text{rejet sinon} \end{cases} \quad (3.2)$$

Où K est le nombre de classifieurs à combiner. Ainsi pour $\alpha = 1$ la classe finale est choisie si tous les classifieurs qui proposent cette réponse sinon la réponse finale est le rejet. Cette méthode restrictive qui accepte le moins de risque possible est appelé majorité unanime. Pour $\alpha = 0.5$, cela signifie que la classe finale est décidée si plus de moitié des classifieurs l'ont proposée. Il s'agit alors du vote à la majorité absolue. Pour $\alpha = 0$ il s'agit de la majorité simple ou le résultat de la combinaison est la classe la plus proposés parmi les K classifieurs. Néanmoins, il y a rejet si toutes les classes ont le même nombre de vote. La combinaison peut aussi être basée soit sur la majorité sans conflit si tous les classifieurs ont choisit la même classe sauf éventuellement ceux qui ont donné une réponse rejet, soit sur le vote pondéré ou la réponse de chaque classifieur est pondérée par un coefficient indiquant son importance dans la combinaison. Dans la forme d'entrée est attribuée à la classe pour laquelle la somme des votes, qui sont pondérés par la fiabilité estimée de chacun des experts, est la plus élevée. Cette méthode n'introduit pas de critère de rejet [29].

3.5 Méthodes avec apprentissage

Contrairement aux méthodes de vote, la théorie de Bayes, la théorie de Dempster-Shafer et la méthode d'espace de connaissance du comportement(BKS) sont des méthodes de combinaison qui utilisent une phase d'apprentissage pour combiner les résultats de type classe. Dans notre cas nous nous intéressons à la théorie de l'évidence et la théorie de Bayes.

3.5.1 Approche bayésienne

Soit un ensemble d'hypothèses mutuellement exclusives H_1, H_2, \dots, H_N . Elles satisfont les conditions suivantes :

$$\begin{cases} \forall i, j, H_i \cap H_j = \emptyset, i \neq j; \\ \bigcup_{i=1}^N H_i = E \end{cases} \quad (3.3)$$

où E représente l'espace des hypothèses (c'est-à-dire l'ensemble des classes de l'image fusionnée). Les hypothèses sont mutuellement exclusives et forment une partition de E . Soit m_1 et m_2 deux primitives caractéristiques issues de deux images différentes représentant le même objet, ou la même hypothèse H_i , la théorie bayésienne nous permet de calculer la probabilité d'obtenir l'hypothèse H_i à partir des deux mesures m_1 et m_2 par l'intermédiaire de la règle de Bayes :

$$P(H_i/m_1, m_2) = \frac{P(H_i) * P(m_1, m_2/H_i)}{\sum_{j=1}^N P(H_j) * P(m_1, m_2/H_j)} \quad (3.4)$$

$P(m_1, m_2/H_i)$ représente la probabilité conjointe d'avoir les deux mesures (m_1, m_2) lorsque l'hypothèse H_i est réalisée, et $P(H_i)$ est la probabilité a priori de l'hypothèse H_i , qui indique la chance d'occurrences de l'hypothèse H_i dans le cas général. Si m_1 et m_2 sont deux variables aléatoires indépendantes, la probabilité conditionnelle $P(m_1, m_2/H_i)$, aussi appelée fonction de vraisemblance, devient une fonction séparable des deux variables m_1 et m_2 .

$$P(m_1, m_2/H_i) = P(m_1/H_i).P(m_2/H_i). \quad (3.5)$$

La condition de séparabilité est appropriée dans le sens où la mesure m_1 provenant d'une acquisition n'a pas d'influence sur la mesure m_2 de l'autre acquisition. L'indépendance statistique est acquise dès lors que les deux mesures n'interfèrent pas l'une avec l'autre. Lorsque les images sont de modalités différentes L'équation 3.4 prend donc la forme suivante :

$$P(H_i/m_1, m_2) = \frac{P(H_i) * P(m_1/H_i) * P(m_2/H_i)}{\sum_{j=1}^N P(H_j) * P(m_1/H_j) * P(m_2/H_j)} \quad (3.6)$$

Ainsi, pour déterminer les probabilités a posteriori $P(H_i/m_1, m_2)$, nous avons besoin de déterminer d'abord les probabilités a priori $P(H_i)$ pour toutes les hypothèses H_i , i allant de 1 à N , et les fonctions de vraisemblance $P(m_j/H_i)$ pour chaque primitive image m_j et pour chaque hypothèse. Une fois la combinaison des probabilités réalisées par l'équation (3.6), nous devons choisir un critère de décision de manière à décider quelle hypothèse H_i doit être sélectionnée en fonction de toutes les probabilités a posteriori. Plusieurs critères sont proposés dans la littérature : le critère le plus couramment employé est celui du maximum a posteriori, (MAP), qui permet de choisir l'hypothèse H_i présentant la probabilité $P(H_i/m_1, m_2)$ la plus importante. D'autres critères ont aussi été utilisés, mais de manière moins courante : il s'agit par exemple du critère de vraisemblance maximale (Maximum of Likelihood). Dans la présente étude, nous choisissons comme critère de choix celui du maximum de vraisemblance [30].

3.5.2 Théorie de l'évidence

Notations et définitions

Soit Θ un ensemble de N hypothèses $\{H_1, H_2, \dots, H_N\}$, pouvant correspondre à différents états d'un système donné. Θ est appelé cadre de discernement. Parmi toutes les hypothèses recensées, on suppose qu'il y en a une et une seule de vérifiée. Toutes les autres sont fausses, il ne peut pas y avoir deux hypothèses vraies en même temps. De plus, la vérité se trouve obligatoirement dans le cadre de discernement. C'est l'hypothèse du monde fermé défini par Smets [31].

Représentation de l'incertain

Shafer [32] a présenté ses travaux comme une généralisation de la théorie de Bayes, au traitement de l'incertain. Il s'appuie sur les travaux de **Dempster**, qui pour chaque probabilité dite « incertaine », définit une probabilité dite supérieure, et une probabilité dite inférieure, définissant ainsi un intervalle de « confiance ». Le principe fondateur de la théorie est la possibilité d'affecter une probabilité basique (bpa : Basic Probability Assignment), non seulement aux différentes hypothèses prises séparément, mais aussi à un ensemble d'hypothèses, appelé hypothèse composée, si nous n'avons de la connaissance que sur cet ensemble, sans pouvoir restreindre notre connaissance à un sous-ensemble plus petit. Pour cela, on définit un jeu de masses m défini sur l'ensemble des sous-ensembles de Θ . Ce jeu de masses est normalisé par la relation : $\sum_{A \subseteq \Theta} m(A) = 1$. Modéliser un jeu de masses consiste à répartir toute notre connaissance disponible (de valeur posée égale à 1 par convention) sur l'ensemble des sous-ensembles de Θ , en fonction de nos observations et de connaissances a priori. Un sous-ensemble qui se voit

affecté d'une masse non nulle est appelé élément focal. Il est par définition impossible, dans le cadre du monde fermé, d'affecter une masse à l'ensemble vide. Cela signifierait que l'on a une portion de croyance sur le fait qu'aucune hypothèse n'est vraie, ce qui est contraire à notre hypothèse de monde fermé disant que Θ contient une hypothèse vraie. La modélisation est très souple, et l'utilisation des hypothèses composées permet d'exprimer de manière fine l'impossibilité à dissocier plusieurs hypothèses. C'est la grande richesse mais aussi la principale difficulté de la méthode, en l'absence de méthode générale pour l'affectation des masses. Nous présentons maintenant des ensembles de jeux de masses particuliers, définissant des formes de connaissances bien précises, de manière à mieux illustrer la finesse de modélisation de cette représentation de la connaissance.

1. Ignorance totale

$$m(\Theta) = 1, m(A) = 0 \text{ et } \forall A \subset \Theta, A \neq \Theta \quad (3.7)$$

Ce jeu de masses est totalement concentré sur l'ensemble de discernement Θ . On n'a pas réussi à séparer la masse de connaissance sur un ensemble plus réduit d'hypothèses, donc on n'arrive pas à discerner entre toutes les hypothèses de Θ . Ce jeu de masses représente donc l'ignorance totale, lorsque le capteur ne donne aucune information sur les différentes hypothèses mises en jeu. Nous verrons qu'il joue un rôle particulier dans le processus de fusion.

2. Certitude

$$\exists H_i \in \Theta / m(H_i) = 1, \forall A \subset \Theta, A \neq H_i, m(A) = 0. \quad (3.8)$$

Ce jeu de masses est totalement concentré sur une seule hypothèse simple, donc exprime la certitude totale de la réalisation de l'hypothèse H_i .

3. Jeu de masse bayésien

Ce jeu de masses est réparti uniquement sur les hypothèses simples. Il est alors équivalent à une probabilité et on se retrouve dans le cas de la théorie bayésienne. On verra dans ce cas que la crédibilité, la masse et la plausibilité d'une hypothèse sont confondues et sont une distribution de probabilité.

4. Jeu de masses consonant

Ce jeu de masses présente la particularité d'avoir tous ses éléments focaux inclus les uns dans les autres. Dubois et Prade ont introduit cette notion dans [33]. Cette modélisation est souvent utilisée pour modéliser la connaissance. Nous pouvons remarquer qu'un jeu de masses consonant est équivalent à une distribution de

possibilité. Ce modèle de jeu de masse est donc souvent utilisé pour permettre la comparaison entre une approche par la théorie de l'évidence et par la théorie des possibilités.

5. Cas intermédiaire

Nous prenons un exemple d'un jeu de masses intermédiaire, afin de montrer comment la répartition des jeux de masses permet de représenter de manière fine la connaissance issue du capteur. $m(H1, H2) = 0.8$, $m(\Theta) = 0.2$. Ce jeu de masses hybride montre la souplesse de modélisation permise par la théorie. Ici, on sait à 80 % que l'une des hypothèses H1 ou H2 est vérifiée, mais on a néanmoins une incertitude totale résiduelle de 20 %.

Notions voisines

Nous avons vu que la modélisation de Dempster-Shafer était très riche. Pour exprimer de manière plus explicite notre connaissance du problème, trois fonctions ont été établies, dérivées de la fonction de masses. Ce sont les fonctions de crédibilité, de plausibilité et de communalité. Ces quatre représentations sont équivalentes, et représentent la même réalité.

1. Crédibilité

La crédibilité (Belief Function) correspond à la quantité d'informations qui est toute entière contenue dans le sous-ensemble considéré. Il contient l'ensemble des connaissances créditant la véracité de ce sous-ensemble. Par définition :

$$Cr(A) = \sum_{B \subset A} m(B). \quad (3.9)$$

2. Plausibilité

La plausibilité d'un sous-ensemble A correspond à la quantité d'information ne discréditant pas A, c'est-à-dire toute l'information contenue dans les sous-ensembles ayant une intersection avec A :

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \quad (3.10)$$

La plausibilité de A est également reliée à la crédibilité du complémentaire de A, et elle correspond à toute l'information ne créditant la véracité du complémentaire de A.

$$Pl(A) = 1 - Cr(\bar{A}). \quad (3.11)$$

On peut représenter ces deux notions (Crédibilité et Plausibilité) par les diagrammes classiques de théorie des ensembles, appelés diagramme de **Venn**. La (figure3.2) explicite les appellations de vraisemblance minimale et maximale dont on affecte parfois la Crédibilité et la Plausibilité. On visualise sur cette figure le fait que la crédibilité regroupe toutes les masses des éléments focaux inclus dans le sous-ensemble, alors que la plausibilité correspond à toute les masses intersectant avec le sous-ensemble considéré.

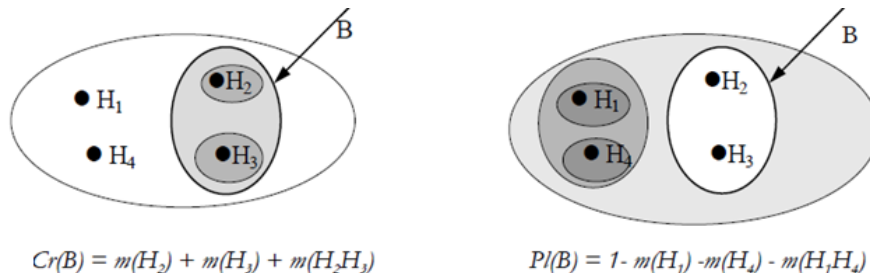


FIGURE 3.2 – Représentation par diagrammes de Venn des notions de crédibilités et de plausibilité

3. Communalité

La communalité d'un sous-ensemble de A est définie par la relation suivante :

$$\vartheta(A) = \sum_{A \subset B} m(B). \quad (3.12)$$

Cela correspond à la quantité d'information de tous les sous-ensembles contenant A. La communalité n'a pas une interprétation physique aussi simple que la crédibilité ou la plausibilité, mais elle est parfois utilisée dans le calcul de combinaison de jeu de masses dans le cadre de la fusion.

3.6 Conclusion

Dans ce chapitre nous avons vu quelques méthodes de fusion avec et sans apprentissage qui nous permettent d'améliorer la qualité des résultats obtenues par les classifieurs en effectuant une complémentarité significative entre eux.

Chapitre 4

Modélisation de la fusion et expérimentations

Préliminaire

Dans la première section de cette partie, nous décrivons la base d'images, nous présentons la construction de la base d'apprentissage en déterminant l'espace couleur hybride. Dans la seconde section nous exposerons quelques modèles de classifications supervisées sur cette base. Et enfin nous appliquerons quelques stratégies de fusion, avec une discussion globale des résultats.

4.1 Description de la base

Notre base d'images a été acquise, au sein du service d'hémobiologie du C.H.U Tlemcen sur des lames avec la coloration de May Grunwald Giemsa (MGG). L'environnement LEICA (caméra et microscope) permet d'obtenir des images couleur RGB de bonne qualité ; au format BMP de dimension 1024x768.

4.2 Construction de la base d'apprentissage

Pour construire une base d'apprentissage, nous avons opté pour les étapes suivantes :

4.2.1 Espaces couleur

Comme il n'existe pas un espace couleur conventionnel qui fournit des résultats prometteur pour la segmentation toutes les images [13] , nous avons proposé de combiner l'ensemble

des espaces couleur en utilisant l'analyse discriminante linéaire pour fournir la meilleure discrimination possible entre les classes avec le moins de redondance d'informations.

4.2.2 Construction de la matrice de passage

L'analyse discriminante linéaire produit une matrice de passage permettant d'obtenir un espace réduit et représentatif des données. Le processus de la construction de la matrice de passage est représenté dans la (figure4.1) suivante :

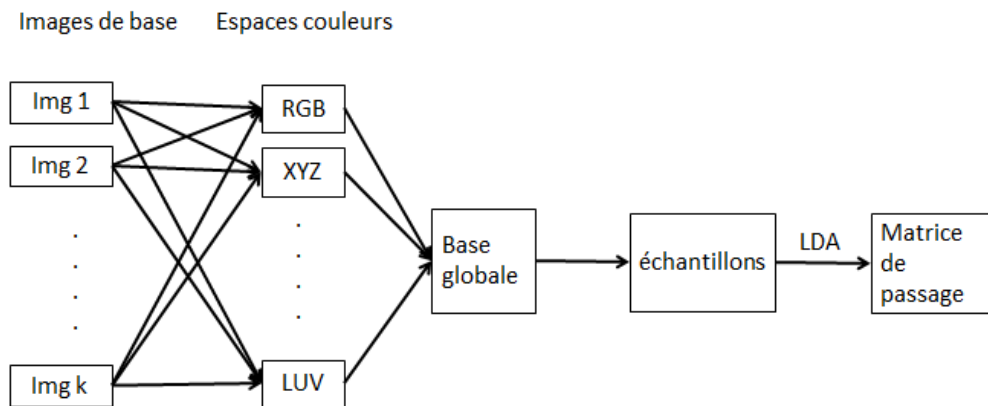


FIGURE 4.1 – Construction de la matrice de passage

4.2.3 Étiquetage

Cette étape est réalisée afin d'obtenir un label pour chaque pixel. Ceci est effectué à l'aide d'une image étiquetée (image vérité terrain) (figure4.2) associé à chaque image.

- C1 : Le noyau prend la couleur verte
- C2 : le cytoplasme, couleur jaune
- C3 : Les globules rouges avec la couleur rouge
- C4 : Et le fond avec la couleur noire.

Pour un bon apprentissage nous avons sélectionné visuellement les images les plus représentatives pour décrire la variabilité des classes.

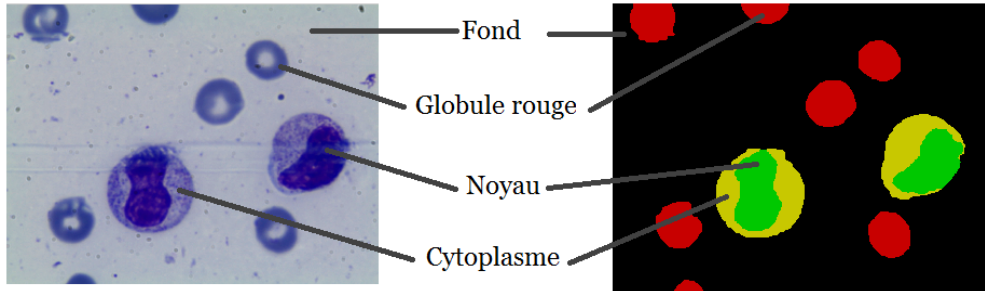


FIGURE 4.2 – Coloration

4.2.4 Échantillonnage

Dans le but de fournir un ensemble d'apprentissage à tout les classifieurs, nous avons sélectionné des échantillons épars de chaque image et de chaque classe (figure4.3). Ainsi nous caractérisons chaque pixel $w(i,j)$ par ses attributs (obtenus après la projection de la matrice de passage) et sa classe d'appartenance C_i .

$$X = \begin{bmatrix} x_{11} & \dots & x_{1,w} & \dots & x_{i,j} & \dots & C_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{21} & \dots & x_{2,w} & \dots & x_{i,j} & \dots & C_2 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{31} & \dots & x_{3,w} & \dots & x_{i,j} & \dots & C_3 \end{bmatrix}$$

Les lignes de la matrice X correspondent aux pixels, tandis que les colonnes correspondent aux composantes couleur de chaque pixel.

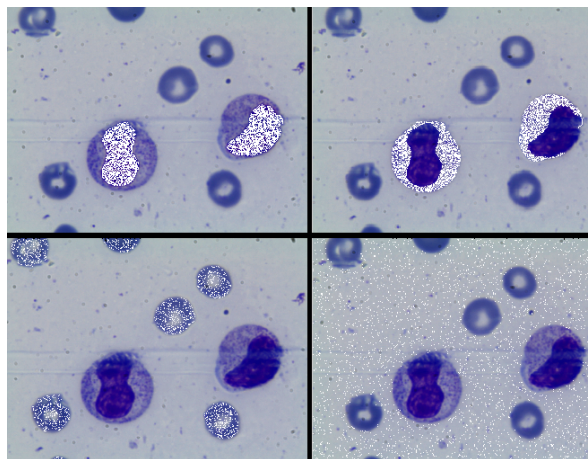


FIGURE 4.3 – Échantillonnage

4.2.5 Base d'apprentissage

Après La construction de la matrice de passage, chaque image sera projeté pour garder le maximum d'informations avec une dimension réduite (figure4.4).

Un prétraitement a été réalisé sur la base d'apprentissage qui consiste à supprimer des pixels conflictuels (même données avec des classes différentes)

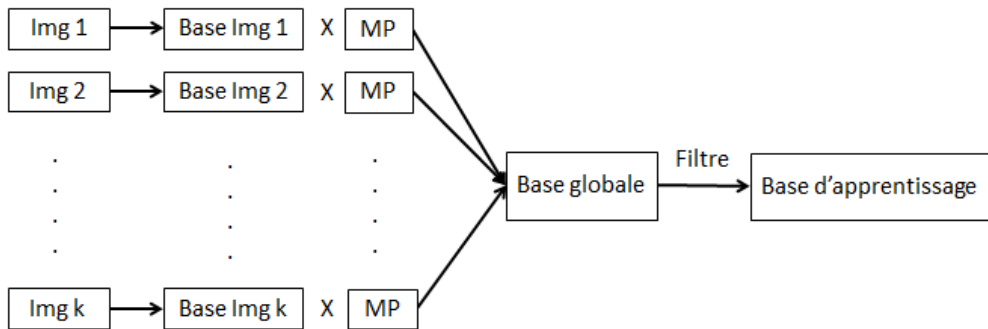


FIGURE 4.4 – Construction de la base d'apprentissage

La projection 3D de notre base d'apprentissage (figure4.5) montre qu'il est difficile de trouver des séparations explicites entre les classes.

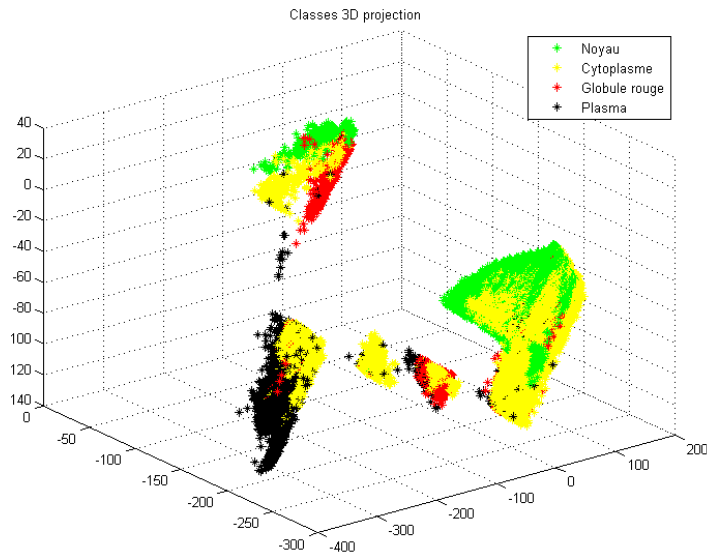


FIGURE 4.5 – Projection 3D de la base d'apprentissage

4.3 Classifications

Dans cette seconde section, nous appliquons cinq méthodes de classification supervisées sur notre base. L'apprentissage de chaque classifieur est optimisé avec ses propres techniques pour fixer leurs paramètres.

- Le réseau de neurones multicouches, l'optimisation est effectuée à l'aide de la validation croisée produisant les paramètres suivants : Il est composé de trois couches contenant respectivement (6, 10, 1) neurones avec un pas d'apprentissage de 0.05, en utilisant l'algorithme de rétro-propagation Levenberg-Marquardt (Trainlm en matlab).
- Pour le classifieur GMM, plusieurs gaussiennes modélisent chaque classe : cinq pour le noyau neuf pour le cytoplasme douze pour les globules rouge et quinze pour le plasma.
- Concernent le KppV, le nombre de voisinage a été fixé à trois avec l'utilisation de la distance euclidienne.
- Pour le SVM, le paramètre de régularisation C est fixé à 512, avec l'utilisation de la fonction à base radiale (RBF), et Gamma avec 0.0001.
- En fin, On a choisie l'algorithme **CART** (*Classification And Regression Trees*) pour l'arbre de décision.

4.4 Fusions

Toutes les classifications obtenues n'ont pas fournit le même résultat sur la classe à attribuer, nous obtenons des zones de contradiction. L'étape de fusion résout le problème d'incohérence des pixels. Dans ce travail, nous avons appliqués deux formalismes :

- Sans apprentissage : Vote majoritaire et vote majoritaire pondéré.
- Avec apprentissage : Théorie de l'évidence, approche bayésienne.

1. **Fusion par vote majoritaire et vote majoritaire pondéré**

Pour le vote majoritaire on affecte chaque pixel à la classe choisit par la majorité des classifieurs.

Pour le vote majoritaire pondéré on a associé des poids à chaque classifieur i pour tout classe j selon leur performances par classe.

2. **Fusion par théorie de l'évidence**

On a appliqué trois modélisations en se basant sur deux hypothèses et deux règles de sommation :

(a) **Hypothèses :**

L'hypothèse1 : consiste à attribuer la classe et son complément dans l'espace de

discernement $H(i)$ et $H(j_1, j_2, j_3)$.

L'hypothèse2 : Chaque sortie avec les autres une à une $H(i)$, $H(i, j_1)$, $H(i, j_2)$ et $H(i, j_3)$. Avec i la sortie du classifieur et j_1, j_2 et j_3 (les autres) $\neq i$.

(b) **Règle de sommation :**

- Associativement (règle1) : combiner les évidences des sorties des classifieurs par l'opérateur de la "somme orthogonale"
- Regroupement (règle2) : répartir les 5 classifieurs en N_g groupes 3 selon leurs réponses. Les classifieurs de chaque groupe (ont tous la même réponse) seront fusionnés et une ultime fusion est réalisée entre les N_g groupes.

Les modélisations effectués sont :

- **H1R1** : fonctionne avec l'hypothèse1 et la règle1.
- **H1R2** : fonctionne avec l'hypothèse1 et la règle2.

Les valeurs des fonctions de masses pour les deux précédentes méthodes ont été prises à partir de la matrice de confusion des classifieurs.

- **H2R1** : fonctionne avec l'hypothèse2 et la règle1 .

4.5 Résultats et évaluations

Sur l'ensemble d'apprentissage, nous avons pu obtenir les taux de reconnaissance suivants (tab4.1)

	noyau	cyto	rouge	fond
RNMC	97,89%	96,97%	98,14%	99,46%
Tree	99,74%	99,50%	99,60%	99,84%
KNN	96,98%	90,16%	92,07%	97,61%
Svm	98,12%	97,57%	98,49%	99,47%
Gmm	97,38%	95,50%	97,35%	98,23%

TABLE 4.1 – Taux d'apprentissage

Pour la phase test , nous avons passé les 27 images sur l'ensemble des classifieurs , les taux de reconnaissances et les précisions de chaque classe par rapport à chaque classifieur sont donnés dans le tableau suivant (tab4.2) :

Précision Tree				Taux Tree			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
80,09%	58,11%	78,5%	98,73%	99,01%	96,52%	94,98%	94,35%
Précision Rnmc				Taux Rnmc			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
88,59%	57,76%	77,32%	99%	99,24%	96,63%	95,24%	95,33%
Précision Gmm				Taux Gmm			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
82,90%	64,4%	82,33%	99,17%	98,93%	97,18%	96,50%	96,17%
Précision Knn				Taux Knn			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
93,85%	73,32%	88,01%	95,29%	99,01%	96,93%	96,93%	95,77%
Précision Svm				Taux Svm			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
92,39%	76,61%	78,91%	95,97%	98,16%	96,94%	96,00%	96,00%

TABLE 4.2 – Résultats obtenus par les classifieurs

Le tableau suivant (tab4.3) décrit les résultats de la fusion par votes (majoritaire et majoritaire pondéré) :

Précision vote majoritaire				Taux vote majoritaire			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
93.42%	78.64%	88.16%	95.72%	99.32%	97.22%	96.83%	96.15%
Précision vote majoritaire pondéré				Taux vote majoritaire pondéré			
noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
92.64%	79.78%	85.22%	96.06%	99.26%	97.22%	96.72%	96.08%

TABLE 4.3 – Résultats obtenus par les votes

Le tableau (tab4.4) suivant décrit les résultats de la fusion par théorie de l'évidence et par la théorie bayésienne :

	Précisions de l'évidence				Taux de l'évidence			
	noyau	cyto	rouge	fond	noyau	cyto	rouge	fond
EvidenceH1R1	93,51%	75,87%	76,34%	96%	99,14%	96,87%	94,98%	94,85%
EvidenceH1R2	86,99%	59,54%	72,53%	93,09%	99,48%	97,25%	95,60%	94,40%
EvidenceH2R1	91,59%	77,67%	83,87%	94,73%	99,24%	96,63%	95,24%	95,33%
bayésienne	93,64%	72,97%	84,71%	94,82%	99,13%	96,5%	95,49%	95,20%

TABLE 4.4 – Résultats obtenus par l'évidence et bayésienne

4.5.1 Discussion des résultats

Sur la totalité des images testées et sur toutes les segmentations réalisées, nous avons constaté que : les résultats obtenus pour la classe "Noyau" et la classe "Fond" ont été satisfaisants en terme de taux de reconnaissances et de précisions. Et cela est due à la légère différence entre les configurations colorimétriques de ces deux classes pour toutes les images ; La classe "Cytoplasme" a été la classe la plus faiblement reconnue (des pourcentages aux alentours des 70%). Si nous regardons l'ensemble des images alors l'écart enregistré des configurations colorimétriques est grand (figure4.6).

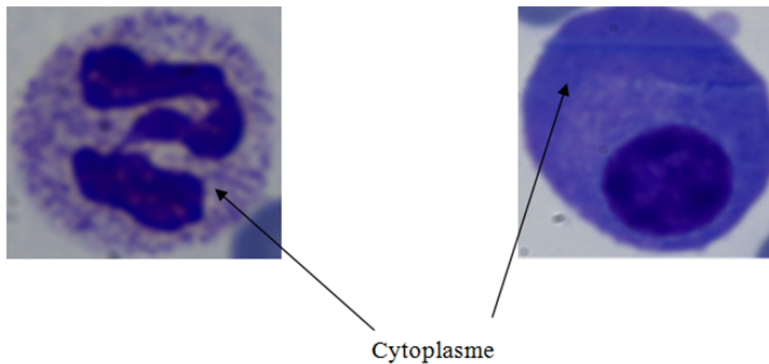


FIGURE 4.6 – Exemple d'écart

En ce qui concerne les zones d'intérêt (noyau, cytoplasme) nous considérons que notre objectif a été atteint puisque une remarquable amélioration a été constaté dans les régions d'intérêt, les taux et précisions des différentes fusions sont supérieur à la moyenne des taux et précisions de l'ensemble des classifieurs. Même pour la classe globule rouge une nette amélioration a été constaté avec les fusions par vote majoritaire, vote majoritaire pondéré, evidenceH2R1 . Pour la classe fond il y'a une amélioration avec la fusion par vote majoritaire et vote majoritaire pondéré (figure4.8) (figure4.7) .

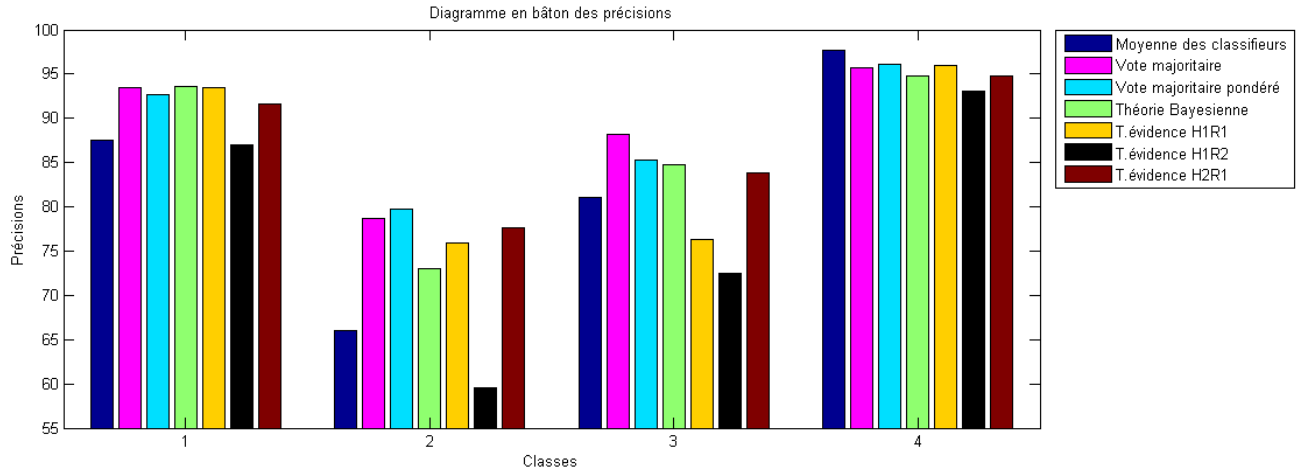


FIGURE 4.7 – Diagramme en bâtons des précisions

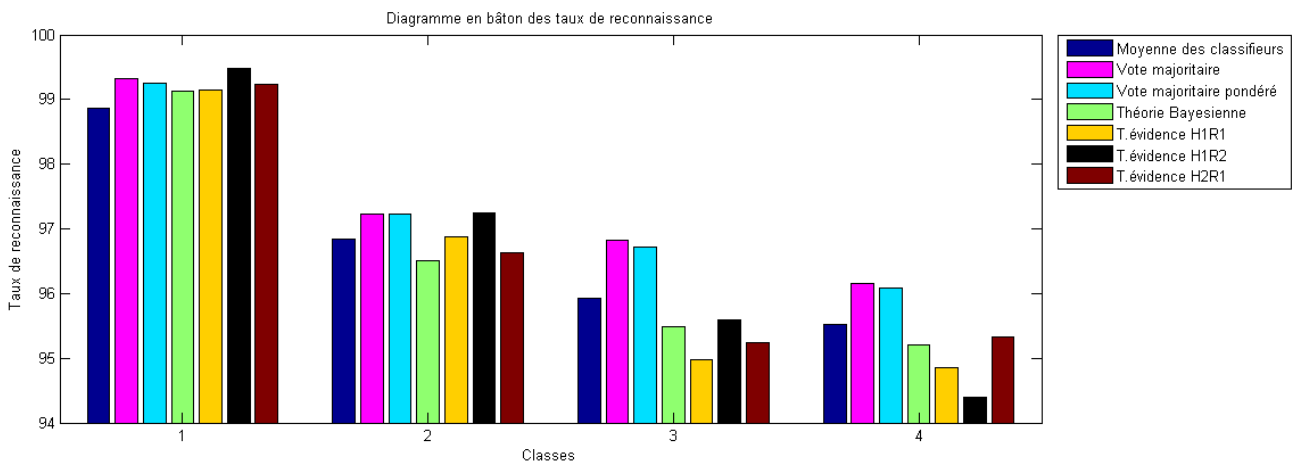


FIGURE 4.8 – Diagramme en bâtons des taux de reconnaissances

En général, nous notons que les différents types de fusion apportent un progrès en terme de taux de reconnaissance et de précision, particulièrement dans les régions d'intérêt (noyau et cytoplasme).

Enfin nous déduisons que la combinaison des classifieurs est une méthode permettant d'améliorer la qualité et la robustesse des systèmes de reconnaissance par rapport à un classifieur seul, dont voici quelques images de test :

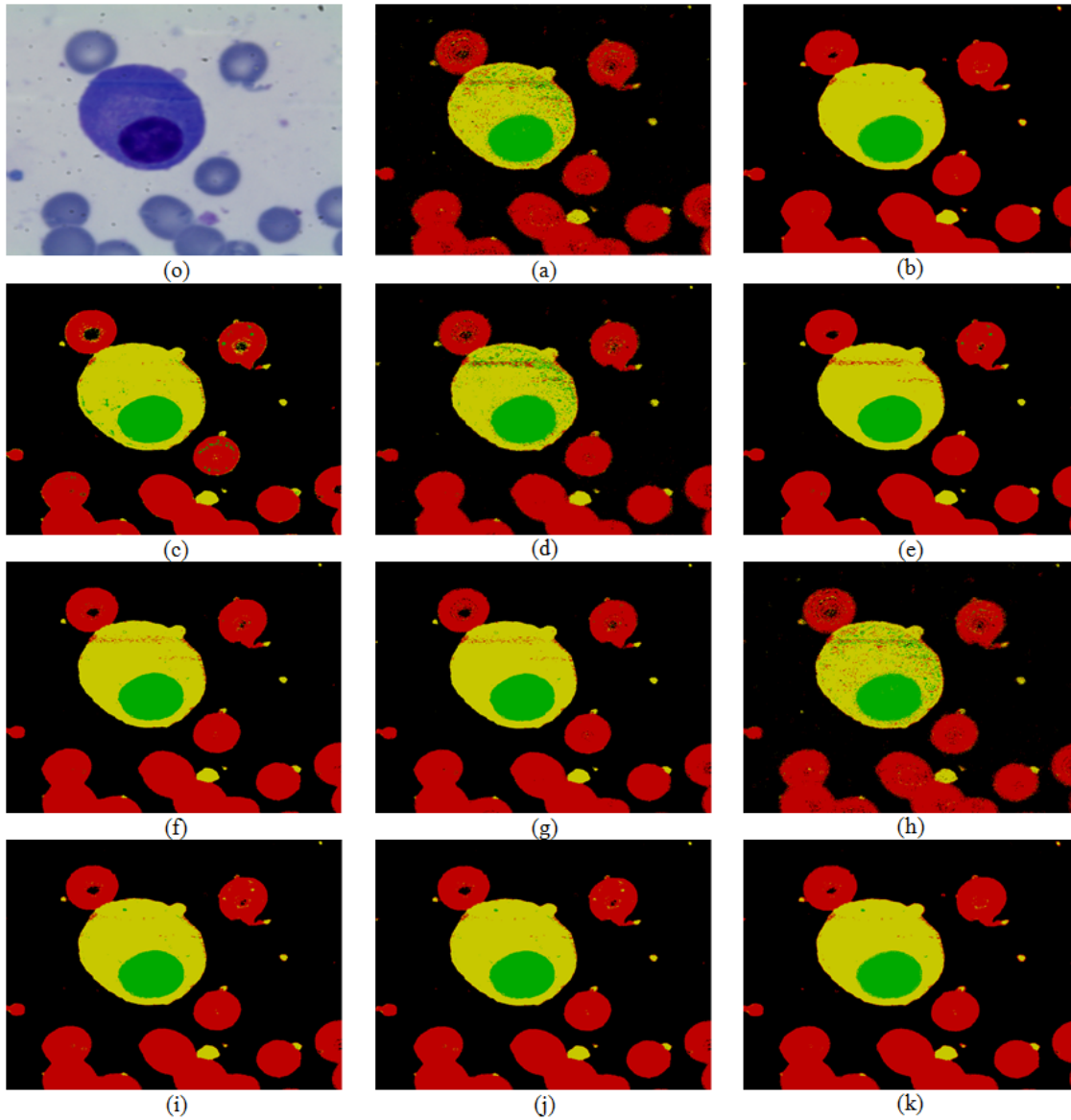


FIGURE 4.9 - image originale(o), résultat arbre de décision(a), résultat RNMC(b),résultat GMM(c) résultat Knn(d), résultat Svm(e),résultat vote maj(f),résultat vote maj pondéré(g), résultat bayésienne(h), résultat T.E H1 R1(i),résultat T.E H1 R2(j), résultat T.E H2 R1(k)

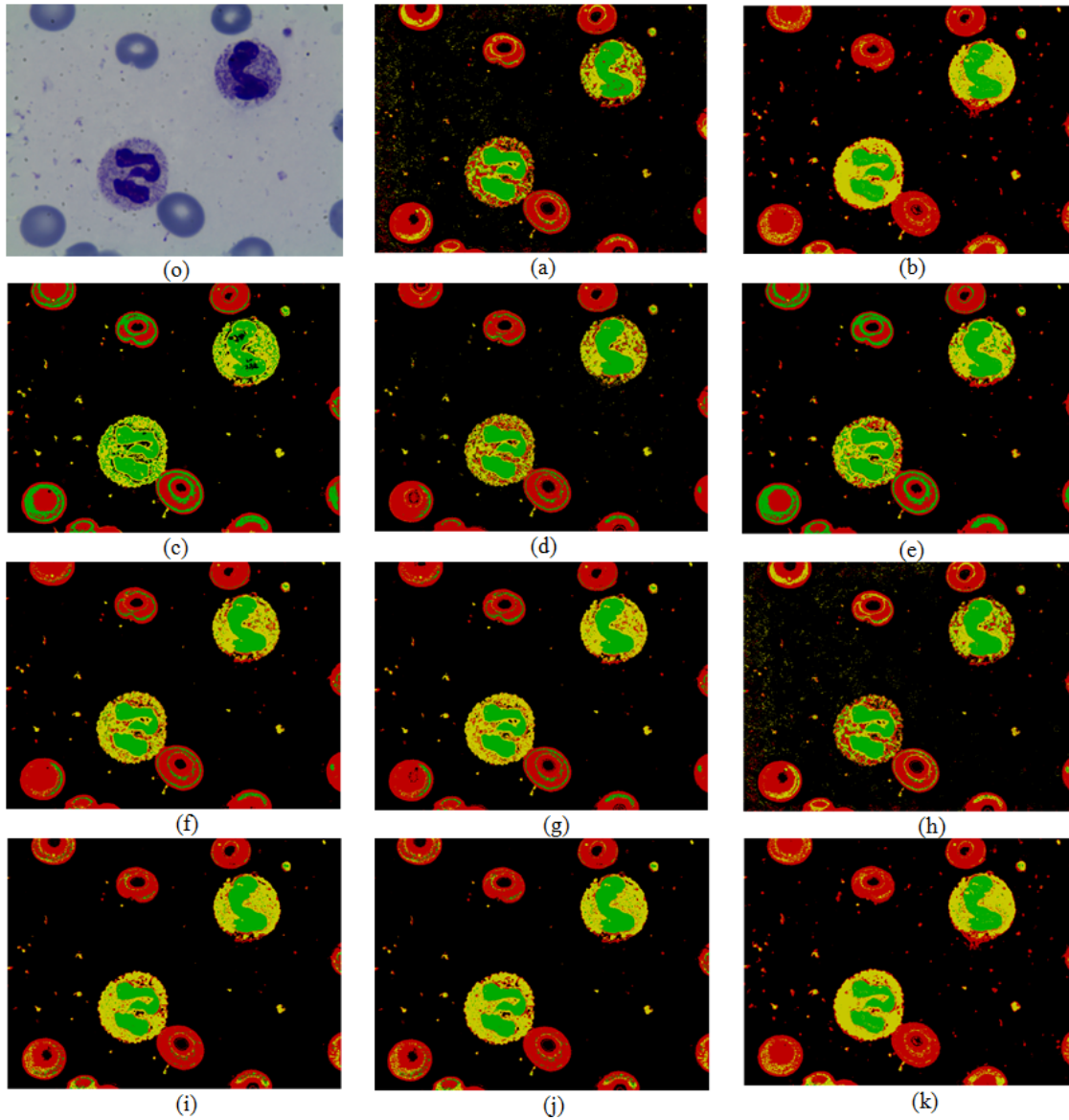


FIGURE 4.10 - image originale(o), résultat arbre de décision(a), résultat RNMC(b),résultat GMM(c) résultat Knn(d), résultat Svm(e),résultat vote maj(f),résultat vote maj pondéré(g), résultat bayésienne(h), résultat T.E H1 R1(i),résultat T.E H1 R2(j), résultat T.E H2 R1(k)

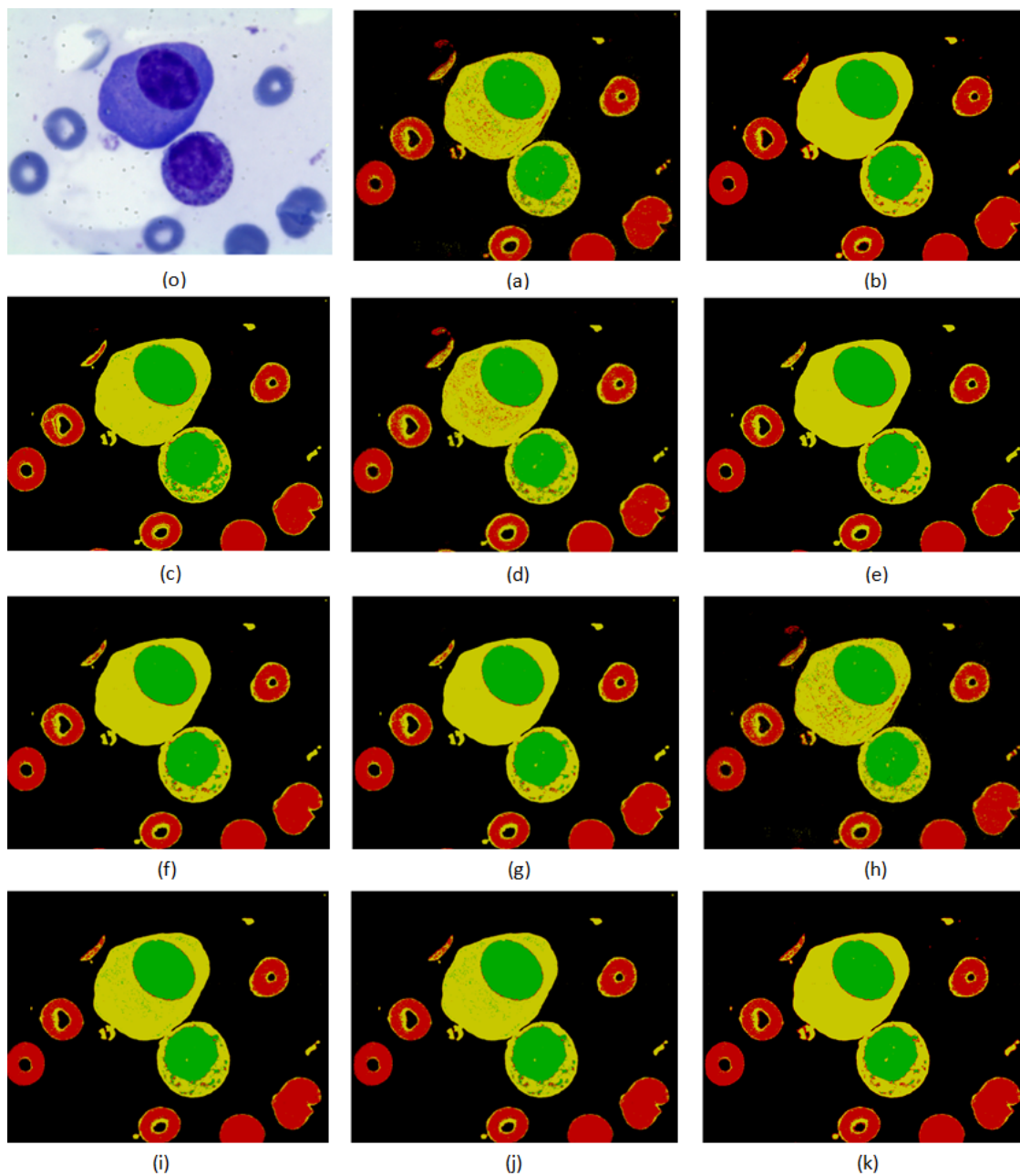


FIGURE 4.11 - image originale(o), résultat arbre de décision(a), résultat RNMC(b),résultat GMM(c) résultat Knn(d), résultat Svm(e),résultat vote maj(f),résultat vote maj pondéré(g), résultat bayésienne(h), résultat T.E H1 R1(i),résultat T.E H1 R2(j), résultat T.E H2 R1(k)

Une comparaison a été réalisé avec le travail de **I.BAGHLI** [34], puisque son travail s'effectue sur la même base d'images. Nos contributions majeure par rapport aux travaux précédents se caractérisent par :

- L'utilisation de quatorze espaces au lieu de cinq espaces pour plus d'informations.
- L'utilisation de cinq classifieurs au lieu d'un seul classifieur (SVM).
- Réduction de la dimension et élimination de la redondance basant sur la méthode LDA.
- Résultat prometteur sans effectuer un post-traitement.

4.5.2 Présentation de l'interface graphique

Les captures ci-dessous (figure4.12) sont issues de notre application.

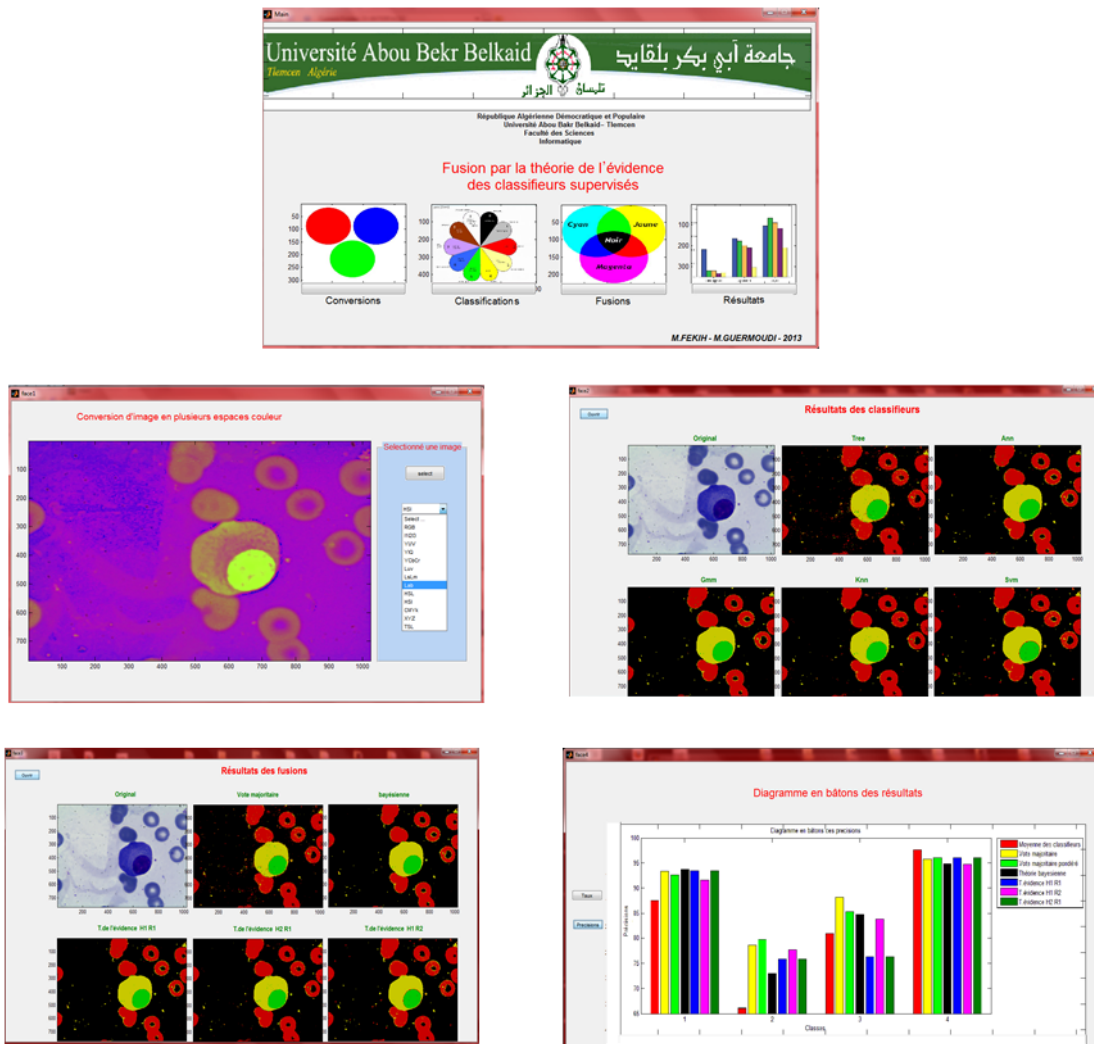


FIGURE 4.12 – Captures de l'interface graphique

Conclusion générale

Le problème traité dans ce travail concerne l'amélioration de la classification d'image dans les conditions d'insuffisance d'informations à priori, déterministes et fiables sur l'état et la nature de la formation de l'image à l'instant de prise de vue.

Notre expérience montre que la combinaison de classifieurs améliore nettement les performances du système de reconnaissance par rapport à chacun des classifieurs pris isolément, ainsi que le choix de classifieurs est important pour aboutir à une meilleure performance avec le minimum de classifieurs.

Une méthode de combinaison de classifieurs d'imagerie pour une meilleure classification des scènes imagées a été proposée. La théorie de l'évidence est basée sur le principe de combinaison parallèle des systèmes répartis. Elle est générale et applicable à tout type de classifieurs.

Perspectives

Nous proposons comme perspectives :

- Ajouter d'autres espaces pour augmenter la quantité d'informations.
- Autres méthodes de réduction de dimension (ICA, MMD ...).
- Appliquer des post traitements sur les résultats obtenus pour améliorer la qualité d'image.
- Introduire la notions de voisinage.
- Explorer d'autres modèles de fusion.

Bibliographie

- [1] V. Lozano. Contribution de l'analyse d'images couleur au traitement des images textile. *Thèse de doctorat*, janvier 1998.
- [2] D. B. Judd. Réduction of data on mixture of color stimuli. *Bureau of Standards Journal Research*, pages 515–548, 1930.
- [3] D. B. Judd. A maxwell triangle yielding uniform chromaticity scales. *Journal of the Optical Society of America*, pages 24–35, 1935.
- [4] D. Eastwood. *A simple modification to improve the visual uniformity of the CIE 1964 U*V*W* colour space*. In *Second congress of the Association Internationale dela Couleur, Colour'73*, pages 293–296, 1973.
- [5] Bureau central de la CIE. Commission internationale de l'éclairage. *Technical Report 15.2*, 1986.
- [6] L. Liang. Contribution à la segmentation chromatique des images. application à l'identification d'objets colorés. *Thèse de doctorat, Institut National Polytechnique de Toulouse*, juillet 1987.
- [7] P. Pujas. Analyse d'images couleur et fusion d'images 3d et couleur. *Thèse de doctorat, Université de Montpellier II Sciences et Techniques du Languedoc*, février 1996.
- [8] T. Carron. Segmentations d'images couleur dans la base teinte-luminance-saturatuion : approche numérique et symbolique. *Thèse de doctorat, Université de Savoie*, décembre 1995.
- [9] T.Y. Shih. The reversibility of six geometric color spaces. *Photogrammetric Engineering and Remote Sensing*, pages 1223–1232, 1995.
- [10] S. Galichet. Etude d'un système d'incrustation d'images vidéo couleur. *Thèse de doctorat, Université de Technologies de Compiègne*, décembre 1989.
- [11] et T. Sakai Y. I. Ohta, T. Kanade. Color information for région segmentation. *Computer Graphics and Image Processing*, pages 222–241, 1980.

- [12] R.D.Dony. Karhunen-Loève transform. *University of Guelph*, 2011.
- [13] L. Macaire ;J. G. Postaire N. Vandenbroucke. Color pixels classification in an hybrid color space. *Dans Proc. of the IEEE Int. Conf. on Image Processing, volume 1*, pages 176–180, 1998.
- [14] Sabrina Tollari. Indexation et recherche d’images par fusion d’informations textuelles et visuelles. *Thèse préparée dans le laboratoire LSIS - UMR CNRS 6168*, octobre 2006.
- [15] A. Dromigny Badin. Fusion d’images par la théorie de l’évidence en vue d’applications médicales et industrielles. ., page 12, 1998.
- [16] A. Cornuéjols Laurent Miclet Avec la participation d’Yves Kodratoff. Apprentissage artificiel. ., pages 7–10, Deuxième tirage 2003.
- [17] Cortes et Vapnik. Support vector networks. machine learning. ., 1995.
- [18] L. Mercier. Les machines à vecteurs support pour la classification en imagerie hyperspectrale : implémentation et mise en oeuvre. ., fevrier 2010.
- [19] M. LaBarre. Résumé : Réseaux de neurones. ., Mai 2002.
- [20] L. Crowley. Systèmes intelligents : Raisonnement et reconnaissance. ., avril 2004.
- [21] S. Caron. Une introduction aux arbres de décision. .
- [22] A. Porebski. Discipline automatique et informatique industrielle. ., page 110, Novembre 2009.
- [23] S. Mathevet L.Trassoudaine P.Checchin J.Alizoni. Application de la théorie de l’évidence à la combinaison de segmentations en régions. ., page 1, Septembre 1999.
- [24] L.N. Kanal. On pattern, categories, and alternate realities. pattern recognition. ., 1993.
- [25] F. Kimura and Shridhar. Handwritten numerical recognition based on multiple algorithms. pattern recognition. ., 1991.
- [26] S. Chitroub. Combinaison de classifieurs :une approche pour l’amélioration de la classification d’images multisources/multidates de télédétection. *Université des Sciences et Technologie Houari Boumediene*, page 3, 2004.
- [27] GDR. Synthèse sur les problèmes de fusion. ., 1991.
- [28] A.Appriou. Perspectives liées à la fusion de données, science et défense. ., mai 1990.
- [29] H.Zouari et al. un panorama de methode de combinaison de classifieur en reconnaissance de forme. *Laboratoire PSI, UFR des sciences*, pages 2–3, 2002.
- [30] Anne Dromigny Badin. Fusion d’images par la théorie de l’évidence en vue d’applications médicales et industrielles. ., 1998.

- [31] F.Sureda I.Bloch C.Pellot A.Herment. Reconstruction 2d de vaisseaux sanguins par fusion de données à partir dimages angiographiques et échographiques, traitement du signal. ., 1994.
- [32] Shafer. G. Mathematical theory of evidence. *Princeton University Press*, page 297, 1976.
- [33] Dubois D. Prade H. Consonant approximation of belief functions. ., 1990.
- [34] I.BAGHLI. Fusion de segmentation d'images cytologiques par svm dans différents espaces couleur. *Mémoire pour l'obtention du diplôme de Master en Informatique*, Juillet 2011.

Annexe A

Nous rappelons ici les changements de système de représentation de la couleur, La relation matricielle utilisant une matrice de passage P.

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} = P \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

A.1 Les systèmes de primaires

Transformation d'un système (R, G, B) vers un autre système (R,G,B)

Les transformations suivantes sont données par Carron [8]. - Transformation du système (Rc, Gc, Bc) vers le système (Rf, Gf, Bf) :

$$P = \begin{bmatrix} 0.842 & 0.156 & 0.091 \\ -0.129 & 1.319 & -0.203 \\ 0.007 & -0.069 & 0.897 \end{bmatrix}$$

A.1.1 Transformation d'un système (R, G, B) vers le système (X, Y, Z) de la CIE

Nous citons ici les cas les plus fréquemment rencontré dans la littérature,

$$P = \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5943 \end{bmatrix} = 5.6508 \times \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.812 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix}$$

A.2 Les systèmes luminance-chrominance

Lorsque les transformations correspondant à un même standard sont légèrement différentes de celles que nous proposons en équation, nous indiquerons les coefficients différents, dans des notes de bas de page se rapportant aux auteurs concernés.

A.2.1 Transformation d'un système (RE, GE, BE) vers le système (Y, U, V)

$$P = \begin{bmatrix} 0.364 & 0.404 & 0.074 \\ 0.705 & -0.331 & -0.503 \\ 0.151 & -0.446 & 0.303 \end{bmatrix}$$

Transformation du système (RE, GE, BE) vers le système (Y, U, V) :

$$P = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.148 & -0.289 & 0.437 \\ 0.615 & -0.515 & -0.100 \end{bmatrix}$$

A.2.2 Transformation du système (R, G, B) vers le système (Y, Ch1, Ch2) de Carron

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{-\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Par ailleurs, Carron [8] met en évidence qu'il est possible de passer d'un système (R, G, B) à un système luminance-chrominance particulier à l'aide d'une matrice P définie par

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix}$$

A partir du système ainsi défini, il est possible de déterminer un système perceptuel (L, C, H) identique au système perceptuel (I, S, T) correspondant au modèle triangulaire.

A.2.3 Transformation du système (R, G, B) vers le système (I1, I2, I3)

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{-1}{2} \\ \frac{-1}{4} & \frac{1}{2} & \frac{-1}{4} \end{bmatrix}$$

Résumé

L'analyse d'images dans le domaine de la cytologie est un outil de diagnostic en voie de développement. L'analyse quantitative de la forme et de la structure des cellules issue d'images couleur de microscopie photonique apporte au cytotechnologiste des informations précieuses pour l'aide au diagnostic. Cette étude de paramètres ne peut se faire qu'à partir d'objets parfaitement segmentés. Dans le cadre de la coloration internationale de MGG, nous proposons une segmentation par la classification pixellaire par fusion de classifieurs supervisés pour l'augmentation de la qualité des résultats. En effet, de plus en plus, les chercheurs se rendent compte qu'il n'existe aucune méthode de classification pouvant manifester une supériorité sur les autres méthodes dans tous les problèmes et toutes les situations. Cette combinaison nous permet d'obtenir des résultats prometteurs.

Mots clés : Cytologie, Diagnostique, Image microscopique, Segmentation, Classification pixellaire, Fusion.

Abstract

Image analysis in the field of cytology is a diagnostic tool developing. The quantitative analysis of the shape and structure of cells derived from color images of light microscopy provides the cytotechnologist valuable information for diagnostic assistance. This study parameters can only be made from perfectly segmented objects. Within the international MGG staining, we propose a segmentation by pixellaire classification fusion supervised classifiers to increase the quality of results. Indeed, increasingly, researchers are realizing that there is no classification method can demonstrate superiority over other methods in all problems and situations. This combination allows us to obtain promising results.

Keywords : Cytology, Diagnostic, Microscopic image, Segmentation, Pixellaire Classification, Fusion.

ملخص
تحليل الصور في مجال علم الخلايا هو أداة تشخيصية في طريق التطور. التحليل الكمي لشكل و بنية الخلايا الملتقطة عن طريق الصورة اللونية المجهرية التي تسهم في إفادة تقني السيتوبلازم بمعلومات قيمة للتشخيص. دراسة هذه المعالم لا تتم إلا بالتقسيم الجيد للمواد. في إطار التلوين الدولي MGG نقترح التقسيم بواسطة تصنيف البكسل بدمج المصنفات الموجهة من أجل زيادة جودة النتائج .
زيادة على ذلك, انتبه العلماء إلى عدم وجود أي منهج للتصنيف يمكنه التفوق على كل المناهج الأخرى في حل كل المشاكل و جميع الوضعيات, إلا أنه بواسطة دمج المصنفات يمكننا الحصول على نتائج واعدة.

الكلمات الرئيسية : تصنيف البكسل, المجهرية التقسيم, تشخيص الصورة, علم الخلايا, الدمج